Asen L. Dontchev and R. Tyrrell Rockafellar

Implicit Functions and Solution Mappings

A View From Variational Analysis

Second Edition

Springer

"Never take the implicit function theorem for granted."

attributed to Carl B. Allendoerfer

Preface to the Second Edition

The preparation of this second edition of our book was triggered by a rush of fresh developments leading to many interesting results. The text has significantly been enlarged by this important new material, but it has also been expanded with coverage of older material, complementary to the results in the the first edition and allowing them to be further extended. We hope in this way to have provided a more comprehensive picture of our subject, from classical to most recent.

Chapter 1 has a new preamble which better explains our approach to the implicit function paradigm for solution mappings of equations, variational problems, and beyond. In the new Section 1H, we present an implicit function theorem for functions that are merely continuous but on the other hand are monotone.

Substantial additions start appearing in Chapter 4, where generalized differentiation is brought in. The coderivative criterion for metric regularity has now a proof in 4C. Section 4D has been reconstituted to follow up with the strict derivative condition for metric regularity and immediately go on to the inverse function theorems of Clarke and Kummer, and an inverse function theorem for nonsmooth generalized equations whose proof is postponed until Chapter 6. In that way, all basic regularity properties are fully supplied with criteria involving generalized derivatives.

Chapter 5, dealing with infinite-dimensional branches of the theory, has been augmented by much more. Section 5G presents parametric inverse function theorems which are later put to use in Chapter 6. Section 5H translates the result to nonlinear metric spaces and furnishes a new proof to the (extended) Lyusternik–Graves theorem. Section 5I links the Lyusternik–Graves theorem fixed point theorems and other results in set-valued analysis. The final section 5K deploys an inverse function theorem in Banach spaces which relies only on selections of the inverse to the directional derivative.

The biggest additions of new material, however, are in Chapter 6. Section 6C has been reworked so that it now provides an easy-to-follow introduction to iterative methods for solving generalized equations. The new Section 6D shows that the paradigm of the Lyusternik–Graves/implicit function theorem extends to mappings whose values are sets of sequences generated by iterative methods. Section 6E focuses on inexact Newton-type methods, while Section 6F deals with a nonsmooth

Newton's method. Section 6G studies the convergence of a path-following method for solving variational inequalities. Section 6I features metric regularity as a tool for obtaining error estimates for approximations in optimal control.

Some of the chapters and some the sections have new titles. The sections are numbered in a new way to facilitate the electronic edition of the book and simultaneously to connect to the labeling used in the first edition; for example, Section 1A in the first edition now becomes Section 1.1 [1A]. A list of statement is added and the notation is moved to the frontmatter. Typos were corrected and new figures and exercises have been added. Also, the list of references has been updated and the index extended.

Ann Arbor and Seattle *Asen L. Dontchev*
March 2014 *R. Tyrrell Rockafellar*

Preface to the First Edition

Setting up equations and solving them has long been so important that, in popular imagination, it has virtually come to describe what mathematical analysis and its applications are all about. A central issue in the subject is whether the solution to an equation involving parameters may be viewed as a function of those parameters, and if so, what properties that function might have. This is addressed by the classical theory of implicit functions, which began with single real variables and progressed through multiple variables to equations in infinite dimensions, such as equations associated with integral and differential operators.

A major aim of the book is to lay out that celebrated theory in a broader way than usual, bringing to light many of its lesser known variants, for instance where standard assumptions of differentiability are relaxed. However, another major aim is to explain how the same constellation of ideas, when articulated in a suitably expanded framework, can deal successfully with many other problems than just solving equations.

These days, forms of modeling have evolved beyond equations, in terms, for example, of problems of minimizing or maximizing functions subject to constraints which may include systems of inequalities. The question comes up of whether the solution to such a problem may be expressed as a function of the problem's parameters, but differentiability no longer reigns. A function implicitly obtainable this manner may only have one-sided derivatives of some sort, or merely exhibit Lipschitz continuity or something weaker. Mathematical models resting on equations are replaced by "variational inequality" models, which are further subsumed by "generalized equation" models.

The key concept for working at this level of generality, but with advantages even in the context of equations, is that of the set-valued *solution mapping* which assigns to each instance of the parameter element in the model *all* the corresponding solutions, if any. The central question is whether a solution mapping can be localized graphically in order to achieve single-valuedness and in that sense produce a function, the desired *implicit function*.

In modern variational analysis, set-valued mappings are an accepted workhorse in problem formulation and analysis, and many tools have been developed for handling them. There are helpful extensions of continuity, differentiability, and regularity of several types, together with powerful results about how they can be applied. A corresponding further aim of this book is to bring such ideas to wider attention by demonstrating their aptness for the fundamental topic at hand.

In line with classical themes, we concentrate primarily on local properties of solution mappings that can be captured metrically, rather than on results derived from topological considerations or involving exotic spaces. In particular, we only briefly discuss the Nash–Moser inverse function theorem. We keep to finite dimensions in Chapters 1 to 4, but in Chapters 5 and 6 provide bridges to infinite dimensions. Global implicit function theorems, including the classical Hadamard theorem, are not discussed in the book.

In Chapter 1 we consider the implicit function paradigm in the classical case of the solution mapping associated with a parameterized equation. We give two proofs of the classical inverse function theorem and then derive two equivalent forms of it: the implicit function theorem and the correction function theorem. Then we gradually relax the differentiability assumption in various ways and even completely exit from it, relying instead on the Lipschitz continuity. We also discuss situations in which an implicit function fails to exist as a graphical localization of the solution mapping, but there nevertheless exists a function with desirable properties serving locally as a selection of the set-valued solution mapping. This chapter does not demand of the reader more than calculus and some linear algebra, and it could therefore be used by both teachers and students in analysis courses.

Motivated by optimization problems and models of competitive equilibrium, Chapter 2 moves into wider territory. The questions are essentially the same as in the first chapter, namely, when a solution mapping can be localized to a function with some continuity properties. But it is no longer an equation that is being solved. Instead it is a condition called a generalized equation which captures a more complicated dependence and covers, as a special case, variational inequality conditions formulated in terms of the set-valued normal cone mapping associated with a convex set. Although our prime focus here is variational models, the presentation is self-contained and again could be handled by students and others without special background. It provides an introduction to a subject of great applicability which is hardly known to the mathematical community familiar with classical implicit functions, perhaps because of inadequate accessibility.

In Chapter 3 we depart from insisting on localizations that yield implicit functions and approach solution mappings from the angle of a "varying set." We identify continuity properties which support the paradigm of the implicit function theorem in a set-valued sense. This chapter may be read independently from the first two. Chapter 4 continues to view solution mappings from this angle but investigates substitutes for classical differentiability. By utilizing concepts of generalized derivatives, we are able to get implicit mapping theorems that reach far beyond the classical scope.

Chapter 5 takes a different direction. It presents extensions of the Banach open mapping theorem which are shown to fit infinite-dimensionally into the paradigm of the theory developed finite-dimensionally in Chapter 3. Some background in basic functional analysis is required. Chapter 6 goes further down that road and illustrates how some of the implicit function/mapping theorems from earlier in the book can be used in the study of problems in numerical analysis.

This book is targeted at a broad audience of researchers, teachers and graduate students, along with practitioners in mathematical sciences, engineering, economics and beyond. In summary, it concerns one of the chief topics in all of analysis, historically and now, an aid not only in theoretical developments but also in methods for solving specific problems. It crosses through several disciplines such as real and functional analysis, variational analysis, optimization, and numerical analysis, and can be used in part as a graduate text as well as a reference. It starts with elementary results and with each chapter, step by step, opens wider horizons by increasing the complexity of the problems and concepts that generate implicit function phenomena.

Many exercises are included, most of them supplied with detailed guides. These exercises complement and enrich the main results. The facts they encompass are sometimes invoked in the subsequent sections.

Each chapter ends with a short commentary which indicates sources in the literature for the results presented (but is not a survey of all the related literature). The commentaries to some of the chapters additionally provide historical overviews of past developments.

Whidbey Island, Washington *Asen L. Dontchev*
August, 2008 *R. Tyrrell Rockafellar*

Acknowledgements

Special thanks are owed to our readers Marius Durea, Shu Lu, Yoshiyuki Sekiguchi and Hristo Sendov, who gave us valuable feedback for the manuscript of the first edition, and to Francisco J. Aragón Artacho, who besides reviewing most of the book helped us masterfully with all the figures. During various stages of the writing we also benefited from discussions with Aris Daniilidis, Darinka Dentcheva, Hélène Frankowska, Michel Geoffroy, Alexander Ioffe, Stephen M. Robinson, Vladimir Veliov, and Constantin Zălinescu. We are also grateful to Mary Anglin for her help with the final copy-editing of the first edition of the book.

During the preparation of drafts of the second edition we received substantial feedback from Jonathan Borwein, Hélène Frankowska, Alexander Ioffe, Stephen Robinson and Vladimir Veliov, to whom we express our sincere gratitude. We are thankful in particular to the readers of the second edition Francisco J. Aragón Artacho, Radek Cibulka, Dmitriy Drusvyatskiy, Michael Hintermüller, Gabor Kassay, Mikhail Krastanov, Michael Lamm, Oana Silvia Serea, Lu Shu, Liang Yin and Nadya Zlateva.

The authors

Contents

List of Statements

Chapter 1

Chapter 2

Chapter 3

3A.1 (distance function characterizations of limits)
3A.2 (characterization of Painlevé–Kuratowski convergence)
3A.3 (characterization of Pompeiu–Hausdorff distance)
3A.4 (Pompeiu–Hausdorff versus Painlevé–Kuratowski)
3A.5 (convergence equivalence under boundedness)
3A.6 (conditions for Pompeiu–Hausdorff convergence)
3A.7 (unboundedness issues)
3B.1 (limit relations as equations)
3B.2 (characterization of semicontinuity)
3B.3 (characterization of Pompeiu–Hausdorff continuity)
3B.4 (solution mapping for a system of inequalities)
3B.5 (basic continuity properties of solution mappings in optimization)
3B.6 (minimization over a fixed set)
3B.7 (continuity of the feasible set versus continuity of the optimal value)
3C.1 (distance characterization of Lipschitz continuity)
3C.2 (polyhedral convex mappings from linear constraint systems)
3C.3 (Lipschitz continuity of polyhedral convex mappings)
3C.4 (Hoffman lemma)
3C.5 (Lipschitz continuity of mappings in linear programming)
3D.1 (outer Lipschitz continuity of polyhedral mappings)
3D.2 (polyhedrality of solution mappings to linear variational inequalities)
3D.3 (isc criterion for Lipschitz continuity)
3D.4 (Lipschitz continuity of polyhedral mappings)
3D.5 (single-valued polyhedral mappings)
3D.6 (single-valued solution mappings)
3D.7 (distance characterization of outer Lipschitz continuity)
3E.1 (local nonemptiness)
3E.2 (single-valued localization from Aubin property)
3E.3 (truncated Lipschitz continuity under convex-valuedness)
3E.4 (convexifying the values)
3E.5 (alternative description of Aubin property)
3E.6 (distance function characterization of Aubin property)
3E.7 (equivalence of metric regularity and the inverse Aubin property)
3E.8 (equivalent formulation)
3E.9 (equivalence of linear openness and metric regularity)
3E.10 (Aubin property of general solution mappings)
3F.1 (inverse mapping theorem with metric regularity)
3F.2 (detailed estimates)
3F.3 (perturbations with Lipschitz modulus 0)
3F.4 (utilization of strict first-order approximations)
3F.5 (utilization of strict differentiability)
3F.6 (affine-polyhedral variational inequalities)
3F.7 (strict differentiability and polyhedral convexity)
3F.8 (implicit mapping theorem with metric regularity)

Chapter 4

Chapter 5

Chapter 6

Notation

2C(4): formula (4) in Section 2C
$I\!R$: the real numbers
$I\!N$: the natural numbers
$I\!R^n$: the n-dimensional Euclidean space
$\mathcal{N}$: the collection of all subsets N of $I\!N$ such that $I\!N \setminus N$ is finite
$\mathcal{N}^\sharp$: the collection of all infinite subsets of $I\!N$
$\{x^k\}$, $\{x_k\}$: a sequence with elements x^k, x_k
$\varepsilon_k \searrow 0$: a sequence of positive numbers ε_k tending to 0
$\limsup_k C^k$: outer limit
$\liminf_k C^k$: inner limit
$|x|$: Euclidean norm
$\|x\|$: any norm
$\langle x, y\rangle$: canonical inner product, bilinear form
$|H|^+$: outer norm
$|H|^-$: inner norm
$I\!B_a(x)$: closed ball with center x and radius r
$I\!B$: closed unit ball
$\operatorname{int} I\!B$: open unit ball
$\operatorname{cl} C$: closure
$\operatorname{int} C$: interior
co: convex hull
$\operatorname{core} C$: core
$\operatorname{rc} C$: recession cone
P_C: projection mapping
$T_C(x)$: tangent cone
$N_C(x)$: normal cone
K^*: polar to cone K, mapping adjoint to K, space dual to K
$K_C(x, v)$: critical cone
A^T: transposition
$\operatorname{rank} A$: rank
$\ker A$: kernel

$\det A$: determinant
$d_C(x)$, $d(x,C)$: distance from x to C
$e(C,D)$: the excess of C beyond D
$h(C,D)$: Pompeiu-Hausforff distance
$\mathbf{d}(C,D)$: the minimal distance between C and D
$\operatorname{dom} F$: domain
$\operatorname{rge} F$: range
$\operatorname{gph} F$: graph
$\operatorname{fix}(F)$: the set of fixed points of F
$\nabla f(x)$: Jacobian
$Df(x)$: derivative
$\bar{\partial} f(x)$ Clarke generalized Jacobian
$\mathscr{C}^k$: the space of k-times continuously differentiable functions
$DF(x\,|\,y)$: graphical derivative
$\tilde{D}F(x\,|\,y)$: convexified graphical derivative
$D^*F(x\,|\,y)$: coderivative
$D_*F(x\,|\,y)$: strict graphical derivative
$\operatorname{clm}(f;x), \operatorname{clm}(S;y\,|\,x)$: calmness modulus
$\operatorname{lip}(f;x), \operatorname{lip}(S;y\,|\,x)$: Lipschitz modulus
$\widehat{\operatorname{clm}}_p(f;(p,x))$: partial calmness modulus
$\widehat{\operatorname{lip}}_p(f;(p,x))$: partial Lipschitz modulus
$\operatorname{reg}(F;x\,|\,y)$: regularity modulus
$\operatorname{subreg}(F;x\,|\,y)$: subregularity modulus

Chapter 1
Introduction and Equation-Solving Background

The idea of solving an equation $f(p,x) = 0$ for x as a function of p, say $x = s(p)$, plays a huge role in classical analysis and its applications. A function obtained in this way is said to be defined *implicitly* by the equation. The closely related idea of solving an equation $f(x) = y$ for x as a function of y concerns the *inversion* of f. The circumstances in which an implicit function or an inverse function exists and has properties like differentiability have long been studied. But all this can now be placed in a greatly expanded setting of modern problems, often involving more than just equations, in which results about implicit functions can be developed along remarkably similar lines.

The general picture at an introductory level is that of a "problem" which depends on a parameter vector p and has "solutions" x. The nature of the "problem" and what constitutes a "solution" matters less for the moment than the idea that for each p there is an associated set $S(p)$ of solutions x; this set could reduce to a single x or in some cases be empty. Fundamentally, we have a *solution mapping* S as a set-valued mapping from $\mathbb{R}^d$ to $\mathbb{R}^n$. For now we restrict our attention to Euclidean spaces, but it is clear that the general pattern goes far beyond finite dimensions. Signaled by the notation

$$S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n,$$

it can be identified graphically with the set

$$\operatorname{gph} S = \left\{ (p,x) \in \mathbb{R}^d \times \mathbb{R}^n \,\middle|\, x \in S(p) \right\}.$$

The traditional case of solving equations has $\operatorname{gph} S = \left\{ (p,x) \,\middle|\, f(p,x) = 0 \right\}$, but in other situations $\operatorname{gph} S$ might be specified very differently. Still, the goal of extracting from S an implicit function s could be essential.

Getting an implicit function typically requires some kind of localization. Consider for instance the elementary case where $\operatorname{gph} S = \left\{ (p,x) \,\middle|\, p - x^2 = 0 \right\}$, that is, S is the solution mapping of the equation $x^2 = p$, as illustrated in Fig. 1.1. The graph of S is not itself the graph of a function, but points $(\bar{p},\bar{x}) \in \operatorname{gph} S$ can have a neighborhood in which $\operatorname{gph} S$ reduces to the graph of a function s. This fails only for $(\bar{p},\bar{x}) = (0,0)$.

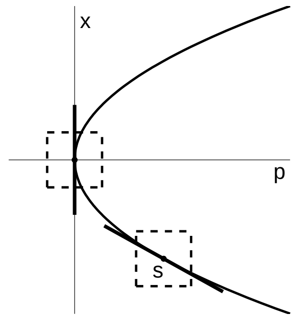

Fig. 1.1 Graphical localizations of the solution mapping of $x^2 = p$.

Broadly speaking, the challenge in each case of a parameterized "problem" and its "solutions" is to determine, for a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$, conditions that support the existence of a neighborhood in which $\operatorname{gph} S$ reduces to $\operatorname{gph} s$ for a function s, and to characterize the properties of that implicit function s. The conditions need to be derived from the given representation of the graph of S and usually rely on a sort of local approximation in that representation. For example, the classical implicit function theorem for $f(p,x) = 0$ utilizes the linearization of $f(\bar{p},x)$ at $\bar{x}$ and requires it to be invertible by asking its Jacobian $\nabla_x f(\bar{p},\bar{x})$ to be nonsingular. In fact, much the same pattern can be traced through a vastly wider territory of solution mappings. This is possible with the help of generalizations of differentiation and approximation promoted by recent advances in variational analysis. Developing that theme as the *implicit function paradigm* for solution mappings is the central aim of this book, along with building a corresponding platform for problem formulation and computation in various applications.

The purpose of this initial chapter is to pave the way with basic notation and terminology and a review of the classical equation-based theory. Beyond that review, however, many new ideas will already be brought into the picture in results which will undergo extension later. In Chapter 2 there will be so-called "generalized equations," covering variational inequalities as a special case. They offer a higher level of structure in which the implicit function paradigm can be propagated.

To set the stage for these developments, we start out with a discussion of general set-valued mappings $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ which are not necessarily to be interpreted as "solution mappings" but could have a role in their formulation. We mean by a such F a correspondence that assigns to each $x \in \mathbb{R}^n$ one or more elements of $\mathbb{R}^m$, or possibly none. The set of elements $y \in \mathbb{R}^m$ assigned by F to x is denoted by $F(x)$. However, instead of regarding F as going from $\mathbb{R}^n$ to a space of subsets of $\mathbb{R}^m$, we identify as the *graph* of F the set

$$\operatorname{gph} F = \left\{ (x,y) \in \mathbb{R}^n \times \mathbb{R}^m \,\middle|\, y \in F(x) \right\}.$$

Every subset of $I\!R^n \times I\!R^m$ serves as $\operatorname{gph} F$ for a uniquely determined $F : I\!R^n \rightrightarrows I\!R^m$, so this concept is very broad indeed and opens up many possibilities.

When F assigns more than one element to x we say it is *multivalued* at x, and when it assigns no element at all, it is *empty-valued* at x. When it assigns exactly one element y to x, it is *single-valued* at x, in which case we allow ourselves to write $F(x) = y$ instead of $F(x) = \{y\}$ and thereby build a bridge to handling functions as special cases of set-valued mappings.

Domains and ranges get flexible treatment in this way. For $F : I\!R^n \rightrightarrows I\!R^m$ the *domain* is the set

$$\operatorname{dom} F = \{\, x \mid F(x) \neq \emptyset \,\},$$

while the *range* is

$$\operatorname{rge} F = \{\, y \mid y \in F(x) \text{ for some } x \,\},$$

so that $\operatorname{dom} F$ and $\operatorname{rge} F$ are the projections of $\operatorname{gph} F$ on $I\!R^n$ and $I\!R^m$ respectively. Any subset of $\operatorname{gph} F$ can freely be regarded then as the graph of a set-valued submapping which likewise projects to some domain in $I\!R^n$ and range in $I\!R^m$.

The *functions* from $I\!R^n$ to $I\!R^m$ are identified in this context with the set-valued mappings $F : I\!R^n \rightrightarrows I\!R^m$ such that F is single-valued at every point of $\operatorname{dom} F$. When F is a function, we can emphasize this by writing $F : I\!R^n \to I\!R^m$, but the notation $F : I\!R^n \rightrightarrows I\!R^m$ doesn't preclude F from actually being a function. Usually, though, we use lower case letters for functions: $f : I\!R^n \to I\!R^m$. Note that in this notation f can still be empty-valued in places; it's single-valued only on the subset $\operatorname{dom} f$ of $I\!R^n$. Note also that, although we employ "mapping" in a sense allowing for potential multivaluedness (as in a "set-valued mapping"), no multivaluedness is ever involved when we speak of a "function."

A clear advantage of the framework of set-valued mappings over that of only functions is that *every set-valued mapping* $F : I\!R^n \rightrightarrows I\!R^m$ *has an inverse*, namely the set-valued mapping $F^{-1} : I\!R^m \rightrightarrows I\!R^n$ defined by

$$F^{-1}(y) = \{\, x \mid y \in F(x) \,\}.$$

The graph of F^{-1} is generated from the graph of F simply by reversing (x,y) to (y,x), which in the case of $m = n = 1$ corresponds to the reflection in Figure 1.1. In this manner a function f always has an inverse f^{-1} as a *set-valued mapping*. The question of an inverse *function* comes down then to passing to some piece of the graph of f^{-1}. For that, the notion of "localization" must come into play, as we are about to explain after a bit more background. Traditionally, a function $f : I\!R^n \to I\!R^m$ is *surjective* when $\operatorname{rge} f = I\!R^m$ and *injective* when $\operatorname{dom} f = I\!R^n$ and f^{-1} is a function; full *invertibility* of f corresponds to the juxtaposition of these two properties.

Graphical localization. *For $F : I\!R^n \rightrightarrows I\!R^m$ and a pair $(\bar{x},\bar{y}) \in \operatorname{gph} F$, a* graphical localization *of F at $\bar{x}$ for $\bar{y}$ is a set-valued mapping $\tilde{F}$ such that*

$$\operatorname{gph} \tilde{F} = (U \times V) \cap \operatorname{gph} F \text{ for some neighborhoods } U \text{ of } \bar{x} \text{ and } V \text{ of } \bar{y},$$

so that

$$\tilde{F} : x \mapsto \begin{cases} F(x) \cap V & \text{when } x \in U, \\ \emptyset & \text{otherwise.} \end{cases}$$

The inverse of $\tilde{F}$ *then has*

$$\tilde{F}^{-1}(y) = \begin{cases} F^{-1}(y) \cap U & \text{when } y \in V, \\ \emptyset & \text{otherwise,} \end{cases}$$

and is thus a graphical localization of the set-valued mapping F^{-1} *at* $\bar{y}$ *for* $\bar{x}$.

Often the neighborhoods U and V can conveniently be taken to be closed balls centered at $\bar{x}$ and $\bar{y}$, respectively. Observe, however, that the domain of a graphical localization $\tilde{F}$ of F with respect to U and V may differ from $U \cap \operatorname{dom} F$ and may well depend on the choice of V.

Single-valuedness in localizations. *By a single-valued localization of* F *at* $\bar{x}$ *for* $\bar{y}$ *will be meant a graphical localization that is a function, its domain not necessarily being a neighborhood of* $\bar{x}$. *The case where the domain is indeed a neighborhood of* $\bar{x}$ *will be indicated by referring to a single-valued localization of* F *around* $\bar{x}$ *for* $\bar{y}$ *instead of just at* $\bar{x}$ *for* $\bar{y}$.

In passing from inverse functions to implicit functions more generally, we need to pass from an equation $f(x) = y$ to one of the form

$$\text{(1)} \qquad f(p,x) = 0 \ \text{ for a function } f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$$

in which p acts as a parameter. The question is no longer that of inverting f, but the framework of set-valuedness is valuable nonetheless because it allows us to immediately introduce the *solution mapping*

$$\text{(2)} \qquad S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n \ \text{ with } S(p) = \{\, x \mid f(p,x) = 0 \,\}.$$

We can then look at pairs $(\bar{p},\bar{x})$ in $\operatorname{gph} S$ and ask whether S has a single-valued localization s around $\bar{p}$ for $\bar{x}$. Such a localization is exactly what constitutes an implicit function coming out of the equation.

Most calculus books present a result going back to Dini, who formulated and proved it in his lecture notes of 1877-78; the cover of Dini's manuscript[1] is displayed in Fig. 1.3. The version typically seen in advanced calculus texts is what we will refer to as the *classical implicit function theorem* or *Dini's theorem*. In those texts the set-valued solution mapping S in (2) never enters the picture directly, but a brief statement in that mode will help to show where we are headed in this book.

Dini classical implicit function theorem. *Let the function* $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ *in (1) be continuously differentiable in a neighborhood of* $(\bar{p},\bar{x})$ *and such that* $f(\bar{p},\bar{x}) = 0$, *and let the partial Jacobian of* f *with respect to* x *at* $(\bar{p},\bar{x})$, *namely* $\nabla_x f(\bar{p},\bar{x})$, *be*

[1] Many thanks to Danielle Ritelli from the University of Bologna for a copy of Dini's manuscript.

Fig. 1.2 Ulisse Dini (1845–1918).

nonsingular. Then the solution mapping S defined in (2) has a single-valued localization s around $\bar{p}$ for $\bar{x}$ which is continuously differentiable in a neighborhood Q of $\bar{p}$ with Jacobian satisfying

$$\nabla s(p) = -\nabla_x f(p,s(p))^{-1}\nabla_p f(p,s(p)) \quad \text{for every } p \in Q.$$

The classical inverse function theorem is the particular case of Dini implicit function theorem in which $f(p,x) = -p + f(x)$ (with some abuse of notation). Actually, these two theorems are equivalent; this will be shown in Section 1.B.

The example we started with illustrated in Fig. 1.1 corresponds to inverting the function $f(x) = x^2$ whose inverse f^{-1} is generally set-valued. Specifically, f^{-1} is single-valued only at 0 with $f^{-1}(0) = 0$, is empty-valued for $p < 0$ and two-valued for $p > 0$. It has a single-valued localization around $\bar{p} = 1$ for $\bar{x} = -1$ since the derivative of f at -1 is nonzero. Note that the derivative at $\bar{x} = 0$ is 0 and the inverse f^{-1} has no single-valued localization around 0 for 0. Indeed, as we will see in Section 1.2 [1B], the invertibility of the Jacobian is not only sufficient, as stated in Dini's theorem, but also necessary, for the existence of a single-valued localization of the inverse.

The Dini classical implicit function theorem and its variants will be taken up in detail in Section 1.2 [1B] after the development in Section 1.1 [1A] of an equivalent inverse function theorem. Later in Chapter 1 we gradually depart from the assumption of continuous differentiability of f to obtain far-reaching extensions of this classical theorem. It will be illuminating, for instance, to reformulate the as-

ANALISI
Infinitesimale
LEZIONI
DETTATE NELLA R. UNIVERSITÀ
PISA
dal Prof: Cav. Ulisse Dini
Anno Accademico 1877-78.
Parte I^a - CALCOLO DIFFERENZIALE
(L'Autore si riserva i diritti di Proprietà Letteraria)
Litografia Gozani
Pisa

Fig. 1.3 The front page of Dini's manuscript from 1877/78.

sumption about the Jacobian $\nabla_x f(\bar{p},\bar{x})$ as an assumption about the function

$$h(x) = f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x - \bar{x})$$

giving the partial linearization of f at $(\bar{p},\bar{x})$ with respect to x and having $h(\bar{x}) = 0$. The condition corresponding to the invertibility of $\nabla_x f(\bar{p},\bar{x})$ can be turned into the condition that the inverse mapping h^{-1}, with $\bar{x} \in h^{-1}(0)$, has a single-valued localization around 0 for $\bar{x}$. In this way the theme of single-valued localizations can be carried forward even into realms where f might not be differentiable and h could be some other kind of "local approximation" of f. We will be able to operate with a broad implicit function paradigm, extending in later chapters to much more than solving equations. It will deal with single-valued localizations s of solution mappings S to "generalized equations." An illustration of such a mapping is given in Fig. 1.4. These localizations s, if not differentiable, will at least have other key properties.

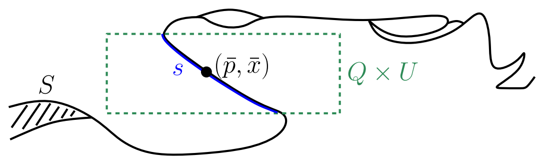

Fig. 1.4 Illustration of a single-valued localization.

At the end of this introductory preamble, some basic background needs to be recalled, and this is also an opportunity to fix some additional notation and terminology for subsequent use.

Terminology and notation

In working with $\mathbb{R}^n$ we will, in the first half of the book, keep to the Euclidean norm $|x|$ associated with the canonical inner product

$$\langle x, x' \rangle = \sum_{j=1}^n x_j x'_j \quad \text{for } x = (x_1, \ldots, x_n) \text{ and } x' = (x'_1, \ldots, x'_n),$$

namely

$$|x| = \sqrt{\langle x, x \rangle} = \Big[\sum_{j=1}^n x_j^2 \Big]^{1/2}.$$

The closed ball around $\bar{x}$ with radius r is then

$$\mathbb{B}_r(\bar{x}) = \{ x \mid |x - \bar{x}| \le r \}.$$

We denote the closed unit ball $\mathbb{B}_1(0)$ by $\mathbb{B}$. A *neighborhood* of $\bar{x}$ is any set U such that $\mathbb{B}_r(\bar{x}) \subset U$ for some $r > 0$. We recall for future needs that the interior of a set $C \subset \mathbb{R}^n$ consists of all points x such that C is a neighborhood of x, whereas the closure of C consists of all points x such that the *complement* of C is *not* a neighborhood of x; C is *open* if it coincides with its interior and *closed* if it coincides with its closure. The interior and closure of C will be denoted by $\operatorname{int} C$ and $\operatorname{cl} C$.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *upper semicontinuous* at a point $\bar{x}$ when $\bar{x} \in \operatorname{int} \operatorname{dom} f$ and for every $\varepsilon > 0$ there exists $\delta > 0$ for which

$$f(x) - f(\bar{x}) < \varepsilon \quad \text{whenever } x \in \operatorname{dom} f \text{ with } |x - \bar{x}| < \delta.$$

If instead we have

$$-\varepsilon < f(x) - f(\bar{x}) \quad \text{whenever } x \in \operatorname{dom} f \text{ with } |x - \bar{x}| < \delta,$$

then f is said to be *lower semicontinuous* at $\bar{x}$. Such upper and lower semicontinuity combine to *continuity*, meaning the existence for every $\varepsilon > 0$ of a $\delta > 0$ for which

$$|f(x) - f(\bar{x})| < \varepsilon \quad \text{whenever } x \in \operatorname{dom} f \text{ with } |x - \bar{x}| < \delta.$$

This condition, in our norm notation, carries over to defining the continuity of a vector-valued function $f : \mathbb{R}^n \to \mathbb{R}^m$ at a point $\bar{x} \in \operatorname{int} \operatorname{dom} f$. However, we also speak more generally then of f being *continuous at $\bar{x}$ relative to a set D* when $\bar{x} \in D \subset \operatorname{dom} f$ and this last estimate holds for $x \in D$; in that case $\bar{x}$ need not belong to $\operatorname{int} \operatorname{dom} f$. When f is continuous relative to D at every point of D, we say it is continuous on D. The graph $\operatorname{gph} f$ of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ with closed domain $\operatorname{dom} f$ that is continuous on $D = \operatorname{dom} f$ is a closed set in $\mathbb{R}^n \times \mathbb{R}^m$.

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *Lipschitz continuous* relative to a set D, or on a set D, if $D \subset \operatorname{dom} f$ and there is a constant $\kappa \geq 0$ such that

$$|f(x') - f(x)| \leq \kappa |x' - x| \quad \text{for all } x', x \in D.$$

If f is Lipschitz continuous relative to a neighborhood of a point $\bar{x} \in \operatorname{int} \operatorname{dom} f$, f is said to be Lipschitz continuous *around* $\bar{x}$. A function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous with respect to x uniformly in p near $(\bar{p}, \bar{x}) \in \operatorname{int} \operatorname{dom} f$ if there is a constant $\kappa \geq 0$ along with neighborhoods U of $\bar{x}$ and Q of $\bar{p}$ such that

$$|f(p, x') - f(p, x)| \leq \kappa |x' - x| \quad \text{for all } x', x \in U \text{ and } p \in Q.$$

Differentiability entails consideration of linear mappings. Although we generally allow for multivaluedness and even empty-valuedness when speaking of "mappings," single-valuedness everywhere is required of a linear mapping, for which we typically use a letter like A. A linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ is thus a function $A : \mathbb{R}^n \to \mathbb{R}^m$ with $\operatorname{dom} A = \mathbb{R}^n$ which obeys the usual rule for linearity:

$$A(\alpha x + \beta y) = \alpha A x + \beta A y \qquad \text{for all } x, y \in \mathbb{R}^n \text{ and all scalars } \alpha, \beta \in \mathbb{R}.$$

The *kernel* of A is

$$\ker A = \{ x \mid Ax = 0 \}.$$

In the finite-dimensional setting, we carefully distinguish between a linear mapping and its matrix, but often use the same notation for both. A linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ is represented then by a matrix A with m rows, n columns, and components $a_{i,j}$:

$$A = (a_{i,j})_{i,j=1}^{m,n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The inverse A^{-1} of a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ always exists in the set-valued sense, but it isn't a linear mapping unless it is actually a function with all of $\mathbb{R}^m$ as its domain, in which case A is said to be *invertible*. From linear algebra, of course, that requires $m = n$ and corresponds to the matrix A being nonsingular. More generally, if $m \le n$ and the rows of the matrix A are linearly independent, then the rank of the matrix A is m and the mapping A is surjective. In terms of the transpose of A, denoted by A^T, the matrix AA^T is in this case nonsingular. On the other hand, if $m \ge n$ and the columns of A are linearly independent then $A^\mathsf{T}A$ is nonsingular.

Both the identity mapping and its matrix will be denoted by I, regardless of dimensionality. By default, $|A|$ is the operator norm of A induced by the Euclidean norm,

$$|A| = \max_{|x|\le 1} |Ax|.$$

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *differentiable* at a point $\bar{x}$ when $\bar{x} \in \operatorname{int}\operatorname{dom} f$ and there is a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ with the property that for every $\varepsilon > 0$ there exists $\delta > 0$ with

$$|f(\bar{x}+h) - f(\bar{x}) - Ah| \le \varepsilon |h| \quad \text{for every } h \in \mathbb{R}^n \text{ with } |h| < \delta.$$

If such a mapping A exists at all, it is unique; it is denoted by $Df(\bar{x})$ and called the *derivative* of f at $\bar{x}$. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *twice differentiable* at a point $\bar{x} \in \operatorname{int}\operatorname{dom} f$ when there is a bilinear mapping $N : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ with the property that for every $\varepsilon > 0$ there exists $\delta > 0$ with

$$|f(\bar{x}+h) - f(\bar{x}) - Df(\bar{x})h - N(h,h)| \le \varepsilon |h|^2 \quad \text{for every } h \in \mathbb{R}^n \text{ with } |h| < \delta.$$

If such a mapping N exists it is unique and is called the *second derivative* of f at $\bar{x}$, denoted by $D^2 f(\bar{x})$. Higher-order derivatives can be defined accordingly.

The $m \times n$ matrix that represents the derivative $Df(\bar{x})$ is called the *Jacobian* of f at $\bar{x}$ and is denoted by $\nabla f(\bar{x})$. In the notation $x = (x_1, \ldots, x_n)$ and $f = (f_1, \ldots, f_m)$, the components of $\nabla f(\bar{x})$ are the partial derivatives of the component functions f_i:

$$\nabla f(\bar{x}) = \left(\frac{\partial f_j}{\partial x_i}(\bar{x}) \right)_{i,j=1}^{m,n}.$$

In distinguishing between $Df(\bar{x})$ as a linear mapping and $\nabla f(\bar{x})$ as its matrix, we can guard better against ambiguities which may arise in some situations. When the Jacobian $\nabla f(x)$ exists and is continuous (with respect to the matrix norms associated with the Euclidean norm) on a set $D \subset \mathbb{R}^n$, then we say that the function f is *continuously differentiable* on D; we also call such a function *smooth* or $\mathscr{C}^1$ on D. Accordingly, we define k times continuously differentiable ($\mathscr{C}^k$) functions.

For a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and a pair $(\bar{p}, \bar{x}) \in \operatorname{int}\operatorname{dom} f$, the *partial derivative* mapping $D_x f(\bar{p}, \bar{x})$ of f with respect to x at $(\bar{p}, \bar{x})$ is the derivative of the function $g(x) = f(\bar{p}, x)$ at $\bar{x}$. If the partial derivative mapping is continuous as a function of the pair (p, x) in a neighborhood of $(\bar{p}, \bar{x})$, then f is said to be continuously differentiable with respect to x around $(\bar{p}, \bar{x})$. The partial derivative $D_x f(\bar{p}, \bar{x})$ is represented by an $m \times n$ matrix, denoted $\nabla_x f(\bar{p}, \bar{x})$ and called the partial Jacobian. Respectively, $D_p f(\bar{p}, \bar{x})$ is represented by the $m \times d$ partial Jacobian $\nabla_p f(\bar{p}, \bar{x})$. It's a standard fact from calculus that if f is differentiable with respect to both p and x around $(\bar{p}, \bar{x})$ and the partial Jacobians $\nabla_x f(p,x)$ and $\nabla_p f(p,x)$ depend continuously on p and x, then f is continuously differentiable around $(\bar{p}, \bar{x})$.

1.1 [1A] The Classical Inverse Function Theorem

In this section of the book, we state and prove the classical inverse function theorem in two ways. In these proofs, and also later in the chapter, we will make use of the following two observations from calculus.

Fact 1 (estimates for differentiable functions)**.** *If a function $f:\mathbb{R}^n \to \mathbb{R}^m$ is differentiable at every point in a convex open neighborhood of $\bar{x}$ and the Jacobian mapping $x \mapsto \nabla f(x)$ is continuous at $\bar{x}$, then for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

(a) $$|f(x') - f(x) - \nabla f(x)(x'-x)| \le \varepsilon |x'-x| \quad \text{for every } x', x \in \mathbb{B}_\delta(\bar{x}).$$

Equivalently, for every $\varepsilon > 0$ there exists $\delta > 0$ such that

(b) $$|f(x') - f(x) - \nabla f(\bar{x})(x'-x)| \le \varepsilon |x'-x| \quad \text{for every } x', x \in \mathbb{B}_\delta(\bar{x}).$$

Proof. For a vector $h \in \mathbb{R}^m$ with $|h| = 1$ and points x, x', $x \neq x'$, in a convex open neighborhood of $\bar{x}$ where f is differentiable, define the function $\varphi:\mathbb{R} \to \mathbb{R}$ as $\varphi(t) = \langle h, f(x+t(x'-x))\rangle$. Then φ is continuous on $[0,1]$ and differentiable in $(0,1)$ and also $\varphi'(t) = \langle h, \nabla f(x+t(x'-x))(x'-x)\rangle$. A basic result in calculus, the *mean value theorem*, says that when a function $\psi:\mathbb{R} \to \mathbb{R}$ is continuous on an interval $[a,b]$ with $a < b$ and differentiable in (a,b), then there exists a point $c \in (a,b)$ such that $\psi(b) - \psi(a) = \psi'(c)(b-a)$; see, e.g., Bartle and Sherbert [1992], p. 197. Applying the mean value theorem to the function φ we obtain that there exists $\bar{t} \in (0,1)$ such that

$$\langle h, f(x')\rangle - \langle h, f(x)\rangle = \langle h, \nabla f(x+\bar{t}(x'-x))(x'-x)\rangle.$$

Then the triangle inequality and the assumed continuity of ∇f at $\bar{x}$ give us (a). The equivalence of (a) and (b) follows from the continuity of ∇f at $\bar{x}$. □

Later in the chapter, in 1D.7 we will show that (a) (and hence (b)) is equivalent to the continuity of the Jacobian mapping ∇f at $\bar{x}$.

Fact 2 (stability of matrix nonsingularity)**.** *Suppose A is a matrix-valued function from $\mathbb{R}^n$ to the space $\mathbb{R}^{m\times m}$ of all $m \times m$ real matrices, such that the determinant of $A(x)$, as well as those of its minors, depends continuously on x around $\bar{x}$ and the matrix $A(\bar{x})$ is nonsingular. Then there is a neighborhood U of $\bar{x}$ such that $A(x)$ is nonsingular for every $x \in U$ and, moreover, the function $x \mapsto A(x)^{-1}$ is continuous in U.*

Proof. Since the nonsingularity of $A(x)$ corresponds to $\det A(x) \neq 0$, it is sufficient to observe that the determinant of $A(x)$ (along with its minors) depends continuously on x. □

The classical inverse function theorem which parallels the classical implicit function theorem described in the introduction to this chapter reads as follows.

Theorem 1A.1 (classical inverse function theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in a neighborhood of a point $\bar{x}$ and let $\bar{y} := f(\bar{x})$. If $\nabla f(\bar{x})$ is nonsingular, then f^{-1} has a single-valued localization s around $\bar{y}$ for $\bar{x}$. Moreover, the function s is continuously differentiable in a neighborhood V of $\bar{y}$, and its Jacobian satisfies*

$$(1) \qquad \nabla s(y) = \nabla f(s(y))^{-1} \quad \text{for every } y \in V.$$

Examples.

1) For the function $f(x) = x^2$ considered in the introduction, the inverse f^{-1} is a set-valued mapping whose domain is $[0, \infty)$. It has two single-valued localizations around any $\bar{y} > 0$ for $\bar{x} \neq 0$, represented by either $x(y) = \sqrt{y}$ if $\bar{x} > 0$ or $x(y) = -\sqrt{y}$ if $\bar{x} < 0$. The inverse f^{-1} has no single-valued localization around $\bar{y} = 0$ for $\bar{x} = 0$.

2) The inverse f^{-1} of the function $f(x) = x^3$ is single-valued everywhere; it is the function $x(y) = \sqrt[3]{y}$. The inverse $f^{-1} = \sqrt[3]{y}$ is not differentiable at 0, which fits with the observation that $f'(0) = 0$.

3) For a higher-dimensional illustration, we look at diagonal real matrices

$$A = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

and the function $f : \mathbb{R}^2 \to \mathbb{R}^2$ which assigns to (λ_1, λ_2) the trace $y_1 = \lambda_1 + \lambda_2$ of A and the determinant $y_2 = \lambda_1 \lambda_2$ of A,

$$f(\lambda_1, \lambda_2) = \begin{pmatrix} \lambda_1 + \lambda_2 \\ \lambda_1 \lambda_2 \end{pmatrix}.$$

What can be said about the inverse of f? The range of f consists of all $y = (y_1, y_2)$ such that $4y_2 \leq y_1^2$. The Jacobian

$$\nabla f(\lambda_1, \lambda_2) = \begin{pmatrix} 1 & 1 \\ \lambda_2 & \lambda_1 \end{pmatrix}$$

has determinant $\lambda_1 - \lambda_2$, so it is nonsingular except along the line where $\lambda_1 = \lambda_2$, which corresponds to $4y_2 = y_1^2$. Therefore, f^{-1} has a smooth single-valued localization around $y = (y_1, y_2)$ for (λ_1, λ_2) as long as $4y_2 < y_1^2$, in fact two such localizations. But it doesn't have such a localization around other (y_1, y_2).

It will be illuminating to look at two proofs[2] of the classical inverse function theorem. The one we lay out first requires no more background than the facts listed

---

[2] These two proofs are not really different, if we take into account that the contraction mapping principle used in the second proof is proved by using an iterative procedure similar to the one used in the first proof.

at the beginning of this section, and it has the advantage of actually "calculating" a single-valued localization of f^{-1} by a procedure which is well known in numerical analysis, namely *Newton's iterative method*[3] for solving nonlinear equations. The second, which we include for the sake of connections with later developments, utilizes a nonconstructive, but very broad, fixed-point argument.

Proof I of Theorem 1A.1. First we introduce some constants. Let $\alpha > 0$ be a scalar so small that, by appeal to Fact 2 in the beginning of this section, the Jacobian matrix $\nabla f(x)$ is nonsingular for every x in $\mathbb{B}_\alpha(\bar{x})$ and the function $x \mapsto \nabla f(x)^{-1}$ is continuous in $\mathbb{B}_\alpha(\bar{x})$. Set

$$c = \max_{x \in \mathbb{B}_\alpha(\bar{x})} |\nabla f(x)^{-1}|.$$

On the basis of the estimate (a) in Fact 1, choose $a \in (0, \alpha]$ such that

$$(2) \qquad |f(x') - f(x) - \nabla f(x)(x' - x)| \leq \frac{1}{2c}|x' - x| \quad \text{for every } x', x \in \mathbb{B}_a(\bar{x}).$$

Let $b = a/(16c)$. Let s be the localization of f^{-1} with respect to the neighborhoods $\mathbb{B}_b(\bar{y})$ and $\mathbb{B}_a(\bar{x})$:

$$(3) \qquad \operatorname{gph} s = \big[\mathbb{B}_b(\bar{y}) \times \mathbb{B}_a(\bar{x})\big] \cap \operatorname{gph} f^{-1}.$$

We will show that s has the properties claimed. The argument is divided into three steps.

STEP 1: The localization s is nonempty-valued on $\mathbb{B}_b(\bar{y})$ with $\bar{x} \in s(\bar{y})$, in particular.

The fact that $\bar{x} \in s(\bar{y})$ is immediate of course from (3), inasmuch as $\bar{x} \in f^{-1}(\bar{y})$. Pick any $y \in \mathbb{B}_b(\bar{y})$ and any $x^0 \in \mathbb{B}_{a/8}(\bar{x})$. We will demonstrate that the iterative procedure

$$(4) \qquad x^{k+1} = x^k - \nabla f(x^k)^{-1}(f(x^k) - y), \quad k = 0, 1, \ldots$$

produces a sequence of vectors $x^1, x^2, \ldots$ which is convergent to a point $x \in f^{-1}(y) \cap \mathbb{B}_a(\bar{x})$. The procedure (4) is the celebrated Newton's iterative method for solving the equation $f(x) = y$ with a starting point x^0. By using induction we will show that this procedure generates an infinite sequence $\{x^k\}$ satisfying for $k = 1, 2, \ldots$ the following two conditions:

$$(5a) \qquad x^k \in \mathbb{B}_a(\bar{x})$$

[3] Isaac Newton (1643–1727). In 1669 Newton wrote his paper *De Analysi per Equationes Numero Terminorum Infinitas*, where, among other things, he describes an iterative procedure for approximating real roots of the equation $x^3 - 2x - 5 = 0$. In 1690 Joseph Raphson proposed a similar iterative procedure for solving more general polynomial equations and attributed it to Newton. It was Thomas Simpson who in 1740 stated the method in today's form (using Newton's fluxions) for an equation not necessarily polynomial, without making connections to the works of Newton and Raphson; he also noted that the method can be used for solving optimization problems by setting the gradient to zero.

and

$$|x^k - x^{k-1}| \leq \frac{a}{2^{k+1}}. \tag{5b}$$

To initialize the induction, we establish (5a) and (5b) for $k = 1$. Since $x^0 \in \mathbb{B}_{a/8}(\bar{x})$, the matrix $\nabla f(x^0)$ is indeed invertible, and (4) gives us x^1. The equality in (4) for $k = 0$ can also be written as

$$x^1 = -\nabla f(x^0)^{-1}(f(x^0) - y - \nabla f(x^0)x^0),$$

which we subtract from the obvious equality

$$\bar{x} = -\nabla f(x^0)^{-1}(f(\bar{x}) - \bar{y} - \nabla f(x^0)\bar{x}),$$

obtaining

$$\bar{x} - x^1 = -\nabla f(x^0)^{-1}(f(\bar{x}) - f(x^0) - \bar{y} + y - \nabla f(x^0)(\bar{x} - x^0)).$$

Taking norms on both sides and utilizing (2) with $x' = \bar{x}$ and $x = x^0$ we get

$$|x^1 - \bar{x}| \leq |\nabla f(x^0)^{-1}|(|f(\bar{x}) - f(x^0) - \nabla f(x^0)(\bar{x} - x^0)| + |y - \bar{y}|) \leq \frac{c}{2c}|x^0 - \bar{x}| + cb.$$

Inasmuch as $|x^0 - \bar{x}| \leq a/8$, this yields

$$|x^1 - \bar{x}| \leq \frac{a}{16} + cb = \frac{a}{8} \leq a.$$

Hence (5a) holds for $k = 1$. Moreover, by the triangle inequality,

$$|x^1 - x^0| \leq |x^1 - \bar{x}| + |\bar{x} - x^0| \leq \frac{a}{8} + \frac{a}{8} = \frac{a}{4},$$

which is (5b) for $k = 1$.

Assume now that (5a) and (5b) hold for $k = 1, 2, \ldots, j$. Then the matrix $\nabla f(x^k)$ is nonsingular for all such k and the iteration (4) gives us for $k = j$ the point x^{j+1}:

$$x^{j+1} = x^j - \nabla f(x^j)^{-1}(f(x^j) - y). \tag{6}$$

Through the preceding iteration, for $k = j - 1$, we have

$$y = f(x^{j-1}) + \nabla f(x^{j-1})(x^j - x^{j-1}).$$

Substituting this expression for y into (6), we obtain

$$x^{j+1} - x^j = -\nabla f(x^j)^{-1}(f(x^j) - f(x^{j-1}) - \nabla f(x^{j-1})(x^j - x^{j-1})).$$

Taking norms, we get from (2) that

$$|x^{j+1}-x^j| \le c|f(x^j)-f(x^{j-1})-\nabla f(x^{j-1})(x^j-x^{j-1})| \le \frac{1}{2}|x^j-x^{j-1}|.$$

The induction hypothesis on (5b) for $k=j$ then yields

$$|x^{j+1}-x^j| \le \frac{1}{2}|x^j-x^{j-1}| \le \frac{1}{2}\left(\frac{a}{2^{j+1}}\right) = \frac{a}{2^{j+2}}.$$

Hence, (5b) holds for $k=j+1$. Further,

$$\begin{aligned}|x^{j+1}-\bar{x}| &\le \sum_{i=1}^{j+1}|x^i-x^{i-1}|+|x^0-\bar{x}| \le \sum_{i=1}^{j+1}\frac{a}{2^{i+1}}+\frac{a}{8}\\ &< \frac{a}{4}\sum_{i=0}^{\infty}\frac{1}{2^i}+\frac{a}{8} = \frac{a}{2}+\frac{a}{8} = \frac{5a}{8} \le a.\end{aligned}$$

This gives (5a) for $k=j+1$ and the induction step is complete. Thus, both (5a) and (5b) hold for all $k=1,2,\ldots$.

To verify that the sequence $\{x^k\}$ converges, we observe next from (5b) that, for every k and j satisfying $k>j$, we have

$$|x^k-x^j| \le \sum_{i=j}^{k-1}|x^{i+1}-x^i| \le \sum_{i=j}^{\infty}\frac{a}{2^{i+2}} = \frac{a}{2^{j+1}}.$$

Hence, the sequence $\{x^k\}$ satisfies the Cauchy criterion, which is known to guarantee that it is convergent.

Let x be the limit of this sequence. Clearly, from (5a), we have $x \in \mathbb{B}_a(\bar{x})$. Through passing to the limit in (4), x must satisfy $x = x-\nabla f(x)^{-1}(f(x)-y)$, which is equivalent to $f(x)=y$. Thus, we have proved that for every $y \in \mathbb{B}_b(\bar{y})$ there exists $x \in \mathbb{B}_a(\bar{x})$ such that $x \in f^{-1}(y)$. In other words, the localization s of the inverse f^{-1} at $\bar{y}$ for $\bar{x}$ specified by (3) has nonempty values. In particular, $\mathbb{B}_b(\bar{y}) \subset \operatorname{dom} f^{-1}$.

STEP 2: The localization s is single-valued on $\mathbb{B}_b(\bar{y})$.

Let $y \in \mathbb{B}_b(\bar{y})$ and suppose x and x' belong to $s(y)$. Then $x,x' \in \mathbb{B}_a(\bar{x})$ and also

$$x = -\nabla f(x)^{-1}\big[f(x)-y-\nabla f(x)x\big] \quad \text{and} \quad x' = -\nabla f(x)^{-1}\big[f(x')-y-\nabla f(x)x'\big].$$

Consequently

$$x'-x = -\nabla f(x)^{-1}\big[f(x')-f(x)-\nabla f(x)(x'-x)\big].$$

Taking norms on both sides and invoking (2), we get

$$|x'-x| \le c|f(x')-f(x)-\nabla f(x)(x'-x)| \le \frac{1}{2}|x'-x|$$

which can only be true if $x'=x$.

STEP 3: The localization s is continuously differentiable in int $\mathbb{B}_b(\bar{y})$ with $\nabla s(y)$ expressed by (1).

An extension of the argument in Step 2 will provide a needed estimate. Consider any y and y' in $\mathbb{B}_b(\bar{y})$ and let $x = s(y)$ and $x' = s(y')$. These elements satisfy

$$x = -\nabla f(x)^{-1}\big[f(x) - y - \nabla f(x)x\big] \quad \text{and} \quad x' = -\nabla f(x)^{-1}\big[f(x') - y' - \nabla f(x)x'\big],$$

so that

$$x' - x = -\nabla f(x)^{-1}\big[f(x') - f(x) - \nabla f(x)(x' - x) - (y' - y)\big].$$

This implies through (2) that

$$|x' - x| \le c|f(x') - f(x) - \nabla f(x)(x' - x)| + c|y' - y| \le \frac{1}{2}|x' - x| + c|y' - y|,$$

hence $|x' - x| \le 2c|y' - y|$. Thus,

$$(7) \qquad |s(y') - s(y)| \le 2c|y' - y| \quad \text{for } y, y' \in \mathbb{B}_b(\bar{y}).$$

This estimate means that the localization s is Lipschitz continuous on $\mathbb{B}_b(\bar{y})$.

Now take any $\varepsilon > 0$. Then, from (a) in Fact 1, there exists $a' \in (0, a]$ such that

$$(8) \qquad |f(x') - f(x) - \nabla f(x)(x' - x)| \le \frac{\varepsilon}{2c^2}|x' - x| \quad \text{for every } x', x \in \mathbb{B}_{a'}(\bar{x}).$$

Let $b' > 0$ satisfy $b' \le \min\{b, a'/(2c)\}$. Then for every $y \in \mathbb{B}_{b'}(\bar{y})$, from (7) we obtain

$$(9) \qquad |s(y) - \bar{x}| = |s(y) - s(\bar{y})| \le 2c|y - \bar{y}| \le 2cb' \le a'.$$

Choose any $y \in \operatorname{int} \mathbb{B}_{b'}(\bar{y})$; then there exists $\tau > 0$ such that $\tau \le \varepsilon/(2c)$ and $y + h \in \mathbb{B}_{b'}(\bar{y})$ for any $h \in \mathbb{R}^n$ with $|h| \le \tau$. Writing the equalities $f(s(y+h)) = y + h$ and $f(s(y)) = y$ as

$$s(y+h) = -\nabla f(s(y))^{-1}(f(s(y+h)) - y - h - \nabla f(s(y))s(y+h))$$

and

$$s(y) = -\nabla f(s(y))^{-1}(f(s(y)) - y - \nabla f(s(y))s(y))$$

and subtracting the second from the first, we obtain

$$\begin{aligned} &s(y+h) - s(y) - \nabla f(s(y))^{-1}h \\ &\qquad = -\nabla f(s(y))^{-1}(f(s(y+h)) - f(s(y)) - \nabla f(s(y))(s(y+h) - s(y))). \end{aligned}$$

Once again taking norms on both sides, and using (7), (8) and (9), we get

$$|s(y+h) - s(y) - \nabla f(s(y))^{-1}h| \le \frac{c\varepsilon}{2c^2}|s(y+h) - s(y)| \le \varepsilon|h| \text{ whenever } h \in \mathbb{B}_\tau(0).$$

By definition, this says that the function s is differentiable at y and that its Jacobian equals $\nabla f(s(y))^{-1}$, as claimed in (1). This Jacobian is continuous in int $\mathbb{B}_b(\bar{y})$; this comes from the continuity of ∇f^{-1} in $\mathbb{B}_a(\bar{x})$ where are the values of s, and the continuity of s in int $\mathbb{B}_b(\bar{y})$, and also taking into account that a composition of continuous functions is continuous. $\square$

We can make a shortcut through Steps 1 and 2 of the Proof I, arriving at the promised Proof II, if we employ a deeper result of analysis far beyond the framework so far, namely the contraction mapping principle. Although we work here in Euclidean spaces, we state this theorem in the framework of a complete metric space, as is standard in the literature. More general versions of this principle for set-valued mappings will be proved in Section 5E, from which we will derive the standard contraction mapping principle given next as Theorem 1A.2. The reader who wants to stick with Euclidean spaces may assume that X is a closed nonempty subset of $\mathbb{R}^n$ with metric $\rho(x,y) = |x-y|$.

Theorem 1A.2 (contraction mapping principle). *Let X be a complete metric space with metric ρ. Consider a point $\bar{x} \in X$ and a function $\Phi : X \to X$ for which there exist scalars $a > 0$ and $\lambda \in [0,1)$ such that:*
(a) $\rho(\Phi(\bar{x}),\bar{x}) \leq a(1-\lambda)$;
(b) $\rho(\Phi(x'),\Phi(x)) \leq \lambda\rho(x',x)$ *for every* $x',x \in \mathbb{B}_a(\bar{x})$.
Then there is a unique $x \in \mathbb{B}_a(\bar{x})$ satisfying $x = \Phi(x)$, that is, Φ has a unique fixed point in $\mathbb{B}_a(\bar{x})$.

Most common in the literature is another formulation of the contraction mapping principle which seems more general but is actually equivalent to 1A.2. To distinguish it from 1A.2, we call it *basic*.

Theorem 1A.3 (basic contraction mapping principle). *Let X be a complete metric space with metric ρ and let $\Phi : X \to X$. Suppose that there exists $\lambda \in [0,1)$ such that*

$$\rho(\Phi(x'),\Phi(x)) \leq \lambda\rho(x',x) \quad \text{for every } x',x \in X.$$

Then there is a unique $x \in X$ satisfying $x = \Phi(x)$.

Another equivalent version of the contraction mapping principle involves a parameter.

Theorem 1A.4 (parametric contraction mapping principle). *Let P be a metric space with metric σ and X be a complete metric space with metric ρ, and let $\Phi : P \times X \to X$. Suppose that there exist $\lambda \in [0,1)$ and $\mu \geq 0$ such that*

$$(10) \qquad \rho(\Phi(p,x'),\Phi(p,x)) \leq \lambda\,\rho(x',x) \quad \text{for every } x',x \in X \text{ and } p \in P$$

and

$$(11) \qquad \rho(\Phi(p',x),\Phi(p,x)) \leq \mu\,\sigma(p',p) \quad \text{for every } p',p \in P \text{ and } x \in X.$$

Then the mapping

$$\psi : p \mapsto \{ x \in X \mid x = \Phi(p,x) \} \quad \textit{for } p \in P \tag{12}$$

is single-valued on P, *which is moreover Lipschitz continuous on* P *with Lipschitz constant* $\mu/(1-\lambda)$.

Exercise 1A.5. *Prove that theorems 1A.2, 1A.3 and 1A.4 are equivalent.*

Guide. Let 1A.2 be true and let Φ satisfy the assumptions in 1A.3 with some $\lambda \in [0,1)$. Choose $\bar{x} \in X$; then $\Phi(\bar{x}) \in X$. Let $a > \rho(\bar{x},\Phi(\bar{x}))/(1-\lambda)$. Then (a) and (b) are satisfied with this a and hence there exists a unique fixed point x of Φ in $\mathbb{B}_a(\bar{x})$. The uniqueness of this fixed point in the whole X follows from the contraction property. To prove the converse implication first use (a)(b) to obtain that Φ maps $\mathbb{B}_a(\bar{x})$ into itself and then use the fact that the closed ball $\mathbb{B}_a(\bar{x})$ equipped with metric ρ is a complete metric space. Another way to have equivalence of 1A.2 and 1A.3 is to reformulate 1A.2 with a being possibly ∞.

Let 1A.3 be true and let Φ satisfy the assumptions (10) and (11) in 1A.4 with corresponding λ and μ. Then, by 1A.3, for every fixed $p \in P$ the set $\{ x \in X \mid x = \Phi(p,x) \}$ is a singleton; that is, the mapping ψ in (12) is a function with domain P. To complete the proof, choose $p', p \in P$ and the corresponding $x' = \Phi(p',x')$, $x = \Phi(p,x)$, and use (10), (11) and the triangle inequality to obtain

$$\begin{aligned} \rho(x',x) &= \rho(\Phi(p',x'),\Phi(p,x)) \\ &\le \rho(\Phi(p',x'),\Phi(p',x)) + \rho(\Phi(p',x),\Phi(p,x)) \le \lambda\, \rho(x',x) + \mu\, \sigma(p',p). \end{aligned}$$

Rearranging the terms gives us the desired Lipschitz continuity. □

Proof II of Theorem 1A.1. Let $A = \nabla f(\bar{x})$ and let $c := |A^{-1}|$. There exists $a > 0$ such that from the estimate (b) in Fact 1 (in the beginning of this section) we have

$$|f(x') - f(x) - \nabla f(\bar{x})(x'-x)| \le \frac{1}{2c}|x'-x| \quad \text{for every } x',x \in \mathbb{B}_a(\bar{x}). \tag{13}$$

Let $b = a/(4c)$. The space $\mathbb{R}^n$ equipped with the Euclidean norm is a complete metric space, so in this case X in Theorem 1A.2 is identified with $\mathbb{R}^n$. Fix $y \in \mathbb{B}_b(\bar{y})$ and consider the function

$$\Phi_y : x \mapsto x - A^{-1}(f(x) - y) \quad \text{for } x \in \mathbb{B}_a(\bar{x}).$$

We have

$$|\Phi_y(\bar{x}) - \bar{x}| = |-A^{-1}(\bar{y} - y)| \le cb = \frac{ca}{4c} < a\left(1 - \frac{1}{2}\right),$$

hence condition (a) in the contraction mapping principle 1A.2 holds with the so chosen a and $\lambda = 1/2$. Further, for any $x, x' \in \mathbb{B}_a(\bar{x})$, from (13) we obtain that

$$|\Phi_y(x) - \Phi_y(x')| = |x - x' - A^{-1}(f(x) - f(x'))|$$

$$\le |A^{-1}||f(x) - f(x') - A(x - x')|$$
$$\le c\frac{1}{2c}|x - x'| = \frac{1}{2}|x - x'|.$$

Thus condition (b) in 1A.2 is satisfied with the same $\lambda = 1/2$. Hence, there is a unique $x \in \mathbb{B}_a(\bar{x})$ such that $\Phi_y(x) = x$; that is equivalent to $f(x) = y$.

Translated into our terminology, this tells us that f^{-1} has a single-valued localization around $\bar{y}$ for $\bar{x}$ whose graph satisfies (3). The continuous differentiability is argued once more through Step 3 of Proof I. □

Exercise 1A.6. *For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ prove that $(f^{-1})^{-1}(x) = f(x)$ for every $x \in \operatorname{dom} f$.*

Guide. Let $x \in \operatorname{dom} f$ and let $y = f(x)$. Then $x \in f^{-1}(y)$ and hence $y \in (f^{-1})^{-1}(x)$; thus $f(x) \in (f^{-1})^{-1}(x)$. Then show by contradiction that the mapping $(f^{-1})^{-1}$ is single-valued. □

Exercise 1A.7. *Prove Theorem 1A.1 by using, instead of iteration (4), the iteration*

$$x^{k+1} = x^k - \nabla f(\bar{x})^{-1}(f(x^k) - y), \quad k = 0, 1, \ldots.$$

Guide. Follow the argument in Proof I with respective adjustments of the constants involved. □

In this and the following chapters we will derive the classical inverse function theorem 1A.1 a number of times and in different ways from more general theorems or utilizing other basic results. For instance, in Section 1.6 [1F] we will show how to obtain 1A.1 from Brouwer's invariance of domain theorem and in Section 4B we will prove 1A.1 again with the help of the Ekeland variational principle.

There are many roads to be taken from here, by relaxing the assumptions in the classical inverse function theorem, that lead to a variety of results. Some of them are paved and easy to follow, others need more advanced techniques, and a few lead to new territories which we will explore later in the book.

1.2 [1B] The Classical Implicit Function Theorem

In this section we give a proof of the classical implicit function theorem stated by Dini and described in the introduction to this chapter. We consider a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ with values $f(p,x)$, where p is the parameter and x is the variable to be determined, and introduce for the equation $f(p,x) = 0$ the associated *solution mapping*

$$(1) \qquad S : p \mapsto \left\{ x \in \mathbb{R}^n \,\middle|\, f(p,x) = 0 \right\} \quad \text{for } p \in \mathbb{R}^d.$$

We restate the result, furnishing it with a label for reference.

Theorem 1B.1 (Dini classical implicit function theorem)**.** *Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$ and such that $f(\bar{p},\bar{x}) = 0$, and let the partial Jacobian of f with respect to x at $(\bar{p},\bar{x})$, namely $\nabla_x f(\bar{p},\bar{x})$, be nonsingular. Then the solution mapping S defined in (1) has a single-valued localization s around $\bar{p}$ for $\bar{x}$ which is continuously differentiable in a neighborhood Q of $\bar{p}$ with Jacobian satisfying*

$$\nabla s(p) = -\nabla_x f(p,s(p))^{-1}\nabla_p f(p,s(p)) \quad \text{for every } p \in Q. \tag{2}$$

The classical inverse function theorem is the particular case of the classical implicit function theorem in which $f(p,x) = -p + f(x)$ (with a slight abuse of notation). However, it will also be seen now that the classical implicit function theorem can be obtained from the classical inverse function theorem. For that, we first state an easy-to-prove fact from linear algebra.

Lemma 1B.2. *Let I be the $d \times d$ identity matrix, 0 be the $d \times n$ zero matrix, B be an $n \times d$ matrix, and A be an $n \times n$ nonsingular matrix. Then the square matrix*

$$J = \begin{pmatrix} I & 0 \\ B & A \end{pmatrix}$$

is nonsingular.

Proof. If J is singular, then there exists

$$y = \begin{pmatrix} p \\ x \end{pmatrix} \neq 0 \ \text{ such that } \ Jy = 0,$$

which reduces to the equation

$$\begin{pmatrix} p \\ Bp + Ax \end{pmatrix} = 0.$$

Hence there exists $x \neq 0$ with $Ax = 0$, which contradicts the nonsingularity of A. □

Proof of Theorem 1B.1. Consider the function

$$\varphi(p,x) = \begin{pmatrix} p \\ f(p,x) \end{pmatrix}$$

acting from $\mathbb{R}^d \times \mathbb{R}^n$ to itself. The inverse of this function is defined by the solutions of the equation

$$\varphi(p,x) = \begin{pmatrix} p \\ f(p,x) \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \tag{3}$$

where the vector $(y_1,y_2) \in I\!R^d \times I\!R^n$ is now the parameter and (p,x) is the dependent variable. The nonsingularity of the partial Jacobian $\nabla_x f(\bar{p},\bar{x})$ implies through Lemma 1B.2 that the Jacobian of the function φ in (3) at the point $(\bar{x},\bar{p})$, namely the matrix

$$J(\bar{p},\bar{x}) = \begin{pmatrix} I & 0 \\ \nabla_p f(\bar{p},\bar{x}) & \nabla_x f(\bar{p},\bar{x}) \end{pmatrix},$$

is nonsingular as well. Then, according to the classical inverse function theorem 1A.1, the inverse φ^{-1} of the function in (3) has a single-valued localization

$$(y_1,y_2) \mapsto (q(y_1,y_2), r(y_1,y_2)) \text{ around } (\bar{p},0) \text{ for } (\bar{p},\bar{x})$$

which is continuously differentiable around $(\bar{p},0)$. To develop formula (2), we note that

$$\begin{cases} q(y_1,y_2) & = y_1, \\ f(y_1,r(y_1,y_2)) & = y_2. \end{cases}$$

Differentiating the second equality with respect to y_1 by using the chain rule, we get

$$\text{(4)} \qquad \nabla_p f(y_1,r(y_1,y_2)) + \nabla_x f(y_1,r(y_1,y_2)) \cdot \nabla_{y_1} r(y_1,y_2) = 0.$$

When (y_1,y_2) is close to $(\bar{p},0)$, the point $(y_1,r(y_1,y_2))$ is close to $(\bar{p},\bar{x})$ and then $\nabla_x f(y_1,r(y_1,y_2))$ is nonsingular (Fact 2 in Section 1.1 [1A]). Thus, solving (4) with respect to $\nabla_{y_1} r(y_1,y_2)$ gives

$$\nabla_{y_1} r(y_1,y_2) = -\nabla_x f(y_1,r(y_1,y_2))^{-1} \nabla_p f(y_1,r(y_1,y_2)).$$

In particular, at points $(y_1,y_2) = (p,0)$ close to $(\bar{p},0)$ we have that the mapping $p \mapsto s(p) := r(p,0)$ is a single-valued localization of the solution mapping S in (1) around $\bar{p}$ for $\bar{x}$ which is continuously differentiable around $\bar{p}$ and its derivative satisfies (2). □

Thus, the classical implicit function theorem 1B.1 is *equivalent* to the classical inverse function theorem 1A.1. We now look at yet another equivalent result.

Theorem 1B.3 (correction function theorem)**.** *Let $f : I\!R^n \to I\!R^n$ be continuously differentiable in a neighborhood of $\bar{x}$. If $\nabla f(\bar{x})$ is nonsingular, then the correction mapping*

$$\Xi : x \mapsto \left\{ u \in I\!R^n \,\middle|\, f(x+u) = f(\bar{x}) + \nabla f(\bar{x})(x-\bar{x}) \right\} \quad \text{for } x \in I\!R^n$$

has a single-valued localization ξ around $\bar{x}$ for 0. Moreover, ξ is continuously differentiable in a neighborhood U of $\bar{x}$ with $\nabla \xi(\bar{x}) = 0$.

Proof. Consider the function

$$\varphi : (x,u) \mapsto f(x+u) - f(\bar{x}) - \nabla f(\bar{x})(x-\bar{x}) \text{ for } (x,u) \in I\!R^n \times I\!R^n$$

in a neighborhood of $(\bar{x},\bar{u})$ for $\bar{u} := 0$. Since $\nabla_u \varphi(\bar{x},\bar{u}) = \nabla f(\bar{x})$ is nonsingular, we apply the classical implicit function theorem 1B.1 obtaining that the solution mapping

$$\Xi : x \mapsto \left\{ u \in \mathbb{R}^n \,\middle|\, \varphi(x,u) = 0 \right\} \quad \text{for } x \in \mathbb{R}^n$$

has a smooth single-valued localization ξ around $\bar{x}$ for 0. The chain rule gives us $\nabla\xi(\bar{x}) = 0$. □

Exercise 1B.4. *Prove that the correction function theorem 1B.3 implies the inverse function theorem 1A.1.*

Guide. Let $\bar{y} := f(\bar{x})$ and assume $A := \nabla f(\bar{x})$ is nonsingular. In these terms the correction function theorem 1B.3 claims that the mapping

$$\Xi : z \mapsto \left\{ \xi \in \mathbb{R}^n \,\middle|\, f(z+\xi) = \bar{y} + A(z-\bar{x}) \right\} \text{ for } z \in \mathbb{R}^n$$

has a single-valued localization ξ around $\bar{x}$ for 0 and that ξ is continuously differentiable around $\bar{x}$ and has zero derivative at $\bar{x}$. The affine function $y \mapsto z(y) := \bar{x} + A^{-1}(y-\bar{y})$ is the solution mapping of the linear equation $\bar{y} + A(z-\bar{x}) = y$ having $z(\bar{y}) = \bar{x}$. The composite function $y \mapsto \xi(z(y))$ hence satisfies

$$f(z(y) + \xi(z(y))) = \bar{y} + \nabla f(\bar{x})(z(y) - \bar{x}) = y.$$

The function $s(y) := z(y) + \xi(z(y))$ is a single-valued localization of the inverse f^{-1} around $\bar{y}$ for $\bar{x}$. To show that $\nabla s(y) = \nabla f(s(y))^{-1}$, use the chain rule. □

Inasmuch as the classical inverse function theorem implies the classical implicit function theorem, and the correction function theorem is a corollary of the classical implicit function theorem, all three theorems — the inverse, the implicit and the correction function theorems, stated in 1A.1, 1B.1 and 1B.3, respectively — are equivalent.

Proposition 1B.5 (higher derivatives)**.** *In Theorem 1B.1, if f is k times continuously differentiable around $(\bar{p},\bar{x})$ then the localization s of the solution mapping S is k times continuously differentiable around $\bar{p}$. Likewise in Theorem 1A.1, if f is k times continuously differentiable around $\bar{x}$, then the localization s of f^{-1} is k times continuously differentiable around $\bar{y}$.*

Proof. For the implicit function theorem 1B.1, this is an immediate consequence of the formula in (2) by way of the chain rule for differentiation. It follows then for the inverse function theorem 1A.1 as a special case. □

If we relax the differentiability assumption for the function f, we obtain a result of a different kind, the origins of which go back to the work of Goursat [1903].

Theorem 1B.6 (Goursat implicit function theorem)**.** *For the solution mapping S defined in (1), consider a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$. Assume that:*

(a) *$f(p,x)$ is differentiable with respect to x in a neighborhood of the point $(\bar{p},\bar{x})$, and both $f(p,x)$ and $\nabla_x f(p,x)$ depend continuously on (p,x) in this neighborhood;*

(b) $\nabla_x f(\bar{p},\bar{x})$ *is nonsingular.*
Then S has a single-valued localization around $\bar{p}$ for $\bar{x}$ which is continuous at $\bar{p}$.

We will prove a far reaching generalization of this result in Section 2B, which we supply with a detailed proof. In the following exercise we give a guide for a direct proof.

Exercise 1B.7. *Prove Theorem 1B.6.*

Guide. Mimic the proof of 1A.1 by choosing α and β sufficiently small so that

$$c = \max_{\substack{x\in \boldsymbol{B}_\alpha(\bar{x})\\ p\in \boldsymbol{B}_\beta(\bar{p})}} |\nabla_x f(p,x)^{-1}|.$$

Then pick $a \in (0,\alpha]$ and $q \in (0,\beta]$ such that, as in the estimate (a) in Fact 1, for every $x',x \in \mathbb{B}_a(\bar{x})$ and $p \in \mathbb{B}_q(\bar{p})$,

$$(5) \qquad |f(p,x') - f(p,x) - \nabla_x f(p,x)(x'-x)| \le \frac{1}{2c}|x'-x|.$$

Then use the iteration

$$x^{k+1} = x^k - \nabla_x f(\bar{p},\bar{x})^{-1} f(p,x^k)$$

to obtain that S has a nonempty graphical localization s around $\bar{p}$ for $\bar{x}$. As in Step 2 in Proof I of 1A.1, show that s is single-valued. To show continuity at $\bar{p}$, for $x = s(p)$ subtract from

$$x = -\nabla_x f(\bar{p},\bar{x})^{-1}(f(p,x) - \nabla_x f(\bar{p},\bar{x})x)$$

the equality

$$\bar{x} = -\nabla_x f(\bar{p},\bar{x})^{-1}(f(\bar{p},\bar{x}) - \nabla_x f(\bar{p},\bar{x})\bar{x}),$$

and, after adding and subtracting terms, use (5). □

It turns out that the nonsingularity of the partial Jacobian $\nabla_x f(\bar{p},\bar{x})$ is not only sufficient but also becomes a necessary condition for the solution mapping of (1) to have a single-valued smooth localization provided that $\nabla_p f(\bar{p},\bar{x})$ is of full rank. In Section 2C we consider in more detail such a parameterization in a much broader context and call it *ample* parameterization. In the classical context the corresponding result is as follows.

Theorem 1B.8 (symmetric implicit function theorem)**.** *Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$ and such that $f(\bar{p},\bar{x}) = 0$, and let $\nabla_p f(\bar{p},\bar{x})$ be of full rank n. Then the solution mapping S defined in (1) has a single-valued localization s around $\bar{p}$ for $\bar{x}$ which is continuously differentiable in a neighborhood Q of $\bar{p}$ if and only if the partial Jacobian $\nabla_x f(\bar{p},\bar{x})$ is nonsingular.*

Proof. We only need to proof the “only if” part. Without loss of generality let $\bar{p} = 0$ and $\bar{x} = 0$ and let $A = \nabla_x f(0,0)$ and $B = \nabla_p f(0,0)$. Consider the function

$$\psi(q,x,y) = f(B^{\mathsf{T}}q,x) - Ax + y.$$

Observe that $\nabla_q \psi(0,0,0) = BB^{\mathsf{T}}$ is nonsingular since B is assumed onto. Hence, by the classical implicit function theorem the solution mapping

$$\mathbb{R}^n \times \mathbb{R}^n \ni (x,y) \mapsto \Psi(x,y) := \left\{ q \in \mathbb{R}^n \,\middle|\, \psi(q,x,y) = 0 \right\}$$

has a single-valued smooth localization at $(0,0)$ for 0; denote this localization by σ.

Now assume that the solution mapping $p \mapsto S(p) = \left\{ x \,\middle|\, f(p,x) = 0 \right\}$ has a single-valued smooth localization s at 0 for 0, i.e. $f(p,s(p)) = 0$ for all p near 0, s is smooth near 0 and $s(0) = 0$. Clearly, we could choose x and y so close to 0 that the norms of $p = B^{\mathsf{T}}q$, $q = \sigma(x,y)$ and $s(p)$ are small enough to satisfy $q = \sigma(x,y) \in \Psi(x,y)$ and $x = s(B^{\mathsf{T}}q) \in S(B^{\mathsf{T}}q)$. Then $s(B^{\mathsf{T}}\sigma(x,y))$ satisfies $-As(B^{\mathsf{T}}\sigma(x,y)) + y = 0$. We obtain that for every y in a neighborhood of 0, hence by linearity, for every $y \in \mathbb{R}^n$ the equation $Ax = y$ has a solution, hence A must be nonsingular. $\square$

The classical inverse function theorem 1A.1 combined with Theorem 1B.8 gives us

Theorem 1B.9 (symmetric inverse function theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable around $\bar{x}$. Then the following are equivalent:*

(a) $\nabla f(\bar{x})$ *is nonsingular;*

(b) f^{-1} *has a single-valued localization s around $\bar{y} := f(\bar{x})$ for $\bar{x}$ which is continuously differentiable around $\bar{y}$.*

The formula for the Jacobian of the single-valued localization s of the inverse,

$$\nabla s(y) = \nabla f(s(y))^{-1} \quad \text{for } y \text{ around } \bar{y},$$

comes as a byproduct of the statement (b) by way of the chain rule.

Exercise 1B.10. *Consider a polynomial of degree $n > 0$,*

$$p(x) = \sum_{i=0}^{n} a_i x^i,$$

where the coefficients $a_0, \ldots, a_n$ are real numbers. For each coefficient vector $a = (a_0, \ldots, a_n) \in \mathbb{R}^{n+1}$ let $S(a)$ be the set of all real zeros of p, so that S is a mapping from $\mathbb{R}^{n+1}$ to $\mathbb{R}$ whose domain consists of the vectors a such that p has at least one real zero. Let $\bar{a}$ be a coefficient vector such that p has a simple real zero $\bar{s}$; thus $p(\bar{s}) = 0$ but $p'(\bar{s}) \neq 0$. Prove that S has a smooth single-valued localization around $\bar{a}$ for $\bar{s}$. Is such a statement correct when $\bar{s}$ is a double zero?

1.3 [1C] Calmness

In this section we introduce a continuity property of functions which will play an important role in the book.

Calmness. *A function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *is said to be* calm *at* $\bar{x}$ *relative to a set* D *in* $\mathbb{R}^n$ *if* $\bar{x} \in D \cap \operatorname{dom} f$ *and there exists a constant* $\kappa \geq 0$ *such that*

$$(1) \qquad |f(x) - f(\bar{x})| \leq \kappa |x - \bar{x}| \quad \textit{for all } x \in D \cap \operatorname{dom} f.$$

The calmness property (1) can alternatively be expressed in the form of the inclusion

$$f(x) \in f(\bar{x}) + \kappa |x - \bar{x}| \mathbb{B} \quad \text{for all } x \in D \cap \operatorname{dom} f.$$

That expression connects with the generalization of the definition of calmness to set-valued mappings, which we will discuss at length in Chapter 3.

Note that a function f which is calm at $\bar{x}$ may have empty values at some points x near $\bar{x}$ when $\bar{x}$ is on the boundary of $\operatorname{dom} f$. If $\bar{x}$ is an isolated point of $D \cap \operatorname{dom} f$, then trivially f is calm at $\bar{x}$ relative to D with $\kappa = 0$.

We will mostly use a local version of the calmness property where the set D in the condition (1) is a neighborhood of $\bar{x}$; if such a neighborhood exists we simply say that f is calm at $\bar{x}$. Calmness of this kind can be identified with the finiteness of the *modulus* which we proceed to define next.

Calmness modulus. *For a function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *and a point* $\bar{x} \in \operatorname{dom} f$*, the calmness modulus of* f *at* $\bar{x}$*, denoted* $\operatorname{clm}(f;\bar{x})$*, is the infimum of the set of values* $\kappa \geq 0$ *for which there exists a neighborhood* D *of* $\bar{x}$ *such that (1) holds.*

According to this, as long as $\bar{x}$ is *not an isolated point* of $\operatorname{dom} f$, the calmness modulus satisfies

$$\operatorname{clm}(f;\bar{x}) = \limsup_{\substack{x \in \operatorname{dom} f, x \to \bar{x}, \\ x \neq \bar{x}}} \frac{|f(x) - f(\bar{x})|}{|x - \bar{x}|}.$$

If $\bar{x}$ is an isolated point we have $\operatorname{clm}(f;\bar{x}) = 0$. When f is not calm at $\bar{x}$, from the definition we get $\operatorname{clm}(f;\bar{x}) = \infty$. In this way,

$$f \text{ is calm at } \bar{x} \iff \operatorname{clm}(f;\bar{x}) < \infty.$$

Examples.

1) The function $f(x) = x$ for $x \geq 0$ is calm at every point of its domain $[0, \infty)$, always with calmness modulus 1.

2) The function $f(x) = \sqrt{|x|}$, $x \in \mathbb{R}$ is not calm at zero but calm everywhere else.

3) The linear mapping $A : x \mapsto Ax$, where A is an $m \times n$ matrix, is calm at every point $x \in \mathbb{R}^n$ and everywhere has the same modulus $\operatorname{clm}(A;x) = |A|$.

Straight from the definition of the calmness modulus, we observe that

(i) $\operatorname{clm}(f;\bar{x}) \geq 0$ for every $\bar{x} \in \operatorname{dom} f$;

(ii) $\operatorname{clm}(\lambda f;\bar{x}) = |\lambda| \operatorname{clm}(f;\bar{x})$ for any $\lambda \in I\!R$ and $\bar{x} \in \operatorname{dom} f$;
(iii) $\operatorname{clm}(f+g;\bar{x}) \leq \operatorname{clm}(f;\bar{x}) + \operatorname{clm}(g;\bar{x})$ for any $\bar{x} \in \operatorname{dom} f \cap \operatorname{dom} g$.

These properties of the calmness modulus resemble those of a norm on a space of functions f, but because $\operatorname{clm}(f;\bar{x}) = 0$ does not imply $f = 0$, one could at most contemplate a seminorm. However, even that falls short, since the modulus can take on ∞, as can the functions themselves, which do not form a linear space because they need not even have the same domain.

Exercise 1C.1 (properties of the calmness modulus)**.** *Prove that*
(a) $\operatorname{clm}(f \circ g;\bar{x}) \leq \operatorname{clm}(f;g(\bar{x})) \cdot \operatorname{clm}(g;\bar{x})$ *whenever* $\bar{x} \in \operatorname{dom} g$ *and* $g(\bar{x}) \in \operatorname{dom} f$;
(b) $\operatorname{clm}(f-g;\bar{x}) = 0 \Rightarrow \operatorname{clm}(f;\bar{x}) = \operatorname{clm}(g;\bar{x})$ *whenever* $\bar{x} \in \operatorname{int}(\operatorname{dom} f \cap \operatorname{dom} g)$, *but the converse is false.*

With the concept of calmness in hand, we can interpret the differentiability of a function $f : I\!R^n \to I\!R^m$ at a point $\bar{x} \in \operatorname{int} \operatorname{dom} f$ as the existence of a linear mapping $A : I\!R^n \to I\!R^m$, represented by an $n \times m$ matrix, such that

$$\operatorname{clm}(e;\bar{x}) = 0 \quad \text{for } e(x) = f(x) - [f(\bar{x}) + A(x - \bar{x})]. \tag{2}$$

According to property (iii) before 1C.1 there is at most one mapping A satisfying (2). Indeed, if A_1 and A_2 satisfy (2) we have for the corresponding approximation error terms $e_1(x)$ and $e_2(x)$ that

$$|A_1 - A_2| = \operatorname{clm}(e_1 - e_2;\bar{x}) \leq \operatorname{clm}(e_1;\bar{x}) + \operatorname{clm}(e_2;\bar{x}) = 0.$$

Thus, A is unique and the associated matrix has to be the Jacobian $\nabla f(\bar{x})$. We conclude further from property (b) in 1C.1 that

$$\operatorname{clm}(f;\bar{x}) = |\nabla f(\bar{x})|.$$

The following theorem complements Theorem 1A.1. It shows that the invertibility of the derivative is a necessary condition to obtain a calm single-valued localization of the inverse.

Theorem 1C.2 (Jacobian nonsingularity from inverse calmness)**.** *Given* $f : I\!R^n \to I\!R^n$ *and* $\bar{x} \in \operatorname{int} \operatorname{dom} f$, *let* f *be differentiable at* $\bar{x}$ *and let* $\bar{y} := f(\bar{x})$. *If* f^{-1} *has a single-valued localization around* $\bar{y}$ *for* $\bar{x}$ *which is calm at* $\bar{y}$, *then the matrix* $\nabla f(\bar{x})$ *must be nonsingular.*

Proof. The assumption that f^{-1} has a calm single-valued localization s around $\bar{y}$ for $\bar{x}$ means several things: first, s is nonempty-valued around $\bar{y}$, that is, $\operatorname{dom} s$ is a neighborhood of $\bar{y}$; second, s is a function; and third, s is calm at $\bar{y}$. Specifically, there exist positive numbers a, b and κ and a function s with $\operatorname{dom} s \supset I\!B_b(\bar{y})$ and values $s(y) \in I\!B_a(\bar{x})$ such that for every $y \in I\!B_b(\bar{y})$ we have $s(y) = f^{-1}(y) \cap I\!B_a(\bar{x})$ and s is calm at $\bar{y}$ with constant κ. Taking b smaller if necessary we have

$$|s(y) - \bar{x}| \leq \kappa |y - \bar{y}| \quad \text{for every } y \in I\!B_b(\bar{y}). \tag{3}$$

Choose τ to satisfy $0 < \tau < 1/\kappa$. Then, since $\bar{x} \in \operatorname{int} \operatorname{dom} f$ and f is differentiable at $\bar{x}$, there exists $\delta > 0$ such that

$$(4) \qquad |f(x) - f(\bar{x}) - \nabla f(\bar{x})(x - \bar{x})| \le \tau |x - \bar{x}| \quad \text{for all } x \in I\!\!B_\delta(\bar{x}).$$

If the matrix $\nabla f(\bar{x})$ were singular, there would exist $d \in I\!\!R^n$, $|d| = 1$, such that $\nabla f(\bar{x})d = 0$. Pursuing this possibility, let ε satisfy $0 < \varepsilon < \min\{a, b/\tau, \delta\}$. Then, by applying (4) with $x = \bar{x} + \varepsilon d$, we get $f(\bar{x} + \varepsilon d) \in I\!\!B_b(\bar{y})$. In terms of $y_\varepsilon := f(\bar{x} + \varepsilon d)$, we then have $\bar{x} + \varepsilon d \in f^{-1}(y_\varepsilon) \cap I\!\!B_a(\bar{x})$, hence $s(y_\varepsilon) = \bar{x} + \varepsilon d$. The calmness condition (3) then yields

$$1 = |d| = \frac{1}{\varepsilon}|\bar{x} + \varepsilon d - \bar{x}| = \frac{1}{\varepsilon}|s(y_\varepsilon) - \bar{x}| \le \frac{\kappa}{\varepsilon}|y_\varepsilon - \bar{y}| = \frac{\kappa}{\varepsilon}|f(\bar{x} + \varepsilon d) - f(\bar{x})|.$$

Combining this with (4) and taking into account that $\nabla f(\bar{x})d = 0$, we arrive at $1 \le \kappa\tau|d| < 1$ which is absurd. Hence $\nabla f(\bar{x})$ is nonsingular. □

Note that in the particular case of an affine function $f(x) = Ax + b$, where A is a square matrix and b is a vector, calmness can be dropped from the set of assumptions of Theorem 1C.2; the existence of a single-valued localization of f^{-1} around any point is already equivalent to the nonsingularity of the Jacobian. This is not always true even for polynomials. Indeed, the inverse of $f(x) = x^3$, $x \in I\!\!R$, has a single-valued localization around the origin (which is not calm), but $\nabla f(0) = 0$.

Exercise 1C.3. *Using 1C.2 give a new proof of the symmetric inverse function theorem 1B.9.*

Exercise 1C.4. *Let $f : I\!\!R^n \to I\!\!R^n$ be differentiable at $\bar{x}$ and suppose that the correction mapping*

$$\Xi : x \mapsto \left\{ u \in I\!\!R^n \,\middle|\, f(x + u) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}) \right\} \quad \text{for } x \in I\!\!R^n$$

has a single-valued localization ξ around $\bar{x}$ for 0 such that ξ is calm at $\bar{x}$ with $\operatorname{clm}(\xi; \bar{x}) = 0$. *Prove that $\nabla f(\bar{x})$ is nonsingular.*

Guide. If $\nabla f(\bar{x})$ is singular, there must exist a vector $d \in I\!\!R^n$ with $|d| = 1$ such that $\nabla f(\bar{x})d = 0$. Then for all sufficiently small $\varepsilon > 0$ we have

$$f(\bar{x} + \varepsilon d + \xi(\bar{x} + \varepsilon d)) = f(\bar{x}).$$

Thus, $\varepsilon d + \xi(\bar{x} + \varepsilon d) \in \Xi(\bar{x})$ for all small $\varepsilon > 0$. Since Ξ has a single-valued localization around $\bar{x}$ for 0 we get $\varepsilon d + \xi(\bar{x} + \varepsilon d) = 0$. Then

$$1 = |d| = \frac{1}{\varepsilon}|\xi(\bar{x} + \varepsilon d)| \to 0 \qquad \text{as} \;\; \varepsilon \to 0,$$

a contradiction. □

Next, we extend the definition of calmness to its partial counterparts.

Partial calmness. *A function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ is said to be calm with respect to x at $(\bar{p},\bar{x}) \in \operatorname{dom} f$ when the function φ with values $\varphi(x) = f(\bar{p},x)$ is calm at $\bar{x}$. Such calmness is said to be uniform in p at $(\bar{p},\bar{x})$ when there exists a constant $\kappa \geq 0$ and neighborhoods Q of $\bar{p}$ and U of $\bar{x}$ such that actually*

$$|f(p,x) - f(p,\bar{x})| \leq \kappa |x-\bar{x}| \quad \text{for all } (p,x) \in (Q \times U) \cap \operatorname{dom} f.$$

Correspondingly, the partial calmness modulus of f with respect to x at $(\bar{p},\bar{x})$ is denoted as $\operatorname{clm}_x(f;(\bar{p},\bar{x}))$, while the uniform partial calmness modulus is

$$\widehat{\operatorname{clm}}_x(f;(\bar{p},\bar{x})) := \limsup_{\substack{x \to \bar{x},\, p \to \bar{p}, \\ (p,x) \in \operatorname{dom} f,\, x \neq \bar{x}}} \frac{|f(p,x) - f(p,\bar{x})|}{|x-\bar{x}|}$$

provided that every neighborhood of $(\bar{p},\bar{x})$ contains points $(p,x) \in \operatorname{dom} f$ with $x \neq \bar{x}$.

Observe in this context that differentiability of $f(p,x)$ with respect to x at $(\bar{p},\bar{x}) \in \operatorname{int} \operatorname{dom} f$ is equivalent to the existence of a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$, the partial derivative of f with respect to x at $(\bar{p},\bar{x})$, which satisfies

$$\operatorname{clm}(e;\bar{x}) = 0 \text{ for } e(x) = f(\bar{p},x) - [f(\bar{p},\bar{x}) + A(x-\bar{x})],$$

and then A is the partial derivative $D_x f(\bar{p},\bar{x})$. In contrast, under the stronger condition that

$$\widehat{\operatorname{clm}}_x(e;(\bar{p},\bar{x})) = 0, \text{ for } e(p,x) = f(p,x) - [f(\bar{p},\bar{x}) + A(x-\bar{x})],$$

we say f is *differentiable with respect to x uniformly in p* at $(\bar{p},\bar{x})$. This means that for every $\varepsilon > 0$ there are neighborhoods Q of $\bar{p}$ and U of $\bar{x}$ such that

$$|f(p,x) - f(\bar{p},\bar{x}) - D_x f(\bar{p},\bar{x})(x-\bar{x})| \leq \varepsilon |x-\bar{x}| \text{ for } p \in Q \text{ and } x \in U.$$

Exercise 1C.5 (joint calmness criterion)**.** *Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ be calm in x uniformly in p and calm in p, both at $(\bar{p},\bar{x})$. Show that f is calm at $(\bar{p},\bar{x})$.*

Exercise 1C.6 (nonsingularity characterization)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be differentiable at $\bar{x}$, let $\bar{y} = f(\bar{x})$, and suppose that f^{-1} has a single-valued localization s around $\bar{y}$ for $\bar{x}$ which is continuous at $\bar{y}$. Prove in this setting that s is differentiable at $\bar{y}$ if and only if the Jacobian $\nabla f(\bar{x})$ is nonsingular.*

Guide. The "only if" part can be obtained from Theorem 1C.2, using the fact that if s is differentiable at $\bar{x}$, it must be calm at $\bar{x}$. In the other direction, starting from the assumption that $\nabla f(\bar{x})$ is nonsingular, argue in a manner parallel to the first part of Step 3 of Proof I of Theorem 1A.1. □

1.4 [1D] Lipschitz Continuity

Calmness is a "one-point" version of the well-known "two-point" property of functions named after Rudolf Otto Sigismund Lipschitz (1832–1903). That property has already entered our deliberations in Section 1.1 [1A] in connection with the Proof II of the classical inverse function theorem by way of the contraction mapping. For convenience we recall the definition:

Lipschitz continuous functions. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be Lipschitz continuous relative to a set D, or on a set D, if $D \subset \operatorname{dom} f$ and there exists a constant $\kappa \geq 0$ (Lipschitz constant) such that*

$$(1) \qquad |f(x') - f(x)| \leq \kappa |x' - x| \quad \text{for all } x', x \in D.$$

It is said to be Lipschitz continuous around $\bar{x}$ when this holds for some neighborhood D of $\bar{x}$. We say further, in the case of an open set C, that f is locally Lipschitz continuous on C if it is a Lipschitz continuous function around every point x of C.

Lipschitz modulus. *For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and a point $\bar{x} \in \operatorname{int} \operatorname{dom} f$, the Lipschitz modulus of f at $\bar{x}$, denoted $\operatorname{lip}(f;\bar{x})$, is the infimum of the set of values of κ for which there exists a neighborhood D of $\bar{x}$ such that (1) holds. Equivalently,*

$$(2) \qquad \operatorname{lip}(f;\bar{x}) := \limsup_{\substack{x', x \to \bar{x}, \\ x \neq x'}} \frac{|f(x') - f(x)|}{|x' - x|}.$$

Note that, by this definition, for the Lipschitz modulus we have $\operatorname{lip}(f;\bar{x}) = \infty$ precisely in the case where, for every $\kappa > 0$ and every neighborhood D of $\bar{x}$, there are points $x', x \in D$ violating (1). Thus,

$$f \text{ is Lipschitz continuous around } \bar{x} \iff \operatorname{lip}(f;\bar{x}) < \infty.$$

A function f with $\operatorname{lip}(f;\bar{x}) < \infty$ is also called *strictly continuous* at $\bar{x}$. For an open set C, a function f is *locally Lipschitz continuous* on C exactly when $\operatorname{lip}(f;x) < \infty$ for every $x \in C$. Every continuously differentiable function on an open set C is locally Lipschitz continuous on C.

Examples.

1) The function $x \mapsto |x|$, $x \in \mathbb{R}^n$, is Lipschitz continuous everywhere with $\operatorname{lip}(|x|;x) = 1$; it is not differentiable at 0.

2) An affine function $f : x \mapsto Ax + b$, corresponding to a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$, has $\operatorname{lip}(f;\bar{x}) = |A|$ for every $\bar{x} \in \mathbb{R}^n$.

3) If f is continuously differentiable in a neighborhood of $\bar{x}$, then $\operatorname{lip}(f;\bar{x}) = |\nabla f(\bar{x})|$.

Like the calmness modulus, the Lipschitz modulus has the properties of a seminorm, except in allowing for ∞:

(i) $\operatorname{lip}(f;\bar{x}) \geq 0$ for every $\bar{x} \in \operatorname{int} \operatorname{dom} f$;
(ii) $\operatorname{lip}(\lambda f;\bar{x}) = |\lambda| \operatorname{lip}(f;\bar{x})$ for every $\lambda \in \mathbb{R}$ and $\bar{x} \in \operatorname{int} \operatorname{dom} f$;
(iii) $\operatorname{lip}(f+g;\bar{x}) \leq \operatorname{lip}(f;\bar{x}) + \operatorname{lip}(g;\bar{x})$ for every $\bar{x} \in \operatorname{int} \operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g$.

Exercise 1D.1 (properties of the Lipschitz modulus). *Prove that*
(a) $\operatorname{lip}(f \circ g;\bar{x}) \leq \operatorname{lip}(f;g(\bar{x})) \cdot \operatorname{lip}(g;\bar{x})$ *when* $\bar{x} \in \operatorname{int} \operatorname{dom} g$ *and* $g(\bar{x}) \in \operatorname{int} \operatorname{dom} f$;
(b) $\operatorname{lip}(f-g;\bar{x}) = 0 \Rightarrow \operatorname{lip}(f;\bar{x}) = \operatorname{lip}(g;\bar{x})$ *when* $\bar{x} \in \operatorname{int} \operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g$;
(c) $\operatorname{lip}(f;\cdot)$ *is upper semicontinuous at every* $\bar{x} \in \operatorname{int} \operatorname{dom} f$ *where it is finite;*
(d) *the set* $\left\{ x \in \operatorname{int} \operatorname{dom} f \,\middle|\, \operatorname{lip}(f;x) < \infty \right\}$ *is open.*

Bounds on the Lipschitz modulus lead to Lipschitz constants relative to sets, as long as convexity is present. First, recall that a set $C \subset \mathbb{R}^n$ is *convex* if

$$(1-\tau)x_0 + \tau x_1 \in C \quad \text{for all } \tau \in (0,1) \text{ when } x_0, x_1 \in C,$$

or in other words, if C contains for any pair of its points the entire line segment that joins them. The most obvious convex set is the ball $\mathbb{B}$ as well as its interior, while the boundary of the ball is of course nonconvex.

Exercise 1D.2 (Lipschitz continuity on convex sets). *Show that if C is a convex subset of* $\operatorname{int} \operatorname{dom} f$ *such that* $\operatorname{lip}(f;x) \leq \kappa$ *for all* $x \in C$, *then f is Lipschitz continuous relative to C with constant κ.*

Guide. It is enough to demonstrate for an arbitrary choice of points x and x' in C and $\varepsilon > 0$ that $|f(x') - f(x)| \leq (\kappa + \varepsilon)|x' - x|$. Argue that the line segment joining x and x' is a compact subset of $\operatorname{int} \operatorname{dom} f$ which can be covered by finitely many balls on which f is Lipschitz continuous with constant $\kappa + \varepsilon$. Moreover these balls can be chosen in such a way that a finite sequence of points $x_0, x_1, \ldots, x_r$ along the segment, starting with $x_0 = x$ and ending with $x_r = x'$, has each consecutive pair in one of them. Get the Lipschitz inequality for x and x' from the Lipschitz inequalities for these pairs. □

Exercise 1D.3 (Lipschitz continuity from differentiability). *If f is continuously differentiable on an open set O and C is a compact convex subset of O, then f is Lipschitz continuous relative to C with constant* $\kappa = \max_{x \in C} |\nabla f(x)|$.

Convexity also provides an important class of examples of Lipschitz continuous functions from $\mathbb{R}^n$ into itself which are not everywhere differentiable, namely distance and projection mappings; for an illustration see Fig. 1.5.

Distance and projection. *For a point $x \in \mathbb{R}^n$ and a set $C \subset \mathbb{R}^n$, the quantity*

$$d_C(x) = d(x,C) = \inf_{y \in C} |x - y| \tag{3}$$

is called the distance *from x to C. (Whether the notation $d_C(x)$ or $d(x,C)$ is used is a matter of convenience in a given context.) Any point y of C which is closest to x in* *the sense of achieving this distance is called a projection of x on C. The set of such projections is denoted by $P_C(x)$. Thus,*

$$P_C(x) = \operatorname*{argmin}_{y \in C} |x - y|. \tag{4}$$

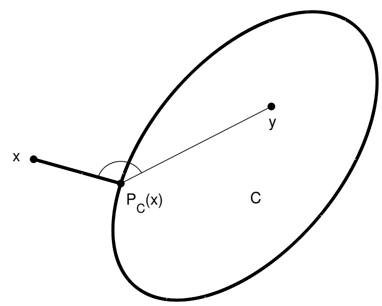

Fig. 1.5 Distance and projection.

In this way, C gives rise to a *distance function* d_C and a *projection mapping* P_C. If C is empty, then trivially $d_C(x) = \infty$ for all x, whereas if C is nonempty, then $d_C(x)$ is finite (and nonnegative) for all x. As for P_C, it is, in general, a set-valued mapping from $\mathbb{R}^n$ into C, but additional properties follow from particular assumptions on C, as we explore next.

Proposition 1D.4 (properties of distance and projection)**.**

(a) *For a nonempty set $C \subset \mathbb{R}^n$, one has $d_C(x) = d_{\operatorname{cl} C}(x)$ for all x. Moreover, C is closed if and only if every x with $d_C(x) = 0$ belongs to C.*

(b) *For a nonempty set $C \subset \mathbb{R}^n$, the distance function d_C is Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant $\kappa = 1$. As long as C is closed, one has*

$$\operatorname{lip}(d_C; \bar{x}) = \begin{cases} 0 & \text{if } \bar{x} \in \operatorname{int} C, \\ 1 & \text{otherwise.} \end{cases}$$

(c) *For a nonempty, closed set $C \subset \mathbb{R}^n$, the projection set $P_C(x)$ is nonempty, closed and bounded for every $x \in \mathbb{R}^n$.*

Proof. For (a), we fix any $x \in \mathbb{R}^n$ and note that $d_{\operatorname{cl} C}(x) \le d_C(x)$. This inequality can't be strict because for every $\varepsilon > 0$ we can find $y \in \operatorname{cl} C$ making $|x-y| < d_{\operatorname{cl} C}(x) + \varepsilon$ but then also find $y' \in C$ with $|y - y'| < \varepsilon$, in which case we have $d_C(x) \le |x - y'| < d_{\operatorname{cl} C}(x) + 2\varepsilon$. In particular, this argument reveals that $d_{\operatorname{cl} C}(x) = 0$ if and only if $x \in \operatorname{cl} C$. Having demonstrated that $d_C(x) = d_{\operatorname{cl} C}(x)$, we may conclude that $C = \{x \mid d_C(x) = 0\}$ if and only if $C = \operatorname{cl} C$.

For (b), consider any points x and x' along with any $\varepsilon > 0$. Take a point $y \in C$ such that $|x - y| \le d_C(x) + \varepsilon$. We have

$$d_C(x') \le |x'-y| \le |x'-x| + |x-y| \le |x'-x| + d_C(x) + \varepsilon,$$

and through the arbitrariness of ε therefore $d_C(x') - d_C(x) \le |x'-x|$. The same thing must hold with the roles of x and x' reversed, so this demonstrates that d_C is Lipschitz continuous with constant 1.

Let C be nonempty and closed. If $\bar{x} \in \operatorname{int} C$, we have $d_C(x) = 0$ for all x in a neighborhood of $\bar{x}$ and consequently $\operatorname{lip}(d_C;\bar{x}) = 0$. Suppose now that $\bar{x} \notin \operatorname{int} C$. We will show that $\operatorname{lip}(d_C;\bar{x}) \ge 1$, in which case equality must actually hold because we already know that d_C is Lipschitz continuous on $\mathbb{R}^n$ with constant 1. According to the property of the Lipschitz modulus displayed in Exercise 1D.1(c), it is sufficient to consider $\bar{x} \notin C$. Let $\tilde{x} \in P_C(\bar{x})$. Then on the line segment from $\tilde{x}$ to $\bar{x}$ the distance increases linearly, that is, $d_C(\tilde{x} + \tau(\bar{x} - \tilde{x})) = \tau d(\bar{x}, C)$ for $0 \le \tau \le 1$ (prove!). Hence, for $\tau, \tau' \in [0,1]$, the two points $x = \tilde{x} + \tau(\bar{x} - \tilde{x})$ and $x' = \tilde{x} + \tau'(\bar{x} - \tilde{x})$ we have $|d_C(x') - d_C(x)| = |\tau' - \tau||\tilde{x} - \bar{x}| = |x' - x|$. Note that $\bar{x}$ can be approached by such pairs of points and hence $\operatorname{lip}(d_C;\bar{x}) \ge 1$.

Turning now to (c), we again fix $x \in \mathbb{R}^n$ and choose a sequence of points $y_k \in C$ such that $|x - y_k| \to d_C(x)$ as $k \to \infty$. This sequence is bounded and therefore has an accumulation point y in C, inasmuch as C is closed. Since $|x - y_k| \ge d_C(x)$ for all k, it follows that $|x - y| = d_C(x)$. Thus, $y \in P_C(x)$, so $P_C(x)$ is not empty. Since by definition $P_C(x)$ is the intersection of C with the closed ball with center x and radius $d_C(x)$, it's clear that $P_C(x)$ is furthermore closed and bounded. □

It has been seen in 1D.4(c) that for any nonempty closed set $C \subset \mathbb{R}^n$ the projection mapping $P_C : \mathbb{R}^n \rightrightarrows C$ is nonempty-compact-valued, but when might it actually be single-valued as well? The convexity of C is the additional property that yields this conclusion, as will be shown in the following proposition[4].

Proposition 1D.5 (Lipschitz continuity of projection mappings). *For a nonempty, closed, convex set $C \subset \mathbb{R}^n$, the projection mapping P_C is single-valued (a function) from $\mathbb{R}^n$ onto C which moreover is Lipschitz continuous with Lipschitz constant $\kappa = 1$. Also,*

$$P_C(\bar{x}) = \bar{y} \iff \langle \bar{x} - \bar{y}, y - \bar{y} \rangle \le 0 \quad \text{for all } y \in C. \tag{5}$$

Proof. We have $P_C(x) \ne \emptyset$ in view of 1D.4(c). Suppose $\bar{y} \in P_C(\bar{x})$. For any $\tau \in (0,1)$, any $y \in \mathbb{R}^n$ and $y_\tau = (1-\tau)\bar{y} + \tau y$ we have the identity

$$\begin{aligned} |\bar{x} - y_\tau|^2 - |\bar{x} - \bar{y}|^2 &= |(y_\tau - \bar{y}) - (\bar{x} - \bar{y})|^2 - |\bar{x} - \bar{y}|^2 \\ &= |y_\tau - \bar{y}|^2 - 2\langle \bar{x} - \bar{y}, y_\tau - \bar{y} \rangle \\ &= \tau^2 |y - \bar{y}|^2 - 2\tau \langle \bar{x} - \bar{y}, y - \bar{y} \rangle. \end{aligned} \tag{6}$$

If $y \in C$, we also have $y_\tau \in C$ by convexity, so the left side is nonnegative. This implies that $\tau |y - \bar{y}|^2 \ge 2\langle \bar{x} - \bar{y}, y - \bar{y} \rangle$ for all $\tau \in (0,1)$. Thus, the inequality in (5)

[4] A set C such that P_C is single-valued is called a Chebyshev set. A nonempty, closed, convex set is always a Chebyshev set, and in $\mathbb{R}^n$ the converse is also true; for proofs of this fact see Borwein and Lewis [2006] and Deutsch [2001]. The question of whether a Chebyshev set in an arbitrary infinite-dimensional Hilbert space must be convex is still open.

holds. On the other hand, let $\langle \bar{x}-\bar{y}, y-\bar{y}\rangle \le 0$ for all $y \in C$. If $y \in C$ is such that $|\bar{x}-y| \le |\bar{x}-\bar{y}|$ then for $\tau \in (0,1)$ from (6) we get $\tau|y-\bar{y}|^2 \le 2\langle \bar{x}-\bar{y}, y-\bar{y}\rangle \le 0$ showing that $y=\bar{y}$. Thus (5) is fully confirmed along with the fact that $P_C(\bar{x})$ can't contain any $y \ne \bar{y}$.

Consider now two points x_0 and x_1 and their projections $y_0 = P_C(x_0)$ and $y_1 = P_C(x_1)$. On applying (5), we see that

$$\langle x_0 - y_0, y_1 - y_0\rangle \le 0 \quad\text{and}\quad \langle x_1 - y_1, y_0 - y_1\rangle \le 0.$$

When added, these inequalities give us

$$0 \ge \langle x_0 - y_0 - x_1 + y_1, y_1 - y_0\rangle = |y_1 - y_0|^2 - \langle x_1 - x_0, y_1 - y_0\rangle$$

and consequently

$$|y_1 - y_0|^2 \le \langle x_1 - x_0, y_1 - y_0\rangle \le |x_1 - x_0||y_1 - y_0|.$$

It follows that

$$|y_1 - y_0| \le |x_1 - x_0|.$$

Thus, P_C is Lipschitz continuous with Lipschitz constant 1. □

Projection mappings have many uses in numerical analysis and optimization. Note that P_C always fails to be differentiable on the boundary of C. As an example, when C is the set of nonpositive reals $\mathbb{R}_-$ one has

$$P_C(x) = \begin{cases} 0 & \text{for } x \ge 0, \\ x & \text{for } x < 0 \end{cases}$$

and this function is not differentiable at $x = 0$.

It is clear from the definitions of the calmness and Lipschitz moduli that we always have

$$\operatorname{clm}(f;\bar{x}) \le \operatorname{lip}(f;\bar{x}).$$

This relation is illustrated in Fig. 1.6.

In the preceding section we showed how to characterize differentiability through calmness. Now we introduce a sharper concept of derivative which is tied up with the Lipschitz modulus.

Strict differentiability. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be strictly differentiable at a point $\bar{x}$ if there is a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\operatorname{lip}(e;\bar{x}) = 0 \quad \textit{for } e(x) = f(x) - [f(\bar{x}) + A(x-\bar{x})].$$

In particular, in this case we have that $\operatorname{clm}(e;\bar{x}) = 0$ and hence f is differentiable at $\bar{x}$ with $A = \nabla f(\bar{x})$, but strictness imposes a requirement on the difference

$$e(x) - e(x') = f(x) - [f(x') + \nabla f(\bar{x})(x - x')]$$

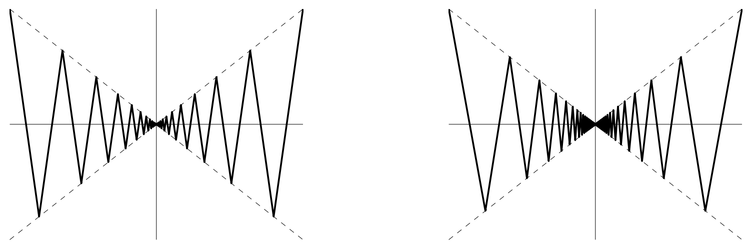

Fig. 1.6 Plots of a calm and a Lipschitz continuous function. On the left is the plot of the function $f(x) = (-1)^{n+1}9x + (-1)^n 2^{2n+1}/5^{n-2}$, $|x| \in [x_{n+1}, x_n]$ for $x_n = 4^{n-1}/5^{n-2}$, $n = 1, 2, \ldots$ for which $\operatorname{clm}(f;0) < \operatorname{lip}(f;0) < \infty$. On the right is the plot of the function $f(x) = (-1)^{n+1}(6+n)x + (-1)^n 210(5+n)!/(6+n)!$, $|x| \in [x_{n+1}, x_n]$ for $x_n = 210(4+n)!/(6+n)!$, $n = 1, 2, \ldots$ for which $\operatorname{clm}(f;0) < \operatorname{lip}(f;0) = \infty$.

also when $x' \neq \bar{x}$. Specifically, it demands the existence for each $\varepsilon > 0$ of a neighborhood U of $\bar{x}$ such that

$$|f(x) - [f(x') + \nabla f(\bar{x})(x - x')]| \leq \varepsilon |x - x'| \quad \text{for every } x, x' \in U.$$

Exercise 1D.6 (strict differentiability from continuous differentiability)**.** *Prove that every function f that is continuously differentiable in a neighborhood of $\bar{x}$ is strictly differentiable at $\bar{x}$.*

Guide. Adopt formula (b) in Fact 1 in the beginning of Section 1.1 [1A]. □

The converse to the assertion in Exercise 1D.6 is false, however: f can be strictly differentiable at $\bar{x}$ without being continuously differentiable around $\bar{x}$. This is demonstrated in Fig. 1.7 showing the graphs of two functions that are both differentiable at origin but otherwise have different properties. On the left is the graph of the continuous function $f : [-1, 1] \to I\!R$ which is even, and on $[0, 1]$ has values $f(0) = 0$, $f(1/n) = 1/n^2$, and is linear in the intervals $[1/n, 1/(n+1)]$. This function is strictly differentiable at 0, but in every neighborhood of 0 there are points where differentiability is lacking. On the right is the graph of the function[5]

$$f(x) = \begin{cases} x/2 + x^2 \sin(1/x) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

which is differentiable at 0 but not strictly differentiable there.

The second of these examples has the interesting feature that, even though $f(0) = 0$ and $f'(0) \neq 0$, no single-valued localization of f^{-1} exists around 0 for

[5] These two examples are from Nijenhuis [1974], where the introduction of strict differentiability is attributed to Leach [1961], but a recent paper by Dolecki and Grecco [2011] suggests that it goes back to Peano [1892]. By the way, Nijenhuis dedicated his paper to Carl Allendoerfer "for *not* taking the implicit function theorem for granted," which is the leading quotation in this book.

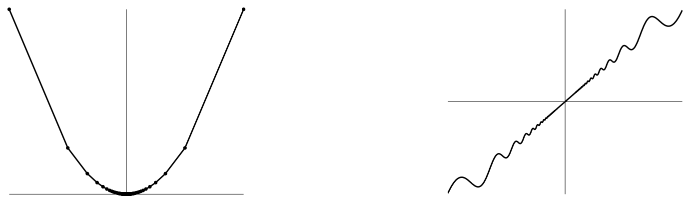

Fig. 1.7 Plots of functions differentiable at the origin. The function on the left is strictly differentiable at the origin but not continuously differentiable. The function on the right is differentiable at the origin but not strictly differentiable there.

0. In contrast, we will see in 1D.9 that strict differentiability would ensure the availability of such a localization.

Proposition 1D.7 (strict differentiability from differentiability)**.** *Consider a function f which is differentiable at every point in a neighborhood of $\bar{x}$. Prove that f is strictly differentiable at $\bar{x}$ if and only if the Jacobian ∇f is continuous at $\bar{x}$.*

Proof. Let f be strictly differentiable at $\bar{x}$ and let $\varepsilon > 0$. Then there exists $\delta_1 > 0$ such that for every $x_1, x_2 \in \mathbb{B}_{\delta_1}(\bar{x})$ we have

$$|f(x_2) - f(x_1) - \nabla f(\bar{x})(x_2 - x_1)| \leq \frac{1}{2}\varepsilon|x_1 - x_2|. \tag{7}$$

Fix an $x_1 \in \mathbb{B}_{\delta_1/2}(\bar{x})$. For this x_1 there exists $\delta_2 > 0$ such that for every $x' \in \mathbb{B}_{\delta_2}(x_1)$,

$$|f(x') - f(x_1) - \nabla f(x_1)(x' - x_1)| \leq \frac{1}{2}\varepsilon|x' - x_1|. \tag{8}$$

Make δ_2 smaller if necessary so that $\mathbb{B}_{\delta_2}(x_1) \subset \mathbb{B}_{\delta_1}(\bar{x})$. By (7) with x_2 replaced by x' and by (8), we have

$$|\nabla f(x_1)(x' - x_1) - \nabla f(\bar{x})(x' - x_1)| \leq \varepsilon|x' - x_1|.$$

This implies

$$|\nabla f(x_1) - \nabla f(\bar{x})| \leq \varepsilon.$$

Since x_1 is arbitrarily chosen in $\mathbb{B}_{\delta_1/2}(\bar{x})$, we obtain that the Jacobian is continuous at $\bar{x}$.

For the opposite direction, use Fact 1 in the beginning of Section 1.1 [1A]. □

Exercise 1D.8 (continuous differentiability from strict differentiability)**.** *Prove that a function f is strictly differentiable at every point of an open set O if and only if it is continuously differentiable on O.*

Guide. Apply 1D.7. □

With the help of the strict derivative we can obtain a new version of the classical inverse function theorem 1A.1.

Theorem 1D.9 (symmetric inverse function theorem under strict differentiability). *Let $f: \mathbb{R}^n \to \mathbb{R}^n$ be strictly differentiable at $\bar{x}$. Then the following are equivalent:*

(a) *$\nabla f(\bar{x})$ is nonsingular;*

(b) *f^{-1} has a single-valued localization s around $\bar{y} := f(\bar{x})$ for $\bar{x}$ which is strictly differentiable at $\bar{y}$. In that case, moreover, $\nabla s(\bar{y}) = \nabla f(\bar{x})^{-1}$.*

Proof. The implication (a) $\Rightarrow$ (b) can be accomplished by combining various pieces already present in the proofs of Theorem 1A.1, since strict differentiability of f at $\bar{x}$ gives us by definition the estimate (b) in Fact 1 in Section 1.1 [1A]. Parallel to Proof II of 1A.1 we find positive constants a and b and a single-valued localization of f^{-1} of the form

$$s : y \mapsto f^{-1}(y) \cap \mathbb{B}_a(\bar{x}) \quad \text{for } y \in \mathbb{B}_b(\bar{y}).$$

Next, by using the equation

$$s(y) = -A^{-1}(f(s(y)) - y - As(y)) \quad \text{for } y \in \mathbb{B}_b(\bar{y}),$$

where $A = \nabla f(\bar{x})$, we demonstrate Lipschitz continuity of s around $\bar{y}$ as in the beginning of Step 3 of Proof I of Theorem 1A.1. Finally, to obtain strict differentiability of s at $\bar{y}$, repeat the second part of Step 3 of Proof I with $\nabla f(s(y))$ replaced with A. For the converse implication we invoke Theorem 1C.2 and the fact that strict differentiability entails calmness. $\square$

Working now towards a corresponding version of the implicit function theorem, we look at additional forms of Lipschitz continuity and strict differentiability.

Partial Lipschitz continuity. *A function $f: \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ is said to be Lipschitz continuous with respect to x around $(\bar{p},\bar{x}) \in \operatorname{int} \operatorname{dom} f$ when the function $x \mapsto f(\bar{p},x)$ is Lipschitz continuous around $\bar{x}$; the associated Lipschitz modulus of f with respect to x is denoted by $\operatorname{lip}_x(f;(\bar{p},\bar{x}))$. We say f is Lipschitz continuous with respect to x uniformly in p around $(\bar{p},\bar{x}) \in \operatorname{int} \operatorname{dom} f$ when there are neighborhoods Q of $\bar{p}$ and U of $\bar{x}$ along with a constant $\kappa \geq 0$ and such that*

$$|f(p,x) - f(p,x')| \leq \kappa|x - x'| \quad \text{for all } x,x' \in U \text{ and } p \in Q.$$

Accordingly, the partial uniform Lipschitz modulus with respect to x has the form

$$\widehat{\operatorname{lip}}_x(f;(\bar{p},\bar{x})) := \limsup_{\substack{x,x' \to \bar{x},\, p \to \bar{p}, \\ x \neq x'}} \frac{|f(p,x') - f(p,x)|}{|x' - x|}.$$

Exercise 1D.10 (partial uniform Lipschitz modulus with differentiability). *Show that if the function $f: \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ is differentiable with respect to x at all points*

(p,x) *in some neighborhood of* $(\bar{p},\bar{x})$, *then*

$$\widehat{\mathrm{lip}}_x(f;(\bar{p},\bar{x})) = \limsup_{(p,x)\to(\bar{p},\bar{x})} |\nabla_x f(p,x)|.$$

Strict partial differentiability. *A function* $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ *is said to be strictly differentiable with respect to* x *at* $(\bar{p},\bar{x})$ *if the function* $x \mapsto f(\bar{p},x)$ *is strictly differentiable at* $\bar{x}$. *It is said to be strictly differentiable with respect to* x *uniformly in* p *at* $(\bar{p},\bar{x})$ *if*

$$\widehat{\mathrm{lip}}_x(e;(\bar{p},\bar{x})) = 0 \quad \text{for } e(p,x) = f(p,x) - [f(\bar{p},\bar{x}) + D_x f(\bar{p},\bar{x})(x-\bar{x})],$$

or in other words, if for every $\varepsilon > 0$ *there are neighborhoods* Q *of* $\bar{p}$ *and* U *of* $\bar{x}$ *such that*

$$|f(p,x) - [f(p,x') + D_x f(\bar{p},\bar{x})(x-x')]| \leq \varepsilon |x-x'| \quad \text{for all } x,x' \in U \text{ and } p \in Q.$$

Exercise 1D.11 (joint differentiability criterion)**.** *Let* $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ *be strictly differentiable with respect to* x *uniformly in* p *and be differentiable with respect to* p, *both at* $(\bar{p},\bar{x})$. *Prove that* f *is differentiable at* $(\bar{p},\bar{x})$.

Exercise 1D.12 (joint strict differentiability criterion)**.** *Prove that* $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ *is strictly differentiable at* $(\bar{p},\bar{x})$ *if and only if it is strictly differentiable with respect to* x *uniformly in* p *and strictly differentiable with respect to* p *uniformly in* x, *both at* $(\bar{p},\bar{x})$.

We state next the implicit function counterpart of Theorem 1D.9.

Theorem 1D.13 (implicit functions under strict partial differentiability)**.** *Given* $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ *and* $(\bar{p},\bar{x})$ *with* $f(\bar{p},\bar{x}) = 0$, *suppose that* f *is strictly differentiable at* $(\bar{p},\bar{x})$ *and let the partial Jacobian* $\nabla_x f(\bar{p},\bar{x})$ *be nonsingular. Then the solution mapping*

$$S : p \mapsto \left\{ x \in \mathbb{R}^n \,\middle|\, f(p,x) = 0 \right\}$$

has a single-valued localization s *around* $\bar{p}$ *for* $\bar{x}$ *which is strictly differentiable at* $\bar{p}$ *with its Jacobian expressed by*

$$\nabla s(\bar{p}) = -\nabla_x f(\bar{p},\bar{x})^{-1} \nabla_p f(\bar{p},\bar{x}).$$

Proof. We apply Theorem 1D.9 in a manner parallel to the way that the classical implicit function theorem 1B.1 was derived from the classical inverse function theorem 1A.1. $\square$

Exercise 1D.14. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be strictly differentiable at* 0 *and let* $f(0) \neq 0$. *Consider the following equation in* x *with a parameter* p:

$$pf(x) = \int_0^x f(pt)dt.$$

Prove that the solution mapping associated with this equation has a strictly differentiable single-valued localization around 0 *for* 0.

Guide. The function $g(p,x) = pf(x) - \int_0^x f(pt)dt$ satisfies $(\partial g/\partial x)(0,0) = -f(0)$, which is nonzero by assumption. For every $\varepsilon > 0$ there exist open intervals Q and U centered at 0 such that for every $p \in Q$ and $x, x' \in U$ we have

$$\begin{aligned}
&|g(p,x) - g(p,x') - \frac{\partial g}{\partial x}(0,0)(x - x')| \\
&\quad = |p(f(x) - f(x')) - \int_{x'}^{x} f(pt)dt + f(0)(x - x')| \\
&\quad = |p(f(x) - f(x')) - (f(p\tilde{x}) - f(0))(x - x')| \\
&\quad \le |p(f(x) - f(x') - f'(0)(x - x'))| + |pf'(0)(x - x')| \\
&\qquad + |(f(p\tilde{x}) - f(0))(x - x')| \le \varepsilon |x - x'|,
\end{aligned}$$

where the mean value theorem guarantees that $\int_{x'}^{x} f(pt)dt = (x - x')f(p\tilde{x})$ for some $\tilde{x}$ between x' and x. Hence, g is strictly differentiable with respect to x uniformly in p at $(0,0)$. Prove in a similar way that g is strictly differentiable with respect to p uniformly in x at $(0,0)$. Then apply 1D.12 and 1D.13. □

1.5 [1E] Lipschitz Invertibility from Approximations

In this section we completely depart from differentiation and develop inverse and implicit function theorems for equations in which the functions are merely Lipschitz continuous. The price to pay is that the single-valued localization of the inverse that is obtained might not be differentiable, but at least it will have a Lipschitz property.

The way to do that is found through notions of how a function f may be "approximated" by another function h around a point $\bar{x}$. Classical theory focuses on f being differentiable at $\bar{x}$ and approximated there by the function h giving its "linearization" at $\bar{x}$, namely $h(x) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x})$. Differentiability corresponds to having $f(x) = h(x) + o(|x - \bar{x}|)$ around $\bar{x}$, which is the same as $\operatorname{clm}(f - h; \bar{x}) = 0$, whereas strict differentiability corresponds to the stronger requirement that $\operatorname{lip}(f - h; \bar{x}) = 0$. The key idea is that conditions like this, and others in a similar vein, can be applied to f and h even when h is not a linearization dependent on the existence of $\nabla f(\bar{x})$. Assumptions on the nonsingularity of $\nabla f(\bar{x})$, corresponding in the classical setting to the invertibility of the linearization, might then be replaced by assumptions on the invertibility of some other approximation h.

First-order approximations of functions. *Consider a function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *and a point* $\bar{x} \in \operatorname{int} \operatorname{dom} f$. *A function* $h : \mathbb{R}^n \to \mathbb{R}^m$ *with* $\bar{x} \in \operatorname{int} \operatorname{dom} h$ *is a* ***first-order***

approximation to f at $\bar{x}$ if $h(\bar{x}) = f(\bar{x})$ and

$$\operatorname{clm}(e;\bar{x}) = 0 \quad \text{for } e(x) = f(x) - h(x),$$

which can also be written as $f(x) = h(x) + o(|x-\bar{x}|)$. It is a strict first-order approximation if the stronger condition holds that

$$\operatorname{lip}(e;\bar{x}) = 0 \quad \text{for } e(x) = f(x) - h(x).$$

In other words, h is a first-order approximation to f at $\bar{x}$ when $f(\bar{x}) = h(\bar{x})$ and for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|f(x) - h(x)| \leq \varepsilon |x - \bar{x}| \quad \text{for every } x \in I\!B_\delta(\bar{x}),$$

and a strict first-order approximation when

$$|[f(x) - h(x)] - [f(x') - h(x')]| \leq \varepsilon |x - x'| \quad \text{for all } x, x' \in I\!B_\delta(\bar{x}).$$

Clearly, if h is a (strict) first-order approximation to f, then f is a (strict) first-order approximation to h.

First-order approximations obey calculus rules which follow directly from the corresponding properties of the calmness and Lipschitz moduli:

(i) *If q is a (strict) first-order approximation to h at $\bar{x}$ and h is a (strict) first-order approximation to f at $\bar{x}$, then q is a (strict) first-order approximation to f at $\bar{x}$.*

(ii) *If f_1 and f_2 have (strict) first-order approximations h_1 and h_2, respectively, at $\bar{x}$, then $h_1 + h_2$ is a (strict) first-order approximation of $f_1 + f_2$ at $\bar{x}$.*

(iii) *If f has a (strict) first-order approximation h at $\bar{x}$, then for any $\lambda \in I\!R$, λh is a (strict) first-order approximation of λf at $\bar{x}$.*

(iv) *If h is a first-order approximation of f at $\bar{x}$, then $\operatorname{clm}(f;\bar{x}) = \operatorname{clm}(h;\bar{x})$. Similarly, if h is a strict first-order approximation of f at $\bar{x}$, then $\operatorname{lip}(f;\bar{x}) = \operatorname{lip}(h;\bar{x})$.*

The next proposition explains how first-order approximations can be chained together.

Proposition 1E.1 (composition of first-order approximations)**.** *Let h be a first-order approximation of f at $\bar{x}$ which is calm at $\bar{x}$. Let v be a first-order approximation of u at $\bar{y}$ for $\bar{y} := f(\bar{x})$ which is Lipschitz continuous around $\bar{y}$. Then $v \circ h$ is a first-order approximation of $u \circ f$ at $\bar{x}$.*

Proof. By the property (iv) of the first-order approximations displayed before the statement, the function f is calm at $\bar{x}$. Choose $\varepsilon > 0$ and let μ and λ be such that $\operatorname{clm}(f;\bar{x}) < \mu$ and $\operatorname{lip}(v;\bar{x}) < \lambda$. Then there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that $f(x) \in V$ and $h(x) \in V$ for $x \in U$,

$$|f(x) - h(x)| \leq \varepsilon |x - \bar{x}| \text{ and } |f(x) - f(\bar{x})| \leq \mu |x - \bar{x}| \quad \text{for } x \in U,$$

and moreover,

$$|u(y) - v(y)| \le \varepsilon |y - \bar{y}| \text{ and } |v(y) - v(y')| \le \lambda |y - y'| \quad \text{for } y, y' \in V.$$

Then for $x \in U$, taking into account that $\bar{y} = f(\bar{x}) = h(\bar{x})$, we have

$$\begin{aligned} |(u \circ f)(x) - (v \circ h)(x)| &= |u(f(x)) - v(h(x))| \\ &\le |u(f(x)) - v(f(x))| + |v(f(x)) - v(h(x))| \\ &\le \varepsilon |f(x) - f(\bar{x})| + \lambda |f(x) - h(x)| \\ &\le \varepsilon \mu |x - \bar{x}| + \lambda \varepsilon |x - \bar{x}| \le \varepsilon(\mu + \lambda)|x - \bar{x}|. \end{aligned}$$

Since ε can be arbitrarily small, the proof is complete. □

Exercise 1E.2 (strict approximations through composition)**.** *Let the function f satisfy $\operatorname{lip}(f;\bar{x}) < \infty$ and let the function g have a strict first-order approximation q at $\bar{y}$, where $\bar{y} := f(\bar{x})$. Then $q \circ f$ is a strict first-order approximation of $g \circ f$ at $\bar{x}$.*

Guide. Mimic the proof of 1E.1. □

For our purposes here, and in later chapters as well, first-order approximations offer an appealing substitute for differentiability, but an even looser notion of approximation will still lead to important conclusions.

Estimators beyond first-order approximations. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and a point $\bar{x} \in \operatorname{int} \operatorname{dom} f$. A function $h : \mathbb{R}^n \to \mathbb{R}^m$ with $\bar{x} \in \operatorname{int} \operatorname{dom} h$ is an* estimator *of f at $\bar{x}$ with constant μ if $h(\bar{x}) = f(\bar{x})$ and*

$$\operatorname{clm}(e;\bar{x}) \le \mu < \infty \quad \text{for } e(x) = f(x) - h(x),$$

which can also be written as $|f(x) - h(x)| \le \mu |x - \bar{x}| + o(|x - \bar{x}|)$. It is a strict estimator *if the stronger condition holds that*

$$\operatorname{lip}(e;\bar{x}) \le \mu < \infty \quad \text{for } e(x) = f(x) - h(x).$$

In this terminology, a first-order approximation is simply an estimator with constant $\mu = 0$. Through that, any result involving estimators can immediately be specialized to a result about first-order approximations.

Estimators can be of interest even when differentiability is present. For instance, in the case of a function f that is strictly differentiable at $\bar{x}$ a strict estimator of f at $\bar{x}$ with constant μ is furnished by $h(x) = f(\bar{x}) + A(x - \bar{x})$ for any matrix A with $|\nabla f(\bar{x}) - A| \le \mu$. Such relations have a role in certain numerical procedures, as will be seen at the end of this section and later in the book.

Theorem 1E.3 (inverse function theorem beyond differentiability)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a function with $\bar{x} \in \operatorname{int} \operatorname{dom} f$, and let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a strict estimator of f at $\bar{x}$ with constant μ. At the point $\bar{y} = f(\bar{x}) = h(\bar{x})$, suppose that h^{-1} has a Lipschitz continuous single-valued localization σ around $\bar{y}$ for $\bar{x}$ with $\operatorname{lip}(\sigma;\bar{y}) \le \kappa$ for a constant κ such that $\kappa\mu < 1$. Then f^{-1} has a Lipschitz continuous single-valued localization s around $\bar{y}$ for $\bar{x}$ with*

$$\operatorname{lip}(s;\bar{y}) \le \frac{\kappa}{1-\kappa\mu}.$$

This result is a particular case of the implicit function theorem 1E.13 presented later in this section, which in turn follows from a more general result proven in Chapter 2. It also goes back to an almost a ninety years old theorem by Hildebrand and Graves which is presented in the commentary to this chapter.

Proof. First, fix any

$$\lambda \in (\kappa,\infty) \text{ and } \nu \in (\mu,\kappa^{-1}) \text{ with } \lambda\nu < 1.$$

Without loss of generality, suppose that $\bar{x}=0$ and $\bar{y}=0$ and take a small enough that the mapping

$$y \mapsto h^{-1}(y) \cap a\mathbb{B} \quad \text{for } y \in a\mathbb{B}$$

is a localization of σ that is Lipschitz continuous with constant λ and also the difference $e = f - h$ is Lipschitz continuous on $a\mathbb{B}$ with constant ν. Next we choose α satisfying

$$0 < \alpha < \frac{1}{4}a(1-\lambda\nu)\min\{1,\lambda\}$$

and let $b := \alpha/(4\lambda)$. Pick any $y \in b\mathbb{B}$ and any $x^0 \in (\alpha/4)\mathbb{B}$; this gives us

$$|y - e(x^0)| \le |y| + |e(x^0) - e(0)| \le |y| + \nu|x^0| \le \frac{\alpha}{4\lambda} + \nu\frac{\alpha}{4} \le \frac{\alpha}{2\lambda}.$$

In particular $|y-e(x^0)| < a$, so the point $y - e(x^0)$ lies in the region where σ is Lipschitz continuous. Let $x^1 = \sigma(y-e(x^0))$; then

$$|x^1| = |\sigma(y-e(x^0))| = |\sigma(y-e(x^0)) - \sigma(0)| \le \lambda|y-e(x^0)| \le \frac{\alpha}{2},$$

so in particular x^1 belongs to the ball $a\mathbb{B}$. Furthermore,

$$|x^1 - x^0| \le |x^1| + |x^0| \le \alpha/2 + \alpha/4 = 3\alpha/4.$$

We also have

$$|y - e(x^1)| \le |y| + \nu|x^1| \le \alpha/(4\lambda) + \alpha/(2\lambda) \le a,$$

so that $y - e(x^1) \in a\mathbb{B}$.

Having started in this pattern, proceed to construct an infinite sequence of points x^k by taking

$$x^{k+1} = \sigma(y - e(x^k))$$

and prove by induction that

$$x^k \in a\mathbb{B}, \quad y - e(x^k) \in a\mathbb{B} \text{ and } |x^k - x^{k-1}| \le (\lambda\nu)^{k-1}|x^1 - x^0| \text{ for } k = 2,3,\ldots.$$

Observe next that, for $k > j$,

$$|x^k - x^j| \le \sum_{i=j}^{k-1} |x^{i+1} - x^i| \le \sum_{i=j}^{k-1} (\lambda\nu)^i a \le \frac{a}{1-\lambda\nu}(\lambda\nu)^j,$$

hence

$$\lim_{\substack{j,k\to\infty \\ k>j}} |x^k - x^j| = 0.$$

Then the sequence $\{x^k\}$ is Cauchy and hence convergent. Let x be its limit. Since all x^k and all $y - e(x^k)$ are in $a\mathbb{B}$, where both e and σ are continuous, we can pass to the limit in the equation $x^{k+1} = \sigma(y - e(x^k))$ as $k \to \infty$, getting

$$x = \sigma(y - e(x)), \text{ that is, } x \in f^{-1}(y).$$

According to our construction, we have

$$|x| \le \lambda(|y| + |e(x) - e(0)|) \le \lambda|y| + \lambda\nu|x|,$$

so that, since $|y| \le b$, we obtain

$$|x| \le \frac{\lambda b}{1-\lambda\nu}.$$

Thus, it is established that for every $y \in b\mathbb{B}$ there exists $x \in f^{-1}(y)$ with $|x| \le \lambda b/(1-\lambda\nu)$. In other words, we have shown the nonempty-valuedness of the localization of f^{-1} given by

$$s : y \mapsto f^{-1}(y) \cap \frac{\lambda b}{1-\lambda\nu}\mathbb{B} \quad \text{for } y \in b\mathbb{B}.$$

Next, demonstrate that this localization s is in fact single-valued and Lipschitz continuous. If for some $y \in b\mathbb{B}$ we have two points $x \ne x'$, both of them in $s(y)$, then subtracting $x = \sigma(y - e(x))$ from $x' = \sigma(y - e(x'))$ gives

$$0 < |x' - x| = |\sigma(y - e(x')) - \sigma(y - e(x))| \le \lambda|e(x') - e(x)| \le \lambda\nu|x' - x| < |x' - x|,$$

which is absurd. Further, considering $y', y \in b\mathbb{B}$ and recalling that $s(y) = \sigma(y - e(s(y)))$, one gets

$$|s(y') - s(y)| = |\sigma(y' - e(s(y'))) - \sigma(y - e(s(y)))| \le \lambda(|y' - y| + \nu|s(y') - s(y)|),$$

and hence s is Lipschitz continuous relative to $b\mathbb{B}$ with constant $\lambda/(1-\lambda\nu)$. This expression is continuous and increasing as a function of λ and ν, which are greater than κ and μ but can be chosen arbitrarily close to them, hence the Lipschitz modulus of s at $\bar{y}$ satisfies the desired inequality. □

Exercise 1E.4. *Prove Theorem 1E.3 by using the contraction mapping principle 1A.2.*

Guide. In parallel to Proof II of 1A.1, show that $\Phi_y(x) = \sigma\big(y - (f-h)(x)\big)$ has a fixed point. □

When $\mu = 0$ in Theorem 1E.3, so that h is a strict first-order approximation of f at $\bar{x}$, the conclusion about the localization s of f^{-1} is that $\operatorname{lip}(s;\bar{y}) \le \kappa$. The strict derivative version 1D.9 of the inverse function theorem corresponds to the case where $h(x) = f(\bar{x}) + Df(\bar{x})(x-\bar{x})$. The assumption on h^{-1} in Theorem 1E.3 is tantamount then to the invertibility of $Df(\bar{x})$, or equivalently the nonsingularity of the Jacobian $\nabla f(\bar{x})$; we have $\operatorname{lip}(\sigma;\bar{y}) = |Df(\bar{x})^{-1}| = |\nabla f(\bar{x})^{-1}|$, and κ can be taken to have this value. Again, though, Theorem 1E.3 does not, in general, insist on h being a first-order approximation of f at $\bar{x}$.

The following example sheds light on the sharpness of the assumptions in 1E.3 about the relative sizes of the Lipschitz modulus of the localization of h^{-1} and the Lipschitz modulus of the "approximation error" $f-h$.

Example 1E.5 (loss of single-valued localization without strict differentiability)**.** With $\alpha \in (0,\infty)$ as a parameter, let $f(x) = \alpha x + g(x)$ for the function

$$g(x) = \begin{cases} x^2 \sin(1/x) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

noting that f and g are differentiable with

$$g'(x) = \begin{cases} 2x\sin(1/x) - \cos(1/x) & \text{for } x \neq 0, \\ 0 & \text{for } x = 0, \end{cases}$$

but f and g are not strictly differentiable at 0, although g is Lipschitz continuous there with $\operatorname{lip}(g;0) = 1$.

Let $h(x) = \alpha x$ and consider applying Theorem 1E.3 to f and h at $\bar{x} = 0$, where $f(0) = h(0) = 0$. Since $f - h = g$, we have for every $\mu > 1$ that $f-h$ is Lipschitz continuous with constant μ on some neighborhood of 0. On the other hand, h^{-1} is Lipschitz continuous with constant $\kappa = 1/\alpha$. Therefore, as long as $\alpha > 1$, the assumptions of Theorem 1E.3 are fulfilled (by taking μ in $(1,\alpha)$ arbitrarily close to 1). We are able to conclude from 1E.3 that f^{-1} has a single-valued localization s around 0 for 0 such that $\operatorname{lip}(s;0) \le (\alpha-1)^{-1}$, despite the inapplicability of 1D.9.

When $0 < \alpha < 1$, however, f^{-1} has no single-valued localization around the origin at all. This comes out of the fact that, for such α, the derivative $f'(x)$ has infinitely many changes of sign in every neighborhood of $\bar{x} = 0$, hence infinitely many consecutive local maximum values and minimum values of f in any such neighborhood, with both values tending to 0 as the origin is approached. Let x_1 and x_2, $0 < x_1 < x_2$, be two consecutive points where f' vanishes and f has a local maximum at x_1 and local minimum at x_2. For a value $y > 0$ the equation $f(x) = y$ must have not only a solution in (x_1,x_2), but also one in $(0,x_1)$. Hence, regardless of the size of the neighborhood U of $\bar{x} = 0$, there will be infinitely many y values near $\bar{y} = 0$ for which $U \cap f^{-1}(y)$ is not a singleton. Both cases are illustrated in Fig. 1.8.

Fig. 1.8 Graphs of the function f in Example 1E.5 when $\alpha = 2$ on the left and $\alpha = 0.5$ on the right.

When the approximation h to f in Theorem 1E.3 is itself strictly differentiable at the point in question, a simpler statement of the result can be made.

Corollary 1E.6 (estimators with strict differentiability)**.** *For $f : \mathbb{R}^n \to \mathbb{R}^n$ with $\bar{x} \in \operatorname{int}\operatorname{dom} f$ and $f(\bar{x}) = \bar{y}$, suppose there is a strict estimator $h : \mathbb{R}^n \to \mathbb{R}^n$ of f at $\bar{x}$ with constant μ which is strictly differentiable at $\bar{x}$ with nonsingular Jacobian $\nabla h(\bar{x})$ satisfying $\mu|\nabla h(\bar{x})^{-1}| < 1$. Then f^{-1} has a Lipschitz continuous single-valued localization s around $\bar{y}$ for $\bar{x}$ with*

$$\operatorname{lip}(s;\bar{y}) \leq \frac{|\nabla h(\bar{x})^{-1}|}{1 - \mu|\nabla h(\bar{x})^{-1}|}.$$

Proof. A localization σ of h^{-1} having $\operatorname{lip}(\sigma;\bar{y}) = |\nabla h(\bar{x})^{-1}|$ is obtained by applying Theorem 1D.9 to h. Theorem 1E.3 can be invoked then with $\kappa = |\nabla h(\bar{x})^{-1}|$. □

An even more special application, to the case where both f and h are linear, yields a well-known estimate for matrices.

Corollary 1E.7 (estimation for perturbed matrix inversion)**.** *Let A and B be $n \times n$ matrices such that A is nonsingular and $|B| < |A^{-1}|^{-1}$. Then $A + B$ is nonsingular with*

$$|(A+B)^{-1}| \leq \left(|A^{-1}|^{-1} - |B|\right)^{-1}.$$

Proof. This comes from Corollary 1E.6 by taking $f(x) = (A+B)x$, $h(x) = Ax$, $\bar{x} = 0$, and writing $\left(|A^{-1}|^{-1} - |B|\right)^{-1}$ as $|A^{-1}|/(1 - |A^{-1}||B|)$. □

We can state 1E.7 in other two equivalent ways, which we give next as an exercise. More about the estimate in 1E.7 will be said in Chapter 5.

Exercise 1E.8 (equivalent estimation rules for matrices)**.** *Prove that the following two statements are equivalent to Corollary 1E.7:*

(a) *For any $n \times n$ matrix C with $|C| < 1$, the matrix $I + C$ is nonsingular and*

$$(1) \qquad |(I+C)^{-1}| \leq \frac{1}{1 - |C|}.$$

(b) *For any* $n \times n$ *matrices* A *and* B *with* A *nonsingular and* $|BA^{-1}| < 1$, *the matrix* $A+B$ *is nonsingular and*

$$|(A+B)^{-1}| \leq \frac{|A^{-1}|}{1-|BA^{-1}|}.$$

Guide. Clearly, (b) implies Corollary 1E.7 which in turn implies (a) with $A = I$ and $B = C$. Let (a) hold and let the matrices A and B satisfy $|BA^{-1}| < 1$. Then, by (a) with $C = BA^{-1}$ we obtain that $I + BA^{-1}$ is nonsingular, and hence $A+B$ is nonsingular, too. Using the equality $A+B = (I+BA^{-1})A$ in (1) we have

$$|(A+B)^{-1}| = |A^{-1}(I+BA^{-1})^{-1}| \leq \frac{|A^{-1}|}{1-|BA^{-1}|},$$

that is, (a) implies (b). □

Corollary 1E.7 implies that, given a nonsingular matrix A,

$$\inf\left\{\, |B| \,\middle|\, A+B \text{ is singular } \right\} \geq |A^{-1}|^{-1}. \tag{2}$$

It turns out that this inequality is actually equality, another classical result in matrix perturbation theory.

Theorem 1E.9 (radius theorem for matrix nonsingularity)**.** *For any nonsingular matrix* A,

$$\inf\left\{\, |B| \,\middle|\, A+B \textit{ is singular } \right\} = |A^{-1}|^{-1}.$$

Proof. It is sufficient to prove the inequality opposite to (2). Choose $\bar{y} \in I\!R^n$ with $|\bar{y}| = 1$ and $|A^{-1}\bar{y}| = |A^{-1}|$. For $\bar{x} = A^{-1}\bar{y}$ we have $|\bar{x}| = |A^{-1}|$. The matrix

$$B = -\frac{\bar{y}\bar{x}^{\mathsf{T}}}{|\bar{x}|^2}$$

satisfies

$$|B| = \max_{|x|=1} \frac{|\bar{y}\bar{x}^{\mathsf{T}}x|}{|\bar{x}|^2} \leq \max_{|x|=1} \frac{|\bar{x}^{\mathsf{T}}x|}{|\bar{x}|^2} = \frac{|\bar{x}^{\mathsf{T}}\bar{x}|}{|\bar{x}|^3} = \frac{1}{|\bar{x}|} = |A^{-1}|^{-1}.$$

On the other hand $(A+B)\bar{x} = A\bar{x} - \bar{y} = 0$, and since $\bar{x} \neq 0$, the matrix $A+B$ is singular. Thus the infimum in (2) is not greater than $|A^{-1}|^{-1}$. □

Exercise 1E.10 (radius theorem for function invertibility)**.** *Consider a function* $f : I\!R^n \to I\!R^n$ *and a point* $\bar{x} \in \operatorname{int} \operatorname{dom} f$. *Call* f *smoothly locally invertible at* $\bar{x}$, *for short, when* f^{-1} *has a Lipschitz continuous single-valued localization around* $f(\bar{x})$ *for* $\bar{x}$ *which is strictly differentiable at* $f(\bar{x})$. *In this terminology, prove in the case of* f *being strictly differentiable at* $\bar{x}$ *with its Jacobian* $\nabla f(\bar{x})$ *nonsingular, that*

$$\inf\left\{\, |B| \,\middle|\, f+B \textit{ is not smoothly locally invertible at } \bar{x} \,\right\} = |A^{-1}|^{-1},$$

where $A = \nabla f(\bar{x})$ and the infimum is taken over all linear mappings $B : \mathbb{R}^n \to \mathbb{R}^n$.

Guide. Combine 1E.9 and 1D.9. □

We will extend the facts in 1E.9 and 1E.10 to the much more general context of set-valued mappings in Chapter 6.

In the case of Theorem 1E.3 with $\mu = 0$, an actual equivalence emerges between the invertibility of f and that of h, as captured by the following statement. The key is the fact that first-order approximation is a *symmetric* relationship between two functions.

Theorem 1E.11 (Lipschitz invertibility with first-order approximations)**.** *Let $h : \mathbb{R}^n \to \mathbb{R}^n$ be a strict first-order approximation to $f : \mathbb{R}^n \to \mathbb{R}^n$ at $\bar{x} \in \operatorname{int} \operatorname{dom} f$, and let $\bar{y}$ denote the common value $f(\bar{x}) = h(\bar{x})$. Then f^{-1} has a Lipschitz continuous single-valued localization s around $\bar{y}$ for $\bar{x}$ if and only if h^{-1} has such a localization σ around $\bar{y}$ for $\bar{x}$, in which case*

$$\operatorname{lip}(s;\bar{y}) = \operatorname{lip}(\sigma;\bar{y}), \tag{3}$$

and moreover σ is then a first-order approximation to s at $\bar{y}$: $s(y) = \sigma(y) + o(|y-\bar{y}|)$.

Proof. Applying Theorem 1E.3 with $\mu = 0$ and $\kappa = \operatorname{lip}(\sigma;\bar{y})$, we see that f^{-1} has a single-valued localization s around $\bar{y}$ for $\bar{x}$ with $\operatorname{lip}(s;\bar{y}) \le \operatorname{lip}(\sigma;\bar{y})$. In these circumstances, though, the symmetry in the relation of first-order approximation allows us to conclude from the existence of this s that h^{-1} has a single-valued localization σ' around $\bar{y}$ for $\bar{x}$ with $\operatorname{lip}(\sigma';\bar{y}) \le \operatorname{lip}(s;\bar{y})$. The two localizations of h have to agree graphically in a neighborhood of $(\bar{y},\bar{x})$, so we can simply speak of σ and conclude the validity of (3).

To argue that σ is a first-order approximation of s, we begin by using the identity $h(\sigma(y)) = y$ to get $f(\sigma(y)) = y + e(\sigma(y))$ for the function $e = f - h$ and then transform that into

$$\sigma(y) = s(y + e(\sigma(y))) \quad \text{for } y \text{ near } \bar{y}. \tag{4}$$

Let $\kappa > \operatorname{lip}(s;\bar{y})$. From (3) there exists $b > 0$ such that

$$\max\Big\{ |\sigma(y) - \sigma(y')|, |s(y) - s(y')| \Big\} \le \kappa|y-y'| \quad \text{for } y,y' \in \mathbb{B}_b(\bar{y}). \tag{5}$$

Because $e(\bar{x}) = 0$ and $\operatorname{lip}(e;\bar{x}) = 0$, we know that for every $\varepsilon > 0$ there exists a positive $a > 0$ for which

$$|e(x) - e(x')| \le \varepsilon|x-x'| \quad \text{for all } x,x' \in \mathbb{B}_a(\bar{x}). \tag{6}$$

Choose

$$0 < \beta \le \min\Big\{ \frac{a}{\kappa}, \frac{b}{(1+\varepsilon\kappa)} \Big\}.$$

Then, for every $y \in \mathbb{B}_\beta(\bar{y})$ from (5) we have

$$|\sigma(y)-\bar{x}| \le \kappa\beta \le a,$$

and

$$|y+e(\sigma(y))-\bar{y}| \le |y-\bar{y}|+\varepsilon|\sigma(y)-\bar{x}| \le \beta+\varepsilon\kappa\beta \le b.$$

Hence, utilizing (4), (5) and (6), we obtain

$$\begin{aligned} |\sigma(y)-s(y)| &= |s(y+e(\sigma(y)))-s(y)| \\ &\le \kappa|e(\sigma(y))-e(\sigma(\bar{y}))| \\ &\le \kappa\varepsilon|\sigma(y)-\sigma(\bar{y})| \le \kappa^2\varepsilon|y-\bar{y}|. \end{aligned}$$

Since ε can be arbitrarily small, we arrive at the equality $\operatorname{clm}(s-\sigma;\bar{y})=0$, and this completes the proof. □

Finally, we observe that these results make it possible to deduce a slightly sharper version of the equivalence in Theorem 1D.9.

Theorem 1E.12 (extended equivalence under strict differentiability)**.** *Let $f:\mathbb{R}^n \to \mathbb{R}^n$ be strictly differentiable at $\bar{x}$ with $f(\bar{x})=\bar{y}$. Then the following are equivalent:*

(a) $\nabla f(\bar{x})$ *is nonsingular;*

(b) f^{-1} *has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$;*

(c) f^{-1} *has a single-valued localization around $\bar{y}$ for $\bar{x}$ that is strictly differentiable at $\bar{y}$.*

In parallel to Theorem 1E.3, it is possible also to state and prove a corresponding implicit function theorem with Lipschitz continuity. For that purpose, we need to introduce the concept of partial first-order approximations for functions of two variables.

Partial first-order estimators and approximations. *For $f:\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^m$ and a point $(\bar{p},\bar{x})\in\operatorname{int}\operatorname{dom} f$, a function $h:\mathbb{R}^n\to\mathbb{R}^m$ is said to be an estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ if $h(\bar{x})=f(\bar{x},\bar{p})$ and*

$$\widehat{\operatorname{clm}}_x(e;(\bar{p},\bar{x})) \le \mu < \infty \quad \text{for } e(p,x)=f(p,x)-h(x).$$

It is a strict estimator in this sense if the stronger condition holds that

$$\widehat{\operatorname{lip}}_x(e;(\bar{p},\bar{x})) \le \mu < \infty \quad \text{for } e(p,x)=f(p,x)-h(x).$$

In the case of $\mu=0$, such an estimator is called a partial first-order approximation.

Theorem 1E.13 (implicit function theorem beyond differentiability)**.** *Consider $f:\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^n$ and $(\bar{p},\bar{x})\in\operatorname{int}\operatorname{dom} f$ with $f(\bar{p},\bar{x})=0$ and $\widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x}))\le\gamma<\infty$. Let h be a strict estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ. Suppose at the point $h(\bar{x})=0$ that h^{-1} has a Lipschitz continuous single-valued localization σ around 0 for $\bar{x}$ with $\operatorname{lip}(\sigma;0)\le\kappa$ for a constant κ such that $\kappa\mu<1$. Then the solution mapping*

$$S : p \mapsto \left\{ x \in \mathbb{R}^n \,\middle|\, f(p,x) = 0 \right\} \qquad \text{for } p \in \mathbb{R}^d$$

has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(s;\bar{p}) \le \frac{\kappa\gamma}{1-\kappa\mu}.$$

Moreover, when $\mu = 0$ the function $\eta(p) = \sigma(-f(p,\bar{x}))$ is a first-order approximation to s at $\bar{p}$: $s(p) = \eta(p) + o(|p-\bar{p}|)$.

This is a special instance of the combination of Theorem 2B.7 (for $\mu > 0$) and of Theorem 2B.8 (for $\mu = 0$) which we will prove in Chapter 2, so there is no purpose in giving a separate argument for it here. If $f(p,x)$ has the form $f(x) - p$, in which case $\widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})) = 1$, taking $\gamma = 1$ we immediately obtain Theorem 1E.3.

Exercise 1E.14 (approximation criteria)**.** *Consider $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ with $f(\bar{p},\bar{x}) = h(\bar{p},\bar{x})$, and the difference $e(p,x) = f(p,x) - h(p,x)$. Prove that*

(a) *If $\widehat{\operatorname{clm}}_x(e;(\bar{p},\bar{x})) = 0$ and $\widehat{\operatorname{clm}}_p(e;(\bar{p},\bar{x})) = 0$, then h is a first-order approximation to f at $(\bar{p},\bar{x})$.*

(b) *If $\widehat{\operatorname{lip}}_x(e;(\bar{p},\bar{x})) = 0$ and $\operatorname{clm}_p(e;(\bar{p},\bar{x})) = 0$, then h is a first-order approximation to f at $(\bar{p},\bar{x})$.*

(c) *If $\widehat{\operatorname{lip}}_x(e;(\bar{p},\bar{x})) = 0$ and $\widehat{\operatorname{lip}}_p(e;(\bar{p},\bar{x})) = 0$, then h is a strict first-order approximation to f at $(\bar{p},\bar{x})$.*

Exercise 1E.15 (partial first-order approximation from differentiability)**.** *Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ be differentiable with respect to x in a neighborhood of $(\bar{p},\bar{x})$ and let f and $\nabla_x f$ be continuous in this neighborhood. Prove that the function $h(x) = f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x})$ is a strict first-order approximation to f with respect to x uniformly in p at $(\bar{p},\bar{x})$. Based on this, derive the Dini classical implicit function theorem 1B.1 from 1E.13.*

Exercise 1E.16 (the zero function as an approximation)**.** *Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ satisfy $\widehat{\operatorname{lip}}_x(f;(\bar{p},\bar{x})) < \infty$, and let $u : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^k$ have a strict first-order approximation v with respect to y at $(\bar{p},\bar{y})$, where $\bar{y} := f(\bar{p},\bar{x})$. Show that the zero function is a strict first-order approximation with respect to x at $(\bar{p},\bar{x})$ to the function $(p,x) \mapsto u(p,f(p,x)) - v(f(p,x))$.*

As another illustration of applicability of Theorem 1E.3 beyond first-order approximations, we sketch now a proof of the quadratic convergence of Newton's method for solving equations, a method we used in Proof I of the classical inverse function theorem 1A.1.

Consider the equation $g(x) = 0$ for a continuously differentiable function $g : \mathbb{R}^n \to \mathbb{R}^n$ with a solution $\bar{x}$ at which the Jacobian $\nabla g(\bar{x})$ is nonsingular. Newton's method consists in choosing a starting point x^0 possibly close to $\bar{x}$ and generating a sequence of points $x^1, x^2, \ldots$ according to the rule

(7) $$x^{k+1} = x^k - \nabla g(x^k)^{-1} g(x^k), \quad k = 0, 1, \ldots.$$

According to the classical inverse function theorem 1A.1, g^{-1} has a smooth single-valued localization around 0 for $\bar{x}$. Consider the function $f(x) = \nabla g(x^0)(x - \bar{x})$ for which $f(\bar{x}) = g(\bar{x}) = 0$ and $f(x) - g(x) = -g(x) + \nabla g(x^0)(x - \bar{x})$. An easy calculation shows that the Lipschitz modulus of $e = f - g$ at $\bar{x}$ can be made arbitrarily small by making x^0 close to $\bar{x}$. However, *this modulus must be nonzero* — but less than $|\nabla g(\bar{x})^{-1}|^{-1}$, the Lipschitz modulus of the single-valued localization of g^{-1} around 0 for $\bar{x}$ — if one wants to choose x^0 as an arbitrary starting point from an open neighborhood of $\bar{x}$. Here Theorem 1E.3 comes into play with $h = g$ and $\bar{y} = 0$, saying that f^{-1} has a Lipschitz continuous single-valued localization s around 0 for $\bar{x}$ with Lipschitz constant, say, μ. (In the simple case considered this also follows directly from the fact that if $\nabla g(\bar{x})$ is nonsingular at $\bar{x}$, then $\nabla g(x)$ is likewise nonsingular for all x in a neighborhood of $\bar{x}$, see Fact 2 in Section 1.1 [1A].) Hence, the Lipschitz constant μ and the neighborhood V of 0 where s is Lipschitz continuous can be determined before the choice of x^0, which is to be selected so that $\nabla g(x^0)(x^0 - \bar{x}) - g(x^0)$ is in V. Noting for the iteration (7) that $x^1 = s(\nabla g(x^0)(x^0 - \bar{x}) - g(x^0))$ and $\bar{x} = s(g(\bar{x}))$, and using the smoothness of g, we obtain

$$|x^1 - \bar{x}| \le \mu |g(\bar{x}) - g(x^0) - \nabla g(x^0)(\bar{x} - x^0)| \le c|x^0 - \bar{x}|^2$$

for a suitable constant c. This kind of argument works for any k, and in that way, through induction, we obtain quadratic convergence for Newton's iteration (7).

In Chapter 6 we will present, in a broader framework of generalized equations in possibly infinite-dimensional spaces, a detailed proof of this quadratic convergence of Newton's method and study its stability properties.

1.6 [1F] Selections of Multi-valued Inverses

Consider a function f which acts between Euclidean spaces of possibly different dimensions, say $f : \mathbb{R}^n \to \mathbb{R}^m$. What can be said then about the inverse mapping f^{-1}? The case of $f(x) = Ax + b$ with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ gives an indication of the differences that must be expected: when $m < n$, the equation $Ax + b = y$ either has no solution x or a continuum of solutions, so that the existence of single-valued localizations is totally hopeless. Anyway, though, if A has full rank m, we know at least that $f^{-1}(y)$ will be nonempty for every y.

Do we really always have to assume that $m = n$ in order to get a single-valued localization of the inverse? Specifically, consider a function f acting from $\mathbb{R}^n$ to $\mathbb{R}^m$ with $m \le n$ and a point $\bar{x}$ in the interior of the domain of f. Suppose that f is continuous in an open neighborhood U of $\bar{x}$ and the inverse f^{-1} has a single-valued localization around $f(\bar{x})$ for $\bar{x}$. It turns out that in this case we necessarily must have

$m = n$. This is due to a basic result in topology, the following version of which, stated here without proof, will serve our purposes.

Theorem 1F.1 (Brouwer invariance of domain theorem). *Let $O \subset \mathbb{R}^n$ be open and for $m \leq n$ let $f : O \to \mathbb{R}^m$ be continuous and such that f^{-1} is single-valued on $f(O)$. Then $f(O)$ is open, f^{-1} is continuous on $f(O)$, and $m = n$.*

This topological result reveals that, for continuous functions f, the dimension of the domain space has to agree with the dimension of the range space, if there is to be any hope of an inverse function theorem claiming the existence of a single-valued localization of f^{-1}. Of course, in the theorems already viewed for differentiable functions f, the dimensions were forced to agree because of a rank condition on the Jacobian matrix, but we see now that this limitation has a deeper source than a matrix condition.

Brouwer's invariance of domain theorem 1F.1 helps us to obtain the following characterization of the existence of a Lipschitz continuous single-valued localization of the inverse:

Theorem 1F.2 (invertibility characterization). *For a function $f : \mathbb{R}^n \to \mathbb{R}^n$ that is continuous around $\bar{x}$, the inverse f^{-1} has a Lipschitz continuous single-valued localization around $f(\bar{x})$ for $\bar{x}$ if and only if, in some neighborhood U of $\bar{x}$, there is a constant $c > 0$ such that*

$$(1) \qquad c|x' - x| \leq |f(x') - f(x)| \quad \text{for all } x', x \in U.$$

Proof. Let (1) hold. There is no loss of generality in supposing that U is open and f is continuous on U. For any $y \in f(U) := \{\, f(x) \,|\, x \in U \,\}$, we have from (1) that if both $f(x) = y$ and $f(x') = y$ with x and x' in U, then $x = x'$; in other words, the mapping s which takes $y \in f(U)$ to $U \cap f^{-1}(y)$ is actually a function on $f(U)$. Moreover $|s(y') - s(y)| \leq (1/c)|y' - y|$ by (1), so that this function is Lipschitz continuous relative to $f(U)$.

But this is not yet enough to get us to the desired conclusion about f^{-1}. For that, we need to know that s, or some restriction of s, is a localization of f^{-1} around $f(\bar{x})$ for $\bar{x}$, with graph equal to $(V \times U) \cap \operatorname{gph} f^{-1}$ for some neighborhood V of $f(\bar{x})$. Brouwer's invariance of domain theorem 1F.1 enters here: as applied to the restriction of f to the open set U, it tells us that $f(U)$ is open. We can therefore take $V = f(U)$ and be done.

Conversely, suppose that f^{-1} has a Lipschitz continuous single-valued localization around $f(\bar{x})$ for $\bar{x}$, its domain being a neighborhood V of $f(\bar{x})$. Let κ be a Lipschitz constant for s on V. Because f is continuous around $\bar{x}$, there is a neighborhood U of $\bar{x}$ such that $f(U) \subset V$. For x and x' in U, we have $s(f(x)) = x$ and $s(f(x')) = x'$, hence $|x' - x| \leq \kappa |f(x') - f(x)|$. Thus, (1) holds for any $c > 0$ small enough such that $c\kappa \leq 1$. $\square$

We will now re-prove the classical inverse function theorem in a somewhat different formulation having an important extra feature, which is here derived from Brouwer's invariance of domain theorem 1F.1.

A continuously differentiable function f acting between some open sets U and V in $\mathbb{R}^n$ and having the property that the inverse mapping f^{-1} is continuously differentiable, is called a *diffeomorphism* (or $\mathscr{C}^1$ diffeomorphism) between U and V. The theorem stated next says that when $\nabla f(\bar{x})$ is nonsingular, then f is a diffeomorphism relative to open neighborhoods U of $\bar{x}$ and V of $f(\bar{x})$.

Theorem 1F.3 (inverse function theorem for local diffeomorphism)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in a neighborhood of a point $\bar{x}$ and let the Jacobian $\nabla f(\bar{x})$ be nonsingular. Then for some open neighborhood U of $\bar{x}$ there exists an open neighborhood V of $\bar{y} := f(\bar{x})$ and a continuously differentiable function $s : V \to U$ which is one-to-one from V onto U and which satisfies $s(y) = f^{-1}(y) \cap U$ for all $y \in V$. Moreover, the Jacobian of s is given by*

$$\nabla s(y) = \nabla f(s(y))^{-1} \quad \textit{for every } y \in V.$$

Proof. First, we utilize a simple observation from linear algebra: for a nonsingular $n \times n$ matrix A, one has $|Ax| \geq |x|/|A^{-1}|$ for every $x \in \mathbb{R}^n$. Thus, let $c > 0$ be such that $|\nabla f(\bar{x})u| \geq 2c|u|$ for every $u \in \mathbb{R}^n$ and choose $a > 0$ to have, on the basis of (b) in Fact 1 of Section 1.1 [1A], that

$$|f(x') - f(x) - \nabla f(\bar{x})(x' - x)| \leq c|x' - x| \quad \text{for every } x', x \in \mathbb{B}_a(\bar{x}).$$

Using the triangle inequality, for any $x', x \in \mathbb{B}_a(\bar{x})$ we then have

$$|f(x') - f(x)| \geq |\nabla f(\bar{x})(x' - x)| - c|x' - x| \geq 2c|x' - x| - c|x' - x| \geq c|x' - x|.$$

We can therefore apply 1F.2, obtaining that there is an open neighborhood U of $\bar{x}$ relative to which f is continuous and f^{-1} is single-valued on $V := f(U)$. Brouwer's theorem 1F.1 then tells us that V is open. Then the mapping $s : V \to U$ whose graph is $\operatorname{gph} s = \operatorname{gph} f^{-1} \cap (V \times U)$ is the claimed single-valued localization of f^{-1} and the rest is argued through Step 3 in Proof I of 1A.1. □

Exercise 1F.4 (implicit function version)**.** *Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in a neighborhood of $(\bar{p}, \bar{x})$ and such that $f(\bar{p}, \bar{x}) = 0$, and let $\nabla_x f(\bar{p}, \bar{x})$ be nonsingular. Then for some open neighborhood U of $\bar{x}$ there exists an open neighborhood Q of $\bar{p}$ and a continuously differentiable function $s : Q \to U$ such that $\{ (p, s(p)) \mid p \in Q \} = \{ (p, x) \mid f(p, x) = 0 \} \cap (Q \times U)$; that is, s is a single-valued localization of the solution mapping $S(p) = \{ x \mid f(p, x) = 0 \}$ with associated open neighborhoods Q for $\bar{p}$ and U for $\bar{x}$. Moreover, the Jacobian of s is given by*

$$\nabla s(p) = -\nabla_x f(p, s(p))^{-1} \nabla_p f(p, s(p)) \quad \textit{for every } p \in Q.$$

Guide. Apply 1F.3 in the same way as 1A.1 is used in the proof of 1B.1. □

Exercise 1F.5. *Derive Theorem 1D.9 from 1F.2.*

Guide. By 1D.7, strict differentiability is equivalent to the assumption in Fact 1 of Section 1.1 [1A]; then repeat the argument in the proof of 1F.3. □

Brouwer's invariance of domain theorem tells us that, for a function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $m < n$, the inverse f^{-1} fails to have a localization which is single-valued. In this case, however, although multivalued, f^{-1} may "contain" a function with the properties of the single-valued localization for the case $m = n$. Such functions are generally called *selections* and their formal definition is as follows.

Selections. *Given a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a set $D \subset \operatorname{dom} F$, a function $w : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a selection of F on D if $\operatorname{dom} w \supset D$ and $w(x) \in F(x)$ for all $x \in D$. If $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, D is a neighborhood of $\bar{x}$ and w is a selection on D which satisfies $w(\bar{x}) = \bar{y}$, then w is said to be a local selection of F around $\bar{x}$ for $\bar{y}$.*

A selection of the inverse f^{-1} of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ might provide a *left inverse* or a *right inverse* to f. A left inverse to f on D is a selection $l : \mathbb{R}^m \to \mathbb{R}^n$ of f^{-1} on $f(D)$ such that $l(f(x)) = x$ for all $x \in D$. Analogously, a right inverse to f on D is a selection $r : \mathbb{R}^m \to \mathbb{R}^n$ of f^{-1} on $f(D)$ such that $f(r(y)) = y$ for all $y \in f(D)$. Commonly known are the right and the left inverses of the linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ represented by a matrix $A \in \mathbb{R}^{m\times n}$ that is of full rank. When $m \le n$, the right inverse of the mapping corresponds to $A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}$, while when $m \ge n$, the left inverse[6] corresponds to $(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$. For $m = n$ they coincide and equal the inverse. In general, of course, whenever a mapping f is one-to-one from a set C to $f(C)$, any left inverse to f on C is also a right inverse, and vice versa, and the restriction of such an inverse to $f(C)$ is uniquely determined.

The following result can be viewed as an extension of the classical inverse function theorem 1A.1 for selections.

Theorem 1F.6 (inverse selections when $m \le n$). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, where $m \le n$, be k times continuously differentiable in a neighborhood of $\bar{x}$ and suppose that its Jacobian $\nabla f(\bar{x})$ is full rank m. Then, for $\bar{y} = f(\bar{x})$, there exists a local selection s of f^{-1} around $\bar{y}$ for $\bar{x}$ which is k times continuously differentiable in a neighborhood V of $\bar{y}$ and whose Jacobian satisfies*

$$\nabla s(\bar{y}) = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}, \quad \text{where } A := \nabla f(\bar{x}). \tag{2}$$

Proof. There are various ways to prove this; here we apply the classical inverse function theorem. Since A has rank m, the $m \times m$ matrix AA^{T} is nonsingular. Then the function $\varphi : \mathbb{R}^m \to \mathbb{R}^m$ defined by $\varphi(u) = f(A^{\mathsf{T}}u)$ is k times continuously differentiable in a neighborhood of the point $\bar{u} := (AA^{\mathsf{T}})^{-1}A\bar{x}$, its Jacobian $\nabla\varphi(\bar{u}) = AA^{\mathsf{T}}$ is nonsingular, and noting that $\bar{x} = A^{\mathsf{T}}\bar{u}$ we get $\varphi(\bar{u}) = \bar{y}$. Then, from Theorem 1A.1 supplemented by Proposition 1B.5, it follows that φ^{-1} has a single-valued localization σ at $\bar{y}$ for $\bar{u}$ which is k times continuously differentiable near $\bar{y}$ with Jacobian $\nabla\sigma(\bar{y}) = (AA^{\mathsf{T}})^{-1}$. But then the function $s(y) = A^{\mathsf{T}}\sigma(y)$ satisfies $s(\bar{y}) = \bar{x}$ and

[6] The left inverse and the right inverse are particular cases of the *Moore-Penrose pseudo-inverse* A^{+} of a matrix A. For more on this, including the singular-value decomposition, see Golub and Van Loan [1996].

$f(s(y)) = y$ for all y near $\bar{y}$ and is k times continuously differentiable near $\bar{y}$ with Jacobian satisfying (2). Thus, $s(y)$ is a solution of the equation $f(x) = y$ for y close to $\bar{y}$ and x close to $\bar{x}$, but perhaps *not the only solution* there, as it would be in the classical inverse function theorem. Therefore, s is a local selection of f^{-1} around $\bar{y}$ for $\bar{x}$ with the desired properties. □

When $m = n$ the Jacobian becomes nonsingular and the right inverse of A in (2) is just A^{-1}. The uniqueness of the localization can be obtained much as in Step 2 of Proof I of the classical theorem 1A.1.

Exercise 1F.7 (parameterization of solution sets). *Let* $M = \{x \mid f(x) = 0\}$ *for a function* $f : \mathbb{R}^n \to \mathbb{R}^m$, *where* $n - m = d > 0$. *Let* $\bar{x} \in M$ *be a point around which* f *is* k *times continuously differentiable, and suppose that the Jacobian* $\nabla f(\bar{x})$ *has full rank* m. *Then for some open neighborhood* U *of* $\bar{x}$ *there is an open neighborhood* O *of the origin in* $\mathbb{R}^d$ *and a* k *times continuously differentiable function* $s : O \to U$ *which is one-to-one from* O *onto* $M \cap U$, *such that the Jacobian* $\nabla s(0)$ *has full rank* d *and*

$$\nabla f(\bar{x})w = 0 \text{ if and only if there exists } q \in \mathbb{R}^d \text{ with } \nabla s(0)q = w.$$

Guide. Choose an $d \times n$ matrix B such that the matrix

$$\begin{pmatrix} \nabla f(\bar{x}) \\ B \end{pmatrix}$$

is nonsingular. Consider the function

$$\bar{f} : (p,x) \mapsto \begin{pmatrix} f(x) \\ B(x - \bar{x}) - p \end{pmatrix} \quad \text{for } (p,x) \text{ near } (0,\bar{x}),$$

and apply 1F.6 (with a modification parallel to 1B.5) to the equation $\bar{f}(p,x) = (0,0)$, obtaining for the solution mapping of this equation a localization s with $\bar{f}(p,s(p)) = (0,0)$, i.e., $Bs(p) = p + B\bar{x}$ and $f(s(p)) = 0$. Show that this function s has the properties claimed. □

Exercise 1F.8 (strictly differentiable selections). *Let* $f : \mathbb{R}^n \to \mathbb{R}^m$, *where* $m \le n$, *be strictly differentiable at* $\bar{x}$ *with Jacobian* $A := \nabla f(\bar{x})$ *of full rank. Then there exists a local selection* s *of the inverse* f^{-1} *around* $\bar{y} := f(\bar{x})$ *for* $\bar{x}$ *which is strictly differentiable at* $\bar{y}$ *and with Jacobian* $\nabla s(\bar{y})$ *satisfying (2).*

Guide. Mimic the proof of 1F.6 taking into account 1D.9. □

Exercise 1F.9 (implicit selections). *Consider a function* $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$, *where* $m \le n$, *along with the associated solution mapping*

$$S : p \mapsto \{x \in \mathbb{R}^n \mid f(p,x) = 0\} \quad \text{for } p \in \mathbb{R}^d.$$

Let $f(\bar{p},\bar{x}) = 0$, so that $\bar{x} \in S(\bar{p})$. Assume that f is strictly differentiable at $(\bar{p},\bar{x})$ and suppose further that the partial Jacobian $\nabla_x f(\bar{p},\bar{x})$ is of full rank m. Then the mapping S has a local selection s around $\bar{p}$ for $\bar{x}$ which is strictly differentiable at $\bar{p}$ with Jacobian

$$\nabla s(\bar{p}) = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}\nabla_p f(\bar{p},\bar{x}), \quad \textit{where } A = \nabla_x f(\bar{p},\bar{x}).$$

Guide. Use 1F.8 and the argument in the proof of 1B.1. □

1.7 [1G] Selections from Nonstrict Differentiability

Even in the case when a function f maps $\mathbb{R}^n$ to itself, the inverse f^{-1} may fail to have a single-valued localization around $\bar{y} = f(\bar{x})$ for $\bar{x}$ if f is not strictly differentiable but merely differentiable at $\bar{x}$ with Jacobian $\nabla f(\bar{x})$ nonsingular. As when $m < n$, we have to deal with just a local selection of f^{-1}.

Theorem 1G.1 (inverse selections from nonstrict differentiability)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuous in a neighborhood of a point $\bar{x} \in \operatorname{int}\operatorname{dom} f$ and differentiable at $\bar{x}$ with $\nabla f(\bar{x})$ nonsingular. Then, for $\bar{y} = f(\bar{x})$, there exists a local selection of f^{-1} around $\bar{y}$ for $\bar{x}$ which is continuous at $\bar{y}$. Moreover, every local selection s of f^{-1} around $\bar{y}$ for $\bar{x}$ which is continuous at $\bar{y}$ has the property that*

(1) $$s \textit{ is differentiable at } \bar{y} \textit{ with Jacobian } \nabla s(\bar{y}) = \nabla f(\bar{x})^{-1}.$$

The verification of this claim relies on the following fixed point theorem, which we state here without proof.

Theorem 1G.2 (Brouwer fixed point theorem)**.** *Let Q be a compact and convex set in $\mathbb{R}^n$, and let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ be a function which is continuous on Q and maps Q into itself. Then there exists a point $x \in Q$ such that $\Phi(x) = x$.*

Proof of Theorem 1G.1. Without loss of generality, we can suppose that $\bar{x} = 0$ and $f(\bar{x}) = 0$. Let $A := \nabla f(0)$ and choose a neighborhood U of $0 \in \mathbb{R}^n$. Take $c \geq |A^{-1}|$. Choose any $\alpha \in (0, c^{-1})$. From the assumed differentiability of f, there exists $a > 0$ such that $x \in a\mathbb{B}$ implies $|f(x) - Ax| \leq \alpha|x|$. By making a smaller if necessary, we can arrange that f is continuous in $a\mathbb{B}$ and $a\mathbb{B} \subset U$. Let $b = a(1 - c\alpha)/c$ and pick any $y \in b\mathbb{B}$. Consider the function

$$\Phi_y : x \mapsto x - A^{-1}\big(f(x) - y\big) \quad \text{for } x \in a\mathbb{B}.$$

This function is of course continuous on the compact and convex set $a\mathbb{B}$. Furthermore, for any $x \in a\mathbb{B}$ we have

$$|\Phi_y(x)| = |x - A^{-1}(f(x) - y)| = |A^{-1}(Ax - f(x) + y)| \le |A^{-1}|(|Ax - f(x)| + |y|)$$
$$\le c|Ax - f(x)| + c|y| \le c\alpha|x| + cb \le c\alpha a + ca(1 - c\alpha)/c = a,$$

so Φ_y maps $a\mathbb{B}$ into itself. Then, by Brouwer's fixed point theorem 1G.2, there exists a point $x \in a\mathbb{B}$ such that $\Phi_y(x) = x$. Note that, in contrast to the contraction mapping principle 1A.2, this point may be not unique in $a\mathbb{B}$. But $\Phi_y(x) = x$ if and only if $f(x) = y$. For each $y \in b\mathbb{B}$, $y \neq 0$, we pick one $x \in a\mathbb{B}$ such that $x = \Phi_y(x)$; then $x \in f^{-1}(y)$. For $y = 0$ we take $x = 0$, which is clearly in $f^{-1}(0)$. Denoting this x by $s(y)$, we deduce the existence of a local selection $s : b\mathbb{B} \to a\mathbb{B}$ of f^{-1} around 0 for 0, also having the property that for any neighborhood U of 0 there exists $b > 0$ such that $s(y) \in U$ for $y \in b\mathbb{B}$, that is, s is continuous at 0.

Let s be a local selection of f^{-1} around 0 for 0 that is continuous at 0. Choose c, α and a as in the beginning of the proof. Then there exists $b' > 0$ with the property that $s(y) \in f^{-1}(y) \cap a\mathbb{B}$ for every $y \in b'\mathbb{B}$. This can be written as

$$s(y) = A^{-1}(As(y) - f(s(y)) + y) \quad \text{for every } y \in b'\mathbb{B},$$

which gives

$$|s(y)| \le |A^{-1}|(|As(y) - f(s(y))| + |y|) \le c\alpha|s(y)| + c|y|,$$

that is,

(2) $$|s(y)| \le \frac{c}{1 - c\alpha}|y| \quad \text{for all } y \in b'\mathbb{B}.$$

In particular, s is calm at 0. But we have even more. Choose any $\varepsilon > 0$. The differentiability of f with $\nabla f(0) = A$ furnishes the existence of $\tau \in (0, a)$ such that

(3) $$|f(x) - Ax| \le \frac{(1 - c\alpha)\varepsilon}{c^2}|x| \quad \text{whenever } |x| \le \tau.$$

Let $\delta = \min\{b', \tau(1 - c\alpha)/c\}$. Then on $\delta\mathbb{B}$ we have our local selection s of f^{-1} satisfying (2) and consequently

$$|s(y)| \le \frac{c}{1 - c\alpha}\delta \le \frac{c}{1 - c\alpha}\frac{(1 - c\alpha)\tau}{c} = \tau \quad \text{when } |y| \le \delta.$$

Taking norms in the identity

$$s(y) - A^{-1}y = -A^{-1}\big(f(s(y)) - As(y)\big),$$

and using (2) and (3), we obtain for $|y| \le \delta$ that

$$\big|s(y) - A^{-1}y\big| \le |A^{-1}||f(s(y)) - As(y)| \le \frac{c(1 - c\alpha)\varepsilon}{c^2}|s(y)| \le \frac{c(1 - c\alpha)\varepsilon c}{c^2(1 - c\alpha)}|y| = \varepsilon|y|.$$

Having demonstrated that for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|s(y) - A^{-1}y| \leq \varepsilon|y| \qquad \text{when} \quad |y| \leq \delta,$$

we conclude that (1) holds, as claimed. □

In order to gain more insight into what Theorem 1G.1 does or does not say, think about the case where the assumptions of the theorem hold and f^{-1} has a localization around $\bar{y}$ for $\bar{x}$ that avoids multivaluedness. This localization must actually be single-valued around $\bar{y}$, coinciding in some neighborhood with the local selection s in the theorem. Then we have a result which appears to be fully analogous to the classical inverse function theorem, but its shortcoming is the need to guarantee that a localization of f^{-1} without multivaluedness does exist. That, in effect, is what strict differentiability of f at $\bar{x}$, in contrast to just ordinary differentiability, is able to provide. An illustration of how inverse multivaluedness can indeed come up when the differentiability is not strict has already been encountered in Example 1E.5 with $\alpha \in (0,1)$. Observe that in this example there are infinitely many (even uncountably many) local selections of the inverse f^{-1} and, as the theorem says, each is continuous and even differentiable at 0, but also each selection is discontinuous at infinitely many points near but different from zero.

We can now partially extend Theorem 1G.1 to the case when $m \leq n$.

Theorem 1G.3 (differentiable inverse selections). *Let $f : I\!\!R^n \to I\!\!R^m$ be continuous in a neighborhood of a point $\bar{x} \in \operatorname{int} \operatorname{dom} f$ and differentiable at $\bar{x}$ with $A := \nabla f(\bar{x})$ of full rank m. Then, for $\bar{y} = f(\bar{x})$, there exists a local selection s of f^{-1} around $\bar{y}$ for $\bar{x}$ which is differentiable at $\bar{y}$.*

Comparing 1G.1 with 1G.3, we see that the equality $m = n$ gives us not only the existence of a local selection which is differentiable at $\bar{y}$ but also that every local selection which is continuous at $\bar{y}$, whose existence is assured also for $m < n$, is differentiable at $\bar{y}$ with the same Jacobian. Of course, if we assume in addition that f is strictly differentiable, we obtain strict differentiability of s at $\bar{y}$. To get this last result, however, we do not have to resort to Brouwer's fixed point theorem 1G.2.

Theorem 1G.1 is in fact a special case of a still broader result in which f does not need to be differentiable.

Theorem 1G.4 (inverse selections from first-order approximation). *Let $f : I\!\!R^n \to I\!\!R^m$ be continuous around $\bar{x}$ with $f(\bar{x}) = \bar{y}$, and let $h : I\!\!R^n \to I\!\!R^m$ be a first-order approximation of f at $\bar{x}$ which is continuous around $\bar{x}$. Suppose h^{-1} has a Lipschitz continuous local selection σ around $\bar{y}$ for $\bar{x}$. Then f^{-1} has a local selection s around $\bar{y}$ for $\bar{x}$ for which σ is a first-order approximation at $\bar{y}$: $s(y) = \sigma(y) + o(|y - \bar{y}|)$.*

Proof. We follow the proof of Theorem 1G.1 with some important modifications. Without loss of generality, take $\bar{x} = 0$, $\bar{y} = 0$. Let U be a neighborhood of the origin in $I\!\!R^n$. Let $\gamma > 0$ be such that σ is Lipschitz continuous on $\gamma I\!\!B$, and let $c > 0$ be a constant for this. Choose α such that $0 < \alpha < c^{-1}$ and $a > 0$ with $aI\!\!B \subset U$ and such that $\alpha a \leq \gamma/2$, f and h are continuous on $aI\!\!B$, and

$$(4) \qquad |e(x)| \leq \alpha|x| \quad \text{for all } x \in aI\!\!B, \quad \text{where } e(x) = f(x) - h(x).$$

Let

$$0 < b \le \min\left\{\frac{a(1-c\alpha)}{c}, \frac{\gamma}{2}\right\}. \tag{5}$$

For $x \in a\mathbb{B}$ and $y \in b\mathbb{B}$ we have

$$|y - e(x)| \le \alpha a + b \le \gamma. \tag{6}$$

Fix $y \in b\mathbb{B}$ and consider the function

$$\Phi_y : x \mapsto \sigma(y - e(x)) \quad \text{for } x \in a\mathbb{B}. \tag{7}$$

This function is of course continuous on $a\mathbb{B}$; moreover, from (6), the Lipschitz continuity of σ on $\gamma\mathbb{B}$ with constant c, the fact that $\sigma(0) = 0$, and the choice of b in (5), we obtain

$$|\Phi_y(x)| = |\sigma(y - e(x))| = |\sigma(y - e(x)) - \sigma(0)| \le c(\alpha a + b) \le a \ \text{ for all } x \in a\mathbb{B}.$$

Hence, by Brouwer's theorem 1G.2, there exists $x \in a\mathbb{B}$ with $x = \sigma(y - e(x))$. Then $h(x) = h(\sigma(y - e(x))) = y - e(x)$, that is, $f(x) = y$. For each $y \in b\mathbb{B}$, $y \neq 0$ we pick one such fixed point x of the function Φ_y in (7) in $a\mathbb{B}$ and call it $s(y)$; for $y = 0$ we set $s(0) = 0 \in f^{-1}(0)$. The function s is a local selection of f^{-1} around 0 for 0 which is, moreover, continuous at 0, since for an arbitrary neighborhood U of 0 we found $b > 0$ such that $s(y) \in U$ whenever $|y| \le b$. Also, for any $y \in b\mathbb{B}$ we have

$$s(y) = \sigma(y - e(s(y))). \tag{8}$$

From the continuity of s at 0 there exists $b' \in (0, b)$ such that $|s(y)| \le a$ for all $y \in b'\mathbb{B}$. For $y \in b'\mathbb{B}$, we see from (4), (8), the Lipschitz continuity of σ with constant c and the equality $\sigma(0) = 0$ that

$$|s(y)| = |\sigma(y - e(s(y))) - \sigma(0)| \le c\alpha|s(y)| + c|y|.$$

Hence, since $c\alpha < 1$,

$$|s(y)| \le \frac{c}{1 - \alpha c}|y| \quad \text{when } |y| \le b'. \tag{9}$$

Now, let $\varepsilon > 0$. By the assumption that h is a first-order approximation of f at 0, there exists $\tau \in (0, a)$ such that

$$|e(x)| \le \frac{(1 - \alpha c)\varepsilon}{c^2}|x| \quad \text{whenever } |x| \le \tau. \tag{10}$$

Finally, taking $b' > 0$ smaller if necessary and using (9) and (10), for any y with $|y| \le b'$ we obtain

$$|s(y) - \sigma(y)| = \big|\sigma(y - e(s(y))) - \sigma(y)\big|$$

$$\le c|e(s(y))| \le c\frac{(1-\alpha c)\varepsilon}{c^2}|s(y)|$$
$$\le c\frac{(1-\alpha c)\varepsilon}{c^2}\frac{c}{1-\alpha c}|y| = \varepsilon|y|.$$

Since for any $\varepsilon > 0$ we found $b' > 0$ for which this holds when $|y| \le b'$, the proof is complete. □

Proof of Theorem 1G.3. Apply Theorem 1G.4 with

$$h(x) = f(\bar{x}) + A(x-\bar{x}) \quad \text{and} \quad \sigma(y) = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}y.$$

□

We state next as an exercise an implicit function counterpart of 1G.3.

Exercise 1G.5 (differentiability of a selection)**.** *Consider a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ with $m \le n$, along with the solution mapping*

$$S : p \mapsto \{x \mid f(p,x) = 0\} \quad \text{for } p \in \mathbb{R}^d.$$

Let $f(\bar{p},\bar{x}) = 0$, so that $\bar{x} \in S(\bar{p})$, and suppose f is continuous around $(\bar{p},\bar{x})$ and differentiable at $(\bar{p},\bar{x})$. Assume further that $\nabla_x f(\bar{p},\bar{x})$ has full rank m. Prove that the mapping S has a local selection s around $\bar{p}$ for $\bar{x}$ which is differentiable at $\bar{p}$ with Jacobian

$$\nabla s(\bar{y}) = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}\nabla_p f(\bar{p},\bar{x}), \quad \text{where } A = \nabla_x f(\bar{p},\bar{x}).$$

The existence of a local selection of the inverse of a function f around $\bar{y} = f(\bar{x})$ for $\bar{x}$ implies in particular that $f^{-1}(y)$ is nonempty for all y in a neighborhood of $\bar{y} = f(\bar{x})$. This weaker property has even deeper significance and is defined next.

Openness. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be* open *at $\bar{x}$ if $\bar{x} \in \operatorname{int} \operatorname{dom} f$ and for every neighborhood U of $\bar{x}$ the set $f(U)$ is a neighborhood of $f(\bar{x})$.*

Thus, f is open at $\bar{x}$ if for every open neighborhood U of $\bar{x}$ there is an open neighborhood V of $\bar{y} = f(\bar{x})$ such that $f^{-1}(y) \cap U \ne \emptyset$ for every $y \in V$. In particular, this corresponds to the localization of f^{-1} relative to V and U being nonempty-valued on V, but goes further than referring just to one such localization at $\bar{y}$ for $\bar{x}$. It actually requires the existence of a nonempty-valued graphical localization for every neighborhood U of $\bar{x}$, no matter how small. From 1G.3 we obtain the following basic result about openness:

Corollary 1G.6 (Jacobian criterion for openness)**.** *For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, where $m \le n$, suppose that f is continuous around $\bar{x}$ and differentiable at $\bar{x}$ with $\nabla f(\bar{x})$ being of full rank m. Then f is open at $\bar{x}$.*

There is much more to say about openness of functions and set-valued mappings, and we will explore this in detail in Chapters 3 and 5.

1.8 [1H] An Implicit Function Theorem for Monotone Functions

In the preceding sections we considered implicit function theorems for for differentiable functions, or functions having first-order approximations at least. What could we say about the inverse f^{-1} of a function $f : \mathbb{R}^n \to \mathbb{R}^n$ when f is merely continuous? A good motivation to have a closer look at this issue is the fact, often mentioned in basic calculus texts, that if a continuous function $f : \mathbb{R} \to \mathbb{R}$ is strictly increasing (decreasing) on an interval $[a,b]$ then the inverse f^{-1} restricted to $[f(a), f(b)]$ $([f(b), f(a)])$ is a continuous function. There is no need to require differentiability here; the condition that the derivative is nonzero on $[a,b]$ guarantees that the graph of the function f is not "flat", which makes it invertible. We can extend the pattern to higher dimensions when we utilize the concept of monotonicity.

Monotone functions. *A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone on a set $C \subset \operatorname{dom} f$ if C is convex and*

$$(1) \qquad \langle f(x') - f(x), x' - x \rangle \geq 0 \quad \text{for all } x, x' \in C.$$

It is strictly monotone on C if (1) holds as a strict inequality for all $x, x' \in C$ with $x \neq x'$. Further, f is strongly monotone on C if there exists $\mu > 0$ such that

$$\langle f(x') - f(x), x' - x \rangle \geq \mu |x' - x|^2 \quad \text{for all } x, x' \in C.$$

The name "monotonicity" comes from the following characterization of the defining property.

Exercise 1H.1 (monotonicity along line segments). *Prove the following statements: A function $f : \mathbb{R}^n \to \mathbb{R}^n$ is monotone on a convex set $C \subset \operatorname{dom} f$ when for every $\hat{x} \in C$ and $w \in \mathbb{R}^n$ with $|w| = 1$, the function $\tau \mapsto \varphi(\tau) = \langle f(\hat{x} + \tau w), w \rangle$ is nondecreasing over the (interval of) τ values such that $\hat{x} + \tau w \in C$. If f is strictly monotone on C, then φ is strictly increasing over the same interval. Strong monotonicity with constant $\mu > 0$ corresponds to the condition that $\varphi(\tau') - \varphi(\tau) \geq \mu(\tau' - \tau)$ when $\tau' > \tau$.*

An affine function $f(x) = a + Ax$ with $a \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$ is monotone on $\mathbb{R}^n$ if and only if $\langle w, Aw \rangle \geq 0$ for all w, i.e., A is positive semidefinite. It is strongly monotone if and only if $\langle w, Aw \rangle > 0$ for all $w \neq 0$, i.e., A is positive definite. These terms make no requirement of symmetry on A. It may be recalled that any square matrix A can be written as a sum $A_s + A_a$ in which A_s is symmetric ($A_s^{\mathsf{T}} = A_s$) and A_a is antisymmetric ($A_a^{\mathsf{T}} = -A_a$), namely with $A_s = \frac{1}{2}[A + A^{\mathsf{T}}]$ and $A_a = \frac{1}{2}[A - A^{\mathsf{T}}]$; then $\langle w, Aw \rangle = \langle w, A_s w \rangle$. The monotonicity of $f(x) = a + Ax$ thus depends only on the symmetric part A_s of A; the antisymmetric part A_a can be anything.

For differentiable functions f that aren't affine, monotonicity has a similar characterization with respect to the Jacobian matrices $\nabla f(x)$.

Exercise 1H.2 (monotonicity from derivatives). *For a function $f : \mathbb{R}^n \to \mathbb{R}^n$ that is continuously differentiable on an open convex set $O \subset \operatorname{dom} f$, verify the following facts.*

(a) *A necessary and sufficient condition for f to be monotone on O is the positive semidefiniteness of $\nabla f(x)$ for all $x \in O$.*

(b) *If $\nabla f(x)$ is positive definite at every point x of a convex set $C \subset O$, then f is strictly monotone on C.*

(c) *If $\nabla f(x)$ is positive definite at every point x of a closed, bounded, convex set $C \subset O$, then f is strongly monotone on C.*

(d) *If C is a convex subset of O such that $\langle \nabla f(x)w, w\rangle \geq 0$ for every $x \in C$ and $w \in C - C$, then f is monotone on C.*

(e) *If C is a convex subset of O such that $\langle \nabla f(x)w, w\rangle > 0$ for every $x \in C$ and $w \in C - C$, then f is strictly monotone on C.*

(f) *If C is a convex subset of O such that $\langle \nabla f(x)w, w\rangle \geq \mu |w|^2$ for every $x \in C$ and every $w \in C - C$ for some $\mu > 0$, then f is strongly monotone on C with constant μ.*

Guide. Derive this from the characterizations in 1H.1 by investigating the derivatives of the function $\varphi(\tau)$ introduced there. In proving (c), argue by way of the mean value theorem that $\langle f(x') - f(x), x' - x\rangle$ equals $\langle \nabla f(\tilde{x})(x' - x), x' - x\rangle$ for some point $\tilde{x}$ on the line segment joining x with x'. □

We will present next an implicit function theorem for strictly monotone functions. Monotonicity will be employed again in Section 2F to obtain inverse function theorems in the broader context of variational inequalities.

Theorem 1H.3 (implicit function theorem for strictly monotone functions)**.** *Consider a function $f : I\!R^d \times I\!R^n \to I\!R^n$ and a point $(\bar{p}, \bar{x}) \in \operatorname{int} \operatorname{dom} f$ satisfying $f(\bar{p}, \bar{x}) = 0$. Suppose that there are neighborhoods Q of $\bar{p}$ and U of $\bar{x}$ such that f is continuous on $Q \times U$ and for each $p \in Q$ the function $f(p, \cdot)$ is strictly monotone on U. Then the solution mapping*

$$S(p) = \left\{ x \in I\!R^n \,\middle|\, f(p, x) = 0 \right\} \quad \text{for} \;\; p \in I\!R^d$$

has a single-valued localization around $\bar{p}$ for $\bar{x}$ which is continuous at $\bar{p}$. If $f(p, \cdot)$ is strongly monotone on U uniformly in $p \in Q$ and $f(\cdot, x)$ is Lipschitz continuous on Q uniformly in $x \in U$, then S has a Lipschitz continuous single-valued localization around $\bar{p}$ for $\bar{x}$.

Proof. First, observe that if $p \in \operatorname{dom} S \cap Q$ then $S(p) \cap U$ consists of one element, if any. Indeed, if there exist two elements $x, x' \in S(p) \cap U$ with $x \neq x'$, then from the strict monotonicity we obtain $0 = \langle f(p,x) - f(p,x'), x - x'\rangle > 0$, a contradiction. Thus, all we need is to establish that $\operatorname{dom} S$ contains a neighborhood of $\bar{p}$.

Without loss of generality, let $\bar{x} = 0$ and choose $\delta > 0$ such that $\delta I\!B \subset U$. For $\varepsilon \in (0, \delta]$ define

$$d(\varepsilon) = \inf_{\varepsilon \leq |x| \leq \delta} \left[\frac{1}{|x|} \langle x, f(\bar{p}, x)\rangle \right].$$

Choose $x \in I\!R^n$ with $|x| \in [\varepsilon, \delta]$. Since $f(\bar{p}, 0) = 0$, from the strict monotonicity of $f(\bar{p}, \cdot)$ we obtain that $\langle f(\bar{p}, x), x\rangle > 0$, hence $d(\varepsilon) \geq 0$ for all $\varepsilon > 0$. Assume that there exists $\varepsilon_0 > 0$ such that $d(\varepsilon_0) = 0$. Then there exists a sequence $\{x_k\}$ with $\varepsilon_0 \le |x_k| \le \delta$ such that

$$\frac{1}{|x_k|}\langle x_k, f(\bar{p}, x_k)\rangle \to 0.$$

Since $x_k \in \delta \mathbb{B}$ the sequence $\{x_k\}$ has an accumulation point $\tilde{x}$ satisfying $\langle \tilde{x}, f(\bar{p}, \tilde{x})\rangle = 0$. But $|\tilde{x}| \ge \varepsilon_0$ which, combined with the strict monotonicity, gives us

$$0 = \langle \tilde{x}, f(\bar{p}, \tilde{x}) - f(\bar{p}, 0)\rangle > 0,$$

a contradiction. Hence $d(\varepsilon) > 0$ for all $\varepsilon \in (0, \delta]$. Clearly, $d(\varepsilon) \searrow 0$ as $\varepsilon \searrow 0$, and $d(\cdot)$ is strictly increasing; thus $d(\delta) > 0$.

Let $\mu \in (0, d(\delta))$ and let $\tau > 0$ be such that $\mathbb{B}_\tau(\bar{p}) \subset Q$. From the continuity of f with respect to $p \in \mathbb{B}_\tau(\bar{p})$ uniformly in $x \in \{x \mid \|x\| = \delta \mathbb{B}\}$ there exists $\nu \in (0, \tau)$ such that for any $p \in \mathbb{B}_\nu(\bar{p})$ and any x on the boundary of $\delta \mathbb{B}$, we have

$$\frac{1}{|x|}\langle x, f(p, x)\rangle \ge \frac{1}{|x|}\langle x, f(\bar{p}, x)\rangle - \mu \ge d(\delta) - \mu > 0. \tag{2}$$

For a fixed $p \in \mathbb{B}_\tau(\bar{p})$ consider the function

$$x \mapsto \Phi(x) = P_{\delta B}(x - f(p, x)) \quad \text{for} \quad x \in \delta \mathbb{B},$$

where, as usual, $P_C(y)$ is the Euclidean projection of y on the set C. By the Brouwer fixed point theorem 1G.2, Φ has a fixed point $\hat{x}$ in $\delta \mathbb{B}$. Let $|\hat{x}| = \delta$, that is, $\hat{x}$ is on the boundary of the ball $\delta \mathbb{B}$. Then, by the properties of the projection we have that

$$\langle 0 - \hat{x}, \hat{x} - f(p, \hat{x}) - \hat{x}\rangle \le 0, \quad \text{that is,} \quad \frac{1}{|\hat{x}|}\langle \hat{x}, f(p, \hat{x})\rangle \le 0.$$

On the other hand, the inequality (2) for the fixed p and $x = \hat{x}$ leads us to

$$\frac{1}{|\hat{x}|}\langle \hat{x}, f(p, \hat{x})\rangle \ge \frac{1}{|\hat{x}|}\langle x, f(\bar{p}, \hat{x})\rangle - \mu \ge d(\delta) - \mu > 0.$$

The obtained contradiction implies that $\hat{x}$ is in the interior of $\delta \mathbb{B}$, in which case we have $\hat{x} = \hat{x} - f(p, \hat{x})$, that is, $\hat{x} \in S(p)$, which means that $S(p) \cap \delta \mathbb{B} \ne \emptyset$. Thus, the mapping $p \mapsto S(p) \cap U$ is a function, to be denoted by $x(\cdot)$, whose domain contains $\mathbb{B}_\nu(\bar{p})$.

To prove the continuity of S at $\bar{p}$ it is sufficient to repeat the preceding argument. Indeed, we have chosen $\delta > 0$ and then found $\nu > 0$ such that for all $p \in \mathbb{B}_\nu(\bar{p})$ the value $x(p)$ of the localization of $S(p)$ is at distance less than δ from $0 = \bar{x} = x(\bar{p})$. Hence $x(\cdot)$ is continuous at $\bar{p}$.

Suppose that $f(p, \cdot)$ is strongly monotone on U uniformly in $p \in Q$. For $p, p' \in \mathbb{B}_\nu(\bar{p})$, $p \ne p'$, let $x(p)$ and $x(p')$ be the values of the single-valued localization of S around $\bar{p}$ for 0. From the strong monotonicity we have that for some $c > 0$,

$$\langle f(p, x(p)) - f(p, x(p')), x(p) - x(p')\rangle \ge c|x(p) - x(p')|^2.$$

Combining this with the equalities $f(p,x(p)) = f(p',x(p')) = 0$, taking norms, and utilizing the Lipschitz continuity of $f(\cdot,x)$ with constant L, we get

$$L|p-p'||x(p)-x(p')| \geq$$
$$|x(p)-x(p')||f(p',x(p'))-f(p,x(p'))| \geq c|x(p)-x(p')|^2.$$

This gives us Lipschitz continuity of $x(\cdot)$, and the proof is complete. $\square$

Commentary

Although functions given implicitly by equations had been considered earlier by Descartes, Newton, Leibnitz, Lagrange, Bernoulli, Euler, and others, Cauchy [1831] is credited by historians to be the first who stated and rigorously proved an implicit function theorem — for analytic functions, by using his calculus of residuals and limits. As we mentioned in the preamble to this chapter, Dini [1877/78] gave the form of the implicit function theorem for continuously differentiable functions which is now used in most calculus books; in his proof he relied on the mean value theorem. More about early history of the implicit function theorem can be found in historical notes of the paper of Hurwicz and Richter [2003] and in the book of Krantz and Parks [2002].

Proof I of the classical inverse function theorem, 1A.1, goes back to Goursat [1903][7]. Most likely not aware of Dini's theorem and inspired by Picard's successive approximation method for proving solution existence of differential equations, Goursat stated an implicit function theorem under assumptions weaker than in Dini's theorem, and supplied it with a new path-breaking proof. With updated notation, Goursat's proof employs the iterative scheme

$$x^{k+1} = x^k - A^{-1} f(p, x^k), \text{ where } A = \nabla_x f(\bar{p}, \bar{x}). \tag{1}$$

This scheme would correspond to Newton's method for solving $f(p,x) = 0$ with respect to x if A were replaced by A^k giving the partial derivative at (p, x^k) instead of $(\bar{p}, \bar{x})$. But Goursat proved anyway that for each p near enough to $\bar{p}$ the sequence $\{x^k\}$ is convergent to a unique point $x(p)$ close to $\bar{x}$, and furthermore that the function $p \mapsto x(p)$ is continuous at $\bar{p}$. Behind the scene, as in Proof I of Theorem 1A.1, is the contraction mapping idea. An updated form of Goursat's implicit function theorem is given in Theorem 1B.6. In the functional analysis text by Kantorovich and Akilov [1964], Goursat's iteration is called a "modified Newton's method."

The rich potential in this proof was seen by Lamson [1920], who used the iterations in (1) to generalize Goursat's theorem to abstract spaces. Especially interesting for our point of view in the present book is the fact that Lamson was motivated by an optimization problem, namely the problem of Lagrange in the calculus of variations with equality constraints, for which he proved a Lagrange multiplier rule by way of his implicit function theorem.

Lamson's work was extended in a significant way by Hildebrand and Graves in their paper from 1927. They first stated a contraction mapping result (their Theorem 1), in which the only difference with the statement of Theorem 1A.2 is the presence of a superfluous parameter. The contraction mapping principle, as formulated in 1A.3, was published five years earlier in Banach [1922] (with some easily fixed typos), but the idea behind the contraction mapping was evidently known to Goursat, Picard and probably even earlier. Hildebrand and Graves cited in their paper

---

[7] Edouard Jean-Baptiste Goursat (1858–1936). Goursat's paper from 1903 is available at http://www.numdam.org/.

Banach's work [1922], but only in the context of the definition of a Banach space[8]. Further, based on their parameterized formulation of 1A.2, they established an implicit function theorem in the classical form of 1B.1 (their Theorem 4) for functions acting in linear metric spaces. More intriguing, however, for its surprising foresight, is their Theorem 3, called by the authors a "neighborhood theorem," where they do not assume differentiability; they say "only an approximate differential ... is required;" this is in contrast to the works of Goursat and Lamson where an "exact differential" is used. In this, Hildebrand and Graves are far ahead of their time. (We will see a similar picture with Graves' theorem later in Section 5D.) Because of the importance of this result of Hildebrand and Graves, we provide a statement of it here in finite dimensions with some adjustments in terminology and notation.

Theorem (Hildebrand–Graves theorem)**.** *Let $Q \subset I\!\!R^d$ and consider a function $f : Q \times I\!\!R^n \to I\!\!R^n$ along with a point $\bar{x} \in I\!\!R^n$. Suppose there are a positive constant a, a linear bounded mapping $A : I\!\!R^n \to I\!\!R^n$ which is invertible, and a positive constant M with $M|A^{-1}| < 1$ such that*

(a) *for all $p \in Q$ and $x, x' \in I\!\!B_a(\bar{x})$, one has $|f(p,x) - f(p,x') - A(x - x')| \leq M|x - x'|$;*

(b) *for every $p \in Q$, one has $|A^{-1}||f(p,\bar{x})| \leq a(1 - M|A^{-1}|)$.*

Then the solution mapping $S : p \mapsto \{x \mid f(p,x) = 0\}$ is single-valued on Q [when its values are restricted to a neighborhood of $\bar{x}$].

The phrase in brackets in the last sentence is our addition: Hildebrand and Graves apparently overlooked the fact, which is still overlooked by some writers, that the implicit function theorem is about a *graphical localization* of the solution mapping. If we assume in addition that f is Lipschitz continuous with respect to p uniformly in x, we will obtain, according to 1E.13, that the solution mapping has a Lipschitz continuous single-valued localization. When f is assumed strictly differentiable at $(\bar{p},\bar{x})$, by taking $A = D_x f(\bar{p},\bar{x})$ we come to 1D.13.

The main novelty in the Hildebrand–Graves theorem is that *differentiability is replaced by Lipschitz continuity*. This is not spelled out in their paper but can be gleaned from their proof. This becomes apparent in the following slightly extended inverse function version of it which can be easily derived from their proof, compare with 1E.3.

Theorem (extended Hildebrand–Graves theorem in inverse form)**.** *Consider a function $f : I\!\!R^n \to I\!\!R^n$ along with a point $\bar{x} \in I\!\!R^n$ such that $f(\bar{x}) = 0$, and a linear bounded mapping $A : I\!\!R^n \to I\!\!R^n$ which is invertible. If*

$$\operatorname{lip}(f - A; \bar{x}) \cdot |A^{-1}| < 1,$$

then the inverse f^{-1} has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$.

[8] They mention in a footnote that the name "Banach spaces" for normed linear spaces that are complete was coined by Fréchet.

The classical implicit function theorem is present in many of the textbooks in calculus and analysis written in the last hundred years. The proofs on the introductory level are mainly variations of Dini's proof, depending on the material covered prior to the theorem's statement. Interestingly enough, in his text Goursat [1904] applies the mean value theorem, as in Dini's proof, and not the contraction mapping iteration he introduced in his paper of 1903. Similar proofs are given as early as the 30s in Courant [1988] and most recently in Fitzpatrick [2006]. In more advanced textbooks from the second half of the last century, such as in the popular texts of Apostol [1962], Schwartz [1967] and Dieudonné [1969], it has become standard to use the contraction mapping principle. We were not able to identify a calculus text in which the implicit function theorem is given in the symmetric form 1B.8.

The material in sections 1.3 [1C] and 1.4 [1D] is mostly known, but the way it is presented is new. First-order approximations of functions were introduced in Robinson [1991].

Theorem 1E.3 can be viewed as an extension of the Hildebrand–Graves theorem where the "approximate differential" is not required to be a linear mapping; we will get back to this result in Chapter 2 and also later in the book. The statement of 1E.8(a) is sometimes called the Banach lemma, see, e.g., Noble and Daniel [1977]. Kahan [1966] and many after him attribute Theorem 1E.9 to Gastinel, without giving a reference. This result can be also found in the literature as the "Eckart–Young theorem" with the citation of Eckart and Young [1936], which however is a related but different kind of result, concerning the distance from a matrix from another matrix with lower rank. That result in turn is much older still and is currently referred to as the Schmidt–Eckart–Young–Mirsky theorem on singular value decomposition. For history see Chipman [1997] and the book by Stewart and Sun [1990].

On the other hand, 1E.9 can be derived with the help of the Schmidt–Eckart–Young–Mirsky theorem, inasmuch as the latter implies that the distance from a nonsingular matrix to the set of rank-one matrices is equal to the smallest singular value, which is the reciprocal to the norm of the inverse. Hence the distance from a nonsingular matrix to the set of singular matrices is not greater than the reciprocal to the norm of the inverse. Combining this with 1E.7 gives us the radius equality. More about stability of perturbed inversions will be presented in Chapters 5 and 6.

Brouwer's invariance of domain theorem 1F.1 can be found, e.g., in Spanier [1966], while Brouwer's fixed point theorem 1G.2 is given in Dunford and Schwartz [1958]. Theorem 1G.3 slightly extends a result in Halkin [1974]; for extensions in other directions, see Hurwicz and Richter [2003].

A one-dimensional version of Theorem 1H.3[9] is given on p. 449 in the calculus textbook by Fikhtengolts [1962]. A more general theorem is given in Kassay and Kolumbán [1988], for extensions to variational inequalities see Section 2F.

[9] The reviewer of the first edition of this book in Zentralblatt für Mathematik named this theorem after Peano. We were unable to find evidence that this theorem is indeed due to Peano.

Chapter 2
Solution Mappings for Variational Problems

Solutions mappings in the classical setting of the implicit function theorem concern problems in the form of parameterized equations. The concept can go far beyond that, however. In any situation where some kind of problem in x depends on a parameter p, there is the mapping S that assigns to each p the corresponding set of solutions x. The same questions then arise about the extent to which a localization of S around a pair $(\bar{p},\bar{x})$ in its graph yields a function s which might be continuous or differentiable, and so forth.

This chapter moves into that much wider territory in replacing equation-solving problems by more complicated problems termed "generalized equations." Such problems arise in constrained optimization, models of equilibrium, and many other areas. An important feature, in contrast to ordinary equations, is that functions obtained implicitly from their solution mappings typically lack differentiability, but often exhibit Lipschitz continuity and sometimes combine that with the existence of one-sided directional derivatives.

The first task is to introduce "generalized equations" and their special case, "variational inequality" problems, which arises from the variational geometry of sets expressing constraints. Problems of optimization and the Lagrange multiplier conditions characterizing their solutions provide key examples. Convexity of sets and functions enters as a valuable ingredient.

From that background, the chapter proceeds to Robinson's implicit function theorem for parameterized variational inequalities and several of its extensions. Subsequent sections introduce concepts of ample parameterization and semidifferentiability, building toward major results in 2E for variational inequalities over convex sets that are polyhedral. A follow-up in 2F looks at a class of monotone variational inequalities, after which, in 2G, a number of applications in optimization are worked out.

2A. Generalized Equations and Variational Problems

By a *generalized equation* in $\mathbb{R}^n$ will be meant a condition on x of the form

$$(1) \qquad f(x) + F(x) \ni 0, \text{ or equivalently } -f(x) \in F(x),$$

for a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and a (generally) set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. The name refers to the fact that (1) reduces to an ordinary equation $f(x) = 0$ when F is the *zero mapping* (with $F(x)$ containing 0 and nothing else, for every x), which we indicate in notation by $F \equiv 0$. Any x satisfying (1) is a *solution* to (1).

Generalized equations take on importance in many situations, but an especially common and useful type arises from normality conditions with respect to convex sets.

Normal cones. *For a convex set $C \subset \mathbb{R}^n$ and a point $x \in C$, a vector v is said to be* normal *to C at x if $\langle v, x' - x\rangle \le 0$ for all $x' \in C$. The set of all such vectors v is called the* normal cone *to C at x and is denoted by $N_C(x)$. For $x \notin C$, $N_C(x)$ is taken to be the empty set. The* normal cone mapping *is thus defined as*

$$N_C : x \mapsto \begin{cases} N_C(x) & \text{for } x \in C, \\ \emptyset & \text{otherwise.} \end{cases}$$

The term *cone* refers to a set of vectors which contains 0 and contains with any of its elements v all positive multiples of v. For each $x \in C$, the normal cone $N_C(x)$ is indeed a cone in this sense. Moreover it is closed and convex. The normal cone mapping $N_C : x \mapsto N_C(x)$ has $\operatorname{dom} N_C = C$. When C is a closed subset of $\mathbb{R}^n$, $\operatorname{gph} N_C$ is a closed subset of $\mathbb{R}^n \times \mathbb{R}^n$.

Variational inequalities. *For a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a closed convex set $C \subset \operatorname{dom} f$, the generalized equation*

$$(2) \qquad f(x) + N_C(x) \ni 0, \text{ or equivalently } -f(x) \in N_C(x),$$

is called the variational inequality *for f and C.*

Note that, because $N_C(x) = \emptyset$ when $x \notin C$, a solution x to (2) must be a point of C. The name of this condition originated from the fact that, through the definition of the normal vectors to C, (2) is equivalent to having

$$(3) \qquad x \in C \text{ and } \langle f(x), x' - x\rangle \ge 0 \text{ for all } x' \in C.$$

Instead of contemplating a system of infinitely many linear inequalities, however, it is usually better to think in terms of the properties of the set-valued mapping N_C, which the formulation in (2) helps to emphasize.

When $x \in \operatorname{int} C$, the only normal vector at x is 0, and the condition in (2) just becomes $f(x) = 0$. Indeed, the variational inequality (2) is totally the same as the equation $f(x) = 0$ in the case of $C = I\!R^n$, which makes $N_C \equiv 0$. In general, though, (2) imposes a relationship between $f(x)$ and the boundary behavior of C at x.

There is a simple connection between the normal cone mapping N_C and the projection mapping P_C introduced in 1D: one has

$$v \in N_C(x) \iff P_C(x+v) = x. \tag{4}$$

Interestingly, this *projection rule* for normals means for the mappings N_C and P_C that

$$N_C = P_C^{-1} - I, \qquad P_C = (I + N_C)^{-1}. \tag{5}$$

A consequence of (5) is that the variational inequality (2) can actually be written as an equation, namely

$$f(x) + N_C(x) \ni 0 \iff P_C(x - f(x)) - x = 0. \tag{6}$$

It should be kept in mind, though, that this doesn't translate the *solving* of variational inequalities into the classical framework of solving nonlinear equations. There, "linearizations" are essential, but P_C often fails to be differentiable, so linearizations generally aren't available for the equation in (6), regardless of the degree of differentiability of f. Other approaches can sometimes be brought in, however, depending on the nature of the set C. Anyway, the characterization in (6) has the advantage of leading quickly to a criterion for the existence of a solution to a variational inequality in a basic case.

Theorem 2A.1 (solutions to variational inequalities). *For a function $f : I\!R^n \to I\!R^n$ and a nonempty, closed convex set $C \subset \operatorname{dom} f$ relative to which f is continuous, the set of solutions to the variational inequality (2) is always closed. It is sure to be nonempty when C is bounded.*

Proof. Let $M(x) = P_C(x - f(x))$. Because C is nonempty, closed and convex, the projection mapping P_C is, by 1D.5, a Lipschitz continuous function from $I\!R^n$ to C. Then M is a continuous function from C to C under our continuity assumption on f. According to (6), the set of solutions x to (2) is the same as the set of points $x \in C$ such that $M(x) = x$, which is closed. When C is bounded, we can apply Brouwer's fixed point theorem 1G.2 to conclude the existence of at least one such point x. $\square$

Other existence theorems which don't require C to be bounded can also be given, especially for situations in which f has the property of monotonicity which we introduced in 1H. This will be taken up in 2F.

The examples and properties to which the rest of this section is devoted will help to indicate the scope of the variational inequality concept. They will also lay the foundations for the generalizations of the implicit function theorem that we are aiming at.

Exercise 2A.2 (some normal cone formulas)**.**

(a) *If M is a linear subspace of $\mathbb{R}^n$, then $N_M(x) = M^\perp$ for every $x \in M$, where $M^\perp$ is the orthogonal complement of M.*

(b) *The unit Euclidean ball $\mathbb{B}$ has $N_{\mathbb{B}}(x) = \{0\}$ when $|x| < 1$, but $N_{\mathbb{B}}(x) = \{\, \lambda x \,|\, \lambda \geq 0 \,\}$ when $|x| = 1$.*

(c) *The nonnegative orthant $\mathbb{R}^n_+ = \{\, x = (x_1,\dots,x_n) \,|\, x_j \geq 0 \text{ for } j = 1,\dots,n \,\}$ has*

$$(v_1,\dots,v_n) \in N_{\mathbb{R}^n_+}(x_1,\dots,x_n) \iff v \leq 0,\ v \perp x \iff \begin{cases} v_j \leq 0 & \text{for } j \text{ with } x_j = 0, \\ v_j = 0 & \text{for } j \text{ with } x_j > 0. \end{cases}$$

Guide. The projection rule (4) provides an easy way of identifying the normal vectors in these examples. □

The formula in 2A.2(c) comes up far more frequently than might be anticipated. A variational inequality (2) in which $C = \mathbb{R}^n_+$ is called a *complementarity problem*; one has

$$-f(x) \in N_{\mathbb{R}^n_+}(x) \iff x \geq 0,\ f(x) \geq 0,\ x \perp f(x).$$

Here the common notation is adopted that a vector inequality like $x \geq 0$ is to be taken componentwise, and that $x \perp y$ means $\langle x, y \rangle = 0$. Many variational inequalities can be recast, after some manipulation, as complementarity problems, and the numerical methodology for solving such problems has therefore received especially much attention.

The orthogonality relation in 2A.2(a) extends to a "polarity" relation for cones which has a major presence in our subject.

Proposition 2A.3 (polar cone)**.** *Let K be a closed, convex cone in $\mathbb{R}^n$ and let K^* be its polar, defined by*

$$K^* = \{\, y \,|\, \langle x, y \rangle \leq 0 \text{ for all } x \in K \,\}. \tag{7}$$

Then K^* is likewise a closed, convex cone, and its polar $(K^*)^*$ is in turn K. Furthermore, the normal vectors to K and K^* are related by

$$y \in N_K(x) \iff x \in N_{K^*}(y) \iff x \in K,\ y \in K^*,\ \langle x, y \rangle = 0. \tag{8}$$

Proof. First consider any $x \in K$ and $y \in N_K(x)$. From the definition of normality in (7) we have $\langle y, x' - x \rangle \leq 0$ for all $x' \in K$, so the maximum of $\langle y, x' \rangle$ over $x' \in K$ is attained at x. Because K contains all positive multiples of each of its vectors, this comes down to having $\langle y, x \rangle = 0$ and $\langle y, x' \rangle \leq 0$ for all $x' \in K$. Therefore $N_K(x) = \{\, y \in K^* \,|\, y \perp x \,\}$.

It's elementary that K^* is a cone which is closed and convex, with $(K^*)^* \supset K$. Consider any $z \notin K$. Let $x = P_K(z)$ and $y = z - x$. Then $y \neq 0$ and $P_K(x+y) = x$, hence $y \in N_K(x)$, so that $y \in K^*$ and $y \perp x$. We have $\langle y, z \rangle = \langle y, y \rangle > 0$, which confirms that $z \notin (K^*)^*$. Therefore $(K^*)^* = K$. The formula for normals to K must hold then equally for K^* through symmetry: $N_{K^*}(y) = \{x \in K \,|\, x \perp y\}$ for any $y \in K^*$. This establishes the normality relations that have been claimed. □

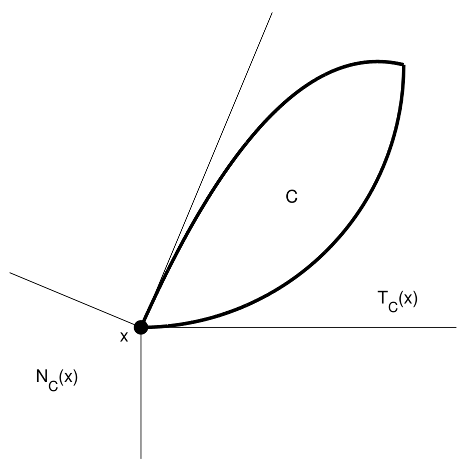

Fig. 2.1 Tangent and normal cone to a convex set.

Polarity has a basic role in relating the normal vectors to a convex set to its "tangent vectors."

Tangent cones. *For a set $C \subset \mathbb{R}^n$ (not necessarily convex) and a point $x \in C$, a vector v is said to be tangent to C at x if*

$$\frac{1}{\tau^k}(x^k - x) \to v \quad \text{for some } x^k \to x,\ x^k \in C,\ \tau^k \searrow 0.$$

The set of all such vectors v is called the tangent cone to C at x and is denoted by $T_C(x)$. For $x \notin C$, $T_C(x)$ is taken to be the empty set.

Although we will mainly be occupied with normal cones to convex sets at present, tangent cones to convex sets and even nonconvex sets will be put to serious use later in the book. Note from the definition that $0 \in T_C(x)$ for all $x \in C$.

Exercise 2A.4 (tangent and normal cones)**.** *The tangent cone $T_C(x)$ to a closed, convex set C at a point $x \in C$ is the closed, convex cone that is polar to the normal cone $N_C(x)$: one has*

$$(9) \qquad T_C(x) = N_C(x)^*, \qquad N_C(x) = T_C(x)^*.$$

Guide. The second of the equations (9) comes immediately from the definition of $N_C(x)$, and the first is then obtained from Proposition 2A.3. □

Variational inequalities are instrumental in capturing conditions for optimality in problems of minimization or maximization and even "equilibrium" conditions such as arise in games and models of conflict. To explain this motivation, it will be helpful to be able to appeal to the convexity of functions, at least in part.

Convex functions. *A function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be convex relative to a convex set C (or just convex, when $C = \mathbb{R}^n$) if*

$$g((1-\tau)x+\tau x') \leq (1-\tau)g(x)+\tau g(x') \text{ for all } \tau \in (0,1) \text{ when } x,x' \in C.$$

It is strictly convex if this holds with strict inequality for $x \neq x'$. It is strongly convex with constant μ when $\mu > 0$ and, for every $x,x' \in C$,

$$g((1-\tau)x+\tau x') \leq (1-\tau)g(x)+\tau g(x') - \mu\tau(1-\tau)|x-x'|^2 \text{ for all } \tau \in (0,1).$$

A function g is concave, strictly concave or strongly concave, if $-g$ is convex, strictly convex or strongly convex, respectively. It is affine relative to C when the inequality is an equation, which corresponds to g being simultaneously convex and concave relative to C.

The following are the standard criteria for convexity or strict convexity of g which can be obtained from the definitions in terms of the *gradient vectors*

$$\nabla g(x) = \left(\frac{\partial g}{\partial x_j}(x_1,\ldots,x_n)\right)_{j=1}^{n}$$

and the *Hessian matrices*

$$\nabla^2 g(x) = \left(\frac{\partial^2 g}{\partial x_i \partial x_j}(x_1,\ldots,x_n)\right)_{i,j=1}^{n,n}.$$

Exercise 2A.5 (characterizations of convexity)**.**

(a) *A differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ on an open convex set O is convex if and only if*

$$g(x') \geq g(x) + \langle \nabla g(x), x'-x\rangle \text{ for all } x,x' \in O.$$

It is strictly convex if and only if this inequality is always strict when $x' \neq x$. It is strongly convex with constant μ, where $\mu > 0$, if and only if

$$g(x') \geq g(x) + \langle \nabla g(x), x'-x\rangle + \frac{\mu}{2}|x'-x|^2 \text{ for all } x,x' \in O.$$

(b) *A twice differentiable function g on an open convex set O is convex if and only if $\nabla^2 g(x)$ is positive semidefinite for every $x \in O$. It is strictly convex if $\nabla^2 g(x)$ is positive definite for every $x \in O$. (This sufficient condition for strict convexity is not necessary, however, in general.) It is strongly convex with constant μ if and only if $\mu > 0$ and $\langle \nabla^2 g(x)w, w\rangle \geq \mu$ for all $x \in O$ and $w \in \mathbb{R}^n$ with $|w| = 1$.*

Guide. Because the definition of convexity revolves only around points that are collinear, the convexity of g can be verified by showing that, for arbitrary $x \in O$ and $w \in \mathbb{R}^n$, the function $\varphi(t) = g(x+tw)$ is convex on the interval $\{ t \,|\, x+tw \in O \}$. The conditions for this on $\varphi'(t)$ and $\varphi''(t)$, known from basic calculus, can be applied by expressing these derivatives in terms of the gradients and Hessian of g. This approach can be used to verify all the claims. □

In Section 1H we introduced the class of monotone functions and presented some of their properties. The following exercise establishes a connection between convex functions and monotonicity of their first derivatives.

Exercise 2A.6 (gradient connections)**.**

(a) *Let g be continuously differentiable from an open set $O \subset \mathbb{R}^n$ to $\mathbb{R}$, and let C be a convex subset of O. Show that the function $f(x) = \nabla g(x)$ is monotone on C if and only if g is convex on C. Show further that f is strictly monotone on C if and only if g is strictly convex on C. Finally, show that f is strongly monotone on C with constant μ if and only if g is strongly convex on C with constant μ.*

(b) *Let h be continuously differentiable from a product $O_1 \times O_2$ of open sets $O_1 \subset \mathbb{R}^{n_1}$ and $O_2 \subset \mathbb{R}^{n_2}$ to $\mathbb{R}$, and let $C_1 \subset O_1$ and $C_2 \subset O_2$ be convex. Show that the function*

$$f(x_1,x_2) = (\nabla_{x_1} h(x_1,x_2), -\nabla_{x_2} h(x_1,x_2))$$

is monotone on $C_1 \times C_2$ if and only if $h(x_1,x_2)$ is convex with respect to $x_1 \in C_1$ for fixed $x_2 \in C_2$, and on the other hand concave with respect to $x_2 \in C_2$ for fixed $x_1 \in C_1$.

Guide. Derive (a) from the characterization of the convexity and strong convexity of g in 2A.5. Proceed similarly in (b), applying also the corresponding characterization of concavity. □

Optimization problems. In this chapter and later, we consider optimization problems which, for a given *objective function* $g : \mathbb{R}^n \to \mathbb{R}$ and a given *constraint set* $C \subset \mathbb{R}^n$, take the form

$$\text{minimize } g(x) \text{ over all } x \in C.$$

The greatest lower bound of the objective function g on C, namely $\inf_{x \in C} g(x)$, is the *optimal value* in the problem, which may or may not be attained, however, and could even be infinite. If it is attained at a point $\bar{x}$, then $\bar{x}$ is said to furnish a *global minimum*, or just a minimum, and to be a *globally optimal solution*; the set of such points is denoted as $\operatorname{argmin}_{x \in C} g(x)$. A point $x \in C$ is said to furnish a *local minimum* of g relative to C and to be a *locally optimal solution* when, at least, $g(x) \leq g(x')$ for every $x' \in C$ belonging to some neighborhood of x. A global or local *maximum* of g corresponds to a global or local minimum of $-g$.

In the context of variational inequalities, the gradient mapping $\nabla g : \mathbb{R}^n \to \mathbb{R}^n$ associated with a differentiable function $g : \mathbb{R}^n \to \mathbb{R}$ will be a focus of attention. Observe that

$$\nabla^2 g(x) = \nabla f(x) \ \text{ when } \ f(x) = \nabla g(x).$$

Theorem 2A.7 (basic variational inequality for minimization)**.** *Let $g : \mathbb{R}^n \to \mathbb{R}$ be differentiable on an open convex set O, and let C be a closed convex subset of O. In minimizing g over C, the variational inequality*

(10) $$\nabla g(x) + N_C(x) \ni 0, \quad \textit{or equivalently} \quad -\nabla g(x) \in N_C(x),$$

is necessary for x to furnish a local minimum. It is both necessary and sufficient for a global minimum if g is convex.

Proof. Along with $x \in C$, consider any other point $x' \in C$ and the function $\varphi(t) = g(x+tw)$ with $w = x' - x$. From convexity we have $x + tw \in C$ for all $t \in [0,1]$. If a local minimum of g occurs at x relative to C, then φ must have a local minimum at 0 relative to $[0,1]$, and consequently $\varphi'(0) \geq 0$. But $\varphi'(0) = \langle \nabla g(x), w \rangle$. Hence $\langle \nabla g(x), x' - x \rangle \geq 0$. This being true for arbitrary $x' \in C$, we conclude through the characterization of (2) in (3) that $-\nabla g(x) \in N_C(x)$.

In the other direction, if g is convex and $-\nabla g(x) \in N_C(x)$ we have for every $x' \in C$ that $\langle \nabla g(x), x' - x \rangle \geq 0$, but also $g(x') - g(x) \geq \langle \nabla g(x), x' - x \rangle$ by the convexity criterion in 2A.5(a). Hence $g(x') - g(x) \geq 0$ for all $x' \in C$, and we have a global minimum at x. $\square$

To illustrate the condition in Theorem 2A.7, we may use it to reconfirm the projection rule for normal vectors in (4), which can be stated equivalently as saying that $P_C(z) = x$ if and only if $z - x \in N_C(x)$. Consider any nonempty, closed, convex set $C \subset \mathbb{R}^n$ and any point $z \in \mathbb{R}^n$. Let $g(x) = \frac{1}{2}|x - z|^2$, which has $\nabla g(x) = x - z$ and $\nabla^2 g(x) \equiv I$, implying strong convexity. The projection $x = P_C(z)$ is the solution to the problem of minimizing g over C. The variational inequality (10) characterizes it by the relation $-(x - z) \in N_C(x)$, which is exactly what was targeted.

According to Theorem 2A.7, minimizing a differentiable convex function g over a closed, convex set C is equivalent to solving a type of variational inequality (2) in which f is the gradient mapping ∇g. When $C = \mathbb{R}^n$, so that we are dealing with unconstrained minimization, this is equivalent to solving $f(x) = 0$ for $f = \nabla g$. The notion of a variational inequality thus makes it possible to pass from unconstrained minimization to constrained minimization. Whether the problem is constrained or unconstrained, there is no guarantee that the minimum will be attained at a unique point (although nonuniqueness is impossible when g is strictly convex, at least), but still, local uniqueness dominates the picture conceptually. For that reason, it does make sense to be thinking of the task as one of "solving a generalized equation."

When g is not convex, solving the variational inequality (2) is no longer equivalent to minimization over C, but nevertheless it has a strong association with identifying a local minimum. Anyway, there's no need really to insist on a minimum. Just as the equation $\nabla g(x) = 0$ describes, in general, a "stationary point" of g (unconstrained), the variational inequality (10) can be viewed as describing a constrained version of a stationary point, which could be of interest in itself.

The minimization rule in Theorem 2A.7 can be employed to deduce a rule for determining normal vectors to intersections of convex sets, as in the second part of the following proposition.

Proposition 2A.8 (normals to products and intersections)**.**
(a) *If $C = C_1 \times C_2$ for closed, convex sets $C_1 \subset \mathbb{R}^{n_1}$ and $C_2 \subset \mathbb{R}^{n_2}$, then for any $x = (x_1, x_2) \in C$ one has $N_C(x) = N_{C_1}(x_1) \times N_{C_2}(x_2)$.*
(b) *If $C = C_1 \cap C_2$ for closed, convex sets C_1 and C_2 in $\mathbb{R}^n$, then the formula*

$$N_C(x) = N_{C_1}(x) + N_{C_2}(x) = \left\{ v_1 + v_2 \,\middle|\, v_1 \in N_{C_1}(x),\ v_2 \in N_{C_2}(x) \right\}$$

holds for every $x \in C$ such that there is no $v \neq 0$ with $v \in N_{C_1}(x)$ and $-v \in N_{C_2}(x)$. This condition is fulfilled in particular for every $x \in C$ if $C_1 \cap \operatorname{int} C_2 \neq \emptyset$ or $C_2 \cap \operatorname{int} C_1 \neq \emptyset$.

Proof. To prove (a), we note that, by definition, a vector $v = (v_1, v_2)$ belongs to $N_C(x)$ if and only if, for every $x' = (x_1', x_2')$ in $C_1 \times C_2$ we have $0 \geq \langle v, x' - x \rangle = \langle v_1, x_1' - x_1 \rangle + \langle v_2, x_2' - x_2 \rangle$. That's the same as having $\langle v_1, x_1' - x_1 \rangle \leq 0$ for all $x_1' \in C_1$ and $\langle v_2, x_2' - x_2 \rangle \leq 0$ for all $x_2' \in C_2$, or in other words, $v_1 \in N_{C_1}(x_1)$ and $v_2 \in N_{C_2}(x_2)$.

In proving (b), it is elementary that if $v = v_1 + v_2$ with $v_1 \in N_{C_1}(x)$ and $v_2 \in N_{C_2}(x)$, then for every x' in $C_1 \cap C_2$ we have both $\langle v_1, x' - x \rangle \leq 0$ and $\langle v_2, x' - x \rangle \leq 0$, so that $\langle v, x' - x \rangle \leq 0$. Thus, $N_C(x) \supset N_{C_1}(x) + N_{C_2}(x)$.

The opposite inclusion takes more work to establish. Fix any $x \in C$ and $v \in N_C(x)$. As we know from (4), this corresponds to x being the projection of $x + v$ on C, which we can elaborate as follows: (x, x) is the unique solution to the problem

$$\text{minimize } |x_1 - (x+v)|^2 + |x_2 - (x+v)|^2 \text{ over all } (x_1, x_2) \in C_1 \times C_2 \text{ with } x_1 = x_2.$$

Consider for $k = 1, 2, \ldots$ the version of this minimization problem in which the constraint $x_1 = x_2$ is relaxed by a penalty expression dependent on k:

$$\text{(11)} \qquad \begin{array}{l} \text{minimize } |x_1 - (x+v)|^2 + |x_2 - (x+v)|^2 + k|x_1 - x_2|^2 \\ \text{over all } (x_1, x_2) \in C_1 \times C_2. \end{array}$$

The expression being minimized here is nonnegative and, as seen from the case of $x_1 = x_2 = x$, has minimum no greater than $2|v|^2$. It suffices therefore in the minimization to consider only points x_1 and x_2 such that $|x_1 - (x+v)|^2 + |x_2 - (x+v)|^2 \leq 2|v|^2$ and $k|x_1 - x_2|^2 \leq 2|v|^2$. For each k, therefore, the minimum in (11) is attained by some (x_1^k, x_2^k), and these pairs form a bounded sequence such that $x_1^k - x_2^k \to 0$. Any accumulation point of this sequence must be of the form $(\tilde{x}, \tilde{x})$ and satisfy $|\tilde{x} - (x+v)|^2 + |\tilde{x} - (x+v)|^2 \leq 2|v|^2$, or in other words $|\tilde{x} - (x+v)| \leq |v|$. But by the projection rule (4), x is the unique closest point of C to $x + v$, the distance being $|v|$, so this inequality implies $\tilde{x} = x$. Therefore, $(x_1^k, x_2^k) \to (x, x)$.

We investigate next the necessary condition for optimality (10) provided by Theorem 2A.7 for problem (11). Invoking the formula in (a) for the normal cone to $C_1 \times C_2$ at (x_1^k, x_2^k), we see that it requires

$$-2[x_1^k - (x+v) + k(x_1^k - x_2^k)] \in N_{C_1}(x_1^k),$$
$$-2[x_2^k - (x+v) - k(x_1^k - x_2^k)] \in N_{C_2}(x_2^k),$$

or equivalently, for $w^k = k(x_2^k - x_1^k)$,

$$(12) \qquad v + (x - x_1^k) + w^k \in N_{C_1}(x_1^k) \text{ and } v + (x - x_2^k) - w^k \in N_{C_2}(x_2^k).$$

Two cases have to be analyzed now separately. In the first case, we suppose that the sequence of vectors w^k is bounded and therefore has an accumulation point w. Let $v_1^k = v + (x - x_1^k) + w^k$ and $v_2^k = v + (x - x_2^k) - w^k$, so that, through (4), we have $P_{C_1}(x_1^k + v_1^k) = x_1^k$ and $P_{C_2}(x_2^k + v_2^k) = x_2^k$. Since $x_1^k \to x$ and $x_2^k \to x$, the sequences of vectors v_1^k and v_2^k have accumulation points $v_1 = v + w$ and $v_2 = v - w$; note that $v_1 + v_2 = 2v$. By the continuity of the projection mappings coming from 1D.5, we get $P_{C_1}(x + v_1) = x$ and $P_{C_2}(x + v_2) = x$. By (6), these relations mean $v_1 \in N_{C_1}(x)$ and $v_2 \in N_{C_2}(x)$ and hence $2v \in N_{C_1}(x) + N_{C_2}(x)$. Since the sum of cones is a cone, we get $v \in N_{C_1}(x) + N_{C_2}(x)$. Thus $N_C(x) \subset N_{C_1}(x) + N_{C_2}(x)$, and since we have already shown the opposite inclusion, we have equality.

In the second case, we suppose that the sequence of vectors w^k is unbounded. By passing to a subsequence if necessary, we can reduce this to having $0 < |w^k| \to \infty$ with $w^k/|w^k|$ converging to some $\bar{v} \neq 0$. Let

$$\bar{v}_1^k = [v + (x - x_1^k) + w^k]/|w^k| \text{ and } \bar{v}_2^k = [v + (x - x_2^k) - w^k]/|w^k|.$$

Then $\bar{v}_1^k \to \bar{v}$ and $\bar{v}_2^k \to -\bar{v}$. By (12) we have $\bar{v}_1^k \in N_{C_1}(x_1^k)$ and $\bar{v}_2^k \in N_{C_1}(x_2^k)$, or equivalently through (4), the projection relations $P_{C_1}(x_1^k + \bar{v}_1^k) = x_1^k$ and $P_{C_2}(x_2^k + \bar{v}_2^k) = x_2^k$. In the limit we get $P_{C_1}(x + \bar{v}) = x$ and $P_{C_2}(x - \bar{v}) = x$, so that $\bar{v} \in N_{C_1}(x)$ and $-\bar{v} \in N_{C_2}(x)$. This contradicts our assumption in (b), and we see thereby that the second case is impossible. □

We turn now to minimization over sets C that might not be convex and are specified by systems of constraints which have to be handled with Lagrange multipliers. This will lead us to other valuable examples of variational inequalities, after some elaborations.

Theorem 2A.9 (Lagrange multiplier rule). *Let $X \subset I\!R^n$ and $D \subset I\!R^m$ be nonempty, closed, convex sets, and consider the problem*

$$(13) \qquad \text{minimize } g_0(x) \text{ over } C = \left\{ x \in X \,\middle|\, g(x) \in D \right\},$$

for $g(x) = (g_1(x), \ldots, g_m(x))$, where the functions $g_i : I\!R^n \to I\!R$, $i = 0, 1, \ldots, m$ are continuously differentiable. Let x be a point of C at which the following constraint qualification condition is fulfilled:

$$(14) \qquad \text{there is no } y \in N_D(g(x)),\ y \neq 0, \text{ such that } -y\nabla g(x) \in N_X(x).$$

If g_0 has a local minimum relative to C at x, then

$$\text{(15)} \qquad \textit{there exists } y \in N_D(g(x)) \textit{ such that } -[\nabla g_0(x) + y\nabla g(x)] \in N_X(x).$$

Proof. Assume that a local minimum occurs at x. Let X' and D' be compact, convex sets which coincide with X and D in neighborhoods of x and $g(x)$, respectively, and are small enough that $g_0(x') \geq g_0(x)$ for all $x' \in X'$ having $g(x') \in D'$. Consider the auxiliary problem

$$\text{(16)} \qquad \begin{array}{l} \text{minimize } \; g_0(x') + \frac{1}{2}|x' - x|^2 \\ \text{over all } \; (x', u') \in X' \times D' \; \text{ satisfying } \; g(x') - u' = 0. \end{array}$$

Obviously the unique solution to this is $(x', u') = (x, g(x))$. Next, for $k \to \infty$, consider the following sequence of problems, which replace the equation in (16) by a penalty expression:

$$\text{(17)} \qquad \text{minimize } \; g_0(x') + \frac{1}{2}|x' - x|^2 + \frac{k}{2}|g(x') - u'|^2 \; \text{ over all } \; (x', u') \in X' \times D'.$$

For each k let (x_k, u_k) give the minimum in this relaxed problem (the minimum being attained because the functions are continuous and the sets X' and D' are compact). The minimum value in (17) can't be greater than the minimum value in (16), as seen by taking (x', u') to be the unique solution $(x, g(x))$ to (16). It's apparent then that the only possible cluster point of the bounded sequence $\{(x_k, u_k)\}_{k=1}^{\infty}$ as $k \to \infty$ is $(x, g(x))$. Thus, $(x_k, u_k) \to (x, g(x))$.

Next we apply the optimality condition in Theorem 2A.7 to problem (17) at its solution (x_k, u_k). We have $N_{X' \times D'}(x_k, u_k) = N_{X'}(x_k) \times N_{D'}(u_k)$, and on the other hand $N_{X'}(x_k) = N_X(x_k)$ and $N_{D'}(u_k) = N_D(u_k)$ by the choice of X' and D', at least when k is sufficiently large. The variational inequality condition in Theorem 2A.7 comes down in this way to

$$\text{(18)} \qquad \begin{cases} -[\nabla g_0(x_k) + (x_k - x) + k(g(x_k) - u_k)\nabla g(x_k)] \in N_X(x_k), \\ k(g(x_k) - u_k) \in N_D(u_k). \end{cases}$$

By passing to subsequences if necessary, we can reduce the rest of the analysis to distinguishing between case (A), where the norms of the vectors $k(g(x_k) - u_k) \in \mathbb{R}^m$ stay bounded as $k \to \infty$, and case (B), where these norms go to ∞.

In case (A) we can arrange, by passing again to a subsequence if necessary, that the sequence of vectors $k(g(x_k) - u_k)$ converges to some y. Then y satisfies the desired relations in (18), inasmuch as $(x_k, u_k) \to (x, g(x))$ and the graphs of the mappings N_X and N_D are closed.

In case (B) we look at the vectors $y_k = k(g(x_k) - u_k)/\rho_k$ with $\rho_k = k|g(x_k) - u_k| \to \infty$, which have $|y_k| = 1$ and, from (18), satisfy

$$\text{(19)} \qquad -\rho_k^{-1}(\nabla g_0(x_k) + (x_k - x)) - y_k \nabla g(x_k) \in N_X(x_k), \qquad y_k \in N_D(u_k).$$

(Here we use the fact that any positive multiple of a vector in $N_X(x_k)$ or $N_D(u_k)$ is another such vector.) By passing to a subsequence, we can arrange that the sequence of vectors y_k converges to some y, necessarily with $|y| = 1$. In this limit, (19) turns into the relation in (14), which has been forbidden to hold for any $y \neq 0$. Hence case (B) is impossible under our assumptions, and we are left with the conclusion (15) obtained from case (A). □

In the first-order optimality condition (15), y is said to be a *Lagrange multiplier vector* associated with x. More can be said about this condition by connecting it with the *Lagrangian function* for problem (13), which is defined by

$$(20) \qquad L(x,y) = g_0(x) + \langle y, g(x)\rangle = g_0(x) + y_1 g_1(x) + \cdots + y_m g_m(x)$$

for $y = (y_1, \ldots, y_m)$.

Theorem 2A.10 (Lagrangian variational inequalities)**.** *In the minimization problem (13), suppose that the set D is a cone, and let Y be the polar cone D^*,*

$$Y = \left\{ y \,\middle|\, \langle u, y\rangle \leq 0 \text{ for all } u \in D \right\}.$$

Then, in terms of the Lagrangian L in (20), the condition on x and y in (15) can be written in the form

$$(21) \qquad -\nabla_x L(x,y) \in N_X(x), \qquad \nabla_y L(x,y) \in N_Y(y),$$

which furthermore can be identified with the variational inequality

$$(22) \qquad -f(x,y) \in N_{X\times Y}(x,y) \;\text{ for } f(x,y) = (\nabla_x L(x,y), -\nabla_y L(x,y)).$$

The existence of $y \in Y$ satisfying this variational inequality with x is thus necessary for the local optimality of x in problem (13) when the constraint qualification (14) is fulfilled. If $L(\cdot, y)$ is convex on X when $y \in Y$, the existence of a y satisfying this variational inequality with x is moreover sufficient for x to give a global minimum in problem (13), without any need for invoking (14).

Proof. We have $\nabla_x L(x,y) = \nabla g_0(x) + y\nabla g(x)$ and $\nabla_y L(x,y) = g(x)$. The N_X condition in (15) amounts therefore to the first condition in (21). The choice of $Y = D^*$ makes it possible through the polarity rule for normal vectors in (8) to express the N_D condition in (15) as $g(x) \in N_Y(y)$ and identify it with the second condition in (21), while deducing from it also that $\langle y, g(x)\rangle = 0$, hence $L(x,y) = g_0(x)$. The recasting of (21) as the variational inequality in (22) comes out of the product rule in 2A.8(a).

When the function $L(\cdot, y)$ is convex on X, the condition $-\nabla_x L(x,y) \in N_X(x)$ implies through Theorem 2A.7 that $L(x', y) \geq L(x,y)$ for all $x' \in X$, where it may be recalled that $L(x,y) = g_0(x)$ because $\langle y, g(x)\rangle = 0$. Thus, $L(x',y) \geq g_0(x)$ for all $x' \in X$. On the other hand, since $y \in Y$ and $Y = D^*$, we have $\langle y, g(x')\rangle \leq 0$ when $g(x') \in D$. Therefore $g_0(x') \geq L(x',y) \geq g_0(x)$ for all x' satisfying the constraints in (13). It follows that all such x' have $g_0(x') \geq g_0(x)$, so x furnishes the global minimum in problem (13). □

Application to nonlinear programming. Theorems 2A.9 and 2A.10 cover the case of a standard problem of *nonlinear programming*, where the task is to

$$\text{(23)} \qquad \text{minimize } g_0(x) \text{ over all } x \text{ satisfying } g_i(x) \begin{cases} \leq 0 & \text{for } i \in [1,s], \\ = 0 & \text{for } i \in [s+1,m]. \end{cases}$$

This problem[1] corresponds in (13) to taking $X = I\!R^n$ and having D be the closed, convex cone in $I\!R^m$ consisting of all $u = (u_1, \ldots, u_m)$ such that $u_i \leq 0$ for $i \in [1,s]$ but $u_i = 0$ for $i \in [s+1,m]$. The polar cone $Y = D^*$ is $Y = I\!R^s_+ \times I\!R^{m-s}$. The optimality condition in (18) can equally well be placed then in the Lagrangian framework in (21), corresponding to the variational inequality (22). The requirements it imposes on x and y come out as

$$\text{(24)} \qquad \begin{array}{l} y \in I\!R^s_+ \times I\!R^{m-s}, \;\; g_i(x) \begin{cases} \leq 0 & \text{for } i \in [1,s] \text{ with } y_i = 0, \\ = 0 & \text{for all other } i \in [1,m], \end{cases} \\ \nabla g_0(x) + y_1 \nabla g_1(x) + \cdots + y_m \nabla g_m(x) = 0. \end{array}$$

These are the *Karush–Kuhn–Tucker conditions* for the nonlinear programming problem (23). According to Theorem 2A.9, the existence of y satisfying these conditions with x is necessary for the local optimality of x under the constraint qualification (14), which insists on the nonexistence of $y \neq 0$ satisfying the same conditions with the term $\nabla g_0(x)$ suppressed. The existence of y satisfying (24) is sufficient for the global optimality of x by Theorem 2A.10 as long as $L(x,y)$ is convex as a function of $x \in I\!R^n$ for each fixed $y \in I\!R^s_+ \times I\!R^{m-s}$, which is equivalent to having

$$g_0, g_1, \ldots, g_s \text{ convex, but } g_{s+1}, \ldots, g_m \text{ affine.}$$

Then (23) is a problem of *convex programming*. The Karush–Kuhn–Tucker conditions correspond then to a saddle point property, as indicated next.

Exercise 2A.11 (variational inequality for a saddle point). *Let $X \subset I\!R^n$ and $Y \subset I\!R^m$ be any nonempty, closed, convex sets, and let L be a $\mathscr{C}^1$ function on $I\!R^n \times I\!R^m$ such that $L(\cdot, y)$ is a convex function on X for each $y \in Y$, and $L(x, \cdot)$ is a concave function on Y for each $x \in X$. The variational inequality (22) is equivalent to having (x,y) be a* saddle point *of L with respect to $X \times Y$ in the sense that*

$$x \in X, \; y \in Y, \text{ and } L(x', y) \geq L(x,y) \geq L(x,y') \text{ for all } x' \in X, \; y' \in Y.$$

Guide. Rely on the equivalence between (21) and (22), plus Theorem 2A.7. □

A saddle point as defined in Exercise 2A.11 represents an equilibrium in the two-person zero-sum game in which Player 1 chooses $x \in X$, Player 2 chooses $y \in Y$, and then Player 1 pays the amount $L(x,y)$ (possibly negative) to Player 2. Other kinds of equilibrium can likewise be captured by other variational inequalities.

For example, in an N-person game there are players $1, \ldots, N$, with Player k having a nonempty strategy set X_k. Each Player k chooses some $x_k \in X_k$, and is then

[1] In (23) and later in the book $[1,s]$ denotes the set of integers $\{1, 2, \ldots, s\}$.

obliged to pay—to an abstract entity (not necessarily another player)—an amount which depends not only on x_k but also on the choices of all the other players; this amount can conveniently be denoted by

$$L_k(x_k, x_{-k}), \text{ where } x_{-k} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_N).$$

(The game is *zero-sum* if $\sum_{k=1}^{N} L_k(x_k, x_{-k}) = 0$.) A choice of strategies $x_k \in X_k$ for $k = 1, \ldots, N$ is said to furnish a *Nash equilibrium* if

$$L_k(x'_k, x_{-k}) \geq L_k(x_k, x_{-k}) \text{ for all } x'_k \in X_k,\ k = 1, \ldots, N.$$

A saddle point as in Exercise 2A.11 corresponds to the case of this where $N = 2$, so x_{-1} and x_{-2} are just x_2 and x_1 respectively, and one has $L_2(x_2, x_1) = -L_1(x_1, x_2)$.

Exercise 2A.12 (variational inequality for a Nash equilibrium)**.** *In an N-person game as described, suppose that X_k is a closed, convex subset of $\mathbb{R}^{n_k}$ and that $L(x_k, x_{-k})$ is differentiable with respect to x_k for every k. Then for $x = (x_1, \ldots, x_N)$ to furnish a Nash equilibrium, it must solve the variational inequality (2) for f and C in the case of*

$$C = X_1 \times \cdots \times X_N, \quad f(x) = f(x_1, \ldots, x_N) = \Big(\nabla_{x_1} L_1(x_1, x_{-1}), \ldots, \nabla_{x_N} L_N(x_N, x_{-N})\Big)^{\mathsf{T}}.$$

This necessary condition is sufficient for a Nash equilibrium if, in addition, the functions $L_k(\cdot, x_{-k})$ on $\mathbb{R}^{n_k}$ are convex.

Guide. Make use of the product rule for normals in 2A.8(a) and the optimality condition in Theorem 2A.7. □

Finally, we look at a kind of generalized equation (1) that is *not* a variational inequality (2), but nonetheless has importance in many situations:

$$(25) \qquad (g_1(x), \ldots, g_m(x)) \in D,$$

which is (1) for $f(x) = -(g_1(x), \ldots, g_m(x))$, $F(x) \equiv D$. Here D is a subset of $\mathbb{R}^m$; the format has been chosen to be that of the constraints in problem (13), or as a special case, problem (23).

Although (25) would reduce to an equation, pure and simple, if D consists of a single point, the applications envisioned for it lie mainly in situations where inequality constraints are involved, and there is little prospect or interest in a solution being locally unique. In the study of generalized equations with parameters, to be taken up next in 2B, our attention will at first be concentrated on issues parallel to those in Chapter 1. Only later, in Chapter 3, will generalized equations like (25) be brought in.

The example in (25) also brings a reminder about a feature of generalized equations which dropped out of sight in the discussion of the variational inequality case. In (2), necessarily f had to go from $\mathbb{R}^n$ to $\mathbb{R}^n$, whereas in (25), and in (1), f may go from $\mathbb{R}^n$ to a space $\mathbb{R}^m$ of different dimension. Later in the book, and in particular in chapters 5 and 6 we will see that the generalized equations cover much larger territory than the variational inequalities.

2B. Implicit Function Theorems for Generalized Equations

With the concept of a generalized equation, and in particular that of a variational inequality problem at our disposal, we are ready to embark on a broad exploration of implicit function theorems beyond those in Chapter 1. The object of study is now a *parameterized* generalized equation

$$(1) \qquad f(p,x)+F(x) \ni 0$$

for a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. Specifically, we consider the properties of the *solution mapping* $S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ defined by

$$(2) \qquad S : p \mapsto \{ x \mid f(p,x)+F(x) \ni 0 \} \quad \text{for } p \in \mathbb{R}^d.$$

The questions we will concentrate on answering, for now, are nevertheless the same as in Chapter 1. To what extent might S be single-valued and possess various properties of continuity or some type of differentiability?

In a landmark paper[2], S. M. Robinson studied the solution mapping S in the case of a parameterized variational inequality, where $m = n$ and F is a normal cone mapping $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$:

$$(3) \qquad f(p,x)+N_C(x) \ni 0, \quad \text{with } C \subset \mathbb{R}^n \text{ convex, closed and nonempty.}$$

His results were, from the very beginning, stated in abstract spaces, and we will come to that in Chapter 5. Here, we confine the exposition to Euclidean spaces, but the presentation is tailored in such a way that, for readers who are familiar with some basic functional analysis, the expansion of the framework from Euclidean spaces to general Banach spaces is straightforward. The original formulation of Robinson's theorem, up to some rewording to fit this setting, is as follows.

Theorem 2B.1 (Robinson implicit function theorem)**.** *For the solution mapping S to a parameterized variational inequality (3), consider a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$. Assume that:*

(a) *$f(p,x)$ is differentiable with respect to x in a neighborhood of the point $(\bar{p},\bar{x})$, and both $f(p,x)$ and $\nabla_x f(p,x)$ depend continuously on (p,x) in this neighborhood;*

(b) *the inverse G^{-1} of the set-valued mapping $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ defined by*

$$(4) \qquad G(x) = f(\bar{p},\bar{x})+\nabla_x f(\bar{p},\bar{x})(x-\bar{x})+N_C(x), \quad \text{with } G(\bar{x}) \ni 0,$$

[2] Cf. Robinson [1980].

has a Lipschitz continuous single-valued localization σ *around* 0 *for* $\bar{x}$ *with*

$$\operatorname{lip}(\sigma;0) \leq \kappa.$$

Then S *has a single-valued localization* s *around* $\bar{p}$ *for* $\bar{x}$ *which is continuous at* $\bar{p}$*, and moreover for every* $\varepsilon > 0$ *there is a neighborhood* Q *of* $\bar{p}$ *such that*

$$(5) \qquad |s(p') - s(p)| \leq (\kappa + \varepsilon)|f(p', s(p)) - f(p, s(p))| \text{ for all } p', p \in Q.$$

An extended version of this result will be stated shortly as Theorem 2B.5, so we can postpone the discussion of its proof until then. Instead, we can draw some immediate conclusions from the estimate (5) which rely on additional assumptions about partial calmness and Lipschitz continuity properties of $f(p,x)$ with respect to p and the modulus notation for such properties that was introduced in 1C and 1D.

Corollary 2B.2 (calmness of solutions)**.** *In the setting of Theorem 2B.1, if* f *is calm with respect to* p *at* $(\bar{p},\bar{x})$*, having* $\operatorname{clm}_p(f;(\bar{p},\bar{x})) \leq \lambda$*, then* s *is calm at* $\bar{p}$ *with* $\operatorname{clm}(s;\bar{p}) \leq \kappa\lambda$*.*

Corollary 2B.3 (Lipschitz continuity of solutions)**.** *In the setting of Theorem 2B.1, if* f *is Lipschitz continuous with respect to* p *uniformly in* x *around* $(\bar{p},\bar{x})$*, having* $\widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})) \leq \lambda$*, then* s *is Lipschitz continuous around* $\bar{p}$ *with* $\operatorname{lip}(s;\bar{p}) \leq \kappa\lambda$*.*

Differentiability of the localization s around $\bar{p}$ can't be deduced from the estimate in (5), not to speak of continuous differentiability around $\bar{p}$, and in fact differentiability may fail. Elementary one-dimensional examples of variational inequalities exhibit solution mappings that are not differentiable, usually in connection with the "solution trajectory" hitting or leaving the boundary of the set C. For such mappings, weaker concepts of differentiability are available. We will touch upon this in 2D.

In the special case where the variational inequality treated by Robinson's theorem reduces to the equation $f(p,x) = 0$ (namely with $C = \mathbb{R}^n$, so $N_C \equiv 0$), the invertibility condition on the mapping G in assumption (b) of Robinson's theorem comes down to the nonsingularity of the Jacobian $\nabla_x f(\bar{p},\bar{x})$ in the Dini classical implicit function theorem 1B.1. But because of the absence of an assertion about the differentiability of s, Theorem 2B.1 falls short of yielding all the conclusions of that theorem. It could, though, be used as an intermediate step in a proof of Theorem 1B.1, which we leave to the reader as an exercise.

Exercise 2B.4. *Supply a proof of the classical implicit function theorem 1B.1 based on Robinson's theorem 2B.1.*

Guide. In the case $C = \mathbb{R}^n$, so $N_C \equiv 0$, use the Lipschitz continuity of the single-valued localization s following from Corollary 2B.3 to show that s is continuously differentiable around $\bar{p}$ when f is continuously differentiable near $(\bar{p},\bar{x})$. □

The invertibility property in assumption (b) of 2B.1 is what Robinson called "strong regularity" of the generalized equation (3). A related term, "strong metric regularity," will be employed in Chapter 3 for set-valued mappings in reference to the existence of Lipschitz continuous single-valued localizations of their inverses.

In the extended version of Theorem 2B.1 which we present next, the differentiability assumptions on f are replaced by assumptions about an estimator h for $f(\bar{p},\cdot)$, which could in particular be a first-order approximation in the x argument. This mode of generalization was initiated in 1E.3 and 1E.13 for equations, but now we use it for a generalized equation (1). In contrast to Theorem 2B.1, which was concerned with the case of a variational inequality (3), the mapping $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ need not be of form N_C and the dimensions n and m could in principle be different. Remarkably, no *direct* assumptions need be made about F, but certain properties of F will implicitly underlie the "invertibility" condition imposed jointly on F and the estimator h.

Theorem 2B.5 (Robinson theorem extended beyond differentiability)**.** *For a generalized equation (1) and its solution mapping S in (2), let $\bar{p}$ and $\bar{x}$ be such that $\bar{x}\in S(\bar{p})$. Assume that:*

(a) *$f(\cdot,\bar{x})$ is continuous at $\bar{p}$, and h is a strict estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ;*

(b) *the inverse G^{-1} of the mapping $G=h+F$, for which $G(\bar{x})\ni 0$, has a Lipschitz continuous single-valued localization σ around 0 for $\bar{x}$ with $\operatorname{lip}(\sigma;0)\le\kappa$ for a constant κ such that $\kappa\mu<1$.*

Then S has a single-valued localization s around $\bar{p}$ for $\bar{x}$ which is continuous at $\bar{p}$, and moreover for every $\varepsilon>0$ there is a neighborhood Q of $\bar{p}$ such that

$$(6)\qquad |s(p')-s(p)|\le\frac{\kappa+\varepsilon}{1-\kappa\mu}\,|f(p',s(p))-f(p,s(p))| \quad \text{for all } p',p\in Q.$$

Theorem 2B.1 follows at once from Theorem 2B.5 by taking F to be N_C and h to be the linearization of $f(\bar{p},\cdot)$ given by $h(x)=f(\bar{p},\bar{x})+\nabla_x f(\bar{p},\bar{x})(x-\bar{x})$, and employing 1E.15. More generally, h could be a strict first-order approximation: the case when $\mu=0$. That case, which has further implications, will be taken up later. However, Theorem 2B.5 is able to extract information from much weaker relationships between f and h than strict first-order approximation, and this information can still have important consequences for the behavior of solutions to a generalized equation, as seen in this pattern already for equations in 1E.

Our proof of Theorem 2B.5 will proceed through an intermediate stage in which we isolate an equivalent formulation of the contracting mapping principle 1A.2, with a somewhat lengthy statement.

Theorem 2B.6 (contracting mapping principle for composition)**.** *Consider a function $\varphi:\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^m$ and a point $(\bar{p},\bar{x})\in\operatorname{int}\operatorname{dom}\varphi$ and let the scalars $\nu\ge 0$, $b\ge 0$, $a>0$, and the set $Q\subset\mathbb{R}^d$ be such that $\bar{p}\in Q$ and*

$$(7)\qquad \begin{cases} |\varphi(p,x')-\varphi(p,x)|\le\nu|x-x'| & \text{for all } x',x\in\mathbb{B}_a(\bar{x}) \text{ and } p\in Q,\\ |\varphi(p,\bar{x})-\varphi(\bar{p},\bar{x})|\le b & \text{for all } p\in Q.\end{cases}$$

Consider also a set-valued mapping $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *with* $(\bar{y}, \bar{x}) \in \operatorname{gph} M$ *where* $\bar{y} := \varphi(\bar{p}, \bar{x})$, *such that for each* $y \in \mathbb{B}_{\nu a+b}(\bar{y})$ *the set* $M(y) \cap \mathbb{B}_a(\bar{x})$ *consists of exactly one point, denoted by* $r(y)$, *and suppose that the function*

$$(8) \qquad r : y \mapsto M(y) \cap \mathbb{B}_a(\bar{x}) \quad \text{for } y \in \mathbb{B}_{\nu a+b}(\bar{y})$$

is Lipschitz continuous on $\mathbb{B}_{\nu a+b}(\bar{y})$ *with a Lipschitz constant* λ. *In addition, suppose that*

(a) $\lambda\nu < 1$;
(b) $\lambda\nu a + \lambda b \le a$.

Then for each $p \in Q$ *the set* $\{x \in \mathbb{B}_a(\bar{x}) \mid x \in M(\varphi(p,x))\}$ *consists of exactly one point, and the associated function*

$$(9) \qquad s : p \mapsto \{x \mid x = M(\varphi(p,x)) \cap \mathbb{B}_a(\bar{x})\} \quad \text{for } p \in Q$$

satisfies

$$(10) \qquad |s(p') - s(p)| \le \frac{\lambda}{1-\lambda\nu} |\varphi(p', s(p)) - \varphi(p, s(p))| \quad \text{for all } p', p \in Q.$$

Proof. Fix $p \in Q$ and consider the function $\Phi_p : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\Phi_p : x \mapsto r(\varphi(p,x)) \text{ for } x \in \mathbb{B}_a(\bar{x}).$$

First, note that for $x \in \mathbb{B}_a(\bar{x})$ from (7) one has $|\bar{y} - \varphi(p,x)| \le b + \nu a$, thus, by (8), $\mathbb{B}_a(\bar{x}) \subset \operatorname{dom} \Phi_p$. Next, if $x \in \mathbb{B}_a(\bar{x})$, we have from the identity $\bar{x} = r(\varphi(\bar{p}, \bar{x}))$, the Lipschitz continuity of r, and conditions (7) and (b) that

$$|\Phi_p(\bar{x}) - \bar{x}| = |r(\varphi(p,\bar{x})) - r(\varphi(\bar{p},\bar{x}))| \le \lambda |\varphi(p,\bar{x}) - \varphi(\bar{p},\bar{x})| \le \lambda b \le a(1-\lambda\nu).$$

For any $x', x \in \mathbb{B}_a(\bar{x})$ we obtain

$$|\Phi_p(x') - \Phi_p(x)| = |r(\varphi(p,x')) - r(\varphi(p,x))| \le \lambda |\varphi(p,x') - \varphi(p,x)| \le \lambda\nu |x' - x|,$$

that is, Φ_p is Lipschitz continuous in $\mathbb{B}_a(\bar{x})$ with constant $\lambda\nu < 1$, from condition (a). We are in position then to apply the contraction mapping principle 1A.2 and to conclude from it that Φ_p has a unique fixed point in $\mathbb{B}_a(\bar{x})$.

Denoting that fixed point by $s(p)$, and doing this for every $p \in Q$, we get a function $s : Q \to \mathbb{B}_a(\bar{x})$. But having $x = \Phi_p(x)$ is equivalent to having $x = r(\varphi(p,x)) = M(\varphi(p,x)) \cap \mathbb{B}_a(\bar{x})$. Hence s is the function in (9). Moreover, since $s(p) = r(\varphi(p, s(p)))$, we have from the Lipschitz continuity of r and (7) that, for any $p', p \in Q$,

$$\begin{aligned}
|s(p') - s(p)| &= |r(\varphi(p', s(p'))) - r(\varphi(p, s(p)))| \\
&\le |r(\varphi(p', s(p'))) - r(\varphi(p', s(p)))| + |r(\varphi(p', s(p))) - r(\varphi(p, s(p)))| \\
&\le \lambda |\varphi(p', s(p')) - \varphi(p', s(p))| + \lambda |\varphi(p', s(p)) - \varphi(p, s(p))| \\
&\le \lambda\nu |s(p') - s(p)| + \lambda |\varphi(p', s(p)) - \varphi(p, s(p))|.
\end{aligned}$$

Since $\lambda\nu < 1$, we see that s satisfies (10), as needed. $\square$

To show that the contraction mapping principle 1A.2 utilized in proving 2B.6 (namely, with $X = I\!R^n$ equipped with the metric induced by the Euclidean norm $|\cdot|$) can in turn be derived from Theorem 2B.6, choose $d = m = n$, $\nu = \lambda$, a unchanged, $b = |\Phi(\bar{x}) - \bar{x}|$, $\bar{p} = 0$, $Q = I\!B_b(0)$, $\varphi(p,x) = \Phi(x) + p$, $\bar{y} = \Phi(\bar{x})$, $M(y) = y + \bar{x} - \Phi(\bar{x})$, and consequently the λ in 1A.2 is 1. All the conditions of 2B.6 hold for such data under the assumptions of the contraction mapping principle 1A.2. Hence for $p = \Phi(\bar{x}) - \bar{x} \in Q$ the set $\{x \in I\!B_a(\bar{x}) \,|\, x = M(\varphi(p,x)) = \Phi(x)\}$ consists of exactly one point; that is, Φ has a unique fixed point in $I\!B_a(\bar{x})$. Thus, Theorem 2B.6 is actually equivalent[3] to the form of the contraction mapping principle 1.A.2 used in its proof.

Proof of Theorem 2B.5. For an arbitrary $\varepsilon > 0$, choose any $\lambda > \operatorname{lip}(\sigma;0)$ and $\nu > \mu$ such that $\lambda\nu < 1$ and

$$\frac{\lambda}{1-\lambda\nu} \le \frac{\kappa+\varepsilon}{1-\kappa\mu}, \tag{11}$$

as is possible under the assumption that $\kappa\mu < 1$. Let a, b and c be positive numbers such that

$$|\sigma(y) - \sigma(y')| \le \lambda|y - y'| \quad \text{for } y, y' \in I\!B_{\nu a+b}(0),$$

$$|e(p,x') - e(p,x)| \le \nu|x - x'| \quad \text{for } x, x' \in I\!B_a(\bar{x}) \text{ and } p \in I\!B_c(\bar{p}),$$

where $e(p,x) = f(p,x) - h(x)$, and

$$|f(p,\bar{x}) - f(\bar{p},\bar{x})| \le b \quad \text{for } p \in I\!B_c(\bar{p}). \tag{12}$$

Take b smaller if necessary so that $b\lambda < a(1 - \lambda\nu)$, and accordingly adjust c to ensure having (12). Now apply Theorem 2B.6 with $r = \sigma$, $M = (h+F)^{-1}$, $\bar{y} = 0$ and $\varphi = -e$, keeping the rest of the notation the same. It's straightforward to check that the estimates in (7) and the conditions (a) and (b) hold for the function in (8). Then, through the conclusion of Theorem 2B.6 and the observation that

$$x \in (h+F)^{-1}(-e(p,x)) \iff x \in S(p), \tag{13}$$

we obtain that the solution mapping S in (2) has a single-valued localization s around $\bar{p}$ for $\bar{x}$. Due to (11), the inequality in (6) holds for $Q = I\!B_c(\bar{p})$. That estimate implies the continuity of s at $\bar{p}$, in particular. $\square$

---

[3] Theorem 2B.6 can be stated in a complete metric space X and then it will be equivalent to the standard formulation of the contraction mapping principle in Theorem 1A.2. There is no point, of course, in giving a fairly complicated equivalent formulation of a classical result unless, as in our case, this formulation would bring some insights and dramatically simplify the proofs of later results. This is yet another confirmation of the common opinion shared also by the authors that the various reincarnations of the contraction mapping principle should be treated as tools for handling specific problems rather than isolated results.

From Theorem 2B.5 we obtain a generalization of Theorem 1E.13, the result in Chapter 1 about implicit functions without differentiability, in which the function f is replaced now by the sum $f+F$ for an *arbitrary* set-valued mapping F. The next statement, 2B.7, covers most of this generalization; the final part of 1E.13 (giving special consequences when $\mu = 0$) will be addressed in the follow-up statement, 2B.8.

Theorem 2B.7 (implicit function theorem for generalized equations)**.** *Consider a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ and a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with $(\bar{p},\bar{x}) \in \operatorname{int}\operatorname{dom} f$ and $f(\bar{p},\bar{x}) + F(\bar{x}) \ni 0$, and suppose that $\widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})) \leq \gamma < \infty$. Let h be a strict estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ. Suppose that $(h+F)^{-1}$ has a Lipschitz continuous single-valued localization σ around 0 for $\bar{x}$ with $\operatorname{lip}(\sigma;0) \leq \kappa$ for a constant κ such that $\kappa\mu < 1$. Then the solution mapping*

$$S : p \mapsto \left\{ x \in \mathbb{R}^n \,\middle|\, f(p,x) + F(x) \ni 0 \right\} \qquad \text{for } p \in \mathbb{R}^d$$

has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(s;\bar{p}) \leq \frac{\kappa\gamma}{1-\kappa\mu}.$$

The inverse function version of 2B.7 has the following simpler form:

Theorem 2B.8 (inverse function theorem for set-valued mappings)**.** *Consider a a mapping $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ with with $G(\bar{x}) \ni \bar{y}$ and suppose that G^{-1} has a Lipschitz continuous single-valued localization σ around $\bar{y}$ for $\bar{x}$ with $\operatorname{lip}(\sigma;\bar{y}) \leq \kappa$ for a constant κ. Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous around $\bar{x}$ with Lipschitz constant μ such that $\kappa\mu < 1$. Then the mapping $(g+G)^{-1}$ has a Lipschitz continuous single-valued localization around $\bar{y} + g(\bar{x})$ for $\bar{x}$ with Lipschitz constant $\kappa/(1-\kappa\mu)$.*

For the case of 2B.7 with $\mu = 0$, in which case h is a partial first-order approximation of f with respect to x at $(\bar{p},\bar{x})$, much more can be said about the single-valued localization s. The details are presented in the next result, which extends the part of 1E.13 for this case, and with it, Corollaries 2B.2 and 2B.3. We see that, by adding some relatively mild assumptions about the function f (while still allowing F to be arbitrary!), we can develop a first-order approximation of the localized solution mapping s in Theorem 2B.5. This opens the way to obtain differentiability properties of s, for example.

Theorem 2B.9 (extended implicit function theorem with first-order approximations)**.** *Specialize Theorem 2B.5 to the case where $\mu = 0$ in 2B.5(a), so that h is a strict first-order approximation of f with respect to x uniformly in p at $(\bar{p},\bar{x})$. Then, with the localization σ in 2B.5(b) we have the following additions to the conclusions of Theorem 2B.5:*

(a) *If $\operatorname{clm}_p(f;(\bar{p},\bar{x})) < \infty$ then the single-valued localization s of the solution mapping S in (2) is calm at $\bar{p}$ with*

$$\text{(14)} \qquad \operatorname{clm}(s;\bar{p}) \le \operatorname{lip}(\sigma;0)\cdot \operatorname{clm}_p(f;(\bar{p},\bar{x})).$$

(b) *If* $\widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})) < \infty$, *then the single-valued localization* s *of the solution mapping* S *in (2) is Lipschitz continuous near* $\bar{p}$ *with*

$$\text{(15)} \qquad \operatorname{lip}(s;\bar{p}) \le \operatorname{lip}(\sigma;0)\cdot \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})).$$

(c) *If, along with (a),* f *has a first-order approximation* r *with respect to* p *at* $(\bar{p},\bar{x})$, *then, for* Q *as in (6), the function* $\eta : Q \to \mathbb{R}^n$ *defined by*

$$\text{(16)} \qquad \eta(p) = \sigma(-r(p) + f(\bar{p},\bar{x})) \quad \text{for } p \in Q$$

is a first-order approximation at $\bar{p}$ *to the single-valued localization* s.

(d) *If, in addition to (b)(c),* σ *is affine, i.e.,* $\sigma(y) = \bar{x} + Ay$ *for some* $n \times m$ *matrix* A, *and furthermore the first-order approximation* r *is strict with respect to* p *uniformly in* x *at* $(\bar{p},\bar{x})$, *then* η *is a strict first-order approximation of* s *at* $\bar{p}$ *in the form*

$$\text{(17)} \qquad \eta(p) = \bar{x} + A(-r(p) + f(\bar{p},\bar{x})) \quad \text{for } p \in Q.$$

Proof. Let the constants a and c be as in the proof of Theorem 2B.5; then $Q = \mathbb{B}_c(\bar{p})$. Let $U = \mathbb{B}_a(\bar{x})$. For $p \in Q$, from (13) we have

$$\text{(18)} \quad s(p) = \sigma(-e(p,s(p))) \text{ for } e(p,x) = f(p,x) - h(x), \text{ and } \bar{x} = s(\bar{p}) = \sigma(0).$$

Let κ equal $\operatorname{lip}(\sigma;0)$ and consider for any $\varepsilon > 0$ the estimate in (6) with $\mu = 0$. Let $p' \in Q$, $p' \neq \bar{p}$ and $p = \bar{p}$ in (6) and divide both sides of (6) by $|p' - \bar{p}|$. Taking the limsup as $p' \to \bar{p}$ and $\varepsilon \to 0$ gives us (14).

Observe that (15) follows directly from 2B.7 under the assumptions of (b). Alternatively, it can be obtained by letting $p', p \in Q$, $p' \neq p$ in (6), dividing both sides of (6) by $|p' - p|$ and passing to the limit.

Consider now any $\lambda > \operatorname{clm}(s;\bar{p})$ and $\varepsilon > 0$. Make the neighborhoods Q and U smaller if necessary so that for all $p \in Q$ and $x \in U$ we have $|s(p) - s(\bar{p})| \le \lambda|p - \bar{p}|$ and

$$\text{(19)} \qquad |e(p,x) - e(p,\bar{x})| \le \varepsilon|x - \bar{x}|, \qquad |f(p,\bar{x}) - r(p)| \le \varepsilon|p - \bar{p}|,$$

and furthermore so that the points $-e(p,x)$ and $-r(p) + f(\bar{p},\bar{x})$ are contained in a neighborhood of 0 on which the function σ is Lipschitz continuous with Lipschitz constant $\kappa + \varepsilon = \operatorname{lip}(\sigma;0) + \varepsilon$. Then, for $p \in Q$, we get by way of (18), along with the first inequality in (19) and the fact that $e(\bar{p},\bar{x}) = 0$, the estimate that

$$\begin{aligned} |s(p) - \eta(p)| &= |s(p) - \sigma(-r(p) + f(\bar{p},\bar{x}))| \\ &= |\sigma(-e(p,s(p))) - \sigma(-r(p) + f(\bar{p},\bar{x}))| \\ &\le (\kappa+\varepsilon)(|-e(p,s(p)) + e(p,\bar{x})| + |f(p,\bar{x}) - r(p)|) \\ &\le (\kappa+\varepsilon)\varepsilon|s(p) - \bar{x}| + (\kappa+\varepsilon)\varepsilon|p - \bar{p}| \le \varepsilon(\kappa+\varepsilon)(\lambda+1)|p - \bar{p}|. \end{aligned}$$

Since ε can be arbitrarily small and also $s(\bar{p}) = \bar{x} = \sigma(0) = \eta(\bar{p})$, the function η defined in (16) is a first-order approximation of s at $\bar{p}$.

Moving on to part (d) of the theorem, suppose that the assumptions in (b)(c) are satisfied and also $\sigma(y) = \bar{x} + Ay$. Again, choose any $\varepsilon > 0$ and further adjust the neighborhoods Q of $\bar{p}$ and U of $\bar{x}$ so that

$$(20)\quad \begin{array}{ll} |e(p,x) - e(p,x')| \le \varepsilon|x - x'| & \text{for all } x,x' \in U \text{ and } p \in Q, \\ |f(p',x) - r(p') - f(p,x) + r(p)| \le \varepsilon|p' - p| & \text{for all } x \in U \text{ and } p',p \in Q, \end{array}$$

and moreover $s(p) \in U$ for $p \in Q$. By part (b), the single-valued localization s is Lipschitz continuous near $\bar{p}$; let $\lambda > \operatorname{lip}(s;\bar{p})$ and shrink Q even more if necessary so as to ensure that s is Lipschitz continuous with constant λ on Q. For $p, p' \in Q$, using (17), (18) and (20), we obtain

$$\begin{aligned} |s(p) - s(p') - \eta(p) + \eta(p')| &= |s(p) - s(p') - A(-r(p) + r(p'))| \\ &= |A(-e(p,s(p)) + e(p',s(p')) + r(p) - r(p'))| \\ &\le |A||-e(p,s(p)) + e(p,s(p'))| \\ &\quad + |A||f(p',s(p')) - r(p') - f(p,s(p')) + r(p)| \\ &\le |A|(\varepsilon|s(p) - s(p')| + \varepsilon|p' - p|) \\ &\le |A|\varepsilon(\lambda + 1)|p - p'|. \end{aligned}$$

Since ε can be arbitrarily small, we see that the first-order approximation of s furnished by η is strict, and the proof is complete. □

Note that the assumption in part (d), that the localization σ of $G^{-1} = (h+F)^{-1}$ around 0 for $\bar{x}$ is affine, can be interpreted as a sort of differentiability condition on G^{-1} at 0 with A giving the derivative mapping.

Corollary 2B.10 (utilization of strict differentiability). *Suppose in the generalized equation (1) with solution mapping S given by (2), that $\bar{x} \in S(\bar{p})$ and f is strictly differentiable at $(\bar{p},\bar{x})$. Assume that the inverse G^{-1} of the mapping*

$$G(x) = f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x - \bar{x}) + F(x), \qquad \textit{with } G(\bar{x}) \ni 0,$$

has a Lipschitz continuous single-valued localization σ around 0 for $\bar{x}$. Then not only do the conclusions of Theorem 2B.5 hold for a solution localization s, but also there is a first-order approximation η to s at $\bar{p}$ given by

$$\eta(p) = \sigma\big(-\nabla_p f(\bar{p},\bar{x})(p - \bar{p})\big).$$

Moreover, if $F \equiv 0$, then the first-order approximation η is strict and given by

$$(21)\qquad \eta(p) = \bar{x} - \nabla_x f(\bar{p},\bar{x})^{-1}\nabla_p f(\bar{p},\bar{x})(p - \bar{p}),$$

so that s is strictly differentiable at $\bar{p}$.

Proof. In this case Theorem 2B.9 is applicable with h taken to be the linearization of $f(\bar{p},\cdot)$ at $\bar{x}$ and r taken to be the linearization of $f(\cdot,\bar{x})$ at $\bar{p}$. When $F\equiv 0$, we get $\sigma(y)=\bar{x}+\nabla_x f(\bar{p},\bar{x})^{-1}y$, so that η as defined in (16) achieves the form in (21). Having a strict first-order approximation by an affine function means strict differentiability. □

The second part of Corollary 2B.10 shows how the implicit function theorem for equations as stated in Theorem 1D.13 is covered as a special case of Theorem 2B.7.

In the case of the generalized equation (1) where $f(p,x)=g(x)-p$ for a function $g:\mathbb{R}^n\to\mathbb{R}^m$ $(d=m)$, so that

$$S(p)=\left\{x\,\middle|\,p\in g(x)+F(x)\right\}=(g+F)^{-1}(p), \tag{22}$$

the inverse function version of Theorem 2B.9 has the following symmetric form.

Theorem 2B.11 (extended inverse function theorem with first-order approximations)**.** *In the framework of the solution mapping (22), consider any pair $(\bar{p},\bar{x})$ with $\bar{x}\in S(\bar{p})$. Let h be any strict first-order approximation to g at $\bar{x}$. Then $(g+F)^{-1}$ has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ if and only if $(h+F)^{-1}$ has such a localization σ around $\bar{p}$ for $\bar{x}$, in which case σ is a first-order approximation of s at $\bar{p}$ and*

$$\operatorname{lip}(s;\bar{p})=\operatorname{lip}(\sigma;\bar{p}). \tag{23}$$

If, in addition, σ is affine, $\sigma(y)=\bar{x}+Ay$, then s is strictly differentiable at $\bar{p}$ with $Ds(\bar{p})=A$.

Proof. For the "if" part, suppose that $(h+F)^{-1}$ has a localization σ as described. Then, from (15) with $f(p,x)=-p+g(x)$ we get $\operatorname{lip}(s;\bar{p})\le\operatorname{lip}(\sigma;0)$. The "only if" part is completely analogous because g and h play symmetric roles in the statement, and yields $\operatorname{lip}(\sigma;\bar{p})\le\operatorname{lip}(s';0)$ for some single-valued localization s' of S. The localizations s and s' have to agree graphically around $(0,\bar{x})$, so we pass to a smaller localization, again called s, and get the equality in (23). Through the observation that $r(p)=g(\bar{x})-p+\bar{p}$, the rest follows from Theorem 2B.9(d). □

We also can modify the results presented so far in this section in the direction indicated in Section 1F, where we considered local selections instead of single-valued localizations. We state such a result here as an exercise.

Exercise 2B.12 (implicit selections)**.** *Let $S(p)=\left\{x\in\mathbb{R}^n\,\middle|\,f(p,x)+F(x)\ni 0\right\}$ for a function $f:\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^m$ and a mapping $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^m$, along with a pair $(\bar{p},\bar{x})$ such that $\bar{x}\in S(\bar{p})$, and suppose that $\widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x}))\le\gamma<\infty$. Let h be a strict first-order approximation of f with respect to x at $(\bar{p},\bar{x})$ for which $(h+F)^{-1}$ has a Lipschitz continuous local selection σ around 0 for $\bar{x}$ with $\operatorname{lip}(\sigma;0)\le\kappa$. Then S has a Lipschitz continuous local selection s around $\bar{p}$ for $\bar{x}$ with*

$$\operatorname{lip}(s;\bar{p})\le\kappa\gamma.$$

If in addition f has a first-order approximation r with respect to p at $(\bar p,\bar x)$, then there exists a neighborhood Q of $\bar p$ such that the function

$$\eta : p \mapsto \sigma(-r(p)+f(\bar p,\bar x)) \quad \text{for } p \in Q$$

is a first-order approximation of s at $\bar p$.

Guide. First verify the following statement, which is a simple modification of Theorem 2B.6. For a function $\varphi : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and a point $(\bar p,\bar x) \in \operatorname{int} \operatorname{dom} \varphi$, let the nonnegative scalars ν, b, the positive scalar a, and the set $Q \subset \mathbb{R}^d$ be such that $\bar p \in Q$ and the conditions (7) hold. Consider also a set-valued mapping $M : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ with $(\bar y,\bar x) \in \operatorname{gph} M$, where $\bar y := \varphi(\bar p,\bar x)$, and assume that there exists a Lipschitz continuous function r on $\mathbb{B}_{\nu a+b}(\bar y)$ such that

$$r(y) \in M(y) \cap \mathbb{B}_a(\bar x) \text{ for } y \in \mathbb{B}_{\nu a+b}(\bar y) \quad \text{and} \quad r(\bar y) = \bar x.$$

In addition, suppose now that the Lipschitz constant λ for the function r is such that the conditions (a) and (b) in the statement of Theorem 2B.6 are fulfilled. Then for each $p \in Q$ the set $\left\{ x \in \mathbb{B}_a(\bar x) \,\middle|\, x \in M(\varphi(p,x)) \right\}$ contains a point $s(p)$ such that the function $p \mapsto s(p)$ satisfies $s(\bar p) = \bar x$ and

$$(24) \qquad |s(p') - s(p)| \le \frac{\lambda}{1-\lambda\nu} |\varphi(p',s(p)) - \varphi(p,s(p))| \text{ for all } p', p \in Q.$$

Thus, the mapping $N := p \mapsto \left\{ x \,\middle|\, x \in M(\varphi(p,x)) \right\} \cap \mathbb{B}_a(\bar x)$ has a local selection s around $\bar p$ for $\bar x$ which satisfies (24). The difference from Theorem 2B.6 is that r is now required only to be a local selection of the mapping M with specified properties, and then we obtain a local selection s of N at $\bar p$ for $\bar x$. For the rest use the proofs of Theorems 2B.5 and 2B.9. □

Exercise 2B.13 (inverting perturbed inverse)**.** *Consider functions f and g from $\mathbb{R}^n$ into itself, a point $\bar x \in \operatorname{int} \operatorname{dom} f$ such that $f(\bar x) \in \operatorname{int} \operatorname{dom} g$, and positive numbers κ and μ such that $\kappa\mu < 1$. Let f be Lipschitz continuous around $\bar x$ with Lipschitz constant κ and g be Lipschitz continuous around $f(\bar x)$ with Lipschitz constant μ. Then $(f^{-1}+g)^{-1}$ has a Lipschitz continuous single-valued localization around $\bar x + g(f(\bar x))$ for $f(\bar x)$ with Lipschitz constant $\kappa/(1-\kappa\mu)$.*

Guide. Apply 2B.9 with $G^{-1} = f$. □

2C. Ample Parameterization and Parametric Robustness

The results in 2B, especially the broad generalization of Robinson's theorem in 2B.5 and its complement in 2B.9 dealing with solution approximations, provide a substantial extension of the classical theory of implicit functions. Equations have been replaced by generalized equations, with variational inequalities as a particular case, and technical assumptions about differentiability have been greatly relaxed. Much of the rest of this chapter will be concerned with working out the consequences in situations where additional structure is available. Here, however, we reflect on the ways that parameters enter the picture and the issue of whether there are "enough" parameters, which emerges as essential in drawing good conclusions about solution mappings.

The differences in parameterization between an inverse function theorem and an implicit function theorem are part of a larger pattern which deserves, at this stage, a closer look. Let's start by considering a generalized equation without parameters,

$$(1) \qquad g(x)+F(x) \ni 0,$$

for a function $g:\mathbb{R}^n \to \mathbb{R}^m$ and a set-valued mapping $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$. We can think of a *parameterization* as the choice of a function

$$(2) \qquad f:\mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m \text{ having } f(\bar{p},x) \equiv g(x) \text{ for a particular } \bar{p} \in \mathbb{R}^d.$$

The specification of such a parameterization leads to an associated solution mapping

$$(3) \qquad S: p \mapsto \left\{ x \,\middle|\, f(p,x)+F(x) \ni 0 \right\},$$

which we proceed to study around $\bar{p}$ and a point $\bar{x} \in S(\bar{p})$ for the presence of a nice localization σ. Different parameterizations yield different solution mappings, which may possess different properties according to the assumptions placed on f.

That's the general framework, but the special kind of parameterization that corresponds to the "inverse function" case has a fundamental role which is worth trying to understand more fully. In that case, we simply have $f(p,x)=g(x)-p$ in (2), so that in (1) we are solving $g(x)+F(x) \ni p$ and the solution mapping is $S=(g+F)^{-1}$. Interestingly, this kind of parameterization comes up even in obtaining "implicit function" results through the way that approximations are utilized. Recall that in Theorem 2B.5, for a function h which is "close" to $f(\bar{p},\cdot)$ near $\bar{x}$, the mapping $(h+F)^{-1}$ having $\bar{x} \in (h+F)^{-1}(0)$ is required to have a Lipschitz continuous single-valued localization around 0 for $\bar{x}$. Only then are we able to deduce that the solution mapping S in (3) has a localization of such type at $\bar{p}$ for $\bar{x}$. In other words, the desired conclusion about S is obtained from an assumption about a simpler solution mapping in the "inverse function" category.

When S itself already belongs to that category, because $f(p,x)=g(x)-p$ and $S=(g+F)^{-1}$, another feature of the situation emerges. Then, as seen in 2B.11, the assumption made about $(h+F)^{-1}$ is not only sufficient for obtaining the desired localization of S but also necessary. This distinction was already observed in the classical setting. In the "symmetric" version of the inverse function theorem in 1B.9, the invertibility of a linearized mapping is both necessary and sufficient for the conclusion, whereas such invertibility acts only as a sufficient condition in the implicit function theorem 1B.2. On an additional assumption on the rank of the Jacobian with respect to the parameter however, as noted in 1B.8, this sufficient condition becomes also necessary. In other words, to achieve necessity as well as sufficiency, the parameterization must be "rich" enough.

Ample parameterization. *A parameterization of the generalized equation (1) as in (2) will be called* ample *at $\bar{x}$ if $f(p,x)$ is strictly differentiable with respect to p uniformly in x at $(\bar{p},\bar{x})$ and the partial Jacobian $\nabla_p f(\bar{p},\bar{x})$ is of full rank:*

$$\text{rank}\ \nabla_p f(\bar{p},\bar{x}) = m. \tag{4}$$

The reason why the rank condition in (4) can be interpreted as ensuring the richness of the parameterization is that it can always be achieved through supplementary parameters. Any parameterization function f having the specified strict differentiability can be extended to a parameterization function $\tilde{f}$ with

$$\tilde{f}(q,x) = f(p,x) - y, \qquad q = (p,y),\ \bar{q} = (\bar{p},0), \tag{5}$$

which does satisfy the ampleness condition, since trivially $\text{rank}\ \nabla_q \tilde{f}(\bar{q},\bar{x}) = m$. The generalized equation being solved then has solution mapping

$$\widetilde{S} : (p,y) \mapsto \left\{ x \,\middle|\, f(p,x) + F(x) \ni y \right\}. \tag{6}$$

Results about localizations of $\widetilde{S}$ can be specialized to results about S by taking $y = 0$.

In order to arrive at the key result about ample parameterization, asserting an equivalence about the existence of several kinds of localizations, we need a lemma about local selections which is related to the results presented in Section 1F.

Lemma 2C.1 local selection from ampleness)**.** *Let $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ with $(\bar{p},\bar{x}) \in \text{int dom}\, f$ afford an ample parameterization of the generalized equation (1) at $\bar{x}$. Suppose that f has a strict first-order approximation $h : \mathbb{R}^n \to \mathbb{R}^m$ with respect to x uniformly in p at $(\bar{p},\bar{x})$. Then the mapping*

$$\Psi : (x,y) \mapsto \left\{ p \,\middle|\, e(p,x) + y = 0 \right\} \quad \textit{for} \ \ (x,y) \in \mathbb{R}^n \times \mathbb{R}^m, \tag{7}$$

where $e(p,x) = f(p,x) - h(x)$, has a local selection ψ around $(\bar{x},0)$ for $\bar{p}$ which satisfies

$$\widehat{\text{lip}}_x(\psi;(\bar{x},0)) = 0 \tag{8a}$$

and

$$\widehat{\text{lip}}_y(\psi;(\bar{x},0)) < \infty. \tag{8b}$$

Proof. Let $A = \nabla_p f(\bar{p},\bar{x})$; then AA^{T} is invertible. Without loss of generality, suppose $\bar{x} = 0$, $\bar{p} = 0$, and $f(0,0) = 0$; then $h(0) = 0$. Let $c = |A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}|$. Let $0 < \varepsilon < 1/(2c)$ and choose a positive a such that for all $x,x' \in a\mathbb{B}$ and $p,p' \in a\mathbb{B}$ we have

$$(9) \qquad |e(p,x') - e(p,x)| \leq \varepsilon |x - x'|$$

and

$$(10) \qquad |f(p,x) - f(p',x) - A(p-p')| \leq \varepsilon |p - p'|.$$

For $b = a(1-2c\varepsilon)/c$, fix $x \in a\mathbb{B}$ and $y \in b\mathbb{B}$, and consider the mapping

$$\Phi_{x,y} : p \mapsto -A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}(e(p,x) + y - Ap) \quad \text{for } p \in a\mathbb{B}.$$

Through (9) and (10), keeping in mind that $e(0,0) = 0$, we see that

$$|\Phi_{x,y}(0)| \leq c|e(0,x) + y| \leq c|e(0,x) - e(0,0)| + c|y| \leq c\varepsilon a + cb = a(1 - c\varepsilon),$$

and for every $p, p' \in a\mathbb{B}$

$$|\Phi_{x,y}(p) - \Phi_{x,y}(p')| \leq c|f(p,x) - f(p',x) - A(p-p')| \leq c\varepsilon|p - p'|.$$

The contraction mapping principle 1A.2 then applies, and we obtain from it the existence of a unique $p \in a\mathbb{B}$ such that

$$(11) \qquad p = -A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}(e(p,x) + y - Ap).$$

We denote by $\psi(x,y)$ the unique solution in $a\mathbb{B}$ of this equation for $x \in a\mathbb{B}$ and $y \in b\mathbb{B}$. Multiplying both sides of (11) by A and simplifying, we get $e(p,x) + y = 0$. This means that for each $(x,y) \in a\mathbb{B} \times b\mathbb{B}$ the equation $e(p,x) + y = 0$ has $\psi(x,y)$ as a solution. From (11), we know that

$$(12) \qquad \psi(x,y) = -A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}(f(\psi(x,y),x) - h(x) + y - A\psi(x,y)).$$

Let $x, x' \in a\mathbb{B}$ and $y, y' \in b\mathbb{B}$. Using (9) and (10) we have

$$\begin{aligned} |\psi(x,y) - \psi(x',y)| &\leq c|e(\psi(x,y),x) - e(\psi(x,y),x')| \\ &\quad + c|f(\psi(x,y),x') - f(\psi(x',y),x') - A(\psi(x,y) - \psi(x',y))| \\ &\leq c\varepsilon|x - x'| + c\varepsilon|\psi(x,y) - \psi(x',y)|. \end{aligned}$$

Hence

$$|\psi(x,y) - \psi(x',y)| \leq \frac{c\varepsilon}{1 - c\varepsilon}|x - x'|.$$

Since ε can be arbitrarily small, we conclude that (8a) holds. Analogously, from (12) and using again (9) and (10) we obtain

$$\begin{aligned} &|\psi(x,y) - \psi(x,y')| \\ &\quad \leq c|f(\psi(x,y),x) + y - A\psi(x,y) - f(\psi(x,y'),x) - y' + A\psi(x,y')| \\ &\quad \leq c|y - y'| + c\varepsilon|\psi(x,y) - \psi(x,y')|, \end{aligned}$$

and then

$$|\psi(x,y)-\psi(x,y')| \leq \frac{c}{1-c\varepsilon}|y-y'|$$

which gives us (8b). □

We are now ready to present the first main result of this section:

Theorem 2C.2 (equivalences from ampleness)**.** *Let f parameterize the generalized equation (1) as in (2). Suppose the parameterization is ample at $(\bar{p},\bar{x})$, and let h be a strict first-order approximation of f with respect to x uniformly in p at $(\bar{p},\bar{x})$. Then the following properties are equivalent:*

(a) *S in (3) has a Lipschitz continuous single-valued localization around $\bar{p}$ for $\bar{x}$;*

(b) *$(h+F)^{-1}$ has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$;*

(c) *$(g+F)^{-1}$ has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$;*

(d) *$\widetilde{S}$ in (6) has a Lipschitz continuous single-valued localization around $(\bar{p},0)$ for $\bar{x}$.*

Proof. If the mapping $(h+F)^{-1}$ has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$, then from Theorem 2B.5 together with Theorem 2B.11 we may conclude that the other three mappings likewise have such localizations at the respective reference points. In other words, (b) is sufficient for (a) and (d). Also, (b) is equivalent to (c) inasmuch as g and h are first-order approximations to each other (Theorem 2B.11). Since (d) implies (a), the issue is whether (b) is necessary for (a).

Assume that (a) holds with a Lipschitz localization s around $\bar{p}$ for $\bar{x}$ and choose $\lambda > \operatorname{lip}(s;\bar{p})$. Let $\nu > 0$ be such that $\lambda\nu < 1$, and consider a Lipschitz continuous local selection ψ of the mapping Ψ in Lemma 2C.1. Then there exist positive a, b and c such that $\lambda\nu a + \lambda b < a$,

$$S(p)\cap \mathbb{B}_a(\bar{x}) = s(p) \quad \text{for } p\in \mathbb{B}_{\nu a+b}(\bar{p}),$$

$$|s(p)-s(p')| \leq \lambda|p-p'| \text{ for } p,p'\in \mathbb{B}_{\nu a+b}(\bar{p}),$$

$$h(x)-f(\psi(x,y),x) = y \text{ for } y\in \mathbb{B}_c(0), x\in \mathbb{B}_a(\bar{x}),$$

$$|\psi(x,y)-\psi(x',y)| \leq \nu|x-x'| \quad \text{for } x,x'\in \mathbb{B}_a(\bar{x}) \text{ and } y\in \mathbb{B}_c(0),$$

the last from (8a), and

$$|\psi(y,\bar{x})-\psi(0,\bar{x})| \leq b \text{ for } y\in \mathbb{B}_c(0).$$

We now apply Theorem 2B.6 with $\varphi(p,x) = \psi(x,y)$ for $p = y$ and $M(p) = S(p)$, thereby obtaining that the mapping

$$\mathbb{B}_c(0) \ni y \mapsto \left\{ x\in \mathbb{B}_a(\bar{x}) \,\middle|\, x\in S(\psi(x,y)) \right\}$$

is a function which is Lipschitz continuous on $\mathbb{B}_c(0)$. Noting that

$$(h+F)^{-1}(y)\cap \mathbb{B}_a(\bar{x}) = \{x \mid x = S(\psi(x,y))\cap \mathbb{B}_a(\bar{x})\},$$

we conclude that $(h+F)^{-1}$ has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$. Thus, (a) implies (b). □

The strict differentiability property with respect to p which is assumed in the definition of ample parameterization is satisfied of course when f is strictly differentiable with respect to (p,x) at $(\bar{p},\bar{x})$. Then, moreover, the linearization of $f(\bar{p},\cdot)$ at $\bar{x}$, which is the same as the linearization of g at $\bar{x}$, can be taken as the function h. This leads to a statement about an entire class of parameterizations.

Theorem 2C.3 (parametric robustness). *Consider the generalized equation (1) under the assumption that $\bar{x}$ is a point where g is strictly differentiable. Let $h(x) = g(\bar{x}) + \nabla g(\bar{x})(x-\bar{x})$. Then the following statements are equivalent:*

(a) $(h+F)^{-1}$ *has a Lipschitz continuous single-valued localization around* 0 *for* $\bar{x}$*;*

(b) *For every parameterization (2) in which f is strictly differentiable at $(\bar{x},\bar{p})$, the mapping S in (3) has a Lipschitz continuous single-valued localization around $\bar{p}$ for $\bar{x}$.*

Proof. The implication from (a) to (b) already follows from Theorem 2B.9. The focus is on the reverse implication. This is valid because, among the parameterizations covered by (b), there will be some that are ample. For instance, one could pass from a given one to an ample parameterization in the mode of (5). For the solution mapping for such a parameterization, we have the implication from (a) to (b) in Theorem 2C.2. That specializes to what we want. □

2D. Semidifferentiable Functions

The notion of a first-order approximation of a function at a given point has already served us for various purposes as a substitute for differentiability, where the approximation is a linearization. We now bring in an intermediate concept in which linearity is replaced by positive homogeneity.

A function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ is *positively homogeneous* if $0 = \varphi(0)$ and $\varphi(\lambda w) = \lambda \varphi(w)$ for all $w \in \operatorname{dom} \varphi$ and $\lambda > 0$. These conditions mean geometrically that the graph of φ is a cone in $\mathbb{R}^n \times \mathbb{R}^m$. A linear function is positively homogeneous in particular, of course. The graph of a linear function $\varphi : \mathbb{R}^n \to \mathbb{R}^m$ is actually a subspace of $\mathbb{R}^n \times \mathbb{R}^m$.

Semiderivatives. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be semidifferentiable*[4] *at $\bar{x}$ if it has a first-order approximation at $\bar{x}$ of the form $h(x) = f(\bar{x}) + \varphi(x-\bar{x})$ with φ continuous and positively homogeneous; when the approximation is strict, f is*

[4] Also called Bouligand differentiable or B-differentiable functions.

strictly semidifferentiable at $\bar{x}$. Either way, the function φ, necessarily unique, is called the semiderivative of f at $\bar{x}$ and denoted by $Df(\bar{x})$, so that $h(x) = f(\bar{x}) + Df(\bar{x})(x-\bar{x})$.

In a first-order approximation we have by definition that $\operatorname{clm}(f-h)(\bar{x}) = 0$, which in the "strict" case is replaced by $\operatorname{lip}(f-h)(\bar{x}) = 0$. The uniqueness of the semiderivative, when it exists, comes from the fact that any two first-order approximations $f(\bar{x})+\varphi(x-\bar{x})$ and $f(\bar{x})+\psi(x-\bar{x})$ of f at $\bar{x}$ must have $\operatorname{clm}(\varphi-\psi)(0) = 0$, and under positive homogeneity that cannot hold without having $\varphi = \psi$. The uniqueness can also be gleaned through comparison with directional derivatives.

One-sided directional derivatives. *For $f : \mathbb{R}^n \to \mathbb{R}^m$, a point $\bar{x} \in \operatorname{dom} f$ and a vector $w \in \mathbb{R}^n$, the limit*

$$f'(\bar{x};w) = \lim_{t \searrow 0} \frac{f(\bar{x}+tw)-f(\bar{x})}{t}, \tag{1}$$

when it exists, is the (one-sided) directional derivative of f at $\bar{x}$ for w; here $t\searrow 0$ means that $t \to 0$ with $t > 0$. If this directional derivative exists for every w, f is said to be directionally differentiable at $\bar{x}$.

Note that $f'(\bar{x};w)$ is positively homogeneous in the w argument. This comes out of the limit definition itself. Directional differentiability is weaker than semidifferentiability in general, but equivalent to it in the presence of Lipschitz continuity, as we demonstrate next.

Proposition 2D.1 (directional differentiability and semidifferentiability)**.** *If a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is semidifferentiable at $\bar{x}$, then f is in particular directionally differentiable at $\bar{x}$ and has*

$$Df(\bar{x})(w) = f'(\bar{x};w) \text{ for all } w, \tag{2}$$

so that the first-order approximation in the definition of semidifferentiability has the form

$$h(x) = f(\bar{x}) + f'(\bar{x};x-\bar{x}).$$

When $\operatorname{lip}(f;\bar{x}) < \infty$, directional differentiability at $\bar{x}$ in turn implies semidifferentiability at $\bar{x}$.

Proof. Having $\operatorname{clm}(f-h)(\bar{x}) = 0$ for $h(x) = f(\bar{x}) + \varphi(x-\bar{x})$ as in the definition of semidifferentiability entails having $[f(\bar{x}+tw)-h(\bar{x}+tw)]/t \to 0$ as $t\searrow 0$ with $h(\bar{x}+tw) = f(\bar{x}) + t\varphi(w)$. Thus, $\varphi(w)$ must be $f'(\bar{x};w)$.

For the converse claim, consider $\lambda > \operatorname{lip}(f;\bar{x})$ and observe that for any u and v,

$$|f'(\bar{x};u) - f'(\bar{x};v)| = \lim_{t \searrow 0} \frac{1}{t} |f(\bar{x}+tu) - f(\bar{x}+tv)| \le \lambda |u-v|. \tag{3}$$

Next, consider an arbitrary sequence $u_k \to 0$ and, without loss of generality, assume that $u_k/|u_k| \to \bar{u}$ with $|\bar{u}| = 1$. Letting $t_k = |u_k|$ and using the positive homogeneity of the directional derivative, we obtain

$$\begin{aligned}
0 &\leq \frac{1}{|u_k|}|f(\bar{x}+u_k) - f(\bar{x}) - f'(\bar{x};u_k)| \\
&\leq \frac{1}{t_k}\Big(|f(\bar{x}+u_k) - f(\bar{x}+t_k\bar{u})| + |f'(\bar{x};t_k\bar{u}) - f'(\bar{x};u_k)| \\
&\quad + |f(\bar{x}+t_k\bar{u}) - f(\bar{x}) - f'(\bar{x};t_k\bar{u})|\Big) \\
&\leq 2\lambda\left|\frac{u_k}{t_k} - \bar{u}\right| + \left|\frac{1}{t_k}(f(\bar{x}+t_k\bar{u}) - f(\bar{x})) - f'(\bar{x};\bar{u})\right|,
\end{aligned}$$

where in the final inequality we invoke (3). Since u_k is arbitrarily chosen, we conclude by passing to the limit as $k \to \infty$ that for $h(x) = f(\bar{x}) + f'(\bar{x};x-\bar{x})$ we do have $\operatorname{clm}(f-h;\bar{x}) = 0$. □

When the semiderivative $Df(\bar{x}) : I\!\!R^n \to I\!\!R^m$ is linear, semidifferentiability turns into differentiability, and strict semidifferentiability turns into strict differentiability. The connections known between $Df(\bar{x})$ and the calmness modulus and Lipschitz modulus of f at $\bar{x}$ under differentiability can be extended to semidifferentiability by adopting the definition that

$$|\varphi| = \sup_{|x|\leq 1} |\varphi(x)| \quad \text{for a positively homogeneous function } \varphi.$$

We then have $\operatorname{clm}(Df(\bar{x});0) = |Df(\bar{x})|$ and consequently $\operatorname{clm}(f;\bar{x}) = |Df(\bar{x})|$, which in the case of strict semidifferentiability becomes $\operatorname{lip}(f;\bar{x}) = |Df(\bar{x})|$. Thus in particular, semidifferentiability of f at $\bar{x}$ implies that $\operatorname{clm}(f;\bar{x}) < \infty$, while strict semidifferentiability at $\bar{x}$ implies that $\operatorname{lip}(f;\bar{x}) < \infty$.

Exercise 2D.2 (alternative characterization of semidifferentiability). *For a function $f : I\!\!R^n \to I\!\!R^m$ and a point $\bar{x} \in \operatorname{dom} f$, semidifferentiability is equivalent to the existence for every $w \in I\!\!R^n$ of*

$$\lim_{\substack{t \searrow 0 \\ w' \to w}} \frac{f(\bar{x}+tw') - f(\bar{x})}{t}. \tag{4}$$

Guide. Directional differentiability of f at $\bar{x}$ corresponds to the difference quotient functions $\Delta_t(w) = [f(\bar{x}+tw) - f(\bar{x})]/t$ converging pointwise to something, namely $f'(\bar{x};\cdot)$, as $t \searrow 0$. Show that the existence of the limits in (4) means that these functions converge to $f'(\bar{x};\cdot)$ not just pointwise, but uniformly on bounded sets. Glean from that the equivalence with having a first-order approximation as in the definition of semidifferentiability. □

Examples.

1) The function $f(x) = e^{|x|}$ for $x \in I\!\!R$ is not differentiable at 0, but it is semidifferentiable there and its semiderivative is given by $Df(0) : w \mapsto |w|$. This is actually

a case of strict semidifferentiability. Away from 0, f is of course continuously differentiable (hence strictly differentiable).

2) The function $f(x_1,x_2)=\min\{x_1,x_2\}$ on $\mathbb{R}^2$ is continuously differentiable at every point away from the line where $x_1=x_2$. On that line, f is strictly semidifferentiable with

$$Df(x_1,x_2)(w_1,w_2)=\min\{w_1,w_2\}.$$

3) A function of the form $f(x)=\max\{f_1(x),f_2(x)\}$, with f_1 and f_2 continuously differentiable from $\mathbb{R}^n$ to $\mathbb{R}$, is strictly differentiable at all points x where $f_1(x)\neq f_2(x)$ and semidifferentiable where $f_1(x)=f_2(x)$, the semiderivative being given there by

$$Df(x)(w)=\max\{Df_1(x)(w),Df_2(x)(w)\}.$$

However, f might not be strictly semidifferentiable at such points; see Example 2D.5.

The semiderivative obeys standard calculus rules, such as semidifferentiation of a sum, product and ratio, and, most importantly, the chain rule. We pose the verification of these rules as exercises.

Exercise 2D.3. *Let f be semidifferentiable at $\bar{x}$ and let g be Lipschitz continuous and semidifferentiable at $\bar{y}:=f(\bar{x})$. Then $g\circ f$ is semidifferentiable at $\bar{x}$ and*

$$D(g\circ f)(\bar{x})=Dg(\bar{y})\circ Df(\bar{x}).$$

Guide. Apply Proposition 1E.1 and observe that a composition of positively homogeneous functions is positively homogeneous. □

Exercise 2D.4. *Let f be strictly semidifferentiable at $\bar{x}$ and g be strictly differentiable at $f(\bar{x})$. Then $g\circ f$ is strictly semidifferentiable at $\bar{x}$.*

Guide. Apply 1E.2. □

Example 2D.5. The functions f and g in Exercise 2D.4 cannot exchange places: the composition of a strictly semidifferentiable function with a strictly differentiable function is not always strictly semidifferentiable. For a counterexample, consider the function $f:\mathbb{R}^2\to\mathbb{R}$ given by

$$f(x_1,x_2)=\min\{x_1^3,x_2\} \quad \text{for } (x_1,x_2)\in\mathbb{R}^2.$$

According to 2D.3, the function f is semidifferentiable at $(0,0)$ with semiderivative $Df(0,0)(w_1,w_2)=\min\{0,w_2\}$. To see that f is not strictly semidifferentiable at $(0,0)$, however, observe for the function $g=f-Df(0,0)$ that

$$\frac{|g(x_1',x_2')-g(x_1,x_2)|}{|(x_1',x_2')-(x_1,x_2)|}=\frac{1}{1+2\varepsilon} \quad \text{for } (x_1,x_2)=(-\varepsilon,-\varepsilon^3/2) \text{ and } (x_1',x_2')=(-\varepsilon,\varepsilon^4).$$

As ε goes to 0 this ratio tends to 1, and therefore $\operatorname{lip}(f - Df(0,0);(0,0)) \geq 1$.

Our aim now is to forge out of Theorem 2B.9 a result featuring semiderivatives. For this purpose, we note that if $f(p,x)$ is (strictly) semidifferentiable at $(\bar{p},\bar{x})$ jointly in its two arguments, it is also "partially (strictly) semidifferentiable" in these arguments separately. In denoting the semiderivative of $f(\bar{p},\cdot)$ at $\bar{x}$ by $D_x f(\bar{p},\bar{x})$ and the semiderivative of $f(\cdot,\bar{x})$ at $\bar{p}$ by $D_p f(\bar{p},\bar{x})$, we have

$$D_x f(\bar{p},\bar{x})(w) = Df(\bar{p},\bar{x})(0,w), \qquad D_p f(\bar{p},\bar{x})(q) = Df(\bar{p},\bar{x})(q,0).$$

In contrast to the situation for differentiability, however, $Df(\bar{p},\bar{x})(q,w)$ isn't necessarily the sum of these two partial semiderivatives.

Theorem 2D.6 (implicit function theorem utilizing semiderivatives). *Let $\bar{x} \in S(\bar{p})$ for the solution mapping*

$$S : p \mapsto \left\{ x \in \mathbb{R}^n \,\middle|\, f(p,x) + F(x) \ni 0 \right\}$$

associated with a choice of $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ such that f is strictly semidifferentiable at $(\bar{p},\bar{x})$. Suppose that the inverse G^{-1} of the mapping

$$G(x) \,=\, f(\bar{p},\bar{x}) + D_x f(\bar{p},\bar{x})(x - \bar{x}) + F(x), \quad \textit{with} \;\; G(\bar{x}) \ni 0,$$

has a Lipschitz continuous single-valued localization σ around 0 for $\bar{x}$ which is semidifferentiable at 0. Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ which is semidifferentiable at $\bar{p}$ with its semiderivative given by

$$Ds(\bar{p}) = D\sigma(0)\circ(-D_p f(\bar{p},\bar{x})).$$

Proof. First, note that $s(\bar{p}) = \sigma(0) = \bar{x}$ and the function r in Theorem 2B.9 may be chosen as $r(p) = f(\bar{x},\bar{p}) + D_p f(\bar{p},\bar{x})(p - \bar{p})$. Then we have

$$\begin{aligned} |s(p) - s(\bar{p}) - (D\sigma(0)\circ(-D_p f(\bar{p},\bar{x})))(p - \bar{p})| &\leq |s(p) - \sigma(-r(p) + r(\bar{p}))| \\ &+ |\sigma(-D_p f(\bar{p},\bar{x})(p - \bar{p})) - \sigma(0) - D\sigma(0)(-D_p f(\bar{p},\bar{x})(p - \bar{p}))|. \end{aligned}$$

According to Theorem 2B.9 the function $p \mapsto \sigma(-r(p) + r(\bar{p}))$ is a first-order approximation to s at $\bar{p}$, hence the first term on the right side of this inequality is of order $o(|p - \bar{p}|)$ when p is close to $\bar{p}$. The same is valid for the second term, since σ is assumed to be semidifferentiable at 0. It remains to observe that the composition of positively homogeneous mappings is positively homogeneous. $\square$

An important class of semidifferentiable functions will be brought in next.

Piecewise smooth functions. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be piecewise smooth on an open set $O \subset \operatorname{dom} f$ if it is continuous on O and for each $x \in O$ there is a finite collection $\{f_i\}_{i \in I}$ of smooth ($\mathscr{C}^1$) functions defined on a neighborhood of x such that, for some $\varepsilon > 0$, one has*

$$(5) \qquad f(y) \in \{ f_i(y) \,|\, i \in I \} \quad \textit{when} \quad |y-x| < \varepsilon.$$

The collection $\{f_i\}_{i\in I(x)}$, *where* $I(x) = \{ i \in I \,|\, f(x) = f_i(x) \}$, *is said then to furnish a local representation of* f *at* x. *A local representation in this sense is minimal if no proper subcollection of it forms a local representation of* f *at* x.

Note that a local representation of f at x characterizes f on a neighborhood of x, and minimality means that this would be lost if any of the functions f_i were dropped.

The piecewise smoothness terminology finds its justification in the following observation.

Proposition 2D.7 (decomposition of piecewise smooth functions)**.** *Let* f *be piecewise smooth on an open set* O *with a minimal local representation* $\{f_i\}_{i\in I}$ *at a point* $\bar{x} \in O$. *Then for each* $i \in I(\bar{x})$ *there is an open set* O_i *such that* $\bar{x} \in \mathrm{cl}\, O_i$ *and* $f(x) = f_i(x)$ *on* O_i.

Proof. Let $\varepsilon > 0$ be as in (5) with $x = \bar{x}$ and assume that $\mathbb{B}_\varepsilon(\bar{x}) \subset O$. For each $i \in I(\bar{x})$, let $U_i = \{ x \in \mathrm{int}\, \mathbb{B}_\varepsilon(\bar{x}) \,|\, f(x) = f_i(x) \}$ and $O_i = \mathrm{int}\, \mathbb{B}_\varepsilon(\bar{x}) \setminus \cup_{j\neq i} U_j$. Because f and f_i are continuous, U_i is closed relative to $\mathrm{int}\, \mathbb{B}_\varepsilon(\bar{x})$ and therefore O_i is open. Furthermore $\bar{x} \in \mathrm{cl}\, O_i$, for if not, the set $\cup_{j\neq i} U_j$ would cover a neighborhood of $\bar{x}$, and then f_i would be superfluous in the local representation, thus contradicting minimality. □

It's not hard to see from this fact that a piecewise smooth function on an open set O must be continuous on O and even locally Lipschitz continuous, since each of the $\mathscr{C}^1$ functions f_i in a local representation is locally Lipschitz continuous, in particular. Semidifferentiability in this situation takes only a little more effort to confirm.

Proposition 2D.8 (semidifferentiability of piecewise smooth functions)**.** *If a function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *is piecewise smooth on an open set* $O \subset \mathrm{dom}\, f$, *then* f *is semidifferentiable on* O. *Furthermore, the semiderivative function* $Df(\bar{x})$ *at any point* $\bar{x} \in O$ *is itself piecewise smooth, in fact with local representation composed by the linear functions* $\{Df_i(\bar{x})\}_{i\in I(\bar{x})}$ *when* f *has a minimal local representation* $\{f_i\}_{i\in I}$ *around* $\bar{x}$.

Proof. Let $\{f_i\}_{i\in I(\bar{x})}$ be a minimal local representation of f around $\bar{x}$. For a suitably small $\delta > 0$ and any $w \in \mathbb{R}^n$, let $\varphi(t) = f(\bar{x}+tw)$ and $\varphi_i(t) = f_i(\bar{x}+tw)$ for $t \in (-\delta, \delta)$. Then φ is piecewise smooth with $\varphi(0) = \varphi_i(0)$ for $i \in I(\bar{x})$ and $\varphi(t) \in \{\varphi_i(t) \,|\, i \in I(\bar{x})\}$ for every $t \in (-\delta, \delta)$. Since φ_i are smooth, taking δ smaller if necessary we have that $\varphi_i(t) \neq \varphi_j(t)$ for all $t \in (0, \delta)$ whenever $\varphi_i'(0) \neq \varphi_j'(0)$. Thus, there must exist a nonempty index set $I \subset I(\bar{x})$ such that $\varphi_i'(0) = \varphi_j'(0)$ for all $i, j \in I$ and also $\varphi(t) \in \{\varphi_i(t) \,|\, i \in I\}$ for every $t \in (0, \delta)$. But then

$$\lim_{t \searrow 0} \frac{1}{t}[\varphi(t) - \varphi(0)] = \varphi_i'(0) \quad \text{for every } i \in I.$$

Hence, φ is directionally differentiable at the origin and $\varphi'(0;1) \in \{\varphi_i'(0) \mid i \in I\}$. Invoking 2D.1, this confirms semidifferentiability and establishes that the semiderivative function is a selection from $\{ Df_i(\bar{x}) \,|\, i \in I(\bar{x}) \}$. □

The functions in the examples given after 2D.2 are not only semidifferentiable but also piecewise smooth. Of course, a semidifferentiable function does not have to be piecewise smooth, e.g., when it is a selection of infinitely many, but not finitely many, smooth functions.

A more elaborate example of a piecewise smooth function is the projection mapping P_C on a nonempty, convex and closed set $C \subset \mathbb{R}^n$ specified by finitely many inequalities.

Exercise 2D.9 (piecewise smoothness of special projection mappings)**.** *For a convex set C of the form*

$$C = \{ x \in \mathbb{R}^n \,|\, g_i(x) \le 0,\ i = 1, \ldots, m \}$$

for convex functions g_i of class $\mathscr{C}^2$ on $\mathbb{R}^n$, let $\bar{x}$ be a point of C at which the gradients $\nabla g_i(\bar{x})$ associated with the active constraints, i.e., the ones with $g_i(\bar{x}) = 0$, are linearly independent. Then there is an open neighborhood O of $\bar{x}$ such that the projection mapping P_C is piecewise smooth on O.

Guide. Since in a sufficiently small neighborhood of $\bar{x}$ the inactive constraints remain inactive, one can assume without loss of generality that $g_i(\bar{x}) = 0$ for all $i = 1, \ldots, m$. Recall that because C is nonempty, closed and convex, P_C is a Lipschitz continuous function from $\mathbb{R}^n$ onto C (see 1D.5). For each u around $\bar{x}$ the projection $P_C(u)$ is the unique solution to the problem of minimizing $\frac{1}{2}|x-u|^2$ in x subject to $g_i(x) \le 0$ for $i = 1, \ldots, m$. The associated Lagrangian variational inequality (Theorem 2A.10) tells us that when u belongs to a small enough neighborhood of $\bar{x}$, the point x solves the problem if and only if x is feasible and there is a subset J of the index set $\{1, 2, \ldots, m\}$ and Lagrange multipliers $y_i \ge 0$, $i \in J$, such that

$$(6) \qquad \begin{cases} x + \sum_{i \in J} y_i \nabla g_i(x)^{\mathsf{T}} = u, \\ \qquad\qquad g_i(x) = 0, \quad i \in J. \end{cases}$$

The linear independence of the gradients of the active constraint gradients yields that the Lagrange multiplier vector y is unique, hence it is zero for $u = x = \bar{x}$. For each fixed subset J of the index set $\{1, 2, \ldots, m\}$ the Jacobian of the function on the left of (6) at $(\bar{x}, 0)$ is

$$Q = \begin{pmatrix} I_n + \sum_{i \in J} y_i \nabla^2 g_i(\bar{x}) & \nabla g_J(\bar{x})^{\mathsf{T}} \\ \nabla g_J(\bar{x}) & 0 \end{pmatrix},$$

where

$$\nabla g_J(\bar{x}) = \left[\frac{\partial g_i}{\partial x_j}(\bar{x}) \right]_{i \in J, j \in \{1, \ldots, n\}} \qquad \text{and } I_n \text{ is the } n \times n \text{ identity matrix.}$$

Since $\nabla g_J(\bar{x})$ has full rank, the matrix Q is nonsingular and then we can apply the classical inverse function theorem (Theorem 1A.1) to the equation (6), obtaining that its solution mapping $u \mapsto (x_J(u), y_J(u))$ has a smooth single-valued localization around $u = \bar{x}$ for $(x,y) = (\bar{x}, 0)$. There are finitely many subsets J of $\{1, \dots, m\}$, and for each u close to $\bar{x}$ we have $P_C(u) = x_J(u)$ for some J. Thus, the projection mapping P_C is a selection of finitely many smooth functions. □

Exercise 2D.10 (projection mapping). *For a set of the form $C = \{x \in \mathbb{R}^n \mid Ax = b \in \mathbb{R}^m\}$, if the $m \times n$ matrix A has linearly independent rows, then the projection mapping is given by*

$$P_C(x) = (I - A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}A)x + A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}b.$$

Guide. The optimality condition (6) in this case leads to the system of equations

$$\begin{pmatrix} x \\ b \end{pmatrix} = \begin{pmatrix} I & A^{\mathsf{T}} \\ A & 0 \end{pmatrix} \begin{pmatrix} P_C(x) \\ \lambda \end{pmatrix}.$$

Use the identity

$$\begin{pmatrix} I & A^{\mathsf{T}} \\ A & 0 \end{pmatrix} = \begin{pmatrix} I - A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}A & A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1} \\ (AA^{\mathsf{T}})^{-1}A & -(AA^{\mathsf{T}})^{-1} \end{pmatrix}^{-1}$$

to reach the desired conclusion. □

2E. Variational Inequalities with Polyhedral Convexity

In this section we apply the theory presented in the preceding sections of this chapter to the parameterized variational inequality

(1) $$f(p,x) + N_C(x) \ni 0$$

where $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ and C is a nonempty, closed and convex subset of $\mathbb{R}^n$. The corresponding solution mapping $S : \mathbb{R}^p \rightrightarrows \mathbb{R}^n$, with

(2) $$S(p) = \{x \mid f(p,x) + N_C(x) \ni 0\},$$

has already been the direct subject of Theorem 2B.1, the implicit function theorem of Robinson. From there we moved on to broader results about solution mappings to generalized equations, but now wish to summarize what those results mean back in the variational inequality setting, and furthermore to explore special features which emerge under additional assumptions on the set C.

Theorem 2E.1 (solution mappings for parameterized variational inequalities)**.** *For a variational inequality (1) and its solution mapping (2), let $\bar{p}$ and $\bar{x}$ be such that $\bar{x} \in S(\bar{p})$. Assume that*

(a) *f is strictly differentiable at $(\bar{p},\bar{x})$;*

(b) *the inverse G^{-1} of the mapping*

$$(3) \qquad G(x) = f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x}) + N_C(x), \quad \text{with } G(\bar{x}) \ni 0,$$

has a Lipschitz continuous single-valued localization σ around 0 for $\bar{x}$.

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(s;\bar{p}) \;\leq\; \operatorname{lip}(\sigma;0) \cdot |\nabla_p f(\bar{p},\bar{x})|,$$

and this localization s has a first-order approximation η at $\bar{p}$ given by

$$(4) \qquad \eta(p) = \sigma(-\nabla_p f(\bar{p},\bar{x})(p-\bar{p})).$$

Moreover, under the ample parameterization condition

$$\operatorname{rank} \nabla_p f(\bar{p},\bar{x}) = n,$$

the existence of a Lipschitz continuous single-valued localization s of S around $\bar{p}$ for $\bar{p}$ not only follows from but also necessitates the existence of a localization σ of G^{-1} having the properties described.

Proof. This comes from the application to S of the combination of Theorem 2B.5 and its specialization in Corollary 2B.10, together with the ample parameterization result in Theorem 2C.2. $\square$

If the localization σ that is assumed to exist in Theorem 2E.1 is actually linear, the stronger conclusion is obtained that s is differentiable at $\bar{p}$. But that's a circumstance which can hardly be guaranteed without supposing, for instance, that C is an affine set (given by a system of linear equations). In some situations, however, s could be at least piecewise smooth, as the projection mapping in 2D.9.

Our special goal here is trying to understand better the circumstances in which the existence of a single-valued localization σ of G^{-1} around 0 for $\bar{x}$ of the kind assumed in (b) of Theorem 2E.1 is assured. It's clear from the formula for G in (3) that everything hinges on how a normal cone mapping $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ may relate to an affine function $x \mapsto a + Ax$. The key lies in the local geometry of the graph of N_C. We will be able to make important progress in analyzing this geometry by restricting our attention to the following class of sets C.

Polyhedral convex sets. *A set C in $\mathbb{R}^n$ is said to be polyhedral convex when it can be expressed as the intersection of finitely many closed half-spaces and/or hyperplanes.*

In other words, C is a polyhedral convex set when it can be described by a finite set of constraints $f_i(x) \leq 0$ or $f_i(x) = 0$ on affine functions $f_i : \mathbb{R}^n \to \mathbb{R}$. Since an equation $f_i(x) = 0$ is equivalent to the pair of inequalities $f_i(x) \le 0$ and $-f_i(x) \le 0$, a polyhedral convex set C is characterized by having a (nonunique) representation of the form

$$C = \left\{ x \,\middle|\, \langle b_i, x\rangle \le \alpha_i \text{ for } i = 1,\ldots,m \right\}. \tag{5}$$

Any such set must obviously be closed. The empty set $\emptyset$ and the whole space $\mathbb{R}^n$ are regarded as polyhedral convex sets, in particular.

Polyhedral convex *cones* are characterized by having a representation (5) in which $\alpha_i = 0$ for all i. A basic fact about polyhedral convex cones is that they can equally well be represented in another way, which we recall next.

Theorem 2E.2 (Minkowski–Weyl theorem). *A set $K \subset \mathbb{R}^n$ is a polyhedral convex cone if and only if there is a collection of vectors $b_1,\ldots,b_m$ such that*

$$K = \left\{ y_1 b_1 + \cdots + y_m b_m \,\middle|\, y_i \ge 0 \text{ for } i = 1,\ldots,m \right\}. \tag{6}$$

It is easy to see that the cone K^* that is polar to a cone K having a representation of the kind in (6) consists of the vectors x satisfying $\langle b_i, x\rangle \le 0$ for $i = 1,\ldots,m$. The polar of a polyhedral convex cone having such an inequality representation must therefore have the representation in (6), inasmuch as $(K^*)^* = K$ for any closed, convex cone K. This fact leads to a special description of the tangent and normal cones to a polyhedral convex set.

Theorem 2E.3 (variational geometry of polyhedral convex sets). *Let C be a polyhedral convex set represented as in (5). Let $x \in C$ and $I(x) = \left\{ i \,\middle|\, \langle b_i, x\rangle = \alpha_i \right\}$, this being the set of indices of the constraints in (5) that are active at x. Then the tangent and normal cones to C at x are polyhedral convex, with the tangent cone having the representation*

$$T_C(x) = \left\{ w \,\middle|\, \langle b_i, w\rangle \le 0 \text{ for } i \in I(x) \right\} \tag{7}$$

and the normal cone having the representation

$$N_C(x) = \left\{ v \,\middle|\, v = \textstyle\sum_{i=1}^m y_i b_i \text{ with } y_i \ge 0 \text{ for } i \in I(x),\, y_i = 0 \text{ for } i \notin I(x) \right\}. \tag{8}$$

Furthermore, the tangent cone has the properties that

$$W \cap [C - x] = W \cap T_C(x) \text{ for some neighborhood } W \text{ of } 0 \tag{9}$$

and

$$T_C(x) \supset T_C(\bar{x}) \text{ for all } x \text{ in some neighborhood } U \text{ of } \bar{x}. \tag{10}$$

Proof. The formula (7) for $T_C(x)$ follows from (5) just by applying the definition of the tangent cone in 2A. Then from (7) and the preceding facts about polyhedral cones and polarity, utilizing also the relation in 2A(8), we obtain (8). The equality

(9) is deduced simply by comparing (5) and (7). To obtain (10), observe that $I(x) \subset I(\bar{x})$ for x close to $\bar{x}$ and then the inclusion follows from (7). □

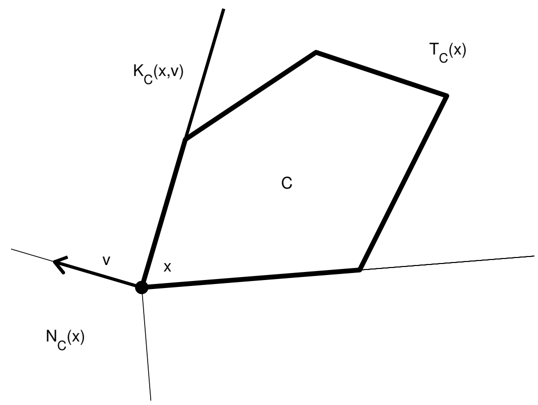

Fig. 2.2 Tangent, normal and critical cones to a polyhedral set.

The normal cone mapping N_C associated with a polyhedral convex set C has a special property which will be central to our analysis. It revolves around the following notion.

Critical cone. *For a convex set C, any $x \in C$ and any $v \in N_C(x)$, the critical cone to C at x for v is*

$$K_C(x,v) = \left\{ w \in T_C(x) \,\middle|\, w \perp v \right\}.$$

If C is polyhedral, then $K_C(x,v)$ is polyhedral as well, as seen immediately from the representation in (7).

Lemma 2E.4 (reduction lemma)**.** *Let C be a polyhedral convex set in $\mathbb{R}^n$, and let*

$$\bar{x} \in C, \quad \bar{v} \in N_C(\bar{x}), \quad K = K_C(\bar{x}, \bar{v}).$$

The graphical geometry of the normal cone mapping N_C around $(\bar{x}, \bar{v})$ reduces then to the graphical geometry of the normal cone mapping N_K around $(0,0)$, in the sense that

$$O \cap [\operatorname{gph} N_C - (\bar{x}, \bar{v})] = O \cap \operatorname{gph} N_K \quad \text{for some neighborhood } O \text{ of } (0,0).$$

In other words, one has

$$(11) \quad \bar{v} + u \in N_C(\bar{x} + w) \iff u \in N_K(w) \quad \text{for } (w,u) \text{ sufficiently near to } (0,0).$$

Proof. Since we are only involved with local properties of C around one of its points $\bar{x}$, and $C - \bar{x}$ agrees with the cone $T_C(\bar{x})$ around 0 by Theorem 2E.3, we can assume without loss of generality that $\bar{x} = 0$ and C is a cone, and $T_C(\bar{x}) = C$. Then, in terms of the polar cone C^* (which likewise is polyhedral on the basis of Theorem 2E.2), we have the characterization from 2A.3 that

$$(12) \qquad v \in N_C(w) \iff w \in N_{C^*}(v) \iff w \in C, \quad v \in C^*, \quad \langle v, w\rangle = 0.$$

In particular for our focus on the geometry of $\operatorname{gph} N_C$ around $(0,\bar{v})$, we have from (12) that

$$(13) \qquad N_{C^*}(\bar{v}) = \{ w \in C \,|\, \langle \bar{v}, w\rangle = 0 \} = K.$$

We know on the other hand from 2E.3 that $U \cap [C^* - \bar{v}] = U \cap T_{C^*}(\bar{v})$ for some neighborhood U of 0, where moreover $T_{C^*}(\bar{v})$ is polar to $N_{C^*}(\bar{v})$, hence equal to K^* by (13). Thus, there is a neighborhood O of $(0,0)$ such that

$$(14) \qquad \text{for } (w,u) \in O: \qquad \bar{v} + u \in N_C(w) \iff w \in C,\ u \in K^*,\ \langle \bar{v} + u, w\rangle = 0.$$

This may be compared with the fact that

$$(15) \qquad u \in N_K(w) \iff w \in K,\ u \in K^*,\ \langle u, w\rangle = 0.$$

Our goal (in the context of $\bar{x} = 0$) is to show that (14) reduces to (11), at least when the neighborhood O in (14) is chosen still smaller, if necessary. Because of (15), this comes down to demonstrating that $\langle \bar{v}, w\rangle = 0$ in the circumstances of (14).

We can take C to be represented by

$$(16) \qquad C = \{ w \,|\, \langle b_i, w\rangle \le 0 \text{ for } i = 1,\dots,m \},$$

in which case, as observed after 2E.2, the polar C^* is represented by

$$(17) \qquad C^* = \{ y_1 b_1 + \cdots + y_m b_m \,|\, y_i \ge 0 \text{ for } i = 1,\dots,m \}.$$

The relations in (12) can be coordinated with these representations as follows. For each index set $I \subset \{1,\dots,m\}$, consider the polyhedral convex cones

$$W_I = \{ w \in C \,|\, \langle b_i, w\rangle = 0 \text{ for } i \in I \}, \qquad V_I = \{ \textstyle\sum_{i\in I} y_i b_i \text{ with } y_i \ge 0 \},$$

with $W_\emptyset = C$ and $V_\emptyset = \{0\}$. Then $v \in N_C(w)$ if and only if, for some I, one has $w \in W_I$ and $v \in V_I$. In other words, $\operatorname{gph} N_C$ is the union of the finitely many polyhedral convex cones $G_I = W_I \times V_I$ in $\mathbb{R}^n \times \mathbb{R}^n$.

Among these cones G_I, we will only be concerned with the ones containing $(0,\bar{v})$. Let $\mathscr{I}$ be the collection of index sets $I \subset \{1,\dots,m\}$ having that property. According to (9) in 2E.3, there exists for each $I \in \mathscr{I}$ a neighborhood O_I of $(0,0)$ such that $O_I \cap [G_I - (0,\bar{v})] = O_I \cap T_{G_I}(0,\bar{v})$. Furthermore, $T_{G_I}(0,\bar{v}) = W_I \times T_{V_I}(\bar{v})$. This has the crucial consequence that when $\bar{v} + u \in N_C(w)$ with (w,u) near enough to $(0,0)$, we also have $\bar{v} + \tau u \in N_C(w)$ for all $\tau \in [0,1]$. Since having $\bar{v} + \tau u \in N_C(w)$ entails having $\langle \bar{v} + \tau u, w\rangle = 0$ through (12), this implies that $\langle \bar{v}, w\rangle = -\tau\langle u, w\rangle$ for all $\tau \in$

$[0,1]$. Hence $\langle \bar{v}, w\rangle = 0$, as required. We merely have to shrink the neighborhood O in (14) to lie within every O_I for $I \in \mathscr{I}$. $\square$

Example 2E.5. The nonnegative orthant $\mathbb{R}^n_+$ is a polyhedral convex cone in $\mathbb{R}^n$, since it consists of the vectors $x = (x_1,\dots,x_n)$ satisfying the linear inequalities $x_j \geq 0$, $j = 1,\dots,n$. For $v = (v_1,\dots,v_n)$, one has

$$v \in N_{\mathbb{R}^n_+}(x) \iff x_j \geq 0,\ v_j \leq 0,\ x_j v_j = 0 \text{ for } j = 1,\dots,n.$$

Thus, whenever $v \in N_{\mathbb{R}^n_+}(x)$ one has in terms of the index sets

$$\begin{aligned} J_1 &= \{\, j \mid x_j > 0,\ v_j = 0 \}, \\ J_2 &= \{\, j \mid x_j = 0,\ v_j = 0 \}, \\ J_3 &= \{\, j \mid x_j = 0,\ v_j < 0 \} \end{aligned}$$

that the vectors $w = (w_1,\dots,w_n)$ belonging to the critical cone to $\mathbb{R}^n_+$ at x for v are characterized by

$$w \in K_{\mathbb{R}^n_+}(x,v) \iff \begin{cases} w_j \text{ free} & \text{for } j \in J_1, \\ w_j \geq 0 & \text{for } j \in J_2, \\ w_j = 0 & \text{for } j \in J_3. \end{cases}$$

In the developments ahead, we will make use of not only critical cones but also certain subspaces.

Critical subspaces. *The smallest linear subspace that includes the critical cone $K_C(x,v)$ will be denoted by $K_C^+(x,v)$, whereas the largest linear subspace that is included in $K_C(x,v)$ will be denoted by $K_C^-(x,v)$, the formulas being*

$$(18)\qquad \begin{aligned} K_C^+(x,v) &= K_C(x,v) - K_C(x,v) = \{\, w - w' \mid w, w' \in K_C(x,v) \}, \\ K_C^-(x,v) &= K_C(x,v) \cap [-K_C(x,v)] = \{\, w \in K_C(x,v) \mid -w \in K_C(x,v) \}. \end{aligned}$$

The formulas follow from the fact that $K_C(x,v)$ is already a convex cone. Obviously, $K_C(x,v)$ is itself a subspace if and only if $K_C^+(x,v) = K_C^-(x,v)$.

Theorem 2E.6 (affine-polyhedral variational inequalities)**.** *For an affine function $x \mapsto a + Ax$ from $\mathbb{R}^n$ into $\mathbb{R}^n$ and a polyhedral convex set $C \subset \mathbb{R}^n$, consider the variational inequality*

$$a + Ax + N_C(x) \ni 0.$$

Let $\bar{x}$ be a solution and let $\bar{v} = -a - A\bar{x}$, so that $\bar{v} \in N_C(\bar{x})$, and let $K = K_C(\bar{x},\bar{v})$ be the associated critical cone. Then for the mappings

$$(19)\qquad \begin{aligned} G(x) &= a + Ax + N_C(x) \text{ with } G(\bar{x}) \ni 0, \\ G_0(w) &= Aw + N_K(w) \text{ with } G_0(0) \ni 0, \end{aligned}$$

the following properties are equivalent:

(a) G^{-1} *has a Lipschitz continuous single-valued localization* σ *around* 0 *for* $\bar{x}$;
(b) G_0^{-1} *is a single-valued mapping with all of* $\mathbb{R}^n$ *as its domain,*

in which case G_0^{-1} *is necessarily Lipschitz continuous globally and the function* $\sigma(v) = \bar{x} + G_0^{-1}(v)$ *furnishes the localization in (a). Moreover, in terms of critical subspaces* $K^+ = K_C^+(\bar{x},\bar{v})$ *and* $K^- = K_C^-(\bar{x},\bar{v})$, *the following condition is sufficient for (a) and (b) to hold:*

$$(20) \qquad w \in K^+, \quad Aw \perp K^-, \quad \langle w, Aw \rangle \leq 0 \quad \Longrightarrow \quad w = 0.$$

Proof. According to reduction lemma 2E.4, we have, for (w,u) in some neighborhood of $(0,0)$, that $\bar{v} + u \in N_C(\bar{x} + w)$ if and only if $u \in N_K(w)$. In the change of notation from u to $v = u + Aw$, this means that, for (w,v) in a neighborhood of $(0,0)$, we have $v \in G(\bar{x} + w)$ if and only if $v \in G_0(w)$. Thus, the existence of a Lipschitz continuous single-valued localization σ of G^{-1} around 0 for $\bar{x}$ as in (a) corresponds to the existence of a Lipschitz continuous single-valued localization σ_0 of G_0^{-1} around 0 for 0; the relationship is given by $\sigma(v) = \bar{x} + \sigma_0(v)$. But whenever $v \in G_0(w)$ we have $\lambda v \in G_0(\lambda w)$ for all $\lambda > 0$, i.e., the graph of G_0 is a cone. Therefore, when σ_0 exists it can be scaled arbitrarily large and must correspond to G_0^{-1} being a single-valued mapping with all of $\mathbb{R}^n$ as its domain.

We claim next that when G_0^{-1} is single-valued everywhere it is necessarily Lipschitz continuous. This comes out of the argument pursued in the proof of 2E.4 in analyzing the graph of N_C, which applies equally well to N_K, inasmuch as K is a polyhedral convex cone. Specifically, the graph of N_K is the union of finitely many polyhedral convex cones in $\mathbb{R}^n \times \mathbb{R}^n$. The same also holds then for the graphs of G_0 and G_0^{-1}. It remains only to observe that if a single-valued mapping has its graph composed of the union of finitely many polyhedral convex sets it has to be Lipschitz continuous (prove or see 3D.6).

This leaves us with verifying that the condition in (20) is sufficient for G_0^{-1} to be single-valued with all of $\mathbb{R}^n$ as its domain. We note in preparation for this that

$$(21) \qquad (K^+)^\perp = K^* \cap (-K^*) = (K^*)^-, \qquad (K^-)^\perp = K^* - K^* = (K^*)^+.$$

We first argue that if $w_1 \in G_0^{-1}(v)$ and $w_2 \in G_0^{-1}(v)$, then $v - Aw_1 \in N_K(w_1)$ and $v - Aw_2 \in N_K(w_2)$. This entails having

$$\begin{aligned} w_1 \in K, \quad v - Aw_1 \in K^*, \quad \langle w_1, v - Aw_1 \rangle = 0, \\ w_2 \in K, \quad v - Aw_2 \in K^*, \quad \langle w_2, v - Aw_2 \rangle = 0, \end{aligned}$$

with $\langle w_1, v - Aw_2 \rangle \leq 0$ and $\langle w_2, v - Aw_1 \rangle \leq 0$. Then $w_1 - w_2 \in K - K = K^+$ and $-A(w_1 - w_2) \in K^* - K^* = (K^-)^\perp$, with $\langle w_1 - w_2, A(w_1 - w_2) \rangle = \langle w_1 - w_2, [v - Aw_2] - [v - Aw_2] \rangle \leq 0$. Under our condition (20), these relations require $w_1 - w_2 = 0$. Thus, (20) guarantees that $G_0^{-1}(v)$ can never contain more than a single w.

Working toward showing that (20) guarantees also that $\operatorname{dom} G_0^{-1} = \mathbb{R}^n$, we next consider the case where $\operatorname{dom} G_0^{-1}$ omits some point $\tilde{v}$ and analyze what that would imply. Again we utilize the fact that the graph of G_0^{-1} is the union of finitely many polyhedral convex cones in $\mathbb{R}^n \times \mathbb{R}^n$. Under the mapping $(v,w) \to v$, each of them projects onto a cone in $\mathbb{R}^n$; the union of these cones is $\operatorname{dom} G_0^{-1}$. Since the image of a polyhedral convex cone under a linear transformation is another polyhedral convex cone, in consequence of 2E.2 (since the image of a cone generated by finitely many vectors is another such cone), and polyhedral convex cones are closed sets in particular, this ensures that $\operatorname{dom} G_0^{-1}$ is closed. Then there is certain to exist a point $v_0 \in \operatorname{dom} G_0^{-1}$ that is closest to $\tilde{v}$; for all $\tau > 0$ sufficiently small, we have $v_0 + \tau(\tilde{v} - v_0) \notin \operatorname{dom} G_0^{-1}$. For each of the polyhedral convex cones D in the finite union making up $\operatorname{dom} G_0^{-1}$, if $v_0 \in D$, then v_0 must be the projection $P_D(\tilde{v})$, so that $\tilde{v} - v_0$ must belong to $N_D(v_0)$ (cf. relation (4) in 2A). It follows that, for some neighborhood U of v_0, we have

$$\langle \tilde{v} - v_0, u - v_0 \rangle \leq 0 \text{ for all } u \in U \cap \operatorname{dom} G_0^{-1}. \tag{22}$$

Consider any $w_0 \in G_0^{-1}(v_0)$; this means $v_0 - Aw_0 \in N_K(w_0)$. Let K_0 be the critical cone to K at w_0 for $v_0 - Aw_0$:

$$K_0 = \left\{ w' \in T_K(w_0) \,\middle|\, w' \perp (v_0 - Aw_0) \right\}. \tag{23}$$

In the line of argument already pursued, the geometry of the graph of N_K around $(w_0, v_0 - Aw_0)$ can be identified with that of the graph of N_{K_0} around $(0,0)$. Equivalently, the geometry of the graph of $G_0^{-1} = (A + N_K)^{-1}$ around (v_0, w_0) can be identified with that of $(A + N_{K_0})^{-1}$ around $(0,0)$; for (v', w') near enough to $(0,0)$, we have $w_0 + w' \in G_0^{-1}(v_0 + v')$ if and only if $w' \in (A + N_{K_0})^{-1}(v')$. Because of (22) holding for the neighborhood U of v_0, this implies that $\langle \tilde{v} - v_0, v' \rangle \leq 0$ for all $v' \in \operatorname{dom}(A + N_{K_0})^{-1}$ close to 0. Thus,

$$\langle \tilde{v} - v_0, Aw' + u' \rangle \leq 0 \text{ for all } w' \in K_0 \text{ and } u' \in N_{K_0}(w'). \tag{24}$$

The case of $w' = 0$ has $N_{K_0}(w') = K_0^*$, so (24) implies in particular that $\langle \tilde{v} - v_0, u' \rangle \leq 0$ for all $u' \in K_0^*$, so that $\tilde{v} - v_0 \in (K_0^*)^* = K_0$. On the other hand, since $u' = 0$ is always one of the elements of $N_{K_0}(w')$, we must have from (24) that $\langle \tilde{v} - v_0, Aw' \rangle \leq 0$ for all $w' \in K_0$. Here $\langle \tilde{v} - v_0, Aw' \rangle = \langle A^{\mathsf{T}}(\tilde{v} - v_0), w' \rangle$ for all $w' \in K_0$, so this means $A^{\mathsf{T}}(\tilde{v} - v_0) \in K_0^*$. In summary, (24) requires, among other things, having

$$\begin{aligned} &\tilde{v} - v_0 \in K_0 \text{ and } A^{\mathsf{T}}(\tilde{v} - v_0) \in K_0^*, \\ &\qquad\qquad \text{hence in particular } \langle A^{\mathsf{T}}(\tilde{v} - v_0), \tilde{v} - v_0 \rangle \leq 0. \end{aligned} \tag{25}$$

We observe now from the formula for K_0 in (23) that $K_0 \subset T_K(w_0)$, where furthermore $T_K(w_0)$ is the cone generated by the vectors $w - w_0$ with $w \in K$ and hence lies in $K - K$. Therefore $K_0 \subset K^+$. On the other hand, because $T_K(w_0)$ and $N_K(w_0)$ are polar to each other by 2E.3, we have from (23) that K_0 is polar to the cone comprised by all differences $v - \tau(v_0 - Aw_0)$ with $v \in N_K(w_0)$ and $\tau \geq 0$, which is again polyhedral. That cone of differences must then in fact be K_0^*. Since we have taken v_0 and w_0 to satisfy $v_0 - Aw_0 \in N_K(w_0)$, and also $N_K(w_0) \subset K^*$, it follows that $K_0^* \subset K^* - K^* = (K^-)^\perp$. Thus, (25) implies that $\tilde{v} - v_0 \in K^+$, $A^{\mathsf{T}}(\tilde{v} - v_0) \in (K^-)^\perp$, with $\langle A^{\mathsf{T}}(\tilde{v}-v_0),\tilde{v}-v_0\rangle \leq 0$. In consequence, (25) would be impossible if we knew that

$$(26)\qquad w\in K^+,\quad A^{\mathsf{T}}w\perp K^-,\quad \langle A^{\mathsf{T}}w,w\rangle\leq 0 \quad\Longrightarrow\quad w=0.$$

Our endgame will be to demonstrate that (26) is actually equivalent to condition (20).

Of course, $\langle A^{\mathsf{T}}w,w\rangle$ is the same as $\langle w,Aw\rangle$. For additional comparison between (20) and (26), we can simplify matters by expressing $\mathbb{R}^n$ as $W_1\times W_2\times W_3$ for the linear subspaces $W_1=K^-$, $W_3=(K^+)^{\perp}$, and W_2 the orthogonal complement of K^- within K^+. Any vector $w\in\mathbb{R}^n$ corresponds then to a triple (w_1,w_2,w_3) in this product, and there are linear transformations $A_{ij}:W_j\to W_i$ such that

$$Aw\longleftrightarrow(A_{11}w_1+A_{12}w_2+A_{13}w_3,A_{21}w_1+A_{22}w_2+A_{23}w_3,A_{31}w_1+A_{32}w_2+A_{33}w_3).$$

In this schematic, (20) has the form

$$(27)\quad A_{11}w_1+A_{12}w_2=0,\quad \langle w_2,A_{21}w_1+A_{22}w_2\rangle\leq 0 \quad\Longrightarrow\quad w_1=0,\ w_2=0,$$

whereas (26) has the form

$$(28)\quad A_{11}^{\mathsf{T}}w_1+A_{21}^{\mathsf{T}}w_2=0,\quad \langle w_2,A_{12}^{\mathsf{T}}w_1+A_{22}^{\mathsf{T}}w_2\rangle\leq 0 \quad\Longrightarrow\quad w_1=0,\ w_2=0.$$

In particular, through the choice of $w_2=0$, (27) insists that the only w_1 with $A_{11}w_1=0$ is $w_1=0$. Thus, A_{11} must be nonsingular. Then the initial equation in (27) can be solved for w_1, yielding $w_1=-A_{11}^{-1}A_{12}w_2$, and this expression can be substituted into the inequality, thereby reducing the condition to

$$\langle w_2,(A_{22}-A_{21}A_{11}^{-1}A_{12})w_2\rangle\leq 0 \quad\Longrightarrow\quad w_2=0.$$

In the same manner, (28) comes out as the nonsingularity of A_{11}^{T} and the property that

$$\langle w_2,(A_{22}^{\mathsf{T}}-A_{12}^{\mathsf{T}}(A_{11}^{\mathsf{T}})^{-1}A_{21}^{\mathsf{T}})w_2\rangle\leq 0 \quad\Longrightarrow\quad w_2=0.$$

Since the nonsingularity of A_{11}^{T} is equivalent to that of A_{11}, and

$$A_{22}^{\mathsf{T}}-A_{12}^{\mathsf{T}}(A_{11}^{\mathsf{T}})^{-1}A_{21}^{\mathsf{T}}=(A_{22}-A_{21}A_{11}^{-1}A_{12})^{\mathsf{T}},$$

the equivalence of (27) and (28) is now evident. □

Examples 2E.7.

(a) When the critical cone K in Theorem 2E.6 is a subspace, the condition in (20) reduces to the nonsingularity of the linear transformation $K\ni w\mapsto P_K(Aw)$, where P_K is the projection onto K.

(b) When the critical cone K in Theorem 2E.6 is pointed, in the sense that $K\cap(-K)=\{0\}$, the condition in (20) reduces to the requirement that $\langle w,Aw\rangle>0$ for all nonzero $w\in K^+$.

(c) Condition (20) always holds when A is the identity matrix.

Theorem 2E.6 tells us that, for a polyhedral convex set C, the assumption in (b) of Theorem 2E.1 is equivalent to the critical cone $K = K_C(\bar{x}, -f(\bar{p},\bar{x}))$ being such that the inverse G_0^{-1} of the mapping $G_0 : w \mapsto \nabla_x f(\bar{x},\bar{p})w + N_K(w)$ is single-valued from all of $\mathbb{R}^n$ into itself and hence Lipschitz continuous globally. Furthermore, Theorem 2E.6 provides a sufficient condition for this to hold. In putting these facts together with observations about the special nature of G_0, we obtain a powerful fact which distinguishes variational inequalities with polyhedral convexity from other variational inequalities.

Theorem 2E.8 (localization criterion under polyhedral convexity)**.** *For a variational inequality (1) and its solution mapping (2) under the assumption that C is polyhedral convex and f is strictly differentiable at $(\bar{p},\bar{x})$, with $\bar{x} \in S(\bar{p})$, let*

$$A = \nabla_x f(\bar{p},\bar{x}) \quad \text{and} \quad K = K_C(\bar{x},\bar{v}) \text{ for } \bar{v} = -f(\bar{p},\bar{x}).$$

Suppose that for each $u \in \mathbb{R}^n$ there is a unique solution $w = \bar{s}(u)$ to the auxiliary variational inequality $Aw - u + N_K(w) \ni 0$, this being equivalent to saying that

$$\text{(29)} \qquad \bar{s} = (A + N_K)^{-1} \text{ is everywhere single-valued,}$$

in which case the mapping $\bar{s}$ is Lipschitz continuous globally. (A sufficient condition for this assumption to hold is the property in (20) with respect to the critical subspaces $K^+ = K_C^+(\bar{x},\bar{v})$ and $K^- = K_C^-(\bar{x},\bar{v})$.)

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ which is semidifferentiable with

$$\text{(30)} \qquad \operatorname{lip}(s;\bar{p}) \le \operatorname{lip}(\bar{s};0) \cdot |\nabla_p f(\bar{p},\bar{x})|, \qquad Ds(\bar{p})(q) = \bar{s}(-\nabla_p f(\bar{p},\bar{x})q).$$

Moreover, under the ample parameterization condition, rank $\nabla_p f(\bar{p},\bar{x}) = n$, *condition (29) is not only sufficient but also necessary for a Lipschitz continuous single-valued localization of S around $\bar{p}$ for $\bar{x}$.*

Proof. We merely have to combine the observation made before this theorem's statement with the statement of Theorem 2E.1. According to formula (4) in that theorem for the first-order approximation η of s at $\bar{p}$, we have $\eta(\bar{p}+q) - \bar{x} = \bar{s}(-\nabla_p f(\bar{p},\bar{x})q)$. Because K is a cone, the mapping N_K is positively homogeneous, and the same is true then for $A + N_K$ and its inverse, which is $\bar{s}$. Thus, the function $q \mapsto \bar{s}(-\nabla_p f(\bar{p},\bar{x})q)$ gives a first-order approximation to $s(\bar{p}+q) - s(\bar{p})$ at $q = 0$ that is positively homogeneous. We conclude that s is semidifferentiable at $\bar{p}$ with this function furnishing its semiderivative, as indicated in (30). □

As a special case of Example 2E.7(a), if $C = \mathbb{R}^n$ the result in Theorem 2E.8 reduces once more to a version of the classical implicit function theorem. Further insights into solution mappings associated with variational inequalities will be gained in Chapter 4.

Exercise 2E.9. *Prove that the projection mapping P_C associated with a polyhedral convex set C is Lipschitz continuous and semidifferentiable everywhere, with its* *semiderivative being given by*

$$DP_C(x)(u) = P_K(u) \quad \text{for } K = K_C(P_C(x), x - P_C(x)).$$

Guide. Use the relation between the projection mapping and the normal cone mapping given in formula 2A(4). □

Additional facts about critical cones, which will be useful later, can be developed from the special geometric structure of polyhedral convex sets.

Proposition 2E.10 (local behavior of critical cones and subspaces)**.** *Let $C \subset \mathbb{R}^n$ be a polyhedral convex set, and let $\bar{v} \in N_C(\bar{x})$. Then the following properties hold:*

(a) $K_C(x,v) \subset K_C^+(\bar{x},\bar{v})$ *for all* $(x,v) \in \operatorname{gph} N_C$ *in some neighborhood of* $(\bar{x},\bar{v})$*;*

(b) $K_C(x,v) = K_C^+(\bar{x},\bar{v})$ *for some* $(x,v) \in \operatorname{gph} N_C$ *in each neighborhood of* $(\bar{x},\bar{v})$*.*

Proof. By appealing to 2E.3 as in the proof of 2E.4, we can reduce to the case where $\bar{x} = 0$ and C is a cone. Theorem 2E.2 then provides a representation in terms of a collection of nonzero vectors $b_1, \dots, b_m$, in which C consists of all linear combinations $y_1 b_1 + \cdots + y_m b_m$ with coefficients $y_i \geq 0$, and the polar cone C^* consists of all v such that $\langle b_i, v\rangle \leq 0$ for all i. We know from 2A.3 that, at any $x \in C$, the normal cone $N_C(x)$ is formed by the vectors $v \in C^*$ such that $\langle x, v\rangle = 0$, so that $N_C(x)$ is the cone that is polar to the one comprised of all vectors $w - \lambda x$ with $w \in C$ and $\lambda \geq 0$. Since the latter cone is again polyhedral (in view of Theorem 2E.2), hence closed, it must in turn be the cone polar to $N_C(x)$ and therefore equal to $T_C(x)$. Thus,

$$T_C(x) = \left\{ y_1 b_1 + \cdots + y_m b_m - \lambda x \,\middle|\, y_i \geq 0,\ \lambda \geq 0 \right\} \quad \text{for any } x \in C.$$

On the other hand, in the notation

$$\text{(31)} \qquad \begin{aligned} I(v) &= \left\{ i \,\middle|\, \langle b_i, v\rangle = 0 \right\}, \\ F(v) &= \left\{ y_1 b_1 + \cdots + y_m b_m \,\middle|\, y_i \geq 0 \text{ for } i \in I(v),\ y_i = 0 \text{ for } i \notin I(v) \right\}, \end{aligned}$$

we see that for $v \in C^*$ we have $F(v) = \left\{ x \in C \,\middle|\, v \in N_C(x) \right\}$, i.e.,

$$v \in N_C(x) \iff x \in F(v).$$

Then too, for such x and v, the critical cone $K_C(x,v) = \left\{ w \in T_C(x) \,\middle|\, \langle w, v\rangle = 0 \right\}$ we have

$$\text{(32)} \qquad K_C(x,v) = \left\{ w - \lambda x \,\middle|\, w \in F(v),\ \lambda \geq 0 \right\},$$

and actually $K_C(\bar{x},\bar{v}) = F(\bar{v})$ (inasmuch as $\bar{x} = 0$). In view of the fact, evident from (31), that

$$I(v) \subset I(\bar{v}) \quad \text{and} \quad F(v) \subset F(\bar{v}) \quad \text{for all } v \text{ near enough to } \bar{v},$$

we have, for v in some neighborhood of $\bar{v}$, that

$$x \in F(\bar{v}) \;\text{ and }\; K_C(x,v) \subset \left\{ w - \lambda x \,\middle|\, w \in F(\bar{v}),\, \lambda \geq 0 \right\} \;\text{ when } v \in N_C(x).$$

In that case $K_C(x,v) \subset F(\bar{v}) - F(\bar{v}) = K_C(\bar{x},\bar{v}) - K_C(\bar{x},\bar{v}) = K_C^+(\bar{x},\bar{v})$, so (a) is valid.

To confirm (b), it will be enough now to demonstrate that, arbitrarily close to $\bar{x} = 0$, we can find a vector $\tilde{x}$ for which $K_C(\tilde{x},\bar{v}) = F(\bar{v}) - F(\bar{v})$. Here $F(\bar{v})$ consists by definition of all nonnegative linear combinations of the vectors b_i with $i \in I(\bar{v})$, whereas $F(\bar{v}) - F(\bar{v})$ is the subspace consisting of all linear combinations. For arbitrary $\varepsilon > 0$, let $\tilde{x} = \tilde{y}_1 b_1 + \cdots + \tilde{y}_m b_m$ with $\tilde{y}_i = \varepsilon$ for $i \in I(\bar{v})$ but $\tilde{y}_i = 0$ for $i \notin I(\bar{v})$. Then $K_C(\tilde{x},\bar{v})$, equaling $\left\{ w - \lambda \tilde{x} \,\middle|\, w \in F(\bar{v}),\, \lambda \geq 0 \right\}$ by (32), consists of all linear combinations of the vectors b_i for $i \in I(\bar{v})$ in which the coefficients have the form $y_i - \lambda\varepsilon$ with $y_i \geq 0$ and $\lambda \geq 0$. Can any given choice of coefficients y_i' for $i \in I(\bar{v})$ be obtained in this manner? Yes, by taking λ high enough that $y_i' + \lambda\varepsilon \geq 0$ for all $i \in I(\bar{v})$ and then setting $y_i = y_i' + \lambda\varepsilon$. This completes the argument. □

2F. Variational Inequalities with Monotonicity

Our attention shifts now from special properties of the set C in a variational inequality to special properties of the function f and their effect on solutions.

In Section 1H we presented an implicit function theorem for equations involving strictly monotone functions. In this section we develop some special results for the variational inequality

$$f(x) + N_C(x) \ni 0, \tag{1}$$

in the case when f is monotone or strongly monotone. Recall that a function $f : \mathbb{R}^n \to \mathbb{R}^n$ is said to be *monotone* on a convex set $C \subset \operatorname{dom} f$ if

$$\langle f(x') - f(x), x' - x \rangle \geq 0 \;\text{ for all } x, x' \in C. \tag{2}$$

It is *strongly* monotone on C with constant $\mu > 0$ when

$$\langle f(x') - f(x), x' - x \rangle \geq \mu |x' - x|^2 \;\text{ for all } x, x' \in C. \tag{3}$$

We work with the basic perturbation scheme in which $f(x)$ is replaced by $f(x) - p$ for a parameter vector $p \in \mathbb{R}^n$. The solution mapping is then

$$S(p) = \left\{ x \,\middle|\, p - f(x) \in N_C(x) \right\} = (f + N_C)^{-1}(p), \tag{4}$$

with the solution set to (1) then being $S(0)$.

Theorem 2F.1 (solution convexity for monotone variational inequalities)**.** *For a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a nonempty closed convex set $C \subset \operatorname{dom} f$ relative to which f is monotone and continuous, the solution mapping S in (4) is closed and* *convex valued. In particular, therefore, the set of solutions (if any) to the variational inequality (1) is not only closed but also convex.*

Proof. It suffices to deal with $S(0)$, since $S(p) = (f_p + N_C)^{-1}(0)$ for $f_p(x) = f(x) - p$ (which is monotone and continuous like f).

The closedness of $S(0)$ already follows from Theorem 2A.1. To see the convexity, consider any two points x_0 and x_1 in $S(0)$. We have $-f(x_0) \in N_C(x_0)$ and $-f(x_1) \in N_C(x_1)$; this is equivalent to

(5) $$\langle f(x_0), x - x_0\rangle \geq 0 \;\text{ and }\; \langle f(x_1), x - x_1\rangle \geq 0 \;\text{ for all } x \in C.$$

Let $\bar{x} = (1-\lambda)x_0 + \lambda x_1$ for any $\lambda \in (0,1)$. Then $\bar{x} \in C$ by convexity. Consider an arbitrary point $\tilde{x} \in C$. The goal is to show that $\langle f(\bar{x}), \tilde{x} - \bar{x}\rangle \geq 0$, which will confirm that $-f(\bar{x}) \in N_C(\bar{x})$, i.e., that $\bar{x} \in S(0)$.

Taking $t \in (0,1)$ as a parameter, let $x(t) = \bar{x} + t(\tilde{x} - \bar{x})$ and note that the convexity of C ensures $x(t) \in C$. From the monotonicity of f and the first inequality in (5) we have

$$0 \leq \langle f(x(t)) - f(x_0), x(t) - x_0\rangle + \langle f(x_0), x(t) - x_0\rangle = \langle f(x(t)), x(t) - x_0\rangle.$$

In parallel from the second inequality in (5), we have $0 \leq \langle f(x(t)), x(t) - x_1\rangle$. Therefore

$$\begin{aligned} 0 &\leq (1-\lambda)\langle f(x(t)), x(t) - x_0\rangle + \lambda\langle f(x(t)), x(t) - x_1\rangle \\ &= \langle f(x(t)), x(t) - (1-\lambda)x_0 - \lambda x_1\rangle, \end{aligned}$$

where the final expression equals $\langle f(x(t)), x(t) - \bar{x}\rangle = t\langle f(x(t)), \tilde{x} - \bar{x}\rangle$, since $x(t) - \bar{x} = t[\tilde{x} - \bar{x}]$. Thus, $0 \leq \langle f(x(t)), \tilde{x} - \bar{x}\rangle$. Because $x(t) \to \bar{x}$ as $t \to 0$, and f is continuous, we conclude that $\langle f(\bar{x}), \tilde{x} - \bar{x}\rangle \geq 0$, as required. □

In order to add nonemptiness of the solution set to the conclusions of Theorem 2F.1 we need an existence theorem for the variational inequality (1). There is already such a result in 2A.1, but only for bounded sets C. The following result goes beyond that boundedness restriction, without yet imposing any monotonicity assumption on f. When combined with monotonicity, it will have particularly powerful consequences.

Theorem 2F.2 (solution existence for variational inequalities without boundedness)**.** *Consider a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a nonempty closed convex set $C \subset$ dom f relative to which f is continuous (but not necessarily monotone). Suppose there exist $\hat{x} \in C$ and $\rho > 0$ such that*

(6) *there is no $x \in C$ with $|x - \hat{x}| \geq \rho$ and $\langle f(x), x - \hat{x}\rangle \leq 0$.*

Then the variational inequality (1) has a solution, and every solution x of (1) satisfies $|x - \hat{x}| < \rho$.

Proof. Any solution x to (1) would have $\langle f(x), x-\hat{x}\rangle \le 0$ in particular, and then necessarily $|x-\hat{x}| < \rho$ under (6). Hence it will suffice to show that (6) guarantees the existence of at least one solution x to (1) with $|x-\hat{x}| < \rho$.

Let $C_\rho = \{ x \in C \mid |x-\hat{x}| \le \rho \}$ and consider the modified variational inequality (1) in which C is replaced by C_ρ. According to Theorem 2A.1, this modified variational inequality has a solution $\bar{x}$. We have $\bar{x} \in C_\rho$ and $-f(\bar{x}) \in N_{C_\rho}(\bar{x})$. From 2A.8(b) we know that $N_{C_\rho}(\bar{x}) = N_C(\bar{x}) + N_B(\bar{x})$ for the ball $B = \mathbb{B}_\rho(\hat{x}) = \{ x \mid |x-\hat{x}| \le \rho \}$. Thus,

$$(7) \qquad -f(\bar{x}) - w \in N_C(\bar{x}) \text{ for some } w \in N_B(\bar{x}).$$

By demonstrating that this implies $w = 0$, we will be able to see that $\bar{x}$ actually satisfies (1).

The normal cone formula for the unit ball in 2A.2(b) extends in an elementary way to the ball B and indicates that w can only be nonzero if $|\bar{x}-\hat{x}| = \rho$ and $w = \lambda[\bar{x}-\hat{x}]$ for some $\lambda > 0$. The normality relation in (7), requiring $0 \ge \langle -f(\bar{x}) - w, x - \bar{x}\rangle$ for all $x \in C$, can be invoked then in the case of $x = \hat{x}$ to obtain $0 \ge \langle -f(\bar{x}) - \lambda[\bar{x}-\hat{x}], \hat{x}-\bar{x}\rangle$, which simplifies to $\langle f(\bar{x}), \bar{x}-\hat{x}\rangle \le -\lambda\rho^2$. But this is impossible under (6). □

The assumption in (6) is fulfilled trivially when C is bounded, and in that way Theorem 2A.1 is seen to be covered by Theorem 2F.2.

Corollary 2F.3 (uniform local existence)**.** *Consider a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a nonempty closed convex set $C \subset \operatorname{dom} f$ relative to which f is continuous (but not necessarily monotone). Suppose there exist $\hat{x} \in C$, $\rho > 0$ and $\eta > 0$ such that*

$$(8) \qquad \textit{there is no } x \in C \textit{ with } |x-\hat{x}| \ge \rho \textit{ and } \langle f(x), x-\hat{x}\rangle \big/ |x-\hat{x}| \le \eta.$$

Then the solution mapping S in (4) has the property that

$$\emptyset \ne S(v) \subset \{ x \in C \mid |x-\hat{x}| < \rho \} \quad \textit{when} \quad |v| \le \eta.$$

Proof. The stronger assumption here ensures that assumption (6) of Theorem 2F.2 is fulfilled by the function $f_v(x) = f(x) - v$ for every v with $|v| \le \eta$. Since $S(v) = (f_v + N_C)^{-1}(0)$, this leads to the desired conclusion. □

We can proceed now to take advantage of monotonicity of f on C through the property in 2F.1 and the observation that

$$\langle f(x), x-\hat{x}\rangle / |x-\hat{x}| = \langle f(\hat{x}+\tau w), w\rangle \quad \text{when } x = \hat{x}+\tau w \text{ with } \tau > 0,\ |w| = 1.$$

Then, for every vector w such that $\hat{x}+\tau w \in C$ for all $\tau \in (0,\infty)$, the expression $\langle f(\hat{x}+\tau w), w\rangle$ is nondecreasing as a function of $\tau \in (0,\infty)$ and thus has a limit (possibly ∞) as $\tau \to \infty$.

Theorem 2F.4 (solution existence for monotone variational inequalities)**.** *Consider a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a nonempty closed convex set $C \subset \operatorname{dom} f$ relative to* *which f is continuous and monotone. Let $\hat{x} \in C$ and let W consist of the vectors w with $|w| = 1$ such that $\hat{x} + \tau w \in C$ for all $\tau \in (0, \infty)$, if any.*

(a) *If $\lim_{\tau \to \infty} \langle f(\hat{x} + \tau w), w \rangle > 0$ for every $w \in W$, then the solution mapping S in (4) is nonempty-valued on a neighborhood of 0.*

(b) *If $\lim_{\tau \to \infty} \langle f(\hat{x} + \tau w), w \rangle = \infty$ for every $w \in W$, then the solution mapping S in (4) is nonempty-valued on all of $\mathbb{R}^n$.*

Proof. To establish (a), we aim at showing that the limit criterion it proposes is enough to guarantee the condition (8) in Corollary 2F.3. Suppose the latter didn't hold. Then there would be a sequence of points $x_k \in C$ and a sequence of scalars $\eta_k > 0$ such that

$$\langle f(x_k), x_k - \hat{x} \rangle \big/ |x_k - \hat{x}| \leq \eta_k \quad \text{with } |x_k - \hat{x}| \to \infty, \ \eta_k \to 0.$$

Equivalently, in terms of $\tau_k = |x_k - \hat{x}|$ and $w_k = \tau_k^{-1}(x_k - \hat{x})$ we have $\langle f(\hat{x} + \tau_k w_k), w_k \rangle \leq \eta_k$ with $|w_k| = 1$, $\hat{x} + \tau_k w_k \in C$ and $\tau_k \to \infty$. Without loss of generality we can suppose that $w_k \to w$ for a vector w again having $|w| = 1$. Then for every $\tau > 0$ and k high enough that $\tau_k \geq \tau$, we have from the convexity of C that $\hat{x} + \tau w_k \in C$ and from the monotonicity of f that $\langle f(\hat{x} + \tau w_k), w_k \rangle \leq \eta_k$. On taking the limit as $k \to \infty$ and utilizing the closedness of C and the continuity of f, we get $\hat{x} + \tau w \in C$ and $\langle f(\hat{x} + \tau w), w \rangle \leq 0$. This being true for any $\tau > 0$, we see that $w \in W$ and the limit condition in (a) is violated. The validity of the claim in (a) is thereby confirmed.

The condition in (b) not only implies the condition in (a) but also, by a slight extension of the argument, guarantees that the criterion in Corollary 2F.3 holds for every $\eta > 0$. □

Exercise 2F.5 (Jacobian criterion for existence and uniqueness). *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ and $C \subset \mathbb{R}^n$ be such that f is continuously differentiable on C and monotone relative to C. Fix $\hat{x} \in C$ and let W consist of the vectors w with $|w| = 1$ such that $\hat{x} + \tau w \in C$ for all $\tau \in (0, \infty)$. Suppose there exists $\mu > 0$ such that $\langle \nabla f(x) w, w \rangle \geq \mu$ for every $w \in W$ and $x \in C$, if any, when $x \in C$. Then the solution mapping S in (4) is single-valued on all of $\mathbb{R}^n$.*

Guide. Argue through the mean value theorem as applied to $\varphi(\tau) = \langle f(\hat{x} + \tau w), w \rangle$ that $\varphi(\tau) = \tau \langle \nabla f(\hat{x} + \theta w) w, w \rangle + \langle f(\hat{x}), w \rangle$ for some $\theta \in (0, \tau)$. Work toward applying the criterion in Theorem 2F.4(b). □

In the perspective of 2F.4(b), the result in 2F.5 seems to come close to invoking strong monotonicity of f in the case where f is continuously differentiable. However, it only involves special vectors w, not every nonzero $w \in \mathbb{R}^n$. For instance, in the affine case where $f(x) = Ax + b$ and $C = \mathbb{R}^n_+$, the criterion obtained from 2F.5 by choosing $\hat{x} = 0$ is simply that $\langle Aw, w \rangle > 0$ for every $w \in \mathbb{R}^n_+$ with $|w| = 1$, whereas strong monotonicity of f would require this for w in $\mathbb{R}^n$, not just $\mathbb{R}^n_+$. In fact, full strong monotonicity has bigger implications than those in 2F.5.

Theorem 2F.6 (variational inequalities with strong monotonicity)**.** *For a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a nonempty closed convex set $C \subset \operatorname{dom} f$, suppose that f is continuous relative to C and strongly monotone on C with constant $\mu > 0$ in the sense of (3). Then the solution mapping S in (4) is single-valued on all of $\mathbb{R}^n$ and moreover Lipschitz continuous with constant μ^{-1}.*

Proof. The strong monotonicity condition in (3) implies for an arbitrary choice of $\hat{x} \in C$ and $w \in \mathbb{R}^n$ with $|w| = 1$ that $\langle f(\hat{x}+\tau w) - f(\hat{x}), \tau w\rangle \geq \mu\tau^2$ when $\hat{x}+\tau w \in C$. Then $\langle f(\hat{x}+\tau w), w\rangle \geq \langle f(\hat{x}), w\rangle + \mu\tau$, from which it's clear that the limit criterion in Theorem 2F.4(b) is satisfied, so that S is nonempty-valued and hence, by the strict monotonicity, single-valued on all of $\mathbb{R}^n$.

Consider now any two vectors v_0 and v_1 in $\mathbb{R}^n$ and the corresponding solutions $x_0 = S(v_0)$ and $x_1 = S(v_1)$. We have $v_0 - f(x_0) \in N_C(x_0)$ and $v_1 - f(x_1) \in N_C(x_1)$, hence in particular $\langle v_0 - f(x_0), x_1 - x_0\rangle \leq 0$ and $\langle v_1 - f(x_1), x_0 - x_1\rangle \leq 0$. The second of these inequalities can also be written as $0 \leq \langle v_1 - f(x_1), x_1 - x_0\rangle$, and from this we see that $\langle v_0 - f(x_0), x_1 - x_0\rangle \leq \langle v_1 - f(x_1), x_1 - x_0\rangle$, which is equivalent to

$$\langle f(x_1) - f(x_0), x_1 - x_0\rangle \leq \langle v_1 - v_0, x_1 - x_0\rangle.$$

Since $\langle f(x_1) - f(x_0), x_1 - x_0\rangle \geq \mu|x_1 - x_0|^2$ by our assumption of strong monotonicity, while $\langle v_1 - v_0, x_1 - x_0\rangle \leq |v_1 - v_0||x_1 - x_0|$, it follows that $|x_1 - x_0| \leq \mu^{-1}|v_1 - v_0|$. This verifies the claimed Lipschitz continuity with constant μ^{-1}. □

We extend our investigations now to the more broadly parameterized variational inequalities of the form

$$f(p,x) + N_C(x) \ni 0 \tag{9}$$

and their solution mappings

$$S(p) = \{\, x \mid f(p,x) + N_C(x) \ni 0 \,\}, \tag{10}$$

with the aim of drawing on the achievements in Section 2E in the presence of monotonicity properties of f with respect to x.

Theorem 2F.7 (strong monotonicity and strict differentiability)**.** *For a variational inequality (9) and its solution mapping (10) in the case of a nonempty, closed, convex set C, let $\bar{x} \in S(\bar{p})$ and assume that f is strictly differentiable at $(\bar{p},\bar{x})$. Suppose for some $\mu > 0$ that*

$$\langle \nabla_x f(\bar{p},\bar{x})w, w\rangle \geq \mu|w|^2 \;\text{ for all }\; w \in C - C. \tag{11}$$

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(s;\bar{p}) \leq \mu^{-1}|\nabla_p f(\bar{p},\bar{x})|. \tag{12}$$

Proof. We apply Theorem 2E.1, observing that its assumption (b) is satisfied on the basis of Theorem 2F.6 and the criterion in 1H.2(f) for strong monotonicity with constant μ. Theorem 2F.6 tells us moreover that the Lipschitz constant for the localization σ in Theorem 2E.1 is no more than μ^{-1}, and we then obtain (12) from Theorem 2E.1. □

Theorem 2F.7 can be compared to the result in Theorem 2E.8. That result requires C to be polyhedral but allows (11) to be replaced by a weaker condition in terms of the critical cone $K = K_C(\bar{x},\bar{v})$ for $\bar{v} = -f(\bar{p},\bar{x})$. Specifically, instead of asking the inequality in (11) to hold for all $w \in C - C$, one only asks it to hold for all $w \in K - K$ such that $\nabla_x f(\bar{p},\bar{x})w \perp K \cap (-K)$. The polyhedral convexity leads in this case to the further conclusion that the localization is semidifferentiable.

2G. Consequences for Optimization

Several types of variational inequalities are closely connected with problems of optimization. These include the basic condition for minimization in Theorem 2A.7 and the Lagrange condition in Theorem 2A.10, in particular. In this section we investigate what the general results obtained for variational inequalities provide in such cases.

Recall from Theorem 2A.7 that in minimizing a continuously differentiable function g over a nonempty, closed, convex set $C \subset \mathbb{R}^n$, the variational inequality

$$(1) \qquad \nabla g(x) + N_C(x) \ni 0$$

stands as a necessary condition for x to furnish a local minimum. When g is convex relative to C, it is sufficient for x to furnish a global minimum, but in the absence of convexity, an x satisfying (1) might not even correspond to a local minimum. However, there is an important case beyond convexity, which we will draw on later, in which an x satisfying (1) can be identified through additional criteria as yielding a local minimum.

In elucidating this case, we will appeal to the fact noted in 2A.4 that the normal cone $N_C(x)$ and the tangent cone $T_C(x)$ are polar to each other, so that (1) can be written equivalently in the form

$$(2) \qquad \langle \nabla g(x), w \rangle \geq 0 \quad \text{for all } w \in T_C(x).$$

This gives a way to think about the first-order condition for a local minimum of g in which the vectors w in (2) making the inequality hold as an equation can be anticipated to have a special role. In fact, those vectors w comprise the critical cone $K_C(x, -\nabla g(x))$ to C at x with respect to the vector $-\nabla g(x)$ in $N_C(x)$, as defined in Section 2E:

$$(3) \qquad K_C(x, -\nabla g(x)) = \left\{ w \in T_C(x) \,\middle|\, \langle \nabla g(x), w \rangle = 0 \right\}.$$

When C is polyhedral, at least, this critical cone is able to serve in the expression of second-order necessary and sufficient conditions for the minimization of g over C.

Theorem 2G.1 (second-order optimality on a polyhedral convex set)**.** *Let C be a polyhedral convex set in $\mathbb{R}^n$ and let $g : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable on C. Let $\bar{x} \in C$ and $\bar{v} = -\nabla g(\bar{x})$.*

(a) (necessary condition) *If g has a local minimum with respect to C at $\bar{x}$, then $\bar{x}$ satisfies the variational inequality (1) and has $\langle w, \nabla^2 g(\bar{x})w\rangle \geq 0$ for all $w \in K_C(\bar{x}, \bar{v})$.*

(b) (sufficient condition) *If $\bar{x}$ satisfies the variational inequality (1) and has $\langle w, \nabla^2 g(\bar{x})w\rangle > 0$ for all nonzero $w \in K_C(\bar{x}, \bar{v})$, then g has a local minimum relative to C at $\bar{x}$, indeed a strong local minimum in the sense of there being an $\varepsilon > 0$ such that*

$$g(x) \geq g(\bar{x}) + \varepsilon|x - \bar{x}|^2 \quad \text{for all } x \in C \text{ near } \bar{x}. \tag{4}$$

Proof. The necessity emerges through the observation that for any $x \in C$ the function $\varphi(t) = g(\bar{x} + tw)$ for $w = x - \bar{x}$ has $\varphi'(0) = \langle \nabla g(\bar{x}), w\rangle$ and $\varphi''(0) = \langle w, \nabla^2 g(\bar{x})w\rangle$. From one-dimensional calculus, it is known that if φ has a local minimum at 0 relative to $[0,1]$, then $\varphi'(0) \geq 0$ and, in the case of $\varphi'(0) = 0$, also has $\varphi''(0) \geq 0$. Having $\varphi'(0) = 0$ corresponds to having $w \in K_C(\bar{x}, -\nabla g(\bar{x}))$.

Conversely, if $\varphi'(0) \geq 0$ and in the case of $\varphi'(0) = 0$ also has $\varphi''(0) > 0$, then φ has a local minimum relative to $[0,1]$ at 0. That one-dimensional sufficient condition is inadequate for concluding (b), however, because (b) requires a neighborhood of $\bar{x}$ relative to C, not just a separate neighborhood relative to each line segment proceeding from x into C.

To get (b), we have to make use of the properties of the second-order Taylor expansion of g at $\bar{x}$ which are associated with g being twice differentiable there: the error expression

$$e(w) = g(\bar{x} + w) - g(\bar{x}) - \langle \nabla g(\bar{x}), w\rangle - \tfrac{1}{2}\langle w, \nabla^2 g(\bar{x})w\rangle$$

is of type $o(|w|^2)$. It will help to translate this into the notation where $w = tz$ with $t \geq 0$ and $|z| = 1$. Let $Z = \{ z \mid |z| = 1 \}$. To say that $e(w)$ is of type $o(|w|^2)$ is to say that the functions

$$f_t(z) = e(tz)/t^2 \text{ for } t > 0$$

converge to 0 on Z uniformly as $t \to 0$.

We furthermore need to rely on the tangent cone property at the end of 2E.3, which is available because C is polyhedral: there is a neighborhood W of the origin in $\mathbb{R}^n$ such that, as long as $w \in W$, we have $\bar{x} + w \in C$ if and only if $w \in T_C(\bar{x})$. Through this, it will be enough to show, on the basis of the assumption in (b), that the inequality in (4) holds for $w \in T_C(\bar{x})$ when $|w|$ is sufficiently small. Equivalently, it will be enough to produce an $\varepsilon > 0$ for which

$$t^{-2}[g(\bar{x} + tz) - g(\bar{x})] \geq \varepsilon \quad \text{for all } z \in Z \cap T_C(\bar{x}) \text{ when } t \text{ is sufficiently small.} \tag{5}$$

The assumption in (b) entails (in terms of $w = tz$) the existence of $\varepsilon > 0$ such that $\langle z, \nabla^2 g(\bar{x})z\rangle > 2\varepsilon$ when $z \in Z \cap K_C(\bar{x}, \bar{v})$. This inequality also holds then for all z in some open set containing $Z \cap K_C(\bar{x}, \bar{v})$. Let Z_0 be the intersection of the complement of that open set with $Z \cap T_C(\bar{x})$. Since (1) is equivalent to (2), we have for $z \in Z_0$ that $\langle \nabla g(\bar{x}), z\rangle > 0$. Because Z_0 is compact, we actually have an $\eta > 0$ such that $\langle \nabla g(\bar{x}), z\rangle > \eta$ for all $z \in Z_0$. We see then, in writing

$$t^{-2}[g(\bar{x}+tz) - g(\bar{x})] = f_t(z) + t^{-1}\langle \nabla g(\bar{x}), z\rangle + \tfrac{1}{2}\langle z, \nabla^2 g(\bar{x})z\rangle$$

and referring to the uniform convergence of the functions f_t to 0 on Z as $t \to 0$, that for t sufficiently small the left side is at least $t^{-1}\eta$ when $z \in Z_0$ and at least 2ε when $z \in Z \cap T_C(\bar{x})$ but $z \notin Z_0$. By taking t small enough that $t^{-1}\eta > 2\varepsilon$, we get (5) as desired. □

When $\bar{x}$ belongs to the interior of C, as for instance when $C = \mathbb{R}^n$ (an extreme case of a polyhedral convex set), the first-order condition in (1) is simply $\nabla g(\bar{x}) = 0$. The second-order conditions in 2G.1 then specify positive semidefiniteness of $\nabla^2 g(\bar{x})$ for necessity and positive definiteness for sufficiency. Second-order conditions for a minimum can also be developed for convex sets that aren't polyhedral, but not in such a simple form. When the boundary of C is "curved," the tangent cone property in 2E.3 fails, and the critical cone $K_C(\bar{x}, \bar{v})$ for $\bar{v} = -\nabla g(\bar{x})$ no longer captures the local geometry adequately. Second-order optimality with respect to nonpolyhedral and even nonconvex sets specified by constraints as in nonlinear programming will be addressed later in this section (Theorem 2G.6).

Stationary points. *An x satisfying (1), or equivalently (2), will be called a stationary point of g with respect to minimizing over C, regardless of whether or not it furnishes a local or global minimum.*

Stationary points attract attention for their own sake, due to the role they have in the design and analysis of minimization algorithms, for example. Our immediate plan is to study how stationary points, as "quasi-solutions" in problems of minimization over a convex set, behave under perturbations. Along with that, we will clarify circumstances in which a stationary point giving a local minimum continues to give a local minimum when perturbed by not too much.

Moving in that direction, we look now at parameterized problems of the form

$$\text{(6)} \qquad \text{minimize } g(p,x) \text{ over all } x \in C,$$

where $g : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable with respect to x (not necessarily convex), and C is a nonempty, closed, convex subset of $\mathbb{R}^n$. In the pattern already seen, the variational inequality

$$\text{(7)} \qquad \nabla_x g(p,x) + N_C(x) \ni 0$$

provides for each p a first-order condition which x must satisfy if it furnishes a local minimum, but only describes, in general, the stationary points x for the minimization in (6). If C is polyhedral the question of a local minimum can be addressed through the second-order conditions provided by Theorem 2G.1 relative to the critical cone

$$(8)\qquad K_C(x,-\nabla_x g(p,x)) = \left\{ w \in T_C(x) \,\middle|\, w \perp \nabla_x g(p,x) \right\}.$$

The basic object of interest to us for now, however, is the *stationary point mapping* $S : I\!R^d \rightrightarrows I\!R^n$ defined by

$$(9)\qquad S(p) = \left\{ x \,\middle|\, \nabla_x g(p,x) + N_C(x) \ni 0 \right\}.$$

With respect to a choice of $\bar{p}$ and $\bar{x}$ such that $\bar{x} \in S(\bar{p})$, it will be useful to consider alongside of (6) an auxiliary problem with parameter $v \in I\!R^n$ in which $g(\bar{p},\cdot)$ is essentially replaced by its second-order expansion at $\bar{x}$:

$$(10)\qquad \begin{array}{l} \text{minimize } \bar{g}(w) - \langle v, w\rangle \text{ over all } w \in W, \\ \quad \text{where } \begin{cases} \bar{g}(w) = g(\bar{p},\bar{x}) + \langle \nabla_x g(\bar{p},\bar{x}), w\rangle + \frac{1}{2}\langle w, \nabla^2_{xx} g(\bar{p},\bar{x}) w\rangle, \\ W = \left\{ w \,\middle|\, \bar{x} + w \in C \right\} = C - \bar{x}. \end{cases} \end{array}$$

The subtraction of $\langle v, w\rangle$ "tilts" $\bar{g}$, and is referred to therefore as a *tilt perturbation*. When $v = 0$, $\bar{g}$ itself is minimized.

For this auxiliary problem the basic first-order condition comes out to be the parameterized variational inequality

$$(11)\qquad \nabla_x g(\bar{p},\bar{x}) + \nabla^2_{xx} g(\bar{p},\bar{x}) w - v + N_W(w) \ni 0, \quad \text{where } N_W(w) = N_C(\bar{x}+w).$$

The stationary point mapping for the problem in (10) is accordingly the mapping $\bar{S} : I\!R^n \rightrightarrows I\!R^n$ defined by

$$(12)\qquad \bar{S}(v) = \left\{ w \,\middle|\, \nabla_x g(\bar{p},\bar{x}) + \nabla^2_{xx} g(\bar{p},\bar{x}) w + N_W(w) \ni v \right\}.$$

The points $w \in \bar{S}(v)$ are sure to furnish a minimum in (10) if, for instance, the matrix $\nabla^2_x g(\bar{p},\bar{x})$ is positive semidefinite, since that corresponds to the convexity of the "tilted" function being minimized. For polyhedral C, Theorem 2G.1 could be brought in for further analysis of a local minimum in (10). Note that $0 \in \bar{S}(0)$.

Theorem 2G.2 (parametric minimization over a convex set)**.** *Suppose in the preceding notation, with $\bar{x} \in S(\bar{p})$, that*

(a) $\nabla_x g$ *is strictly differentiable at* $(\bar{p},\bar{x})$*;*

(b) $\bar{S}$ *has a Lipschitz continuous single-valued localization* $\bar{s}$ *around* 0 *for* 0*.*

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(s;\bar{p}) \leq \operatorname{lip}(\bar{s};0) \cdot |\nabla^2_{xp} g(\bar{p},\bar{x})|,$$

and s has a first-order approximation η at $\bar{p}$ given by

$$(13)\qquad \eta(p) = \bar{x} + \bar{s}\Big(-\nabla^2_{xp} g(\bar{p},\bar{x})(p-\bar{p}) \Big).$$

On the other hand, (b) is necessary for S to have a Lipschitz continuous single-valued localization around $\bar p$ for $\bar x$ when the $n\times d$ matrix $\nabla^2_{xp}g(\bar p,\bar x)$ has rank n.

Under the additional assumption that C is polyhedral, condition (b) is equivalent to the condition that, for the critical cone $K = K_C(\bar x,-\nabla_x g(\bar p,\bar x))$, the mapping

$$v\mapsto \bar S_0(v) = \left\{ w \,\middle|\, \nabla^2_{xx}g(\bar p,\bar x)w + N_K(w)\ni v\right\} \quad \text{is everywhere single-valued.}$$

Moreover, a sufficient condition for this can be expressed in terms of the critical subspaces $K^+_C(\bar x,\bar v) = K_C(\bar x,\bar v) - K_C(\bar x,\bar v)$ and $K^-_C(\bar x,\bar v) = K_C(\bar x,\bar v)\cap[-K_C(\bar x,\bar v)]$ for $\bar v = -\nabla_x g(\bar p,\bar x)$, namely

$$(14)\qquad \langle w,\nabla^2_{xx}g(\bar p,\bar x)w\rangle > 0 \quad \begin{cases}\text{for every nonzero } w\in K^+_C(\bar x,\bar v)\\ \text{with } \nabla^2_{xx}g(\bar p,\bar x)w\perp K^-_C(\bar x,\bar v).\end{cases}$$

Furthermore, in this case the localization s is semidifferentiable at $\bar p$ with semiderivative given by

$$Ds(\bar p)(q) = \bar s(-\nabla^2_{xp}g(\bar p,\bar x)q).$$

Proof. We apply Theorem 2E.1 with $f(p,x) = \nabla_x g(p,x)$. The mapping G in that result coincides with $\nabla\bar g + N_C$, so that G^{-1} is $\bar S$. Assumptions (a) and (b) cover the corresponding assumptions in 2E.1, with $\sigma(v) = \bar x + \bar s(v)$, and then (13) follows from 2E(4). In the polyhedral case we also have Theorems 2E.6 and 2E.8 at our disposal, and this gives the rest. $\square$

Theorem 2G.3 (stability of a local minimum on a polyhedral convex set)**.** *Suppose in the setting of the parameterized minimization problem in (6) and its stationary point mapping S in (9) that C is polyhedral and $\nabla_x g(p,x)$ is strictly differentiable with respect to (p,x) at $(\bar p,\bar x)$, where $\bar x\in S(\bar p)$. With respect to the critical subspace $K^+_C(\bar x,\bar v)$ for $\bar v = -\nabla_x g(\bar p,\bar x)$, assume that*

$$(15)\qquad \langle w,\nabla^2_{xx}g(\bar p,\bar x)w\rangle > 0 \text{ for every nonzero } w\in K^+_C(\bar x,\bar v).$$

Then S has a localization s not only with the properties laid out in Theorem 2G.2, but also with the property that, for every p in some neighborhood of $\bar p$, the point $x = s(p)$ furnishes a strong local minimum in (6). Moreover, (15) is necessary for the existence of a localization s with all these properties, when the $n\times d$ matrix $\nabla^2_{xp}g(\bar p,\bar x)$ has rank n.

Proof. Obviously (15) implies (14), which ensures according to Theorem 2G.2 that S has a Lipschitz continuous single-valued localization s around $\bar p$ for $\bar x$. Applying 2E.10(a), we see then that

$$K_C(x,-\nabla_x g(p,x)) \subset K^+_C(\bar x,-\nabla_x g(\bar p,\bar x)) = K^+_C(\bar x,\bar v)$$

when $x = s(p)$ and p is near enough to $\bar p$. Since the matrix $\nabla^2_{xx}g(p,x)$ converges to $\nabla^2_{xx}g(\bar p,\bar x)$ as (p,x) tends to $(\bar p,\bar x)$, it follows that $\langle w,\nabla^2_{xx}g(p,x)w\rangle > 0$ for all nonzero $w\in K_C(x,-\nabla_x g(p,x))$ when $x = s(p)$ and p is close enough to $\bar p$. Since having $x = s(p)$ corresponds to having the first-order condition in (7), we conclude that from Theorem 2G.1 that x furnishes a strong local minimum in this case.

Arguing now toward the necessity of (15) under the rank condition on $\nabla^2_{xp} g(\bar{p},\bar{x})$, we suppose S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ such that $x = s(p)$ gives a local minimum when p is close enough to $\bar{p}$. For any $x \in C$ near $\bar{x}$ and $v \in N_C(x)$, the rank condition gives us a p such that $v = -\nabla_x g(p,x)$; this follows e.g. from 1F.6. Then $x = s(p)$ and, because we have a local minimum, it follows that $\langle w, \nabla^2_{xx} g(p,x)w\rangle \geq 0$ for every nonzero $w \in K_C(x,v)$. We know from 2E.10(b) that $K_C(x,v) = K_C^+(\bar{x},\bar{v})$ for choices of x and v arbitrarily close to $(\bar{x},\bar{v})$, where $\bar{v} = -\nabla_x g(\bar{p},\bar{x})$. Through the continuous dependence of $\nabla^2_{xx} g(p,x)$ on (p,x), we therefore have

$$(16) \qquad \langle w, Aw\rangle \geq 0 \;\text{ for all } w \in K_C^+(\bar{x},\bar{v}), \;\text{ where } A = \nabla^2_{xx} g(\bar{p},\bar{x}) \text{ is symmetric.}$$

For this reason, we can only have $\langle w, Aw\rangle = 0$ if $Aw \perp K_C^+(\bar{x},\bar{v})$, i.e., $\langle w', Aw\rangle = 0$ for all $w' \in K_C^+(\bar{x},\bar{v})$.

On the other hand, because the rank condition corresponds to the ample parameterization property, we know from Theorem 2E.8 that the existence of the single-valued localization s requires for A and the critical cone $K = K_C(\bar{x},\bar{v})$ that the mapping $(A + N_K)^{-1}$ be single-valued. This would be impossible if there were a nonzero w such that $Aw \perp K_C^+(\bar{x},\bar{v})$, because we would have $\langle w', Aw\rangle = 0$ for all $w' \in K$ in particular (since $K \subset K_C^+(\bar{x},\bar{v})$), implying that $-Aw \in N_K(w)$. Then $(A + N_K)^{-1}(0)$ would contain w along with 0, contrary to single-valuedness. Thus, the inequality in (16) must be strict when $w \neq 0$. □

Next we provide a complementary, global result for the special case of a tilted strongly convex function.

Proposition 2G.4 (tilted minimization of strongly convex functions). *Let $g : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on an open set O, and let $C \subset O$ be a nonempty, closed, convex set on which g is strongly convex with constant $\mu > 0$. Then for each $v \in \mathbb{R}^n$ the problem*

$$(17) \qquad \text{minimize } g(x) - \langle v, x\rangle \text{ over } x \in C$$

has a unique solution $s(v)$, and the solution mapping s is a Lipschitz continuous function on $\mathbb{R}^n$ (globally) with Lipschitz constant μ^{-1}.

Proof. Let g_v denote the function being minimized in (17). Like g, this function is continuously differentiable and strongly convex on C with constant μ; we have $\nabla g_v(x) = \nabla g(x) - v$. According to Theorem 2A.7, the condition $\nabla g_v(x) + N_C(x) \ni 0$, or equivalently $x \in (\nabla g + N_C)^{-1}(v)$, is both necessary and sufficient for x to furnish the minimum in (17). The strong convexity of g makes the mapping $f = \nabla g$ strongly monotone on C with constant μ; see 2A.7(a). The conclusion follows now by applying Theorem 2F.6 to this mapping f. □

When the function g in Proposition 2G.4 is twice continuously differentiable, the strong monotonicity can be identified through 2A.5(b) with the inequality $\langle\nabla^2 g(x)w,w\rangle \geq \mu|w|^2$ holding for all $x\in C$ and $w\in C-C$.

Exercise 2G.5. *In the setting of Theorem 2G.2, condition (b) is fulfilled in particular if there exists $\mu>0$ such that*

$$\langle\nabla^2_{xx} g(\bar{p},\bar{x})w,w\rangle \geq \mu|w|^2 \quad \text{for all } w\in C-C, \tag{18}$$

and then $\operatorname{lip}(s;\bar{p})\leq\mu^{-1}$. If C is polyhedral, the additional conclusion holds that, for all p in some neighborhood of $\bar{p}$, there is a strong local minimum in problem (6) at the point $x=s(p)$.

Guide. Apply Proposition 2G.4 to the function $\bar{g}$ in the auxiliary minimization problem (10). Get from this that $\bar{s}$ coincides with $\bar{S}$, which is single-valued and Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant μ^{-1}. In the polyhedral case, also apply Theorem 2G.3, arguing that (18) entails (15). □

Observe that because $C-C$ is a convex set containing 0, the condition in (18) holds for all $w\in C-C$ if it holds for all $w\in C-C$ with $|w|$ sufficiently small.

We turn now to minimization over sets which need not be convex but are specified by a system of constraints. A first-order necessary condition for a minimum in that case was developed in a very general manner in Theorem 2A.9. Here, we restrict ourselves to the most commonly treated problem of nonlinear programming, where the format is to

$$\text{minimize } g_0(x) \text{ over all } x \text{ satisfying } g_i(x)\begin{cases}\leq 0 & \text{for } i\in[1,s],\\ =0 & \text{for } i\in[s+1,m].\end{cases} \tag{19}$$

In order to bring second-order conditions for optimality into the picture, we assume that the functions $g_0,g_1,\ldots,g_m$ are *twice* continuously differentiable on $\mathbb{R}^n$.

The basic first-order condition in this case has been worked out in detail in Section 2A as a consequence of Theorem 2A.9. It concerns the existence, relative to x, of a multiplier vector $y=(y_1,\ldots,y_m)$ fulfilling the Karush–Kuhn–Tucker conditions:

$$\begin{aligned} &y\in\mathbb{R}^s_+\times\mathbb{R}^{m-s},\ g_i(x)\begin{cases}\leq 0 & \text{for } i\in[1,s] \text{ with } y_i=0,\\ =0 & \text{for all other } i\in[1,m],\end{cases}\\ &\nabla g_0(x)+y_1\nabla g_1(x)+\cdots+y_m\nabla g_m(x)=0.\end{aligned} \tag{20}$$

This existence is necessary for a local minimum at x as long as x satisfies the constraint qualification requiring that the same conditions, but with the term $\nabla g_0(x)$ suppressed, can't be satisfied with $y\neq 0$. It is sufficient for a global minimum at x if $g_0,g_1,\ldots,g_s$ are convex and $g_{s+1},\ldots,g_m$ are affine. However, we now wish to take a second-order approach to local sufficiency, rather than rely on convexity for global sufficiency.

The key for us will be the fact, coming from Theorem 2A.10, that (20) can be identified in terms of the Lagrangian function

(21) $$L(x,y) = g_0(x) + y_1 g_1(x) + \cdots + y_m g_m(x)$$

with a certain variational inequality for a continuously differentiable function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ and polyhedral convex cone $E \subset \mathbb{R}^n \times \mathbb{R}^m$, namely

(22) $$f(x,y) + N_E(x,y) \ni (0,0), \text{ where } \begin{cases} f(x,y) = (\nabla_x L(x,y), -\nabla_y L(x,y))^{\mathsf{T}}, \\ E = \mathbb{R}^n \times [\mathbb{R}^s_+ \times \mathbb{R}^{m-s}]. \end{cases}$$

Because our principal goal is to illustrate the application of the results in the preceding sections, rather than push consequences for optimization theory to the limit, we will only deal with this variational inequality under an assumption of linear independence for the gradients of the active constraints. A constraint in (19) is *inactive* at x if it is an inequality constraint with $g_i(x) < 0$; otherwise it is *active* at x.

Theorem 2G.6 (second-order optimality in nonlinear programming)**.** *Let $\bar{x}$ be a point satisfying the constraints in (19). Let $I(\bar{x})$ be the set of indices i of the active constraints at $\bar{x}$, and suppose that the gradients $\nabla g_i(\bar{x})$ for $i \in I(\bar{x})$ are linearly independent. Let K consist of the vectors $w \in \mathbb{R}^n$ satisfying*

(23) $$\langle \nabla g_i(\bar{x}), w \rangle \begin{cases} \leq 0 & \text{for } i \in I(\bar{x}) \text{ with } i \leq s, \\ = 0 & \text{for all other } i \in I(\bar{x}) \text{ and also for } i = 0. \end{cases}$$

(a) (necessary condition) *If $\bar{x}$ furnishes a local minimum in problem (19), then a multiplier vector $\bar{y}$ exists such that $(\bar{x},\bar{y})$ not only satisfies the variational inequality (22) but also has*

(24) $$\langle w, \nabla^2_{xx} L(\bar{x},\bar{y}) w \rangle \geq 0 \text{ for all } w \in K.$$

(b) (sufficient condition) *If a multiplier vector $\bar{y}$ exists such that $(\bar{x},\bar{y})$ satisfies the conditions in (20), or equivalently (22), and if (24) holds with strict inequality when $w \neq 0$, then $\bar{x}$ furnishes a local minimum in (19). Indeed, it furnishes a strong local minimum in the sense of there being an $\varepsilon > 0$ such that*

(25) $$g_0(x) \geq g_0(\bar{x}) + \varepsilon |x - \bar{x}|^2 \text{ for all } x \text{ near } \bar{x} \text{ satisfying the constraints.}$$

Proof. The linear independence of the gradients of the active constraints guarantees, among other things, that $\bar{x}$ satisfies the constraint qualification under which (22) is necessary for local optimality.

In the case of a local minimum, as in (a), we do therefore have the variational inequality (22) fulfilled by $\bar{x}$ and some vector $\bar{y}$; and of course (22) holds by assumption in (b). From this point on, therefore, we can concentrate on just the second-order parts of (a) and (b) in the framework of having $\bar{x}$ and $\bar{y}$ satisfying (20). In particular then, we have

(26) $$-\nabla g_0(\bar{x}) = \bar{y}_1 \nabla g_1(\bar{x}) + \cdots + \bar{y}_m \nabla g_m(\bar{x}),$$

where the multiplier vector $\bar{y}$ is moreover uniquely determined by the linear independence of the gradients of the active constraints and the stipulation in (20) that inactive constraints get coefficient 0.

Actually, the inactive constraints play no role around $\bar{x}$, so we can just as well assume, for simplicity of exposition in our local analysis, that *every* constraint is active at $\bar{x}$: we have $g_i(\bar{x}) = 0$ for $i = 1,\dots,m$. Then, on the level of first-order conditions, we just have the combination of (20), which corresponds to $\nabla_x L(\bar{x},\bar{y}) = 0$, and the requirement that $\bar{y}_i \geq 0$ for $i = 1,\dots,s$. In this simplified context, let

$$\text{(27)} \quad T = \text{ set of all } w \in I\!\!R^n \text{ satisfying } \langle \nabla g_i(\bar{x}), w\rangle \begin{cases} \leq 0 & \text{for } i = 1,\dots,s, \\ = 0 & \text{for } i = s+1,\dots,m, \end{cases}$$

so that the cone K described by (23) can be expressed in the form

$$\text{(28)} \qquad K = \big\{\, w \in T \,\big|\, \langle \nabla g_0(\bar{x}), w\rangle = 0 \big\}.$$

The rest of our argument will rely heavily on the classical inverse function theorem, 1A.1. Our assumption that the vectors $\nabla g_i(\bar{x})$ for $i = 1,\dots,m$ are linearly independent in $I\!\!R^n$ entails of course that $m \leq n$. These vectors can be supplemented, if necessary, by vectors a_k for $k = 1,\dots,n-m$ so as to form a basis for $I\!\!R^n$. Then, by setting $g_{m+k}(x) = \langle a_k, x-\bar{x}\rangle$, we get functions g_i for $i = m+1,\dots,n$ such that for

$$g : I\!\!R^n \to I\!\!R^n \text{ with } g(x) = (g_1(x),\dots,g_m(x),g_{m+1}(x),\dots,g_n(x))$$

we have $g(\bar{x}) = 0$ and $\nabla g(\bar{x})$ nonsingular. We can view this as providing, at least locally around $\bar{x}$, a change of coordinates $g(x) = u = (u_1,\dots,u_n)$, $x = s(u)$ (for a localization s of g^{-1} around 0 for $\bar{x}$) in which $\bar{x}$ corresponds to 0 and the constraints in (19) correspond to linear constraints

$$u_i \leq 0 \text{ for } i = 1,\dots,s, \qquad u_i = 0 \text{ for } i = s+1,\dots,m$$

(with no condition on u_i for $i = m+1,\dots,n$), which specify a *polyhedral convex* set D in $I\!\!R^n$. Problem (19) is thereby transformed in a local sense into minimizing over this set D the twice continuously differentiable function $f(u) = g_0(s(u))$, and we are concerned with whether or not there is a local minimum at $\bar{u} = 0$. The necessary and sufficient conditions in Theorem 2G.1 are applicable to this and entail having $-f(0)$ belong to $N_D(0)$. It will be useful to let $\tilde{y}$ stand for $(\bar{y},0,\dots,0) \in I\!\!R^n$.

The inverse function theorem reveals that the Jacobian $\nabla s(0)$ is $\nabla g(\bar{x})^{-1}$. We have $\nabla f(0) = \nabla g_0(\bar{x})\nabla s(0)$ by the chain rule, and on the other hand $-\nabla g_0(\bar{x}) = \tilde{y}\nabla g(\bar{x})$ by (26), and therefore $\nabla f(0) = -\tilde{y}$. The vectors w belonging to the set T in (27) correspond one-to-one with the vectors $z \in D$ through $\nabla g(\bar{x})w = z$, and under this, through (28), the vectors $w \in K$ correspond to the vectors $z \in D$ such that $\langle z, \tilde{y}\rangle = 0$, i.e., the vectors in the critical cone $K_D(0,\tilde{y}) = K_D(0,-\nabla f(0))$.

The second-order conditions in Theorem 2G.1, in the context of the transformed version of problem (19), thus revolve around the nonnegativity or positivity of $\langle z, \nabla^2 f(0)z\rangle$ for vectors $z \in K_D(0,\tilde{y})$. It will be useful that this is the same as the nonnegativity or positivity or $\langle z, \nabla^2 h(0) z\rangle$ for the function

$$h(u) = f(u) + \langle \tilde{y}, u\rangle = f(u) + \langle \bar{y}, Pu\rangle = L(s(u), \bar{y}),$$

where P is the projection from $(u_1, \ldots, u_m, u_{m+1}, \ldots, u_n)$ to $(u_1, \ldots, u_m)$. Furthermore,

$$\langle z, \nabla^2 h(0) z\rangle = \varphi''(0) \text{ for the function } \varphi(t) = h(tz) = L(s(tz), \bar{y}).$$

Fix any nonzero $z \in K_D(0, \tilde{y})$ and the corresponding $w \in K$, given by $w = \nabla s(0) z = \nabla g(\bar{x})^{-1} z$. Our task is to demonstrate that actually

$$\varphi''(0) = \langle w, \nabla^2_{xx} L(\bar{x}, \bar{y}) w\rangle. \tag{29}$$

Let $x(t) = s(tz)$, so that $x(0) = \bar{x}$ and $x'(0) = w$. We have

$$\begin{aligned}
&\varphi(t) = L(x(t), \bar{y}),\\
&\varphi'(t) = \langle \nabla_x L(x(t), \bar{y}), x'(t)\rangle,\\
&\varphi''(t) = \langle w, \nabla^2_{xx} L(x(t), \bar{y}), x'(t)\rangle + \langle \nabla_x L(x(t), \bar{y}), x''(t)\rangle,
\end{aligned}$$

hence $\varphi''(0) = \langle w, \nabla^2_{xx} L(\bar{x}, \bar{y}), w\rangle + \langle \nabla_x L(\bar{x}, \bar{y}), x''(0)\rangle$. But $\nabla_x L(\bar{x}, \bar{y}) = 0$ from the first-order conditions. Thus, (29) holds, as claimed.

The final assertion of part (b) automatically carries over from the corresponding assertion of part (b) of Theorem 2G.1 under the local change of coordinates that we utilized. □

Exercise 2G.7. *In the context of Theorem 2G.6, let $\bar{y}$ be a multiplier associated with $\bar{x}$ through the first-order condition (22). Let $I_0(\bar{x}, \bar{y})$ be the set of indices $i \in I(\bar{x})$ such that $i \le s$ and $\bar{y}_i = 0$. Then an equivalent description of the cone K in the second-order conditions is that*

$$w \in K \iff \langle \nabla g_i(\bar{x}), w\rangle \begin{cases} \le 0 & \text{for } i \in I_0(\bar{x}, \bar{y}),\\ = 0 & \text{for } i \in I(\bar{x}) \backslash I_0(\bar{x}, \bar{y}). \end{cases}$$

Guide. Utilize the fact that $-\nabla g_0(\bar{x}) = \bar{y}_1 \nabla g_1(\bar{x}) + \cdots + \bar{y}_m \nabla g_m(\bar{x})$ with $\bar{y}_i \ge 0$ for $i = 1, \ldots, s$. □

The alternative description in 2G.7 lends insights in some situations, but it makes K appear to depend on $\bar{y}$, whereas in reality it doesn't.

Next we take up the study of a parameterized version of the nonlinear programming problem in the form

$$\text{minimize } g_0(p, x) \text{ over all } x \text{ satisfying } g_i(p, x) \begin{cases} \le 0 & \text{for } i \in [1, s],\\ = 0 & \text{for } i \in [s+1, m], \end{cases} \tag{30}$$

where the functions $g_0, g_1, \ldots, g_m$ are twice continuously differentiable from $\mathbb{R}^d \times \mathbb{R}^n$ to $\mathbb{R}$. The Lagrangian function is now

$$L(p,x,y) = g_0(p,x) + y_1 g_1(p,x) + \cdots + y_m g_m(p,x) \tag{31}$$

and the variational inequality capturing the associated first-order conditions is

$$f(p,x,y) + N_E(x,y) \ni (0,0), \tag{32}$$

where

$$\begin{cases} f(p,x,y) = (\nabla_x L(p,x,y), -\nabla_y L(p,x,y))^{\mathsf{T}}, \\ E = \mathbb{R}^n \times [\mathbb{R}^s_+ \times \mathbb{R}^{m-s}]. \end{cases}$$

The pairs (x,y) satisfying this variational inequality are the Karush–Kuhn–Tucker pairs for the problem specified by p in (30). The x components of such pairs might or might not give a local minimum according to the circumstances in Theorem 2G.6 (or whether certain convexity assumptions are fulfilled), and indeed we are not imposing a linear independence condition on the constraint gradients in (30) of the kind on which Theorem 2G.6 was based. But these x's serve anyway as stationary points and we wish to learn more about their behavior under perturbations by studying the *Karush–Kuhn–Tucker mapping* $S : \mathbb{R}^d \to \mathbb{R}^n \times \mathbb{R}^m$ defined by

$$S(p) = \left\{ (x,y) \,\middle|\, f(p,x,y) + N_E(x,y) \ni (0,0) \right\}. \tag{33}$$

Once more, an auxiliary problem will be important with respect to a choice of $\bar{p}$ and a pair $(\bar{x},\bar{y}) \in S(\bar{p})$. To formulate it, we let

$$\begin{aligned} \bar{g}_0(w) &= L(\bar{p},\bar{x},\bar{y}) + \langle \nabla_x L(\bar{p},\bar{x},\bar{y}), w\rangle + \tfrac{1}{2}\langle w, \nabla^2_{xx} L(\bar{p},\bar{x},\bar{y}) w\rangle, \\ \bar{g}_i(w) &= g_i(\bar{p},\bar{x}) + \langle \nabla_x g_i(\bar{p},\bar{x}), w\rangle \ \text{ for } i = 1,\ldots,m, \end{aligned}$$

and introduce the notation

$$\begin{aligned} I &= \left\{ i \in [1,m] \,\middle|\, g_i(\bar{p},\bar{x}) = 0 \right\} \supset \{s+1,\ldots,m\}, \\ I_0 &= \left\{ i \in [1,s] \,\middle|\, g_i(\bar{p},\bar{x}) = 0 \text{ and } \bar{y}_i = 0 \right\} \subset I, \\ I_1 &= \left\{ i \in [1,s] \,\middle|\, g_i(\bar{p},\bar{x}) < 0 \right\}. \end{aligned} \tag{34}$$

The auxiliary problem, depending on a tilt parameter vector v but also now an additional parameter vector $u = (u_1,\ldots,u_m)$, is to

$$\begin{aligned} &\text{minimize } \bar{g}_0(w) - \langle v, w\rangle \text{ over all } w \text{ satisfying} \\ &\bar{g}_i(w) + u_i \begin{cases} = 0 & \text{for } i \in I \backslash I_0, \\ \leq 0 & \text{for } i \in I_0, \\ \text{free} & \text{for } i \in I_1, \end{cases} \end{aligned} \tag{35}$$

where "free" means unrestricted. (The functions $\bar{g}_i$ for $i \in I_1$ play no role in this problem, but it will be more convenient to carry them through in this scheme than to drop them.)

In comparison with the auxiliary problem introduced earlier in (10) with respect to minimization over a set C, it's apparent that a second-order expansion of L rather than g_0 has entered, but merely first-order expansions of the constraint functions $g_1,\ldots,g_m$. In fact, only the quadratic part of the Lagrangian expansion matters, inasmuch as $\nabla_x L(\bar{p},\bar{x},\bar{y}) = 0$ by the first-order conditions. The Lagrangian for the problem in (35) depends on the parameter pair (v,u) and involves a multiplier vector $z = (z_1,\dots,z_m)$:

$$\bar{g}_0(w) - \langle v,w\rangle + z_1[\bar{g}_1(w)+u_1] + \cdots + z_m[\bar{g}_m(w)+u_m] =: \bar{L}(w,z) - \langle v,w\rangle + \langle z,u\rangle.$$

The corresponding first-order conditions are given by the variational inequality

$$\begin{aligned} &\bar{f}(w,z) - (v,u) + N_{\bar{E}}(w,z) \ni (0,0), \quad \text{where} \\ &\bar{f}(w,z) = (\nabla_w \bar{L}(w,z), -\nabla_z \bar{L}(w,z)), \quad \bar{E} = \mathbb{R}^n \times W, \text{ with} \\ &z = (z_1,\dots,z_m) \in W \iff z_i \begin{cases} \geq 0 & \text{for } i \in I_0, \\ = 0 & \text{for } i \in I_1, \end{cases} \end{aligned} \tag{36}$$

which translate into the requirements that

$$\begin{aligned} &\nabla^2_{xx} L(\bar{p},\bar{x},\bar{y})w + z_1 \nabla_x g_1(\bar{p},\bar{x}) + \cdots + z_m \nabla_x g_m(\bar{p},\bar{x}) - v = 0, \\ &\text{with } z_i \begin{cases} \geq 0 & \text{for } i \in I_0 \text{ having } \bar{g}_i(w) + u_i = 0, \\ = 0 & \text{for } i \in I_0 \text{ having } \bar{g}_i(w) + u_i < 0 \text{ and for } i \in I_1. \end{cases} \end{aligned} \tag{37}$$

We need to pay heed to the auxiliary solution mapping $\bar{S} : \mathbb{R}^n \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ defined by

$$\bar{S}(v,u) = \{ (w,z) \mid \bar{f}(w,z) + N_{\bar{E}}(w,z) \ni (v,u) \} = \{ (w,z) \mid \text{satisfying (37)} \}, \tag{38}$$

which has

$$(0,0) \in \bar{S}(0,0).$$

The following subspaces will enter our analysis of the properties of the mapping $\bar{S}$:

$$\begin{aligned} M^+ &= \{ w \in \mathbb{R}^n \mid w \perp \nabla_x g_i(\bar{p},\bar{x}) \text{ for all } i \in I \backslash I_0 \}, \\ M^- &= \{ w \in \mathbb{R}^n \mid w \perp \nabla_x g_i(\bar{p},\bar{x}) \text{ for all } i \in I \}. \end{aligned} \tag{39}$$

Theorem 2G.8 (implicit function theorem for stationary points)**.** *Let $(\bar{x},\bar{y}) \in S(\bar{p})$ for the mapping S in (33), constructed from functions g_i that are twice continuously differentiable. Assume for the auxiliary mapping $\bar{S}$ in (38) that*

$$\begin{cases} \bar{S} \text{ has a Lipschitz continuous single-valued} \\ \text{localization } \bar{s} \text{ around } (0,0) \text{ for } (0,0). \end{cases} \tag{40}$$

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $(\bar{x},\bar{y})$, and this localization s is semidifferentiable at $\bar{p}$ with semiderivative given by

(41) $$Ds(\bar{p})(q) = \bar{s}(-Bq), \text{ where } B = \nabla_p f(\bar{p},\bar{x},\bar{y}) = \begin{pmatrix} \nabla^2_{xp} L(\bar{p},\bar{x},\bar{y}) \\ -\nabla_p g_1(\bar{p},\bar{x}) \\ \vdots \\ -\nabla_p g_m(\bar{p},\bar{x}) \end{pmatrix}.$$

Moreover the condition in (40) is necessary for the existence of a Lipschitz continuous single-valued localization of S around $\bar{p}$ for $(\bar{x},\bar{y})$ when the $(n+m)\times d$ matrix B has rank $n+m$. In particular, $\bar{S}$ is sure to have the property in (40) when the following conditions are both fulfilled:

(a) *the gradients $\nabla_x g_i(\bar{p},\bar{x})$ for $i \in I$ are linearly independent;*

(b) *$\langle w, \nabla^2_{xx} L(\bar{p},\bar{x},\bar{y})w\rangle > 0$ for every nonzero $w \in M^+$ with $\nabla^2_{xx} L(\bar{p},\bar{x},\bar{y})w \perp M^-$, with M^+ and M^- as in (39).*

On the other hand, (40) always entails at least (a).

Proof. This is obtained by applying 2E.1 with the additions in 2E.6 and 2E.8 to the variational inequality (32). Since

$$\nabla_y L(p,x,y) = g(p,x) \quad \text{for} \quad g(p,x) = (g_1(p,x),\dots,g_m(p,x)),$$

the Jacobian in question is

(42) $$\nabla_{(x,y)} f(\bar{p},\bar{x},\bar{y}) = \begin{pmatrix} \nabla^2_{xx} L(\bar{p},\bar{x},\bar{y}) & \nabla_x g(\bar{p},\bar{x})^{\mathsf{T}} \\ -\nabla_x g(\bar{p},\bar{x}) & 0 \end{pmatrix}.$$

In terms of polyhedral convex cone $Y = \mathbb{R}^s_+ \times \mathbb{R}^{m-s}$, the critical cone to the polyhedral convex cone set E is

(43) $$K_E(\bar{x},\bar{y},-f(\bar{p},\bar{x},\bar{y})) = \mathbb{R}^n \times W$$

for the polyhedral cone W in (36). By taking A to be the matrix in (42) and K to be the cone in (43), the auxiliary mapping $\bar{S}$ can be identified with $(A+N_K)^{-1}$ in the framework of Theorem 2E.8. (The w and u in that result have here turned into pairs (w,z) and (v,u).)

This leads to all the conclusions except for establishing that (40) implies (a) and working out the details of the sufficient condition provided by 2E.8. To verify that (40) implies (a), consider any $\varepsilon > 0$ and let

$$v^\varepsilon = \sum_{i=1}^m z_i^\varepsilon \nabla_x g_i(\bar{p},\bar{x}), \qquad z_i^\varepsilon = \begin{cases} \varepsilon & \text{for } i \in I_0, \\ 0 & \text{otherwise.} \end{cases}$$

Then, as seen from the conditions in (37), we have $(0,z^\varepsilon) \in \bar{S}(v^\varepsilon,0)$. If (a) didn't hold, we would also have $\sum_{i=1}^m \zeta_i \nabla_x g(\bar{p},\bar{x}) = 0$ for some coefficient vector $\zeta \neq 0$ with $\zeta_i = 0$ when $i \in I_1$. Then for every $\delta > 0$ small enough that $\varepsilon + \delta\zeta_i \geq 0$ for all $i \in I_0$, we would also have $(0, z^\varepsilon + \delta\zeta) \in \bar{S}(v^\varepsilon,0)$. Since ε and δ can be chosen arbitrarily small, this would contradict the single-valuedness in (40). Thus, (a) is necessary for (40).

To come next to an understanding of what the sufficient condition in 2E.8 means here, we observe in terms of $Y = \mathbb{R}^s_+ \times \mathbb{R}^{m-s}$ that

$$\begin{aligned} K_E^+(\bar{x},\bar{y},-f(\bar{p},\bar{x},\bar{y})) &= \mathbb{R}^n \times K_Y^+(\bar{y},g(\bar{p},\bar{x})),\\ K_E^-(\bar{x},\bar{y},-f(\bar{p},\bar{x},\bar{y})) &= \mathbb{R}^n \times K_Y^-(\bar{y},g(\bar{p},\bar{x})),\end{aligned}$$

where

$$\begin{aligned} z \in K_Y^+(\bar{y},g(\bar{p},\bar{x})) &\iff z_i = 0 \text{ for } i \in I_1,\\ z \in K_Y^-(\bar{y},g(\bar{p},\bar{x})) &\iff z_i = 0 \text{ for } i \in I_0 \cup I_1.\end{aligned}$$

In the shorthand notation

$$H = \nabla^2_{xx}L(\bar{p},\bar{x},\bar{y}), \quad K^+ = K_E^+(\bar{x},\bar{y},-f(\bar{p},\bar{x},\bar{y})), \quad K^- = K_E^-(\bar{x},\bar{y},-f(\bar{p},\bar{x},\bar{y})),$$

our concern is to have $\langle (w,z), A(w,z)\rangle > 0$ for every $(w,z) \in K^+$ with $A(w,z) \perp K^-$ and $(w,z) \neq (0,0)$. It's clear from (42) that

$$\begin{aligned} &\langle (w,z), A(w,z)\rangle = \langle w, Hw\rangle,\\ &(w,z) \in K^+ \iff z_i = 0 \text{ for } i \in I_1,\\ &A(w,z) \perp K^- \iff Hw + \nabla_x g(\bar{p},\bar{x})^{\mathsf{T}} z = 0 \text{ and } \nabla_x g(\bar{p},\bar{x})w \perp K_Y^-.\end{aligned}$$

Having $\nabla_x g(\bar{p},\bar{x})w \perp K_Y^-$ corresponds to having $w \perp \nabla_x g_i(\bar{p},\bar{x})$ for all $i \in I \setminus I_0$, which means $w \in M^+$. On the other hand, having $Hw + \nabla_x g(\bar{p},\bar{x})^{\mathsf{T}} z = 0$ corresponds to having $Hw = -(z_1\nabla_x g_1(\bar{p},\bar{x}) + \cdots + z_m \nabla_x g_m(\bar{p},\bar{x}))$. The sufficient condition in 2E.8 boils down, therefore, to the requirement that

$$\langle w, Hw\rangle > 0 \text{ when } (w,z) \neq (0,0),\ w \in M^+,\ Hw = -\textstyle\sum_{i\in I} z_i \nabla_x g_i(\bar{p},\bar{x}).$$

In particular this requirement has to cover cases where $w = 0$ but $z \neq 0$. That's obviously equivalent to the linear independence in (a). Beyond that, we need only observe that expressing Hw in the manner indicated corresponds simply to having $Hw \perp M^-$. Thus, the sufficient condition in Theorem 2E.8 turns out to be the combination of (a) with (b). □

Our final topic concerns the conditions under which the mapping S in (33) describes perturbations not only of stationarity, but also of local minima.

Theorem 2G.9 (implicit function theorem for local minima). *Suppose in the setting of the parameterized nonlinear programming problem (30) for twice continuously differentiable functions g_i and its Karush–Kuhn–Tucker mapping S in (33) that the following conditions hold in the notation coming from (34):*

(a) *the gradients $\nabla_x g_i(\bar{p},\bar{x})$ for $i \in I$ are linearly independent;*

(b) *$\langle w, \nabla^2_{xx}L(\bar{p},\bar{x},\bar{y})w\rangle > 0$ for every $w \neq 0$ in the subspace M^+ in (39).*

Then not only does S have a localization s with the properties laid out in Theorem 2G.8, but also, for every p in some neighborhood of $\bar{p}$, the x component of $s(p)$ furnishes a strong local minimum in (30). Moreover, (a) and (b) are necessary for this additional conclusion when $n+m$ is the rank of the $(n+m)\times d$ matrix B in (41).

Proof. Sufficiency. Condition (b) here is a stronger assumption than (b) of Theorem 2G.8, so we can be sure that (a) and (b) guarantee the existence of a localization s possessing the properties in that result. Moreover (b) implies satisfaction of the sufficient condition for a local minimum at $\bar{x}$ in Theorem 2G.6, inasmuch as the cone K in that theorem is obviously contained in the set of w such that $\langle \nabla g_i(\bar{p},\bar{x}), w\rangle = 0$ for all $i \in I \setminus I_0$. We need to demonstrate, however, that this local minimum property persists in passing from $\bar{p}$ to nearby p.

To proceed with that, denote the two components of $s(p)$ by $x(p)$ and $y(p)$, and let $I(p)$, $I_0(p)$ and $I_1(p)$ be the index sets which correspond to $x(p)$ as I, I_0 and I_1 do to $\bar{x}$, so that $I(p)$ consists of the indices $i \in \{1,\ldots,m\}$ with $g_i(p,x(p)) = 0$, and $I(p) \setminus I_0(p)$ consists of the indices $i \in I(p)$ having $y_i(p) > 0$ for inequality constraints, but $I_1(p)$ consists of the indices of the inequality constraints having $g_i(p,x(p)) < 0$. Consider the following conditions, which reduce to (a) and (b) when $p = \bar{p}$:

(a(p)) the gradients $\nabla_x g_i(p,x(p))$ for $i \in I(p)$ are linearly independent;

(b(p)) $\langle w, \nabla^2_{xx} L(p,x(p),y(p))w\rangle > 0$ for every $w \neq 0$ such that $w \perp \nabla_x g_i(p,x(p))$ for all $i \in I(p) \setminus I_0(p)$.

Since $x(p)$ and $y(p)$ tend toward $x(\bar{p}) = \bar{x}$ and $y(\bar{p}) = \bar{y}$ as $p \to \bar{p}$, the fact that $y_i(p) = 0$ for $i \in I_1(p)$ and the continuity of the g_i's ensure that

$$I(p) \subset I \text{ and } I(p) \setminus I_0(p) \supset I \setminus I_0 \text{ for } p \text{ near enough to } \bar{p}.$$

Through this and the fact that $\nabla_x g_i(p,x(p))$ tends toward $\nabla_x g_i(\bar{p},\bar{x})$ as p goes to $\bar{p}$, we see that the linear independence in (a) entails the linear independence in (a(p)) for p near enough to $\bar{p}$. Indeed, not only (a(p)) but also (b(p)) must hold, in fact in the stronger form that there exist $\varepsilon > 0$ and a neighborhood Q of $\bar{p}$ for which

$$\langle w, \nabla^2_{xx} L(p,x(p),y(p))w\rangle > \varepsilon$$

when $|w| = 1$ and $w \perp \nabla_x g_i(p,x(p))$ for all $i \in I(p) \setminus I_0(p)$. Indeed, otherwise there would be sequences of vectors $p_k \to \bar{p}$ and $w_k \to w$ violating this condition for $\varepsilon_k \to 0$, and this would lead to a contradiction of (b) in view of the continuous dependence of the matrix $\nabla^2_{xx} L(p,x(p),y(p))$ on p.

Of course, with both (a(p)) and (b(p)) holding when p is in some neighborhood of $\bar{p}$, we can conclude through Theorem 2G.6, as we did for $\bar{x}$, that $x(p)$ furnishes a strong local minimum for problem (30) for such p, since the cone

$$K(p) = \text{ set of } w \text{ satisfying } \begin{cases} \langle \nabla_x g_i(p,x(p)), w\rangle \le 0 & \text{for } i \in I(p) \text{ with } i \le s, \\ \langle \nabla_x g_i(p,x(p)), w\rangle = 0 & \text{for } i = s+1,\ldots,m \text{ and } i = 0 \end{cases}$$

lies in the subspace formed by the vectors w with $\langle \nabla_x g_i(p,x(p)), w\rangle = 0$ for all $i \in I(p) \setminus I_0(p)$.

Necessity. Suppose that S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $(\bar{x},\bar{y})$. We already know from Theorem 2G.8 that, under the rank condition in question, the auxiliary mapping $\bar{S}$ in (38) must have such a localization around $(0,0)$ for $(0,0)$, and that this requires the linear independence in (a). Under the further assumption now that $x(p)$ gives a local minimum in problem (30) when p is near enough to $\bar{p}$, we wish to deduce that (b) must hold as well. Having a local minimum at $x(p)$ implies that the second-order necessary condition for optimality in Theorem 2G.6 is satisfied with respect to the multiplier vector $y(p)$:

$$(44) \qquad \langle w, \nabla^2_{xx} L(p, x(p), y(p))w\rangle \geq 0 \text{ for all } w \in K(p) \text{ when } p \text{ is near to } \bar{p}.$$

We will now find a value of the parameter p close to $\bar{p}$ such that $(x(p), y(p)) = (\bar{x}, \bar{y})$ and $K(p) = M^+$. If $I_0 = \emptyset$ there is nothing to prove. Let $I_0 \neq \emptyset$. The rank condition on the Jacobian $B = \nabla_p f(\bar{p}, \bar{x}, \bar{y})$ provides through Theorem 1F.6 (for $k = 1$) the existence of $p(v,u)$, depending continuously on some (v,u) in a neighborhood of $(0, -g(\bar{p}, \bar{x}))$, such that $f(p(v,u), \bar{x}, \bar{y}) = (v,u)$, i.e., $\nabla_x L(p(v,u), \bar{x}, \bar{y}) = v$ and $-g(p(v,u), \bar{x}) = u$. For an arbitrarily small $\varepsilon > 0$, let the vector u^ε have $u_i^\varepsilon = -\varepsilon$ for $i \in I_0$ but $u_i^\varepsilon = 0$ for all other i. Let $p^\varepsilon = p(0, u^\varepsilon)$. Then $\nabla_x L(p^\varepsilon, \bar{x}, \bar{y}) = 0$ with $g_i(p^\varepsilon, \bar{x}) = 0$ for $i \in I \setminus I_0$ but $g_i(p^\varepsilon, \bar{x}) < 0$ for $i \in I_0$ as well as for $i \in I_1$. Thus, $I(p^\varepsilon) = I \setminus I_0$, $I_0(p^\varepsilon) = \emptyset$, $I_1(p^\varepsilon) = I_0 \cup I_1$, and $(\bar{x}, \bar{y}) \in S(p^\varepsilon)$ and, moreover, $(\bar{x}, \bar{y})$ furnishes a local minimum in (30) for $p = p^\varepsilon$, moreover with $K(p^\varepsilon)$ coming out to be the subspace

$$M^+(p^\varepsilon) = \left\{ w \,\middle|\, w \perp \nabla_x g_i(p^\varepsilon, \bar{x}) \text{ for all } i \in I \setminus I_0 \right\}.$$

In consequence of (44) we therefore have

$$\langle w, \nabla^2_{xx} L(p^\varepsilon, \bar{x}, \bar{y})w\rangle \geq 0 \text{ for all } w \in M^+(p^\varepsilon),$$

whereas we are asking in (b) for this to hold with strict inequality for $w \neq 0$ in the case of $M^+ = M^+(\bar{p})$.

We know that $p^\varepsilon \to \bar{p}$ as $\varepsilon \to 0$. Owing to (a) and the continuity of the functions g_i and their derivatives, the gradients $\nabla_x g_i(p^\varepsilon, \bar{x})$ for $i \in I$ must be linearly independent when ε is sufficiently small. It follows from this that any w in M^+ can be approximated as $\varepsilon \to 0$ by vectors w^ε belonging to the subspaces $M^+(p^\varepsilon)$. In the limit therefore, we have at least that

$$(45) \qquad \langle w, \nabla^2_{xx} L(\bar{p}, \bar{x}, \bar{y})w\rangle \geq 0 \text{ for all } w \in M^+.$$

How are we to conclude strict inequality when $w \neq 0$? It's important that the matrix $H = \nabla^2_{xx} L(\bar{p}, \bar{x}, \bar{y})$ is symmetric. In line with the positive semidefiniteness in (45), any $\bar{w} \in M^+$ with $\langle \bar{w}, H\bar{w}\rangle = 0$ must have $H\bar{w} \perp M^+$. But then in particular, the auxiliary solution mapping $\bar{S}$ in (38) would have $(t\bar{w}, 0) \in \bar{S}(0,0)$ for all $t \geq 0$, in contradiction to the fact, coming from Theorem 2G.8, that $\bar{S}(0,0)$ contains only $(0,0)$ in the current circumstances. $\square$

Condition (a) in Theorem 2G.9 is commonly called the *linear independence constraint qualification condition* while condition (b) is the *strong second order sufficient condition*.

Commentary

The basic facts about convexity, polyhedral sets, and tangent and normal cones given in Section 2A are taken mainly from Rockafellar [1970]. Robinson's implicit function theorem was stated and proved in Robinson [1980], where the author was clearly motivated by the problem of how the solutions of the standard nonlinear programming problem depend on parameters, and he pursued this goal in the same paper.

At that time it was already known from the work of Fiacco and McCormick [1968] that under the linear independence of the constraint gradients and the standard second-order sufficient condition, together with *strict complementarity slackness* at the reference point (which means that there are no inequality constraints satisfied as equalities that are associated with zero Lagrange multipliers), the solution mapping for the standard nonlinear programming problem has a *smooth* single-valued localization around the reference point. The proof of this result was based on the classical implicit function theorem, inasmuch as under strict complementarity slackness the Karush–Kuhn–Tucker system turns into a system of equations locally. Robinson looked at the case when the strict complementarity slackness is violated, which may happen, as already noted in 2B, when the "stationary point trajectory" hits the constraints. Based on his implicit function theorem, which actually reached far beyond his immediate goal, Robinson proved, still in his paper from 1980, that under a stronger form of the second-order sufficient condition, together with linear independence of the constraint gradients, the solution mapping of the standard nonlinear programming problem has a *Lipschitz continuous* single-valued localization around the reference point; see Theorem 2G.9 for an updated statement. In Chapter 3 we will look again at this result from much more general perspective.

This result was a stepping stone to the subsequent extensive development of stability analysis in optimization, whose maturity came with the publication of the books of Bank, Guddat, Klatte, Kummer and Tammer [1983], Levitin [1992], Bonnans and Shapiro [2000], Klatte and Kummer [2002] and Facchinei and Pang [2003].

Robinson's breakthrough in the stability analysis of nonlinear programming was in fact much needed for the emerging numerical analysis of variational problems more generally. In his paper from 1980, Robinson noted the thesis of his Ph.D. student Josephy [1979], who proved that the condition used in his implicit function theorem yields local quadratic convergence of Newton's method for solving variational inequalities, a method whose version for constrained optimization problems is well known as the sequential quadratic programming (SQP) method.

Quite a few years after Robinson's theorem was published, it was realized that the result could be used as a tool in the analysis of a variety of variational problems, and beyond. Alt [1990] applied it to optimal control, while in Dontchev and Hager [1993], and further in Dontchev [1995b], the statement of Robinson's theorem was observed actually to hold for generalized equations of the form 2B(1) for an *arbitrary* mapping F, not just a normal cone mapping. Variational inequalities thus serve as an example, not a limitation. Important applications, e.g. to convergence analysis of algorithms and discrete approximations to infinite-dimensional variational problems, came later. In the explosion of works in this area in the 80's and 90's Robinson's contribution, if not forgotten, was sometimes taken for granted. More about this will come out in Chapter 6.

One may derive stability estimates for solutions of optimization problems without relying on optimality conditions, but dealing directly with the objective function and the constraints under, for instance, strong convexity. Earlier results of this kind were obtained in Dontchev [1981], for a recent follow up, see Shvartsman [2012].

The presentation of the material in Section 2B mainly follows Dontchev and Rockafellar [2009], while that in Section 2C comes from Dontchev and Rockafellar [2001]. In Section 2D, we used some facts from the books of Facchinei and Pang [2003] (in particular, 2D.5) and Scholtes [2013][5] (in particular, 2D.7). The result in 2D.2 is a specialization of Theorem 5.43 in the book Rockafellar and Wets [1998] to the case of difference quotient functions. Reduction Lemma 2E.4 first appeared as Proposition 4.4 in Robinson [1991]. Theorem 2E.6 is a special case of a result in Dontchev and Rockafellar [1996]. For far reaching extensions of this result for variational inequalities over *perturbed* polyhedral convex sets, see Lu and Robinson [2008] and Robinson [2012].

Section 2F gives an introduction to the theory of monotone mappings which for its application to optimization problems goes back to Rockafellar [1976a] and [1976b]. Much more about this kind of monotonicity and its earlier history can be found in Chapter 12 of the book of Rockafellar and Wets [1998]. Related results are presented in Kassay and Kolumbán [1989], and Alt and Kolumbán [1993]. As mentioned above, the stability analysis in 2G goes back to Robinson [1980] who was the first to prove the sufficiency part of Theorem 2G.9. About the same time Kojima [1980] introduced the concept of strong stability in nonlinear programming which is closely related to strong regularity, and gave a characterization of this property by using degree theory. Cornet and Laroque [1987], and later Jongen, Klatte and Tammer [1990/91] combined Kojima's approach with Clarke's implicit function theorem. The presentation in 2G uses material from Dontchev and Rockafellar [1996,1998], but some versions of these results could be extracted from earlier works.

[5] This book is a reprint of Scholtes' habilitation thesis of 1994.

Chapter 3
Set-valued Analysis of Solution Mappings

In the concept of a solution mapping for a problem dependent on parameters, whether formulated with equations or something broader like variational inequalities, we have always had to face the possibility that solutions might not exist, or might not be unique when they do exist. This goes all the way back to the setting of the classical implicit function theorem. In letting $S(p)$ denote the set of all x satisfying $f(p,x)=0$, where f is a given function from $\mathbb{R}^d \times \mathbb{R}^n$ to $\mathbb{R}^m$, we cannot expect to be defining a *function* S from $\mathbb{R}^d$ to $\mathbb{R}^n$, even when $m=n$. In general, we only get a set-valued mapping S. However, this mapping S could have a single-valued localization s with properties of continuity or differentiability. The study of such localizations, as "subfunctions" within a set-valued mapping, has been our focus so far, but now we open up to a wider view.

There are plenty of reasons, already in the classical context, to be interested in localizations of solution mappings without insisting on single-valuedness. For instance, in the case of $S(p)=\{x \mid f(p,x)=0\}$ with f going from $\mathbb{R}^d \times \mathbb{R}^n$ to $\mathbb{R}^m$ and $m<n$, it can be anticipated for a choice of $\bar{p}$ and $\bar{x}$ with $\bar{x} \in S(\bar{p})$, under assumptions on $\nabla_x f(\bar{p},\bar{x})$, that a graphical localization S_0 of S exists around $(\bar{p},\bar{x})$ such that $S_0(p)$ is an $(n-m)$-dimensional manifold which varies with p. What generalizations of the usual notions of continuity and differentiability might help in understanding, and perhaps quantifying, this dependence on p?

Such challenges in dealing with the dependence of a set on the parameters which enter its definition carry over to solution mappings of problems centered on constraint systems. Just as the vector equation $f(p,x)=0$ for

$$f:\mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m \text{ with } f(p,x)=(f_1(p,x),\dots,f_m(p,x))$$

can be viewed as standing for a system of scalar equations

$$f_i(p,x)=0 \text{ for } i=1,\dots,m,$$

we can contemplate vector representations of mixed systems of inequalities and equations like

$$(1) \qquad f_i(p,x) \begin{cases} \leq 0 & \text{for } i = 1,\ldots,s, \\ = 0 & \text{for } i = s+1,\ldots,m, \end{cases}$$

which are important in optimization. Such a system takes the form

$$f(p,x) - K \ni 0 \ \text{ for } K = \mathbb{R}^s_- \times \{0\}^{m-s}.$$

In fact this is an instance of a parameterized generalized equation:

$$(2) \qquad f(p,x) + F(x) \ni 0 \ \text{ with } F \text{ a constant mapping, } F(x) \equiv -K.$$

In studying the behavior of the corresponding solution mapping $S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ given by

$$(3) \qquad S(p) = \{\, x \mid x \text{ satisfies } (2) \,\} \ \text{ (covering (1) as a special case)},$$

we are therefore still, very naturally, in the realm of the "extended implicit function theory" we have been working to build up.

In (2), F is not a normal cone mapping N_C, so we are not dealing with a variational inequality. The results in Chapter 2 for solution mappings to parameterized generalized equations would anyway, in principle, be applicable, but in this framework they miss the mark. The trouble is that those results focus on the prospects of finding single-valued localizations of a solution mapping, especially ones that exhibit Lipschitz continuity. For a solution mapping S as in (3), coming from a generalized equation as in (2), single-valued localizations are unlikely to exist at all (apart from the pure equation case with $m = n$) and aren't even a topic of serious concern. Rather, we are confronted with a "varying set" $S(p)$ which cannot be reduced locally to a "varying point." That could be the case even if, in (2), $F(x)$ is not a constant set but a sort of continuously moving or deforming set. What we are looking for is not so much a generalized implicit function theorem, but an *implicit mapping theorem*, the distinction being that "mappings" truly embrace set-valuedness.

To understand the behavior of such a solution mapping S, whether qualitatively or quantitatively, we have to turn to other concepts, beyond those in Chapter 2. Our immediate task, in Sections 3A, 3B, 3C and 3D, is to introduce the notions of Painlevé–Kuratowski convergence and Pompeiu–Hausdorff convergence for sequences of sets, and to utilize them in developing properties of continuity and Lipschitz continuity for set-valued mappings. In tandem with this, we gain important insights into the solution mappings (3) associated with constraint systems as in (1) and (2). We also obtain by-products concerning the behavior of various mappings associated with problems of optimization.

In Section 3E, however, we open a broader investigation in which the Aubin property, serving as a sort of localized counterpart to Lipschitz continuity for set-valued mappings, is tied to the concept of metric regularity, which directly relates to estimates of distances to solutions. The natural context for this is the study of how properties of a set-valued mapping correspond to properties of its set-valued inverse, or in other words, the paradigm of the inverse function theorem. We are able nevertheless to return in Section 3F to the paradigm of the implicit function theorem, based on a stability property of metric regularity fully developed later in Chapter 5. Powerful results, applicable to fully set-valued solution mappings (2)(3) even when F is *not* just a constant mapping, are thereby obtained. Sections 3G and 3I then take these ideas back to situations where single-valuedness is available in a localization of a solution mapping, at least at the reference point, showing how previous results such as those in 2B can thereby be amplified. Section 3H reveals that a set-valued version of calmness does not similarly submit to the implicit function theorem paradigm.

3A. Set Convergence

Various continuity properties of set-valued mappings will be essential for the developments in this chapter. To lay the foundation for them we must first introduce two basic concepts: convergence of sets and distance between sets.

The set of all natural numbers $k = 1,2,\ldots$, will be denoted by $\mathbb{N}$. The collection of all subsets N of $\mathbb{N}$ such that $\mathbb{N} \setminus N$ is finite will be denoted by $\mathcal{N}$, whereas the collection of all infinite $N \subset \mathbb{N}$ will be denoted by $\mathcal{N}^\sharp$. This scheme is designed for convenience in handling subsequences of a given sequence. For instance, if we have a sequence $\{x^k\}_{k=1}^\infty$ of points in $\mathbb{R}^n$, the notation $\{x^k\}_{k\in N}$ for either $N \in \mathcal{N}$ or $N \in \mathcal{N}^\sharp$ designates a subsequence. In the first case it is a subsequence which coincides with the full sequence beyond some k_0, whereas in the second case it is a general subsequence. Limits as $k \to \infty$ with $k \in N$ will be indicated by $\lim_{k\in N}$, or in terms of arrows by $\overset{N}{\to}$, and so forth.

For a sequence $\{r^k\}_{k=1}^\infty$ in $\mathbb{R}$, the limit as $k \to \infty$ may or may not exist—even though we always include ∞ and $-\infty$ as possible limit values in the obvious sense. However, the upper limit, or "limsup," and the lower limit, or "liminf," do always exist, as defined by

$$\limsup_{k\to\infty} r^k = \lim_{k\to\infty} \sup_{m\geq k} r^m,$$
$$\liminf_{k\to\infty} r^k = \lim_{k\to\infty} \inf_{m\geq k} r^m.$$

An alternative description of these values is that $\limsup_{k\to\infty} r^k$ is the highest r for which there exists $N \in \mathcal{N}^\sharp$ such that $r^k \overset{N}{\to} r$, whereas $\liminf_{k\to\infty} r^k$ is the lowest such r. The limit itself exists if and only if these upper and lower limits coincide. For simplicity, we often just write $\limsup_k$, $\liminf_k$ and $\lim_k$, with the understanding that this refers to $k \to \infty$.

In working with sequences of sets, a similar pattern is encountered in which "outer" and "inner" limits always exist and give a "limit" when they agree.

Outer and inner limits. *Consider a sequence $\{C^k\}_{k=1}^\infty$ of subsets of $\mathbb{R}^n$.*

(a) *The* outer limit *of this sequence, denoted by $\limsup_k C^k$, is the set of all $x \in \mathbb{R}^n$ for which*

$$\text{there exist } N \in \mathcal{N}^\sharp \text{ and } x^k \in C^k \text{ for } k \in N \text{ such that } x^k \overset{N}{\to} x.$$

(b) *The* inner limit *of this sequence, denoted by $\liminf_k C^k$, is the set of all $x \in \mathbb{R}^n$ for which*

$$\text{there exist } N \in \mathcal{N} \text{ and } x^k \in C^k \text{ for } k \in N \text{ such that } x^k \overset{N}{\to} x.$$

(c) *When the inner and outer limits are the same set C, this set is defined to be the* limit *of the sequence $\{C^k\}_{k=1}^\infty$:*

$$C = \lim_k C^k = \limsup_k C^k = \liminf_k C^k.$$

In this case C^k is said to converge to C in the sense of Painlevé–Kuratowski convergence.

Note that although the outer and inner limit sets always exist by this definition, they might be empty. When $C^k \neq \emptyset$ for all k, these sets can be described equivalently in terms of the sequences $\{x^k\}_{k=1}^{\infty}$ that can be formed by selecting an $x_k \in C_k$ for each k: the set of all cluster points of such sequences is $\limsup_k C^k$, while the set of all limits of such sequences is $\liminf_k C^k$. Obviously $\liminf_k C^k \subset \limsup_k C^k$. When each C^k is a singleton, $\liminf_k C^k$ can at most be another singleton, but $\limsup_k C^k$ might have multiple elements.

Examples.

1) The sequence of doubletons $C^k = \{0, \frac{1}{k}\}$ in $\mathbb{R}$ has $\lim_k C^k = \{0\}$. Indeed, every sequence of elements $x_k \in C_k$ converges to 0.

2) The sequence of balls $\mathbb{B}(x^k, \rho^k)$ converges to $\mathbb{B}(x, \rho)$ when $x^k \to x$ and $\rho^k \to \rho$.

3) A sequence of sets C^k which alternates between two different closed sets D_1 and D_2, that is, $C^k = D_1$ when k is odd and $C^k = D_2$ when k is even, has $D_1 \cap D_2$ as its inner limit and $D_1 \cup D_2$ as its outer limit. Such a sequence is not convergent if $D_1 \neq D_2$.

Outer and inner limits can also be described with the help of neighborhoods:

$$\text{(1a)}\quad \limsup_{k\to\infty} C^k = \Big\{ x \,\Big|\, \forall \text{ neighborhood } V \text{ of } x, \exists N \in \mathcal{N}^\sharp, \forall k \in N : C^k \cap V \neq \emptyset \Big\},$$

$$\text{(1b)}\quad \liminf_{k\to\infty} C^k = \Big\{ x \,\Big|\, \forall \text{ neighborhood } V \text{ of } x, \exists N \in \mathcal{N}, \forall k \in N : C^k \cap V \neq \emptyset \Big\}.$$

Without loss of generality the neighborhoods in (1a,b) can be taken to be closed balls; then we obtain the following more transparent definitions:

$$\text{(2a)}\qquad \limsup_{k\to\infty} C^k = \big\{ x \,\big|\, \forall \varepsilon > 0, \exists N \in \mathcal{N}^\sharp : x \in C^k + \varepsilon \mathbb{B} \ (k \in N) \big\},$$

$$\text{(2b)}\qquad \liminf_{k\to\infty} C^k = \big\{ x \,\big|\, \forall \varepsilon > 0, \exists N \in \mathcal{N} : x \in C^k + \varepsilon \mathbb{B} \ (k \in N) \big\}.$$

Both the outer and inner limits of a sequence $\{C^k\}_{k\in N}$ are *closed sets*. Indeed, if $x \notin \limsup_k C^k$, then, from (2a), there exists $\varepsilon > 0$ such that for every $N \in \mathcal{N}^\sharp$ we have $x \notin C^k + \varepsilon\mathbb{B}$, that is, $\mathbb{B}(x, \varepsilon) \cap C^k = \emptyset$, for some $k \in N$. But then a neighborhood of x can meet C^k for finitely many k only. Hence no points in this neighborhood can be cluster points of sequences $\{x^k\}$ with $x^k \in C^k$ for infinitely many k. This implies that the complement of $\limsup_k C^k$ is an open set and therefore that $\limsup_k C^k$ is closed. An analogous argument works for $\liminf_k C^k$ (this could also be derived by the following Proposition 3A.1).

Recall from Section 1D that the distance from a point $x \in \mathbb{R}^n$ to a subset C of $\mathbb{R}^n$ is

$$d_C(x) = d(x,C) = \inf_{y \in C} |x-y|.$$

Recall Proposition 1D.4, according to which a set C is closed if and only if $d(x,C) = 0$ is equivalent to having $x \in C$.

Proposition 3A.1 (distance function characterizations of limits)**.** *Outer and inner limits of sequences of sets are described alternatively by the following formulas:*

(3a) $$\limsup_{k\to\infty} C^k = \left\{ x \,\Big|\, \liminf_{k\to\infty} d(x,C^k) = 0 \right\},$$

(3b) $$\liminf_{k\to\infty} C^k = \left\{ x \,\Big|\, \lim_{k\to\infty} d(x,C^k) = 0 \right\}.$$

Proof. If $x \in \limsup_k C^k$ then, by (2a), for any $\varepsilon > 0$ there exists $N \in \mathcal{N}^\sharp$ such that $d(x,C^k) \le \varepsilon$ for all $k \in N$. But then, by the definition of the lower limit for a sequence of real numbers, as recalled in the beginning of this section, we have $\liminf_{k\to\infty} d(x,C^k) = 0$. The left side of (3a) is therefore contained in the right side. Conversely, if x is in the set on the right side of (3a), then there exists $N \in \mathcal{N}^\sharp$ and $x^k \in C^k$ for all $k \in N$ such that $x^k \xrightarrow{N} x$; then, by definition, x must belong to the left side of (3a).

If x is not in the set on the right side of (3b), then there exist $\varepsilon > 0$ and $N \in \mathcal{N}^\sharp$ such that $d(x,C^k) > \varepsilon$ for all $k \in N$. Then $x \notin C^k + \varepsilon \mathbb{B}$ for all $k \in N$ and hence by (2b) x is not in $\liminf_k C^k$. In a similar way, from (2b) we obtain that $x \notin \liminf_k C^k$ only if $\limsup_k d(x,C^k) > 0$. This gives us (3b). □

Observe that the distance to a set does not distinguish whether this set is closed or not. Therefore, in the context of convergence, there is no difference whether the sets in a sequence are closed or not. (But limits of all types are closed sets.)

More examples.

1) The limit of the sequence of intervals $[k,\infty)$ as $k \to \infty$ is the empty set, whereas the limit of the sequence of intervals $[1/k,\infty)$ is $[0,\infty)$.

2) More generally for monotone sequences of subsets $C^k \subset \mathbb{R}^n$, if $C^k \supset C^{k+1}$ for all $k \in \mathbb{N}$, then $\lim_k C^k = \cap_k \operatorname{cl} C^k$, whereas if $C^k \subset C^{k+1}$ for all k, then $\lim_k C^k = \operatorname{cl} \cup_k C^k$.

3) The constant sequence $C^k = D$, where D is the set of vectors in $\mathbb{R}^n$ whose coordinates are rational numbers, converges not to D, which isn't closed, but to the closure of D, which is $\mathbb{R}^n$. More generally, if $C^k = C$ for all k, then $\lim_k C^k = \operatorname{cl} C$.

Theorem 3A.2 (characterization of Painlevé–Kuratowski convergence)**.** *For a sequence C^k of sets in $\mathbb{R}^n$ and a closed set $C \subset \mathbb{R}^n$ one has:*

(a) *$C \subset \liminf_k C^k$ if and only if for every open set $O \subset \mathbb{R}^n$ with $C \cap O \neq \emptyset$ there exists $N \in \mathcal{N}$ such that $C^k \cap O \neq \emptyset$ for all $k \in N$;*

(b) *$C \supset \limsup_k C^k$ if and only if for every compact set $B \subset \mathbb{R}^n$ with $C \cap B = \emptyset$ there exists $N \in \mathcal{N}$ such that $C^k \cap B = \emptyset$ for all $k \in N$;*

(c) $C \subset \liminf_k C^k$ *if and only if for every* $\rho > 0$ *and* $\varepsilon > 0$ *there is an index set* $N \in \mathcal{N}$ *such that* $C \cap \rho I\!B \subset C^k + \varepsilon I\!B$ *for all* $k \in N$*;*

(d) $C \supset \limsup_k C^k$ *if and only if for every* $\rho > 0$ *and* $\varepsilon > 0$ *there is an index set* $N \in \mathcal{N}$ *such that* $C^k \cap \rho I\!B \subset C + \varepsilon I\!B$ *for all* $k \in N$*;*

(e) $C \subset \liminf_k C^k$ *if and only if* $\limsup_k d(x, C^k) \le d(x, C)$ *for every* $x \in I\!R^n$*;*

(f) $C \supset \limsup_k C^k$ *if and only if* $d(x, C) \le \liminf_k d(x, C^k)$ *for every* $x \in I\!R^n$*.*

Thus, from (c)(d) $C = \lim_k C^k$ *if and only if for every* $\rho > 0$ *and* $\varepsilon > 0$ *there is an index set* $N \in \mathcal{N}$ *such that*

$$C^k \cap \rho I\!B \subset C + \varepsilon I\!B \text{ and } C \cap \rho I\!B \subset C^k + \varepsilon I\!B \quad \text{for all } k \in N.$$

Also, from (e)(f), $C = \lim_k C^k$ *if and only if* $\lim_k d(x, C^k) = d(x, C)$ *for every* $x \in I\!R^n$*.*

Proof. (a): Necessity comes directly from (1b). To show sufficiency, assume that there exists $x \in C \setminus \liminf_k C^k$. But then, by (1b), there exists an open neighborhood V of x such that for every $N \in \mathcal{N}$ there exists $k \in N$ with $V \cap C^k = \emptyset$ and also $V \cap C \neq \emptyset$. This is the negation of the condition on the right.

(b): Let $C \supset \limsup_k C^k$ and let there exist a compact set B with $C \cap B = \emptyset$, such that for every $N \in \mathcal{N}$ one has $C^k \cap B \neq \emptyset$ for some $k \in N$. But then there exist $N \in \mathcal{N}^\sharp$ and a convergent sequence $x^k \in C^k$ for $k \in N$ whose limit is not in C, a contradiction. Conversely, if there exists $x \in \limsup_k C^k$ which is not in C then, from (2a), a ball $I\!B_\varepsilon(x)$ with sufficiently small radius ε does not meet C yet meets C^k for infinitely many k; this contradicts the condition on the right.

Sufficiency in (c): Consider any point $x \in C$, and any $\rho > |x|$. For an arbitrary $\varepsilon > 0$, there exists, by assumption, an index set $N \in \mathcal{N}$ such that $C \cap \rho I\!B \subset C^k + \varepsilon I\!B$ for all $k \in N$. Then $x \in C^k + \varepsilon I\!B$ for all $k \in N$. By (2b), this yields $x \in \liminf_k C^k$. Hence, $C \subset \liminf_k C^k$.

Necessity of (c): It will be demonstrated that if the condition fails, there must be a point $\bar{x} \in C$ lying outside of $\liminf_k C^k$. To say that the condition fails is to say that there exist $\rho > 0$ and $\varepsilon > 0$, such that, for each $N \in \mathcal{N}$, the inclusion $C \cap \rho I\!B \subset C^k + \varepsilon I\!B$ is false for at least one $k \in N$. Then there is an index set $N_0 \in \mathcal{N}^\sharp$ such that this inclusion is false for all $k \in N_0$; there are points $x^k \in [C \cap \rho I\!B] \setminus [C^k + \varepsilon I\!B]$ for all $k \in N_0$. Such points form a bounded sequence in the closed set C with the property that $d(x^k, C^k) \ge \varepsilon$. A subsequence $\{x^k\}_{k \in N_1}$, for an index set $N_1 \in \mathcal{N}^\sharp$ within N_0, converges in that case to a point $\bar{x} \in C$. Since $d(x^k, C^k) \le d(\bar{x}, C^k) + |\bar{x} - x^k|$, we must have

$$d(\bar{x}, C^k) \ge \varepsilon/2 \text{ for all } k \in N_1 \text{ large enough.}$$

It is impossible then for $\bar{x}$ to belong to $\liminf_k C^k$, because that requires $d(\bar{x}, C^k)$ to converge to 0, cf. (3b).

Sufficiency in (d): Let $\bar{x} \in \limsup_k C^k$; then for some $N_0 \in \mathcal{N}^\sharp$ there are points $x^k \in C^k$ such that $x^k \overset{N_0}{\to} \bar{x}$. Fix any $\rho > |\bar{x}|$, so that $x^k \in \rho I\!B$ for $k \in N_0$ large enough. By assumption, there exists for every $\varepsilon > 0$ an index set $N \in \mathcal{N}$ such that $C^k \cap \rho I\!B \subset C + \varepsilon I\!B$ when $k \in N$. Then for large enough $k \in N_0 \cap N$ we have $x^k \in C + \varepsilon I\!B$, hence $d(x^k, C) \le \varepsilon$. Because $d(\bar{x}, C) \le d(x^k, C) + |x^k - \bar{x}|$ and $x^k \overset{N_0}{\to} \bar{x}$, it follows from the arbitrary choice of ε that $d(\bar{x}, C) = 0$, which means $\bar{x} \in C$ (since C is closed).

Necessity in (d): Suppose to the contrary that one can find $\rho > 0$, $\varepsilon > 0$ and $N \in \mathcal{N}^\sharp$ such that, for all $k \in N$, there exists $x^k \in [C^k \cap \rho I\!B] \setminus [C + \varepsilon I\!B]$. The sequence $\{x^k\}_{k \in N}$ is then bounded, so it has a cluster point $\bar{x}$ which, by definition, belongs to $\limsup_k C^k$. On the other hand, since each x^k lies outside of $C + \varepsilon I\!B$, we have $d(x^k, C) \geq \varepsilon$ and, in the limit, $d(\bar{x}, C) \geq \varepsilon$. Hence $\bar{x} \notin C$, and therefore $\limsup_k C^k$ is not a subset of C.

(e): Sufficiency follows from (3b) by taking $x \in C$. To prove necessity, choose $x \in I\!R^n$ and let $y \in C$ be a projection of x on C: $|x - y| = d(x, C)$. By the definition of liminf there exist $N \in \mathcal{N}$ and $y^k \in C^k$, $k \in N$ such that $y^k \xrightarrow{N} y$. For such y^k we have $d(x, C^k) \leq |y^k - x|$, $k \in N$ and passing to the limit with $k \to \infty$ we get the condition on the right.

(f): Sufficiency follows from (3a) by taking $x \in \limsup_k C^k$. Choose $x \in I\!R^n$. If $x \in C$ there is nothing to prove. If not, note that for any nonnegative α the condition $d(x, C) > \alpha$ is equivalent to $C \cap I\!B_\alpha(x) = \emptyset$. But then from (b) there exists $N \in \mathcal{N}$ with $C^k \cap I\!B_\alpha(x) = \emptyset$ for $k \in N$, which is the same as $d(x, C^k) > \alpha$ for $k \in N$. This implies the condition on the right. □

Observe that in parts (c)(d) of 3A.2 we can replace the phrase "for every ρ" by "there is some $\rho_0 \geq 0$ such that for every $\rho \geq \rho_0$".

Set convergence can also be characterized in terms of concepts of distance between sets.

Excess and Pompeiu–Hausdorff distance. *For sets C and D in $I\!R^n$, the excess of C beyond D is defined by*

$$e(C, D) = \sup_{x \in C} d(x, D),$$

where the convention is used that

$$e(\emptyset, D) = \begin{cases} 0 & \text{when } D \neq \emptyset, \\ \infty & \text{otherwise.} \end{cases}$$

The Pompeiu–Hausdorff distance between C and D is the quantity

$$h(C, D) = \max\{e(C, D), e(D, C)\}.$$

Equivalently, these quantities can be expressed by

$$e(C, D) = \inf\{\, \tau \geq 0 \,|\, C \subset D + \tau I\!B \}$$

and

$$h(C, D) = \inf\{\, \tau \geq 0 \,|\, C \subset D + \tau I\!B,\, D \subset C + \tau I\!B \}.$$

The excess and the Pompeiu–Hausdorff distance are illustrated in Fig. 3.1. They are unaffected by whether C and D are closed or not, but in the case of closed sets the infima in the alternative formulas are attained. Note that both $e(C, D)$ and $h(C, D)$ can sometimes be ∞ when unbounded sets are involved. For that reason in particular, the Pompeiu–Hausdorff distance does not furnish a metric on the space of nonempty closed subsets of $I\!R^n$, although it does on the space of nonempty closed subsets of a bounded set $X \subset I\!R^n$. Also note that $e(C,\emptyset) = \infty$ for any set C, including the empty set.

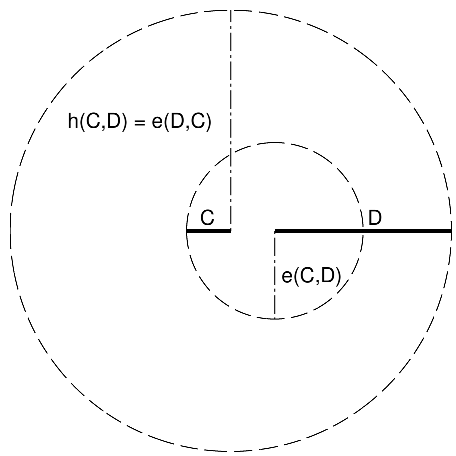

Fig. 3.1 Illustration of the excess and Pompeiu–Hausdorff distance.

Proposition 3A.3 (characterization of Pompeiu–Hausdorff distance)**.** *For any nonempty sets C and D in $I\!R^n$, one has*

$$h(C,D) = \sup_{x \in I\!R^n} |d(x,C) - d(x,D)|. \tag{4}$$

Proof. Since the distance to a set doesn't distinguish whether the set is closed or not, we may assume that C and D are nonempty closed sets.

According to 1D.4(c), for any $x \in I\!R^n$ we can pick $u \in C$ such that $d(x,u) = d(x,C)$. For any $v \in D$, the triangle inequality tells us that $d(x,v) \leq d(x,u) + d(u,v)$. Taking the infimum on both sides with respect to $v \in D$, we see that $d(x,D) \leq d(x,u) + d(u,D)$, where $d(u,D) \leq e(C,D)$. Therefore, $d(x,D) - d(x,C) \leq e(C,D)$, and by symmetry in exchanging the roles of C and D, also $d(x,C) - d(x,D) \leq e(D,C)$, so that

$$|d(x,C) - d(x,D)| \leq \max\{e(C,D), e(D,C)\} = h(C,D).$$

Hence "$\geq$" holds in (4).

On the other hand, since $d(x,C) = 0$ when $x \in C$, we have

$$e(C,D) = \sup_{x \in C} d(x,D) = \sup_{x \in C} |d(x,D) - d(x,C)| \leq \sup_{x \in I\!R^n} |d(x,D) - d(x,C)|$$

and likewise $e(D,C) \leq \sup_{x \in I\!R^n} |d(x,C) - d(x,D)|$, so that

$$\max\{e(C,D), e(D,C)\} \le \sup_{x\in\mathbb{R}^n} |d(x,C) - d(x,D)|.$$

This confirms that "$\le$" also holds in (4). □

Pompeiu–Hausdorff convergence. *A sequence of sets $\{C^k\}_{k=1}^\infty$ is said to converge with respect to Pompeiu–Hausdorff distance to a set C when C is closed and $h(C^k, C) \to 0$ as $k \to \infty$.*

From the definition of the Pompeiu-Hausdorff distance it follows that when a sequence C^k converges to C, then the set C must be nonempty and only finitely many C^k can be empty. Note that this is *not* the case when Painlevé–Kuratowski convergence is considered (see the first example after 3A.1). The following theorem exhibits the main relationship between these two types of convergence.

Theorem 3A.4 (Pompeiu–Hausdorff versus Painlevé–Kuratowski)**.** *If a sequence of closed sets $\{C^k\}_{k=1}^\infty$ converges to C with respect to Pompeiu–Hausdorff distance then it also converges to C in the Painlevé–Kuratowski sense. The opposite implication holds if there is a bounded set X which contains C and every C^k.*

Proof. By definition, C^k converges to C with respect to Pompeiu–Hausdorff distance if and only if, for every $\varepsilon > 0$, there exists $N \in \mathscr{N}$ with

$$C^k \subset C + \varepsilon\mathbb{B} \;\text{ and }\; C \subset C^k + \varepsilon\mathbb{B} \quad \text{for all } k \in N. \tag{5}$$

Since (5) implies (2a,b), the Painlevé–Kuratowski convergence of C^k to C then follows from the convergence with respect to Pompeiu–Hausdorff distance.

Suppose now that C^k converges to C in the Painlevé–Kuratowski sense, with C and every C^k included in a bounded set X. Then there exists $\rho_0 > 0$ such that $C^k = C^k \cap \rho\mathbb{B}$ and $C = C \cap \rho\mathbb{B}$ for every $\rho \ge \rho_0$. We obtain that for every $\rho > \rho_0$ and $k \in \mathbb{N}$,

$$C^k = C^k \cap \rho\mathbb{B} \subset C + \varepsilon\mathbb{B} \;\text{ and }\; C = C \cap \rho\mathbb{B} \subset C^k + \varepsilon\mathbb{B}.$$

But then, for every $\rho > 0$ we have

$$C^k \cap \rho\mathbb{B} \subset C^k \subset C + \varepsilon\mathbb{B} \;\text{ and }\; C \cap \rho\mathbb{B} \subset C \subset C^k + \varepsilon\mathbb{B} \;\text{ for } k \in \mathbb{N}$$

and hence (5) holds and we have convergence of C^k to C with respect to Pompeiu–Hausdorff distance. □

Exercise 3A.5 (convergence equivalence under boundedness)**.** *For a sequence of sets C^k in $\mathbb{R}^n$ and a nonempty closed set C, the following are equivalent:*

(a) *C^k converges to C in the Pompeiu–Hausdorff sense and C is bounded;*

(b) *C^k converges to C in the Painlevé–Kuratowski sense and there is a bounded set X along with an index set $N \in \mathscr{N}$ such that $C^k \subset X$ for all $k \in N$.*

Theorem 3A.6 (conditions for Pompeiu–Hausdorff convergence). *A sequence C^k of sets in $\mathbb{R}^n$ is convergent with respect to Pompeiu–Hausdorff distance to a closed set $C \subset \mathbb{R}^n$ if both of the following conditions hold:*

(a) *for every open set $O \subset \mathbb{R}^n$ with $C \cap O \neq \emptyset$ there exists $N \in \mathscr{N}$ such that $C^k \cap O \neq \emptyset$ for all $k \in N$;*

(b) *for every open set $O \subset \mathbb{R}^n$ with $C \subset O$ there exists $N \in \mathscr{N}$ such that $C^k \subset O$ for all $k \in N$.*

Moreover, condition (a) is always necessary for Pompeiu–Hausdorff convergence, while (b) is necessary when the set C is bounded.

Proof. Let (a)(b) hold and let the second inclusion in (5) be violated, that is, there exist $x \in C$, a scalar $\varepsilon > 0$ and a sequence $N \in \mathscr{N}^\sharp$ such that $x \notin C^k + \varepsilon \mathbb{B}$ for $k \in N$. Then an open neighborhood of x does not meet C^k for infinitely many k; this contradicts condition (a). Furthermore, (b) implies that for any $\varepsilon > 0$ there exists $N \in \mathscr{N}$ such that $C^k \subset C + \varepsilon \mathbb{B}$ for all $k \in N$, which is the first inclusion in (5). Then Pompeiu–Hausdorff convergence follows from (5).

According to 3A.2(a), condition (a) is equivalent to $C \subset \liminf_k C^k$, and hence it is necessary for Painlevé–Kuratowski convergence, and then also for Pompeiu–Hausdorff convergence. To show necessity of (b), let $C \subset O$ for some open set $O \subset \mathbb{R}^n$. For $k \in \mathbb{N}$ let there exist points $x^k \in C$ and y^k in the complement of O such that $|x^k - y^k| \to 0$ as $k \to \infty$. Since C is compact, there exists $N \in \mathscr{N}^\sharp$ and $x \in C$ such that $x^k \xrightarrow{N} x$, hence $y^k \xrightarrow{N} x$ as well. But then x must be also in the complement of O, which is impossible. The contradiction so obtained shows there is an $\varepsilon > 0$ such that $C + \varepsilon \mathbb{B} \subset O$; then, from (5), for some $N \subset \mathscr{N}$ we have $C^k \subset O$ for $k \in N$. □

Examples 3A.7 (unboundedness issues). As an illustration of the troubles that may occur when we deal with unbounded sets, consider first the sequence of bounded sets $C^k \subset \mathbb{R}^2$ in which C^k is the segment having one end at the origin and the other at the point $(\cos \frac{1}{k}, \sin \frac{1}{k})$; that is,

$$C^k = \Big\{ x \in \mathbb{R}^2 \,\Big|\, x_1 = t \cos \frac{1}{k},\ x_2 = t \sin \frac{1}{k},\ 0 \le t \le 1 \Big\}.$$

Both the Painlevé–Kuratowski and Pompeiu–Hausdorff limits exist and are equal to the segment having one end at the origin and the other at the point $(1,0)$. Also, both conditions (a) and (b) in 3A.6 are satisfied.

Let us now modify this example by taking as C^k, instead of a segment, the whole unbounded ray with its end at the origin. That is,

$$C^k = \Big\{ x \in \mathbb{R}^2 \,\Big|\, x_1 = t \cos \frac{1}{k},\ x_2 = t \sin \frac{1}{k},\ t \ge 0 \Big\}.$$

The Painlevé–Kuratowski limit is the ray $\{ x \in \mathbb{R}^2 \,|\, x_1 \ge 0, x_2 = 0 \}$, whereas the Pompeiu–Hausdorff limit fails to exist. In this case condition (a) in 3A.6 holds, whereas (b) is violated.

As another example demonstrating issues with unboundedness, consider the sequence of sets

$$C^k = \left\{ x \in \mathbb{R}^2 \,\middle|\, x_1 > 0,\; x_2 \geq \frac{1}{x_1} - \frac{1}{k} \right\},$$

which is obviously convergent with respect to Pompeiu–Hausdorff distance to the set $C = \left\{ x \in \mathbb{R}^2 \,\middle|\, x_1 > 0, x_2 \geq 1/x_1 \right\}$ (choose $k > 1/\varepsilon$ in (5)). On the other hand, condition (b) in 3A.6 fails, since the open set $O = \left\{ x \in \mathbb{R}^2 \,\middle|\, x_1 > 0, x_2 > 0 \right\}$ contains C but does not contain C^k for any k.

3B. Continuity of Set-valued Mappings

Continuity properties of a set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ can be defined on the basis of Painlevé–Kuratowski set convergence. Alternatively they can be defined on the basis of Pompeiu–Hausdorff set convergence, which is the same in a context of boundedness but otherwise is more stringent and only suited to special situations, as explained at the end of Section 3A. Following the pattern of inner and outer limits used in introducing Painlevé–Kuratowski convergence, we let

$$\begin{aligned}
\limsup_{y \to \bar{y}} S(y) &= \bigcup_{y^k \to \bar{y}} \limsup_{k \to \infty} S(y^k) \\
&= \left\{ x \,\middle|\, \exists y^k \to \bar{y},\ \exists x^k \to x \ \text{ with } \ x^k \in S(y^k) \right\}
\end{aligned}$$

and

$$\begin{aligned}
\liminf_{y \to \bar{y}} S(y) &= \bigcap_{y^k \to \bar{y}} \liminf_{k \to \infty} S(y^k) \\
&= \left\{ x \,\middle|\, \forall y^k \to \bar{y},\ \exists N \in \mathcal{N}, x^k \xrightarrow{N} x \ \text{ with } \ x^k \in S(y^k) \right\}.
\end{aligned}$$

In other words, the limsup is the set of all possible limits of sequences $x^k \in S(y^k)$ when $y^k \to \bar{y}$, while the liminf is the set of points x for which there exists a sequence $x^k \in S(y^k)$ when $y^k \to \bar{y}$ such that $x^k \to x$.

Semicontinuity and continuity. *A set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is outer semicontinuous (osc) at $\bar{y}$ when*

$$\limsup_{y \to \bar{y}} S(y) \subset S(\bar{y})$$

and inner semicontinuous (isc) at $\bar{y}$ when

$$\liminf_{y \to \bar{y}} S(y) \supset S(\bar{y}).$$

It is called Painlevé–Kuratowski continuous at $\bar{y}$ when it is both osc and isc at $\bar{y}$, as expressed by

$$\lim_{y \to \bar{y}} S(y) = S(\bar{y}).$$

On the other hand, S is called Pompeiu–Hausdorff continuous *at $\bar{y}$ when*

$$S(\bar{y}) \text{ is closed and } \lim_{y \to \bar{y}} h(S(y), S(\bar{y})) = 0.$$

These terms are invoked relative to a subset D in $\mathbb{R}^m$ *when the properties hold for limits taken with $y \to \bar{y}$ in D (but not necessarily for limits $y \to \bar{y}$ without this restriction). Continuity is taken to refer to Painlevé–Kuratowski continuity, unless otherwise specified.*

For single-valued mappings both definitions of continuity reduce to the usual definition of continuity of a function. Note that when S is isc at $\bar{y}$ relative to D then there must exist a neighborhood V of $\bar{y}$ such that $D \cap V \subset \operatorname{dom} S$. When $D = \mathbb{R}^m$, this means $\bar{y} \in \operatorname{int}(\operatorname{dom} S)$.

Exercise 3B.1 (limit relations as equations)**.**

(a) *Show that S is osc at $\bar{y}$ if and only if actually* $\limsup_{y \to \bar{y}} S(y) = S(\bar{y})$.

(b) *Show that, when $S(\bar{y})$ is closed, S is isc at $\bar{y}$ if and only if* $\liminf_{y \to \bar{y}} S(y) = S(\bar{y})$.

Although the closedness of $S(\bar{y})$ is automatic from this when S is continuous at $\bar{y}$ in the Painlevé–Kuratowski sense, it needs to be assumed directly for Pompeiu–Hausdorff continuity because the distance concept utilized for that concept is unable to distinguish whether sets are closed or not.

Recall that a set M is closed relative to a set D when any sequence $y^k \in M \cap D$ has its cluster points in M. A set M is open relative to D if the complement of M is closed relative to D. Also, recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is lower semicontinuous on a closed set $D \subset \mathbb{R}^n$ when the lower level set $\{ x \in D \mid f(x) \leq \alpha \}$ is closed for every $\alpha \in \mathbb{R}$. We defined this property at the beginning of Chapter 1 for the case of $D = \mathbb{R}^n$ and for functions with values in $\mathbb{R}$, and now are merely echoing that for $D \subset \mathbb{R}^n$.

Theorem 3B.2 (characterization of semicontinuity)**.** *For $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, a set $D \subset \mathbb{R}^m$ and $\bar{y} \in \operatorname{dom} S$ we have:*

(a) *S is osc at $\bar{y}$ relative to D if and only if for every $x \notin S(\bar{y})$ there are neighborhoods U of x and V of $\bar{y}$ such that $D \cap V \cap S^{-1}(U) = \emptyset$;*

(b) *S is isc at $\bar{y}$ relative to D if and only if for every $x \in S(\bar{y})$ and every neighborhood U of x there exists a neighborhood V of $\bar{y}$ such that $D \cap V \subset S^{-1}(U)$;*

(c) *S is osc at every $y \in \operatorname{dom} S$ if and only if $\operatorname{gph} S$ is closed;*

(d) *S is osc relative to a set $D \subset \mathbb{R}^m$ if and only if $S^{-1}(B)$ is closed relative to D for every compact set $B \subset \mathbb{R}^n$;*

(e) *S is isc relative to a set $D \subset \mathbb{R}^m$ if and only if $S^{-1}(O)$ is open relative to D for every open set $O \subset \mathbb{R}^n$;*

(f) *S is osc at $\bar{y}$ relative to a set $D \subset \mathbb{R}^m$ if and only if the distance function $y \mapsto d(x, S(y))$ is lower semicontinuous at $\bar{y}$ relative to D for every $x \in \mathbb{R}^n$;*

(g) *S is isc at $\bar{y}$ relative to a set $D \subset \mathbb{R}^m$ if and only if the distance function $y \mapsto d(x, S(y))$ is upper semicontinuous at $\bar{y}$ relative to D for every $x \in \mathbb{R}^n$.*

Thus, S is continuous relative to D at $\bar{y}$ if and only if the distance function $y \mapsto d(x, S(y))$ is continuous at $\bar{y}$ relative to D for every $x \in \mathbb{R}^n$.

Proof. Necessity in (a): Suppose that there exists $x \notin S(\bar{y})$ such that for every neighborhood U of x and every neighborhood V of $\bar{y}$ we have $S(y) \cap U \neq \emptyset$ for some $y \in V \cap D$. But then there exists a sequence $y^k \to \bar{y}$, $y^k \in D$ and $x^k \in S(y^k)$ such that $x^k \to x$. This implies that $x \in \limsup_k S(y^k)$, hence $x \in S(\bar{y})$ since S is osc, a contradiction.

Sufficiency in (a): Let $x \notin S(\bar{y})$. Then there exists $\rho > 0$ such that $S(\bar{y}) \cap \mathbb{B}_\rho(x) = \emptyset$; the condition in the second half of (a) then gives a neighborhood V of $\bar{y}$ such that for every $N \in \mathcal{N}$, every sequence $y^k \xrightarrow{N} \bar{y}$ with $y^k \in D \cap V$ has $S(y^k) \cap \mathbb{B}_\rho(x) = \emptyset$. But in that case $d(x, S(y^k)) > \rho/2$ for all large k, which implies, by Proposition 3A.1 and the definition of limsup, that $x \notin \limsup_{y \to \bar{y}} S(y)$. This means that S is osc at $\bar{y}$.

Necessity in (b): Suppose that there exists $x \in S(\bar{y})$ such that for some neighborhood U of x and any neighborhood V of $\bar{y}$ we have $S(y) \cap U = \emptyset$ for some $y \in V \cap D$. Then there is a sequence y^k convergent to $\bar{y}$ in D such that for every sequence $x^k \to x$ one has $x^k \notin S(y^k)$. This means that $x \notin \liminf_{y \to \bar{y}} S(y)$. But then S is not isc at $\bar{y}$.

Sufficiency in (b): If S is not isc at $\bar{y}$ relative to D, then, according to 3A.2(a), there exist an infinite sequence $y^k \to \bar{y}$ in D, a point $x \in S(\bar{y})$ and an open neighborhood U of x such that $S(y^k) \cap U = \emptyset$ for infinitely many k. But then there exists a neighborhood V of $\bar{y}$ such that $D \cap V$ is not in $S^{-1}(U)$ which is the opposite of (b).

(c): S has closed graph if and only if for any $(y, x) \notin \operatorname{gph} S$ there exist open neighborhoods V of y and U of x such that $V \cap S^{-1}(U) = \emptyset$. From (a), this comes down to S being osc at every $y \in \operatorname{dom} S$.

(d): Every sequence in a compact set B has a convergent subsequence, and on the other hand, a set consisting of a convergent sequence and its limit is a compact set. Therefore the condition in the second part of (d) is equivalent to the condition that if $x^k \to \bar{x}$, $y^k \in S^{-1}(x^k)$ and $y^k \to \bar{y}$ with $y^k \in D$, one has $\bar{y} \in S^{-1}(\bar{x})$. But this is precisely the condition for S to be osc relative to D.

(e): Failure of the condition in (e) means the existence of an open set O and a sequence $y^k \to \bar{y}$ in D such that $\bar{y} \in S^{-1}(O)$ but $y^k \notin S^{-1}(O)$; that is, $S(\bar{y}) \cap O \neq \emptyset$ yet $S(y^k) \cap O = \emptyset$ for all k. This last property says that $\liminf_k S(y^k) \not\supset S(\bar{y})$, by 3A.2(a). Hence the condition in (e) fails precisely when S is not isc.

The equivalences in (f) and (g) follow from 3A.2(e) and 3A.2(f). □

Theorem 3B.3 (characterization of Pompeiu–Hausdorff continuity)**.** *A set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is Pompeiu–Hausdorff continuous at $\bar{y}$ if $S(\bar{y})$ is closed and both of the following conditions hold:*

(a) *for every open set $O \subset \mathbb{R}^n$ with $S(\bar{y}) \cap O \neq \emptyset$ there exists a neighborhood V of $\bar{y}$ such that $S(y) \cap O \neq \emptyset$ for all $y \in V$;*

(b) *for every open set $O \subset \mathbb{R}^n$ with $S(\bar{y}) \subset O$ there exists a neighborhood V of $\bar{y}$ such that $S(y) \subset O$ for all $y \in V$.*

Moreover, if S is Pompeiu–Hausdorff continuous at $\bar{y}$, then it is continuous at $\bar{y}$. On the other hand, when $S(\bar{y})$ is nonempty and bounded, Pompeiu–Hausdorff continuity of S at $\bar{y}$ reduces to continuity together with the existence of a neighborhood V of $\bar{y}$ such that $S(V)$ is bounded; in this case conditions (a) and (b) are not only sufficient but also necessary for continuity of S at $\bar{y}$.

Observe that we can define inner semicontinuity of a mapping S at $\bar{y}$ in the Pompeiu–Hausdorff sense by $\lim_{y\to\bar{y}} e(S(\bar{y}),S(y)) = 0$, but this is simply equivalent to inner semicontinuity in the Painlevé–Kuratowski sense (compare 3B.2(b) and 3B.3(b)). In contrast, if we define outer semicontinuity in the Pompeiu–Hausdorff sense by $\lim_{y\to\bar{y}} e(S(y),S(\bar{y})) = 0$, we get a generally much more restrictive concept than outer semicontinuity in the Painlevé–Kuratowski sense.

We present next two applications of these concepts to mappings that play central roles in optimization.

Example 3B.4 (solution mapping for a system of inequalities)**.** Consider a mapping defined implicitly by a parameterized system of inequalities, that is,

$$S : p \mapsto \left\{ x \,\middle|\, f_i(p,x) \le 0, i = 1,\dots,m \right\} \text{ for } p \in I\!R^d.$$

Assume that each f_i is a continuous real-valued function on $I\!R^d \times I\!R^n$. Then S is osc at every point of its domain. If moreover each f_i is convex in x for each p and $\bar{p}$ is such that there exists $\bar{x}$ with $f_i(\bar{p},\bar{x}) < 0$ for each $i = 1,\dots,m$, then S is continuous at $\bar{p}$.

Detail. The graph of S is the intersection of the sets $\left\{ (p,x) \,\middle|\, f_i(p,x) \le 0 \right\}$, which are closed by the continuity of f_i. Then $\operatorname{gph} S$ is closed, and the osc property comes from Theorem 3B.2(c). The isc part will follow from a much more general result (Robinson-Ursescu theorem) which we present in Chapter 5.

Applications in optimization. Consider the following general problem of minimization, involving a parameter p which ranges over a set $P \subset I\!R^d$, a function $f_0 : I\!R^d \times I\!R^n \to I\!R$, and a mapping $S_{\rm feas} : P \rightrightarrows I\!R^n$:

$$\text{minimize } f_0(p,x) \text{ over all } x \in I\!R^n \text{ satisfying } x \in S_{\rm feas}(p).$$

Here f_0 is the *objective function* and $S_{\rm feas}$ is the *feasible set mapping* (with $S_{\rm feas}(p)$ taken to be the empty set when $p \notin P$). In particular, $S_{\rm feas}$ could be specified by constraints in the manner of Example 3B.4, but we now allow it to be more general.

Our attention is focused now on two other mappings in this situation: the *optimal value* mapping acting from $I\!R^d$ to $I\!R$ and defined by

$$S_{\rm val} : p \mapsto \inf_x \left\{ f_0(p,x) \,\middle|\, x \in S_{\rm feas}(p) \right\} \text{ when the inf is finite,}$$

and the *optimal set* mapping acting from P to $I\!R^n$ and defined by

$$S_{\rm opt} : p \mapsto \left\{ x \in S_{\rm feas}(p) \,\middle|\, f_0(p,x) = S_{\rm val}(p) \right\}.$$

Theorem 3B.5 (basic continuity properties of solution mappings in optimization)**.** *In the preceding notation, let $\bar{p} \in P$ be fixed with the feasible set $S_{\text{feas}}(\bar{p})$ nonempty and bounded, and suppose that:*

(a) *the mapping S_{feas} is Pompeiu–Hausdorff continuous at $\bar{p}$ relative to P, or equivalently, S_{feas} is continuous at $\bar{p}$ relative to P with $S_{\text{feas}}(Q \cap P)$ bounded for some neighborhood Q of $\bar{p}$,*

(b) *the function f_0 is continuous relative to $P \times \mathbb{R}^n$ at $(\bar{p}, \bar{x})$ for every $\bar{x} \in S_{\text{feas}}(\bar{p})$.*

Then the optimal value mapping S_{val} is continuous at $\bar{p}$ relative to P, whereas the optimal set mapping S_{opt} is osc at $\bar{p}$ relative to P.

Proof. The equivalence in assumption (a) comes from the final statement in Theorem 3B.3. In particular (a) implies $S_{\text{feas}}(\bar{p})$ is closed, hence from boundedness actually compact. Then too, since $f_0(\bar{p}, \cdot)$ is continuous on $S_{\text{feas}}(\bar{p})$ by (b), the set $S_{\text{opt}}(\bar{p})$ is nonempty.

Let $\bar{x} \in S_{\text{opt}}(\bar{p})$. From (a) we get for any sequence $p^k \to \bar{p}$ in P the existence of a sequence of points x^k with $x^k \in S_{\text{feas}}(p^k)$ such that $x^k \to \bar{x}$ as $k \to \infty$. But then, for every $\varepsilon > 0$ there exists $N \in \mathcal{N}$ such that

$$S_{\text{val}}(p^k) \le f_0(p^k, x^k) \le f_0(\bar{p}, \bar{x}) + \varepsilon = S_{\text{val}}(\bar{p}) + \varepsilon \qquad \text{for } k \in N.$$

This gives us

$$\limsup_{p \to \bar{p}} S_{\text{val}}(p) \le S_{\text{val}}(\bar{p}). \tag{1}$$

On the other hand, let us assume that

$$\liminf_{p \to \bar{p}} S_{\text{val}}(p) < S_{\text{val}}(\bar{p}). \tag{2}$$

Then there exist $\varepsilon > 0$ and sequences $p_k \to \bar{p}$ in P and $x^k \in S_{\text{feas}}(p^k)$, $k \in \mathbb{N}$, such that

$$f_0(p^k, x^k) < S_{\text{val}}(\bar{p}) - \varepsilon \quad \text{for all } k. \tag{3}$$

From (a) we see that $d(x^k, S_{\text{feas}}(\bar{p})) \to 0$ as $k \to \infty$. This provides the existence of a sequence of points $\bar{x}^k \in S_{\text{feas}}(\bar{p})$ such that $|x^k - \bar{x}^k| \to 0$ as $k \to \infty$. Because $S_{\text{feas}}(\bar{p})$ is compact, there must be some $\bar{x} \in S_{\text{feas}}(\bar{p})$ along with an index set $N \in \mathcal{N}^\sharp$ such that $\bar{x}^k \xrightarrow{N} \bar{x}$, in which case $x^k \xrightarrow{N} \bar{x}$ as well. Then, from the continuity of f_0 at $(\bar{p}, \bar{x})$, we have $f_0(\bar{p}, \bar{x}) \le f_0(p^k, x^k) + \varepsilon$ for $k \in N$ and sufficiently large, which, together with (3), implies for such k that

$$S_{\text{val}}(\bar{p}) \le f_0(\bar{p}, \bar{x}) \le f_0(p^k, x^k) + \varepsilon < S_{\text{val}}(\bar{p}).$$

The contradiction obtained proves that (2) is false. Thus,

$$S_{\text{val}}(\bar{p}) \le \liminf_{p \to \bar{p}} S_{\text{val}}(p),$$

which, combined with (1), gives us the continuity of the optimal value mapping S_{val} at $\bar{p}$ relative to P.

To show that S_{opt} is osc at $\bar{p}$ relative to P, we use the equivalent condition in 3B.2(a). Suppose there exists $x \notin S_{\mathrm{opt}}(\bar{p})$ such that for every neighborhoods U of x and Q of $\bar{p}$ there exist $u \in U$ and $p \in Q \cap P$ such that $u \in S_{\mathrm{opt}}(p)$. This is the same as saying that there are sequences $p^k \to \bar{p}$ in P and $u^k \to x$ as $k \to \infty$ such that $u^k \in S_{\mathrm{opt}}(p^k)$. Note that since $u^k \in S_{\mathrm{feas}}(p^k)$ we have that $x \in S_{\mathrm{feas}}(\bar{p})$. But then, as we already proved,

$$f_0(p^k, u^k) = S_{\mathrm{val}}(p^k) \to S_{\mathrm{val}}(\bar{p}) \quad \text{as } k \to \infty.$$

By continuity of f_0, the left side tends to $f_0(\bar{p}, x)$ as $k \to \infty$, which means that $x \in S_{\mathrm{opt}}(\bar{p})$, a contradiction. □

Example 3B.6 (minimization over a fixed set)**.** Let X be a nonempty, compact subset of $\mathbb{R}^n$ and let f_0 be a continuous function from $P \times X$ to $\mathbb{R}$, where P is a nonempty subset of $\mathbb{R}^d$. For each $p \in P$, let

$$S_{\mathrm{val}}(p) = \min_{x \in X} f_0(p,x), \qquad S_{\mathrm{opt}}(p) = \operatorname*{argmin}_{x \in X} f_0(p,x).$$

Then the function $S_{\mathrm{val}} : P \to \mathbb{R}$ is continuous relative to P, and the mapping $S_{\mathrm{opt}} : P \rightrightarrows \mathbb{R}^n$ is osc relative to P.

Detail. This exploits the case of Theorem 3B.5 where S_{feas} is the constant mapping $p \mapsto X$. □

Example 3B.7 (continuity of the feasible set versus continuity of the optimal value)**.** Consider the minimization of $f_0(x_1, x_2) = e^{x_1} + x_2^2$ on the set

$$S_{\mathrm{feas}}(p) = \left\{ (x_1, x_2) \in \mathbb{R}^2 \;\middle|\; -\frac{x_1}{(1+x_1^2)} - p \le x_2 \le \frac{x_1}{(1+x_1^2)} + p \right\}.$$

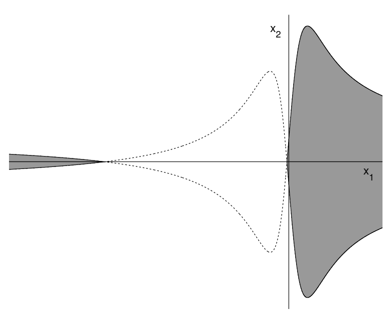

Fig. 3.2 The feasible set in Example 3B.7 for $p = 0.1$.

For parameter value $p=0$, the optimal value $S_{\mathrm{val}}(0)=1$ and occurs at $S_{\mathrm{opt}}(0)=(0,0)$, but for $p>0$, see Fig. 3.2, the asymptotics of the function $x_1/(1+x_1^2)$ open up a "phantom" portion of the feasible set along the negative x_1-axis, and the optimal value is 0. The feasible set nonetheless does depend continuously on p in $[0,\infty)$ in the Painlevé–Kuratowski sense.

3C. Lipschitz Continuity of Set-valued Mappings

A quantitative notation of continuity for set-valued mappings can be formulated with the help of the Pompeiu–Hausdorff distance between sets in the same way that Lipschitz continuity is defined for functions. It has important uses, although it suffers from shortcomings when the sets may be unbounded. Here we invoke the terminology that a set-valued mapping S is *closed-valued* on a set D when $S(y)$ is a closed set for each $y \in D$.

Lipschitz continuity of set-valued mappings. *A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to be Lipschitz continuous relative to a (nonempty) set D in $\mathbb{R}^m$ if $D \subset \operatorname{dom} S$, S is closed-valued on D, and there exists $\kappa \geq 0$ (Lipschitz constant) such that*

(1) $$h(S(y'),S(y)) \leq \kappa|y'-y| \quad \text{for all } y', y \in D,$$

or equivalently, there exists $\kappa \geq 0$ such that

(2) $$S(y') \subset S(y) + \kappa|y'-y|\mathbb{B} \quad \text{for all } y', y \in D.$$

When S is single-valued on D, we obtain from this definition the notion of Lipschitz continuity of a function introduced in Section 1D.

One could contemplate defining Lipschitz continuity of a set-valued mapping S without requiring S to be closed-valued, relying in that case simply on (1) or (2). That might be workable, although κ in (2) could then be slightly larger than the κ in (1), but a fundamental objection arises. A mapping that is continuous necessarily does have closed values, so we would be in the position of having a concept of Lipschitz continuity which did not entail continuity. That is a paradox we prefer to avoid. The issue is absent for single-valued mappings, since they are trivially closed-valued.

When a mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is Lipschitz continuous on a closed set D then clearly the set $\operatorname{gph} S \cap (D \times \mathbb{R}^n)$ is closed. Lipschitz continuity of a set-valued mapping can be characterized by a property which relates the distances to its value and the value of the inverse mapping:

Proposition 3C.1 (distance characterization of Lipschitz continuity)**.** *Consider a closed-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ and a nonempty subset D of $\operatorname{dom} S$. Then S is Lipschitz continuous relative to D with constant κ if and only if*

(3) $$d(x,S(y)) \le \kappa d(y,S^{-1}(x)\cap D) \quad \text{for all } x \in \mathbb{R}^n \text{ and } y \in D.$$

Proof. Let S be Lipschitz continuous relative to D with a constant κ and let $x \in \mathbb{R}^n$ and $y \in D$. If $S^{-1}(x) \cap D = \emptyset$, the inequality (3) holds automatically. Let $S^{-1}(x) \cap D \neq \emptyset$ and choose $\varepsilon > 0$. Then there exists $y' \in S^{-1}(x) \cap D$ with $|y' - y| \le d(y,S^{-1}(x)\cap D) + \varepsilon$. By (1),

$$d(x,S(y)) \le h(S(y'),S(y)) \le \kappa|y'-y| \le \kappa d(y,S^{-1}(x)\cap D) + \kappa\varepsilon.$$

Since the left side of this inequality does not depend on ε, passing to zero with ε we conclude that (3) holds with the κ of (1).

Conversely, let (3) hold, let $y, y' \in D \subset \operatorname{dom} S$ and let $x \in S(y)$. Then

$$d(x,S(y')) \le \kappa d(y',S^{-1}(x)\cap D) \le \kappa|y-y'|,$$

since $y \in S^{-1}(x) \cap D$. Taking the supremum with respect to $x \in S(y)$, we obtain $e(S(y),S(y')) \le \kappa|y-y'|$ and, by symmetry, we get (1). □

For the inverse mapping $F = S^{-1}$ the property described in (3) can be written as

$$d(x,F^{-1}(y)) \le \kappa d(y,F(x)\cap D) \quad \text{for all } x \in \mathbb{R}^n,\ y \in D,$$

and when $\operatorname{gph} F$ is closed this can be interpreted in the following manner. Whenever we pick a $y \in D$ and an $x \in \operatorname{dom} F$, the distance from x to the set of solutions u of the inclusion $y \in F(u)$ is proportional to $d(y,F(x)\cap D)$, which measures the extent to which x itself fails to solve this inclusion. In Section 3E we will introduce a local version of this property which plays a major role in variational analysis and is known as "metric regularity."

The difficulty with the concept of Lipschitz continuity for set-valued mappings S with values $S(y)$ that may be unbounded comes from the fact that usually $h(C_1,C_2) = \infty$ when C_1 or C_2 is unbounded, the only exceptions being cases where both C_1 and C_2 are unbounded and "the unboundedness points in the same direction." For instance, when C_1 and C_2 are lines in $\mathbb{R}^2$, one has $h(C_1,C_2) < \infty$ only when these lines are parallel.

In the remainder of this section we consider a particular class of set-valued mappings, with significant applications in variational analysis, which are automatically Lipschitz continuous even when their values are unbounded sets.

Polyhedral convex mappings. *A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to be polyhedral convex if its graph is a polyhedral convex set.*

Here it should be recalled from Section 2E that a set is polyhedral convex if it can be expressed as the intersection of a finite collection of closed half-spaces and/or hyperplanes.

Example 3C.2 (polyhedral convex mappings from linear constraint systems)**.** A solution mapping S of the form in Example 3B.4 is polyhedral convex when the f_i

there are all affine; furthermore, this continues to be true when some or all of the constraints are equations instead of inequalities.

In the notational context of elements $x \in S(y)$ for a mapping $S : I\!R^m \rightrightarrows I\!R^n$, polyhedral convexity of S is equivalent to the existence of a positive integer r, matrices $D \in I\!R^{r\times n}$, $E \in I\!R^{r\times m}$, and a vector $q \in I\!R^r$ such that

$$S(y) = \left\{ x \in I\!R^n \,\middle|\, Dx + Ey \le q \right\} \quad \text{for all } y \in I\!R^m. \tag{4}$$

Note for instance that any mapping S whose graph is a linear subspace is a polyhedral convex mapping.

Theorem 3C.3 (Lipschitz continuity of polyhedral convex mappings). *Any polyhedral convex mapping $S : I\!R^m \rightrightarrows I\!R^n$ is Lipschitz continuous relative to its domain.*

We will prove this theorem by using a fundamental result due to A. J. Hoffman regarding approximate solutions of systems of linear inequalities. For a vector $a = (a_1, a_2, \ldots, a_n) \in I\!R^n$, we use the vector notation that

$$a_+ = (\max\{0, a_1\}, \ldots, \max\{0, a_n\}).$$

Also, recall that the *convex hull* of a set $C \subset I\!R^n$, which will be denoted by $\operatorname{co} C$, is the smallest convex set that includes C. (It can be identified as the intersection of all convex sets that include C, but also can be described as consisting of all linear combinations $\lambda_0 x_0 + \lambda_1 x_1 + \cdots + \lambda_n x_n$ with $x_i \in C$, $\lambda_i \ge 0$, and $\lambda_0 + \lambda_1 + \cdots + \lambda_n = 1$; this is Carathéodory's theorem.) The *closed convex hull* of C is the closure of the convex hull of C and denoted $\operatorname{cl}\operatorname{co} C$; it is the smallest closed convex set that contains C.

Lemma 3C.4 (Hoffman lemma). *For the set-valued mapping*

$$S : y \mapsto \left\{ x \in I\!R^n \,\middle|\, Ax \le y \right\} \quad \text{for } y \in I\!R^m,$$

where A is a nonzero $m \times n$ matrix, there exists a constant L such that

$$d(x, S(y)) \le L|(Ax - y)_+| \quad \text{for every } y \in \operatorname{dom} S \text{ and every } x \in I\!R^n. \tag{5}$$

Proof. For any $y \in \operatorname{dom} S$ the set $S(y)$ is nonempty, convex and closed, hence every point $x \notin S(y)$ has a unique (Euclidean) projection $u = P_{S(y)}(x)$ on $S(y)$ (Proposition 1D.5):

$$u \in S(y),\ |u - x| = d(x, S(y)). \tag{6}$$

As noted in Section 2A, the projection mapping satisfies

$$P_{S(y)} = (I + N_{S(y)})^{-1},$$

where $N_{S(y)}$ is the normal cone mapping to the convex set $S(y)$. In these terms, the problem of projecting x on $S(y)$ is equivalent to that of finding the unique $u \neq x$ such that

$$x \in u + N_{S(y)}(u).$$

The formula in 2E(8) gives us a representation of the normal cone to a polyhedral convex set specified by affine inequalities, which here comes out as

$$N_{S(y)}(u) = \left\{ v \,\middle|\, v = \textstyle\sum_{i=1}^m \lambda_i a_i \text{ with } \lambda_i \geq 0,\ \lambda_i(\langle a_i, u\rangle - y_i) = 0,\ i = 1, \ldots, m \right\},$$

where the a_i's are the rows of the matrix A regarded as vectors in $\mathbb{R}^n$. Thus, the projection u of x on $S(y)$, as described by (6), can be obtained by finding a pair (u, λ) such that

$$\text{(7)} \qquad \begin{cases} x - u - \sum_{i=1}^m \lambda_i a_i = 0, \\ \lambda_i \geq 0,\ \lambda_i(\langle a_i, u\rangle - y_i) = 0, \quad i = 1, \ldots, m. \end{cases}$$

While the projection u exists and is unique, this variational inequality might not have a unique solution (u, λ) because the λ component might not be unique. But since $u \neq x$ (through our assumption that $x \notin S(y)$), we can conclude from the first relation in (7) that for any solution (u, λ) the vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ is not the zero vector. Consider the family $\mathscr{J}$ of subsets J of $\{1, \ldots, m\}$ for which there are real numbers $\lambda_1, \ldots, \lambda_m$ with $\lambda_i > 0$ for $i \in J$ and $\lambda_i = 0$ for $i \notin J$ and such that (u, λ) satisfies (7). Of course, if $\langle a_i, u\rangle - y_i < 0$ for some i, then $\lambda_i = 0$ according to the second relation (complementarity) in (7), and then this i cannot be an element of any J. That is,

$$\text{(8)} \qquad J \in \mathscr{J} \text{ and } i \in J \quad \Longrightarrow \quad \langle a_i, u\rangle = y_i \text{ and } \lambda_i > 0.$$

Since the set of vectors λ such that (u, λ) solves (7) does not contain the zero vector, we have $\mathscr{J} \neq \emptyset$.

We will now prove that there is a nonempty index set $\bar{J} \in \mathscr{J}$ for which there are no numbers β_i, $i \in \bar{J}$ satisfying

$$\text{(9)} \qquad \beta_i \geq 0,\ i \in \bar{J}, \quad \sum_{j \in \bar{J}} \beta_i > 0 \text{ and } \sum_{i \in \bar{J}} \beta_i a_i = 0.$$

On the contrary, suppose that for every $J \in \mathscr{J}$ this is not the case, that is, (9) holds with $\bar{J} = J$ for some β_i, $i \in J$. Let J' be a set in $\mathscr{J}$ with a minimal number of elements (J' might be not unique). Note that the number of elements in any J' is greater than 1. Indeed, if there were just one element i' in J', then we would have $\beta_{i'} a_{i'} = 0$ and $\beta_{i'} > 0$, hence $a_{i'} = 0$, and then, since (7) holds for (u, λ) such that $\lambda_i = \beta_i$, $i = i'$, $\lambda_i = 0$, $i \neq i'$, from the first equality in (7) we would get $x = u$ which contradicts the assumption that $x \notin S(y)$. Since $J' \in \mathscr{J}$, there are $\lambda_i' > 0$, $i \in J'$ such that

$$\text{(10)} \qquad x - u = \sum_{i \in J'} \lambda_i' a_i.$$

By assumption, there are also real numbers $\beta_i' \geq 0$, $i \in J'$, such that

$$\sum_{i\in J'} \beta_i' > 0 \quad \text{and} \quad \sum_{i\in J'} \beta_i' a_i = 0. \tag{11}$$

Multiplying both sides of the equality in (11) by a positive scalar t and adding to (10), we obtain

$$x - u = \sum_{i\in J'} (\lambda_i' - t\beta_i')a_i.$$

Let

$$t_0 = \min_i \left\{ \frac{\lambda_i'}{\beta_i'} \;\middle|\; i \in J' \text{ with } \beta_i' > 0 \right\}.$$

Then for any $k \in J'$ for which this minimum is attained, we have

$$\lambda_i' - t_0\beta_i' \geq 0 \quad \text{for every } i \in J' \setminus k \quad \text{and} \quad x - u = \sum_{i\in J'\setminus k} (\lambda_i' - t\beta_i')a_i.$$

Thus, the vector $\lambda \in \mathbb{R}^m$ with components $\lambda_i' - t_0\beta_i'$ when $i \in J'$ and $\lambda_i' = 0$ when $i \notin J'$ is such that (u,λ) satisfies (7). Hence, we found a nonempty index set $J'' \in \mathscr{J}$ having fewer elements than J', which contradicts the choice of J'. The contradiction obtained proves that there is a nonempty index set $\bar{J} \in \mathscr{J}$ for which there are no numbers β_i, $i \in J$, satisfying (8). In particular, the zero vector in $\mathbb{R}^n$ is not in the convex hull $\operatorname{co}\{a_j, j \in \bar{J}\}$.

Let $\bar{\lambda}_i > 0$, $i \in \bar{J}$, be the corresponding vector of multipliers such that, if we set $\bar{\lambda}_i = 0$ for $i \notin \bar{J}$, we have that $(u,\bar{\lambda})$ is a solution of (7). Since $\sum_{j\in\bar{J}} \bar{\lambda}_i a_i \neq 0$, because otherwise (9) would hold for $\beta_i = \bar{\lambda}_i$, we have

$$\gamma := \sum_{i\in\bar{J}} \bar{\lambda}_i > 0.$$

Because (7) holds with $(u,\bar{\lambda})$, using (7) and (8) we have

$$\begin{aligned}
d(0, \operatorname{co}\{a_j, j \in \bar{J}\})|x-u| &\leq \Big|\sum_{i\in\bar{J}} \frac{\bar{\lambda}_i}{\gamma} a_i\Big||x-u| = \frac{1}{\gamma}|x-u||x-u| \\
&= \Big\langle \frac{1}{\gamma}(x-u), x-u \Big\rangle = \Big\langle \frac{1}{\gamma}\Big(\sum_{i\in\bar{J}} \bar{\lambda}_i a_i\Big), x-u \Big\rangle \\
&= \sum_{i\in\bar{J}} \frac{\bar{\lambda}_i}{\gamma}(\langle a_i, x\rangle - \langle a_i, u\rangle) = \sum_{i\in\bar{J}} \frac{\bar{\lambda}_i}{\gamma}(\langle a_i, x\rangle - y_i) \\
&\leq \max_{i\in\bar{J}} \{(\langle a_i, x\rangle - y_i)_+\}.
\end{aligned}$$

Hence, for some constant c independent of x and y we have

$$d(x, S(y)) = |x-u| \leq c \max_{1\leq i\leq m} \{(\langle a_i, x\rangle - y_i)_+\}.$$

This inequality remains valid (perhaps with a different constant c) after passing from the max vector norm to the equivalent Euclidean norm. This proves (5). □

Proof of Theorem 3C.3. Let $y, y' \in \operatorname{dom} S$ and let $x \in S(y)$. Since S is polyhedral, from the representation (4) we have $Dx + Ey - q \le 0$ and then

$$Dx + Ey' - q = Dx + Ey - q - Ey + Ey' \le -Ey + Ey'. \tag{12}$$

Then from Lemma 3C.4 we obtain the existence of a constant L such that

$$d(x, S(y')) \le L|(Dx + Ey' - q)_+|,$$

and hence, by (12),

$$d(x, S(y')) \le L|(E(y' - y))_+| \le L|E(y - y')|.$$

Since x is arbitrarily chosen in $S(y)$, this leads to

$$e(S(y), S(y')) \le \kappa |y - y'|$$

with $\kappa = L|E|$. The same must hold with the roles of y and y' reversed, and in consequence S is Lipschitz continuous on $\operatorname{dom} S$. □

Applications to solution mappings in linear programming. Consider the following problem of linear programming in which y acts as a parameter:

$$\text{minimize } \langle c, x\rangle \ \text{ over all } \ x \in I\!R^n \ \text{ satisfying } Ax \le y. \tag{13}$$

Here c is a fixed vector in $I\!R^n$, A is a fixed matrix in $I\!R^{m\times n}$. Define the solution mappings associated with (13) as in Section 3B, that is, the feasible set mapping

$$S_{\text{feas}} : y \mapsto \{\, x \,|\, Ax \le y \,\}, \tag{14}$$

the optimal value mapping

$$S_{\text{val}} : y \mapsto \inf_x \{\, \langle c, x\rangle \,|\, Ax \le y \,\} \ \text{ when the inf is finite,} \tag{15}$$

and the optimal set mapping by

$$S_{\text{opt}} : y \mapsto \{\, x \in S_{\text{feas}}(y) \,|\, \langle c, x\rangle = S_{\text{val}}(y) \,\}. \tag{16}$$

It is known from the theory of linear programming that $S_{\text{opt}}(y) \ne \emptyset$ when the infimum in (15) is finite (and only then).

Exercise 3C.5 (Lipschitz continuity of mappings in linear programming)**.** *Establish that the mappings in (14), (15) and (16) are Lipschitz continuous relative to their domains, the domain in the case of (15) and (16) being the set D consisting of all y for which the infimum in (15) is finite.*

Guide. Derive the Lipschitz continuity of S_{feas} from Theorem 3C.3, out of the connection with Example 3C.2. Let κ be a Lipschitz constant for S_{feas}.

Next, for the case of S_{val}, consider any $y, y' \in D$ and any $x \in S_{\text{opt}}(y)$, which exists because S_{opt} is nonempty when $y \in D$. In particular we have $x \in S_{\text{feas}}(y)$. From the Lipschitz continuity of S_{feas}, there exists $x' \in S_{\text{feas}}(y')$ such that $|x - x'| \leq \kappa|y - y'|$. Use this along with the fact that $S_{\text{val}}(y') \leq \langle c, x'\rangle$ but $S_{\text{val}}(y) = \langle c, x\rangle$ to get a bound on $S_{\text{val}}(y') - S_{\text{val}}(y)$ which confirms the Lipschitz continuity claimed for S_{val}.

For the case of S_{opt}, consider the set-valued mapping

$$G : (y,t) \mapsto \left\{ x \in \mathbb{R}^n \,\middle|\, Ax \leq y,\ \langle c, x\rangle \leq t \right\} \ \text{ for } (y,t) \in \mathbb{R}^m \times \mathbb{R}.$$

Confirm that this mapping is polyhedral convex and apply Theorem 3C.3 to it. Observe that $S_{\text{opt}}(y) = G(y, S_{\text{val}}(y))$ for $y \in D$ and invoke the Lipschitz continuity of S_{val}. □

3D. Outer Lipschitz Continuity

In this section we define a "one-point" property of set-valued mappings by fixing one of the points y and y' in the definition of Lipschitz continuity at its reference value $\bar{y}$. Then these points no longer play symmetric roles, so we use the excess instead of the Pompeiu–Hausdorff distance.

Outer Lipschitz continuity. *A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to be outer Lipschitz continuous at $\bar{y}$ relative to a set D if $\bar{y} \in D \subset \operatorname{dom} S$, $S(\bar{y})$ is a closed set, and there is a constant $\kappa \geq 0$ along with a neighborhood V of $\bar{y}$ such that*

$$(1) \qquad e(S(y), S(\bar{y})) \leq \kappa|y - \bar{y}| \quad \text{for all } y \in V \cap D,$$

or equivalently

$$(2) \qquad S(y) \subset S(\bar{y}) + \kappa|y - \bar{y}|\mathbb{B} \quad \text{for all } y \in V \cap D.$$

If S is outer Lipschitz continuous at every point $y \in D$ relative to D with the same κ, then S is said to be outer Lipschitz continuous relative to D.

It is clear that any mapping which is Lipschitz continuous relative to a set D with constant κ is also outer Lipschitz continuous relative to D with constant κ, but the converse may not be true. Also, outer Lipschitz continuity at a point $\bar{y}$ implies outer semicontinuity at $\bar{y}$. For single-valued mappings, outer Lipschitz continuity becomes the property of calmness which we considered in Section 1C. The examples in Section 1D show how very different this property is from the Lipschitz continuity.

The condition in the definition that the mapping is closed-valued at $\bar{y}$ could be dropped; but then the constant κ in (2) might be slightly larger than the one in (1), and furthermore outer Lipschitz continuity might not entail outer semicontinuity (where closed-valuedness is essential). Therefore, we hold back from such an extension.

We present next a result which historically was the main motivation for introducing the property of outer Lipschitz continuity and which complements Theorem 3C.3. It uses the following concept.

Polyhedral mappings. *A set-valued mapping* $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *will be called polyhedral if* gph S *is the union of finitely many sets that are polyhedral convex in* $\mathbb{R}^n \times \mathbb{R}^m$.

Clearly, a polyhedral mapping has closed graph, since polyhedral convex sets are closed, and hence is osc and in particular closed-valued everywhere. Any polyhedral *convex* mapping as defined in 3C is obviously a polyhedral mapping, but the graph then is comprised of only one "piece," whereas now we are allowing a multiplicity of such polyhedral convex "pieces," which furthermore could overlap.

Theorem 3D.1 (outer Lipschitz continuity of polyhedral mappings)**.** *Any polyhedral mapping* $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *is outer Lipschitz continuous at every point of its domain.*

Proof. Let gph $S = \bigcup_{i=1}^k G_i$ where the G_i's are polyhedral convex sets in $\mathbb{R}^m \times \mathbb{R}^n$. For each i define the mapping

$$S_i : y \mapsto \left\{ x \,\middle|\, (y,x) \in G_i \right\} \quad \text{for } y \in \mathbb{R}^m.$$

Then each S_i is Lipschitz continuous on its domain, according to Theorem 3C.3. Let $\bar{y} \in \operatorname{dom} S$ and let

$$\mathscr{J} = \left\{ i \,\middle|\, \text{there exists } x \in \mathbb{R}^n \text{ with } (\bar{y},x) \in G_i \right\}.$$

Then $\bar{y} \in \operatorname{dom} S_i$ for each $i \in \mathscr{J}$, and moreover,

$$S(\bar{y}) = \bigcup_{i \in \mathscr{J}} S_i(\bar{y}). \tag{3}$$

For any $i \notin \mathscr{J}$, since the sets $\{\bar{y}\} \times \mathbb{R}^n$ and G_i are disjoint and polyhedral convex, there is a neighborhood V_i of $\bar{y}$ such that $(V_i \times \mathbb{R}^n) \bigcap G_i = \emptyset$. Let $V = \bigcap_{i \notin \mathscr{J}} V_i$. Then of course V is a neighborhood of $\bar{y}$ and we have

$$(V \times \mathbb{R}^n) \bigcap \operatorname{gph} S \subset \bigcup_{i=1}^k G_i \setminus \bigcup_{i \notin \mathscr{J}} G_i \subset \bigcup_{i \in \mathscr{J}} G_i. \tag{4}$$

Let $y \in V$. If $S(y) = \emptyset$, then the relation (1) holds trivially. Let x be any point in $S(y)$. Then from (4),

$$(y,x) \in (V \times \mathbb{R}^n) \bigcap \operatorname{gph} S \subset \bigcup_{j \in \mathscr{J}} G_i,$$

hence for some $i \in \mathscr{J}$ we have $(y,x) \in G_i$, that is, $x \in S_i(y)$. Since each S_i is Lipschitz continuous and $\bar{y} \in \operatorname{dom} S_i$, with constant κ_i, say, we obtain by using (3) that

$$d(x,S(\bar{y})) \le \max_i d(x,S_i(\bar{y})) \le \max_i e(S_i(y),S_i(\bar{y})) \le \max_i \kappa_i|y-\bar{y}|.$$

Since x is an arbitrary point in $S(y)$, we conclude that S is outer Lipschitz continuous at $\bar{y}$ with constant $\kappa := \max_i \kappa_i$. □

Exercise 3D.2 (polyhedrality of solution mappings to linear variational inequalities). *Given an $n \times n$ matrix A and a polyhedral convex set C in $\mathbb{R}^n$, show that the solution mapping of the linear variational inequality*

$$y \mapsto S(y) = \{\, x \mid y \in Ax + N_C(x) \,\} \quad \textit{for } y \in \mathbb{R}^n$$

is polyhedral, and therefore it is outer Lipschitz continuous relative to its domain.

Guide. Any polyhedral convex set C is representable (in a non-unique manner) by a system of affine inequalities:

$$C = \{\, x \mid \langle a_i, x\rangle \le \alpha_i \quad \text{for } i = 1,2,\dots,m \,\}.$$

We know from Section 2E that the normal cone to C at the point $x \in C$ is the set

$$N_C(x) = \left\{ u \,\middle|\, u = \textstyle\sum_{i=1}^m y_i a_i,\ y_i \ge 0 \text{ for } i \in I(x),\ y_i = 0 \text{ for } i \notin I(x) \right\},$$

where $I(x) = \{\, i \mid \langle a_i, x\rangle = \alpha_i \,\}$ is the active index set for $x \in C$. The graph of the normal cone mapping N_C is not convex, unless C is a translate of a subspace, but it is the union, with respect to all possible subsets J of $\{1,\dots,m\}$, of the polyhedral convex sets

$$\left\{ (x,u) \,\middle|\, u = \textstyle\sum_{i=1}^m y_i a_i,\ \langle a_i, x\rangle = \alpha_i,\ y_i \ge 0 \text{ if } i \in J,\ \langle a_i, x\rangle < \alpha_i,\ y_i = 0 \text{ if } i \notin J \right\}.$$

It remains to observe that the graph of the sum $A + N_C$ is also the union of polyhedral convex sets. □

Outer Lipschitz continuity becomes automatically Lipschitz continuity when the mapping is inner semicontinuous, a property we introduced in the preceding section.

Theorem 3D.3 (isc criterion for Lipschitz continuity). *Consider a set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ and a convex set $D \subset \operatorname{dom} S$ such that $S(y)$ is closed for every $y \in D$. Then S is Lipschitz continuous relative to D with constant κ if and only if S is both inner semicontinuous (isc) relative to D and outer Lipschitz continuous relative to D with constant κ.*

Proof. Let S be inner semicontinuous and outer Lipschitz continuous with constant κ, both relative to D. Choose $y, y' \in D$ and let $y_t = (1-t)y + ty'$. The assumed outer Lipschitz continuity together with the closedness of the values of S implies that for each $t \in [0,1]$ there exists a positive r_t such that

$$S(u) \subset S(y_t) + \kappa|u - y_t|\mathbb{B} \quad \text{for all } u \in D \cap \mathbb{B}_{r_t}(y_t).$$

Let

$$\tau = \sup\big\{\, t \in [0,1] \,\big|\, S(y_s) \subset S(y) + \kappa|y_s - y|\mathbb{B} \quad \text{for each } s \in [0,t] \,\big\}. \tag{5}$$

We will show that the supremum in (5) is attained at $\tau = 1$.

First, note that $\tau > 0$ because $r_0 > 0$. Since $S(y)$ is closed, the set $S(y) + \kappa|y_\tau - y|\mathbb{B}$ is closed too, thus its complement, denoted O, is open. Suppose that $y_\tau \in S^{-1}(O)$; then, applying Theorem 3B.2(e) to the isc mapping S, we obtain that there exists $\sigma \in [0,\tau)$ such that $y_\sigma \in S^{-1}(O)$ as well. But this is impossible since from $\sigma < \tau$ we have

$$S(y_\sigma) \subset S(y) + \kappa|y_\sigma - y|\mathbb{B} \subset S(y) + \kappa|y_\tau - y|\mathbb{B}.$$

Hence, $y_\tau \notin S^{-1}(O)$, that is, $S(y_\tau) \cap O = \emptyset$ and therefore $S(y_\tau)$ is a subset of $S(y) + \kappa|y_\tau - y|\mathbb{B}$. This implies that the supremum in (5) is attained.

Let us next prove that $\tau = 1$. If $\tau < 1$ there must exist $\eta \in (\tau, 1)$ with $|y_\eta - y_\tau| < r_\tau$ such that

$$S(y_\eta) \not\subset S(y) + \kappa|y_\eta - y|\mathbb{B}. \tag{6}$$

But then, from the definition of r_τ,

$$S(y_\eta) \subset S(y_\tau) + \kappa|y_\eta - y_\tau|\mathbb{B} \subset S(y) + \kappa(|y_\eta - y_\tau| + |y_\tau - y|)\mathbb{B} = S(y) + \kappa|y_\eta - y|\mathbb{B},$$

where the final equality holds because y_τ is a point in the segment $[y, y_\eta]$. This contradicts (6), hence $\tau = 1$. Putting $\tau = 1$ into (5) results in $S(y') \subset S(y) + \kappa|y' - y|$. By the symmetry of y and y', we obtain that S is Lipschitz continuous relative to D.

Conversely, if S is Lipschitz continuous relative to D, then S is of course outer Lipschitz continuous. Let now $y \in D$ and let O be an open set such that $y \in S^{-1}(O)$. Then there is $x \in S(y)$ and $\varepsilon > 0$ such that $x \in S(y) \cap O$ and $x + \varepsilon\mathbb{B} \subset O$. Let $0 < \rho < \varepsilon/\kappa$ and pick a point $y' \in D \cap \mathbb{B}_\rho(y)$. Then

$$x \in S(y) \subset S(y') + \kappa|y - y'|\mathbb{B} \subset S(y') + \varepsilon\mathbb{B}.$$

Hence there exists $x' \in S(y')$ with $|x' - x| \le \varepsilon$ and thus $x' \in S(y') \cap O$, that is $y' \in S^{-1}(O)$. This means that $S^{-1}(O)$ is open relative to D, and from Theorem 3B.2(e) we conclude that S is isc relative to D. $\square$

We obtain from Theorems 3D.1 and 3D.3 some further insights.

Corollary 3D.4 (Lipschitz continuity of polyhedral mappings). *Let $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ be polyhedral and let $D \subset \operatorname{dom} S$ be convex. Then S is isc relative to D if and only if S is actually Lipschitz continuous relative to D. Thus, for a polyhedral mapping, continuity relative to its domain implies Lipschitz continuity.*

Proof. This is immediate from 3D.3 in the light of 3D.1 and the fact that polyhedral mappings are osc and in particular closed-valued everywhere. □

Corollary 3D.5 (single-valued polyhedral mappings). *Let $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ be polyhedral and let $D \subset \operatorname{dom} S$ be convex. If S is not multivalued on D, then S must be a Lipschitz continuous function on D.*

Proof. It is sufficient to show that S is isc relative to D. Let $y \in D$ and O be an open set such that $y \in S^{-1}(O)$; then $x := S(y) \in O$. Since S is outer Lipschitz at y, there exists a neighborhood U of y such that if $y' \in U \cap D$ then $S(y') \in x + \kappa|y' - y|\mathbb{B}$. Taking U smaller if necessary so that $x' := S(y') \in x + \kappa|y' - y|\mathbb{B} \subset O$ for $y' \in U$, we obtain that for every $y' \in U \cap D$ one has $y' \in S^{-1}(x') \subset S^{-1}(O)$. But then $S^{-1}(O)$ must be open relative to D and, from Theorem 3B.2(e), S is isc relative to D. □

In the proof of 2E.6 we used the fact that if a function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $\operatorname{dom} f = \mathbb{R}^n$ has its graph composed by finitely many polyhedral convex sets, then it must be Lipschitz continuous. Now this is a particular case of the preceding result.

Corollary 3D.6 (single-valued solution mappings). *If the solution mapping S of the linear variational inequality in Exercise 3D.2 is single-valued everywhere in $\mathbb{R}^n$, then it must be Lipschitz continuous globally.*

Exercise 3D.7 (distance characterization of outer Lipschitz continuity). *Prove that a mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is outer Lipschitz continuous at $\bar{y}$ relative to a set D with constant $\kappa > 0$ and neighborhood V if and only if $S(\bar{y})$ is closed and*

$$d(x, S(\bar{y})) \leq \kappa d(\bar{y}, S^{-1}(x) \cap D \cap V) \quad \textit{for all } x \in \mathbb{R}^n.$$

Guide. Mimic the proof of 3C.1. □

In parallel with outer Lipschitz continuity we can introduce *inner Lipschitz continuity* of a set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ relative to a set $D \subset \mathbb{R}^m$ at $\bar{y}$ when $\bar{y} \in D \subset \operatorname{dom} S$, $S(\bar{y})$ is a closed set and there exist a constant $\kappa \geq 0$ and a neighborhood V of $\bar{y}$ such that

$$S(\bar{y}) \subset S(y) + \kappa|y - \bar{y}|\mathbb{B} \quad \text{for all } y \in V \cap D.$$

Inner Lipschitz continuity might be of interest on its own, but no significant application of this property in variational analysis has come to light, as yet. Even very simple polyhedral (nonconvex) mappings don't have this property (e.g., consider the mapping from $\mathbb{R}$ to $\mathbb{R}$, which graph is the union of the axes, and choose the origin as reference point) as opposed to the outer Lipschitz continuity which holds for every polyhedral mapping. In addition, a local version of this property does not obey the general implicit function theorem paradigm, as we will show in Section 3H. We therefore drop inner Lipschitz continuity from further consideration.

3E. Aubin Property, Metric Regularity and Linear Openness

A way to localize the concept of Lipschitz continuity of a set-valued mapping is to focus on a neighborhood of a reference point of the graph of the mapping and to use the Pompeiu–Hausdorff distance to a truncation of the mapping with such a neighborhood. More instrumental turns out to be to take the excess and truncate just one part of it. To do that we need the following concept.

Locally closed sets. *A set C is said to be locally closed at $x \in C$ if there exists a neighborhood U of x such that the intersection $C \cap U$ is closed.*

Local closedness of a set C at $x \in C$ can be equivalently defined as the existence of a scalar $a > 0$ such that the set $C \cap \mathbb{B}_a(x)$ is closed.

Aubin property. *A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to have the Aubin property at $\bar{y} \in \mathbb{R}^m$ for $\bar{x} \in \mathbb{R}^n$ if $\bar{x} \in S(\bar{y})$, the graph of S is locally closed at $(\bar{y}, \bar{x})$, and there is a constant $\kappa \geq 0$ together with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$e(S(y') \cap U, S(y)) \leq \kappa |y' - y| \quad \text{for all } y', y \in V, \tag{1}$$

or equivalently, there exist κ, U and V, as described, such that

$$S(y') \cap U \subset S(y) + \kappa |y' - y| \mathbb{B} \quad \text{for all } y', y \in V. \tag{2}$$

The infimum of κ over all such combinations of κ, U and V is called the Lipschitz modulus of S at $\bar{y}$ for $\bar{x}$ and denoted by $\operatorname{lip}(S; \bar{y} | \bar{x})$. The absence of this property is signaled by $\operatorname{lip}(S; \bar{y} | \bar{x}) = \infty$.

It is not claimed that (1) and (2) are themselves equivalent, although that is true when $S(y)$ is closed for every $y \in V$. Nonetheless, the infimum furnishing $\operatorname{lip}(S; \bar{y} | \bar{x})$ is the same whichever formulation is adopted. When S is single-valued on a neighborhood of $\bar{y}$, then the Lipschitz modulus $\operatorname{lip}(S; \bar{y} | S(\bar{y}))$ equals the usual Lipschitz modulus $\operatorname{lip}(S; \bar{y})$ for functions.

In contrast to Lipschitz continuity, the Aubin property is tied to a particular point in the graph of the mapping. As an example, consider the set-valued mapping $S : \mathbb{R} \rightrightarrows \mathbb{R}$ defined as

$$S(y) = \begin{cases} \{0, 1 + \sqrt{y}\} & \text{for } y \geq 0, \\ 0 & \text{for } y < 0. \end{cases}$$

At 0, the value $S(0)$ consists of two points, 0 and 1. This mapping has the Aubin property at 0 for 0 but not at 0 for 1. Also, S is not Lipschitz continuous relative to any interval containing 0.

If $S(y)$ were replaced in (1) and (2) by $S(y) \cap U$, with U and V small enough to ensure this intersection is closed when $y \in V$, we would be looking at *truncated Lipschitz continuity*. This is stronger in general than the Aubin property, but the two are equivalent when S is convex-valued, as will be seen in 3E.3.

Observe that when a set-valued mapping S has the Aubin property at $\bar{y}$ for $\bar{x}$, then, for every point $(y, x) \in \operatorname{gph} S$ which is sufficiently close to $(\bar{y}, \bar{x})$, it has the Aubin property at y for x as well. It is also important to note that the Aubin property of S at $\bar{y}$ for $\bar{x}$ implicitly requires $\bar{y}$ be an element of $\operatorname{int} \operatorname{dom} S$; this is exhibited in the following proposition.

Proposition 3E.1 (local nonemptiness)**.** *If $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ has the Aubin property at $\bar{y}$ for $\bar{x}$, then for every neighborhood U of $\bar{x}$ there exists a neighborhood V of $\bar{y}$ such that $S(y) \cap U \neq \emptyset$ for all $y \in V$.*

Proof. The inclusion (2) for $y' = \bar{y}$ yields

$$\bar{x} \in S(y) + \kappa|y - \bar{y}|\mathbb{B} \quad \text{for every } y \in V,$$

which is the same as

$$(\bar{x} + \kappa|y - \bar{y}|\mathbb{B}) \cap S(y) \neq \emptyset \quad \text{for every } y \in V.$$

That is, $S(y)$ intersects every neighborhood of $\bar{x}$ when y is sufficiently close to $\bar{y}$. □

The property displayed in Proposition 3E.1 is a local version of the inner semicontinuity. Further, if S is Lipschitz continuous relative to an open set D, then S has the Aubin property at any $y \in D \cap \operatorname{int} \operatorname{dom} S$ for any $x \in S(y)$. In particular, the inverse A^{-1} of a linear mapping A has the Aubin property at any point provided that $\operatorname{rge} A = \operatorname{dom} A^{-1}$ has nonempty interior, that is, A is surjective. The converse is also true, since the inverse A^{-1} of a surjective linear mapping A is Lipschitz continuous on the whole space, by Theorem 3C.3, and hence A^{-1} has the Aubin property at any point.

Proposition 3E.2 (single-valued localization from Aubin property)**.** *A set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$ with constant κ if and only if it has a localization at $\bar{y}$ for $\bar{x}$ that is not multivalued and has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant κ.*

Proof. Let s be a localization that is not multivalued and has the Aubin property at $\bar{y}$ for $\bar{x}$. From Proposition 3E.1 we have $\bar{y} \in \operatorname{int} \operatorname{dom} S$, so s is a single-valued localization of S around $\bar{y}$ for $\bar{x}$. Let a and b be positive constants such that $y \mapsto s(y) := S(y) \cap \mathbb{B}_a(\bar{x})$ is a function defined on $\mathbb{B}_b(\bar{y})$ and let $b' > 0$ satisfy $b' < b$ and $8\kappa b' < a$. Then for $y, y' \in \mathbb{B}_{b'}(\bar{y})$ we have

$$\begin{aligned} d(s(y), S(y')) &= d(S(y) \cap \mathbb{B}_a(\bar{x}), S(y')) \\ &\leq \kappa|y - y'| \leq \kappa|y - \bar{y}| + \kappa|y' - \bar{y}| \leq 2\kappa b' < a/4. \end{aligned}$$

Hence, there exists $x' \in S(y')$ such that $|x' - s(y)| \leq d(s(y), S(y')) + a/4 < a/2$. Since $|s(y) - \bar{x}| = d(\bar{x}, S(y))$ and

$$|x' - \bar{x}| \leq |x' - s(y)| + |s(y) - \bar{x}| < a/2 + d(\bar{x}, S(y)) \leq a/2 + \kappa|y - \bar{y}| < a,$$

we obtain $d(s(y), S(y') \cap \mathbb{B}_a(\bar{x})) = d(s(y), S(y'))$ and therefore

$$\begin{aligned}\kappa|y-y'| &\geq d(S(y)\cap \mathbb{B}_a(\bar{x}), S(y')) \\ &= d(s(y), S(y')) = d(s(y), S(y')\cap \mathbb{B}_a(\bar{x})) = |s(y)-s(y')|.\end{aligned}$$

Thus, s is a Lipschitz continuous single-valued localization of S around $\bar{y}$ for $\bar{x}$ with constant κ. □

Proposition 3E.2 is actually a special case of the following, more general result in which convexity enters, inasmuch as singletons are convex sets in particular.

Theorem 3E.3 (truncated Lipschitz continuity under convex-valuedness). *A set-valued mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, whose graph is locally closed at $(\bar{y},\bar{x})$ and whose values are convex sets, has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant $\kappa > 0$ if and only if it has a Lipschitz continuous graphical localization (not necessarily single-valued) around $\bar{y}$ for $\bar{x}$, or in other words, there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that the truncated mapping $y \mapsto S(y)\cap U$ is Lipschitz continuous on V.*

Proof. The "if" part holds even without the convexity assumption. Indeed, if $y \mapsto S(y)\cap U$ is Lipschitz continuous on V and moreover we have

$$S(y')\cap U \subset S(y)\cap U + \kappa|y'-y|\mathbb{B} \subset S(y) + \kappa|y'-y|\mathbb{B} \quad \text{for all } y', y \in V.$$

that is, S has the desired Aubin property. For the "only if" part, suppose now that S has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant $\kappa > 0$, and let $a > 0$ and $b > 0$ be such that

$$(3) \qquad S(y')\cap \mathbb{B}_a(\bar{x}) \subset S(y) + \kappa|y'-y|\mathbb{B} \quad \text{for all } y', y \in \mathbb{B}_b(\bar{y}),$$

and the set $S(y)\cap \mathbb{B}_a(\bar{x})$ is closed for all $y \in \mathbb{B}_b(\bar{y})$. Adjust a and b so that, by 3E.1,

$$(4) \qquad S(y)\cap \mathbb{B}_{a/2}(\bar{x}) \neq \emptyset \text{ for all } y \in \mathbb{B}_b(\bar{x}) \text{ and } 4\kappa b < a.$$

Pick $y, y' \in \mathbb{B}_b(\bar{y})$ and let $x' \in S(y')\cap \mathbb{B}_a(\bar{x})$. Then from (3) there exists $x \in S(y)$ such that

$$(5) \qquad |x - x'| \leq \kappa|y-y'|.$$

If $x \in \mathbb{B}_a(\bar{x})$, there is nothing more to prove, so assume that $r := |x-\bar{x}| > a$. By (4), we can choose a point $\tilde{x} \in S(y)\cap \mathbb{B}_{a/2}(\bar{x})$. Since $S(y)$ is convex, there exists a point $z \in S(y)$ on the segment $[x,\tilde{x}]$ such that $|z-\bar{x}| = a$ and then $z \in S(y)\cap \mathbb{B}_a(\bar{x})$. We will now show that

$$(6) \qquad |z - x'| \leq 5\kappa|y-y'|,$$

which yields that the mapping $y \mapsto S(y)\cap \mathbb{B}_a(\bar{x})$ is Lipschitz continuous on $\mathbb{B}_b(\bar{y})$ with constant 5κ.

By construction, there exists $t \in (0,1)$ such that $z = (1-t)x + t\tilde{x}$. Then

$$a = |z-\bar{x}| = |(1-t)(x-\bar{x}) + t(\tilde{x}-\bar{x})| \leq (1-t)r + t|\tilde{x}-\bar{x}|$$

and in consequence $t(r-|\tilde{x}-\bar{x}|) \le r-a$. Since $\tilde{x} \in \mathbb{B}_{a/2}(\bar{x})$, we get

$$t \le \frac{r-a}{r-a/2}.$$

Using the triangle inequality $|\tilde{x}-x| \le |x-\bar{x}|+|\tilde{x}-\bar{x}| \le r+a/2$, we obtain

$$|z-x| = t|\tilde{x}-x| \le \frac{r-a}{r-a/2}(r+a/2). \tag{7}$$

Also, in view of (4) and (5), we have that

$$r = |x-\bar{x}| \le |x'-x|+|x'-\bar{x}| \le \kappa|y-y'|+a \le \kappa 2b+a \le \frac{3a}{2}. \tag{8}$$

From (7), (8) and the inequality $r > a$ we obtain

$$|z-x| \le (r-a)\frac{r+a/2}{r-a/2} \le (r-a)\frac{3a/2+a/2}{a-a/2} = 4(r-a). \tag{9}$$

Note that $d := r-a$ is exactly the distance from x to the ball $\mathbb{B}_a(\bar{x})$, hence $d \le |x-x'|$ because $x' \in \mathbb{B}_a(\bar{x})$. Combining this with (9) and taking into account (5), we arrive at

$$|z-x'| \le |z-x|+|x-x'| \le 4d+|x'-x| \le 5|x-x'| \le 5\kappa|y-y'|.$$

But this is (6), and we are done. □

Example 3E.4 (convexifying the values)**.** Convexifying the values may change significantly the Lipschitz properties of a mapping. As a simple example consider the solution mapping Σ of the equation $x^2 = p$ (see Fig. 1.1 in the introduction to Chapter 1) and the mapping

$$p \mapsto S(p) = \Sigma(p) \cup \{1\} \cup \{-1\} \quad \text{for} \quad p \in \mathbb{R}.$$

The mapping S is not Lipschitz continuous on any interval $[0,a]$, $a > 0$, while $p \mapsto \operatorname{co} S(p)$ is Lipschitz continuous on $\mathbb{R}$.

The Aubin property could alternatively be defined with one variable "free," as shown in the next proposition.

Proposition 3E.5 (alternative description of Aubin property)**.** *A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant $\kappa > 0$ if and only if $\bar{x} \in S(\bar{y})$, gph S is locally closed at $(\bar{y},\bar{x})$, and there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$e(S(y') \cap U, S(y)) \le \kappa|y'-y| \quad \text{for all } y' \in \mathbb{R}^m \text{ and } y \in V. \tag{10}$$

Proof. Clearly, (10) implies (1). Assume (1) with corresponding U and V and choose positive a and b such that $\mathbb{B}_a(\bar{x}) \subset U$ and $\mathbb{B}_b(\bar{y}) \subset V$. Let $0 < a' < a$ and $0 < b' < b$ be such that

$$(11) \qquad 2\kappa b' + a' \le \kappa b.$$

For any $y \in \mathbb{B}_{b'}(\bar{y})$ we have from (1) that

$$d(\bar{x}, S(y)) \le \kappa|y - \bar{y}| \le \kappa b',$$

hence

$$(12) \qquad e(\mathbb{B}_{a'}(\bar{x}), S(y)) \le \kappa b' + a'.$$

Take any $y' \in \mathbb{R}^m$. If $y' \in \mathbb{B}_b(\bar{y})$ the inequality in (10) comes from (1) and there is nothing more to prove. Assume $|y' - \bar{y}| > b$. Then $|y - y'| > b - b'$ and from (11), $\kappa b' + a' \le \kappa(b - b') \le \kappa|y - y'|$. Using this in (12) we obtain

$$e(\mathbb{B}_{a'}(\bar{x}), S(y)) \le \kappa|y' - y|$$

and since $S(y') \cap \mathbb{B}_{a'}(\bar{x})$ is obviously a subset of $\mathbb{B}_{a'}(\bar{x})$, we come again to (10). $\square$

The Aubin property of a mapping is characterized by Lipschitz continuity of the distance function associated with it.

Theorem 3E.6 (distance function characterization of Aubin property)**.** *For a mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ with $(\bar{y}, \bar{x}) \in \operatorname{gph} S$ and $\operatorname{gph} S$ locally closed at $(\bar{y}, \bar{x})$, let $s(y, x) = d(x, S(y))$. Then S has the Aubin property at $\bar{y}$ for $\bar{x}$ if and only if the function s is Lipschitz continuous with respect to y uniformly in x around $(\bar{y}, \bar{x})$, in which case one has*

$$(13) \qquad \operatorname{lip}(S; \bar{y}\,|\,\bar{x}) = \widehat{\operatorname{lip}}_y(s; (\bar{y}, \bar{x})).$$

Proof. Let $\kappa > \operatorname{lip}(S; \bar{y}\,|\,\bar{x})$. Then, from 3E.1, there exist positive constants a and b such that

$$(14) \qquad \emptyset \ne S(y) \cap \mathbb{B}_a(\bar{x}) \subset S(y') + \kappa|y - y'|\mathbb{B} \ \text{ for all } y, y' \in \mathbb{B}_b(\bar{y}).$$

Without loss of generality, let $a/(4\kappa) \le b$. Let $y \in \mathbb{B}_{a/(4\kappa)}(\bar{y})$ and $x \in \mathbb{B}_{a/4}(\bar{x})$ and let $\tilde{x}$ be a projection of x on $\operatorname{cl} S(y)$. Using 1D.4(b) and (14) with $y = \bar{y}$ we have

$$\begin{aligned} |x - \tilde{x}| &= d(x, S(y)) \le |x - \bar{x}| + d(\bar{x}, S(y)) \\ &\le |x - \bar{x}| + e(S(\bar{y}) \cap \mathbb{B}_a(\bar{x}), S(y)) \le \frac{a}{4} + \kappa|y - \bar{y}| \le \frac{a}{4} + \kappa\frac{a}{4\kappa} = a/2. \end{aligned}$$

Hence

$$|\bar{x} - \tilde{x}| \le |\bar{x} - x| + |x - \tilde{x}| \le \frac{a}{4} + \frac{a}{2} = \frac{3a}{4} < a.$$

This gives us that

$$(15) \qquad |x - \tilde{x}| = d(x, S(y)) = d(x, S(y) \cap \mathbb{B}_a(\bar{x})).$$

Now, let $y' \in \mathbb{B}_{a/(4\kappa)}(\bar{y})$. The inclusion in (14) yields

$$d\left(x, S(y') + \kappa|y - y'|\mathbb{B}\right) \le d(x, S(y) \cap \mathbb{B}_a(\bar{x})). \tag{16}$$

Using the fact that for any set C and for any $r \ge 0$ one has

$$d(x,C) - r \le d(x, C + r\mathbb{B}),$$

from (15) and (16) we obtain

$$d(x, S(y')) - \kappa|y - y'| \le d(x, S(y) \cap \mathbb{B}_a(\bar{x})) = d(x, S(y)).$$

By the symmetry of y and y', we conclude that $\widehat{\operatorname{lip}}_y(s;(\bar{y},\bar{x})) \le \kappa$. Since κ can be arbitrarily close to $\operatorname{lip}(S;\bar{y}\,|\,\bar{x})$, it follows that

$$\widehat{\operatorname{lip}}_y(s;(\bar{y},\bar{x})) \le \operatorname{lip}(S;\bar{y}\,|\,\bar{x}). \tag{17}$$

Conversely, let $\kappa > \widehat{\operatorname{lip}}_y(s;(\bar{y},\bar{x}))$. Then there exist neighborhoods U and V of $\bar{x}$ and $\bar{y}$, respectively, such that $s(\cdot,x)$ is Lipschitz continuous relative to V with a constant κ for any given $x \in U$. Let $y, y' \in V$. Since $V \subset \operatorname{dom} s(\cdot,x)$ for any $x \in U$ we have that $S(y') \cap U \ne \emptyset$. Pick any $x \in S(y') \cap U$; then $s(y',x) = 0$ and, by the assumed Lipschitz continuity of $s(\cdot,x)$, we get

$$d(x, S(y)) = s(y,x) \le s(y',x) + \kappa|y - y'| = \kappa|y - y'|.$$

Taking supremum with respect to $x \in S(y') \cap U$ on the left, we obtain that S has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant κ. Since κ can be arbitrarily close to $\widehat{\operatorname{lip}}_y(s;(\bar{y},\bar{x}))$, we get

$$\widehat{\operatorname{lip}}_y(s;(\bar{y},\bar{x})) \ge \operatorname{lip}(S;\bar{y}\,|\,\bar{x}).$$

This, combined with (17), gives us (13). □

The Aubin property of a mapping is closely tied with a property of its inverse, called *metric regularity*. The concept of metric regularity goes back to the classical Banach open mapping principle. We will devote most of Chapter 5 to studying the metric regularity of set-valued mappings acting in infinite-dimensional spaces.

Metric regularity. *A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be metrically regular at $\bar{x}$ for $\bar{y}$ when $\bar{y} \in F(\bar{x})$, the graph of F is locally closed at $(\bar{x},\bar{y})$, and there is a constant $\kappa \ge 0$ together with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$d(x, F^{-1}(y)) \le \kappa d(y, F(x)) \quad \textit{for all } (x,y) \in U \times V. \tag{18}$$

The infimum of κ over all such combinations of κ, U and V is called the regularity modulus for F at $\bar{x}$ for $\bar{y}$ and denoted by $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$. The absence of metric regularity is signaled by $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \infty$.

Metric regularity is a valuable concept in its own right, especially for numerical purposes. For a general set-valued mapping F and a vector y, it gives an estimate for how far a point x is from being a solution to the generalized equation $F(x) \ni y$ in terms of the "residual" $d(y, F(x))$.

To be specific, let $\bar{x}$ be a solution of the inclusion $\bar{y} \in F(x)$, let F be metrically regular at $\bar{x}$ for $\bar{y}$, and let x_{a} and y_{a} be approximations to $\bar{x}$ and $\bar{y}$, respectively. Then from (18), the distance from x_{a} to the set of solutions of the inclusion $y_{\mathrm{a}} \in F(x)$ is bounded by the constant κ times the residual $d(y_{\mathrm{a}}, F(x_{\mathrm{a}}))$. In applications, the residual is typically easy to compute or estimate, whereas finding a solution might be considerably more difficult. Metric regularity says that there exists a solution to the inclusion $y_{\mathrm{a}} \in F(x)$ at distance from x_{a} proportional to the residual. In particular, if we know the rate of convergence of the residual to zero, then we will obtain the rate of convergence of approximate solutions to an exact one.

Proposition 3C.1 for a mapping S, when applied to $F = S^{-1}$, $F^{-1} = S$, ties the Lipschitz continuity of F^{-1} relative to a set D to a condition resembling (18), but with $F(x)$ replaced by $F(x) \cap D$ on the right, and with $U \times V$ replaced by $\mathbb{R}^n \times D$. We demonstrate now that metric regularity of F in the sense of (18) corresponds to the Aubin property of F^{-1} for the points in question.

Theorem 3E.7 (equivalence of metric regularity and the inverse Aubin property)**.** *A set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is metrically regular at $\bar{x}$ for $\bar{y}$ if and only if its inverse $F^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ has the Aubin property at $\bar{y}$ for $\bar{x}$, in which case*

$$\text{(19)} \qquad \operatorname{lip}(F^{-1}; \bar{y} \,|\, \bar{x}) = \operatorname{reg}(F; \bar{x} \,|\, \bar{y}).$$

Proof. Clearly, the local closedness of the graph of F at $(\bar{x}, \bar{y})$ is equivalent to the local closedness of the graph of F^{-1} at $(\bar{y}, \bar{x})$. Let $\kappa > \operatorname{reg}(F; \bar{x} \,|\, \bar{y})$; then there are positive constants a and b such that (18) holds with $U = \mathbb{B}_a(\bar{x})$, $V = \mathbb{B}_b(\bar{y})$ and with this κ. Without loss of generality, assume $b < a/\kappa$. We will prove next that

$$\text{(20)} \qquad e(F^{-1}(y) \cap U, F^{-1}(y')) \le \kappa |y - y'| \quad \text{for all } y, y' \in V.$$

with $U = \mathbb{B}_a(\bar{x})$ and $V = \mathbb{B}_b(\bar{y})$. Choose $y, y' \in \mathbb{B}_b(\bar{y})$. If $F^{-1}(y) \cap \mathbb{B}_a(\bar{x}) = \emptyset$, then $d(\bar{x}, F^{-1}(y)) \ge a$. But then the inequality (18) with $x = \bar{x}$ yields

$$a \le d(\bar{x}, F^{-1}(y)) \le \kappa d(y, F(\bar{x})) \le \kappa |y - \bar{y}| \le \kappa b < a,$$

a contradiction. Hence there exists $x \in F^{-1}(y) \cap \mathbb{B}_a(\bar{x})$, and for any such x we have from (18) that

$$\text{(21)} \qquad d(x, F^{-1}(y')) \le \kappa d(y', F(x)) \le \kappa |y - y'|.$$

Taking the supremum with respect to $x \in F^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ we obtain (20) with $U = \mathbb{B}_a(\bar{x})$ and $V = \mathbb{B}_b(\bar{y})$, and therefore

$$\text{(22)} \qquad \operatorname{reg}(F; \bar{x} \,|\, \bar{y}) \ge \operatorname{lip}(F^{-1}; \bar{y} \,|\, \bar{x}).$$

Conversely, suppose there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ along with a constant $\kappa > 0$ such that (20) is satisfied. Take U and V smaller if necessary so that, according to Proposition 3E.5, we have

(23) $$e(F^{-1}(y)\cap U, F^{-1}(y')) \leq \kappa|y-y'| \quad \text{for all } y\in \mathbb{R}^m \text{ and } y'\in V.$$

Let $x\in U$ and $y\in V$. If $F(x)\neq\emptyset$, then for any $y\in F(x)$ we have $x\in F^{-1}(y)\cap U$. From (23), we obtain

$$d(x,F^{-1}(y')) \leq e(F^{-1}(y)\cap U, F^{-1}(y')) \leq \kappa|y-y'|.$$

This holds for every $y\in F(x)$, hence, by taking the infimum with respect to $y\in F(x)$ in the last expression we get

$$d(x,F^{-1}(y')) \leq \kappa d(y',F(x)).$$

(If $F(x)=\emptyset$, then because of the convention $d(y,\emptyset)=\infty$, this inequality holds automatically.) Hence, F is metrically regular at $\bar{x}$ for $\bar{y}$ with a constant κ. Then we have $\kappa \geq \operatorname{reg}(F;\bar{x}|\bar{y})$ and hence $\operatorname{reg}(F;\bar{x}|\bar{y}) \leq \operatorname{lip}(F^{-1};\bar{y}|\bar{x})$. This inequality together with (22) results in (19). □

Observe that metric regularity of F at $\bar{x}$ for $\bar{y}$ does *not* require that $\bar{x}\in\operatorname{int}\operatorname{dom} F$. Indeed, if $\bar{x}$ is an isolated point of $\operatorname{dom} F$ then the right side in (18) is ∞ for all $x\in U$, $x\neq\bar{x}$, and then (18) holds automatically. On the other hand, for $x=\bar{x}$ the right side of (18) is always finite (since by assumption $\bar{x}\in\operatorname{dom} F$), and then $F^{-1}(y)\neq\emptyset$ for $y\in V$. This also follows from 3E.1 via 3E.7.

Exercise 3E.8 (equivalent formulation). *Prove that a mapping F, with $(\bar{x},\bar{y})\in \operatorname{gph} F$ at which the graph of F is locally closed, is metrically regular at $\bar{x}$ for $\bar{y}$ with constant $\kappa>0$ if and only if there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

(24) $$d(x,F^{-1}(y)) \leq \kappa d(y,F(x)) \quad \textit{for all } x\in U \textit{ having } F(x)\cap V\neq\emptyset \textit{ and all } y\in V.$$

Guide. First, note that (18) implies (24). Let (24) hold with constant $\kappa>0$ and neighborhoods $\mathbb{B}_a(\bar{x})$ and $\mathbb{B}_b(\bar{y})$ having $b<a/\kappa$. Choose $y,y'\in\mathbb{B}_b(\bar{y})$. As in the proof of 3E.7 show first that $F^{-1}(y)\cap\mathbb{B}_a(\bar{x})\neq\emptyset$ by noting that $F(\bar{x})\cap\mathbb{B}_b(\bar{y})\neq\emptyset$ and hence the inequality in (24) holds for $\bar{x}$ and y. Then for every $x\in F^{-1}(y)\cap\mathbb{B}_a(\bar{x})$ we have that $y\in F(x)\cap\mathbb{B}_b(\bar{y})$, that is, $F(x)\cap\mathbb{B}_b(\bar{y})\neq\emptyset$. Thus, the inequality in (24) holds with y' and any $x\in F^{-1}(y)\cap\mathbb{B}_a(\bar{x})$, which leads to (21) and hence to (20), in the same way as in the proof of 3E.7. The rest follows from the equivalence of (18) and (20) established in 3E.7. □

There is a third property, which we introduced for functions in Section 1F, and which is closely related to both metric regularity and the Aubin property.

Openness. *A mapping $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be open at $\bar{x}$ for $\bar{y}$ if $\bar{y}\in F(\bar{x})$ and for every neighborhood U of $\bar{x}$, $F(U)$ is a neighborhood of $\bar{y}$.*

From the equivalence of metric regularity of F at $\bar{x}$ for $\bar{y}$ and the Aubin property of F^{-1} at $\bar{y}$ for $\bar{x}$, and Proposition 3E.1, we obtain that if a mapping F is metrically regular at $\bar{x}$ for $\bar{y}$, then F is open at $\bar{x}$ for $\bar{y}$. Metric regularity is actually equivalent to the following stronger version of the openness property:

Linear openness. *A mapping $F : I\!\!R^n \rightrightarrows I\!\!R^m$ is said to be linearly open at $\bar{x}$ for $\bar{y}$ when $\bar{y} \in F(\bar{x})$, the graph of F is locally closed at $(\bar{x},\bar{y})$, and there is a constant $\kappa > 0$ together with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$(25) \qquad F(x+\kappa r \operatorname{int} I\!\!B) \supset \big[F(x) + r \operatorname{int} I\!\!B\big] \cap V \quad \textit{for all } x \in U \textit{ and all } r > 0.$$

Linear openness is a particular case of openness which is obtained from (25) for $x = \bar{x}$. Linear openness postulates openness *around* the reference point with balls having proportional radii.

Theorem 3E.9 (equivalence of linear openness and metric regularity)**.** *A set-valued mapping $F : I\!\!R^n \rightrightarrows I\!\!R^m$ is linearly open at $\bar{x}$ for $\bar{y}$ if and only if F is metrically regular at $\bar{x}$ for $\bar{y}$. In this case the infimum of κ for which (25) holds is equal to* $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$.

Proof. Both properties require local closedness of the graph at the reference point. Let (25) hold with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ and a constant $\kappa > 0$. Choose $y \in V$ and $x' \in U$. Let $y' \in F(x')$ (if there is no such y' there is nothing to prove). Since $y = y' + |y-y'|w$ for some $w \in I\!\!B$, denoting $r = |y-y'|$, for every $\varepsilon > 0$ we have $y \in (F(x') + r(1+\varepsilon)\operatorname{int} I\!\!B) \cap V$. From (25), there exists $x \in F^{-1}(y)$ with $|x-x'| \le \kappa(1+\varepsilon)r = \kappa(1+\varepsilon)|y'-y|$. Then $d(x',F^{-1}(y)) \le \kappa(1+\varepsilon)|y'-y|$. Taking infimum with respect to $y' \in F(x')$ on the right and passing to zero with ε (since the left side does not depend on ε), we obtain that F is metrically regular at $\bar{x}$ for $\bar{y}$ with constant κ.

For the converse, let F be metrically regular at $\bar{x}$ for $\bar{y}$ with a constant $\kappa > 0$. We use the characterization of the Aubin property given in Proposition 3E.5. Let $x \in U$, $r > 0$, and let $y' \in (F(x) + r \operatorname{int} I\!\!B) \cap V$. Then there exists $y \in F(x)$ such that $|y-y'| < r$. If $y = y'$ then $y' \in F(x) \subset F(x+\kappa r \operatorname{int} I\!\!B)$, which yields (25) with constant κ. Let $y \ne y'$ and let $\varepsilon > 0$ be so small that $(\kappa+\varepsilon)|y-y'| < \kappa r$. From (10) we obtain $d(x,F^{-1}(y')) \le \kappa|y-y'| < (\kappa+\varepsilon)|y-y'|$. Then there exists $x' \in F^{-1}(y')$ such that $|x-x'| \le (\kappa+\varepsilon)|y-y'|$. But then

$$y' \in F(x') \subset F(x+(\kappa+\varepsilon)|y-y'|I\!\!B) \subset F(x+\kappa r \operatorname{int} I\!\!B),$$

which again yields (25) with constant κ. □

In the classical setting, of course, the equation $f(p,x) = 0$ is solved for x in terms of p, and the goal is to determine when this reduces to x being a function of p through a localization, moreover one with some kind of property of differentiability, or at least Lipschitz continuity. Relinquishing single-valuedness entirely, we can look at "solving" the relation

$$(26) \qquad G(p,x) \ni 0 \ \text{ for a mapping } G : I\!\!R^d \times I\!\!R^n \rightrightarrows I\!\!R^m,$$

or in other words studying the solution mapping $S : I\!\!R^d \rightrightarrows I\!\!R^n$ defined by

$$S(p) = \{ x \mid G(p,x) \ni 0 \}. \tag{27}$$

Fixing a pair $(\bar p, \bar x)$ such that $\bar x \in S(\bar p)$, we can raise questions about local behavior of S as might be deduced from assumptions on G. Note that $\operatorname{gph} S$ will be locally closed at $(\bar p, \bar x)$ when $\operatorname{gph} G$ is locally closed at $((\bar p, \bar x), 0)$.

We will concentrate here on the extent to which S can be guaranteed to have the Aubin property at $\bar p$ for $\bar x$. This turns out to be true when G has the Aubin property with respect to p and a weakened metric regularity property with respect to x, but we have to formulate exactly what we need about this in a local sense.

Partial Aubin property. *The mapping $G : \mathbb{R}^d \times \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to have the partial Aubin property with respect to p uniformly in x at $(\bar p, \bar x)$ for $\bar y$ if $\bar y \in G(\bar p, \bar x)$, $\operatorname{gph} G$ is locally closed at $((\bar p, \bar x), \bar y)$, and there is a constant $\kappa \geq 0$ together with neighborhoods Q for $\bar p$, U of $\bar x$ and V of $\bar y$ such that*

$$e(G(p,x) \cap V, G(p',x)) \leq \kappa |p - p'| \quad \text{for all } p, p' \in Q \text{ and } x \in U, \tag{28}$$

or equivalently, there exist κ, Q, U and V, as described, such that

$$G(p,x) \cap V \subset G(p',x) + \kappa |p - p'| \mathbb{B} \quad \text{for all } p, p' \in Q \text{ and } x \in U.$$

The infimum of κ over all such combinations of κ, Q, U and V is called the partial Lipschitz modulus of G with respect to p uniformly in x at $(\bar p, \bar x)$ for $\bar y$ and denoted by $\widehat{\operatorname{lip}}_p(G; \bar p, \bar x \mid \bar y)$. The absence of this property is signaled by $\widehat{\operatorname{lip}}_p(G; \bar p, \bar x \mid \bar y) = \infty$.

The basic result we are able now to state about the solution mapping in (27) could be viewed as an "implicit function" complement to the "inverse function" result in Theorem 3E.7, rather than as a result in the pattern of the implicit function theorem (which features approximations of one kind or another).

Theorem 3E.10 (Aubin property of general solution mappings)**.** *In (26), let $G : \mathbb{R}^d \times \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, with $G(\bar p, \bar x) \ni 0$, have the partial Aubin property with respect to p uniformly in x at $(\bar p, \bar x)$ for 0 with constant κ. Furthermore, in the notation (27), let G enjoy the existence of a constant λ such that*

$$d(x, S(p)) \leq \lambda\, d(0, G(p,x)) \quad \text{for all } (p,x) \text{ close to } (\bar p, \bar x). \tag{29}$$

Then the solution mapping S in (27) has the Aubin property at $\bar p$ for $\bar x$ with constant $\lambda \kappa$.

Proof. As mentioned earlier, $\operatorname{gph} S$ will be locally closed at $(\bar p, \bar x)$ when $\operatorname{gph} G$ is locally closed at $((\bar p, \bar x), 0)$. Take $p, p' \in Q$ and $x \in S(p) \cap U$ so that (28) holds for a neighborhood V of 0. From (29) and then (28) we have

$$d(x, S(p')) \leq \lambda\, d(0, G(p',x)) \leq \lambda\, e(G(p,x) \cap V, G(p',x)) \leq \lambda \kappa |p - p'|.$$

Taking the supremum of the left side with respect to $x \in S(p) \cap U$, we obtain that S has the Aubin property with constant $\lambda \kappa$. □

Example 3E.11. Theorem 3E.10 cannot be extended to a two-way characterization parallel to Theorem 3E.7. Indeed, consider the "saddle" function of two real variables $f(p,x) = x^2 - p^2$. In this case f does not satisfy (29) at the origin of $I\!R^2$, yet the solution mapping $S(p) = \{ x \mid f(p,x) = 0 \} = \{-p, p\}$ has the Aubin property at 0 for 0.

At the end of this section we will take a closer look at the following question. If a mapping F is simultaneous metrically regular and has the Aubin property, both at $\bar{x}$ for $\bar{y}$ for some $(\bar{x},\bar{y}) \in \operatorname{gph} F$, then what is the relation, if any, between $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$ and $\operatorname{lip}(F;\bar{x}\,|\,\bar{y})$? Having in mind 3E.7, it is the same as asking what is the relation between $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$ and $\operatorname{reg}(F^{-1};\bar{y}\,|\,\bar{x})$ or between $\operatorname{lip}(F;\bar{x}\,|\,\bar{y})$ and $\operatorname{lip}(F^{-1};\bar{y}\,|\,\bar{x})$. When we exclude the trivial case when $(\bar{x},\bar{y}) \in \operatorname{int} \operatorname{gph} F$, in which case both moduli would be zero, an answer to this question is stated in the following exercise.

Exercise 3E.12. *Consider a mapping $F : I\!R^n \rightrightarrows I\!R^m$ with closed graph and a point $(\bar{x},\bar{y}) \in \operatorname{gph} F \setminus \operatorname{int} \operatorname{gph} F$. Then*

$$\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \cdot \operatorname{lip}(F;\bar{x}\,|\,\bar{y}) \geq 1,$$

including the limit cases when either of these moduli is 0 and then the other is ∞ under the convention $0 \cdot \infty = \infty$.

Guide. Let $\kappa > \operatorname{reg}(F;\bar{x}\,|\,\bar{y})$ and $\gamma > \operatorname{lip}(F;\bar{x}\,|\,\bar{y})$. Then there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ corresponding to metric regularity and the Aubin property of F with constants κ and γ, respectively. Let $(x,y) \in U \times V$ be such that $d(x, F^{-1}(y)) > 0$ (why does such a point exist?). Then there exists $x' \in F^{-1}(y)$ such that $0 < |x - x'| = d(x, F^{-1}(y))$. We have

$$|x - x'| = d(x, F^{-1}(y)) \leq \kappa d(y, F(x)) \leq \kappa e(F(x') \cap V, F(x)) \leq \kappa\gamma|x - x'|.$$

Hence, $\kappa\gamma \geq 1$. $\square$

For a solution mapping $S = F^{-1}$ of an inclusion $F(x) \ni y$ with a parameter y, the quantity $\operatorname{lip}(S;\bar{y}\,|\,\bar{x})$ measures how "stable" solutions near $\bar{x}$ are under changes of the parameter near $\bar{y}$. In this context, the smaller this modulus is, the "better" stability we have. In view of 3E.12, better stability means larger values of the regularity modulus $\operatorname{reg}(S;\bar{y}\,|\,\bar{x})$. In the limit case, when S is a constant function near $\bar{y}$, that is, when the solution is not sensitive at all with respect to small changes of the parameter y near $\bar{y}$, then $\operatorname{lip}(S;\bar{y}\,|\,S(\bar{y})) = 0$ while the metric regularity modulus of S there is infinity. In Section 6A we will see that the "larger" the regularity modulus of a mapping is, the "easier" it is to perturb the mapping so that it looses its metric regularity.

3F. Implicit Mapping Theorems with Metric Regularity

In the paradigm of the implicit function theorem, as applied to a generalized equation $f(p,x)+F(x)\ni 0$ with solutions $x\in S(p)$, the focus is on some $\bar{p}$ and $\bar{x}\in S(\bar{p})$, and on some kind of approximation of the mapping $x\mapsto f(\bar{p},x)+F(x)$. Assumptions about this approximation lead to conclusions about the solution mapping S relative to $\bar{p}$ and $\bar{x}$. Stability properties of the approximation are the key to progress in this direction. Our aim now is to study such stability with respect to metric regularity and to show that this leads to implicit function theorem type results which apply to set-valued solution mappings S beyond any framework of single-valued localization.

We start this section with a particular case of a fundamental result in variational analysis and beyond, stated next as Theorem 3F.1, which goes back to works by Lyusternik and Graves. We will devote most of Chapter 5 to the full theory behind this result—in an infinite-dimensional setting.

Theorem 3F.1 (inverse mapping theorem with metric regularity)**.** *Consider a mapping $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^m$, a point $(\bar{x},\bar{y})\in\operatorname{gph}F$, and a function $g:\mathbb{R}^n\to\mathbb{R}^m$ with $\bar{x}\in\operatorname{int}\operatorname{dom}g$. Let κ and μ be nonnegative constants such that*

$$\kappa\mu<1,\quad \operatorname{reg}(F;\bar{x}\,|\,\bar{y})\le\kappa\quad \textit{and}\quad \operatorname{lip}(g;\bar{x})\le\mu.$$

Then

$$\text{(1)}\qquad \operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y})\le\frac{\kappa}{1-\kappa\mu}.$$

Although formally there is no inversion of a mapping in Theorem 3F.1, if this result is stated equivalently in terms of the Aubin property of the inverse mapping F^{-1}, it fits then into the pattern of the inverse function theorem paradigm. It can also be viewed as a result concerning stability of metric regularity under perturbations by functions with small Lipschitz constants. We can actually deduce the classical inverse function theorem 1A.1 from 3F.1. Indeed, let $f:\mathbb{R}^n\to\mathbb{R}^n$ be a smooth function around $\bar{x}$ and let $\nabla f(\bar{x})$ be nonsingular. Then $F=Df(\bar{x})$ is metrically regular everywhere and from 3F.1 for the function $g(x)=f(x)-Df(\bar{x})(x-\bar{x})$ with $\operatorname{lip}(g;\bar{x})=0$ we obtain that $g+F=f$ is metrically regular at $\bar{x}$ for $f(\bar{x})$. But then f must be open (cf. 3E.9). Establishing this fact is the main part of all proofs of 1A.1 presented so far.

We will postpone proving Theorem 3F.1 to Chapter 5, where we will do it for a mapping F acting from a complete metric space to a linear metric space. In this section we focus on some consequences of this result and its implicit function version. Several corollaries of 3F.1 will lead the way.

Corollary 3F.2 (detailed estimates)**.** *Consider a mapping $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^m$ and any pair $(\bar{x},\bar{y})\in\operatorname{gph}F$. If $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})>0$, then for every $g:\mathbb{R}^n\to\mathbb{R}^m$ such that $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})\cdot\operatorname{lip}(g;\bar{x})<1$, one has*

(2) $$\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) \le (\operatorname{reg}(F;\bar{x}\,|\,\bar{y})^{-1} - \operatorname{lip}(g;\bar{x}))^{-1}.$$

If $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = 0$, *then* $\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) = 0$ *for any* $g:\mathbb{R}^n \to \mathbb{R}^m$ *with* $\operatorname{lip}(g;\bar{x}) < \infty$. *If* $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \infty$, *then* $\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) = \infty$ *for any* $g:\mathbb{R}^n \to \mathbb{R}^m$ *with* $\operatorname{lip}(g;\bar{x}) = 0$.

Proof. If $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) < \infty$, then by choosing κ and μ appropriately and passing to limits in (1) we obtain the claimed inequality (2) also for the case where $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = 0$. Let $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \infty$, and suppose that $\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) < \kappa$ for some κ and a function g with $\operatorname{lip}(g;\bar{x}) = 0$. Since the graph of $g+F$ is locally closed at $(\bar{x}, g(\bar{x})+\bar{y})$ and g is continuous around $\bar{x}$, we get that $\operatorname{gph} F$ is locally closed at $(\bar{x},\bar{y})$ (prove it!). Applying Theorem 3F.1 to the mapping $g+F$ with perturbation $-g$, and noting that $\operatorname{lip}(-g;\bar{x}) = 0$, we obtain $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \le \kappa$, which constitutes a contradiction. □

When the perturbation g has zero Lipschitz modulus at the reference point, we obtain another interesting fact.

Corollary 3F.3 (perturbations with Lipschitz modulus 0)**.** *Consider a mapping* $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *and a point* $(\bar{x},\bar{y}) \in \operatorname{gph} F$. *Then for every* $g:\mathbb{R}^n \to \mathbb{R}^m$ *with* $\operatorname{lip}(g;\bar{x}) = 0$ *one has*

$$\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) = \operatorname{reg}(F;\bar{x}\,|\,\bar{y}).$$

Proof. The cases with $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = 0$ or $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \infty$ are already covered by Corollary 3F.2. If $0 < \operatorname{reg}(F;\bar{x}\,|\,\bar{y}) < \infty$, we get from (2) that

$$\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) \le \operatorname{reg}(F;\bar{x}\,|\,\bar{y}).$$

By exchanging the roles of F and $g+F$, we also get

$$\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \le \operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}),$$

and in that way arrive at the claimed equality. □

An elaboration of Corollary 3F.3 employs first-order approximations of a function as were introduced in Section 1E.

Corollary 3F.4 (utilization of strict first-order approximations)**.** *Consider a mapping* $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *and a point* $(\bar{x},\bar{y}) \in \operatorname{gph} F$. *Let* $f:\mathbb{R}^n \to \mathbb{R}^m$ *be continuous in a neighborhood of* $\bar{x}$. *Then, for every* $h:\mathbb{R}^n \to \mathbb{R}^m$ *which is a strict first-order approximation to* f *at* $\bar{x}$, *one has*

$$\operatorname{reg}(f+F;\bar{x}\,|\,f(\bar{x})+\bar{y}) = \operatorname{reg}(h+F;\bar{x}\,|\,h(\bar{x})+\bar{y}).$$

In particular, when the strict first-order approximation is represented by the linearization coming from strict differentiability, we get something even stronger.

Corollary 3F.5 (utilization of strict differentiability)**.** *Consider* $M = f+F$ *for a function* $f:\mathbb{R}^n \to \mathbb{R}^m$ *and a mapping* $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$, *and let* $\bar{y} \in M(\bar{x})$. *Suppose that*

f is strictly differentiable at $\bar{x}$. Then, for the linearization

$$M_0(x) = f(\bar{x}) + \nabla f(\bar{x})(x - \bar{x}) + F(x)$$

one has

$$\operatorname{reg}(M;\bar{x}\,|\,\bar{y}) \;=\; \operatorname{reg}(M_0;\bar{x}\,|\,\bar{y}).$$

In the case when $m = n$ and the mapping F is the normal cone mapping to a polyhedral convex set, we can likewise employ a "first-order approximation" of F. When f is linear the corresponding result parallels 2E.6.

Corollary 3F.6 (affine-polyhedral variational inequalities)**.** *For an $n \times n$ matrix A and a polyhedral convex set $C \subset \mathbb{R}^n$, consider the variational inequality*

$$Ax + N_C(x) \ni y.$$

Let $\bar{x}$ be a solution for $\bar{y}$, let $\bar{v} = \bar{y} - A\bar{x}$, so that $\bar{v} \in N_C(\bar{x})$, and let $K = K_C(\bar{x},\bar{v})$ be the critical cone to C at $\bar{x}$ for $\bar{v}$. Then, for the mappings

$$\begin{aligned} G(x) &= Ax + N_C(x) \;\text{ with }\; G(\bar{x}) \ni \bar{y},\\ G_0(w) &= Aw + N_K(w) \;\text{ with }\; G_0(0) \ni 0, \end{aligned}$$

we have

$$\operatorname{reg}(G;\bar{x}\,|\,\bar{y}) \;=\; \operatorname{reg}(G_0;0\,|\,0).$$

Proof. From reduction lemma 2E.4, for (w,u) in a neighborhood of $(0,0)$, we have that $\bar{v} + u \in N_C(\bar{x} + w)$ if and only if $u \in N_K(w)$. Then, for (w,v) in a neighborhood of $(0,0)$, we obtain $\bar{y} + v \in G(\bar{x} + w)$ if and only if $v \in G_0(w)$. Thus, metric regularity of $A + N_C$ at $\bar{x}$ for $\bar{y}$ with a constant κ implies metric regularity of $A + N_K$ at 0 for 0 with the same constant κ, and conversely. □

Combining 3F.5 and 3F.6 we obtain the following corollary:

Corollary 3F.7 (strict differentiability and polyhedral convexity)**.** *Consider $H = f + N_C$ for a function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a polyhedral convex set $C \subset \mathbb{R}^n$, let $\bar{y} \in H(\bar{x})$ and let f be strictly differentiable at $\bar{x}$. For $\bar{v} = \bar{y} - f(\bar{x})$, let $K = K_C(\bar{x},\bar{v})$ be the critical cone to the set C at $\bar{x}$ for $\bar{v}$. Then, for $H_0(x) = \nabla f(\bar{x})x + N_K(x)$ one has*

$$\operatorname{reg}(H;\bar{x}\,|\,\bar{y}) \;=\; \operatorname{reg}(H_0;0\,|\,0).$$

We are ready now to take up once more the study of a generalized equation having the form

$$f(p,x) + F(x) \ni 0 \tag{3}$$

for $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, and its solution mapping $S : \mathbb{R}^d \to \mathbb{R}^n$ defined by

$$S(p) = \big\{\, x \,\big|\, f(p,x) + F(x) \ni 0 \big\}. \tag{4}$$

This time, however, we are not looking for single-valued localizations of S but aiming at a better understanding of situations in which S may not have any such localization, as in the example of parameterized constraint systems. Recall from Chapter 1 that, for $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and a point $(\bar{p},\bar{x}) \in \operatorname{int}\operatorname{dom} f$, a function $h : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a strict estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ if $h(\bar{x}) = f(\bar{x},\bar{p})$ and

$$\widehat{\operatorname{lip}}_x(e;(\bar{p},\bar{x})) \leq \mu < \infty \ \text{ for } e(p,x) = f(p,x) - h(x).$$

Theorem 3F.8 (implicit mapping theorem with metric regularity). *For the generalized equation (3) and its solution mapping S in (4), and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$, let $h : \mathbb{R}^n \to \mathbb{R}^m$ be a strict estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ, let $h+F$ be metrically regular at $\bar{x}$ for 0 with $\operatorname{reg}(h+F;\bar{x}|0) \leq \kappa$. Assume*

$$\kappa\mu < 1 \ \text{ and } \ \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})) \leq \lambda < \infty. \tag{5}$$

Then S has the Aubin property at $\bar{p}$ for $\bar{x}$, and moreover

$$\operatorname{lip}(S;\bar{p}|\bar{x}) \leq \frac{\kappa\lambda}{1-\kappa\mu}. \tag{6}$$

This theorem will be established in an infinite-dimensional setting in Theorem 5E.5 in Chapter 5, so we will not prove it separately here. An immediate consequence is obtained by specializing the function h in Theorem 3F.8 to a linearization of f with respect to x. We add to this the effect of ample parameterization, in parallel to the case of single-valued localization in Theorem 2C.2.

Theorem 3F.9 (using strict differentiability and ample parameterization). *For the generalized equation (3) and its solution mapping S in (4), and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$, suppose that f is strictly differentiable at $(\bar{p},\bar{x})$. If the mapping*

$$h+F \ \text{ for } h(x) = f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x})$$

is metrically regular at $\bar{x}$ for 0, then S has the Aubin property at $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(S;\bar{p}|\bar{x}) \leq \operatorname{reg}(h+F;\bar{x}|0)\cdot|\nabla_p f(\bar{p},\bar{x})|. \tag{7}$$

Furthermore, when f satisfies the ample parameterization condition

$$\operatorname{rank} \nabla_p f(\bar{p},\bar{x}) = m, \tag{8}$$

then the converse implication holds as well: the mapping $h+F$ is metrically regular at $\bar{x}$ for 0 provided that S has the Aubin property at $\bar{p}$ for $\bar{x}$.

Proof of 3F.9, initial part. In these circumstances with this choice of h, the conditions in (5) are satisfied because, for $e = f - h$,

$$\widehat{\mathrm{lip}}_x(e;(\bar{p},\bar{x})) = 0 \ \text{ and } \ \widehat{\mathrm{lip}}_p(f;(\bar{p},\bar{x})) = |\nabla_p f(\bar{p},\bar{x})|.$$

Thus, (7) follows from (6) with $\mu = 0$. In the remainder of the proof, regarding ample parameterization, we will make use of the following fact.

Proposition 3F.10 (Aubin property in composition). *For a mapping $M : I\!R^d \rightrightarrows I\!R^n$ and a function $\psi : I\!R^n \times I\!R^m \to I\!R^d$ consider the composite mapping $N : I\!R^m \rightrightarrows I\!R^n$ of the form*

$$y \mapsto N(y) = \{\, x \,|\, x \in M(\psi(x,y)) \,\} \quad \textit{for } y \in I\!R^m.$$

Let ψ satisfy

$$\text{(9)} \qquad \widehat{\mathrm{lip}}_x(\psi;(\bar{x},0)) = 0 \ \textit{ and } \ \widehat{\mathrm{lip}}_y(\psi;(\bar{x},0)) < \infty,$$

and, for $\bar{p} = \psi(\bar{x},0)$, let $(\bar{p},\bar{x}) \in \mathrm{gph}\, M$. Under these conditions, if M has the Aubin property at $\bar{p}$ for $\bar{x}$, then N has the Aubin property at 0 for $\bar{x}$.

Proof. Let the mapping M have the Aubin property at $\bar{p}$ for $\bar{x}$ with neighborhoods Q of $\bar{p}$ and U of $\bar{x}$ and constant $\kappa > \mathrm{lip}\,(M;\bar{p}\,|\,\bar{x})$. Choose $\lambda > 0$ with $\lambda < 1/\kappa$ and let $\gamma > \widehat{\mathrm{lip}}_y(\psi;(\bar{x},0))$. By (9) there exist positive constants a and b such that for any $y \in I\!B_a(0)$ the function $\psi(\cdot,y)$ is Lipschitz continuous on $I\!B_b(\bar{x})$ with Lipschitz constant λ and for every $x \in I\!B_b(\bar{x})$ the function $\psi(x,\cdot)$ is Lipschitz continuous on $I\!B_a(0)$ with Lipschitz constant γ. Pick a positive constant c and make a and b smaller if necessary so that:

(a) $I\!B_c(\bar{p}) \subset Q$ and $I\!B_b(\bar{x}) \subset U$,
(b) the set $\mathrm{gph}\, M \cap (I\!B_c(\bar{p}) \times I\!B_b(\bar{x}))$ is closed, and
(c) the following inequalities are satisfied:

$$\text{(10)} \qquad \frac{4\kappa\gamma a}{1-\kappa\lambda} \le b \ \text{ and } \ \gamma a + \lambda b \le c.$$

Let $y', y \in I\!B_a(0)$ and let $x' \in N(y') \cap I\!B_{b/2}(\bar{x})$. Then $x' \in M(\psi(x',y')) \cap I\!B_{b/2}(\bar{x})$. Further, we have

$$|\psi(x',y') - \bar{p}| \le |\psi(x',y') - \psi(x',0)| + |\psi(x',0) - \psi(\bar{x},0)| \le \gamma a + \lambda b/2 \le c$$

and the same for $\psi(x',y)$. From the Aubin property of M we obtain the existence of $x^1 \in M(\psi(x',y))$ such that

$$|x^1 - x'| \le \kappa|\psi(x',y') - \psi(x',y)| \le \kappa\gamma|y'-y|.$$

Thus, through the first inequality in (10),

$$|x^1 - \bar{x}| \le |x^1 - x'| + |x' - \bar{x}| \le \kappa\gamma|y'-y| + |x'-\bar{x}| \le \kappa\gamma(2a) + \frac{b}{2} \le b,$$

and consequently

$$|\psi(x^1,y)-\bar{p}| = |\psi(x^1,y)-\psi(\bar{x},0)| \le \lambda b+\gamma a \le c,$$

utilizing the second inequality in (10). Hence again, from the Aubin property of M applied to $x^1 \in M(\psi(x',y))\cap \mathbb{B}_b(\bar{x})$, there exists $x^2 \in M(\psi(x^1,y))$ such that

$$|x^2-x^1| \le \kappa|\psi(x^1,y)-\psi(x',y)| \le \kappa\lambda|x^1-x'| \le (\kappa\lambda)\kappa\gamma|y'-y|.$$

Employing induction, assume that we have a sequence $\{x^j\}$ with

$$x^j \in M(\psi(x^{j-1},y)) \;\text{ and }\; |x^j-x^{j-1}| \le (\kappa\lambda)^{j-1}\kappa\gamma|y'-y| \;\text{ for }\; j=1,\dots,k.$$

Setting $x^0=x'$, we get

$$\begin{aligned}|x^k-\bar{x}| &\le |x^0-\bar{x}|+\sum_{j=1}^{k}|x^j-x^{j-1}|\\ &\le \frac{b}{2}+\sum_{j=0}^{k-1}(\kappa\lambda)^j\kappa\gamma|y'-y| \le \frac{b}{2}+\frac{2a\kappa\gamma}{1-\kappa\lambda} \le b,\end{aligned}$$

where we use the first inequality in (10). Hence $|\psi(x^k,y)-\bar{p}| \le \lambda b+\gamma a \le c$. Then there exists $x^{k+1} \in M(\psi(x^k,y))$ such that

$$|x^{k+1}-x^k| \le \kappa|\psi(x^k,y)-\psi(x^{k-1},y)| \le \kappa\lambda|x^k-x^{k-1}| \le (\kappa\lambda)^k\kappa\gamma|y'-y|,$$

and the induction step is complete.

The sequence $\{x^k\}$ is Cauchy, hence convergent to some $x \in \mathbb{B}_b(\bar{x}) \subset U$. From the local closedness of $\operatorname{gph} M$ and the continuity of ψ we deduce that $x \in M(\psi(x,y))$, hence $x \in N(y)$. Furthermore, using the estimate

$$|x^k-x^0| \le \sum_{j=1}^{k}|x^j-x^{j-1}| \le \sum_{j=0}^{k-1}(\kappa\lambda)^j\kappa\gamma|y'-y| \le \frac{\kappa\gamma}{1-\kappa\lambda}|y'-y|$$

and passing to the limit with respect to $k\to\infty$, we obtain that

$$|x-x'| \le \frac{\kappa\gamma}{1-\kappa\lambda}|y'-y|.$$

Thus, for any $\kappa' \ge (\kappa\gamma)/(1-\kappa\lambda)$ the mapping N has the Aubin property at 0 for $\bar{x}$ with constant κ'. □

Proof of 3F.9, final part. Under the ample parameterization condition (8), Lemma 2C.1 guarantees the existence of neighborhoods U of $\bar{x}$, V of 0, and Q of $\bar{p}$, as well as a local selection $\psi: U\times V \to Q$ around $(\bar{x},0)$ for $\bar{p}$ of the mapping

$$(x,y) \mapsto \{\, p \mid y+f(p,x)=h(x)\,\}$$

for $h(x) = f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x})$ which satisfies the conditions in (9). Hence,

$$y + f(\psi(x,y),x) = h(x) \quad \text{and} \quad \psi(x,y) \in Q \quad \text{for } x \in U, y \in V.$$

Fix $y \in V$. If $x \in (h+F)^{-1}(y) \cap U$ and $p = \psi(x,y)$, then $p \in Q$ and $y + f(p,x) = h(x)$, hence $x \in S(p) \cap U$. Conversely, if $x \in S(\psi(x,y)) \cap U$, then clearly $x \in (h+F)^{-1}(y) \cap U$. Thus,

$$(11) \qquad (h+F)^{-1}(y) \cap U = \{ x \mid x \in S(\psi(x,y)) \cap U \}.$$

Since the Aubin property of S at $\bar{p}$ for $\bar{x}$ is a local property of the graph of S relative to the point $(\bar{p},\bar{x})$, it holds if and only if the same holds for the truncated mapping $S_U : p \mapsto S(p) \cap U$ (see Exercise 3F.11). That equivalence is valid for $(h+F)^{-1}$ as well. Thus, if the mapping S_U has the Aubin property at $\bar{p}$ for $\bar{x}$, from Proposition 3F.10 in the context of (11), we obtain that $(h+F)^{-1}$ has the Aubin property at 0 for $\bar{x}$, hence, by 3E.7, $h+F$ is metrically regular at $\bar{x}$ for 0 as desired. □

Exercise 3F.11. *Let $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ have the Aubin property at $\bar{y}$ for $\bar{x}$ with constant κ. Show that for any neighborhood U of $\bar{x}$ the mapping $S_U : y \mapsto S(y) \cap U$ also has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant κ.*

Guide. Choose sufficiently small $a > 0$ and $b > 0$ such that $\mathbb{B}_a(\bar{x}) \subset U$ and $4\kappa b \le a$. Then for every $y, y' \in \mathbb{B}_b(\bar{y})$ and every $x \in S(y) \cap \mathbb{B}_{a/2}(\bar{x})$ there exists $x' \in S(y')$ with $|x'-x| \le \kappa |y'-y| \le 2\kappa b \le a/2$. Then both x and x' are from U. □

Let us now look at the case of 3F.9 in which F is a constant mapping, $F(x) \equiv K$, which was featured at the beginning of this chapter as a motivation for investigating real set-valuedness in solution mappings. Solving $f(p,x) + F(x) \ni 0$ for a given p then means finding an x such that $-f(p,x) \in K$. For particular choices of K this amounts to solving some mixed system of equations and inequalities, for example.

Example 3F.12 (application to general constraint systems)**.** For $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and a closed set $K \subset \mathbb{R}^m$, let

$$S(p) = \{ x \mid 0 \in f(p,x) + K \}.$$

Fix $\bar{p}$ and $\bar{x} \in S(\bar{p})$. Suppose that f is continuously differentiable on a neighborhood of $(\bar{p},\bar{x})$, and consider the solution mapping for an associated linearized system:

$$\bar{S}(y) = \{ x \mid y \in f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x}) + K \}.$$

If $\bar{S}$ has the Aubin property at 0 for $\bar{x}$, then S has the Aubin property at $\bar{p}$ for $\bar{x}$. The converse implication holds under the ample parameterization condition (8).

The key to applying this result, of course, is being able to ascertain when the linearized system does have the Aubin property in question. In the important case of $K = \mathbb{R}^s_+ \times \{0\}^{m-s}$, a necessary and sufficient condition will emerge in the so-called Mangasarian–Fromovitz constraint qualification. This will be seen in Section 4D.

Example 3F.13 (application to polyhedral variational inequalities)**.** For $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n$ and a convex polyhedral set $C \subset \mathbb{R}^n$, let

$$S(p) = \{ x \mid f(p,x) + N_C(x) \ni 0 \}.$$

Fix $\bar{p}$ and $\bar{x} \in S(\bar{p})$ and for $\bar{v} = -f(\bar{p},\bar{x})$ let $K = K_C(\bar{x},\bar{v})$ be the associated critical cone to C. Suppose that f is continuously differentiable on a neighborhood of $(\bar{p},\bar{x})$, and consider the solution mapping for an associated reduced system:

$$\bar{S}(y) = \{ x \mid \nabla_x f(\bar{p},\bar{x})x + N_K(x) \ni y \}.$$

If $\bar{S}$ has the Aubin property at 0 for 0, then S has the Aubin property at $\bar{p}$ for $\bar{x}$. The converse implication holds under the ample parameterization condition (8).

If a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ has the Aubin property at $\bar{x}$ for $\bar{y}$, and a function $f : \mathbb{R}^n \to \mathbb{R}^m$ satisfies $\operatorname{lip}(f;\bar{x}) < \infty$, then the mapping $f + F$ has the Aubin property at $\bar{x}$ for $f(\bar{x}) + \bar{y}$ as well. This is a particular case of the following observation which utilizes ample parameterization.

Theorem 3F.14 (Aubin property of the inverse to the solution mapping)**.** *Consider a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with $(\bar{x},\bar{y}) \in \operatorname{gph} F$ and a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ having $\bar{y} = -f(\bar{p},\bar{x})$ and which is strictly differentiable at $(\bar{p},\bar{x})$ and satisfies the ample parameterization condition (8). Then the mapping*

$$x \mapsto P(x) = \{ p \mid 0 \in f(p,x) + F(x) \}$$

has the Aubin property at $\bar{x}$ for $\bar{p}$ if and only if F has the Aubin property at $\bar{x}$ for $\bar{y}$.

Proof. First, from 3F.9 it follows that under the ample parameterization condition (8) the mapping

$$(x,y) \mapsto \Omega(x,y) = \{ p \mid y + f(p,x) = 0 \}$$

has the Aubin property at $(\bar{x},\bar{y})$ for $\bar{p}$. Let F have the Aubin property at $\bar{x}$ for $\bar{y}$ with neighborhoods U of $\bar{x}$ of V for $\bar{y}$ and constant $\kappa > 0$. Choose a neighborhood Q of $\bar{p}$ and adjust U and V accordingly so that Ω has the Aubin property with constant λ and neighborhoods $U \times V$ and Q. Let $b > 0$ be such that $\mathbb{B}_b(\bar{y}) \subset V$, then choose $a > 0$ and adjust Q such that $\mathbb{B}_a(\bar{x}) \subset U$, $a \leq b/(4\kappa)$ and also $-f(p,x) \in \mathbb{B}_{b/2}(\bar{y})$ for $x \in \mathbb{B}_a(\bar{x})$ and $p \in Q$. Let $x, x' \in \mathbb{B}_a(\bar{x})$ and $p \in P(x) \cap Q$. Then $y = -f(p,x) \in F(x) \cap V$ and by the Aubin property of F there exists $y' \in F(x')$ such that $|y - y'| \leq \kappa|x - x'|$. But then $|y' - \bar{y}| \leq \kappa(2a) + b/2 \leq b$. Thus $y' \in V$ and hence, by the Aubin property of Ω, there exists p' satisfying $y' + f(p',x') = 0$ and

$$|p' - p| \leq \lambda(|y' - y| + |x' - x|) \leq \lambda(\kappa + 1)|x' - x|.$$

Noting that $p' \in P(x')$ we get that P has the Aubin property at $\bar{x}$ for $\bar{p}$.

Conversely, let P have the Aubin property at $\bar{x}$ for $\bar{p}$ with associated constant κ and neighborhoods U and Q of $\bar{x}$ and $\bar{p}$, respectively. Let f be Lipschitz continuous on $Q \times U$ with constant μ. We already know that the mapping Ω has the Aubin property at $(\bar{x},\bar{y})$ for $\bar{p}$; let λ be the associated constant and $U \times V$ and Q the neighborhoods of $(\bar{x},\bar{y})$ and of $\bar{p}$, respectively. Choose $c > 0$ such that $\mathbb{B}_c(\bar{p}) \subset Q$ and let $a > 0$ satisfy

$$\mathbb{B}_a(\bar{x}) \subset U,\ \mathbb{B}_a(\bar{y}) \subset V \ \text{ and } a\max\{\kappa,\lambda\} \leq c/4.$$

Let $x,x' \in \mathbb{B}_a(\bar{x})$ and $y \in F(x) \cap \mathbb{B}_a(\bar{y})$. Since Ω has the Aubin property and $\bar{p} \in \Omega(\bar{x},\bar{y}) \cap \mathbb{B}_c(\bar{p})$, there exists $p \in \Omega(x,y)$ such that $|p - \bar{p}| \leq \lambda(2a) \leq c/2$. This means that $p \in P(x) \cap \mathbb{B}_{c/2}(\bar{p})$ and from the Aubin property of P there exists $p' \in P(x')$ so that $|p' - p| \leq \kappa|x' - x|$. Thus, $|p' - \bar{p}| \leq \kappa(2a) + c/2 \leq c$. Let $y' = -f(p',x')$. Then $y' \in F(x')$ because $p' \in P(x')$ and the Lipschitz continuity of f gives us

$$|y - y'| = |f(p,x) - f(p',x')| \leq \mu(|p - p'| + |x - x'|) \leq \mu(\kappa + 1)|x - x'|.$$

Hence, F has the Aubin property at $\bar{x}$ for $\bar{y}$. □

In none of the directions of the statement of 3F.14 we can replace the Aubin property by metric regularity. Indeed, the function $f(p,x) = x + p$ satisfies the assumptions, but if we add to it the zero mapping $F \equiv 0$, which is not metrically regular (anywhere), we get the mapping $P(x) = -x$ which is metrically regular (everywhere). Taking the same f and $F(x) = -x$ contradicts the other direction.

3G. Strong Metric Regularity

Although our chief goal in this chapter has been the treatment of solution mappings for which Lipschitz continuous single-valued localizations need not exist or even be a topic of interest, the concepts and results we have built up can shed new light on our earlier work with such localizations through their connection with metric regularity.

Proposition 3G.1 (single-valued localizations and metric regularity)**.** *For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a pair $(\bar{x},\bar{y}) \in \operatorname{gph} F$, the following properties are equivalent:*

(a) *F^{-1} has a Lipschitz continuous single-valued localization s around $\bar{y}$ for $\bar{x}$;*

(b) *F is metrically regular at $\bar{x}$ for $\bar{y}$ and F^{-1} has a localization at $\bar{y}$ for $\bar{x}$ that is nowhere multivalued.*

Indeed, in the circumstances of (b) the localization s in (a) has $\operatorname{lip}(s;\bar{y}) = \operatorname{reg}(F;\bar{x}\,|\,\bar{y})$.

Proof. According to 3E.2 as applied to $S = F^{-1}$, condition (a) is equivalent to F^{-1} having the Aubin property at $\bar{y}$ for $\bar{x}$ and a localization around $\bar{y}$ for $\bar{x}$ that is nowhere multivalued. When F^{-1} has the Aubin property at $\bar{y}$ for $\bar{x}$, by 3E.1 the domain of

F^{-1} contains a neighborhood of $\bar{y}$, hence any localization of F^{-1} at $\bar{y}$ for $\bar{x}$ is actually a localization around $\bar{y}$ for $\bar{x}$. On the other hand, we know from 3E.7 that F^{-1} has the Aubin property at $\bar{y}$ for $\bar{x}$ if and only if F is metrically regular at $\bar{x}$ for $\bar{y}$. That result also relates the constants κ in the two properties and yields for us the final statement. □

Proposition 3G.2 (stability of single-valuedness under perturbation). *Let ν and λ be positive constants such that $\nu\lambda < 1$. Consider a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a pair $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, such that F^{-1} has a Lipschitz continuous single-valued localization s around $\bar{y}$ for $\bar{x}$ with $\operatorname{lip}(s;\bar{y}) < \lambda$. Consider also a function $g : \mathbb{R}^n \to \mathbb{R}^m$ with $\bar{x} \in \operatorname{int}\operatorname{dom} g$ and such that $\operatorname{lip}(g;\bar{x}) < \nu$. Then the mapping $(g+F)^{-1}$ has a localization around $g(\bar{x}) + \bar{y}$ for $\bar{x}$ which is nowhere multivalued.*

Proof. Our hypothesis says that there are neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that for any $y \in V$ the set $F^{-1}(y) \cap U$ consists of exactly one point, $s(y)$, and that the function $s : y \mapsto F^{-1}(y) \cap U$ is Lipschitz continuous on V with Lipschitz constant λ. Let $0 < \nu < \lambda^{-1}$ and choose a function $g : \mathbb{R}^n \to \mathbb{R}^m$ and a neighborhood U' of $\bar{x}$ on which g is Lipschitz continuous with constant ν. We can find neighborhoods $U_0 = \mathbb{B}_\tau(\bar{x}) \subset U \cap U'$ and $V_0 = \mathbb{B}_\varepsilon(g(\bar{x}) + \bar{y}) \subset (g(\bar{x}) + V)$ such that

$$(1) \qquad x \in U_0,\ y \in V_0 \quad \Longrightarrow \quad y - g(x) \in V.$$

Consider now the graphical localization of $V \ni y \mapsto (g+F)^{-1}(y) \cap U_0$. It will be demonstrated that the set $(g+F)^{-1}(y) \cap U_0$ can have at most one element, and that will finish the proof.

Suppose to the contrary that $y \in V_0$ and $x, x' \in U_0$, $x \neq x'$, are such that both x and x' belong to $(g+F)^{-1}(y)$. Clearly $x \in (g+F)^{-1}(y) \cap U_0$ if and only if $x \in U_0$ and $y \in g(x) + F(x)$, or equivalently $y - g(x) \in F(x)$. The latter, in turn, is the same as having $x \in F^{-1}(y - g(x)) \cap U_0 \subset F^{-1}(y - g(x)) \cap U = s(y - g(x))$, where $y - g(x) \in V$ by (1). Then

$$0 < |x - x'| = |s(y - g(x)) - s(y - g(x'))| \leq \lambda |g(x) - g(x')| \leq \lambda\nu|x - x'| < |x - x'|,$$

which is absurd. □

The observation in 3G.1 leads to a definition.

Strong metric regularity. *A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ whose inverse F^{-1} has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$ will be called strongly metrically regular at $\bar{x}$ for $\bar{y}$.*

For a linear mapping represented by an $m \times n$ matrix A, strong metric regularity comes out as the nonsingularity of A and thus requires that $m = n$. Moreover, for any single-valued function $f : \mathbb{R}^n \to \mathbb{R}^m$, strong metric regularity requires $m = n$ by Theorem 1F.1 on the invariance of domain. This property can be seen therefore as corresponding closely to the one in the classical implicit function theorem, except for its focus on Lipschitz continuity instead of continuous differentiability. It was the central property in fact, if not in name, in Robinson's implicit function theorem 2B.1.

The terminology of *strong* metric regularity offers a way of gaining new perspectives on earlier results by translating them into the language of metric regularity. Indeed, strong metric regularity is just metric regularity plus the existence of a single-valued localization of the inverse. According to Theorem 3F.1, metric regularity of a mapping F is stable under addition of a function g with a "small" Lipschitz constant, and so too is local single-valuedness, according to 3G.2. Thus, strong metric regularity must be stable under perturbation in the same way as metric regularity. The corresponding result is a version of the inverse function result in 2B.10 corresponding to the extended form of Robinson's implicit function theorem in 2B.5.

Theorem 3G.3 (inverse function theorem with strong metric regularity)**.** *Let κ and μ be nonnegative constants such that $\kappa\mu < 1$. Consider a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and any $(\bar{x},\bar{y}) \in \operatorname{gph} F$ such that F is strongly metrically regular at $\bar{x}$ for $\bar{y}$ with $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \leq \kappa$ and a function $g : \mathbb{R}^n \to \mathbb{R}^m$ with $\bar{x} \in \operatorname{int}\operatorname{dom} g$ and $\operatorname{lip}(g;\bar{x}) \leq \mu$. Then the mapping $g+F$ is strongly metrically regular at $\bar{x}$ for $g(\bar{x})+\bar{y}$. Moreover,*

$$\operatorname{reg}\Big(g+F;\bar{x}\Big|g(\bar{x})+\bar{y}\Big) \leq \frac{\kappa}{1-\kappa\mu}.$$

Proof. Our hypothesis that F is strongly metrically regular at $\bar{x}$ for $\bar{y}$ implies that a graphical localization of F^{-1} around $(\bar{y},\bar{x})$ is single-valued near $\bar{y}$. Further, by fixing $\lambda > \kappa$ such that $\lambda\mu < 1$ and using Proposition 3G.1, we can get neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that for every $y \in V$ the set $F^{-1}(y) \cap U$ consists of exactly one point, which we may denote by $s(y)$ and know that the function $s : y \mapsto F^{-1}(y) \cap U$ is Lipschitz continuous on V with Lipschitz constant λ. Let $\mu < \nu < \lambda^{-1}$ and choose a neighborhood $U' \subset U$ of $\bar{x}$ on which g is Lipschitz continuous with constant ν. Applying Proposition 3G.2 we obtain that the mapping $(g+F)^{-1}$ has a localization around $g(\bar{x})+\bar{y}$ for $\bar{x}$ which is nowhere multivalued. On the other hand, we know from Theorem 3F.1 that for such g the mapping $g+F$ is metrically regular at $g(\bar{x})+\bar{y}$ for $\bar{x}$. Applying Proposition 3G.1 once more, we complete the proof. $\square$

In much the same way we can state in terms of strong metric regularity an implicit function result paralleling Theorem 2B.7.

Theorem 3G.4 (implicit function theorem with strong metric regularity)**.** *For the generalized equation $f(p,x)+F(x) \ni 0$ with $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and its solution mapping*

$$S : p \mapsto \big\{\, x \,\big|\, f(p,x)+F(x) \ni 0 \big\},$$

consider a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$. Let $h : \mathbb{R}^n \to \mathbb{R}^m$ be a strict estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ and let $h+F$ be strongly metrically regular at $\bar{x}$ for 0 with $\operatorname{reg}(h+F;\bar{x}\,|\,0) \leq \kappa$. Suppose that

$$\kappa\mu < 1 \ \text{ and } \ \widehat{\text{lip}}_p(f;(\bar{p},\bar{x})) \leq \lambda < \infty.$$

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$, moreover with

$$\text{lip}\,(s;\bar{p}) \leq \frac{\kappa\lambda}{1-\kappa\mu}.$$

Many corollaries of this theorem could be stated in a mode similar to that in Section 3F, but the territory has already been covered essentially in Chapter 2. We will get back to this result in Section 5F.

In some situations, metric regularity automatically entails strong metric regularity. That is the case, for instance, for a linear mapping from $\mathbb{R}^n$ to itself represented by an $n \times n$ matrix A. Such a mapping is metrically regular if and only if it is surjective, which means that A has full rank, but then A is nonsingular, so that we have strong metric regularity. More generally, for any mapping which describes the Karush-Kuhn-Tucker optimality system in a nonlinear programming problem, metric regularity implies strong metric regularity. We will prove this fact in Section 4H.

We will describe now another class of mappings for which metric regularity and strong metric regularity come out to be the same thing. This class depends on a localized, set-valued form of the monotonicity concept which appeared in sections 1H and 2F.

Locally monotone mappings. *A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be locally monotone at $\bar{x}$ for $\bar{y}$ if $(\bar{x},\bar{y}) \in \text{gph}\, F$ and for some neighborhood W of $(\bar{x},\bar{y})$, one has*

$$\langle y'-y, x'-x\rangle \geq 0 \quad \text{whenever } (x',y'),(x,y) \in \text{gph}\, F \cap W.$$

Theorem 3G.5 (strong metric regularity of locally monotone mappings)**.** *If a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ that is locally monotone at $\bar{x}$ for $\bar{y}$ is metrically regular at $\bar{x}$ for $\bar{y}$, then it must be strongly metrically regular at $\bar{x}$ for $\bar{y}$.*

Proof. According to 3G.1, all we need to show is that a mapping F which is locally monotone and metrically regular at $\bar{x}$ for $\bar{y}$ must have a localization around $\bar{y}$ for $\bar{x}$ which is nowhere multivalued. Suppose to the contrary that every graphical localization of F^{-1} at $\bar{y}$ for $\bar{x}$ is multivalued. Then there are infinite sequences $y_k \to \bar{y}$ and $x_k, z_k \in F^{-1}(y_k)$, $x_k \to \bar{x}$, $z_k \to \bar{x}$ such that $x_k \neq z_k$ for all k. Let $b_k = |z_k - x_k| > 0$ and $h_k = (z_k - x_k)/b_k$. Then we have

$$(2) \qquad \langle z_k, h_k\rangle = b_k + \langle x_k, h_k\rangle \quad \text{for all } k = 1,2,\ldots.$$

Since the metric regularity of F implies through 3E.7 the Aubin property of F^{-1} at $\bar{y}$ for $\bar{x}$, there exist $\kappa > 0$ and $a > 0$ such that

$$F^{-1}(y) \cap \mathbb{B}_a(\bar{x}) \subset F^{-1}(y') + \kappa|y-y'|\mathbb{B} \quad \text{for all } y, y' \in \mathbb{B}_a(\bar{y}).$$

Choose a sequence of positive numbers τ_k satisfying

$$(3) \qquad \tau_k \searrow 0 \quad \text{and} \quad \tau_k < b_k/(2\kappa).$$

Then for k large, we have $y_k, y_k + \tau_k h_k \in \mathbb{B}_a(\bar{y})$ and $x_k \in F^{-1}(y_k) \cap \mathbb{B}_a(\bar{x})$, and hence there exists $u_k \in F^{-1}(y_k + \tau_k h_k)$ satisfying

$$(4) \qquad |u_k - x_k| \leq \kappa \tau_k.$$

By the local monotonicity of F at $(\bar{x}, \bar{y})$ we have

$$\langle u_k - z_k, y_k + \tau_k h_k - y_k \rangle \geq 0.$$

This, combined with (2), yields

$$(5) \qquad \langle u_k, h_k \rangle \geq \langle z_k, h_k \rangle = b_k + \langle x_k, h_k \rangle.$$

We get from (3), (4) and (5) that

$$b_k + \langle x_k, h_k \rangle \leq \langle u_k, h_k \rangle \leq \langle x_k, h_k \rangle + \kappa \tau_k < \langle x_k, h_k \rangle + (b_k/2),$$

which is impossible. Therefore, F^{-1} must indeed have a localization around $\bar{y}$ for $\bar{x}$ which is not multivalued. □

Observe that strong metric regularity of a mapping F at $\bar{x}$ for $\bar{y}$ automatically implies that $\operatorname{gph} F$ is locally closed at $(\bar{x}, \bar{y})$. The following proposition shows that this remains true when the mapping F is perturbed by a Lipschitz function.

Proposition 3G.6 (local graph closedness from strong metric regularity)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a function which is Lipschitz continuous around $\bar{x}$, let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a set-valued mapping, and suppose that $f + F$ is strongly metrically regular at $\bar{x}$ for $\bar{y}$. Then* $\operatorname{gph} F$ *is locally closed at $(\bar{x}, \bar{y} - f(\bar{x}))$.*

Proof. The assumption that $f + F$ is strongly metrically regular at $\bar{x}$ for $\bar{y}$ means that there exist positive α and β such that the mapping $\mathbb{B}_\beta(\bar{y}) \ni y \mapsto s(y) := (f + F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$ is a Lipschitz continuous function. Let $a > 0$ and $b > 0$ be such that f is Lipschitz continuous in $\mathbb{B}_a(\bar{x})$ with constant l, and moreover $b + la \leq \beta$ and $a \leq \alpha$. Then $\operatorname{dom} s \supset \mathbb{B}_{b+la}(\bar{y})$. Let $(x_k, z_k) \in \operatorname{gph} F \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(\bar{y} - f(\bar{x}))$ converge to (x, z) as $k \to \infty$. Then we have

$$|z_k + f(x_k) - \bar{y}| \leq |z_k - (\bar{y} - f(\bar{x}))| + |f(x_k) - f(\bar{x})| \leq b + la.$$

Then $z_k + f(x_k) \in \mathbb{B}_{b+la}(\bar{y})$ and also $(x_k, z_k + f(x_k)) \in \operatorname{gph}(f + F)$; hence, $x_k = s(z_k + f(x_k))$. Passing to the limit with $k \to \infty$ we get $x = s(z + f(x)) = (f + F)^{-1}(z + f(x)) \cap \mathbb{B}_a(\bar{x})$, that is, $z + f(x) \in f(x) + F(x)$, hence $(x, z) \in \operatorname{gph} F$. Thus, $\operatorname{gph} F$ is locally closed at $(\bar{x}, \bar{y} - f(\bar{x}))$. □

3H. Calmness and Metric Subregularity

A "one-point" variant of the Aubin property can be defined for set-valued mappings in the same way as calmness of functions, and this leads to another natural topic of investigation.

Calmness. *A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said be calm at $\bar{y}$ for $\bar{x}$ if $(\bar{y},\bar{x}) \in \operatorname{gph} S$, and there is a constant $\kappa \geq 0$ along with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$(1) \qquad e(S(y) \cap U, S(\bar{y})) \leq \kappa|y - \bar{y}| \quad \text{for all } y \in V.$$

Equivalently, the property in (1) can be also written as

$$(2) \qquad S(y) \cap U \subset S(\bar{y}) + \kappa|y - \bar{y}|\mathbb{B} \quad \text{for all } y \in V$$

although perhaps with larger constant κ. The infimum of κ over all such combinations of κ, U and V is called the calmness modulus of S at $\bar{y}$ for $\bar{x}$ and denoted by $\operatorname{clm}(S;\bar{y}\,|\,\bar{x})$. The absence of this property is signaled by $\operatorname{clm}(S;\bar{y}\,|\,\bar{x}) = \infty$.

As in the case of the Lipschitz modulus $\operatorname{lip}(S;\bar{y}\,|\,\bar{x})$ in 3E, it is not claimed that (1) and (2) are themselves equivalent; anyway, the infimum furnishing $\operatorname{clm}(S;\bar{y}\,|\,\bar{x})$ is the same with respect to (2) as with respect to (1).

In the case when S is not multivalued, the definition of calmness reduces to that of a function given in Section 1C relative to a neighborhood V of $\bar{y}$; $\operatorname{clm}(S;\bar{y}\,|\,S(\bar{y})) = \operatorname{clm}(S;\bar{y})$. Indeed, for any $y \in V \setminus \operatorname{dom} S$ the inequality (1) holds automatically.

Clearly, for closed-valued mappings outer Lipschitz continuity implies calmness. In particular, we get the following fact from Theorem 3D.1.

Proposition 3H.1 (calmness of polyhedral mappings)**.** *Any mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ whose graph is the union of finitely many polyhedral convex sets is calm with the same constant κ at any $\bar{y}$ for any $\bar{x}$ whenever $(\bar{y},\bar{x}) \in \operatorname{gph} S$.*

In particular, any linear mapping is calm at any point of its graph, and this is also true for its inverse. For comparison, the inverse of a linear mapping has the Aubin property at some point if and only if the mapping is surjective.

Exercise 3H.2 (local outer Lipschitz continuity under truncation)**.** *Show that a mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ with $(\bar{y},\bar{x}) \in \operatorname{gph} S$ and with $S(\bar{y})$ closed and convex is calm at $\bar{y}$ for $\bar{x}$ if and only if there is a neighborhood U of $\bar{x}$ such that the truncated mapping $y \mapsto S(y) \cap U$ is outer Lipschitz continuous at $\bar{y}$.*

Guide. Mimic the proof of 3E.3 with $y = \bar{y}$. □

Is there a "one-point" variant of the metric regularity which would characterize calmness of the inverse, in the way metric regularity characterizes the Aubin property of the inverse? Yes, as we explore next.

Metric subregularity. *A mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is called* metrically subregular *at* $\bar{x}$ *for* $\bar{y}$ *if* $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ *and there exists* $\kappa \geq 0$ *along with neighborhoods* U *of* $\bar{x}$ *and* V *of* $\bar{y}$ *such that*

$$(3) \qquad d(x, F^{-1}(\bar{y})) \leq \kappa d(\bar{y}, F(x) \cap V) \quad \textit{for all } x \in U.$$

The infimum of all κ *for which this holds is the* modulus *of metric subregularity, denoted by* $\operatorname{subreg}(F; \bar{x} \,|\, \bar{y})$*. The absence of metric subregularity is signaled by* $\operatorname{subreg}(F; \bar{x} \,|\, \bar{y}) = \infty$.

The main difference between metric subregularity and metric regularity is that the data input $\bar{y}$ is now fixed and not perturbed to a nearby y. Since $d(\bar{y}, F(x)) \leq d(\bar{y}, F(x) \cap V)$, it is clear that subregularity is a weaker condition than metric regularity, and

$$\operatorname{subreg}(F; \bar{x} \,|\, \bar{y}) \leq \operatorname{reg}(F; \bar{x} \,|\, \bar{y}).$$

The following result reveals the equivalence of metric subregularity of a mapping with calmness of its inverse:

Theorem 3H.3 (characterization by calmness of the inverse)**.** *For a mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$*, let* $F(\bar{x}) \ni \bar{y}$*. Then* F *is metrically subregular at* $\bar{x}$ *for* $\bar{y}$ *if and only if its inverse* $F^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *is calm at* $\bar{y}$ *for* $\bar{x}$*, in which case*

$$\operatorname{clm}(F^{-1}; \bar{y} \,|\, \bar{x}) = \operatorname{subreg}(F; \bar{x} \,|\, \bar{y}).$$

Proof. Assume first that there exist a constant $\kappa > 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$(4) \qquad F^{-1}(y) \cap U \subset F^{-1}(\bar{y}) + \kappa |y - \bar{y}| \mathbb{B} \quad \text{for all } y \in V.$$

Let $x \in U$. If $F(x) \cap V = \emptyset$, then the right side of (3) is ∞ and we are done. If not, having $x \in U$ and $y \in F(x) \cap V$ is the same as having $x \in F^{-1}(y) \cap U$ and $y \in V$. For such x and y, the inclusion in (4) requires the ball $x + \kappa |y - \bar{y}| \mathbb{B}$ to have nonempty intersection with $F^{-1}(\bar{y})$. Then $d(x, F^{-1}(\bar{y})) \leq \kappa |y - \bar{y}|$. Thus, for any $x \in U$, we must have $d(x, F^{-1}(\bar{y})) \leq \inf_y \{ \kappa |y - \bar{y}| \,\big|\, y \in F(x) \cap V \}$, which is (3). This shows that (4) implies (3) and that

$$\inf \{ \kappa \,\big|\, U, V, \kappa \text{ satisfying } (4) \} \geq \inf \{ \kappa \,\big|\, U, V, \kappa \text{ satisfying } (3) \},$$

the latter being by definition $\operatorname{subreg}(F; \bar{x} \,|\, \bar{y})$.

For the opposite direction, we have to demonstrate that if $\operatorname{subreg}(F; \bar{x} \,|\, \bar{y}) < \kappa < \infty$, then (4) holds for some choice of neighborhoods U and V. Consider any κ' with $\operatorname{subreg}(F; \bar{x} \,|\, \bar{y}) < \kappa' < \kappa$. For this κ', there exist U and V such that $d(x, F^{-1}(\bar{y})) \leq \kappa' d(\bar{y}, F(x) \cap V)$ for all $x \in U$. Then we have $d(x, F^{-1}(\bar{y})) \leq \kappa' |y - \bar{y}|$ when $x \in U$ and $y \in F(x) \cap V$, or equivalently $y \in V$ and $x \in F^{-1}(y) \cap U$. Fix $y \in V$. If $y = \bar{y}$ there is nothing to prove; let $y \neq \bar{y}$. If $x \in F^{-1}(y) \cap U$, then $d(x, F^{-1}(\bar{y})) \leq \kappa' |y - \bar{y}| < \kappa |y - \bar{y}|$. Then there must be a point of $x' \in F^{-1}(\bar{y})$ having $|x' - x| \leq \kappa |y - \bar{y}|$. Hence we have (4), as required, and the proof if complete. $\square$

As we will see next, there is no need at all to mention a neighborhood V of $\bar{y}$ in the description of calmness and subregularity in (2) and (3).

Exercise 3H.4 (equivalent formulations)**.** *For a mapping $F : I\!\!R^n \rightrightarrows I\!\!R^m$ and a point $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ metric subregularity of F at $\bar{x}$ for $\bar{y}$ with constant $\kappa > 0$ is equivalent simply to the existence of a neighborhood U of $\bar{x}$ such that*

$$(5) \qquad d(x, F^{-1}(\bar{y})) \leq \kappa d(\bar{y}, F(x)) \quad \text{for all } x \in U,$$

whereas the calmness of F^{-1} at $\bar{y}$ for $\bar{x}$ with constant $\kappa > 0$ can be identified with the existence of a neighborhood U of $\bar{x}$ such that

$$(6) \qquad F^{-1}(y) \cap U \subset F^{-1}(\bar{y}) + \kappa|y - \bar{y}| I\!\!B \quad \text{for all } y \in I\!\!R^m.$$

Guide. Assume that (3) holds with $\kappa > 0$ and associated neighborhoods U and V. We can choose within V a neighborhood of the form $V' = I\!\!B_\varepsilon(\bar{y})$ for some $\varepsilon > 0$. Let $U' := U \cap (\bar{x} + \varepsilon\kappa I\!\!B)$ and pick $x \in U'$. If $F(x) \cap V' \neq \emptyset$ then $d(\bar{y}, F(x) \cap V') = d(\bar{y}, F(x))$ and (3) becomes (5) for this x. Otherwise, $F(x) \cap V' = \emptyset$ and then

$$d(\bar{y}, F(x)) \geq \varepsilon \geq \frac{1}{\kappa}|x - \bar{x}| \geq \frac{1}{\kappa} d(x, F^{-1}(\bar{y})),$$

which is (5).

Similarly, (6) entails the calmness in (4), so attention can be concentrated on showing that we can pass from (4) to (6) under an adjustment in the size of U. We already know from 3H.3 that the calmness condition in (4) leads to the metric subregularity in (3), and further, from the argument just given, that such subregularity yields the condition in (5). But that condition can be plugged into the argument in the proof of 3H.3, by taking $V = I\!\!R^m$, to get the corresponding calmness property with $V = I\!\!R^m$ but with U replaced by a smaller neighborhood of $\bar{x}$. □

Although we could take (5) as a redefinition of metric subregularity, we prefer to retain the neighborhood V in (3) in order to underscore the parallel with metric regularity; similarly for calmness.

Does metric subregularity enjoy stability properties under perturbation resembling those of metric regularity and strong metric regularity? In other words, does metric subregularity obey the general paradigm of the implicit function theorem, as developed in chapters 1 and 2? The answer to this question turns out to be *no* even for simple functions. Indeed, the function $f(x) = x^2$ is clearly not metrically subregular at 0 for 0, but its derivative $Df(0)$, which is the zero mapping, is metrically subregular.

More generally, every linear mapping $A : I\!\!R^n \to I\!\!R^m$ is metrically subregular, and hence the derivative mapping of any smooth function is metrically subregular. But of course, not every smooth function is subregular. *For this reason, there cannot be an implicit mapping theorem in the vein of 3F.8 in which metric regularity is replaced by metric subregularity, even for the classical case of an equation with smooth f and no set-valued F.*

An illuminating but more intricate counterexample of instability of metric subregularity of set-valued mappings is as follows. In $\mathbb{R} \times \mathbb{R}$, let $\operatorname{gph} F$ be the set of all (x,y) such that $x \geq 0$, $y \geq 0$ and $yx = 0$. Then $F^{-1}(0) = [0,\infty) \supset F^{-1}(y)$ for all y, so F is metrically subregular at $\bar{x} = 0$ for $\bar{y} = 0$, even "globally" with $\kappa = 0$. By Theorem 3H.3, $\operatorname{subreg}(F;0\,|\,0) = 0$.

Consider, however the function $f(x) = -x^2$ for which $f(0) = 0$ and $\nabla f(0) = 0$. The perturbed mapping $f + F$ has $(f+F)^{-1}$ single-valued everywhere: $(f+F)^{-1}(y) = 0$ when $y \geq 0$, and $(f+F)^{-1}(y) = \sqrt{|y|}$ when $y \leq 0$. This mapping is not calm at 0 for 0. Then, from Theorem 3H.3 again, $f+F$ is not metrically subregular; we have $\operatorname{subreg}(f+F;0\,|\,0) = \infty$.

To conclude this section, we point out some other properties which are, in a sense, derived from metric regularity but, like subregularity, lack such kind of stability. One of these properties is the openness which we introduced in Section 1F: a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be open at $\bar{x} \in \operatorname{dom} f$ when for every $a > 0$ there exists $b > 0$ such that $f(\bar{x} + a \operatorname{int} \mathbb{B}) \supset f(\bar{x}) + b \operatorname{int} \mathbb{B}$. It turns out that this property likewise fails to be preserved when f is perturbed to $f + g$ by a function g with $\operatorname{lip}(g;\bar{x}) = 0$. As an example, consider the zero function $f \equiv 0$ acting from $\mathbb{R}$ to $\mathbb{R}$ which is not open at the origin, and its perturbation $f + g$ with $g = x^3$ having $\operatorname{lip}(g;0) = |g'(0)| = 0$, which is open at the origin. A "metric regularity variant" of the openness property, equally failing to be preserved under small Lipschitz continuous perturbations, as shown by this same example, is the requirement that $(\bar{x},\bar{y}) \in \operatorname{gph} F$ and $d(\bar{x}, F^{-1}(y)) \leq \kappa |y - \bar{y}|$ for y close to $\bar{y}$.

If we consider calmness as a local version of the outer Lipschitz continuity, then it might seem to be worthwhile to define a local version of inner Lipschitz continuity, introduced in Section 3D. For a mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ with $(\bar{y},\bar{x}) \in \operatorname{gph} S$, this would refer to the existence of neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$(7) \qquad S(\bar{y}) \cap U \subset S(y) + \kappa |y - \bar{y}| \mathbb{B} \quad \text{for all } y \in V.$$

We will not give a name to this property here, or a name to the associated property of the inverse of a mapping satisfying (7). We will only demonstrate, by an example, that the property of the inverse associated to (7), similar to metric subregularity, is not stable under perturbation, in the sense we have been exploring, and hence does not support the implicit function theorem paradigm.

Consider the mapping $S : \mathbb{R} \rightrightarrows \mathbb{R}$ whose values are the set of three points $\{-\sqrt{y}, 0, \sqrt{y}\}$ for all $y \geq 0$ and the empty set for $y < 0$. This mapping has the property in (7) at $\bar{y} = 0$ for $\bar{x} = 0$. Now consider the inverse S^{-1} and add to it the function $g(x) = -x^2$, which has zero derivative at $\bar{x} = 0$. The sum $S^{-1} + g$ is the mapping whose value at $x = 0$ is the interval $[0,\infty)$ but is just zero for $x \neq 0$. The inverse $(S^{-1} + g)^{-1}$ has $(-\infty,\infty)$ as its value for $y = 0$, but 0 for $y > 0$ and the empty set for $y < 0$. Clearly, this inverse does not have the property displayed in (7) at $\bar{y} = 0$ for $\bar{x} = 0$. It should be noted that for special cases of mappings with particular perturbations one might still obtain stability of metric subregularity, or the property associated to (7), but we shall not go into this further.

3I. Strong Metric Subregularity

The handicap of serious instability of calmness and metric subregularity can be obviated by passing to strengthened forms of these properties.

Isolated calmness. *A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to have the isolated calmness property if it is calm at $\bar{y}$ for $\bar{x}$ and, in addition, S has a graphical localization at $\bar{y}$ for $\bar{x}$ that is single-valued at $\bar{y}$ itself (with value $\bar{x}$). Specifically, this refers to the existence of a constant $\kappa \geq 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$(1) \qquad |x-\bar{x}| \leq \kappa|y-\bar{y}| \quad \text{when } x \in S(y) \cap U \text{ and } y \in V.$$

Observe that in this definition $S(\bar{y}) \cap U$ is a singleton, namely the point $\bar{x}$, so $\bar{x}$ is an isolated point in $S(\bar{y})$, hence the terminology. Isolated calmness can equivalently be defined as the existence of a (possibly slightly larger) constant κ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$(2) \qquad S(y) \cap U \subset \bar{x} + \kappa|y-\bar{y}|\mathbb{B} \quad \text{when } y \in V.$$

For a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^n$, isolated calmness holds at every point, whereas isolated calmness of A^{-1} holds at some point of $\operatorname{dom} A^{-1}$ if and only if A is nonsingular. More generally we have the following fact through Theorem 3D.1 for polyhedral mappings, as defined there.

Proposition 3I.1 (isolated calmness of polyhedral mappings)**.** *A polyhedral mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ has the isolated calmness property at $\bar{y}$ for $\bar{x}$ if and only if $\bar{x}$ is an isolated point of $S(\bar{y})$.*

Once again we can ask whether there is a property of a mapping that corresponds to isolated calmness of its inverse. Such a property exists and, parallel to strong metric regularity in 3G.1, can be deduced from an equivalence relation.

Proposition 3I.2 (isolated calmness and metric subregularity)**.** *For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a pair $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, the following properties are equivalent:*

(a) *F^{-1} has the isolated calmness property at $\bar{y}$ for $\bar{x}$;*

(b) *F is metrically subregular at $\bar{x}$ for $\bar{y}$ and F^{-1} has no localization at $\bar{y}$ for $\bar{x}$ that is multivalued at $\bar{y}$.*

Proof. According to 3H.3 condition (a) implies that F is metrically subregular at $\bar{x}$ for $\bar{y}$; furthermore, the definition of isolated calmness yields that F^{-1} has no localization at $\bar{y}$ for $\bar{x}$ that is multivalued at $\bar{y}$. Hence (a) implies (b). The proof of the converse implication is symmetric. □

In order to unify the terminology, we name the new property as follows.

Strong metric subregularity. *A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be strongly metrically subregular at $\bar{x}$ for $\bar{y}$ if $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ and there is a constant $\kappa \geq 0$ along with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that*

$$(3) \qquad |x-\bar{x}| \le \kappa d(\bar{y}, F(x)\cap V) \quad \textit{for all } x \in U.$$

Clearly, the infimum of κ for which (3) holds is equal to $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$.

Note however that, in general, the condition $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y}) < \infty$ is *not* a characterization of strong metric subregularity, but becomes such a criterion under the isolatedness assumption. As an example, observe that, for a linear mapping A is always metrically subregular at 0 for 0, but it is strongly metrically subregular at 0 for 0 if and only if $\ker A$ consists of just 0, which corresponds to A being injective. The equivalence of strong metric subregularity and isolated calmness of the inverse is shown next:

Theorem 3I.3 (characterization by isolated calmness of the inverse)**.** *A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ if and only if its inverse F^{-1} has the isolated calmness property at $\bar{y}$ for $\bar{x}$.*

Specifically, for any $\kappa > \operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$ there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$(4) \qquad F^{-1}(y)\cap U \subset \bar{x} + \kappa|y-\bar{y}|\mathbb{B} \quad \textit{when } y \in V.$$

Moreover, the infimum of all κ such that the inclusion (4) holds for some neighborhoods U and V actually equals $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$.

Proof. Assume first that F is strongly subregular at $\bar{x}$ for $\bar{y}$. Let $\kappa > \operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$. Then there are neighborhoods U for $\bar{x}$ and V for $\bar{y}$ such that (3) holds with the indicated κ. Consider any $y \in V$. If $F^{-1}(y)\cap U = \emptyset$, then (4) holds trivially. If not, let $x \in F^{-1}(y)\cap U$. This entails $y \in F(x)\cap V$, hence $d(\bar{y}, F(x)\cap V) \le |y-\bar{y}|$ and consequently $|x-\bar{x}| \le \kappa|y-\bar{y}|$ by (3). Thus, $x \in \bar{x}+\kappa|y-\bar{y}|\mathbb{B}$, and we conclude that (4) holds. Also, we see that $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$ is not less than the infimum of all κ such that (4) holds for some choice of U and V.

For the converse, suppose (4) holds for some κ and neighborhoods U and V. Consider any $x \in U$. If $F(x)\cap V = \emptyset$ the right side of (3) is ∞ and there is nothing more to prove. If not, for an arbitrary $y \in F(x)\cap V$ we have $x \in F^{-1}(y)\cap U$, and therefore $x \in \bar{x}+\kappa|y-\bar{y}|\mathbb{B}$ by (4), which means $|x-\bar{x}| \le \kappa|y-\bar{y}|$. This being true for all $y \in F(x)\cap V$, we must have $|x-\bar{x}| \le \kappa d(\bar{y}, F(x)\cap V)$. Thus, (3) holds, and in particular we have $\kappa \ge \operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$. Therefore, the infimum of κ in (4) equals $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$. □

Observe also, through 3H.4, that the neighborhood V in (2) and (3) can be chosen to be the entire space $\mathbb{R}^m$, by adjusting the size of U; that is, strong metric subregularity as in (3) with constant κ is equivalent to the existence of a neighborhood U' of $\bar{x}$ such that

$$(5) \qquad |x-\bar{x}| \le \kappa d(\bar{y}, F(x)) \quad \text{for all } x \in U'.$$

Accordingly, the associated isolated calmness of the inverse is equivalent to the existence of a neighborhood U' of $\bar{x}$ such that

$$(6) \qquad F^{-1}(y)\cap U' \subset \bar{x}+\kappa|y-\bar{y}|\mathbb{B} \quad \text{when } y \in \mathbb{R}^m.$$

Exercise 3I.4. *Provide direct proofs of the equivalence of (3) and (5), and (4) and (6), respectively.*

Guide. Use the argument in the proof of 3H.4. □

Similarly to the distance function characterization in Theorem 3E.6 for the Aubin property, the isolated calmness property is characterized by uniform calmness of the distance function associated with the inverse mapping:

Theorem 3I.5 (distance function characterization of strong metric subregularity)**.** *For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a point $(\bar{x},\bar{y}) \in \operatorname{gph} F$, suppose that $\bar{x}$ is an isolated point in $F^{-1}(\bar{y})$ and moreover*

$$\text{(7)} \qquad \bar{x} \in \liminf_{y \to \bar{y}} F^{-1}(y).$$

Consider the function $s(y,x) = d(x, F^{-1}(y))$. Then the mapping F is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ if and only if s is calm with respect to y uniformly in x at $(\bar{y},\bar{x})$, in which case

$$\widehat{\operatorname{clm}}_y(s;(\bar{y},\bar{x})) = \operatorname{subreg}(F;\bar{x}\,|\,\bar{y}).$$

Proof. Let F be strongly metrically subregular at $\bar{x}$ for $\bar{y}$ and let $\kappa > \operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$. Let (5) and (6) hold with $U' = \mathbb{B}_a(\bar{x})$ and also $F^{-1}(\bar{y}) \cap \mathbb{B}_a(\bar{x}) = \bar{x}$. Let $b > 0$ be such that, according to (7), $F^{-1}(y) \cap \mathbb{B}_a(\bar{x}) \neq \emptyset$ for all $y \in \mathbb{B}_b(\bar{y})$. Make b smaller if necessary so that $b \le a/(10\kappa)$. Choose $y \in \mathbb{B}_b(\bar{y})$ and $x \in \mathbb{B}_{a/4}(\bar{x})$; then from (6) we have

$$\text{(8)} \qquad d(\bar{x}, F^{-1}(y) \cap \mathbb{B}_a(\bar{x})) \le \kappa|y - \bar{y}|.$$

Since all points in $F^{-1}(\bar{y})$ except $\bar{x}$ are at distance from x more than $a/4$ we obtain

$$\text{(9)} \qquad d(x, F^{-1}(\bar{y})) = |x - \bar{x}|.$$

Utilizing (8), we get

$$\text{(10)} \qquad \begin{aligned} d(x, F^{-1}(y)) &\le |x - \bar{x}| + d(\bar{x}, F^{-1}(y)) \\ &\le |x - \bar{x}| + d(\bar{x}, F^{-1}(y) \cap \mathbb{B}_a(\bar{x})) \le |x - \bar{x}| + \kappa|y - \bar{y}|. \end{aligned}$$

Then, taking (9) into account, we have

$$\text{(11)} \qquad s(y,x) - s(\bar{y},x) = d(x, F^{-1}(y)) - d(x, F^{-1}(\bar{y})) \le \kappa|y - \bar{y}|.$$

Let $\tilde{x}$ be a projection of x on $\operatorname{cl} F^{-1}(y)$. Using (10), we obtain

$$|x - \tilde{x}| = d(x, F^{-1}(y)) \le |x - \bar{x}| + \kappa|y - \bar{y}| \le a/4 + \kappa b \le a/4 + \kappa a/(10\kappa) < a/2,$$

and consequently

$$|\bar{x} - \tilde{x}| \le |\bar{x} - x| + |x - \tilde{x}| \le a/4 + a/2 = 3a/4 < a.$$

Therefore,

$$d(x, F^{-1}(y) \cap \mathbb{B}_a(\bar{x})) = d(x, F^{-1}(y)). \tag{12}$$

According to (6),

$$d(x, \bar{x} + \kappa|y - \bar{y}|\mathbb{B}) \leq d(x, F^{-1}(y) \cap \mathbb{B}_a(\bar{x}))$$

and then, by (12),

$$|x - \bar{x}| - \kappa|y - \bar{y}| \leq d(x, F^{-1}(y) \cap \mathbb{B}_a(\bar{x})) = d(x, F^{-1}(y)). \tag{13}$$

Plugging (9) into (13), we conclude that

$$s(\bar{y}, x) - s(y, x) = d(x, F^{-1}(\bar{y})) - d(x, F^{-1}(y)) \leq \kappa|y - \bar{y}|. \tag{14}$$

Since x and y were arbitrarily chosen in dom s and close to $\bar{x}$ and $\bar{y}$, respectively, we obtain by combining (11) and (14) that $\widehat{\mathrm{clm}}_y(s; (\bar{y}, \bar{x})) \leq \kappa$, hence

$$\widehat{\mathrm{clm}}_y(s; (\bar{y}, \bar{x})) \leq \mathrm{subreg}(F; \bar{x}\,|\,\bar{y}). \tag{15}$$

To show the converse inequality, let $\kappa > \widehat{\mathrm{clm}}_y(s; (\bar{y}, \bar{x}))$; then there exists $a > 0$ such that $s(\cdot, x)$ is calm on $\mathbb{B}_a(\bar{y})$ with constant κ uniformly in $x \in \mathbb{B}_a(\bar{x})$. Adjust a so that $F^{-1}(\bar{y}) \cap \mathbb{B}_a(\bar{x}) = \bar{x}$. Pick any $x \in \mathbb{B}_{a/3}(\bar{x})$. If $F(x) = \emptyset$, (5) holds automatically. If not, choose any $y \in \mathbb{R}^m$ such that $(x, y) \in \mathrm{gph}\, F$. Since $s(y, x) = 0$, we have

$$|x - \bar{x}| = d(x, F^{-1}(\bar{y})) = s(\bar{y}, x) \leq s(y, x) + \kappa|y - \bar{y}| = \kappa|y - \bar{y}|.$$

Since y is arbitrarily chosen in $F(x)$, this gives us (5). This means that F is strongly subregular at $\bar{x}$ for $\bar{y}$ with constant κ and hence

$$\widehat{\mathrm{clm}}_y(s; (\bar{y}, \bar{x})) \geq \mathrm{subreg}(F; \bar{x}\,|\,\bar{y}).$$

Combining this with (15) brings the proof to a finish. □

Exercise 3I.6 (counterexample). *Show that the mapping $F : \mathbb{R} \rightrightarrows \mathbb{R}$ given by*

$$F(x) = \begin{cases} x & \text{if } 0 \leq x < 1, \\ \mathbb{R} & \text{if } x \geq 1, \\ \emptyset & \text{if } x < 0 \end{cases}$$

does not satisfy condition (7) and has $\mathrm{subreg}(F; 0\,|\,0) = 1$ *while* $\widehat{\mathrm{clm}}_y(s; (0, 0)) = \infty$.

We look next at perturbations of F by single-valued mappings g in the pattern that was followed for the other regularity properties considered in the preceding sections.

Theorem 3I.7 (inverse mapping theorem for strong metric subregularity)**.** *Consider a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a point $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ such that F is strongly metrically subregular at $\bar{x}$ for $\bar{y}$. Consider also a function $g : \mathbb{R}^n \to \mathbb{R}^m$ with $\bar{x} \in \operatorname{dom} g$. Let κ and μ be nonnegative constants such that*

$$\kappa\mu < 1, \quad \operatorname{subreg}(F;\bar{x}\,|\,\bar{y}) \leq \kappa \quad \textit{and} \quad \operatorname{clm}(g;\bar{x}) \leq \mu.$$

Then

$$\operatorname{subreg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) \leq \frac{\kappa}{1-\kappa\mu}.$$

Proof. Choose κ and μ as in the statement of the theorem and let $\lambda > \kappa$, $\nu > \mu$ be such that $\lambda\nu < 1$. Pick $g : \mathbb{R}^n \to \mathbb{R}^m$ with $\operatorname{clm}(g;\bar{x}) < \nu$. Without loss of generality, let $g(\bar{x}) = 0$; then there exists $a > 0$ such that

$$(16) \qquad |g(x)| \leq \nu|x-\bar{x}| \quad \text{when } x \in \mathbb{B}_a(\bar{x}).$$

Since $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y}) < \lambda$, we can arrange, by taking a smaller if necessary, that

$$(17) \qquad |x-\bar{x}| \leq \lambda|y-\bar{y}| \quad \text{when } (x,y) \in \operatorname{gph} F \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_a(\bar{y})).$$

Let $\nu' = \max\{1, \nu\}$ and consider any

$$(18) \qquad z \in \mathbb{B}_{a/2}(\bar{y}) \text{ with } x \in (g+F)^{-1}(z) \cap \mathbb{B}_{a/2\nu'}(\bar{x}).$$

These relations entail $z \in g(x)+F(x)$, hence $z = y+g(x)$ for some $y \in F(x)$. From (16) and since $x \in \mathbb{B}_{a/(2\nu')}(\bar{x})$, we have $|g(x)| \leq \nu a/(2\nu') \leq a/2$ (inasmuch as $\nu' \geq \nu$). Using the equality $y-\bar{y} = z-g(x)-\bar{y}$ we get $|y-\bar{y}| \leq |z-\bar{y}|+|g(x)| \leq (a/2)+(a/2) = a$. However, because $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_a(\bar{y}))$, through (17),

$$|x-\bar{x}| \leq \lambda|(z-g(x))-\bar{y}| \leq \lambda|z-\bar{y}|+\lambda|g(x)| \leq \lambda|z-\bar{y}|+\lambda\nu|x-\bar{x}|,$$

hence $|x-\bar{x}| \leq \lambda/(1-\lambda\nu)|z-\bar{y}|$. Since x and z are chosen as in (18) and λ and ν could be arbitrarily close to κ and μ, respectively, the proof is complete. □

Corollaries that parallel those for metric regularity given in Section 3F can immediately be derived.

Corollary 3I.8 (detailed estimate)**.** *Consider a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ which is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ and a function $g : \mathbb{R}^n \to \mathbb{R}^m$ such that*

$$\operatorname{subreg}(F;\bar{x}\,|\,\bar{y}) > 0 \quad \textit{and} \quad \operatorname{subreg}(F;\bar{x}\,|\,\bar{y}) \cdot \operatorname{clm}(g;\bar{x}) < 1.$$

Then the mapping $g+F$ is strongly metrically subregular at $\bar{x}$ for $g(\bar{x})+\bar{y}$, and one has

$$\operatorname{subreg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) \leq \left(\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})^{-1} - \operatorname{clm}(g;\bar{x})\right)^{-1}.$$

This result implies in particular that the property of strong metric subregularity is preserved under perturbations with zero calmness moduli. The only difference with the corresponding results for metric regularity in Section 3F is that now a larger class of perturbation is allowed with first-order approximations replacing the strict first-order approximations.

Corollary 3I.9 (utilizing first-order approximations). *Consider $F : I\!R^n \rightrightarrows I\!R^m$, a point $(\bar{x},\bar{y}) \in \operatorname{gph} F$ and two functions $f : I\!R^n \to I\!R^m$ and $g : I\!R^n \to I\!R^m$ with $\bar{x} \in \operatorname{int} \operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g$ which are first-order approximations to each other at $\bar{x}$. Then the mapping $f+F$ is strongly metrically subregular at $\bar{x}$ for $f(\bar{x})+\bar{y}$ if and only if $g+F$ is strongly metrically subregular at $\bar{x}$ for $g(\bar{x})+\bar{y}$, in which case*

$$\operatorname{subreg}(f+F;\bar{x}\,|\,f(\bar{x})+\bar{y}) \;=\; \operatorname{subreg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}).$$

This corollary takes a more concrete form when the first-order approximation is represented by a linearization:

Corollary 3I.10 (linearization). *Let $M = f+F$ for mappings $f : I\!R^n \to I\!R^m$ and $F : I\!R^n \rightrightarrows I\!R^m$, and let $\bar{y} \in M(\bar{x})$. Suppose f is differentiable at $\bar{x}$, and let*

$$M_0 = h+F \;\text{ for }\; h(x) = f(\bar{x}) + \nabla f(\bar{x})(x-\bar{x}).$$

Then M is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ if and only if M_0 has this property. Moreover $\operatorname{subreg}(M;\bar{x}\,|\,\bar{y}) = \operatorname{subreg}(M_0;\bar{x}\,|\,\bar{y})$.

Through 3I.1, the result in Corollary 3I.10 could equally well be stated in terms of the isolated calmness property of M^{-1} in relation to that of M_0^{-1}. We can specialize that result in the following way.

Corollary 3I.11 (linearization with polyhedrality). *Let $M : I\!R^n \rightrightarrows I\!R^m$ with $\bar{y} \in M(\bar{x})$ be of the form $M = f+F$ for $f : I\!R^n \to I\!R^m$ and $F : I\!R^n \rightrightarrows I\!R^m$ such that f is differentiable at $\bar{x}$ and F is polyhedral. Let $M_0(x) = f(\bar{x}) + \nabla f(\bar{x})(x-\bar{x}) + F(x)$. Then M^{-1} has the isolated calmness property at $\bar{y}$ for $\bar{x}$ if and only if $\bar{x}$ is an isolated point of $M_0^{-1}(\bar{y})$.*

Proof. This applies 3I.1 in the framework of the isolated calmness restatement of 3I.10 in terms of the inverses. □

Applying Corollary 3I.10 to the case where F is the zero mapping, we obtain yet another inverse function theorem in the classical setting:

Corollary 3I.12 (an inverse function result). *Let $f : I\!R^n \to I\!R^m$ be differentiable at $\bar{x}$ and such that $\ker \nabla f(\bar{x}) = \{0\}$. Then there exist $\kappa > 0$ and a neighborhood U of $\bar{x}$ such that*

$$|x-\bar{x}| \le \kappa |f(x) - f(\bar{x})| \quad \text{for every } x \in U.$$

Proof. This comes from (5). □

Next, we state and prove an implicit function theorem for strong metric subregularity:

Theorem 3I.13 (implicit mapping theorem with strong metric subregularity)**.** *For the generalized equation $f(p,x)+F(x)\ni 0$ and its solution mapping*

$$S : p \mapsto \{x \mid f(p,x)+F(x)\ni 0\},$$

consider a pair $(\bar{p},\bar{x})$ with $\bar{x}\in S(\bar{p})$. Let $h:\mathbb{R}^n\to\mathbb{R}^m$ be an estimator of f with respect to x at $(\bar{p},\bar{x})$ with constant μ and let $h+F$ be strongly metrically subregular at $\bar{x}$ for 0 with $\operatorname{subreg}(h+F;\bar{x}\,|\,0)\le\kappa$. Suppose that

$$\kappa\mu<1 \quad \textit{and} \quad \widehat{\operatorname{clm}}_p(f;(\bar{p},\bar{x}))\le\lambda<\infty. \tag{19}$$

Then S has the isolated calmness property at $\bar{p}$ for $\bar{x}$, moreover with

$$\operatorname{clm}(S;\bar{p}\,|\,\bar{x}) \le \frac{\kappa\lambda}{1-\kappa\mu}.$$

Proof. The proof goes along the lines of the proof of Theorem 3I.7 with different choice of constants. Let κ, μ and λ be as required and let $\delta>\kappa$ and $\nu>\mu$ be such that $\delta\nu<1$. Let $\gamma>\lambda$. By the assumptions for the mapping $h+F$ and the functions f and h, there exist positive scalars a and r such that

$$|x-\bar{x}|\le\delta|y| \quad \text{for all } x\in(h+F)^{-1}(y)\cap\mathbb{B}_a(\bar{x}) \text{ and } y\in\mathbb{B}_{\nu a+\gamma r}(0), \tag{20}$$

$$|f(p,x)-f(\bar{p},x)|\le\gamma|p-\bar{p}| \quad \text{for all } p\in\mathbb{B}_r(\bar{p}) \text{ and } x\in\mathbb{B}_a(\bar{x}), \tag{21}$$

and also, for $e=f-h$,

$$|e(p,x)-e(p,\bar{x})|\le\nu|x-\bar{x}| \quad \text{for all } x\in\mathbb{B}_a(\bar{x}) \text{ and } p\in\mathbb{B}_r(\bar{p}). \tag{22}$$

Let $x\in S(p)\cap\mathbb{B}_a(\bar{x})$ for some $p\in\mathbb{B}_r(\bar{p})$. Then, since $h(\bar{x})=f(\bar{p},\bar{x})$, we obtain from (21) and (22) that

$$\begin{aligned} |e(p,x)| &\le |e(p,x)-e(p,\bar{x})|+|f(p,\bar{x})-f(\bar{p},\bar{x})| \\ &\le \nu|x-\bar{x}|+\gamma|p-\bar{p}|\le\nu a+\gamma r. \end{aligned} \tag{23}$$

Observe that $x\in(h+F)^{-1}(-f(p,x)+h(x))\cap\mathbb{B}_a(\bar{x})$, and then from (20) and (23) we have

$$|x-\bar{x}|\le\delta|-f(p,x)+h(x)|\le\delta\nu|x-\bar{x}|+\delta\gamma|p-\bar{p}|.$$

In consequence,

$$|x-\bar{x}|\le\frac{\delta\gamma}{1-\delta\nu}|p-\bar{p}|.$$

Since δ is arbitrarily close to κ, ν is arbitrarily close to μ and γ is arbitrarily close to λ, we arrive at the desired result. □

In the theorem we state next, we can get away with a property of f at $(\bar{p},\bar{x})$ which is weaker than local continuous differentiability, namely a kind of uniform differentiability. We say that $f(p,x)$ is differentiable in x *uniformly with respect to* p at $(\bar{p},\bar{x})$ if f is differentiable with respect to (p,x) at $(\bar{p},\bar{x})$ and for every $\varepsilon > 0$ there is a (p,x)-neighborhood of $(\bar{p},\bar{x})$ in which

$$|f(p,x) - f(p,\bar{x}) - \nabla_x f(\bar{p},\bar{x})(x-\bar{x})| \leq \varepsilon|x-\bar{x}|.$$

Symmetrically, we define what it means for $f(p,x)$ to be differentiable in p *uniformly with respect to* x at $(\bar{p},\bar{x})$. Note that the combination of these two properties is implied by, yet weaker than, the continuous differentiability of f at $(\bar{p},\bar{x})$. For instance, the two uniformity properties hold when $f(p,x) = f_1(p) + f_2(x)$ and we simply have f_1 differentiable at $\bar{p}$ and f_2 differentiable at $\bar{x}$.

Theorem 3I.14 (utilizing differentiability and ample parameterization). *For the generalized equation in Theorem 3I.13 and its solution mapping S, and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$, suppose that f is differentiable in x uniformly with respect to p at $(\bar{p},\bar{x})$, and at the same time differentiable in p uniformly with respect to x at $(\bar{p},\bar{x})$. If the mapping*

$$h+F \text{ for } h(x) = f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x})$$

is strongly metrically subregular at $\bar{x}$ for 0, then S has the isolated calmness property at $\bar{p}$ for $\bar{x}$ with

$$\text{(24)} \qquad \operatorname{clm}(S;\bar{p}\,|\,\bar{x}) \leq \operatorname{subreg}(h+F;\bar{x}\,|\,0)\cdot|\nabla_p f(\bar{p},\bar{x})|.$$

Furthermore, when f is continuously differentiable on a neighborhood of $(\bar{p},\bar{x})$ and satisfies the ample parameterization condition

$$\operatorname{rank} \nabla_p f(\bar{p},\bar{x}) = m,$$

then the converse implication holds as well: the mapping $h+F$ is strongly metrically subregular at $\bar{x}$ for 0 provided that S has the isolated calmness property at $\bar{p}$ for $\bar{x}$.

Proof. With this choice of h, the assumption (19) of 3I.13 holds and then (24) follows from the conclusion of this theorem. To handle the ample parameterization we employ Lemma 2C.1 by repeating the argument in the proof of 3F.10, simply replacing the composition rule there with the one in the following proposition. □

Proposition 3I.15 (isolated calmness in composition). *For a mapping $M : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ and a function $\psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^d$ consider the composite mapping $N : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ of the form*

$$y \mapsto N(y) = \left\{ x \,\middle|\, x \in M(\psi(x,y)) \right\} \text{ for } y \in \mathbb{R}^m.$$

Let ψ satisfy

$$\text{(25)} \qquad \widehat{\operatorname{clm}}_x(\psi;(\bar{x},0)) = 0 \quad \text{and} \quad \widehat{\operatorname{clm}}_y(\psi;(\bar{x},0)) < \infty,$$

and let $(\psi(\bar{x},0),\bar{x}) \in \operatorname{gph} M$. *If* M *has the isolated calmness property at* $\psi(\bar{x},0)$ *for* $\bar{x}$, *then* N *has the isolated calmness property at* 0 *for* $\bar{x}$.

Proof. Let M have the isolated calmness property with neighborhoods $I\!B_b(\bar{x}), I\!B_c(\bar{p})$ and constant $\kappa > \operatorname{clm}(M;\bar{p}\,|\,\bar{x})$, where $\bar{p} = \psi(\bar{x},0)$. Choose $\lambda \in (0,1/\kappa)$ and $a > 0$ such that for any $y \in I\!B_a(0)$ the function $\psi(\cdot,y)$ is calm on $I\!B_b(\bar{x})$ with calmness constant λ. Pick $\gamma > \widehat{\operatorname{clm}}_y(\psi;(\bar{x},0))$ and make a and b smaller if necessary so that the function $\psi(x,\cdot)$ is calm on $I\!B_a(0)$ with constant γ and also

$$\lambda b + \gamma a \le c. \tag{26}$$

Let $y \in I\!B_a(0)$ and $x \in N(y) \cap I\!B_b(\bar{x})$. Then $x \in M(\psi(x,y)) \cap I\!B_b(\bar{x})$. Using the assumed calmness properties (25) of ψ and utilizing (26) we see that

$$|\psi(x,y) - \bar{p}| = |\psi(x,y) - \psi(\bar{x},0)| \le \lambda b + \gamma a \le c.$$

From the isolated calmness of M we then have

$$|x-\bar{x}| \le \kappa|\psi(x,y) - \psi(\bar{x},0)| \le \kappa\lambda|x-\bar{x}| + \kappa\gamma|y|,$$

hence

$$|x-\bar{x}| \le \frac{\kappa\gamma}{1-\kappa\lambda}|y|.$$

This establishes that the mapping N has the isolated calmness property at 0 for $\bar{x}$ with constant $\kappa\gamma/(1-\kappa\lambda)$. □

Example 3I.16 (complementarity problem)**.** For $f : I\!R^d \times I\!R^n \to I\!R^n$, consider the complementarity problem of finding for given p an x such that

$$x \ge 0, \qquad f(p,x) \ge 0, \qquad x \perp f(p,x). \tag{27}$$

This corresponds to solving $f(p,x) + N_{I\!R^n_+}(x) \ni 0$, as seen in 2A. Let $\bar{x}$ be a solution for $\bar{p}$, and suppose that f is continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$. Consider now the linearized problem

$$x \ge 0, \qquad Ax + y \ge 0, \qquad x \perp Ax + y, \quad \text{with } A = \nabla_x f(\bar{p},\bar{x}), \tag{28}$$

where y is a parameter in a neighborhood of $\bar{y} = f(\bar{p},\bar{x}) - A\bar{x}$. Then, from 3I.14 we obtain that if the solution mapping of (28) has the isolated calmness property at $\bar{y}$ for $\bar{x}$, then the solution mapping of (27) has the isolated calmness property at $\bar{p}$ for $\bar{x}$. Under the ample parameterization condition, rank $\nabla_p f(\bar{p},\bar{x}) = n$, the converse implication holds as well.

Commentary

The inner and outer limits of sequences of sets were introduced by Painlevé in his lecture notes as early as 1902, but the idea could be traced back to Peano according to Dolecki and Greco [2011]. These limits were later popularized by Hausdorff [1927] and Kuratowski [1933]. The definition of excess was first given by Pompeiu [1905], who also defined the distance between sets C and D as $e(C,D)+e(D,C)$. Hausdorff [1927] gave the definition we use here. These two definitions are equivalent in the sense that they induce the same convergence of sets. The reader can find much more about set-convergence and continuity properties of set-valued mappings together with extended historical commentary in Rockafellar and Wets [1998], for more advanced treatment see Beer [1993]. This includes the reason why we prefer "inner and outer" in contrast to the more common terms "lower and upper," so as to avoid certain conflicts in definition that unfortunately pervade the literature.

Theorem 3B.4 is a particular case of a result sometimes referred to as the Berge theorem; see Section 8.1 in Dontchev and Zolezzi [1993] for a general statement. Theorem 3C.3 comes from Walkup and Wets [1969], while the Hoffman lemma, 3C.4, is due to Hoffman [1952].

The concept of outer Lipschitz continuity was introduced by Robinson [1979, 1981] under the name "upper Lipschitz continuity" and adjusted to "outer Lipschitz continuity" later in Robinson [2007]. Theorem 3D.1 is due to Robinson [1981] while 3D.3 is a version, given in Robinson [2007], of a result due to Wu Li [1994].

The Aubin property of set-valued mappings was introduced by J.-P. Aubin [1984], who called it "pseudo-Lipschitz continuity"; it was renamed after Aubin in Dontchev and Rockafellar [1996]. In the literature one can also find it termed "Aubin continuity," but we do not use that here since the Aubin property does not imply continuity. Theorem 3E.3 is from Bessis, Ledyaev and Vinter [2001]. The name "metric regularity" was coined by J. M. Borwein [1986a], but the origins of this concept go back to the Banach open mapping theorem and even earlier. Theorem 3E.5 is from Rockafellar [1985]. Theorem 3E.10 comes from Ledyaev and Zhu [1999]. For historical remarks regarding inverse and implicit mapping theorems with metric regularity, see the commentary to Chapter 5 where more references are given.

As we mentioned earlier in Chapter 2, the term "strong regularity" comes from Robinson [1980], who used it in the framework of variational inequalities. Theorem 3F.5 is a particular case of a more general result due to Kenderov [1975]; see also Levy and Poliquin [1997]. Theorem 3F.14 is a simplified version of a result in Aragón Artacho and Mordukhovich [2010].

Calmness and metric subregularity, as well as isolated calmness and metric subregularity, have been considered in various contexts and under various names in the literature; here we follow the terminology of Dontchev and Rockafellar [2004]. Isolated calmness was formally introduced in Dontchev [1995a], where its stability (Theorem 3I.7) was first proved. The equivalent property of strong metric subregularity was considered earlier, without giving it a name, by Rockafellar [1989]; see also the commentary to Chapter 4.

Chapter 4
Metric Regularity Through Generalized Derivatives

In the wide-ranging generalizations we have been developing of the inverse function theorem and implicit function theorem, we have followed the idea that conclusions about a solution mapping, concerning the Aubin property, say, or the existence of a single-valued localization, can be drawn by confirming that some auxiliary solution mapping, obtained from a kind of approximation, has the property in question. In the classical framework, we can appeal to a condition like the invertibility of a Jacobian matrix and thus tie in with standard calculus. Now we are far away in another world where even a concept of differentiability seems to be lacking. However, substitutes for classical differentiability can very well be introduced and put to work. In this chapter we show the way to that and explain numerous consequences.

First, graphical differentiation of a set-valued mapping is defined through the variational geometry of the mapping's graph. A characterization of the Aubin property is derived and applied to the case of a solution mapping. Another criterion for metric regularity is derived utilizing coderivatives. Strong metric regularity is characterized next through a strict graphical derivative. Finally, the graphical derivative is used again to characterize strong metric subregularity. Applications are made to parameterized constraint systems and special features of solution mappings for variational inequalities, with emphasis on variational inequalities over polyhedral convex sets. Finally, it is shown that for the Karush-Kuhn-Tucker system metric regularity and strong metric regularity are equivalent properties.

4A. Graphical Differentiation

The concept of the tangent cone $T_C(x)$ to a set C in $\mathbb{R}^n$ at a point $x \in C$ was introduced in 2A, but it was only utilized there in the case of C being closed and convex. In 2E, the geometry of tangent cones to polyhedral convex sets received special attention and led to significant insights in the study of variational inequalities. Now, tangent cones to possibly nonconvex sets will come strongly onto the stage as well, serving as a tool for a kind of generalized differentiation. The definition of the tangent cone is the same as before.

Tangent cones. *A vector $v \in \mathbb{R}^n$ is said to be tangent to a set $C \subset \mathbb{R}^n$ at a point $x \in C$ if*

$$\frac{1}{\tau^k}(x^k - x) \to v \quad \text{for some } x^k \to x,\ x^k \in C,\ \tau^k \searrow 0.$$

The set of all such vectors v is called the tangent cone to C at x and is denoted $T_C(x)$. The tangent cone mapping is defined as

$$T_C : x \mapsto \begin{cases} T_C(x) & \text{for } x \in C, \\ \emptyset & \text{otherwise.} \end{cases}$$

A description equivalent to this definition is that $v \in T_C(x)$ if and only if there are sequences $v^k \to v$ and $\tau^k \searrow 0$ with $x + \tau^k v^k \in C$, or equivalently, if there are sequences $x^k \in C$, $x^k \to x$ and $\tau^k \searrow 0$ such that $v^k := (x^k - x)/\tau^k \to v$ as $k \to \infty$.

Note that $T_C(x)$ is indeed a *cone*: it contains $v = 0$ (as seen from taking $x^k \equiv x$), and contains along with any vector v all positive multiples of v. The definition can also be recast in the notation of set convergence:

$$(1) \qquad T_C(x) = \limsup_{\tau \searrow 0} \tau^{-1}(C - x).$$

Described as an outer limit in this way, it is clear in particular that $T_C(x)$ is always a closed set. When C is a "smooth manifold" in $\mathbb{R}^n$, $T_C(x)$ is the usual tangent subspace, but in general, of course, $T_C(x)$ need not even be convex. The tangent cone mapping T_C has $\operatorname{dom} T_C = C$ but $\operatorname{gph} T_C$ is not necessarily a closed subset of $\mathbb{R}^n \times \mathbb{R}^n$ even when C is closed.

As noted in 2A.4, when the set C is convex, the tangent cone $T_C(x)$ is also convex for every $x \in C$. In this case the limsup in (1) can be replaced by lim, as shown in the following proposition.

Proposition 4A.1 (tangent cones to convex sets)**.** *For a convex set $C \subset \mathbb{R}^n$ and a point $x \in C$,*

$$(2) \qquad T_C(x) = \lim_{\tau \searrow 0} \tau^{-1}(C - x).$$

Proof. Consider the set

$$K_C(x) = \{ v \mid \exists\, \tau > 0 \text{ with } x + \tau v \in C \}.$$

Let $v \in T_C(x)$. Then there exist sequences $\tau^k \searrow 0$ and $x^k \in C$, $x^k \to x$, such that $v^k := (x^k - x)/\tau^k \to v$. Hence $v^k \in K_C(x)$ for all k and therefore $v \in \mathrm{cl}\, K_C(x)$. Thus, we obtain

$$T_C(x) \subset \mathrm{cl}\, K_C(x). \tag{3}$$

Now let $v \in K_C(x)$. Then $v = (\tilde{x} - x)/\tau$ for some $\tau > 0$ and $\tilde{x} \in C$. Take an arbitrary sequence $\tau^k \searrow 0$ as $k \to \infty$. Since C is convex, we have

$$x + \tau^k v = (1 - \frac{\tau^k}{\tau})x + \frac{\tau^k}{\tau}\tilde{x} \in C \quad \text{for all } k.$$

But then $v \in (C - x)/\tau^k$ for all k and hence $v \in \liminf_k (C - x)/\tau^k$. Since τ^k was arbitrarily chosen, we conclude that

$$K_C(x) \subset \liminf_{\tau \searrow 0} \tau^{-1}(C - x) \subset \limsup_{\tau \searrow 0} \tau^{-1}(C - x) = T_C(x).$$

This combined with (3) gives us (2). □

We should note that, in order to have the equality (2), the set C does not need to be convex. Generally, sets C for which (2) is satisfied are called *geometrically derivable*. Proposition 4A.1 simply says that all convex sets are geometrically derivable.

Starting in elementary calculus, students are taught to view differentiation in terms of tangents to the graph of a function. This can be formulated in the notation of tangent cones as follows. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a function which is differentiable at x with derivative mapping $Df(x) : \mathbb{R}^n \to \mathbb{R}^m$. Then

$$(u, v) \in \mathrm{gph}\, Df(x) \iff (u, v) \in T_{\mathrm{gph}\, f}(x, f(x)).$$

In other words, the derivative is completely represented geometrically by the tangent cone to the set $\mathrm{gph}\, f$ at the point $(x, f(x))$. In fact, differentiability is more or less equivalent to having $T_{\mathrm{gph}\, f}(x, f(x))$ turn out to be the graph of a linear mapping.

By adopting such a geometric characterization as a definition, while not insisting on linearity, we can introduce derivatives for an arbitrary set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. However, because $F(x)$ may have more than one element y, it is essential for the derivative mapping to depend not just on x but also on a choice of $y \in F(x)$.

Graphical derivative. *For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a pair (x, y) with $y \in F(x)$, the graphical derivative of F at x for y is the mapping $DF(x \mid y) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ whose graph is the tangent cone $T_{\mathrm{gph}\, F}(x, y)$ to $\mathrm{gph}\, F$ at (x, y):*

$$v \in DF(x \mid y)(u) \iff (u, v) \in T_{\mathrm{gph}\, F}(x, y).$$

Thus, $v \in DF(x \mid y)(u)$ if and only if there exist sequences $u^k \to u$, $v^k \to v$ and $\tau^k \searrow 0$ such that $y + \tau^k v^k \in F(x + \tau^k u^k)$ for all k.

On this level, derivative mappings may no longer even be single-valued. But because their graphs are cones, they do always belong to the following class of mappings, at least.

Positively homogeneous mappings. *A mapping $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called positively homogeneous when* gph H *is a cone, which is equivalent to H satisfying*

$$0 \in H(0) \quad \text{and} \quad H(\lambda x) = \lambda H(x) \text{ for } \lambda > 0.$$

Clearly, the inverse of a positively homogeneous mapping is another positively homogeneous mapping. Linear mappings are positively homogeneous as a special case, their graphs being not just cones but linear subspaces.

Since the graphical differentiation comes from an operation on graphs, and the graph of a mapping F can be converted to the graph of its inverse F^{-1} just by interchanging variables, we immediately have the rule that

$$D(F^{-1})(y|x) = DF(x|y)^{-1}.$$

Another useful relation is available for sums.

Proposition 4A.2 (sum rule)**.** *For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ which is differentiable at x, a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and any $y \in F(x)$, one has*

$$D(f+F)(x|f(x)+y) = Df(x) + DF(x|y).$$

Proof. If $v \in D(f+F)(x|f(x)+y)(u)$ there exist sequences $\tau^k \searrow 0$ and $u^k \to u$ and $v^k \to v$ such that

$$f(x) + y - f(x+\tau^k u^k) + \tau^k v^k \in F(x+\tau^k u^k) \text{ for every } k.$$

By using the definition of the derivative for f we get

$$y + \tau^k(-Df(x)u + v^k) + o(\tau^k) \in F(x+\tau^k u^k).$$

Hence, by the definition of the graphical derivative, $v \in Df(x)u + DF(x|y)(u)$.

Conversely, if $v - Df(x)u \in DF(x|y)(u)$ then there exist sequences $\tau^k \searrow 0$, and $u^k \to u$ and $w^k \to v - Df(x)u$ such that $y + \tau^k w^k \in F(x+\tau^k u^k)$. Again, by the differentiability of f,

$$y + f(x) + \tau^k v^k + o(\tau^k) \in (f+F)(x+\tau^k u^k) \text{ for } v^k = w^k + Df(x)u^k,$$

which yields $v \in D(f+F)(x|f(x)+y)(u)$. □

Example 4A.3 (graphical derivative for a constraint system)**.** Consider a general constraint system of the form

$$f(x) - D \ni y, \tag{4}$$

for a function $f:\mathbb{R}^n\to\mathbb{R}^m$, a set $D\subset\mathbb{R}^m$ and a parameter vector y, and let $\bar{x}$ be a solution of (4) for $\bar{y}$ at which f is differentiable. Then for the mapping

$$G: x\mapsto f(x)-D,\ \text{ with }\ \bar{y}\in G(\bar{x}),$$

one has

$$DG(\bar{x}\,|\,\bar{y})(u)\ = Df(\bar{x})u - T_D(f(\bar{x})-\bar{y}). \tag{5}$$

Detail. This applies the sum rule to the case of a constant mapping $F\equiv -D$, for which the definition of the graphical derivative gives $DF(x\,|\,z)=T_{-D}(z)=-T_D(-z)$. □

In the special but important case of Example 4A.3 in which $D=\mathbb{R}^s_-\times\{0\}^{m-s}$ with $f=(f_1,\dots,f_m)$, the constraint system (4) with respect to $y=(y_1,\dots,y_m)$ takes the form

$$f_i(x)\begin{cases}\le y_i & \text{for } i=1,\dots,s,\\ = y_i & \text{for } i=s+1,\dots,m.\end{cases}$$

The graphical derivative formula (5) says then that a vector $v=(v_1,\dots,v_m)$ is in $DG(x\,|\,y)(u)$ if and only if

$$\nabla f_i(x)u\begin{cases}\le v_i & \text{for } i\in[1,s] \text{ with } f_i(x)=y_i,\\ = v_i & \text{for } i=s+1,\dots,m.\end{cases}$$

Example 4A.4 (graphical derivative for a variational inequality)**.** For a function $f:\mathbb{R}^n\to\mathbb{R}^n$ and a convex set $C\subset\mathbb{R}^n$ that is polyhedral, consider the variational inequality

$$f(x)+N_C(x)\ni y \tag{6}$$

in which y is a parameter. Let x be a solution of (6) at which f is differentiable. Let $v=y-f(x)\in N_C(x)$ and let $K_C(x,v)$ be the corresponding critical cone, this being the polyhedral convex cone $T_C(x)\cap[v]^\perp$. Then for the mapping

$$G: x\mapsto f(x)+N_C(x),\ \text{ with }\ y\in G(x),$$

one has

$$DG(x\,|\,y)(u)=Df(x)u+N_{K_C(x,v)}(u). \tag{7}$$

Detail. From the sum rule in 4A.2 we have $DG(x\,|\,y)(u)=Df(x)u+DN_C(x\,|\,v)(u)$. According to Lemma 2E.4 (the reduction lemma for normal cone mappings to polyhedral convex sets), for every $(x,v)\in\operatorname{gph} N_C$ there exists a neighborhood O of the origin in $\mathbb{R}^n\times\mathbb{R}^n$ such that for $(x',v')\in O$ one has

$$v+v'\in N_C(x+x')\iff v'\in N_{K_C(x,v)}(x').$$

This reveals in particular that the tangent cone to $\operatorname{gph} N_C$ at (x,v) is just $\operatorname{gph} N_{K_C(x,v)}$, or in other words, that $DN_C(x\,|\,v)$ is the normal cone mapping $N_{K_C(x,v)}$. Thus we have (7). $\square$

Because graphical derivative mappings are positively homogeneous, general properties of positively homogeneous mappings can be applied to them. Norm concepts are available in particular for capturing quantitative characteristics.

Outer and inner norms. *For any positively homogeneous mapping $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, the outer norm and the inner norm are defined, respectively, by*

$$(8) \qquad |H|^+ = \sup_{|x|\le 1} \sup_{y\in H(x)} |y| \quad \textit{and} \quad |H|^- = \sup_{|x|\le 1} \inf_{y\in H(x)} |y|$$

with the convention $\inf_{y\in\emptyset} |y| = \infty$ *and* $\sup_{y\in\emptyset} |y| = -\infty$.

When H is a linear mapping, both $|H|^+$ and $|H|^-$ reduce to the operator (matrix) norm $|H|$ associated with the Euclidean norm. However, it must be noted that neither $|H|^+$ nor $|H|^-$ satisfies the conditions in the definition of a true "norm," inasmuch as set-valued mappings do not even form a vector space.

The inner and outer norms have simple interpretations when $H = A^{-1}$ for a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$. Let the $m\times n$ matrix for this linear mapping be denoted likewise by A, for simplicity. If $m < n$, we have A surjective (the associated matrix being of rank m) if and only if $|A^{-1}|^-$ is finite, this expression being the norm of the right inverse of A: $|A^{-1}|^- = |A^{\mathsf T}(AA^{\mathsf T})^{-1}|$. Then $|A^{-1}|^+ = \infty$. On the other hand, if $m > n$, we have $|A^{-1}|^+ < \infty$ if and only if A is injective (the associated matrix has rank n), and then $|A^{-1}|^+ = |(A^{\mathsf T}A)^{-1}A^{\mathsf T}|$ but $|A^{-1}|^- = \infty$. For $m = n$, of course, both norms agree with the usual matrix norm $|A^{-1}|$, and the finiteness of this quantity is equivalent to nonsingularity of A.

Proposition 4A.5 (domains of positively homogeneous mappings)**.** *For a positively homogeneous mapping $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$,*

$$(9) \qquad \operatorname{dom} H = \mathbb{R}^n \quad\Longrightarrow\quad |H|^+ \ge |H|^-.$$

Moreover,

$$(10) \qquad |H|^- < \infty \quad\Longrightarrow\quad \operatorname{dom} H = \mathbb{R}^n;$$

thus, if $|H^{-1}|^- < \infty$ then H must be surjective.

Proof. The implications (9) and (10) are immediate from the definition (8) and its conventions concerning the empty set. $\square$

In cases where $\operatorname{dom} H$ is not all of $\mathbb{R}^n$, it is possible for the inequality in (9) to fail. As an illustration, this occurs for the positively homogeneous mapping $H : \mathbb{R} \rightrightarrows \mathbb{R}$ defined by

$$H(x) = \begin{cases} 0 & \text{for } x \ge 0, \\ \emptyset & \text{for } x < 0, \end{cases}$$

for which $|H|^+ = 0$ while $|H|^- = \infty$.

Proposition 4A.6 (norm characterizations). *The inner norm of a positively homogeneous mapping $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ satisfies*

$$(11) \qquad |H|^- = \inf\left\{ \kappa > 0 \,\middle|\, H(x) \cap \kappa\mathbb{B} \neq \emptyset \text{ for all } x \in \mathbb{B} \right\}.$$

In parallel, the outer norm satisfies

$$(12) \qquad |H|^+ \;=\; \inf\left\{ \kappa \in (0,\infty) \,\middle|\, y \in H(x) \Rightarrow |y| \leq \kappa|x| \right\} \;=\; \sup_{|y|=1} \frac{1}{d(0,H^{-1}(y))}.$$

If H has closed graph, then furthermore

$$(13) \qquad |H|^+ < \infty \iff H(0) = \{0\}.$$

If H has closed and convex graph, then the implication (10) becomes equivalence:

$$(14) \qquad |H|^- < \infty \iff \operatorname{dom} H = \mathbb{R}^n$$

and in that case $|H^{-1}|^- < \infty$ if and only if H is surjective.

Proof. We get (11) and the first part of (12) simply by rewriting the formulas in terms of $\mathbb{B} = \left\{ x \,\middle|\, |x| \leq 1 \right\}$ and utilizing the positive homogeneity. The infimum so obtained in (12) is unchanged when y is restricted to have $|y| = 1$, and in this way it can be identified with the infimum of all $\kappa \in (0,\infty)$ such that $\kappa \geq 1/|x|$ whenever $x \in H^{-1}(y)$ and $|y| = 1$. (It is correct in this to interpret $1/|x| = \infty$ when $x = 0$.) This shows that the middle expression in (12) agrees with the final one.

Moving on to (13), we observe that when $|H|^+ < \infty$ the middle expression in (12) implies that if $(0,y) \in \operatorname{gph} H$ then y must be 0. To prove the converse implication we will need the assumption that $\operatorname{gph} H$ is closed. Suppose that $H(0) = \{0\}$. If $|H|^+ = \infty$, there has to be a sequence of points $(x_k, y_k) \in \operatorname{gph} H$ such that $0 < |y_k| \to \infty$ but x_k is bounded. Consider then the sequence of pairs (w_k, u_k) in which $w_k = x_k/|y_k|$ and $u_k = y_k/|y_k|$. We have $w_k \to 0$, while u_k has a cluster point $\bar{u}$ with $|\bar{u}| = 1$. Moreover, $(w_k, u_k) \in \operatorname{gph} H$ by the positive homogeneity and hence, through the closedness of $\operatorname{gph} H$, we must have $H(0) \ni \bar{u}$. This contradicts our assumption that $H(0) = \{0\}$ and terminates the proof of (13). The equivalence (14) follows from (9) and a general result (Robinson–Ursescu theorem) which we will prove in Section 5B. □

Corollary 4A.7 (norms of linear-constraint-type mappings). *Suppose*

$$H(x) = Ax - K$$

for a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ and a closed convex cone $K \subset \mathbb{R}^m$. Then H is positively homogeneous with closed and convex graph. Moreover

$$(15) \qquad |H^{-1}|^- = \sup_{|y| \leq 1} d(0, A^{-1}(y+K))$$

and

$$|H^{-1}|^{-} < \infty \iff \operatorname{rge} A - K = \mathbb{R}^m. \tag{16}$$

On the other hand,

$$|H^{-1}|^{+} = \sup_{|x|=1} \frac{1}{d(Ax,K)} \tag{17}$$

and

$$|H^{-1}|^{+} < \infty \iff \Big[Ax - K \ni 0 \implies x = 0 \Big]. \tag{18}$$

Proof. Formula (15) follows from the definition (8) while (16) comes from (14) applied to this case. Formula (17) follows from (12) while (18) is the specification of (13). □

We will come back to the general theory of positively homogeneous mappings and their norms in Section 5A. In the meantime there will be applications to the case of derivative mappings.

Some properties of the graphical derivatives of convex-valued mappings under Lipschitz continuity are displayed in the following exercise.

Exercise 4A.8. *Consider a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ that is convex-valued and Lipschitz continuous in its domain and let $(x,y) \in \operatorname{gph} F$. Prove that in this case $DF(x|y)$ is convex-valued and*

$$DF(x|y)(u) = \lim_{\tau \searrow 0} \tau^{-1}(F(x+\tau u) - y), \tag{19}$$

and in particular,

$$DF(x|y)(0) = T_{F(x)}(y). \tag{20}$$

Guide. Observe that, by definition

$$DF(x|y)(u) = \limsup_{\tau \searrow 0,\, u' \to u} \tau^{-1}(F(x+\tau u') - y).$$

Since F is Lipschitz continuous, this equality reduces to

$$DF(x|y)(u) = \limsup_{\tau \searrow 0} \tau^{-1}(F(x+\tau u) - y). \tag{21}$$

Then use the convexity of the values of F as in the proof of Proposition 4A.1 to show that limsup in (21) can be replaced by lim and use this to obtain convexity of $DF(x|y)(u)$ from the convexity of $F(x+\tau u)$. Lastly, to show (20) apply 4A.1 to (19) in the case $u = 0$. □

Exercise 4A.9. *For a positively homogeneous mapping* $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, *show that*

$$|H|^- = 0 \iff \operatorname{cl} H(x) \ni 0 \textit{ for all } x \in \mathbb{R}^n,$$

$$|H|^+ = 0 \iff \operatorname{rge} H = \{0\}.$$

Guide. Apply the norm characterizations in (11) and (12). □

4B. Graphical Derivative Criterion for Metric Regularity

Conditions will next be developed which characterize metric regularity and the Aubin property in terms of graphical derivatives. From these conditions, new forms of implicit mapping theorems will be obtained. First, we state a fundamental fact.

Theorem 4B.1 (graphical derivative criterion for metric regularity)**.** *For a mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *and a point* $(\bar{x},\bar{y}) \in \operatorname{gph} F$ *at which the* $\operatorname{gph} F$ *is locally closed, one has*

$$\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \limsup_{\substack{(x,y)\to(\bar{x},\bar{y})\\(x,y)\in\operatorname{gph} F}} |DF(x\,|\,y)^{-1}|^-. \tag{1}$$

Thus, F *is metrically regular at* $\bar{x}$ *for* $\bar{y}$ *if and only if the right side of (1) is finite.*

The proof of Theorem 4B.1 will be furnished later in this section. Note that in the case when $m \le n$ and F is a function f which is differentiable on a neighborhood of $\bar{x}$, the representation of the regularity modulus in (1) says that f is metrically regular precisely when the Jacobians $\nabla f(x)$ for x near $\bar{x}$ are of full rank and the inner norms of their inverses $\nabla f(x)^{-1}$ are uniformly bounded. This holds automatically when f is continuously differentiable around $\bar{x}$ with $\nabla f(\bar{x})$ of full rank, in which case we get not only metric regularity but also existence of a continuously differentiable local selection of f^{-1}, as in 1F.6. When $m = n$ this becomes nonsingularity and we come to the classical inverse function theorem.

Also to be kept in mind here is the connection between metric regularity and the Aubin property in 3E.7. This allows Theorem 4B.1 to be formulated equivalently as a statement about the Aubin property.

Theorem 4B.2 (graphical derivative criterion for the Aubin property)**.** *For a mapping* $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *and a point* $(\bar{y},\bar{x}) \in \operatorname{gph} S$ *at which* $\operatorname{gph} S$ *is locally closed, one has*

$$\operatorname{lip}(S;\bar{y}\,|\,\bar{x}) = \limsup_{\substack{(y,x)\to(\bar{y},\bar{x})\\(y,x)\in\operatorname{gph} S}} |DS(y\,|\,x)|^-. \tag{2}$$

Thus, S has the Aubin property at $\bar{y}$ for $\bar{x}$ if and only if the right side of (2) is finite.

Solution mappings of much greater generality can also be handled with these ideas. For this, we return to the framework introduced briefly at the end of Section 3E and delve into it much further. We consider the parameterized relation

$$(3) \qquad G(p,x) \ni 0 \quad \text{for a mapping} \quad G : \mathbb{R}^d \times \mathbb{R}^n \rightrightarrows \mathbb{R}^m$$

and its solution mapping $S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ defined by

$$(4) \qquad S(p) = \left\{ x \,\middle|\, G(p,x) \ni 0 \right\}.$$

In Theorem 3E.10, a result was presented in which a partial Aubin property of G with respect to p, combined with other assumptions, led to a conclusion that S has the Aubin property. We are looking now toward finding derivative criteria for these Aubin properties, so as to obtain a different type of statement about the "implicit mapping" S.

The following theorem will be our stepping stone to progress and will have many other interesting consequences as well. It makes use of the *partial* graphical derivative of $G(p,x)$ with respect to x, which is defined as the graphical derivative of the mapping $x \mapsto G(p,x)$ with p fixed and denoted by $D_x G(p,x \,|\, y)$. Of course, $D_p G(p,x \,|\, y)$ has a similar meaning.

Theorem 4B.3 (solution mapping estimate)**.** *For the inclusion (3) and its solution mapping S in (4), let $\bar{x} \in S(\bar{p})$, so that $(\bar{p},\bar{x},0) \in \operatorname{gph} G$. Suppose that $\operatorname{gph} G$ is locally closed at $(\bar{p},\bar{x},0)$ and that the distance mapping $p \mapsto d(0,G(p,\bar{x}))$ is upper semicontinuous at $\bar{p}$. Then for every $c \in (0,\infty)$ satisfying*

$$(5) \qquad \limsup_{\substack{(p,x,y)\to(\bar{p},\bar{x},0)\\ (p,x,y)\in \operatorname{gph} G}} |D_x G(p,x \,|\, y)^{-1}|^- < c$$

there are neighborhoods V of $\bar{p}$ and U of $\bar{x}$ such that

$$(6) \qquad d(x,S(p)) \le c\, d(0,G(p,x)) \quad \text{for } x \in U \text{ and } p \in V.$$

Proof. Let c satisfy (5). Then there exists $\eta > 0$ such that

$$(7) \qquad \begin{cases} \text{for every } (p,x,y) \in \operatorname{gph} G \text{ with } |p-\bar{p}| + \max\{|x-\bar{x}|, c|y|\} \le 2\eta, \\ \text{and for every } v \in \mathbb{R}^m, \text{ there exists } u \in D_x G(p,x \,|\, y)^{-1}(v) \text{ with } |u| \le c|v|. \end{cases}$$

We can always choose η smaller so that the intersection

$$(8) \qquad \operatorname{gph} G \bigcap \left\{ (p,x,y) \,\middle|\, |p-\bar{p}| + \max\{|x-\bar{x}|, c|y|\} \le 2\eta \right\} \quad \text{is closed.}$$

The next part of the proof is developed as a lemma.

Lemma 4B.4 (intermediate estimate)**.** *For c and η as above, let $\varepsilon > 0$ and $s > 0$ be such that*

(9) $$c\varepsilon < 1 \quad \text{and} \quad s < \varepsilon\eta$$

and let $(p,\omega,\nu) \in \operatorname{gph} G$ *satisfy*

(10) $$|p-\bar{p}| + \max\{|\omega - \bar{x}|, c|\nu|\} \le \eta.$$

Then for every $y' \in \mathbb{B}_s(\nu)$ *there exists* $\hat{x}$ *with* $y' \in G(p,\hat{x})$ *such that*

(11) $$|\hat{x} - \omega| \le \frac{1}{\varepsilon}|y' - \nu|.$$

In the proof of the lemma we apply a fundamental result in variational analysis, which is stated next:

Theorem 4B.5 (Ekeland variational principle)**.** *Let* (X,ρ) *be a complete metric space and let* $f : X \to (-\infty,\infty]$ *be a lower semicontinuous function on* X *which is bounded from below. Let* $\bar{u} \in \operatorname{dom} f$*. Then for every* $\delta > 0$ *there exists* u_δ *such that*

$$f(u_\delta) + \delta\rho(u_\delta,\bar{u}) \le f(\bar{u}),$$

and

$$f(u_\delta) < f(u) + \delta\rho(u,u_\delta) \quad \textit{for every } u \in X,\ u \ne u_\delta.$$

Proof of Lemma 4B.4. On the product space $Z := \mathbb{R}^n \times \mathbb{R}^m$ we introduce the norm

$$\|(x,y)\| := \max\{|x|, c|y|\},$$

which is equivalent to the Euclidean norm. Pick ε, s and $(p,\omega,\nu) \in \operatorname{gph} G$ as required in (9) and (10) and let $y' \in \mathbb{B}_s(\nu)$. By (8) the set

$$E_p := \left\{ (x,y) \,\middle|\, (p,x,y) \in \operatorname{gph} G, |p-\bar{p}| + \|(x,y) - (\bar{x},0)\| \le 2\eta \right\} \subset \mathbb{R}^n \times \mathbb{R}^m$$

is closed, hence, equipped with the metric induced by the norm in question, it is a complete metric space. The function $V_p : E_p \to \mathbb{R}$ defined by

(12) $$V_p : (x,y) \mapsto |y' - y| \qquad \text{for } (x,y) \in E_p$$

is continuous on its domain E_p. Also, $(\omega,\nu) \in \operatorname{dom} V_p$. We apply Ekeland's variational principle 4B.5 to V_p with $\bar{u} = (\omega,\nu)$ and the indicated ε to obtain the existence of $(\hat{x},\hat{y}) \in E_p$ such that

(13) $$V_p(\hat{x},\hat{y}) + \varepsilon\|(\omega,\nu) - (\hat{x},\hat{y})\| \le V_p(\omega,\nu)$$

and

(14) $$V_p(\hat{x},\hat{y}) \le V_p(x,y) + \varepsilon\|(x,y) - (\hat{x},\hat{y})\| \qquad \text{for every } (x,y) \in E_p.$$

With V_p as in (12), the inequalities (13) and (14) come down to

$$(15) \qquad |y' - \hat{y}| + \varepsilon\|(\omega, v) - (\hat{x}, \hat{y})\| \le |y' - v|$$

and

$$(16) \qquad |y' - \hat{y}| \le |y' - y| + \varepsilon\|(x,y) - (\hat{x}, \hat{y})\| \quad \text{for every } (x,y) \in E_p.$$

Through (15) we obtain in particular that

$$(17) \qquad \|(\omega, v) - (\hat{x}, \hat{y})\| \le \frac{1}{\varepsilon}|y' - v|.$$

Since $y' \in \mathbb{B}_s(v)$, we then have

$$\|(\omega, v) - (\hat{x}, \hat{y})\| \le \frac{s}{\varepsilon}$$

and consequently, from the choice of (p, ω, v) in (10) and s in (9),

$$(18) \qquad \begin{aligned} &|p - \bar{p}| + \|(\hat{x}, \hat{y}) - (\bar{x}, 0)\| \\ &\quad \le |p - \bar{p}| + \|(\omega, v) - (\bar{x}, 0)\| + \|(\omega, v) - (\hat{x}, \hat{y})\| \le \eta + \frac{s}{\varepsilon} < 2\eta. \end{aligned}$$

Thus, $(p, \hat{x}, \hat{y})$ satisfies the condition in (7), so there exists $u \in \mathbb{R}^n$ for which

$$(19) \qquad y' - \hat{y} \in D_x G(p, \hat{x} \,|\, \hat{y})(u) \ \text{ and } \ |u| \le c|y' - \hat{y}|.$$

By the definition of the partial graphical derivative, there exist sequences $\tau^k \searrow 0$, $u^k \to u$, and $v^k \to y' - \hat{y}$ such that

$$\hat{y} + \tau^k v^k \in G(p, \hat{x} + \tau^k u^k) \ \text{ for all } k.$$

Also, from (18) we know that, for sufficiently large k,

$$|p - \bar{p}| + \|(\hat{x} + \tau^k u^k, \hat{y} + \tau^k v^k) - (\bar{x}, 0)\| \le 2\eta,$$

implying $(\hat{x} + \tau^k u^k, \hat{y} + \tau^k v^k) \in E_p$. If we now plug the point $(\hat{x} + \tau^k u^k, \hat{y} + \tau^k v^k)$ into (16) in place of (x,y), we get

$$|y' - \hat{y}| \le |y' - (\hat{y} + \tau^k v^k)| + \varepsilon\|(\hat{x} + \tau^k u^k, \hat{y} + \tau^k v^k) - (\hat{x}, \hat{y})\|.$$

This gives us

$$|y' - \hat{y}| \le (1 - \tau^k)|y' - \hat{y}| + \tau^k|v^k - (y' - \hat{y})| + \varepsilon\tau^k\|(u^k, v^k)\|,$$

that is,

$$|y' - \hat{y}| \le |v^k - (y' - \hat{y})| + \varepsilon\|(u^k, v^k)\|.$$

Passing to the limit with $k \to \infty$ leads to $|y' - \hat{y}| \le \varepsilon\|(u, y' - \hat{y})\|$ and then, taking into account the second relation in (19), we conclude that $|y' - \hat{y}| \le \varepsilon c|y' - \hat{y}|$. Since

$\varepsilon c < 1$ by (9), the only possibility here is that $y' = \hat{y}$. But then $y' \in G(p,\hat{x})$ and (17) yields (11). This proves the lemma. $\square$

We continue now with the proof of Theorem 4B.3. Let $\tau = \eta/(4c)$. Since the function $p \to d(0, G(p,\bar{x}))$ is upper semicontinuous at $\bar{p}$, there exists a positive $\delta \le c\tau$ such that $d(0, G(p,\bar{x})) \le \tau/2$ for all p with $|p-\bar{p}| < \delta$. Set $V := \mathbb{B}_\delta(\bar{p})$, $U := \mathbb{B}_{c\tau}(\bar{x})$ and pick any $p \in V$ and $x \in U$. We can find y such that $y \in G(p,\bar{x})$ with $|y| \le d(0, G(p,\bar{x})) + \tau/3 < \tau$. Note that

$$|p-\bar{p}| + \|(\bar{x},y) - (\bar{x},0)\| = |p-\bar{p}| + c|y| \le \delta + c\tau \le \eta. \tag{20}$$

Choose $\varepsilon > 0$ such that $1/2 < \varepsilon c < 1$ and let $s = \varepsilon\eta/2$. Then $s > \tau$. We apply Lemma 4B.4 with the indicated ε and s, and with $(p,\omega,\nu) = (p,\bar{x},y)$ which, as seen in (20), satisfies (10), and with $y' = 0$, since $0 \in \mathbb{B}_s(y)$. Thus, there exists $\hat{x}$ such that $0 \in G(p,\hat{x})$, that is, $\hat{x} \in S(p)$, and also, from (11), $|\hat{x} - \bar{x}| \le |y|/\varepsilon$. Therefore, in view of the choice of y, we have $\hat{x} \in \mathbb{B}_{\tau/\varepsilon}(\bar{x})$. We now consider two cases.

CASE 1. $d(0, G(p,x)) \ge 2\tau$. We just proved that there exists $\hat{x} \in S(p)$ with $\hat{x} \in \mathbb{B}_{\tau/\varepsilon}(\bar{x})$; then

$$\begin{aligned} d(x,S(p)) &\le d(\bar{x},S(p)) + |x-\bar{x}| \\ &\le |\bar{x}-\hat{x}| + |x-\bar{x}| \le \frac{\tau}{\varepsilon} + c\tau \le \frac{2\tau}{\varepsilon} \le \frac{1}{\varepsilon} d(0,G(p,x)). \end{aligned} \tag{21}$$

CASE 2. $d(0, G(p,x)) < 2\tau$. In this case, for every y with $|y| \le 2\tau$ we have

$$|p-\bar{p}| + \max\{|x-\bar{x}|, c|y|\} \le \delta + \max\{c\tau, 2c\tau\} \le 3c\tau \le \eta$$

and then, by (8), the nonempty set $G(p,x) \cap 2\tau\mathbb{B}$ is closed. Hence, there exists $\tilde{y} \in G(p,x)$ such that $|\tilde{y}| = d(0, G(p,x)) < 2\tau$ and therefore

$$c|\tilde{y}| < 2c\tau = \frac{\eta}{2}\ .$$

We conclude that the point $(p,x,\tilde{y}) \in \operatorname{gph} G$ satisfies

$$|p-\bar{p}| + \max\{|x-\bar{x}|, c|\tilde{y}|\} \le \delta + \max\{c\tau, \eta/2\} \le \eta.$$

Thus, the assumptions of Lemma 4B.4 hold for $(p,\omega,\nu) = (p,x,\tilde{y})$, $s = 2\tau$, and $y' = 0$. Hence there exists $\tilde{x} \in S(p)$ such that

$$|\tilde{x} - x| \le \frac{1}{\varepsilon}|\tilde{y}|.$$

Then, by the choice of $\tilde{y}$,

$$d(x,S(p)) \le |x-\tilde{x}| \le \frac{1}{\varepsilon}|\tilde{y}| = \frac{1}{\varepsilon} d(0,G(p,x)).$$

Hence, by (21), for both cases 1 and 2, and therefore for any p in V and $x \in U$, we have

$$d(x,S(p)) \leq \frac{1}{\varepsilon} d(0,G(p,x)).$$

Since U and V do not depend on ε, and $1/\varepsilon$ can be arbitrarily close to c, this gives us (6). □

With this result in hand, we can confirm the criterion for metric regularity presented at the beginning of this section.

Proof of Theorem 4B.1. For short, let d_{DF} denote the right side of (1). We will start by showing that $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \leq d_{DF}$. If $d_{DF} = \infty$ there is nothing to prove. Let $d_{DF} < c < \infty$. Applying Theorem 4B.3 to $G(p,x) = F(x) - p$ and this c, letting y take the place of p, we have $S(y) = F^{-1}(y)$ and $d(0,G(y,x)) = d(y,F(x))$. Condition (6) becomes the definition of metric regularity of F at $\bar{x}$ for $\bar{y} = \bar{p}$, and therefore $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \leq c$. Since c can be arbitrarily close to d_{DF} we conclude that $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \leq d_{DF}$.

We turn now to demonstrating the opposite inequality,

$$\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \geq d_{DF}. \tag{22}$$

If $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \infty$ we are done. Suppose therefore that F is metrically regular at $\bar{x}$ for $\bar{y}$ with respect to a constant κ and neighborhoods U for $\bar{x}$ and V for $\bar{y}$. Then

$$d(x',F^{-1}(y)) \leq \kappa|y - y'| \ \text{ whenever } (x',y') \in \operatorname{gph} F,\, x' \in U,\, y \in V. \tag{23}$$

We know from 3E.1 that V can be chosen so small that $F^{-1}(y) \cap U \neq \emptyset$ for every $y \in V$. Pick any $y' \in V$ and $x' \in F^{-1}(y') \cap U$, and let $v \in \mathbb{B}$. Take a sequence $\tau^k \searrow 0$ such that $y^k := y' + \tau^k v \in V$ for all k. By (23) and the local closedness of $\operatorname{gph} F$ at $(\bar{x},\bar{y})$ there exists $x^k \in F^{-1}(y' + \tau^k v)$ such that

$$|x' - x^k| = d(x',F^{-1}(y^k)) \leq \kappa|y^k - y'| = \kappa\tau^k|v|.$$

For $u^k := (x^k - x')/\tau^k$ we obtain

$$|u^k| \leq \kappa|v|. \tag{24}$$

Thus, u^k is bounded, so $u^{k_i} \to u$ for a subsequence $k_i \to \infty$. Since $(x^{k_i}, y' + \tau^{k_i} v) \in \operatorname{gph} F$, we obtain $(u,v) \in T_{\operatorname{gph} F}(x',y')$. Hence, by the definition of the graphical derivative, we have $u \in DF^{-1}(y'\,|\,x')(v) = DF(x'\,|\,y')^{-1}(v)$. The bound (24) guarantees that

$$|DF(x\,|\,y)^{-1}|^{-} \leq \kappa.$$

Since $(y,x) \in \operatorname{gph} S$ is arbitrarily chosen near $(\bar{x},\bar{y})$, and κ is independent of this choice, we conclude that (22) holds and hence we have (1). □

We apply Theorem 4B.3 now to obtain for the implicit mapping result in Theorem 3E.10 an elaboration in which graphical derivatives provide estimates. Recall here the definition of $\widehat{\mathrm{lip}}_p(G;\bar{p},\bar{x}\,|\,\bar{y})$, the modulus of the partial Aubin property introduced just before 3E.10.

Theorem 4B.6 (implicit mapping theorem with graphical derivatives)**.** *For the general inclusion (3) and its solution mapping S in (4), let $\bar{x}\in S(\bar{p})$, so that $(\bar{p},\bar{x},0)\in \operatorname{gph} G$. Suppose that the distance $d(0,G(p,\bar{x}))$ depends upper semicontinuously on p at $\bar{p}$. Assume further that G has the partial Aubin property with respect to p uniformly in x at $(\bar{p},\bar{x})$, and that*

$$(25)\qquad \limsup_{\substack{(p,x,y)\to(\bar{p},\bar{x},0)\\ (p,x,y)\in \operatorname{gph} G}} |D_xG(p,x\,|\,y)^{-1}|^- \le \lambda < \infty.$$

Then S has the Aubin property at $\bar{p}$ for $\bar{x}$ with

$$(26)\qquad \mathrm{lip}\,(S;\bar{p}\,|\,\bar{x}) \le \lambda\, \widehat{\mathrm{lip}}_p(G;\bar{p},\bar{x}\,|\,0).$$

Proof. This just combines Theorem 3E.10 with the estimate now available from Theorem 4B.3. □

Note from Proposition 4A.5 that finiteness in condition (25) necessitates, in particular, having the range of $D_xG(p,x\,|\,y)$ be all of $\mathbb{R}^m$ when (p,x,y) is sufficiently close to $(\bar{p},\bar{x},0)$ in $\operatorname{gph} G$.

Next we specialize Theorem 4B.6 to the generalized equations we studied in detail in Chapters 2 and 3, or in other words, to a solution mapping of the type

$$(27)\qquad S(p) = \{\, x \mid f(p,x) + F(x) \ni 0 \,\},$$

where $f:\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^m$ and $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$. In the next two corollaries we take a closer look at the Aubin property of the solution mapping (27).

Corollary 4B.7 (derivative criterion for generalized equations)**.** *For the solution mapping S in (27), and a pair $(\bar{p},\bar{x})$ with $\bar{x}\in S(\bar{p})$, suppose that f is continuous around $(\bar{p},\bar{x})$, F has locally closed graph at $\bar{x}$ for $-f(\bar{p},\bar{x})$ and also $\widehat{\mathrm{lip}}_p(f;(\bar{p},\bar{x})) < \infty$. Then the mapping $G(p,x) := f(p,x)+F(x)$ has the partial Aubin property with respect to p uniformly in x at $(\bar{p},\bar{x})$ for 0 with*

$$(28)\qquad \widehat{\mathrm{lip}}_p(G;\bar{p},\bar{x}\,|\,0) \le \widehat{\mathrm{lip}}_p(f;(\bar{p},\bar{x})).$$

In addition, if f is differentiable in a neighborhood of $(\bar{p},\bar{x})$, $\operatorname{gph} F$ is locally closed at $(\bar{x},-f(\bar{p},\bar{x}))$ and

$$(29)\qquad \limsup_{\substack{(p,x,y)\to(\bar{p},\bar{x},0)\\ y\in f(p,x)+F(x)}} |(D_xf(p,x)+DF(x\,|\,y-f(p,x)))^{-1}|^- \le \lambda < \infty,$$

then S has the Aubin property at $\bar{p}$ for $\bar{x}$ with

$$(30)\qquad \mathrm{lip}\,(S;\bar{p}\,|\,\bar{x}) \le \lambda\, \widehat{\mathrm{lip}}_p(f;\bar{p}\,|\,\bar{x}).$$

Proof. By definition, the mapping G has $(\bar{p},\bar{x},0) \in \operatorname{gph} G$. Let $\mu > \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x}))$ and let Q and U be neighborhoods of $\bar{p}$ and $\bar{x}$ such that f is Lipschitz continuous with respect to $p \in Q$ uniformly in $x \in U$ with Lipschitz constant μ. Let $p, p' \in Q$, $x \in U$ and $y \in G(p,x)$; then $y - f(p,x) \in F(x)$ and we have

$$d(y, G(p',x)) = d(y - f(p',x), F(x)) \leq |f(p,x) - f(p',x)| \leq \mu |p - p'|.$$

Thus,

$$e(G(p,x), G(p',x)) \leq \mu |p' - p|$$

and hence G has the partial Aubin (actually, Lipschitz) property with respect to p uniformly in x at $(\bar{p},\bar{x})$ with modulus satisfying (28). The assumptions that f is differentiable near $(\bar{p},\bar{x})$ and $\operatorname{gph} F$ is locally closed at $(\bar{x}, -f(\bar{p},\bar{x}))$ yield that $\operatorname{gph} G$ is locally closed at $(\bar{p},\bar{x},0)$ as well. Further, observe that the function $p \mapsto d(0, G(p,\bar{x})) = d(-f(p,\bar{x}), F(\bar{x}))$ is Lipschitz continuous near $\bar{p}$ and therefore upper semicontinuous at $\bar{p}$. Then we can apply Theorem 4B.6 where, by using the sum rule 4A.2, the condition (25) comes down to (29) while (26) yields (30). □

From Section 3F we know that when the function f is continuously differentiable, the Aubin property of the solution mapping in (27) can be obtained by passing to the linearized generalized equation, in which case we can also utilize the ample parameterization condition. Specifically, we have the following result:

Corollary 4B.8 (derivative criterion with differentiability and ample parameterization). *For the solution mapping S in (27), and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$, suppose that f is continuously differentiable on a neighborhood of $(\bar{p},\bar{x})$ and that $\operatorname{gph} F$ is locally closed at $(\bar{x}, -f(\bar{p},\bar{x}))$. If*

$$(31) \qquad \limsup_{\substack{(x,y)\to(\bar{x},-f(\bar{p},\bar{x}))\\ y\in D_x f(\bar{p},\bar{x})(x-\bar{x})+F(x)}} |(\nabla_x f(\bar{p},\bar{x}) + DF(x\,|\,y - D_x f(\bar{p},\bar{x})(x-\bar{x})))^{-1}|^{-} \leq \lambda < \infty,$$

then S has the Aubin property at $\bar{p}$ for $\bar{x}$, moreover with

$$(32) \qquad \operatorname{lip}(S;\bar{p}\,|\,\bar{x}) \leq \lambda\, |\nabla_p f(\bar{p},\bar{x})|.$$

Furthermore, when f satisfies the ample parameterization condition

$$(33) \qquad \operatorname{rank} \nabla_p f(\bar{p},\bar{x}) = m,$$

then the converse implication holds as well; that is, S has the Aubin property at $\bar{p}$ for $\bar{x}$ if and only if condition (31) is satisfied.

Proof. According to Theorem 3F.9, the mapping S has the Aubin property at $\bar{p}$ for $\bar{x}$ provided that the linearized mapping

$$(34) \qquad h + F \quad \text{for} \quad h(x) = f(\bar{p},\bar{x}) + D_x f(\bar{p},\bar{x})(x - \bar{x})$$

is metrically regular at $\bar{x}$ for 0, and the converse implication holds under the ample parameterization condition (33). Further, according to the derivative criterion for metric regularity 4B.1, metric regularity of the mapping $h+F$ in (34) is equivalent to condition (31) and its regularity modulus is bounded by λ. Then the estimate (32) follows from formula 3F(7) in the statement of 3F.9. □

The purpose of the next exercise is to understand what condition (29) means in the setting of the classical implicit function theorem.

Exercise 4B.9 (application to classical implicit functions)**.** *For a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$, consider the solution mapping*

$$S : p \mapsto \{ x \,|\, f(p,x) = 0 \}$$

and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$. Suppose that f is differentiable in a neighborhood of $(\bar{p},\bar{x})$ with Jacobians satisfying

$$\limsup_{(p,x)\to(\bar{p},\bar{x})} |\nabla_x f(p,x)^{-1}|^- < \lambda \quad \textit{and} \quad \limsup_{(p,x)\to(\bar{p},\bar{x})} |\nabla_p f(p,x)| < \kappa.$$

Show that then S has the Aubin property at $\bar{p}$ for $\bar{x}$ with constant $\lambda\kappa$.

When f is continuously differentiable we can apply Corollary 4B.8, and the assumptions in 4B.9 can in that case be captured by conditions on the Jacobian $\nabla f(\bar{p},\bar{x})$. Then 4B.8 goes a long way toward the classical implicit function theorem, 1A.1. But Steps 2 and 3 of Proof I of that theorem would afterward need to be carried out to reach the conclusion that S has a single-valued localization that is smooth around $\bar{p}$.

Applications of Theorem 4B.6 and its corollaries to constraint systems and variational inequalities will be worked out in Section 4F.

We end this section with yet another proof of the classical inverse function theorem 1A.1. This time it is based on the Ekeland principle given in 4B.5.

Proof of Theorem 1A.1. Without loss of generality, let $\bar{x} = 0$, $f(\bar{x}) = 0$. Let $A = \nabla f(0)$ and let $\delta = |A^{-1}|^{-1}$. Choose $a > 0$ such that

$$(35) \qquad |f(x) - f(x') - A(x - x')| \le \frac{\delta}{2}|x - x'| \quad \text{for all } x, x' \in a\mathbb{B},$$

and let $b = a\delta/2$. We now redo Step 1 in Proof I that the localization s of f^{-1} with respect to the neighborhoods $b\mathbb{B}$ and $a\mathbb{B}$ is nonempty-valued. The other two steps remain the same as in Proof I.

Fix $y \in b\mathbb{B}$ and consider the function $|f(x) - y|$ with domain containing the closed ball $a\mathbb{B}$, which we view as a complete metric space equipped with the Euclidean metric. This function is continuous and bounded below, hence, by Ekeland principle 4B.5 with the indicated δ and $\bar{u} = 0$ there exists $x_\delta \in a\mathbb{B}$ such that

$$(36) \qquad |y - f(x_\delta)| < |y - f(x)| + \frac{\delta}{2}|x - x_\delta| \quad \text{for all } x \in a\mathbb{B},\ x \ne x_\delta.$$

Let us assume that $y \neq f(x_\delta)$. Then $\tilde{x} := A^{-1}(y - f(x_\delta)) + x_\delta \neq x_\delta$. Moreover, from (35) with $x = x_\delta$ and $x' = 0$ and the choice of δ and b we get

$$|\tilde{x}| \leq |A^{-1}|(|y| + |-f(x_\delta) + Ax_\delta|) \leq |A^{-1}|\left(b + \frac{a\delta}{2}\right) = |A^{-1}|a\delta = a.$$

Hence we can set $x = \tilde{x}$ in (36), obtaining

$$|y - f(x_\delta)| < |y - f(\tilde{x})| + \frac{\delta}{2}|\tilde{x} - x_\delta|. \tag{37}$$

Using (35), we have

$$\begin{aligned} |y - f(\tilde{x})| &= |f(A^{-1}(y - f(x_\delta)) + x_\delta) - y| \\ &= |f(A^{-1}(y - f(x_\delta)) + x_\delta) - f(x_\delta) - A(A^{-1}(y - f(x_\delta)))| \\ &\leq \tfrac{\delta}{2}|A^{-1}(y - f(x_\delta))| \end{aligned} \tag{38}$$

and also

$$|\tilde{x} - x_\delta| = |A^{-1}(y - f(x_\delta))|. \tag{39}$$

Plugging (38) and (39) into (37), we arrive at

$$|y - f(x_\delta)| < \left(\frac{\delta}{2} + \frac{\delta}{2}\right)|A^{-1}(y - f(x_\delta))| \leq \delta|A^{-1}||y - f(x_\delta)| = |y - f(x_\delta)|$$

which furnishes a contradiction. Thus, our assumption that $y \neq f(x_\delta)$ is voided, and we have $x_\delta \in f^{-1}(y) \cap (a\mathbb{B})$. This means that s is nonempty-valued, and the proof is complete. □

4C. Coderivative Criterion for Metric Regularity

Normal cones $N_C(x)$ have already been prominent, of course, in our work with optimality conditions and variational inequalities, starting in Section 2A, but only in the case of convex sets. To arrive at coderivatives for a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, we wish to make use of normal cones to $\operatorname{gph} F$ at points (x, y), but to keep the door open to significant applications we need to deal with graph sets that are not convex. The first task, therefore, is generalizing $N_C(x)$ to the case of nonconvex C.

General normal cones. *For a set $C \subset \mathbb{R}^n$ and a point $x \in C$ at which C is locally closed, a vector v is said to be a regular normal at x to C if $\langle v, x' - x\rangle \leq o(|x' - x|)$ for $x' \in C$. The set of all such vectors v is called the regular normal cone to C at x and is denoted by $\hat{N}_C(x)$. A vector v is said to be a general normal to C at x if there* *are sequences* $\{x^k\}$ *and* $\{v^k\}$ *with* $x^k \in C$, *such that*

$$x^k \to x \text{ and } v^k \to v \text{ with } v^k \in \hat{N}_C(x^k).$$

The set of all such vectors v *is called the general normal cone to* C *at* x *and is denoted by* $N_C(x)$. *For* $x \notin C$, $N_C(x)$ *is the empty set.*

Very often, the limit process in the definition of the general normal cone $N_C(x)$ is superfluous: no additional vectors v are produced in that manner, and one merely has $N_C(x) = \hat{N}_C(x)$. This circumstance is termed the *Clarke regularity* of C at x. When C is convex, for instance, it is Clarke regular at every one of its points x, and the generalized normal cone $N_C(x)$ agrees with the normal cone defined earlier, in 2A. Anyway, $N_C(x)$ is always a closed cone.

Coderivative. *For a mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *and a pair* $(x,y) \in \operatorname{gph} F$ *at which* $\operatorname{gph} F$ *is locally closed, the coderivative of* F *at* x *for* y *is the mapping* $D^*F(x\,|\,y) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *defined by*

$$w \in D^*F(x\,|\,y)(z) \iff (w,-z) \in N_{\operatorname{gph} F}(x,y).$$

Obviously this is a “dual” sort of notion, but where does it fit in with classical differentiation? The answer can be seen by specializing to the case where F is single-valued, thus reducing to a function $f : \mathbb{R}^n \to \mathbb{R}^m$. Suppose f is strictly differentiable at x; then for $y = f(x)$, the graphical derivative $Df(x\,|\,y)$ is of course the linear mapping $Df(x)$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ with matrix $\nabla f(x)$. In contrast, the coderivative $D^*f(x\,|\,y)$ comes out as the adjoint linear mapping $Df(x)^*$ from $\mathbb{R}^m$ to $\mathbb{R}^n$ with matrix $\nabla f(x)^{\mathsf{T}}$.

Exercise 4C.1 (sum rule for coderivatives)**.** *For a function* $f : \mathbb{R}^n \to \mathbb{R}^m$ *which is strictly differentiable at* $\bar{x}$ *and a mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *with* $\bar{y} \in F(\bar{x})$, *prove that*

$$D^*(f+F)(\bar{x}\,|\,f(\bar{x})+\bar{y})(u) = Df(\bar{x})^*u + D^*F(\bar{x}\,|\,\bar{y})(u) \quad \text{for all } \ u \in \mathbb{R}^m.$$

An important fact about coderivatives in our context is the following characterization of metric regularity.

Theorem 4C.2 (coderivative criterion for metric regularity)**.** *For a mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *and a pair* $(\bar{x},\bar{y}) \in \operatorname{gph} F$ *at which* $\operatorname{gph} F$ *is locally closed, one has*

$$\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = |D^*F(\bar{x}\,|\,\bar{y})^{-1}|^+. \tag{1}$$

Thus, F *is metrically regular at* $\bar{x}$ *for* $\bar{y}$ *if and only if the right side of (1) is finite, which is equivalent to*

$$D^*F(\bar{x}\,|\,\bar{y})(u) \ni 0 \implies u = 0.$$

If F is single-valued, that is, a function $f : \mathbb{R}^n \to \mathbb{R}^m$ which is strictly differentiable at $\bar{x}$, then the coderivative criterion means that the adjoint to the derivative mapping $Df(\bar{x})$ is injective, that is, $\ker \nabla f(\bar{x})^{\mathsf{T}} = \{0\}$. This is equivalent to surjectivity of $Df(\bar{x})$ which, as we know from 3F or 4B, is equivalent to metric regularity of f at $\bar{x}$ for $f(\bar{x})$.

We will present a proof of Theorem 4C.2 based on the following more general result:

Theorem 4C.3 (basic equality)**.** *Let $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued map, let $\bar{y} \in F(\bar{x})$ and assume that* gph F *is locally closed at $(\bar{x}, \bar{y})$. Then*

(2) $$\limsup_{\substack{(x,y)\to(\bar{x},\bar{y}),\\ (x,y)\in \operatorname{gph} F}} |DF(x\,|\,y)^{-1}|^{-} = |D^*F(\bar{x}\,|\,\bar{y})^{-1}|^{+}.$$

In the proof of Theorem 4C.3 we employ the following lemma:

Lemma 4C.4 (intersection with tangent cone)**.** *Let C be a convex and compact set in $\mathbb{R}^d$, $K \subset \mathbb{R}^d$ be a closed set and $\bar{x} \in K$. Then $C \cap T_K(x) \neq \emptyset$ for all $x \in K$ near $\bar{x}$ if and only if $C \cap \operatorname{clco} T_K(x) \neq \emptyset$ for all $x \in K$ near $\bar{x}$.*

Proof. Clearly, $C \cap T_K(x) \neq \emptyset$ implies $C \cap \operatorname{clco} T_K(x) \neq \emptyset$. Assume that there exists an open neighborhood U of $\bar{x}$ such that $C \cap \operatorname{clco} T_K(x) \neq \emptyset$ for all $x \in K \cap U$. Let $\varepsilon > 0$ be such that $\mathbb{B}_\varepsilon(\bar{x}) \subset U$. Take any $x \in \mathbb{B}_{\varepsilon/3}(\bar{x})$ and let v be a projection of x on K. Then $|v - x| \leq |\bar{x} - x| \leq \varepsilon/3$ and hence

$$|v - \bar{x}| \leq |v - x| + |x - \bar{x}| \leq \varepsilon/3 + \varepsilon/3 < \varepsilon.$$

Thus, there exists an open neighborhood W of $\bar{x}$ such that any metric projection of a point $x \in W$ on K belongs to $K \cap U$.

Fix $x \in K \cap W$. For all $t \geq 0$ define $\varphi(t) := \min\{|u - v| \mid u \in x + tC, v \in K\}$. The function φ is Lipschitz continuous. Indeed, for every $t_i \geq 0$, $i = 1,2$ there exist $c_i \in C$ and $k_i \in K$ such that $\varphi(t_i) = |x + t_i c_i - k_i|$, $i = 1,2$. Then

$$\begin{aligned}\varphi(t_1) - \varphi(t_2) &= |x + t_1 c_1 - k_1| - |x + t_2 c_2 - k_2| \\ &\leq |x + t_1 c_2 - k_2| - |x + t_2 c_2 - k_2| \leq |c_2||t_1 - t_2|.\end{aligned}$$

Hence φ is absolutely continuous, that is, its derivative φ' exists almost everywhere and $\varphi(s) = \varphi(t) + \int_t^s \varphi'(\tau)d\tau$ for all $s \geq t \geq 0$. We will prove next that

(3) $$\varphi(t) = 0 \quad \text{for all sufficiently small} \quad t > 0.$$

If this holds, then for every small $t > 0$ there exists $v_t \in C$ such that $x + tv_t \in K$. Consider sequences $t_k \searrow 0$ and $v_{t_k} \in C$ such that v_{t_k} converges to some v. Then $v \in T_K(x) \cap C$ and since $x \in K \cap W$ is arbitrary, we arrive at the claim of the lemma.

To prove (3), let $\gamma > 0$ be such that $x + [0,\gamma]C \subset W$. Assume that there exists $t_0 \in (0,\gamma]$ such that $\varphi(t_0) > 0$. Define $\bar{t} = \max\{t \mid \varphi(t) = 0 \text{ and } 0 \leq t < t_0\}$. Let $t \in (\bar{t}, t_0)$ be such that $\varphi'(t)$ exists. Then for some $v_t \in C$ and $x_t \in K$ we have $\varphi(t) = |x + tv_t - x_t| > 0$. Since x_t is a projection of $x + tv_t$ on K, by the observation in the beginning of the proof we have $x_t \in K \cap U$. By assumption, there exists $w_t \in \operatorname{clco} T_K(x_t)$ such that $w_t \in C$. Then, for any $h > 0$ sufficiently small,

$$x + tv_t + hw_t = x + (t+h)\left(\frac{t}{t+h}v_t + \frac{h}{t+h}w_t\right) \in x + (t+h)C \subset W$$

because the set C is assumed convex. Thus

$$\varphi(t+h) - \varphi(t) \le |x + tv_t + hw_t - x_t| - |x + tv_t - x_t|.$$

Dividing both sides of this inequality by $h > 0$ and passing to the limit when $h \searrow 0$, we get

$$\varphi'(t) \le \left\langle \frac{x + tv_t - x_t}{|x + tv_t - x_t|}, w_t \right\rangle.$$

Recall that x_t is a projection of $x + tv_t$ on K and also the elementary fact[1] that in this case $x + tv_t - x_t \in \hat{N}_K(x_t)$. Since $w_t \in \operatorname{clco} T_K(x_t)$, we obtain from the inequality above that $\varphi'(t) \le 0$. Having in mind that t is any point of differentiability of φ in $(\bar{t}, t_0)$, we get $\varphi(t_0) \le \varphi(\bar{t}) = 0$. This contradicts the choice of t_0 according to which $\varphi(t_0) > 0$. Hence (3) holds and the lemma is proved. □

Proof of Theorem 4C.3. Since the graphical derivative and the coderivative are defined only locally around $(\bar{x}, \bar{y})$, we can assume without loss of generality that the graph of the mapping F is closed. We will show first that

(4)
$$\limsup_{\substack{(x,y)\to(\bar{x},\bar{y}),\\ (x,y)\in \operatorname{gph} F}} |DF(x\,|\,y)^{-1}|^{-} \ge |D^*F(\bar{x}\,|\,\bar{y})^{-1}|^{+}.$$

If the left side of (4) equals ∞ there is nothing to prove. Suppose that a constant c satisfies

$$\limsup_{\substack{(x,y)\to(\bar{x},\bar{y}),\\ (x,y)\in \operatorname{gph} F}} |DF(x\,|\,y)^{-1}|^{-} < c.$$

From the properties of the outer norm, see (11) in 4A.6, there exists $\delta > 0$ such that for all $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_\delta(\bar{x}) \times \mathbb{B}_\delta(\bar{y}))$ and for every $v \in \mathbb{B}$ there exists $u \in DF(x\,|\,y)^{-1}(v)$ such that $|u| < c$. Also, note that

$$(u,v) \in T_{\operatorname{gph} F}(x,y) \subset \operatorname{clco} T_{\operatorname{gph} F}(x,y) = T^{**}_{\operatorname{gph} F}(x,y).$$

Fix $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_\delta(\bar{x}) \times \mathbb{B}_\delta(\bar{y}))$ and let $v \in \mathbb{B} \subset \mathbb{R}^m$. Then there exists u with $(u,v) \in T^{**}_{\operatorname{gph} F}(x,y)$ such that $u = cw$ for some $w \in \mathbb{B}$. Let $(p,q) \in \hat{N}_{\operatorname{gph} F}(x,y) = T^{*}_{\operatorname{gph} F}(x,y)$. From the inequality $\langle u,p\rangle + \langle v,q\rangle \le 0$ we get

$$c \min_{w\in \mathbb{B}} \langle w,p\rangle + \langle v,q\rangle \le 0 \quad \text{which yields} \quad -c|p| + \langle v,q\rangle \le 0.$$

Since v is arbitrarily chosen in $\mathbb{B}$, we conclude that

---

[1] See Example 6.16 in the book Rockafellar and Wets [1998].

(5) $$|q| \leq c|p| \quad \text{whenever} \quad (p,q) \in \hat{N}_{\mathrm{gph}\, F}(x,y).$$

Now, let $(-p,q) \in N_{\mathrm{gph}\, F}(\bar{x},\bar{y})$; then there exist sequences $(x_k,y_k) \in \mathrm{gph}\, F$, $(x_k,y_k) \to (\bar{x},\bar{y})$ and $(-p_k,q_k) \in \hat{N}_{\mathrm{gph}\, F}(x_k,y_k)$ such that $(-p_k,q_k) \to (-p,q)$. But then, from (5), $|q_k| \leq c|p_k|$ and in the limit $|q| \leq c|p|$. Thus, $|q| \leq c|p|$ whenever $(-p,q) \in N_{\mathrm{gph}\, F}(\bar{x},\bar{y})$ and therefore we have $|q| \leq c|p|$ whenever $(q,-p) \in N_{\mathrm{gph}\, F^{-1}}(\bar{y},\bar{x})$. By the definition of the coderivative,

$$|q| \leq c|p| \quad \text{whenever} \quad q \in D^*F(\bar{x}\,|\,\bar{y})^{-1}(p).$$

This, together with the property of the outer norm given in 4A.6(12), implies that $c \geq |D^*F(\bar{x}\,|\,\bar{y})^{-1}|^+$ and we obtain (4) since c is arbitrary.

For the converse inequality, it is enough to consider the case when there is a constant c such that

(6) $$|D^*F(\bar{x}\,|\,\bar{y})^{-1}|^+ < c.$$

We first show there exists $\delta > 0$ such that for $(x,y) \in \mathrm{gph}\, F \cap (I\!B_\delta(\bar{x}) \times I\!B_\delta(\bar{y}))$ we have that

(7) $$(0,v) \in \hat{N}_{\mathrm{gph}\, F}(x,y) \quad \Longrightarrow \quad v = 0.$$

On the contrary, assume that there exist sequences $(x_k,y_k) \in \mathrm{gph}\, F$ with $(x_k,y_k) \to (\bar{x},\bar{y})$ and $v_k \in I\!R^m$ with $|v_k| = 1$ such that $(0,v_k) \in \hat{N}_{\mathrm{gph}\, F}(x_k,y_k)$ for all k. Then there is $v \neq 0$ such that $(0,v) \in N_{\mathrm{gph}\, F}(\bar{x},\bar{y})$, that is, $-v \in D^*F(\bar{x}\,|\,\bar{y})^{-1}(0)$. Taking into account 4A.6(12), this contradicts (6).

We will now prove a statement more general than (7); namely, there exists $\delta > 0$ such that for every $(x,y) \in \mathrm{gph}\, F \cap (I\!B_\delta(\bar{x}) \times I\!B_\delta(\bar{y}))$ we have

(8) $$(v,-u) \in \hat{N}_{\mathrm{gph}\, F^{-1}}(y,x) \quad \Longrightarrow \quad |v| \leq c|u|.$$

On the contrary, assume that there exists a sequence $(y_k,x_k) \to (\bar{y},\bar{x})$ such that for each k we can find $(v_k,-u_k) \in \hat{N}_{\mathrm{gph}\, F^{-1}}(y_k,x_k)$ satisfying $|v_k| > c|u_k|$. If $u_k = 0$ for some k, then from (7) we get $v_k = 0$, a contradiction. Thus, without loss of generality we assume that $|u_k| = 1$. Let v_k be unbounded and let w be a cluster point of $\frac{1}{|v_k|}v_k$; then $|w| = 1$. Since $(\frac{1}{|v_k|}v_k, -\frac{1}{|v_k|}u_k) \in \hat{N}_{\mathrm{gph}\, F^{-1}}(y_k,x_k)$, passing to the limit we get $(w,0) \in N_{\mathrm{gph}\, F^{-1}}(\bar{y},\bar{x})$ which contradicts (6) because of (12) in 4A.6. Further, if v_k is bounded, then $(v_k,u_k) \to (v,u)$ for a subsequence, where $|u| = 1$, $(v,-u) \in N_{\mathrm{gph}\, F^{-1}}(\bar{y},\bar{x})$, and $|v| \geq c$. This again contradicts (6). Thus, (8) holds for all $(y,x) \in \mathrm{gph}\, F^{-1}$ close to $(\bar{y},\bar{x})$.

Let $\delta > 0$ be such that (8) is satisfied for $(x,y) \in \mathrm{gph}\, F \cap (I\!B_\delta(\bar{x}) \times I\!B_\delta(\bar{y}))$. Pick such (x,y). We will show that

(9) $$(cI\!B \times \{w\}) \cap T^{**}_{\mathrm{gph}\, F}(x,y) \neq \emptyset \quad \text{for every } w \in I\!B.$$

On the contrary, assume that there exists $w \in \mathbb{B}$ such that $(c\mathbb{B} \times \{w\}) \cap T^{**}_{\mathrm{gph}\,F}(x,y) = \emptyset$. Then, by the theorem on separation of convex sets (see Theorem 5C.12 for a general statement), there exists a nonzero $(p,q) \in T^{*}_{\mathrm{gph}\,F}(x,y) = \hat{N}_{\mathrm{gph}\,F}(x,y)$ such that

$$\min_{u \in \mathbb{B}} \langle p, cu \rangle + \langle q, w \rangle > 0.$$

Indeed, it is sufficient to choose a hyperplane through the origin that separates the sets $(c\mathbb{B} \times \{w\})$ and $T^{**}_{\mathrm{gph}\,F}(x,y)$; then take the normal (p,q) pointing toward the half-space containing $(c\mathbb{B} \times \{w\})$. If $p = 0$, then $q \neq 0$ and then $(q,0) \in \hat{N}_{\mathrm{gph}\,F^{-1}}(y,x)$ in contradiction with (8). Hence, $p \neq 0$. Without loss of generality, let $|p| = 1$. Then $(q,p) \in \hat{N}_{\mathrm{gph}\,F^{-1}}(y,x)$ and

$$\langle q, w \rangle > \max_{u \in \mathbb{B}} \langle p, cu \rangle = c|p| = c. \tag{10}$$

By (8), $|q| \leq c$ and since $w \in \mathbb{B}$, this contradicts (10). Thus, (9) is satisfied.

By Lemma 4C.4, for all $(x,y) \in \mathrm{gph}\,F$ sufficiently close to $(\bar{x},\bar{y})$, we have that (9) holds when the set $T^{**}_{\mathrm{gph}\,F}(x,y) = \mathrm{cl\,co}\, T_{\mathrm{gph}\,F}(x,y)$ is replaced with $T_{\mathrm{gph}\,F}(x,y)$. This means that for every $w \in \mathbb{B}$ there exists $u \in DF(x\,|\,y)^{-1}(w)$ such that $|u| \leq c$. But then $c \geq |DF(x\,|\,y)^{-1}|^{-}$ for all $(x,y) \in \mathrm{gph}\,F$ sufficiently close to $(\bar{x},\bar{y})$. This combined with the arbitrariness of c in (6) implies the inequality opposite to (4) and hence the proof of the theorem if complete. $\square$

Exercise 4C.5 (coderivative criterion for generalized equations)**.** *For the solution mapping $S(p) = \{x \mid f(p,x) + F(x) \ni 0\}$ and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$, suppose that f is differentiable in a neighborhood of $(\bar{p},\bar{x})$, $\mathrm{gph}\,F$ is locally closed at $(\bar{x}, -f(\bar{p},\bar{x}))$, and*

$$|(D^*_x f(\bar{p},\bar{x}) + D^*F(\bar{x}\,|\,-f(\bar{p},\bar{x})))^{-1}|^{+} \leq \lambda < \infty.$$

Then S has the Aubin property at $\bar{p}$ for $\bar{x}$ with

$$\mathrm{lip}\,(S;\bar{p}\,|\,\bar{x}) \leq \lambda\, \widehat{\mathrm{lip}}_p(f;(\bar{p},\bar{x})).$$

We conclude the present section with a variant of the graphical derivative formula for the modulus of metric regularity, which will be put to use in the numerical variational analysis of Chapter 6.

Convexified graphical derivative. *For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a pair (x,y) with $y \in F(x)$, the convexified graphical derivative of F at x for y is the mapping $\tilde{D}F(x\,|\,y) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ whose graph is the closed convex hull of the tangent cone $T_{\mathrm{gph}\,F}(x,y)$ to $\mathrm{gph}\,F$ at (x,y):*

$$v \in \tilde{D}F(x\,|\,y)(u) \iff (u,v) \in \mathrm{cl\,co}\, T_{\mathrm{gph}\,F}(x,y).$$

Theorem 4C.6 (alternative characterization of regularity modulus)**.** *For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a point $(\bar{x},\bar{y}) \in \operatorname{gph} F$ at which the graph of F is locally closed, one has*

$$\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \limsup_{\substack{(x,y)\to(\bar{x},\bar{y})\\(x,y)\in\operatorname{gph} F}} |\tilde{D}F(x\,|\,y)^{-1}|^{-}.$$

Proof. Since $\tilde{D}F(x\,|\,y)^{-1}(v) \supset DF(x\,|\,y)^{-1}(v)$ we obtain

$$|\tilde{D}F(x\,|\,y)^{-1}|^{-} \le |DF(x\,|\,y)^{-1}|^{-}.$$

Thus, from (1),

$$\limsup_{\substack{(x,y)\to(\bar{x},\bar{y})\\(x,y)\in\operatorname{gph} F}} |\tilde{D}F(x\,|\,y)^{-1}|^{-} \le |D^*F(\bar{x}\,|\,\bar{y})^{-1}|^{+}.$$

The converse inequality follows from the first part of the proof of Theorem 4C.3, by limiting the argument to the convexified graphical derivative. □

Exercise 4C.7 (sum rule for convexified derivatives)**.** *For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ which is differentiable at x and a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, prove that*

$$\tilde{D}(f+F)(x\,|\,f(x)+y)(u) = Df(x)(u) + \tilde{D}F(x\,|\,y)(u).$$

Guide. Let $v \in \tilde{D}(f+F)(x\,|\,f(x)+y)(u)$. By Carathéodory's theorem on convex hull representation of cones, there are sequences $\{u_i^k\}$, $\{v_i^k\}$ and $\{\lambda_i^k\}$ for $i = 0,1,\ldots,n+m$ and $k = 1,2,\ldots$ with $\lambda_i^k \ge 0$, $\sum_{i=0}^{n+m} \lambda_i^k = 1$, such that $v_i^k \in D(f+F)(x\,|\,f(x)+y)(u_i^k)$ for all i and k and $\sum_{i=0}^{n+m} \lambda_i^k(u_i^k, v_i^k) \to (u,v)$ as $k \to \infty$. From 4A.2, get $v_i^k \in Df(x)u_i^k + DF(x\,|\,y)(u_i^k)$ for all i and k. Hence $\sum_{i=0}^{n+m} \lambda_i^k(u_i^k, v_i^k - Df(x)u_i^k) \in \operatorname{cl}\operatorname{co}\operatorname{gph} DF(x\,|\,y)$. Then pass to the limit. □

Exercise 4C.8 (convexified derivative criterion for generalized equations)**.** *State and prove a result parallel to 4B.7 with the graphical derivative replaced by the convexified graphical derivative.*

4D. Strict Derivative Criterion for Strong Metric Regularity

In order to characterize the strong metric regularity by "differentiation" we need to appeal to another type of derivative of a set-valued mapping.

Strict graphical derivative for a set-valued mapping. *For a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ the strict graphical derivative mapping $D_*F(\bar{x}\,|\,\bar{y})$ at $\bar{x}$ for $\bar{y}$, where $\bar{y} \in F(\bar{x})$, is defined as a mapping whose graph is the collection of vectors (u,v) for which there ex-*

ist sequences $(x^k, y^k) \in \operatorname{gph} F$, $(x^k, y^k) \to (\bar{x}, \bar{y})$, *as well as* $\tau^k \searrow 0$ *and* $(u^k, v^k) \to (u, v)$ *such that* $(x^k + \tau^k u^k, y^k + \tau^k v^k) \in \operatorname{gph} F$.

We are now ready to state a prove a criterion for strong metric regularity based on the strict graphical derivative.

Theorem 4D.1 (strict derivative criterion for strong metric regularity)**.** *Consider a set-valued mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *and* $(\bar{x}, \bar{y}) \in \operatorname{gph} F$. *If* F *is strongly metrically regular at* $\bar{x}$ *for* $\bar{y}$, *then*

$$(1) \qquad |D_*F(\bar{x}\,|\,\bar{y})^{-1}|^+ < \infty.$$

On the other hand, if the graph of F *is locally closed at* $(\bar{x}, \bar{y})$ *and*

$$(2) \qquad \bar{x} \in \liminf_{y \to \bar{y}} F^{-1}(y),$$

then condition (1) is also sufficient for strong metric regularity of F *at* $\bar{x}$ *for* $\bar{y}$. *In this case the quantity on the left side of (1) equals* $\operatorname{reg}(F; \bar{x}\,|\,\bar{y})$.

Proof. Proposition 3G.1 says that a mapping F is strongly metrically regular at $\bar{x}$ for $\bar{y}$ if and only if it is metrically regular there and F^{-1} has a localization around $\bar{y}$ for $\bar{x}$ which is nowhere multivalued. Furthermore, in this case for every $c > \operatorname{reg}(F; \bar{x}\,|\,\bar{y})$ there exists a neighborhood V of $\bar{y}$ such that F^{-1} has a localization around $\bar{y}$ for $\bar{x}$ which is a Lipschitz continuous function on V with constant c.

Let F be strongly metrically regular at $\bar{x}$ for $\bar{y}$, let $c > \operatorname{reg}(F; \bar{x}\,|\,\bar{y})$ and let U and V be open neighborhoods of $\bar{x}$ and $\bar{y}$, respectively, such that the localization $V \ni y \mapsto \varphi(y) := F^{-1}(y) \cap U$ is a Lipschitz continuous function on V with a Lipschitz constant c. We will show first that for every $v \in \mathbb{R}^m$ the set $D_*F(\bar{x}\,|\,\bar{y})^{-1}(v)$ is nonempty. Let $v \in \mathbb{R}^m$. Since $\operatorname{dom} \varphi \supset V$, we can choose sequences $\tau^k \searrow 0$ and u^k such that $\bar{x} + \tau^k u^k = \varphi(\bar{y} + \tau^k v)$ for large k. Then, from the Lipschitz continuity of φ with Lipschitz constant c we conclude that $|u^k| \leq c|v|$, hence u^k has a cluster point u which, by definition, is from $D_*F(\bar{x}\,|\,\bar{y})^{-1}(v)$. Now choose any $v \in \mathbb{R}^m$ and $u \in D_*F(\bar{x}\,|\,\bar{y})^{-1}(v)$; then there exist sequences $(x^k, y^k) \in \operatorname{gph} F$, $(x^k, y^k) \to (\bar{x}, \bar{y})$, $\tau^k \searrow 0$, $u^k \to u$ and $v^k \to v$ such that $y^k + \tau^k v^k \in V$, $x^k = \varphi(y^k)$ and $x^k + \tau^k u^k = \varphi(y^k + \tau^k v^k)$ for k sufficiently large. But then, again from the Lipschitz continuity of φ with Lipschitz constant c, we obtain that $|u^k| \leq c|v^k|$. Passing to the limit we conclude that $|u| \leq c|v|$ which implies that $|D_*F(\bar{x}\,|\,\bar{y})^{-1}|^+ \leq c$. Hence (1) is satisfied; moreover, the quantity on the left side of (1) is less than or equal to $\operatorname{reg}(F; \bar{x}\,|\,\bar{y})$.

To prove the second statement, we first show that F^{-1} has a single-valued bounded localization, that is there exist a bounded neighborhood U of $\bar{x}$ and a neighborhood V of $\bar{y}$ such that $V \ni y \mapsto F^{-1}(y) \cap U$ is single valued. On the contrary, assume that for any bounded neighborhood U of $\bar{x}$ and any neighborhood V of $\bar{y}$ the intersection $\operatorname{gph} F^{-1} \cap (V \times U)$ is the graph of a multivalued mapping. This means that there exist sequences $\varepsilon_k \searrow 0$, $x_k \to \bar{x}$, $x'_k \to \bar{x}$, $x_k \neq x'_k$ for all k such that $F(x_k) \cap F(x'_k) \cap \mathbb{B}_{\varepsilon_k}(\bar{y}) \neq \emptyset$ for all k. Let $t_k = |x_k - x'_k|$ and let $u_k = (x_k - x'_k)/t_k$. Then $t_k \searrow 0$ and $|u_k| = 1$ for all k. Hence $\{u_k\}$ has a cluster point $u \neq 0$. Consider any

$y_k \in F(x_k) \cap F(x'_k) \cap \mathbb{B}_{\varepsilon_k}(\bar{y})$. Then, $y_k + t_k 0 \in F(x'_k + t_k u_k)$ for all k. By the definition of the strict graphical derivative, $0 \in D_*F(\bar{x}|\bar{y})(u)$. Hence $|D_*F(\bar{x}|\bar{y})^{-1}|^+ = \infty$, which contradicts (1). Thus, there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that $\varphi(y) := F^{-1}(y) \cap U$ is at most single-valued on V and U is bounded. By assumption (2), there exists a neighborhood $V' \subset V$ of $\bar{y}$ such that $F^{-1}(y) \cap U \neq \emptyset$ for any $y \in V'$, hence $V' \subset \operatorname{dom} \varphi$. Further, since $\operatorname{gph} F$ is locally closed at $(\bar{x}, \bar{y})$ and φ is bounded, there exists an open neighborhood $V'' \subset V'$ of $\bar{y}$ such that φ is a continuous function on V''.

From the definition of the strict graphical derivative we obtain that the set-valued mapping $(x,y) \mapsto D_*F(x|y)$ has closed graph. We claim that condition (1) implies that

$$\limsup_{\substack{(x,y)\to(\bar{x},\bar{y}),\\ (x,y)\in \operatorname{gph} F}} |D_*F(x|y)^{-1}|^+ < \infty. \tag{3}$$

On the contrary, assume that there exist sequences $(x_k, y_k) \in \operatorname{gph} F$ converging to $(\bar{x}, \bar{y})$, $v_k \in \mathbb{B}$ and $u_k \in D_*F(x_k|y_k)^{-1}(v_k)$ such that $|u_k| > k|v_k|$.

Case 1: There exists a subsequence $v_{k_i} = 0$ for all k_i. Since $\operatorname{gph} D_*F(x_{k_i}|y_{k_i})^{-1}$ is a cone, we may assume that $|u_{k_i}| = 1$. Let u be a cluster point of $\{u_{k_i}\}$. Then, passing to the limit we get $0 \neq u \in D_*F(\bar{x}|\bar{y})^{-1}(0)$ which, combined with formula (12) in 4A.6, contradicts (1).

Case 2: For all large k, $v_k \neq 0$. Since $\operatorname{gph} D_*F(x_k|y_k)^{-1}$ is a cone, we may assume that $|v_k| = 1$. Then $\lim_{k\to\infty} |u_k| = \infty$. Define

$$w_k := \frac{1}{|u_k|} u_k \in D_*F(x_k|y_k)^{-1}\left(\frac{1}{|u_k|} v_k\right)$$

and let w be a cluster point of w_k. Then, passing to the limit we obtain $0 \neq w \in D_*F(\bar{x}|\bar{y})^{-1}(0)$ which, combined with 4A(12), again contradicts (1).

Hence (3) is satisfied. Therefore, there exists an open neighborhood $\tilde{V} \subset V''$ of $\bar{y}$ such that $|D_*F(\varphi(y)|y)^{-1}|^+ < \infty$ for all $y \in \tilde{V}$. We will now prove that for every $(x,y) \in \operatorname{gph} F$ near $(\bar{x}, \bar{y})$ and every $v \in \mathbb{R}^m$ we have that $DF(x|y)^{-1}(v) \neq \emptyset$. Fix $(x,y) \in \operatorname{gph} F \cap (U \times \tilde{V})$ and $v \in \mathbb{R}^m$, and let $t^k \searrow 0$; then there exist $u^k \in \mathbb{R}^n$ such that $x + t^k u^k = F^{-1}(y + t^k v) \cap U = \varphi(y + t^k v)$ for all large k and we also have that $t^k u^k \to 0$ by the continuity of φ. Assume that $|u^k| \to \infty$ for some subsequence (which is denoted in the same way without loss of generality). Set $\tau^k = t^k |u^k|$ and $w^k = \frac{1}{|u^k|} u^k$. Then $\tau^k \searrow 0$ and, for a further subsequence, $w^k \to w$ for some w with $|w| = 1$. Since $(y + \tau^k \frac{1}{|u^k|} v, x + \tau^k w^k) \in \operatorname{gph} F^{-1}$ we obtain that $w \in DF(x|y)^{-1}(0) \subset D_*F(x|y)^{-1}(0)$ for some $w \neq 0$. Thus $|D_*F(x|y)^{-1}|^+ = \infty$ contradicting the choice of $\tilde{V}$. Hence the sequence $\{u^k\}$ cannot be unbounded and since $y + t^k v \in F(x + t^k u^k)$ for all k, any cluster point u of $\{u^k\}$ satisfies $u \in DF(x|y)^{-1}(v)$. Hence $DF(x|y)^{-1}$ is nonempty-valued. From this, 4A.5, and the inclusion $DF(x|y)^{-1}(v) \subset D_*F(x|y)^{-1}(v)$ we obtain

(4) $$|DF(x\,|\,y)^{-1}|^{-} \leq |D_*F(x\,|\,y)^{-1}|^{+}.$$

Putting together (3) and (4), and utilizing the derivative criterion for metric regularity in 4B.1, we obtain that F is metrically regular at $\bar{x}$ for $\bar{y}$ with $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$ bounded by the quantity on the left side of (3). But since F^{-1} has a single-valued localization at $\bar{y}$ for $\bar{x}$ we conclude that F is strongly metrically regular at $\bar{x}$ for $\bar{y}$. Moreover, $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$ equals $|D_*F(\bar{x}\,|\,\bar{y})^{-1}|^{+}$. The proof is complete. □

The following corollary is an application of the strict derivative criterion to the solution mapping of a generalized equation,

(5) $$S(p) = \left\{ x \,\middle|\, f(p,x) + F(x) \ni 0 \right\},$$

where $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$.

Corollary 4D.2 (strict derivative rule)**.** *For the solution mapping S in (5), and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$, suppose that f is continuously differentiable around $(\bar{p},\bar{x})$,* gph F *is locally closed at $(\bar{x}, -f(\bar{p},\bar{x}))$, $\bar{x} \in \liminf_{p\to\bar{p}} S(p)$, and*

(6) $$|(D_x f(\bar{p},\bar{x}) + D_*F(\bar{x}\,|-f(\bar{p},\bar{x})))^{-1}|^{+} \leq \lambda < \infty.$$

Then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ with

(7) $$\operatorname{lip}(s;\bar{p}) \leq \lambda\, \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})).$$

Furthermore, when f satisfies the ample parameterization condition

(8) $$\operatorname{rank} \nabla_p f(\bar{p},\bar{x}) = m,$$

then the converse implication holds as well; that is, S has Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ if and only if (6) is satisfied.

Proof. By 2B.9 the mapping S has a Lipschitz single-valued localization around $\bar{p}$ for $\bar{x}$ provided that the mapping

$$x \mapsto D_x f(\bar{p},\bar{x})(x-\bar{x}) + F(x)$$

is strongly metrically regular at $\bar{x}$ for $-f(\bar{p},\bar{x})$, and the converse implication holds under (8). From 4D.1, the latter is equivalent to (6) by noting that

$$D_*(D_x f(\bar{p},\bar{x}) + F)(\bar{x}\,|-f(\bar{p},\bar{x})) = D_x f(\bar{p},\bar{x}) + D_*F(\bar{x}\,|-f(\bar{p},\bar{x})).$$

Then (7) follows from (6) and 2B(15). □

Although the characterization of strong metric regularity in 4D.1 looks relative simple, the price to be paid still lies ahead: we have to be able to *calculate the strict derivative* in every case of interest. This task could be quite hard for set-valued mappings. A better hold on the existence of single-valued Lipschitz continuous localizations can be gained if we limit our attention to (single-valued) functions.

We first introduce an alternative concept of differentiation of functions, Clarke's generalized Jacobian. It relies on a theorem by Rademacher, according to which a function $f : \mathbb{R}^n \to \mathbb{R}^m$ that is Lipschitz continuous on an open set O is differentiable almost everywhere in O, hence at a set of points that is dense in O. If κ is a Lipschitz constant, then from the definition of Jacobian $\nabla f(x)$ we have $|\nabla f(x)| \leq \kappa$ at those points of differentiability. Recall that a set $C \subset \mathbb{R}^n$ is said to be dense in the closed set D when $\operatorname{cl} C = D$, or equivalently, when, for every $x \in D$, any neighborhood U of x contains elements of C. One of the simplest examples of a dense set is the set of rational numbers relative to the set of real numbers.

Consider now any function $f : \mathbb{R}^n \to \mathbb{R}^m$ and any point $\bar{x} \in \operatorname{int} \operatorname{dom} f$ where $\operatorname{lip}(f;\bar{x}) < \infty$. For any $\kappa > \operatorname{lip}(f;\bar{x})$ we have f Lipschitz continuous with constant κ in some neighborhood U of $\bar{x}$, and hence, from Rademacher's theorem, there is a dense set of points x in U where f is differentiable with $|\nabla f(x)| \leq \kappa$. Hence there exist sequences $x_k \to \bar{x}$ such that f is differentiable at x_k, in which case the corresponding sequence of norms $|\nabla f(x_k)|$ is bounded by the Lipschitz constant κ and hence has at least one cluster point. This leads to the following definition.

Clarke generalized Jacobian. *For $f : \mathbb{R}^n \to \mathbb{R}^m$ and any $\bar{x} \in \operatorname{int} \operatorname{dom} f$ where $\operatorname{lip}(f;\bar{x}) < \infty$, denote by $\bar{\nabla} f(\bar{x})$ the set consisting of all matrices $A \in \mathbb{R}^{m \times n}$ for which there is a sequence of points $x_k \to \bar{x}$ such that f is differentiable at x_k and $\nabla f(x_k) \to A$. The Clarke generalized Jacobian of f at $\bar{x}$, denoted $\bar{\partial} f(\bar{x})$, is the convex hull of this set: $\bar{\partial} f(\bar{x}) = \operatorname{co} \bar{\nabla} f(\bar{x})$.*

Note that $\bar{\nabla} f(\bar{x})$ is a nonempty, closed, bounded subset of $\mathbb{R}^{m \times n}$. This ensures that the convex set $\bar{\partial} f(\bar{x})$ is nonempty, closed, and bounded as well. Furthermore, the mapping $x \mapsto \bar{\partial} f(x)$ has closed graph and is outer semicontinuous, meaning that, given $\bar{x} \in \operatorname{int} \operatorname{dom} f$, for every $\varepsilon > 0$ there exists $\delta > 0$ such that, for all $x \in \mathbb{B}_\delta(\bar{x})$,

$$\bar{\partial} f(x) \subset \bar{\partial} f(\bar{x}) + \varepsilon \mathbb{B}_{m \times n},$$

where $\mathbb{B}_{m \times n}$ is the set consisting of all $m \times n$ matrices whose norm is less or equal to one (the unit ball in the space of $m \times n$ matrices). Strict differentiability of f at $\bar{x}$ is known to be characterized by having $\bar{\nabla} f(\bar{x})$ consist of a single matrix A (or equivalently by having $\bar{\partial} f(\bar{x})$ consist of a single matrix A), in which case $A = \nabla f(\bar{x})$. We will use in this section a mean value theorem for the generalized Jacobian[2] according to which, if f is Lipschitz continuous in an open convex set containing the points x and x', then one has

$$f(x) - f(x') = A(x - x') \quad \text{for some} \quad A \in \operatorname{co} \bigcup_{t \in [0,1]} \bar{\partial} f(tx + (1-t)x').$$

The inverse function theorem which we state next says roughly that a Lipschitz continuous function can be inverted locally around a point $\bar{x}$ when all elements of the generalized Jacobian at $\bar{x}$ are nonsingular. Compared with the classical inverse

[2] This is Theorem 2.6.5 in Clarke [1983].

function theorem, the main difference is that the single-valued graphical localization so obtained can only be claimed to be Lipschitz continuous.

Theorem 4D.3 (Clarke inverse function theorem). *Consider $f : \mathbb{R}^n \to \mathbb{R}^n$ and a point $\bar{x} \in \operatorname{int} \operatorname{dom} f$ where $\operatorname{lip}(f;\bar{x}) < \infty$. Let $\bar{y} = f(\bar{x})$. If all of the matrices in the generalized Jacobian $\bar{\partial} f(\bar{x})$ are nonsingular, then f^{-1} has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$.*

For illustration, we provide elementary examples. The function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x + x^3 & \text{for } x < 0, \\ 2x - x^3 & \text{for } x \geq 0 \end{cases}$$

has generalized Jacobian $\bar{\partial} f(0) = [1,2]$, which does not contain 0. According to Theorem 2H.1, f^{-1} has a Lipschitz continuous single-valued localization around 0 for 0.

In contrast, the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = |x|$ has $\bar{\partial} f(0) = [-1,1]$, which does contain 0. Although the theorem makes no claims about this case, there is no graphical localization of f^{-1} around 0 for 0 that is single-valued.

We will deduce Clarke's theorem from the following more general result concerning mappings of the form $f + F$ that describe generalized equations:

Theorem 4D.4 (inverse function theorem for nonsmooth generalized equations). *Consider a function $f : \mathbb{R}^n \to \mathbb{R}^n$, a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and a point $(\bar{x},\bar{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $\bar{x} \in \operatorname{int} \operatorname{dom} f$, $\operatorname{lip}(f;\bar{x}) < \infty$ and $\bar{y} \in f(\bar{x}) + F(\bar{x})$. Suppose that for every $A \in \bar{\partial} f(\bar{x})$ the inverse G_A^{-1} of the mapping G_A given by*

$$G_A : x \mapsto f(\bar{x}) + A(x - \bar{x}) + F(x)$$

has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$. Then the mapping $(f+F)^{-1}$ has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$ as well.

When F is the zero mapping, from 4D.4 we obtain Clarke's theorem 4D.3. If f is strictly differentiable at $\bar{x}$, 4D.4 reduces to the inverse function version of 2B.10. We supply Theorem 4D.4 with a proof in Section 6F, where, after a preliminary analysis, we employ an iteration resembling Newton's method, in analogy to Proof I of Theorem 1A.1.

Clarke's theorem is a particular case of an inverse function theorem due to Kummer, which relies on the strict graphical derivative. For a function, the definition of strict graphical derivative can be restated as follows:

Strict graphical derivative for a function. *For a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and any point $\bar{x} \in \operatorname{dom} f$, the strict graphical derivative at $\bar{x}$ is the set-valued mapping $D_* f(\bar{x}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined by*

$$D_*f(\bar{x})(u) = \left\{ w \,\middle|\, w = \lim_{k\to\infty} \frac{f(x^k + \tau^k u^k) - f(x^k)}{\tau^k} \text{ for some } (x^k, u^k) \to (\bar{x}, u), \tau^k \searrow 0 \right\}.$$

When $\operatorname{lip}(f;\bar{x}) < \infty$, the set $D_*f(\bar{x})(u)$ is nonempty, closed and bounded in $I\!R^m$ for each $u \in I\!R^n$. Then too, the definition of $D_*f(\bar{x})(u)$ can be simplified by taking $u_k \equiv u$. In this Lipschitzian setting the strict graphical derivative can be expressed in terms of the generalized Jacobian. Specifically, it can be shown that $D_*f(\bar{x})(u) = \{ Au \,|\, A \in \bar{\partial} f(\bar{x}) \}$ for all u if $m = 1$, but that fails for higher dimensions. In general, it is known[3] that for a function $f : I\!R^n \to I\!R^m$ with $\operatorname{lip}(f;\bar{x}) < \infty$, one has

$$\text{(9)} \qquad \bar{\partial} f(\bar{x})(u) \supset D_*f(\bar{x})(u) \quad \text{for all} \quad u \in I\!R^n.$$

Note that f is strictly graphically differentiable at $\bar{x}$ if and only if $D_*f(\bar{x})$ is a linear mapping, with the matrix for that mapping then being $\nabla f(\bar{x})$. Anyway, the strict derivatives can be used without having to assume even that $\operatorname{lip}(f;\bar{x}) < \infty$.

The computation of the strict graphical derivatives will be illustrated now in a special case of nonsmoothness which has a basic role in various situations.

Example 4D.5. Consider the function

$$\theta^+ : x \mapsto x^+ := \max\{x, 0\}, \quad x \in I\!R.$$

Directly from the definition, we have, for any real u, that

$$D_*\theta^+(\bar{x})(u) = \begin{cases} u & \text{for } \bar{x} > 0, \\ \{ \lambda u \,|\, \lambda \in [0,1] \} & \text{for } \bar{x} = 0, \\ 0 & \text{for } \bar{x} < 0 \end{cases}$$

and is the same as $\bar{\partial}\theta^+(\bar{x})(u)$. Similarly, the function

$$\theta^- : x \mapsto x^- := \min\{x, 0\}, \quad x \in I\!R$$

satisfies

$$\theta^-(x) = x - \theta^+(x).$$

Then, just by applying the definition, we get

$$v \in D_*\theta^+(\bar{x})(u) \iff (u - v) \in D_*\theta^-(\bar{x})(u) \text{ for every real } u.$$

We will now present an inverse function theorem, due to Kummer, which furnishes a complete characterization of the existence of a Lipschitz continuous localization of the inverse, and thus sharpens the theorem of Clarke 4D.3.

Theorem 4D.6 (Kummer inverse function theorem)**.** *Let $f : I\!R^n \to I\!R^n$ be continuous around $\bar{x}$, with $f(\bar{x}) = \bar{y}$. Then f^{-1} has a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$ if and only if*

---
[3] Cf. Klatte and Kummer [2002], Section 6.3.

$$(10) \qquad 0 \in D_*f(\bar{x})(u) \quad \Longrightarrow \quad u = 0.$$

Proof. Recall Theorem 1F.2, which says that for a function $f : \mathbb{R}^n \to \mathbb{R}^n$ that is continuous around $\bar{x}$, the inverse f^{-1} has a Lipschitz continuous localization around $f(\bar{x})$ for $\bar{x}$ if and only if, in some neighborhood U of $\bar{x}$, there is a constant $c > 0$ such that

$$(11) \qquad c|x' - x| \leq |f(x') - f(x)| \quad \text{for all} \quad x', x \in U.$$

We will show first that (10) implies (11), from which the sufficiency of the condition will follow. With the aim of reaching a contradiction, let us assume there are sequences $c^k \searrow 0$, $x^k \to \bar{x}$ and $\tilde{x}^k \to \bar{x}$, $x^k \neq \tilde{x}^k$ such that

$$|f(x^k) - f(\tilde{x}^k)| < c^k |x^k - \tilde{x}^k|.$$

Then the sequence of points

$$u^k := \frac{\tilde{x}^k - x^k}{|x^k - \tilde{x}^k|}$$

satisfies $|u^k| = 1$ for all k, hence a subsequence u^{k_i} of it is convergent to some $u \neq 0$. Restricting ourselves to such a subsequence, we obtain for $t^{k_i} = |x^{k_i} - \tilde{x}^{k_i}|$ that

$$\lim_{k_i \to \infty} \frac{f(x^{k_i} + t^{k_i} u^{k_i}) - f(x^{k_i})}{t^{k_i}} = 0.$$

By definition, the limit on the left side belongs to $D_*f(\bar{x})(u)$, yet $u \neq 0$, which is contrary to (10). Hence (10) does imply (11).

For the converse, we argue that if (10) were violated, there would be sequences $\tau^k \searrow 0$, $x^k \to \bar{x}$, and $u^k \to u$ with $u \neq 0$, for which

$$(12) \qquad \lim_{k \to \infty} \frac{f(x^k + \tau^k u^k) - f(x^k)}{\tau^k} = 0.$$

On the other hand, under (11) however, one has

$$\frac{|f(x^k + \tau^k u^k) - f(x^k)|}{\tau^k} \geq c|u^k|,$$

which combined with (12) and the assumption that u^k is away from 0 leads to an absurdity for large k. Thus (11) guarantees that (10) holds. □

The property recorded in (9) indicates clearly that Clarke's inverse function theorem follows from that of Kummer. In the last section 4I of this chapter we apply Kummer's theorem to the nonlinear programming problem we studied in 2G.

4E. Derivative Criterion for Strong Metric Subregularity

Strong metric subregularity of a mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ at $\bar{x}$ for $\bar{y}$, where $\bar{y} \in F(\bar{x})$, was defined in Section 3I to mean the existence of a constant κ along with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$(1) \qquad |x - \bar{x}| \leq \kappa d(\bar{y}, F(x) \cap V) \quad \text{for all } x \in U.$$

This property is equivalent to the combination of two other properties: that F is *metrically subregular* at $\bar{x}$ for $\bar{y}$, and $\bar{x}$ is an *isolated point* of $F^{-1}(\bar{y})$. The associated modulus, the infimum of all $\kappa > 0$ for which this holds for some U and V, is thus the same as the modulus of subregularity, $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$.

It was shown in 3I.2 that F is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ if and only if F^{-1} has the isolated calmness property at $\bar{y}$ for $\bar{x}$. As an illustration, a linear mapping A is strongly metrically subregular at $\bar{x}$ for $\bar{y} = A\bar{x}$ if and only if $A^{-1}(\bar{y})$ consists only of $\bar{x}$, i.e., A is injective. More generally, a mapping F that is polyhedral, as defined in 3D, is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ if and only if $\bar{x}$ is an isolated point of $F^{-1}(\bar{y})$; this follows from 3I.1.

What makes the strong metric subregularity attractive, along the same lines as metric regularity and strong metric regularity, is its stability with respect to approximation as established in 3I.7. In particular, a function $f : \mathbb{R}^n \to \mathbb{R}^m$ which is differentiable at $\bar{x}$ is strongly metrically subregular at $\bar{x}$ for $f(\bar{x})$ if and only if its Jacobian $\nabla f(\bar{x})$ has rank n, so that $\nabla f(\bar{x})u = 0$ implies $u = 0$. According to 4A.6, this is also characterized by $|Df(\bar{x})^{-1}|^{+} < \infty$. It turns out that such an outer norm characterization can be provided also for set-valued mappings by letting graphical derivatives take over the role of ordinary derivatives.

Theorem 4E.1 (graphical derivative criterion for strong metric subregularity). *A mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ whose graph is locally closed at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ if and only if*

$$(2) \qquad DF(\bar{x}\,|\,\bar{y})^{-1}(0) = \{0\},$$

this being equivalent to

$$(3) \qquad |DF(\bar{x}\,|\,\bar{y})^{-1}|^{+} < \infty,$$

and in that case

$$(4) \qquad \operatorname{subreg}(F;\bar{x}\,|\,\bar{y}) = |DF(\bar{x}\,|\,\bar{y})^{-1}|^{+}.$$

Proof. The equivalence between (2) and (3) comes from 4A.6. To get the equivalence of these conditions with strong metric regularity, suppose first that $\kappa > \operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$ so that F is strongly metrically subregular at $\bar{x}$ for $\bar{y}$ and (1) holds for some neighborhoods U and V. By definition, having $v \in DF(\bar{x}\,|\,\bar{y})(u)$ refers to the existence of sequences $u^k \to u$, $v^k \to v$ and $\tau^k \searrow 0$ such that $\bar{y} + \tau^k v^k \in F(\bar{x} + \tau^k u^k)$. Then $\bar{x} + \tau^k u^k \in U$ and $\bar{y} + \tau^k v^k \in V$ eventually, so that (1) yields $|(\bar{x} + \tau^k u^k) - \bar{x}| \leq$

$\kappa|(\bar{y}+\tau^k v^k)-\bar{y}|$, which is the same as $|u^k| \leq \kappa|v^k|$. In the limit, this implies $|u| \leq \kappa|v|$. But then, by 4A.6, $|DF(\bar{x}\,|\,\bar{y})^{-1}|^{+} \leq \kappa$ and hence

$$\text{subreg}\,(F;\bar{x}\,|\,\bar{y}) \geq |DF(\bar{x}\,|\,\bar{y})^{-1}|^{+}. \tag{5}$$

In the other direction, (3) implies the existence of a $\kappa > 0$ such that

$$\sup_{v\in B}\ \sup_{u\in DF(\bar{x}\,|\,\bar{y})^{-1}(v)} |u| < \kappa.$$

This in turn implies that $|x-\bar{x}| \leq \kappa|y-\bar{y}|$ for all $(x,y) \in \text{gph}\, F$ close to $(\bar{x},\bar{y})$. That description fits with (1). Further, κ can be chosen arbitrarily close to $|DF(\bar{x}\,|\,\bar{y})^{-1}|^{+}$, and therefore $|DF(\bar{x}\,|\,\bar{y})^{-1}|^{+} \geq \text{subreg}\,(F;\bar{x}\,|\,\bar{y})$. This, combined with (5), finishes the argument. □

Corollary 4E.2 (graphical derivative criterion for isolated calmness)**.** *For a mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ and a point $(\bar{y},\bar{x}) \in \text{gph}\, S$ at which $\text{gph}\, S$ is locally closed. one has*

$$\text{clm}\,(S;\bar{y}\,|\,\bar{x}) = |DS(\bar{y}\,|\,\bar{x})|^{+}.$$

Theorem 4E.1 immediately gives us the linearization result in Corollary 3I.11 by using the sum rule in 4A.2. Implicit function theorems could be developed for the isolated calmness of solution mappings to general inclusions $G(p,x) \ni 0$ in parallel to the results in 4B, but we shall not do this here. In the following corollary we utilize Theorem 3I.14 and the ample parameterization condition.

Corollary 4E.3 (derivative rule for isolated calmness of solution mappings)**.** *For the solution mapping of the generalized equation,*

$$S(p) = \left\{\, x \,\middle|\, f(p,x) + F(x) \ni 0 \right\}$$

with $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a pair $(\bar{p},\bar{x})$ satisfying $\bar{x} \in S(\bar{p})$, suppose that f is differentiable with respect to x uniformly in p at $(\bar{p},\bar{x})$ and also differentiable with respect to p uniformly in x at $(\bar{p},\bar{x})$. Also, suppose that $\text{gph}\, F$ is locally closed at $(\bar{x},-f(\bar{p},\bar{x}))$. If

$$|(D_x f(\bar{p},\bar{x}) + DF(\bar{x}\,|\,-f(\bar{p},\bar{x})))^{-1}|^{+} \leq \lambda < \infty, \tag{6}$$

then S has the isolated calmness property at $\bar{p}$ for $\bar{x}$, moreover with

$$\text{clm}\,(S;\bar{p}\,|\,\bar{x}) \leq \lambda\,|\nabla_p f(\bar{p},\bar{x})|.$$

Furthermore, when f is continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$ and satisfies the ample parameterization condition

$$\text{rank}\ \nabla_p f(\bar{p},\bar{x}) = m,$$

then the converse implication holds as well; that is, S has isolated calmness property at $\bar{p}$ for $\bar{x}$ if and only if (6) is satisfied.

The following simple example in terms of graphical derivatives illustrates further the distinction between metric regularity and strong metric subregularity.

Example 4E.4 (strong metric subregularity without metric regularity)**.** Let $F : \mathbb{R} \rightrightarrows \mathbb{R}$ be defined by

$$F(x) = \begin{cases} [\sqrt{1-(x-1)^2}, \infty) & \text{for } 0 \le x \le 1, \\ [\sqrt{1-(x+1)^2}, \infty) & \text{for } -1 \le x \le 0, \\ \emptyset & \text{elsewhere,} \end{cases}$$

as shown in Figure 4.1. Then

$$DF(0\,|\,0)(u) = \begin{cases} [0, \infty) & \text{for } u = 0, \\ \emptyset & \text{for } u \neq 0, \end{cases}$$

and therefore

$$|DF(0\,|\,0)^{-1}|^{+} = 0, \qquad |DF(0\,|\,0)^{-1}|^{-} = \infty.$$

This fits with F being strongly metrically subregular, but not metrically regular at 0 for 0.

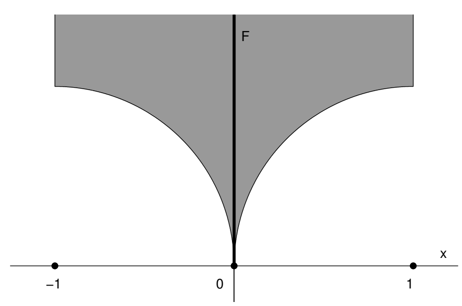

Fig. 4.1 Graph of the mapping in Example 4E.4.

4F. Applications to Parameterized Constraint Systems

Next we look at what the graphical derivative results given in Section 4B have to say about a constraint system

(1) $$f(p,x) - D \ni 0, \quad \text{or equivalently} \quad f(p,x) \in D,$$

and its solution mapping

(2) $$S : p \mapsto \left\{ x \,\middle|\, f(p,x) \in D \right\}$$

for a function $f : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^m$ and a set $D \subset \mathbb{R}^m$.

Theorem 4F.1 (implicit mapping theorem for a constraint system)**.** *Let $\bar{x} \in S(\bar{p})$ for solution mapping S of the constraint system (1) and suppose that f is continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$ and that the set D is closed. If*

(3) $$\limsup_{\substack{(p,x,y)\to(\bar{p},\bar{x},0)\\ f(p,x)-y\in D}} \sup_{|v|\leq 1} d\Big(0, D_x f(p,x)^{-1}(v + T_D(f(p,x)-y))\Big) \leq \lambda < \infty,$$

then S has the Aubin property at $\bar{p}$ for $\bar{x}$, with

(4) $$\operatorname{lip}(S;\bar{p},\bar{x}) \leq \lambda \, \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})).$$

Proof. The assumed closedness of D and continuous differentiability of f around $(\bar{p},\bar{x})$ allow us to apply Corollary 4B.8 to the case of $F(x) \equiv -D$. Further, according to 4A.3 we have

$$D_x G(p,x\,|\,y) = D_x f(p,x) - T_D(f(p,x)-y).$$

Next, we use the definition of inner norm in 4A(8) to write 4B(29) as (3) and apply 4B.7 to obtain that S has the Aubin property at $\bar{p}$ for $\bar{x}$. The estimate (4) follows immediately from 4B(30). □

A much sharper result can be obtained when f is continuously differentiable and the set D in the system (1) is polyhedral convex.

Theorem 4F.2 (constraint systems with polyhedral convexity)**.** *Let $\bar{x} \in S(\bar{p})$ for the solution mapping S of the constraint system (1) in the case of a polyhedral convex set D. Suppose that f is continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$. Then for S to have the Aubin property at $\bar{x}$ for $\bar{p}$, it is sufficient that*

(5) $$\operatorname{rge} \nabla_x f(\bar{p},\bar{x}) - T_D(f(\bar{p},\bar{x})) = \mathbb{R}^m,$$

in which case the corresponding modulus satisfies $\operatorname{lip}(S;\bar{p}\,|\,\bar{x}) \leq \lambda\, |\nabla_p f(\bar{p},\bar{x})|$ for

(6) $$\lambda = \sup_{|v|\leq 1} d\Big(0, \nabla_x f(\bar{p},\bar{x})^{-1}(v + T_D(f(\bar{p},\bar{x})))\Big).$$

Moreover (5) is necessary for S to have the Aubin property at $\bar{p}$ for $\bar{x}$ under the ample parameterization condition

$$\text{rank}\, \nabla_p f(\bar{p},\bar{x}) = m. \tag{7}$$

Proof. We invoke Theorem 4F.1 but make special use of the fact that D is polyhedral. That property implies that $T_D(w) \supset T_D(\bar{w})$ for all w sufficiently near to $\bar{w}$, as seen in 2E.3; we apply this to $w = f(p,x) - y$ and $\bar{w} = f(\bar{p},\bar{x})$ in the formulas (3) and (4) of 4F.1. The distances in question are greatest when the cone is as small as possible; this, combined with the continuous differentiability of f, allows us to drop the limit in (3). Further, from the equivalence relation 4A(16) in Corollary 4A.7, we obtain that the finiteness of λ in (6) is equivalent to (5).

For the necessity, we bring in a further argument which makes use of the ample parameterization condition (7). According to Theorem 3F.9, under (7) the Aubin property of S at $\bar{p}$ for $\bar{x}$ implies metric regularity of the linearized mapping $h - D$ for $h(x) = f(\bar{p},\bar{x}) + D_x f(\bar{p},\bar{x})(x - \bar{x})$. The derivative criterion for metric regularity 4B.1 tells us then that

$$\begin{aligned} \limsup_{\substack{(x,y)\to(\bar{x},0)\\ f(\bar{p},\bar{x})+D_x f(\bar{p},\bar{x})(x-\bar{x})-y\in D}} \sup_{|v|\le 1} d\Big(0,& \\ \nabla_x f(\bar{p},\bar{x})^{-1}(v + T_D(f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x}) - y))\Big) &< \infty. \end{aligned} \tag{8}$$

Taking $x = \bar{x}$ and $y = 0$ instead of limsup in (8) gives us the expression for λ in (6) and may only decrease the left side of this inequality. We already know that the finiteness of λ in (6) yields (5), and so we are done. $\square$

Example 4F.3 (application to systems of inequalities and equalities)**.** For $D = \mathbb{R}^s_- \times \{0\}^{m-s}$, the solution mapping $S(p)$ in (2) consists, in terms of $f(p,x) = (f_1(p,x),\dots,f_m(p,x))$ of all solutions x to

$$f_i(p,x) \begin{cases} \le 0 & \text{for } i \in [1,s], \\ = 0 & \text{for } i \in [s+1,m]. \end{cases}$$

Let $\bar{x}$ solve this for $\bar{p}$ and let each f_i be continuously differentiable around $(\bar{p},\bar{x})$. Then a sufficient condition for S to have the Aubin property for $\bar{p}$ for $\bar{x}$ is the *Mangasarian–Fromovitz condition*:

$$\exists\, w \in \mathbb{R}^n \;\; \text{with} \;\; \begin{cases} \nabla_x f_i(\bar{p},\bar{x})w < 0 & \text{for } i \in [1,s] \text{ with } f_i(\bar{p},\bar{x}) = 0, \\ \nabla_x f_i(\bar{p},\bar{x})w = 0 & \text{for } i \in [s+1,m], \end{cases} \tag{9}$$

and

$$\text{the vectors } \nabla_x f_i(\bar{p},\bar{x}) \text{ for } i \in [s+1,m] \text{ are linearly independent.} \tag{10}$$

Moreover, the combination of (9) and (10) is also necessary for S to have the Aubin property under the ample parameterization condition (7). In particular, when f is independent of p and then $0 \in f(\bar{x}) - D$, the Mangasarian–Fromovitz condition (9)–

(10) is a necessary and sufficient condition for metric regularity of the mapping $f-D$ at $\bar{x}$ for 0.

Detail. According to 4F.2, it is enough to show that (5) is equivalent to the combination of (9) and (10) in the case of $D = \mathbb{R}^s_- \times \{0\}^{m-s}$. Observe that the tangent cone to the set D at $f(\bar{p},\bar{x})$ has the following form:

$$(11)\qquad v \in T_D(f(\bar{p},\bar{x})) \iff v_i \begin{cases} \leq 0 & \text{for } i \in [1,s] \text{ with } f_i(\bar{p},\bar{x}) = 0, \\ = 0 & \text{for } i \in [s+1,m]. \end{cases}$$

Let (5) hold. Then, using (11), we obtain that the matrix with rows the vectors $\nabla_x f_{s+1}(\bar{p},\bar{x}),\ldots,\nabla_x f_m(\bar{p},\bar{x})$ must be of full rank, hence (10) holds. If (9) is violated, then for every $w \in \mathbb{R}^n$ either $\nabla_x f_i(\bar{p},\bar{x})w \geq 0$ for some $i \in [1,s]$ with $f_i(\bar{p},\bar{x}) = 0$, or $\nabla_x f_i(\bar{p},\bar{x})w \neq 0$ for some $i \in [s+1,m]$, which contradicts (5) in an obvious way.

The combination of (9) and (10) implies that for every $y \in \mathbb{R}^m$ there exist $w, v \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$ with $z_i \leq 0$ for $i \in [1,s]$ with $f_i(\bar{p},\bar{x}) = 0$ such that

$$\begin{cases} \nabla_x f_i(\bar{p},\bar{x})w - z_i = y_i & \text{for } i \in [1,s] \text{ with } f_i(\bar{p},\bar{x}) = 0, \\ \nabla_x f_i(\bar{p},\bar{x})(w+v) = y_i & \text{for } i \in [s+1,m]. \end{cases}$$

But then (5) follows directly from the form (11) of the tangent cone.

If f is independent of p, by 3E.7 the metric regularity of $-f+D$ is equivalent to the Aubin property of the inverse $(-f+D)^{-1}$, which is the same as the solution mapping

$$S(p) = \{\, x \mid p + f(x) \in D \,\}$$

for which the ample parameterization condition (7) holds automatically. Then, from 4F.2, for $\bar{x} \in S(\bar{p})$, the Aubin property of S at $\bar{p}$ for $\bar{x}$ and hence metric regularity of $f-D$ at $\bar{x}$ for $\bar{p}$ is equivalent to (5) and therefore to (9)–(10). □

Exercise 4F.4. *Consider the constraint system in 4F.3 with $f(p,x) = g(x) - p$, $\bar{p} = 0$ and g continuously differentiable near $\bar{x}$. Show that the existence of a Lipschitz continuous local selection of the solution mapping S at 0 for $\bar{x}$ implies the Mangasarian–Fromovitz condition. In other words, the existence of a Lipschitz continuous local selection of S at 0 for $\bar{x}$ implies metric regularity of the mapping $g-D$ at $\bar{x}$ for 0.*

Guide. Utilizing 2B.12, from the existence of a local selection of S at 0 for $\bar{x}$ we obtain that the inverse F_0^{-1} of the linearization $F_0(x) := g(\bar{x}) + Dg(\bar{x})(x-\bar{x}) - D$ has a Lipschitz continuous local selection at 0 for $\bar{x}$. Then, in particular, for every $v \in \mathbb{R}^m$ there exists $w \in \mathbb{R}^n$ such that

$$\begin{cases} \nabla g_i(\bar{x})w \leq v_i & \text{for } i \in [1,s] \text{ with } g_i(\bar{x}) = 0, \\ \nabla g_i(\bar{x})w = v_i & \text{for } i \in [s+1,m]. \end{cases}$$

This is the same as (5). □

We will present next an application of the coderivative criterion for metric regularity to the constraint system (1).

Theorem 4F.5 (coderivative criterion for constraint systems). *Let $\bar{x} \in S(\bar{p})$ for the solution mapping S of the constraint system (1) and suppose that f is continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$ and the set D is closed. Then for S to have the Aubin property at $\bar{x}$ for $\bar{p}$, it is sufficient that*

$$y \in N_D(f(\bar{p},\bar{x})) \;\text{ and }\; \nabla_x f(\bar{p},\bar{x})^{\mathsf{T}} y = 0 \quad \Longrightarrow \quad y = 0. \tag{12}$$

in which case the corresponding Lipschitz modulus satisfies

$$\operatorname{lip}(S;\bar{p}\,|\,\bar{x}) \le \lambda\, |\nabla_p f(\bar{p},\bar{x})|$$

for

$$\lambda = \sup_{\substack{|y|\le 1\\ y\in N_D(f(\bar{p},\bar{x}))}} |\nabla_x f(\bar{p},\bar{x})^{\mathsf{T}} y|^{-1}. \tag{13}$$

Moreover (13) is necessary for S to have the Aubin property at $\bar{p}$ for $\bar{x}$ under the ample parameterization condition (7).

Proof. The solution mapping S has the Aubin property at $\bar{x}$ for 0 if and only if the linearized mapping

$$L : x \mapsto f(\bar{p},\bar{x}) + \nabla_x f(\bar{p},\bar{x})(x-\bar{x}) - D$$

is metrically regular at $\bar{x}$ for 0. The general normal cone to the graph of L has the form[4]

$$N_{\operatorname{gph} L}(\bar{x},0) = \{(v,-y) \mid y \in N_D(f(\bar{p},\bar{x})),\; v = \nabla_x f(\bar{p},\bar{x})^{\mathsf{T}} y\}.$$

It remains to use Theorem 4C.2 and the definition of the outer norm. □

A result parallel to 4F.2 can be formulated also for isolated calmness instead of the Aubin property.

Proposition 4F.6 (isolated calmness of constraint systems). *In the setting of Theorem 4F.1, for S to have the isolated calmness property at $\bar{p}$ for $\bar{x}$ it is sufficient that*

$$D_x f(\bar{p},\bar{x})u \in T_D(f(\bar{p},\bar{x})) \quad \Longrightarrow \quad u = 0.$$

Moreover, this condition is necessary for S to have the isolated calmness property at $\bar{p}$ for $\bar{x}$ under the ample parameterization condition (8).

Proof. This is a special case of Corollary 4F.3 in which we utilize 4A.2. □

We note that the isolated calmness property offers little of interest in the case of solution mappings for constraint systems beyond equations, inasmuch as it necessitates $\bar{x}$ being an isolated point of the solution set $S(\bar{p})$; this restricts significantly the class of constraint systems for which such a property may occur. In the following

---

[4] This is 9.44 in Rockafellar and Wets [1998].

section we will consider mappings associated with variational inequalities for which the isolated calmness is a more natural property.

4G. Isolated Calmness for Variational Inequalities

Now we take up once more the topic of variational inequalities, to which serious attention was already devoted in Chapter 2. This revolves around a generalized equation of the form

$$(1) \qquad f(p,x)+N_C(x)\ni 0, \quad \text{or} \quad -f(p,x)\in N_C(x),$$

for a function $f:\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^n$ and the normal cone mapping N_C associated with a nonempty, closed, convex set $C\subset\mathbb{R}^n$, and the solution mapping $S:\mathbb{R}^d\rightrightarrows\mathbb{R}^n$ defined by

$$(2) \qquad S(p)=\left\{\, x \,\middle|\, f(p,x)+N_C(x)\ni 0 \,\right\}.$$

Especially strong results were obtained in 2E for the case in which C is a *polyhedral* convex set, and that will also persist here. Of special importance in that setting is the critical cone associated with C at a point x with respect to a vector $v\in N_C(x)$, defined by

$$(3) \qquad K_C(x,v)\;=\;T_C(x)\cap[v]^{\perp},$$

which is always polyhedral convex as well. Recall here that for any vector $v\in\mathbb{R}^n$ we denote $[v]=\left\{\,\tau v \,\middle|\, \tau\in\mathbb{R}\,\right\}$; then $[v]$ is a subspace of dimension 1 if $v\neq 0$ and 0 if $v=0$. Accordingly, $[v]^{\perp}$ is a hyperplane through the origin if $v\neq 0$ and the whole $\mathbb{R}^n$ when $v=0$.

In this section we examine the isolated calmness property, which is inverse to strong metric subregularity by 3I.3.

Theorem 4G.1 (isolated calmness for variational inequalities)**.** *For the variational inequality (1) and its solution mapping (2) under the assumption that the convex set C is polyhedral, let $\bar{x}\in S(\bar{p})$ and suppose that f is continuously differentiable around $(\bar{p},\bar{x})$. Let $A=\nabla_x f(\bar{p},\bar{x})$ and let $K=K_C(\bar{x},\bar{v})$ be the corresponding critical cone in (3) for $\bar{v}=-f(\bar{p},\bar{x})$. If*

$$(4) \qquad (A+N_K)^{-1}(0)=\{0\},$$

then the solution mapping S has the isolated calmness property at $\bar{p}$ for $\bar{x}$ with

$$\operatorname{clm}(S;\bar{p}\,|\,\bar{x})\;\le\;|(A+N_K)^{-1}|^{+}\cdot|\nabla_p f(\bar{p},\bar{x})|.$$

Moreover, under the ample parameterization condition $\operatorname{rank} \nabla_p f(\bar{p},\bar{x}) = n$, *the property in (4) is not just sufficient but also necessary for S to have the isolated calmness property at $\bar{p}$ for $\bar{x}$.*

Proof. Utilizing the specific form of the graphical derivative established in 4A.4 and the equivalence relation 4A(13) in 4A.6, we see that (4) is equivalent to the condition 4E(7) in Corollary 4E.3. Everything then follows from the claim of that corollary. □

Exercise 4G.2 (alternative cone condition)**.** *In terms of the cone K^* that is polar to K, show that the condition in (4) is equivalent to*

(5) $$w \in K,\ -Aw \in K^*,\ w \perp Aw \quad \Longrightarrow \quad w = 0.$$

Guide. Make use of 2A.3. □

In the important case when $C = \mathbb{R}^n_+$, the variational inequality (1) turns into the *complementarity* relation

(6) $$x \geq 0, \quad f(p,x) \geq 0, \quad x \perp f(p,x).$$

This will serve to illustrate the result in Theorem 4G.1. Using the notation introduced in Section 2E for the analysis of a complementarity problem, we associate with the reference point $(\bar{x},\bar{v}) \in \operatorname{gph} N_{\mathbb{R}^n_+}$ the index sets J_1, J_2 and J_3 in $\{1,\ldots,n\}$ given by

$$J_1 = \left\{\, j \,\middle|\, \bar{x}_j > 0, \bar{v}_j = 0 \right\}, \quad J_2 = \left\{\, j \,\middle|\, \bar{x}_j = 0, \bar{v}_j = 0 \right\}, \quad J_3 = \left\{\, j \,\middle|\, \bar{x}_j = 0, \bar{v}_j < 0 \right\}.$$

Then, by 2E.5, the critical cone $K = K_C(\bar{x},\bar{v}) = T_{\mathbb{R}^n_+}(\bar{x}) \cap [f(\bar{p},\bar{x})]^\perp$ is described by

(7) $$w \in K \iff \begin{cases} w_j \text{ free} & \text{for } i \in J_1, \\ w_j \geq 0 & \text{for } i \in J_2, \\ w_j = 0 & \text{for } i \in J_3. \end{cases}$$

Example 4G.3 (isolated calmness for complementarity problems)**.** In the case of $C = \mathbb{R}^n_+$ in which the variational inequality (1) reduces to the complementarity relation (6) and the critical cone K is given by (7), the condition (4) in Theorem 4G.1 reduces through (5) to having the following hold for the entries a_{ij} of the matrix A. If w_j for $j \in J_1 \cup J_2$ are real numbers satisfying

$$w_j \geq 0 \text{ for } j \in J_2 \quad \text{and} \quad \sum_{j \in J_1 \cup J_2} a_{ij} w_j \begin{cases} = 0 & \text{for } i \in J_1 \text{ and for } i \in J_2 \text{ with } w_i > 0, \\ \geq 0 & \text{for } i \in J_2 \text{ with } w_i = 0, \end{cases}$$

then $w_j = 0$ for all $j \in J_1 \cup J_2$.

In the particular case when $J_1 = J_3 = \emptyset$, the matrices satisfying the condition in 4G.3 are called R_0-matrices[5].

As another application of Theorem 4G.1, consider the tilted minimization problem from Section 2G:

$$\text{(8)} \qquad \text{minimize } g(x) - \langle v, x \rangle \text{ over } x \in C,$$

where C is a nonempty *polyhedral* convex subset of $\mathbb{R}^n$, $v \in \mathbb{R}^n$ is a parameter, and the function $g : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable everywhere. We first give a brief summary of the optimality conditions from 2G.

If x is a local optimal solution of (8) for v then x satisfies the basic first-order necessary optimality condition

$$\text{(9)} \qquad \nabla g(x) + N_C(x) \ni v.$$

Any solution x of (9) is a stationary point for problem (8), denoted $S(v)$, and the associated stationary point mapping is $v \mapsto S(v) = (Dg + N_C)^{-1}(v)$. The set of local minimizers of (8) for v is a subset of $S(v)$. If the function g is convex, every stationary point is not only local but also a global minimizer. For the variational inequality (9), the critical cone to C associated with a solution x for v has the form

$$K_C(x, v - \nabla g(x)) = T_C(x) \cap [v - \nabla g(x)]^{\perp}.$$

If x furnishes a local minimum of (8) for v, then, according to 2G.1(a), x must satisfy the second-order necessary condition

$$\text{(10)} \qquad \langle u, \nabla^2 g(x) u \rangle \geq 0 \quad \text{for all } u \in K_C(x, v - \nabla g(x)).$$

In addition, from 2G.1(b), when $x \in S(v)$ satisfies the second-order sufficient condition

$$\text{(11)} \qquad \langle u, \nabla^2 g(x) u \rangle > 0 \quad \text{for all nonzero } u \in K_C(x, v - \nabla g(x)),$$

then x is a local optimal solution of (8) for v. Having x to satisfy (9) and (11) is equivalent to the existence of $\varepsilon > 0$ and $\delta > 0$ such that

$$\text{(12)} \qquad g(y) - \langle v, y \rangle \geq g(x) - \langle v, x \rangle + \varepsilon |y - x|^2 \text{ for all } y \in C \text{ with } |y - x| \leq \delta,$$

meaning by definition that x furnishes a strong local minimum in (8).

We know from 2G.3 that the stationary point mapping S has a Lipschitz localization s around v for x with the property that $s(u)$ furnishes a strong local minimum for u near v if and only if the following stronger form of the second-order sufficient optimality condition holds:

$$\langle w, \nabla^2 g(\bar{x}) w \rangle > 0 \text{ for all nonzero } w \in K_C^+(x, v),$$

[5] For a detailed description of the classes of matrices appearing in the theory of linear complementarity problems, see the book Cottle, Pang and Stone [1992].

where $K_C^+(x,v) = K_C(x, v-\nabla g(x)) - K_C(x, v-\nabla g(x))$ is the critical subspace associated with x and v. We now complement this result with a necessary and sufficient condition for isolated calmness of S combined with local optimality at the reference point.

Theorem 4G.4 (role of second-order sufficiency)**.** *Consider the stationary point mapping S for problem (8), that is, the solution mapping for (9), and let $\bar{x} \in S(\bar{v})$. Then the following are equivalent:*

(a) *the second-order sufficient condition (11) holds at $\bar{x}$ for $\bar{v}$;*

(b) *the point $\bar{x}$ is a local minimizer of (8) for $\bar{v}$ and the mapping S has the isolated calmness property at $\bar{v}$ for $\bar{x}$.*

Moreover, in either case, $\bar{x}$ is actually a strong local minimizer of (7) for $\bar{v}$.

Proof. Let $A := \nabla^2 g(\bar{x})$. According to Theorem 4G.1 complemented with 4G.2, the mapping S has the isolated calmness property at $\bar{v}$ for $\bar{x}$ if and only if

$$u \in K, \quad -Au \in K^*, \quad u \perp Au \quad \Longrightarrow \quad u = 0, \tag{13}$$

where $K = K_C(\bar{x}, \bar{v} - \nabla g(\bar{x}))$. Let (a) hold. Then of course $\bar{x}$ is a local optimal solution as described. If (b) doesn't hold, there must exist some $u \neq 0$ satisfying the conditions in the left side of (13), and that would contradict the inequality $\langle u, Au\rangle > 0$ in (11).

Conversely, assume that (b) is satisfied. Then the second-order necessary condition (10) must hold; this can be written as

$$u \in K \quad \Longrightarrow \quad -Au \in K^*.$$

The isolated calmness property of S at $\bar{v}$ for $\bar{x}$ is identified with (13), which in turn eliminates the possibility of there being a nonzero $u \in K$ such that the inequality in (10) fails to be strict. Thus, the necessary condition (10) turns into the sufficient condition (11). We already know that (11) implies (12), so the proof is complete. $\square$

4H. Variational Inequalities over Polyhedral Convex Sets

In this section we investigate applications of graphical derivatives and coderivatives to characterize (strong) metric regularity of the following mapping:

$$x \mapsto f(x) + N_C(x) \subset \mathbb{R}^n \quad \text{for } x \in \mathbb{R}^n, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is a function and N_C is the normal cone mapping associated with a *polyhedral* convex set $C \subset \mathbb{R}^n$. Thus, the inclusion $0 \in (f + N_C)(x)$ is the standard variational inequality which we introduced in Section 2A. Central in this section is the following theorem:

Theorem 4H.1 (characterization of (strong) metric regularity). *For the mapping in (1) under the assumption that the convex set C is polyhedral, let $\bar{x} \in (f+N_C)^{-1}(0)$ and suppose that f is continuously differentiable near $\bar{x}$. Let*

$$A = D_x f(\bar{x}) \quad \text{and} \quad K = \left\{ w \in T_C(\bar{x}) \,\middle|\, w \perp f(\bar{x}) \right\}, \tag{2}$$

that is, K is the critical cone to the set C at $\bar{x}$ for $-f(\bar{x})$. Then the mapping $f+N_C$ is metrically regular at $\bar{x}$ for 0 if and only if the mapping $A+N_K$ is metrically regular at 0 for 0, in which case

$$\operatorname{reg}(f+N_C;\bar{x}\,|\,0) = \operatorname{reg}(A+N_K;0\,|\,0). \tag{3}$$

Furthermore, metric regularity of $A+N_K$ at 0 for 0 implies, and hence is equivalent to, strong metric regularity of this mapping at 0 for 0, which is actually equivalent to the property that

$$(A+N_K)^{-1} \quad \text{is everywhere single-valued,} \tag{4}$$

in which case $(A+N_K)^{-1}$ is a function which is Lipschitz continuous globally with Lipschitz modulus equal to $\operatorname{reg}(A+N_K;0\,|\,0)$. Thus, metric regularity of $f+N_C$ at $\bar{x}$ for 0 is equivalent to strong metric regularity of $f+N_C$ at $\bar{x}$ for 0.

The statement involving (3) comes from the combination of 3F.7, while the one regarding (4) is from 2E.8. The most important part of this theorem that, for the mapping $f+N_C$ with a smooth f and a polyhedral convex set C, metric regularity is equivalent to strong metric regularity, will not be proved here in full generality. To prove this fact we need tools that go beyond the scope of this book, see the commentary to this chapter. We will however give a proof of this equivalence for a particular but important case, namely, for the mapping appearing in the Karush-Kuhn-Tucker (KTT) optimality condition in nonlinear programming. This will be done in Section 4I. First, we focus on characterizing metric regularity of (1) by applying the derivative and coderivative criteria in Theorems 4B.1 and 4C.2 to $A+N_K$. Before that can be done, however, we put some effort into a better understanding of the normal cone mapping N_K.

Faces and critical superfaces of a cone. *For a polyhedral convex cone K, a closed face is a set F of the form*

$$F = K \cap [v]^{\perp} \text{ for some } v \in K^*.$$

A superface *is a set Q of the form $F_1 - F_2$ coming from two closed faces F_1 and F_2. A superface Q will be called* critical *if it arises from closed faces F_1 and F_2 such that $F_1 \supset F_2$.*

The collection of all faces of K will be denoted by $\mathscr{F}_K$ and the collection of all *superfaces of K will be denoted by $\mathscr{Q}_K$. The collection of all critical superfaces will be* $\operatorname{crit}\mathscr{Q}_K$.

Since K is polyhedral, we obtain from the Minkowski-Weyl theorem 2E.2 and its surroundings that $\mathscr{F}_K$ is a finite collection of polyhedral convex cones. This collection contains K itself and the zero cone, in particular. The same holds for the collections $\mathscr{Q}_K$ and $\operatorname{crit}\mathscr{Q}_K$. The importance of critical superfaces is due to the following fact.

Lemma 4H.2 (critical superface lemma)**.** *Let C be a convex polyhedral set, let $v \in N_C(x)$ and let $K = K_C(x,v)$ be the critical cone for C at x for v, that is,*

$$K = T_C(x) \cap [v]^{\perp}.$$

Then there exists a neighborhood O of (x,v) such that for every choice of $(x',v') \in \operatorname{gph} N_C \cap O$ the corresponding critical cone $K_C(x',v')$ has the form

$$K_C(x',v') = F_1 - F_2$$

for some faces F_1, F_2 in $\mathscr{F}_K$ with $F_2 \subset F_1$. In particular, $K_C(x',v') \subset K - K$ for every $(x',v') \in \operatorname{gph} N_C \cap O$. Conversely, for every two faces F_1, F_2 in $\mathscr{F}_K$ with $F_2 \subset F_1$ and every neighborhood O of (x,v) there exists $(x',v') \in \operatorname{gph} N_C \cap O$ such that $K_C(x',v') = F_1 - F_2$.

In other words, the collection $\operatorname{crit}\mathscr{Q}_K$ of all critical superfaces Q of the cone $K = K_C(x,v)$ is identical to the collection of all critical cones $K_C(x',v')$ of K associated with pairs (x',v') in a small enough neighborhood of (x,v).

Proof. Because C is polyhedral, all vectors of the form $x'' = x' - x$ with $x' \in C$ close to x are the vectors $x'' \in T_C(x)$ having sufficiently small norm. Also, for such x'',

$$T_C(x') = T_C(x) + [x''] \supset T_C(x) \tag{5}$$

and

$$N_C(x') = N_C(x) \cap [x'']^{\perp} \subset N_C(x). \tag{6}$$

Now, let $(x',v') \in \operatorname{gph} N_C$ be close to (x,v) and let $x'' = x' - x$. Then from (5) we have

$$K_C(x',v') = T_C(x') \cap [v']^{\perp} = \Big(T_C(x) + [x'']\Big) \cap [v']^{\perp}.$$

Further, from (6) it follows that $v' \perp x''$ and then we obtain

$$K_C(x',v') = T_C(x) \cap [v']^{\perp} + [x''] = K_C(x,v') + [x'']. \tag{7}$$

We will next show that $K_C(x,v') \subset K$ for v' sufficiently close to v. If this were not so, there would be a sequence $v_k \to v$ and another sequence $w_k \in K_C(x,v_k)$ such that $w_k \notin K$ for all k. Each set $K_C(x,v_k)$ is a face of $T_C(x)$, but since $T_C(x)$ is polyhedral, the set of its faces is finite, hence for some face F of $T_C(x)$ we have $K_C(x,v_k) = F$

for infinitely many k. Note that the set $\operatorname{gph} K_C(x,\cdot)$ is closed, hence for any $w \in F$, since (v_k, w) is in this graph, the limit (v,w) belongs to it as well. But then $w \in K$ and since $w \in F$ is arbitrarily chosen, we have $F \subset K$. Thus the sequence $w_k \in K$ for infinitely many k, which is a contradiction. Hence $K_C(x,v') \subset K$.

Let $(x',v') \in \operatorname{gph} N_C$ be close to (x,v). Relation (7) tells us that $K_C(x',v') = K_C(x,v') + [x'']$ for $x'' = x' - x$. Let $F_1 = T_C(x) \cap [v']^{\perp}$, this being a face of $T_C(x)$. The critical cone $K = K_C(x,v) = T_C(x) \cap [v]^{\perp}$ is itself a face of $T_C(x)$, and any face of $T_C(x)$ within K is also a face of K. Then F_1 is a face of the polyhedral cone K. Let F_2 be the face of F_1 having x'' in its relative interior. Then F_2 is also a face of K and therefore $K_C(x',v') = F_1 - F_2$, furnishing the desired representation.

Conversely, let F_1 be a face of K. Then there exists $v' \in K^* = N_K(0)$ such that $F_1 = K \cap [v']^{\perp}$. The size of v' does not matter; hence we may assume that $v + v' \in N_C(x)$ by the Reduction lemma 2E.4. By repeating the above argument we have $F_1 = T_C(x) \cap [v'']^{\perp}$ for $v'' := v + v'$. Now let F_2 be a face of F_1. Let x' be in the relative interior of F_2. In particular, $x' \in T_C(x)$, so by taking the norm of x' sufficiently small we can arrange that the point $x'' = x + x'$ lies in C. We have $x' \perp v'$ and, as in (7),

$$F_1 - F_2 = T_C(x) \cap [v'']^{\perp} + [x'] = \Big(T_C(x) + [x']\Big) \cap [v'']^{\perp} = T_C(x'') \cap [v'']^{\perp} = K_C(x'',v'').$$

This gives us the form required. □

Our next step is to specify the derivative criterion for metric regularity in 4B.1 for the reduced mapping $A + N_K$.

Theorem 4H.3 (regularity modulus from derivative criterion)**.** *For $A + N_K$ with A and K as in (2), we have*

$$\text{(8)} \qquad \operatorname{reg}(A + N_K; 0\,|\,0) \;=\; \max_{Q \in \operatorname{crit} \mathcal{Q}_K} |(A + N_Q)^{-1}|^{-}.$$

Thus, $A + N_K$ is metrically regular at 0 for 0 if and only if $|(A + N_Q)^{-1}|^{-} < \infty$ for every critical superface Q of K.

Proof. From Theorem 4B.1, combined with Example 4A.4, we have that

$$\operatorname{reg}(A + N_K; 0\,|\,0) \;=\; \limsup_{\substack{(x,y)\to(0,0)\\ (x,y)\in \operatorname{gph}(A+N_K)}} |(A + N_{T_K(x)\cap[y-Ax]^{\perp}})^{-1}|^{-}.$$

Lemma 4H.2 with $(x,v) = (0,0)$ gives us the desired representation $N_{T_K(x)\cap[y-Ax]^{\perp}} = N_{F_1 - F_2}$ for (x,y) near zero and hence (8). □

Example 4H.4 (critical superfaces for complementarity problems)**.** Consider the complementarity problem

$$f(p,x) + N_{\mathbb{R}^n_+}(x) \ni 0,$$

with a solution $\bar{x}$ for $\bar{p}$. Let K and A be as in (2) (with $C = \mathbb{R}^n_+$) and consider the index sets

$$J_1 = \{\, j \,|\, \bar{x}_j > 0, \bar{v}_j = 0 \}, \quad J_2 = \{\, j \,|\, \bar{x}_j = 0, \bar{v}_j = 0 \}, \quad J_3 = \{\, j \,|\, \bar{x}_j = 0, \bar{v}_j < 0 \}$$

for $\bar{v} = -f(\bar{p}, \bar{x})$. Then the critical superfaces Q of K have the following description. There is a partition of $\{1, \dots, n\}$ into index sets J_1', J_2', J_3' with

$$J_1 \subset J_1' \subset J_1 \cup J_2, \qquad J_3 \subset J_3' \subset J_2 \cup J_3,$$

such that

$$(9) \qquad x' \in Q \iff \begin{cases} x_i' \text{ free} & \text{for } i \in J_1', \\ x_i' \geq 0 & \text{for } i \in J_2', \\ x_i' = 0 & \text{for } i \in J_3'. \end{cases}$$

Detail. Each face F of K has the form $K \cap [v']^\perp$ for some vector $v' \in K^*$. The vectors v' in question are those with

$$\begin{cases} v_i' = 0 & \text{for } i \in J_1, \\ v_i' \leq 0 & \text{for } i \in J_2, \\ v_i' \text{ free} & \text{for } i \in J_3. \end{cases}$$

The closed faces F of K correspond one-to-one therefore with the subsets of J_2: the face F corresponding to an index set J_2^F consists of the vectors x' such that

$$\begin{cases} x_i' \text{ free} & \text{for } i \in J_1, \\ x_i' \geq 0 & \text{for } i \in J_2 \setminus J_2^F, \\ x_i' = 0 & \text{for } i \in J_3 \cup J_2^F. \end{cases}$$

If such faces F_1 and F_2 have $J_2^{F_1} \subset J_2^{F_2}$, so that $F_1 \supset F_2$, then $F_1 - F_2$ is given by (9) with $J_1' = J_1 \cup [J_2 \setminus J_2^{F_2}]$, $J_2' = J_2^{F_2} \setminus J_2^{F_1}$, $J_3' = J_3 \cup J_2^{F_1}$. □

Theorem 4H.5 (critical superface criterion from graphical derivative criterion)**.** *For a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}^n$ and a polyhedral convex set $C \subset \mathbb{R}^n$, let $f(\bar{x}) + N_C(\bar{x}) \ni 0$ and let K be the critical cone to C at $\bar{x}$ for $-f(\bar{x})$. Then the mapping $f + N_C$ is metrically regular at $\bar{x}$ for 0 if and only if, for all critical superfaces Q of K, i.e., $Q \in \operatorname{crit}\mathscr{Q}_K$, the following condition holds with $A = \nabla f(\bar{x})$:*

$$\forall\, v \in \mathbb{R}^n \quad \exists\, u \in Q \text{ such that } (v - Au) \in Q^* \text{ and } (v - Au) \perp u.$$

Proof. From 3F.7, metric regularity of $f + N_C$ at $\bar{x}$ for 0 is equivalent to metric regularity of $A + N_K$ at 0 for 0. Then it is enough to apply 4H.3 together with 4A.7(16) and using the fact that $w \in N_Q(u)$ whenever $u \in Q$, $w \in Q^*$, and $u \perp w$. □

Exercise 4H.6 (variational inequality over a subspace)**.** *Show that when the critical cone K in 4H.5 is a subspace of $\mathbb{R}^n$ of dimension $m \leq n$, then the matrix BAB^T is nonsingular, where B is the matrix whose columns form an orthonormal basis in K.*

To illuminate the structure of the superfaces and their polars that underlies the criterion in Theorem 4H.5, we record some additional properties. In polarity they make use of the following concept.

Complementary faces. *A closed face F of a polyhedral cone $K \subset \mathbb{R}^n$ and a closed face F' of the polar cone K^* are complementary to each other when*

$$F \perp F' \text{ and } \dim F + \dim F' = n.$$

The unique face F' of K^* complementary to F is in fact $(F-F)^\perp \cap K^*$, and then in turn F is given by $(F'-F')^\perp \cap K$, because $K = (K^*)^*$. In the definition of a face F, as having a representation $F = K \cap [v]^\perp$ with $v \in K^*$, the vectors v that fill this role are the ones in the relative interior of F'. Likewise, the vectors w in the relative interior of F are the vectors $w \in K$ such that $F' = K^* \cap [w]^\perp$.

Proposition 4H.7 (superface properties). *Let K be a polyhedral cone in $\mathbb{R}^n$ with polar K^*.*

(a) *In the expression $Q = F_1 - F_2$ for a superface in terms of closed faces F_1 and F_2 of K, both F_1 and F_2 are uniquely determined by Q, namely $F_1 = Q \cap K$ and $F_2 = (-Q) \cap K$. In particular, a superface is a subspace if and only if it is of the form $F - F$ for some face F of K.*

(b) *The superfaces Q of K correspond one-to-one with the superfaces $Q^\#$ in terms of $Q = F_1 - F_2$ and $Q^\# = F_2' - F_1'$ for the faces F_1' and F_2' complementary to F_1 and F_2. In this correspondence, Q is a critical superface of K if and only if $Q^\#$ is a critical superface of K^*.*

(c) *For a critical superface $Q = F_1 - F_2$ of K, the polar cone Q^* is $Q^\#$ for the corresponding critical superface $Q^\#$ of K^* as described in (b).*

Proof. For (a) with $Q = F_1 - F_2$, consider any nonzero v in the relative interior of the face F_1' complementary to F_1, so as to get

$$F_1 = \left\{ w_1 \in K \,\middle|\, \langle v, w_1 \rangle \geq 0 \right\} = \left\{ w_1 \in K \,\middle|\, \langle v, w_1 \rangle = 0 \right\} = K \cap [v]^\perp,$$

while having

$$\langle v, w_2 \rangle \leq 0 \text{ for all } w_2 \in F_2, \text{ with } \langle v, w_2 \rangle < 0 \text{ unless } w_2 \in F_1 \cap F_2.$$

Let $w = w_1 - w_2$ with $w_1 \in F_1$ and $w_2 \in F_2$. Then $\langle v, w \rangle - \langle v, w_2 \rangle > 0$ unless $w_2 \in F_1 \cap F_2$. Consequently $w \notin K$ except perhaps if $w_1 \in F_1 \cap F_2$, or in other words $Q \cap K = (F_1 - F_1 \cap F_2) \cap K$. Since $F_1 - F_1 \cap F_2 \subset [v]^\perp$, this comes out as $Q \cap K = K \cap [v]^\perp = F_1$. The other claim, that $(-Q) \cap K = F_2$, follow then by symmetry in reversing the roles of F_1 and F_2.

Continuing to (b) with the complementary faces F_1' and F_2', we see that these are uniquely determined by Q, and the same is true then of $Q^\#$. It follows too by symmetry that Q can be recovered uniquely from $Q^\#$ along with F_1 and F_2. In this correspondence, having $F_1 \supset F_2$ corresponds to having $F_2' \supset F_1'$, so criticality is preserved. Indeed, this comes from the facts in the definition of complementarity, namely that for $i = 1,2$, we have $F_i = K \cap (F_i - F_i)$ and $F_i' = K^* \cap (F_i' - F_i')$ with $F_i' - F_i' = (F_i - F_i)^\perp$ and $F_i - F_i = (F_i' - F_i')^\perp$. Since $F_1 \supset F_2$, we have $F_1 - F_1 \supset F_2 - F_2$ and therefore $F_2' \supset F_1'$ in particular.

To justify (c) for a critical superface $Q = F_1 - F_2$, we observe next that

$$(10)\qquad F_i^* = [K \cap (F_i - F_i)]^* = K^* + (F_i - F_i)^\perp = K^* + (F_i' - F_i') = K^* - F_i'.$$

Also, $F_1 - F_2 = F_1 + (F_2 - F_2)$, so that

$$(11)\qquad (F_1 - F_2)^* = F_1^* \cap (F_2 - F_2)^\perp = (K^* - F_1') \cap (F_2' - F_2')$$

by (10). Because $F_1' \subset F_2'$, any element $y - z$ with $y \in K^*$ and $z \in F_1'$ that belongs to the final set in (11) must have $y \in F_2' - F_2'$. Hence $(K^* - F_1') \cap (F_2' - F_2') = K^* \cap (F_2' - F_2') - F_1' = F_2' - F_1'$, and we get $(F_1 - F_2)^* = F_2' - F_1'$ as claimed. Through symmetry, we likewise have $(F_2' - F_1')^* = F_1 - F_2$. □

We will now apply the coderivative criterion for metric regularity in 4C.2 to the mapping in (1). According to Theorem 4H.1, for that purpose we have to compute the coderivative of the mapping in $A + N_K$. The first step to do that is easy and we will give it as an exercise.

Exercise 4H.8 (reduced coderivative formula)**.** *Show that, for a linear mapping $A : \mathbb{R}^n \to \mathbb{R}^n$ and a closed convex cone $K \subset \mathbb{R}^n$ one has*

$$D^*(A + N_K)(\bar{x}\,|\,\bar{y}) = A^* + D^*N_K(\bar{x}\,|\,\bar{y} - A\bar{x}).$$

Guide. Apply the definition of the general normal cone in Section 4C. □

Thus, everything hinges on determining the coderivative $D^*N_K(0\,|\,0)$ of the mapping N_K at the point $(0,0) \in G = \operatorname{gph} N_K$. By definition, the graph of the coderivative mapping consists of all pairs $(w, -z)$ such that $(w,z) \in N_G(0,0)$ where N_G is the general normal cone to the nonconvex set G. In these terms, for A and K as in (2), the coderivative criterion becomes

$$(12)\qquad (u, A^{\mathsf{T}}u) \in N_G(0,0) \quad\Longrightarrow\quad u = 0.$$

Everything depends then on determining $N_G(0,0)$.

We will next appeal to the known fact[6] that $N_G(0,0)$ is the limsup of polar cones $T_G(x,v)^*$ at $(x,v) \in G$ as $(x,v) \to (0,0)$. Because G is the union of finitely many polyhedral convex sets in $\mathbb{R}^{2n}$ (due to K being polyhedral), only finitely many cones can be manifested as $T_G(x,v)$ at points $(x,v) \in G$ near $(0,0)$. Thus, for a sufficiently small neighborhood O of the origin in $\mathbb{R}^{2n}$ we have that

$$(13)\qquad N_G(0,0) = \bigcup_{(x,v)\in O\cap G} T_G(x,v)^*.$$

[6] See Proposition 6.5 in Rockafellar and Wets [1998].

It follows from reduction lemma 2E.4 that $T_G(x,v) = \operatorname{gph} N_{K(x,v)}$, where $K(x,v) = \{x' \in T_K(x) \,|\, x' \perp v\}$. Therefore,

$$T_G(x,v) = \{(x',v') \,|\, x' \in K(x,v),\ v' \in K(x,v)^*, x' \perp v'\},$$

and we have

$$\begin{aligned} T_G(x,v)^* &= \{(r,u) \,|\, \langle (r,u),(x',v')\rangle \le 0 \text{ for all } (x',v') \in T_G(x,v)\} \\ &= \{(r,u) \,|\, \langle r,x'\rangle + \langle u,v'\rangle \le 0 \text{ for all} \\ &\qquad\qquad x' \in K(x,v),\ v' \in K(x,v)^* \text{ with } x' \perp v'\}. \end{aligned}$$

It is evident from this (first in considering $v' = 0$, then in considering $x' = 0$) that actually

$$T_G(x,v)^* = K(x,v)^* \times K(x,v). \tag{14}$$

Hence $N_G(0,0)$ is the union of all product sets $\hat{K}^* \times \hat{K}$ associated with cones $\hat{K}$ such that $\hat{K} = K(x,v)$ for some $(x,v) \in G$ near enough to $(0,0)$.

It remains to observe that the form of the critical cones $\hat{K} = K(x,v)$ at points (x,v) close to $(0,0)$ is already derived in Lemma 4H.2, namely, for every choice of $(x,v) \in \operatorname{gph} K$ near $(0,0)$ (this last requirement is actually not needed) the corresponding critical cone $\hat{K} = K(x,v)$ is a critical superface of K. To see this, all one has to do is to replace C by K and (x,v) by $(0,0)$ in the proof of 4H.2. Summarizing, from (12), (13) and (14), and the coderivative criterion in 4C.2, we come to the following result:

Theorem 4H.9 (critical superface criterion from coderivative criterion)**.** *For the mapping $A + N_K$ we have*

$$\operatorname{reg}(A + N_K; 0\,|\,0) \;=\; \max_{Q \in \operatorname{crit} \mathscr{Q}_K} \sup_{\substack{u \in Q \\ |u|=1}} \frac{1}{d(A^{\mathsf{T}}u, Q^*)}.$$

Thus, $A + N_K$ is metrically regular at 0 for 0 if and only if for every superface $Q \in \operatorname{crit} \mathscr{Q}_K$,

$$u \in Q \text{ and } A^{\mathsf{T}}u \in Q^* \quad \Longrightarrow \quad u = 0.$$

To conclude this section we look at the strong regularity in Theorem 4H.1 in terms of the strict derivative criterion in Theorem 4D.1 and provide an elaboration that utilizes the results we have developed about the facial structure of polyhedral cones. For A and K as in Theorem 4H.1, the strict derivative criterion for $A + N_K$ to be strongly metrically regular at 0 for 0 comes down to having

$$0 \in D_*(A + N_K)(0|0)(w) \quad \Longrightarrow \quad w = 0. \tag{15}$$

Note that the assumption 4D(2) may not hold in the case considered, and then (15) is only a necessary condition for strong metric regularity. Since $D_*(A + N_K) = A +$

D_*N_K, (15) can be written equivalently as

$$(w, -Aw) \in \operatorname{gph} D_*N_K(0|0) \quad \Longrightarrow \quad w = 0.$$

Understanding $\operatorname{gph} D_*N_K(0|0)$ then becomes the issue, and this is where more can now be said. The additional information revolves around the superfaces Q of K forming the collection $\mathscr{Q}_K$.

Proposition 4H.10 (strict graphical derivative structure in polyhedral convexity)**.** *For a polyhedral cone K, one has*

$$\operatorname{gph} D_*N_K(0|0) = \operatorname{gph} N_K - \operatorname{gph} N_K = \bigcup \left\{ Q \times [-Q^\#] \,\middle|\, Q \in \mathscr{Q}_K \right\}$$

where $Q^\#$ is the superface of K^* that is dual to Q in the sense of 4H.7(b).

Proof. Let $G = \operatorname{gph} D_*N_K(0|0)$. By definition, G consists of all pairs (w, z) obtainable as

$$(w,z) = \lim_k \frac{1}{t_k}[(w_1^k, z_1^k) - (w_2^k, z_2^k)] \text{ with } (w_i^k, z_i^k) \in \operatorname{gph} N_K,\ (w_i^k, z_i^k) \to (0,0),\ t_k \searrow 0.$$

Since $\operatorname{gph} N_K$ is a closed cone (in fact the union of finitely many polyhedral convex cones), both t_k and the limits in k are superfluous: we simply have $G = \operatorname{gph} N_K - \operatorname{gph} N_K$.

In general, we know that $(w, z) \in \operatorname{gph} N_K$ if and only if $w \in K$, $z \in K^*$, and $w \perp z$. This signals that complementary faces of K and K^* can be brought in. The pairs in question belong to $F \times F'$ for some closed face F of K and its complement F'. Thus,

$$\operatorname{gph} N_K = \bigcup \left\{ F \times F' \,\middle|\, \text{ for complementary face pairs } (F, F') \right\}.$$

This furnishes a representation of $\operatorname{gph} N_K$ as the union of a *finite* collection of *n-dimensional* polyhedral convex cones with disjoint relative interiors. It follows then that

$$G = \bigcup \left\{ (F_1 - F_2) \times (F_1' - F_2') \,\middle|\, \text{ for complementary face pairs } (F_i, F_i'),\ i = 1, 2 \right\}.$$

It remains only to recall that $F_1 - F_2$ is a superface $Q \in \mathscr{Q}_K$ that corresponds to $Q^\# = F_2' - F_1'$ as a superface of K^*. $\square$

Corollary 4H.11 (strict derivative criterion, elaborated)**.** *The strict derivative criterion (15) is equivalent to the following. For each superface Q of K and the corresponding dual superface $Q^\#$ of K^*, one has $w \in Q$ and $Aw \in Q^\#$ only for $w = 0$.*

4I. Strong Metric Regularity of the KKT Mapping

In this section we apply first Kummer's theorem 4D.6 to the Karush-Kuhn-Tucker (KKT) system associated with the following nonlinear programming problem with inequality constraints and special *canonical* perturbations:

(1) minimize $g_0(x) - \langle v, x\rangle$ over all x satisfying $g_i(x) \le u_i$ for $i \in [1,m]$,

where the functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 0,1,\dots,m$ are twice continuously differentiable and $v \in \mathbb{R}^n$, $u = (u_1,\dots,u_m)^{\mathsf{T}} \in \mathbb{R}^m$ are parameters. According to the basic first-order optimality conditions established in Section 2A, if x is a solution to (1) and and the constraint qualification condition 2A(13) holds, then there exists a multiplier vector $y = (y_1,\dots,y_m)$ such that the pair (x,y) satisfies the so-called Karush–Kuhn–Tucker conditions 2A(24), which are in the form of the variational inequality

$$(2) \qquad \begin{pmatrix} -v + \nabla g_0(x) + y\nabla g(x) \\ -u + g(x) \end{pmatrix} \in N_E(x,y) \text{ for } E = \mathbb{R}^n \times \mathbb{R}^m_+.$$

We wish to apply Kummer's theorem to characterize the strong metric regularity of the mapping

$$(3) \qquad G : (x,y) \mapsto \begin{pmatrix} \nabla g_0(x) + y\nabla g(x) \\ -g(x) \end{pmatrix} + N_E(x,y),$$

in terms of which (2) becomes the inclusion

$$G(x,y) \ni \begin{pmatrix} v \\ u \end{pmatrix}.$$

Choose a reference value $(\bar v,\bar u)$ of the parameters and let $(\bar x,\bar y)$ solve (2) for $(\bar v,\bar u)$, that is, $(\bar v,\bar u) \in G(\bar x,\bar y)$. By definition, G is strongly metrically regular at $(\bar x,\bar y)$ for $(\bar v,\bar u)$ exactly when G^{-1} has a Lipschitz continuous localization around $(\bar v,\bar u)$ for $(\bar x,\bar y)$.

To apply Kummer's theorem 4D.6, we first convert the variational inequality (2) into an equation involving the function $H : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ defined as follows:

$$(4) \qquad H(x,y) = \begin{pmatrix} \nabla g_0(x) + \sum_{i=1}^m y_i^+ \nabla g_i(x) \\ -g_1(x) + y_1^- \\ \vdots \\ -g_m(x) + y_m^- \end{pmatrix}.$$

Let $((x,y),(v,u)) \in \operatorname{gph} H$; then for $z_i = y_i^+$, $i = 1,\dots,m$, we have that $((x,z),(v,u)) \in \operatorname{gph} G$. Indeed, for each $i = 1,\dots,m$, if $y_i \le 0$, then $u_i + g_i(x) = y_i^- \le 0$ and $(u_i + g_i(x))y_i^+ = 0$; otherwise $u_i + g_i(x) = y_i^- = 0$. Conversely, if $((x,z),(v,u)) \in \operatorname{gph} G$, then for

$$y_i = \begin{cases} z_i & \text{if } z_i > 0, \\ u_i + g_i(x) & \text{if } z_i = 0, \end{cases} \tag{5}$$

we obtain $(x,y) \in H^{-1}(v,u)$. In particular, if H^{-1} has a Lipschitz continuous localization around $(\bar{v},\bar{u})$ for $(\bar{x},\bar{y})$ then G^{-1} has the same property at $(\bar{v},\bar{u})$ for $(\bar{x},\bar{z})$ where $\bar{z}_i = \bar{y}_i^+$, and if G^{-1} has a Lipschitz continuous localization around $(\bar{v},\bar{u})$ for $(\bar{x},\bar{z})$ then H^{-1} has the same property at $(\bar{v},\bar{u})$ for $(\bar{x},\bar{y})$, where $\bar{y}$ satisfies (5).

Next, we need to determine the strict derivative of H. There is no trouble in differentiating the expressions $-g_i(x)+y_i^-$, inasmuch as we already know from 4D.5 the strict derivative of y^-. A little bit more involved is the determination of the strict derivative of $\varphi_i(x,y) := \nabla g_i(x) y_i^+$ for $i = 1,\dots,m$. Adding and subtracting the same expressions, passing to the limit as in the definition, and using 4D.5, we obtain

$$z \in D_*\varphi_i(\bar{x},\bar{y})(u,v) \iff z = \bar{y}_i^+ \nabla^2 g_i(\bar{x})u + \lambda_i v_i \nabla g_i(\bar{x}), \quad i = 1,\dots,m,$$

where the coefficients λ_i for $i = 1,\dots,m$ satisfy

$$\lambda_i \begin{cases} = 1 & \text{for } \bar{y}_i > 0, \\ \in [0,1] & \text{for } \bar{y}_i = 0, \\ = 0 & \text{for } \bar{y}_i < 0. \end{cases} \tag{6}$$

Right from the definition, the form thereby obtained for the strict graphical derivative of the function H in (4) at $(\bar{x},\bar{y})$ is as follows:

$$(\xi,\eta) \in D_*H(\bar{x},\bar{y})(u,v) \iff \begin{cases} \xi = Au + \sum_{i=1}^m \lambda_i v_i \nabla g_i(\bar{x}), \\ \eta_i = -\nabla g_i(\bar{x})u + (1-\lambda_i)v_i \quad \text{for } i = 1,\dots,m, \end{cases}$$

where the λ_i's are as in (6), and

$$A = \nabla^2 g_0(\bar{x}) + \sum \bar{y}_i^+ \nabla^2 g_i(\bar{x}).$$

Denoting by Λ the $m \times m$ diagonal matrix with elements λ_i on the diagonal, by I_m the $m \times m$ identity matrix, and setting

$$B = \begin{pmatrix} \nabla g_1(\bar{x}) \\ \vdots \\ \nabla g_m(\bar{x}) \end{pmatrix},$$

we obtain that

$$M(\Lambda) \in D_*H(\bar{x},\bar{y}) \iff M(\Lambda) = \begin{pmatrix} A & B^{\mathsf{T}}\Lambda \\ -B & I_m - \Lambda \end{pmatrix}. \tag{7}$$

This formula can be simplified by re-ordering the functions g_i according to the sign of $\bar{y}_i$. We first introduce some notation. Let $I = \{1,\dots,m\}$ and, without loss of generality, suppose that

$$\{\, i \in I \mid \bar{y}_i > 0 \,\} = \{1,\dots,k\} \quad \text{and} \quad \{\, i \in I \mid \bar{y}_i = 0 \,\} = \{k+1,\dots,l\}.$$

Let

$$B_+ = \begin{pmatrix} \nabla g_1(\bar{x}) \\ \vdots \\ \nabla g_k(\bar{x}) \end{pmatrix} \quad \text{and} \quad B_0 = \begin{pmatrix} \nabla g_{k+1}(\bar{x}) \\ \vdots \\ \nabla g_l(\bar{x}) \end{pmatrix},$$

let Λ_0 be the $(l-k)\times(l-k)$ diagonal matrix with diagonal elements $\lambda_i \in [0,1]$, let I_0 be the identity matrix for $\mathbb{R}^{l-k}$, and let I_{m-l} be the identity matrix for $\mathbb{R}^{m-l}$. Then, since $\lambda_i = 1$ for $i = 1,\ldots,k$ and $\lambda_i = 0$ for $i = l+1,\ldots,m$, the matrix $M(\Lambda)$ in (7) takes the form

$$M(\Lambda_0) = \begin{pmatrix} A & B_+^{\mathsf{T}} & B_0^{\mathsf{T}}\Lambda_0 & 0 \\ & 0 & 0 & 0 \\ -B & 0 & I_0 - \Lambda_0 & 0 \\ & 0 & 0 & I_{m-l} \end{pmatrix}.$$

Each column of $M(\Lambda_0)$ depends on at most one λ_i, hence there are numbers

$$a_{k+1}, b_{k+1}, \ldots, a_l, b_l$$

such that

$$\det M(\Lambda_0) = (a_{k+1} + \lambda_{k+1} b_{k+1}) \cdots (a_l + \lambda_l b_l).$$

Therefore, $\det M(\Lambda_0) \neq 0$ for all $\lambda_i \in [0,1]$, $i = k+1,\ldots,l$, if and only if the following condition holds:

$$a_i \neq 0,\ a_i + b_i \neq 0 \ \text{and} \ [\operatorname{sign} a_i = \operatorname{sign}(a_i+b_i) \ \text{or} \ \operatorname{sign} b_i \neq \operatorname{sign}(a_i+b_i)], \quad \text{for } i = k+1,\ldots,l.$$

Here we invoke the convention that

$$\operatorname{sign} a = \begin{cases} 1 & \text{for } a > 0, \\ 0 & \text{for } a = 0, \\ -1 & \text{for } a < 0. \end{cases}$$

One can immediately note that it is not possible to have simultaneously $\operatorname{sign} b_i \neq \operatorname{sign}(a_i+b_i)$ and $\operatorname{sign} a_i \neq \operatorname{sign}(a_i+b_i)$ for some i. Therefore, it suffices to have

(8) $\quad a_i \neq 0,\ a_i + b_i \neq 0,\ \operatorname{sign} a_i = \operatorname{sign}(a_i+b_i) \ \text{for all} \ i = k+1,\ldots,l.$

Now, let J be a subset of $\{k+1,\ldots,l\}$ and for $i = k+1,\ldots,l$, and let

$$\lambda_i^J = \begin{cases} 1 & \text{for } i \in J, \\ 0 & \text{otherwise.} \end{cases}$$

Let Λ^J be the diagonal matrix composed by these λ_i^J, and let $B_0(J) = \Lambda^J B_0$. Then we can write

$$M(J) := M(\Lambda^J) = \begin{pmatrix} A & B_+^{\mathsf{T}} & B_0(J)^{\mathsf{T}} & 0 \\ & 0 & 0 & 0 \\ -B & 0 & I_0^J & 0 \\ & 0 & 0 & I_l \end{pmatrix},$$

where I_0^J is the diagonal matrix having 0 as $(i-k)$-th element if $i \in J$ and 1 otherwise. Clearly, all the matrices $M(J)$ are obtained from $M(\Lambda_0)$ by taking each λ_i either 0 or 1. The condition (8) can be then written equivalently as

$$\det M(J) \neq 0 \text{ and } \operatorname{sign} \det M(J) \text{ is the same for all } J. \tag{9}$$

Let the matrix $B(J)$ have as rows the row vectors $\nabla g_i(\bar{x})$ for $i \in J$. Reordering the last $m-k$ columns and rows of $M(J)$, if necessary, we obtain

$$M(J) = \begin{pmatrix} A & B_+^{\mathsf{T}} & B(J)^{\mathsf{T}} & 0 \\ -B & 0 & 0 & I \end{pmatrix},$$

where I is now the identity for $\mathbb{R}^{\{k+1,\ldots,m\}\setminus J}$. The particular form of the matrix $M(J)$ implies that $M(J)$ fulfills (9) if and only if (9) holds for just a part of it, namely for the matrix

$$N(J) := \begin{pmatrix} A & B_+^{\mathsf{T}} & B(J)^{\mathsf{T}} \\ -B_+ & 0 & 0 \\ -B(J) & 0 & 0 \end{pmatrix}. \tag{10}$$

By applying Kummer's theorem, we arrive finally at the following result.

Theorem 4I.1 (characterization of KKT strong metric regularity)**.** *The solution mapping of the KKT variational inequality (2) has a Lipschitz continuous single-valued localization around* $(\bar{v},\bar{u})$ *for* $(\bar{x},\bar{y})$ *if and only if, for the matrix* $N(J)$ *in (10),* $\det N(J)$ *has the same nonzero sign for all* $J \subset \left\{ i \in I \,\middle|\, \bar{y}_i = 0 \right\}$.

We will prove next that for the KKT mapping, *metric regularity and strong metric regularity are equivalent properties*. To make things different, we consider a slightly more general problem than (1), which is not explicitly parameterized but now involves both equality and inequality constraints:

$$\text{minimize } g_0(x) \text{ over all } x \text{ satisfying } g_i(x) \begin{cases} = 0 & \text{for } i \in [1,r], \\ \leq 0 & \text{for } i \in [r+1,m] \end{cases} \tag{11}$$

with twice continuously differentiable functions $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 0,1,\ldots,m$. The associated KKT optimality system has the form

$$f(x,y) + N_{\bar{E}}(x,y) \ni (0,0), \tag{12}$$

where

$$f(x,y) = \begin{pmatrix} \nabla g_0(x) + \sum_{i=1}^m \nabla g_i(x) y_i \\ -g_1(x) \\ \vdots \\ -g_m(x) \end{pmatrix} \tag{13}$$

and

(14) $$\bar{E} = \mathbb{R}^n \times [\mathbb{R}^r \times \mathbb{R}^{m-r}_+].$$

Theorem 2A.8 tells us that, under the constraint qualification condition 2A(18), for every local minimum x of (11) there exists a Lagrange multiplier y, with $y_i \geq 0$ for $i = r+1, \ldots, m$, such that (x, y) is a solution of (12). We will now establish an important fact for the mapping on the left side of (12).

Theorem 4I.2 (KKT metric regularity implies strong metric regularity)**.** *Consider the mapping $F : \mathbb{R}^{n+m} \rightrightarrows \mathbb{R}^{n+m}$ defined as*

(15) $$F : z \mapsto f(z) + N_{\bar{E}}(z)$$

with f as in (13) for $z = (x, y)$ and $\bar{E}$ as in (14), and let $\bar{z} = (\bar{x}, \bar{y})$ solve (12), that is, $F(\bar{z}) \ni 0$. If F is metrically regular at $\bar{z}$ for 0, then F is strongly metrically regular there.

We already showed in Theorem 3G.5 that this kind of equivalence holds for locally monotone mappings, but here F need not be monotone even locally, although it is a special kind of mapping in another way.

The claimed equivalence is readily apparent in a simple case of (15) when F is an affine mapping, which corresponds to problem (17) with no constraints and with g_0 being a quadratic function, $g_0(x) = \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle$ for an $n \times n$ matrix A and a vector $b \in \mathbb{R}^n$. Then $F(x, y) = Ax + b$ and metric regularity of F (at any point) means that A has full rank. But then A must be nonsingular, so F is in fact strongly regular.

The general argument for $F = f + N_{\bar{E}}$ is lengthy and proceeds through a series of reductions. First, since our analysis is local, we can assume without loss of generality that all inequality constraints are active at $\bar{x}$. Indeed, if for some index $i \in [r+1, m]$ we have $g_i(\bar{x}) < 0$, then $\bar{y}_i = 0$. For $q \in \mathbb{R}^{n+m}$ consider the solution set of the inclusion $F(z) \ni q$. Then for any q near zero and all x near $\bar{x}$ we will have $g_i(x) < q_i$, and hence any Lagrange multiplier y associated with such an x must have $y_i = 0$; thus, for q close to zero the solution set of $F(z) \ni q$ will not change if we drop the constraint with index i. Further, if there exists an index i such that $\bar{y}_i > 0$, then we can always rearrange the constraints so that $\bar{y}_i > 0$ for $i \in [r+1, s]$ for some $r < s \leq m$. Under these simplifying assumptions the critical cone $K = K_{\bar{E}}(\bar{z}, \bar{v})$ to the set $\bar{E}$ in (14) at $\bar{z} = (\bar{x}, \bar{y})$ for $\bar{v} = -f(\bar{z})$ is the product $\mathbb{R}^n \times \mathbb{R}^s \times \mathbb{R}^{m-s}_+$. (Show that this form of the critical cone can be also derived by utilizing Example 2E.5.) The normal cone mapping N_K to the critical cone K has then the form $N_K = \{0\}^n \times \{0\}^s \times N^{m-s}_+$.

We next recall that metric regularity of F is equivalent to metric regularity of the mapping
$$L : z \mapsto \nabla f(\bar{z})z + N_{\bar{E}}(z) \quad \text{for } z = (x, y) \in \mathbb{R}^{n+m}$$
at 0 for 0 and the same equivalence holds for strong metric regularity. This reduction to a simpler situation has already been highlighted several times in this book, e.g. in 2E.8 for strong metric regularity and 3F.9 for metric regularity. Thus, to achieve our goal of confirming the claimed equivalence between metric regularity and strong regularity for F, it is enough to focus on the mapping L which, in terms of the

functions g_i in (11), has the form

$$(16) \qquad L = \begin{pmatrix} A & B^{\mathsf{T}} \\ -B & 0 \end{pmatrix} + N_{\bar{E}},$$

where

$$A = \nabla^2 g_0(\bar{x}) + \sum_{i=1}^{m} \nabla^2 g_i(\bar{x})\bar{y}_i \quad \text{and} \quad B = \begin{pmatrix} \nabla_x g_1(\bar{x}) \\ \vdots \\ \nabla g_m(\bar{x}) \end{pmatrix}.$$

Taking into account the specific form of $N_{\bar{E}}$, the inclusion $(v,w) \in L(x,y)$ becomes

$$(17) \qquad \begin{cases} v = Ax + B^{\mathsf{T}} y, & \\ (w+Bx)_i = 0 & \text{for } i \in [1,s], \\ (w+Bx)_i \le 0,\ y_i \ge 0,\ y_i(w+Bx)_i = 0 & \text{for } i \in [s+1,m]. \end{cases}$$

In further preparation for proving Theorem 4I.2, next we state and prove three lemmas. From now on any kind of regularity is at 0 for 0, unless specified otherwise.

Lemma 4I.3 (KKT metric regularity implies strong metric subregularity)**.** *If the mapping L in (16) is metrically regular, then it is strongly subregular.*

Proof. Suppose that L is metrically regular. Then the critical superface criterion displayed in 4H.5 with critical superfaces given in 4H.4 takes the following form: for every partition J_1', J_2', J_3' of $\{s+1,\ldots,m\}$ and for every $(v,w) \in I\!\!R^n \times I\!\!R^m$ there exists $(x,y) \in I\!\!R^n \times I\!\!R^m$ satisfying

$$(18) \qquad \begin{cases} v = Ax + B^{\mathsf{T}} y, & \\ (w+Bx)_i = 0 & \text{for } i \in [1,s], \\ (w+Bx)_i = 0 & \text{for } i \in J_1', \\ (w+Bx)_i \le 0,\ y_i \ge 0,\ y_i(w+Bx)_i = 0 & \text{for } i \in J_2', \\ y_i = 0 & \text{for } i \in J_3'. \end{cases}$$

In particular, denoting by B_0 the submatrix of B composed by the first s rows of B, for any index set $J \subset \{s+1,\ldots,m\}$, including the empty set, if $B(J)$ is the submatrix of B whose rows have indices in J, then the condition involving (18) implies that

$$(19) \qquad \text{the matrix } N(J) = \begin{pmatrix} A & B_0^{\mathsf{T}} & B^{\mathsf{T}}(J) \\ -B_0 & 0 & 0 \\ -B(J) & 0 & 0 \end{pmatrix} \text{ is nonsingular.}$$

Indeed, to reach such a conclusion it is enough to take $J = J_1'$ and $J_2' = \emptyset$ in (18). By 4E.1, the mapping L in (16) is strongly subregular if and only if

$$(20) \qquad \text{the only solution of (18) with } (v,w) = 0 \text{ is } (x,y) = 0.$$

Now, suppose that L is not strongly subregular. Then, by (20), for some index set $J \subset \{s+1,\ldots,m\}$, possibly the empty set, there exists a nonzero vector $(x,y) \in I\!\!R^n \times I\!\!R^m$

satisfying (18) for $v = 0, w = 0$. Note that this y has $y_j = 0$ for $j \in \{s+1,\dots,m\} \setminus J$. But then the nonzero vector $z = (x,y)$ with y having components in $\{1,\dots,s\} \times J$ solves $N(J)z = 0$ where the matrix $N(J)$ is defined in (19). Hence, $N(J)$ is singular, and then the condition involving (18) is violated; thus, the mapping L is not metrically regular. This contradiction means that L must be strongly subregular. □

The next two lemmas present general facts that are separate from the specific circumstances of nonlinear programming problem (11) considered. The second lemma is a simple consequence of Brouwer's invariance of domain theorem 1F.1:

Lemma 4I.4 (single-valued localization from continuous local selection)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuous and let there exist an open neighborhood V of $\bar{y} := f(\bar{x})$ and a continuous function $h : V \to \mathbb{R}^n$ such that $h(y) \in f^{-1}(y)$ for $y \in V$. Then f^{-1} has a single-valued graphical localization around $\bar{y}$ for $\bar{x}$.*

Proof. Since f is a function, we have $f^{-1}(y) \cap f^{-1}(y') = \emptyset$ for every $y, y' \in V$, $y \neq y'$. But then h is one-to-one and hence h^{-1} is a function defined in $U := h(V)$. Note that $x = h(h^{-1}(x)) \in f^{-1}(h^{-1}(x))$ implies $f(x) = h^{-1}(x)$ for all $x \in U$. Since h is a function, we have that $h^{-1}(x) \neq h^{-1}(x')$ for all $x, x' \in U$, $x \neq x'$. Thus, f is one-to-one on U, implying that the set $f^{-1}(y) \cap U$ consists of one point, $h(y)$. Hence, by Theorem 1F.1 applied to h, U is an open neighborhood of $\bar{x}$. Therefore, h is a single-valued graphical localization of f^{-1} around $\bar{y}$ for $\bar{x}$. □

Lemma 4I.5 (properties of optimal solutions)**.** *Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and let $Q : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ have the Aubin property at $\bar{p}$ for $\bar{x}$. Then any graphical localization around $\bar{p}$ for $\bar{x}$ of the solution mapping S_{opt} of the problem*

$$\textit{minimize } \varphi(x) \quad \textit{subject to } \; x \in Q(p)$$

is either multi-valued or a continuous function on a neighborhood of $\bar{p}$.

Proof. Suppose that S_{opt} has a single-valued localization $\hat{x}(p) = S_{\text{opt}}(p) \cap U$ for all $p \in V$ for some neighborhoods U of $\bar{x}$ and V of $\bar{p}$ and, without loss of generality, that Q has the Aubin property at $\bar{p}$ for $\bar{x}$ with the same neighborhoods U and V. Let $p \in V$ and $V \ni p_k \to p$ as $k \to \infty$. Since $\hat{x}(p) \in Q(p) \cap U$ and $\hat{x}(p_k) \in Q(p_k) \cap U$, there exist $x_k \in Q(p_k)$ and $x'_k \in Q(p)$ such that $x_k \to \hat{x}(p)$ and also $|x'_k - \hat{x}(p_k)| \to 0$ as $k \to \infty$. From optimality,

$$\varphi(x_k) \geq \varphi(\hat{x}(p_k)) \quad \text{and} \quad \varphi(x'_k) \geq \varphi(\hat{x}(p)).$$

These two inequalities, combined with the continuity of φ, give us

$$\varphi(\hat{x}(p_k)) \to \varphi(\hat{x}(p)) \text{ as } k \to \infty.$$

Hence any limit of a sequence of minimizers $\hat{x}(p_k)$ is a minimizer for p, which implies that $\hat{x}$ is continuous at p. □

We are now ready to complete the proof of 4I.2.

Proof of Theorem 4I.2 (final part). We already know from the argument displayed after the statement of the theorem that metric regularity of the mapping F in (15) at $\bar{z} = (\bar{x}, \bar{y})$ for 0 is equivalent to the metric regularity of the mapping L in (16) at 0 for 0, and the same holds for the strong metric regularity. Our next step is to associate with the mapping L the function

$$(x,y) \mapsto R(x,y) = \begin{pmatrix} Ax + \sum_{i=1}^{s} b_i y_i + \sum_{i=s+1}^{m} b_i y_i^+ \\ -\langle b_1, x\rangle + y_1 \\ \vdots \\ -\langle b_s, x\rangle + y_s \\ -\langle b_{s+1}, x\rangle + y_{s+1}^- \\ \vdots \\ -\langle b_m, x\rangle + y_m^- \end{pmatrix}, \tag{21}$$

from $\mathbb{R}^n \times \mathbb{R}^m$ to itself, where b_i are the rows of the matrix B and where we let $y^+ = \max\{0, y\}$ and $y^- = y - y^+$.

In the beginning of this section, we showed that the strong metric regularity of the mapping H in (4) is equivalent to the same property for the mapping G in (3). The same argument works for the mappings R and L Indeed, for a given $(v,u) \in \mathbb{R}^n \times \mathbb{R}^m$, let $(x,y) \in R^{-1}(v,u)$. Then for $z_i = y_i^+$, $i = s+1, \ldots, q$, we have $(x,z) \in L^{-1}(v,u)$. Indeed, for each $i = s+1, \ldots, m$, if $y_i \le 0$, then $u_i + \langle b_i, x\rangle = y_i^- \le 0$ and $(u_i + \langle b_i, x\rangle) y_i^+ = 0$; otherwise $u_i + \langle b_i, x\rangle = y_i^- = 0$. Conversely, if $(x,z) \in L^{-1}(v,u)$ then for

$$y_i = \begin{cases} z_i & \text{if } z_i > 0, \\ u_i + \langle b_i, x\rangle & \text{if } z_i = 0, \end{cases} \tag{22}$$

we obtain $(x,y) \in R^{-1}(v,u)$. Thus, in order to achieve our goal for the mapping L, we can focus on the same question for the equivalence between metric regularity and strong metric regularity for the *function* R in (21).

Suppose that R is metrically regular but not strongly metrically regular. Then, from 4I.3 and the equivalence between regularity properties of L and R, R is strongly subregular. Consequently, its inverse R^{-1} has both the Aubin property and the isolated calmness property, both at 0 for 0. In particular, since R is positively homogeneous and has closed graph, for each w sufficiently close to 0, $R^{-1}(w)$ is a compact set contained in an arbitrarily small ball around 0. Let $a > 0$. For every $w \in a\mathbb{B}$ the problem

$$\text{minimize } y_m \;\text{ subject to } (x,y) \in R^{-1}(w) \tag{23}$$

has a solution $(x(w), y(w))$ which, from the property of R^{-1} mentioned just above (22), has a nonempty-valued graphical localization around 0 for 0. According to Lemma 4I.5, this localization is either a continuous function or a multi-valued mapping. If it is a continuous function, Lemma 4I.3 implies that R^{-1} has a continuous single-valued localization around 0 for 0. But then, since R^{-1} has the Aubin property at that point, we conclude that R must be strongly metrically regular, which contradicts the assumption made. Hence, any graphical localization of the solution mapping of (22) is multi-valued. Thus, there exists a sequence $z^k = (v^k, u^k) \to 0$ and two sequences $(x^k, y^k) \to 0$ and $(\xi^k, \eta^k) \to 0$, whose k-terms are both in $R^{-1}(z^k)$, such that the m-components of y^k and η^k are the same, $y^k_m = \eta^k_m$, but $(x^k, y^k) \neq (\xi^k, \eta^k)$ for all k. Remove from y^k the final component y^k_m and denote the remaining vector by y^k_{-m}. Do the same for η^k. Then (x^k, y^k_{-m}) and (ξ^k, η^k_{-m}) are both solutions of

$$\begin{aligned}
v^k - b_m y^k_m &= Ax^k + \sum_{i=1}^{s} b_i y_i + \sum_{i=s+1}^{m-1} b_i y_i^+ \\
u^k_1 &= -\langle b_1, x\rangle + y_1 \\
&\vdots \\
u^k_s &= -\langle b_s, x\rangle + y_s \\
u^k_{s+1} &= -\langle b_{s+1}, x\rangle + y^-_{s+1} \\
&\vdots \\
u^k_{m-1} &= -\langle b_{m-1}, x\rangle + y^-_{m-1}\,.
\end{aligned}$$

This relation concerns the reduced mapping R_{-m} with $m-1$ vectors b_i, and accordingly a vector y of dimension $m-1$:

$$R_{-m}(x,y) = \begin{pmatrix} Ax + \sum_{i=1}^{s} b_i y_i + \sum_{i=s+1}^{m-1} b_i y_i^+ \\ -\langle b_1, x\rangle + y_1 \\ \vdots \\ -\langle b_s, x\rangle + y_s \\ -\langle b_{s+1}, x\rangle + y^-_{s+1} \\ \vdots \\ -\langle b_{m-1}, x\rangle + y^-_{m-1} \end{pmatrix}.$$

We obtain that the mapping R_{-m} cannot be strongly metrically regular because for the same value $z^k = (v^k - b_m y^k_m, u^k_{-m})$ of the parameter arbitrarily close to 0, we have two solutions (x^k, y^k_{-m}) and (ξ^k, η^k_{-m}). On the other hand, R_{-m} is metrically regular as a submapping of R; this follows e.g. from the characterization in (18) for metric regularity of the mapping L, which is equivalent to the metric regularity of R_{-m} if we choose J_3' in (18) always to include the index m.

Thus, our assumption for the mapping R leads to a submapping R_{-m}, of one less variable y associated with the "inequality" part of L, for which the same assumption is satisfied. By proceeding further with "deleting inequalities" we will end up with no inequalities at all, and then the mapping L becomes just the linear mapping represented by the square matrix

$$\begin{pmatrix} A & B_0^{\top} \\ -B_0 & 0 \end{pmatrix}.$$

But this linear mapping cannot be simultaneously metrically regular and not strongly metrically regular, because a square matrix of full rank is automatically nonsingular. Hence, our assumption that the mapping R is metrically regular and not strongly regular is void. □

Commentary

Graphical derivatives of set-valued mappings were introduced by Aubin [1981]; for more, see Aubin and Frankowska [1990]. The material in Section 4B is mainly from Dontchev, Quincampoix and Zlateva [2006], where results of Aubin and Frankowska [1987, 1990] were used.

The statement 4B.5 of the Ekeland principle is from Ekeland [1990]. A detailed presentation of this principle along with various forms and extensions is given in Borwein and Zhu [2005]. The proof of the classical implicit function theorem 1A.1 given at the end of Section 4B is close, but not identical, to that in Ekeland [1990].

The coderivative criterion in 4C.2 goes back to the early works of Ioffe [1981, 1984], Kruger [1982] and Mordukhovich [1984]. Theorem 4C.3 is from Dontchev and Frankowska [2013]; for a predecessor see Frankowska and Quincampoix [2010]. Lemma 4C.4 is a particular case of results proved in Aubin and Frankowska [1995] and Dal Maso and Frankowska [2000]. Broad reviews of the role of coderivatives in variational analysis are given in Rockafellar and Wets [1998] and Mordukhovich [2006]. Detailed treatments of the topic of generalized differentiation are given in the recent books Clarke [2013] and Penot [2013].

The strict graphical derivative was introduced by Bouligand, but may well go back to Peano, under the name *paratingent derivative*; for more about its history see Aubin and Frankowska [1990] and Dolecki and Greco [2011]. Theorem 4D.1 is from Dontchev and Frankowska [2013]. It sharpens Theorem 9.54 in Rockafellar and Wets [1998] and also Lemma 3.1 in Klatte and Kummer [2002]. Theorem 4D.4 is from Izmailov [2013], but we supply it in Section 6F with a different proof using ideas from Páles [1974]. Theorem 4D.6 is from Kummer [1991], see also Klatte and Kummer [2002] and Páles [1997].

The derivative criterion for metric subregularity in 4E.1 was obtained by Rockafellar [1989], but the result itself was embedded in a proof of a statement requiring additional assumptions. The necessity without those assumptions was later noted in King and Rockafellar [1992] and in the case of sufficiency by Levy [1996]. The statement and the proof of 4E.1 are from Dontchev and Rockafellar [2004].

Sections 4F and 4G give a unified presentation of various results scattered in the literature. Section 4H is partially based on Dontchev and Rockafellar [1996] but also contains some new results. The critical superface lemma 4H.2 is a particular case of Lemma 3.5 in Robinson [1984]; see also Theorem 5.6 in Rockafellar [1989].

Theorem 4I.1 originates from Robinson [1980], see the commentary to Section 2. The proof given here uses some ideas from Kojima [1980] and Jongen et al. [1987]. Theorem 4I.2 is a particular case of Theorem 3 in Dontchev and Rockafellar [1996] which in turn is based on a deeper result in Robinson [1992], see also Ralph [1993]. The presented proof uses a somewhat modified version of a reduction argument from the book Klatte and Kummer [2002], Section 7.5. For more recent results in this direction, see Klatte and Kummer [2013].

Chapter 5
Metric Regularity in Infinite Dimensions

The theme of this chapter has origins in the early days of functional analysis and the Banach open mapping theorem, which concerns continuous linear mappings from one Banach space to another. The graphs of such mappings are subspaces of the product of the two Banach spaces, but remarkably much of the classical theory extends to set-valued mappings whose graphs are convex sets or cones instead of subspaces. Openness connects up then with metric regularity and interiority conditions on domains and ranges, as seen in the Robinson–Ursescu theorem. Infinite-dimensional inverse function theorems and implicit function theorems due to Lyusternik, Graves, and Bartle and Graves can be derived and extended. Banach spaces can even be replaced to some degree by more general metric spaces.

Before proceeding we review some notation and terminology. Already in the first section of Chapter 1 we stated the contraction mapping principle in metric spaces. Given a set X, a function $\rho : X \times X \to I\!\!R_+$ is said to be a *metric* in X when

(i) $\rho(x,y) = 0$ if and only if $x = y$;
(ii) $\rho(x,y) = \rho(y,x)$;
(iii) $\rho(x,y) \le \rho(x,z) + \rho(z,y)$ (triangle inequality).

A set X equipped with a metric ρ is called a *metric space*. In a metric space, a sequence $\{x_k\}$ is called a *Cauchy sequence* if for every $\varepsilon > 0$ there exists $n \in I\!\!N$ such that $\rho(x_k, x_j) < \varepsilon$ for all $k, j > n$. A metric space is *complete* if every Cauchy sequence converges to an element of the space. Any closed set in a Euclidean space is a complete metric space with the metric $\rho(x,y) = |x - y|$.

A *linear (vector) space* over the reals is a set X in which addition and scalar multiplication are defined obeying the standard algebraic laws of commutativity, associativity and distributivity. A linear space X with elements x is *normed* if it is furnished with a real-valued expression $\|x\|$, called the *norm* of x, having the properties

(i) $\|x\| \ge 0$ and $\|x\| = 0$ if and only if $x = 0$;
(ii) $\|\alpha x\| = |\alpha| \, \|x\|$ for $\alpha \in I\!\!R$;
(iii) $\|x + y\| \le \|x\| + \|y\|$.

Any normed space is a metric space with the metric $\rho(x,y) = \|x - y\|$. A complete normed vector space is called a *Banach space*. On a finite-dimensional space, all norms are equivalent, but when we refer specifically to $\mathbb{R}^n$ we ordinarily have in mind the Euclidean norm denoted by $|\cdot|$. Regardless of the particular norm being employed in a Banach space, the closed unit ball for that norm will be denoted by $\mathbb{B}$, and the distance from a point x to a set C will be denoted by $d(x,C)$, and so forth.

As in finite dimensions, a function A acting from a Banach space X into a Banach space Y is called a *linear mapping* if $\operatorname{dom} A = X$ and $A(\alpha x + \beta y) = \alpha Ax + \beta Ay$ for all $x, y \in X$ and all scalars α and β. The range of a linear mapping A from X to Y is always a subspace of Y, but it might not be a closed subspace, even if A is continuous. A linear mapping $A : X \to Y$ is *surjective* if $\operatorname{rge} A = Y$ and *injective* if $\ker A = \{0\}$.

Although in finite dimensions a linear mapping $A : X \to Y$ is automatically continuous, this fails in infinite dimensions; neither does surjectivity of A when $X = Y$ necessarily yield invertibility, in the sense that A^{-1} is single-valued. However, if A is continuous at any one point of X, then it is continuous at every point of X. That, moreover, is equivalent to A being *bounded*, in the sense that A carries bounded subsets of X into bounded subsets of Y, or what amounts to the same thing due to linearity, the image of the unit ball in X is included in some multiple of the unit ball in Y, i.e., the value

$$\|A\| = \sup_{\|x\| \le 1} \|Ax\|$$

is finite. This expression defines the *operator norm* on the space $\mathscr{L}(X,Y)$, consisting of all continuous linear mappings $A : X \to Y$, which is then another Banach space.

Special and important in this respect is the Banach space $\mathscr{L}(X,\mathbb{R})$, consisting of all linear and continuous real-valued functions on X. It is the space *dual* to X, symbolized by X^*, and its elements are typically denoted by x^*; the value that an $x^* \in X^*$ assigns to an $x \in X$ is written as $\langle x^*, x\rangle$. The dual of the Banach space X^* is the bidual X^{**} of X; when every function $x^{**} \in X^{**}$ on X^* can be represented as $x^* \mapsto \langle x^*, x\rangle$ for some $x \in X$, the space X is called *reflexive*. This holds in particular when X is a *Hilbert* space with $\langle x, y\rangle$ as its *inner product*, and each $x^* \in X^*$ corresponds to a function $x \mapsto \langle x, y\rangle$ for some $y \in X$, so that X^* can be identified with X itself.

Another thing to be mentioned for a pair of Banach spaces X and Y and their duals X^* and Y^* is that any $A \in \mathscr{L}(X,Y)$ has an *adjoint* $A^* \in \mathscr{L}(Y^*,X^*)$ such that $\langle Ax, y^*\rangle = \langle x, A^*y^*\rangle$ for all $x \in X$ and $y^* \in Y^*$. Furthermore, $\|A^*\| = \|A\|$. A generalization of this to set-valued mappings having convex cones as their graphs will be seen later.

In fact most of the definitions, and even many of the results, in the preceding chapters will carry over with hardly any change, the major exception being results with proofs which truly depended on the compactness of $\mathbb{B}$. Our initial task, in Section 5A, will be to formulate various facts in this broader setting while coordinating them with classical theory. In the remainder of the chapter, we present inverse and implicit mapping theorems with metric regularity and strong metric regularity in abstract spaces. Parallel results for metric subregularity are not considered.

5A. Positively Homogeneous Mappings

We begin this section with a review of basic regularity properties introduced in Chapter 3, now in infinite dimensions. Then we focus on the class of positively homogeneous mapping, which are a natural generalization of linear mappings.

At the end of Chapter 1 we introduced the concept of openness of a function and presented a Jacobian criterion for openness, but did not elaborate further. We now return to this property in the broader context of Banach spaces X and Y. A function $f : X \to Y$ is called *open* at $\bar{x}$ if $\bar{x} \in \operatorname{int} \operatorname{dom} f$ and, for every neighborhood U of $\bar{x}$ in X, the set $f(U)$ is a neighborhood of $f(\bar{x})$ in Y. This definition extends to set-valued mappings $F : X \rightrightarrows Y$:

Openness. *A mapping $F : X \rightrightarrows Y$ is said to be open at $\bar{x}$ for $\bar{y}$ if $\bar{y} \in F(\bar{x})$*

$$(1) \qquad \begin{array}{l} \bar{x} \in \operatorname{int} \operatorname{dom} F \text{ and for every neighborhood } U \text{ of } \bar{x}, \\ \text{the set } F(U) = \bigcup_{x \in U} F(x) \text{ is a neighborhood of } \bar{y}. \end{array}$$

We will also be concerned with another property, introduced for mappings $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ in 3E but likewise directly translatable to mappings $F : X \rightrightarrows Y$:

Metric regularity. *A mapping $F : X \rightrightarrows Y$ is said to be metrically regular at $\bar{x}$ for $\bar{y}$ when $\bar{y} \in F(\bar{x})$, the set $\operatorname{gph} F$ is locally closed at $(\bar{x}, \bar{y})$, and*

$$(2) \qquad \begin{array}{l} \text{there exists } \kappa \geq 0 \text{ with neighborhoods } U \text{ of } \bar{x} \text{ and } V \text{ of } \bar{y} \\ \text{such that } d(x, F^{-1}(y)) \leq \kappa d(y, F(x)) \text{ for all } (x, y) \in U \times V. \end{array}$$

As before, the infimum of all such κ associated with choices of U and V is denoted by $\operatorname{reg}(F; \bar{x} \,|\, \bar{y})$ and called the modulus of metric regularity of F at $\bar{x}$ for $\bar{y}$.

The classical Banach open mapping theorem addresses linear mappings. There are numerous versions of it available in the literature; we provide the following formulation:

Theorem 5A.1 (Banach open mapping theorem). *For any $A \in \mathscr{L}(X, Y)$ the following properties are equivalent:*

(a) *A is surjective;*
(b) *A is open (at every point);*
(c) $0 \in \operatorname{int} A(\operatorname{int} \mathbb{B})$;
(d) *there is a $\kappa > 0$ such that for all $y \in Y$ there exists $x \in X$ with $Ax = y$ and $\|x\| \leq \kappa \|y\|$.*

This theorem will be derived in Section 5B from a far more general result about set-valued mappings F than just linear mappings A. Our immediate interest lies in connecting it with the ideas in previous chapters, so as to shed light on where we have arrived and where we are going.

The first observation to make is that (d) of Theorem 5A.1 is the same as the existence of a $\kappa > 0$ such that $d(0, A^{-1}(y)) \leq \kappa \|y\|$ for all y. Clearly (d) does imply this, but the converse holds also by passing to a slightly higher κ if need be. But the linearity of A can also be brought in. For $x \in X$ and $y \in Y$ in general, we have $d(x, A^{-1}(y)) = d(0, A^{-1}(y) - x)$, and since $z \in A^{-1}(y) - x$ corresponds to $A(x+z) = y$, we have $d(0, A^{-1}(y) - x) = d(0, A^{-1}(y - Ax)) \leq \kappa \|y - Ax\|$. Thus, (d) of Theorem 5A.1 is actually equivalent to:

(3) there exists $\kappa > 0$ such that $d(x, A^{-1}(y)) \leq \kappa d(y, Ax)$ for all $x \in X$, $y \in Y$.

Obviously this is the same as the metric regularity property in (2) as specialized to A, with the local character property becoming global through the arbitrary scaling made available because $A(\lambda x) = \lambda Ax$. In fact, due to linearity, metric regularity of A with respect to any pair $(\bar{x}, \bar{y})$ in its graph is identical to metric regularity with respect to $(0, 0)$, and the same modulus of metric regularity prevails everywhere. We can simply denote this modulus by $\operatorname{reg} A$ and use the formula that

$$\operatorname{reg} A = \sup_{\|y\| \leq 1} d(0, A^{-1}(y)) \quad \text{for } A \in \mathscr{L}(X, Y). \tag{4}$$

What we see then is that the condition

(e) *A is metrically regular (everywhere):* $\operatorname{reg} A < \infty$

could be added to the equivalences in Theorem 5A.1 as a way of relating it to the broader picture we now have of the subject of openness.

Corollary 5A.2 (invertibility of linear mappings)**.** *If a continuous linear mapping $A : X \to Y$ is both surjective and injective, then its inverse is a continuous linear mapping $A^{-1} : Y \to X$ with $\|A^{-1}\| = \operatorname{reg} A$.*

Proof. When A is both surjective and injective, then A^{-1} is single-valued everywhere and linear. Observe that the right side of (4) reduces to $\|A^{-1}\|$. The finiteness of $\|A^{-1}\|$ corresponds to A^{-1} being bounded, hence continuous. □

It is worth noting also that if the range of $A \in \mathscr{L}(X, Y)$ is a *closed* subspace Y' of Y, then Y' is a Banach space in its own right, and the facts we have recorded can be applied to A as a surjective mapping from X to Y'.

A result is available for set-valued mappings $F : X \rightrightarrows Y$ which has close parallels to the version of Theorem 5A.1 with (e) added, although it misses some aspects. This result corresponds in the case of $X = \mathbb{R}^n$ and $Y = \mathbb{R}^n$ to Theorems 3E.7 and 3E.9, where an equivalence was established between metric regularity, *linear* openness, and the inverse mapping having the Aubin property. For completeness we give the definitions.

Aubin property. *A mapping $S : Y \rightrightarrows X$ is said to have the Aubin property at $\bar{y}$ for $\bar{x}$ when $\bar{x} \in S(\bar{y})$, the set $\operatorname{gph} S$ is locally closed at $(\bar{y}, \bar{x})$, and*

(5) *there is a* $\kappa \geq 0$ *with neighborhoods* U *of* $\bar{x}$ *and* V *of* $\bar{y}$ *such that* $e(S(y)\cap U, S(y')) \leq \kappa\|y-y'\|$ *for all* $y,y' \in V$,

where the excess $e(C,D)$ *is defined in Section 3A. The infimum of all* κ *in (5) over various choices of* U *and* V *is the modulus* $\operatorname{lip}(S;\bar{y}\,|\,\bar{x})$.

Linear openness. *A mapping* $F: X \rightrightarrows Y$ *is said to be linearly open at* $\bar{x}$ *for* $\bar{y}$ *when* $\bar{y} \in F(\bar{x})$, *the set* $\operatorname{gph} F$ *is locally closed at* $(\bar{x},\bar{y})$, *and*

(6) *there is a* $\kappa > 0$ *along with neighborhoods* U *of* $\bar{x}$ *and* V *of* $\bar{y}$ *such that* $F(x+\kappa r \operatorname{int} \mathbb{B}) \supset \big[F(x)+r\operatorname{int}\mathbb{B}\big]\cap V$ *for all* $x \in U$, $r > 0$.

The statements and proofs of Theorems 3E.7 and 3E.9 carry over in the obvious manner to our present setting to yield the following combined result:

Theorem 5A.3 (equivalence of metric regularity, linear openness and Aubin property)**.** *For Banach spaces* X *and* Y, *a mapping* $F: X \rightrightarrows Y$ *and a constant* $\kappa > 0$, *the following properties with respect to a pair* $(\bar{x},\bar{y}) \in \operatorname{gph} F$ *are equivalent:*

(a) F *is linearly open at* $\bar{x}$ *for* $\bar{y}$ *with constant* κ;

(b) F *is metrically regular at* $\bar{x}$ *for* $\bar{y}$ *with constant* κ;

(c) F^{-1} *has the Aubin property at* $\bar{y}$ *for* $\bar{x}$ *with constant* κ.

Moreover $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \operatorname{lip}(F^{-1};\bar{y}\,|\,\bar{x})$.

We should note here that although the same positive constant κ appear in all three properties in 5A.3(a)(b)(c), the associated neighborhoods could be different; for more on this see Section 5H.

When F is taken to be a mapping $A \in \mathscr{L}(X,Y)$, how does the content of Theorem 5A.3 compare with that of Theorem 5A.1? With linearity, the openness in 5A.1(b) comes out the same as the linear openness in 5A.3(a) and is easily seen to reduce as well to the interiority condition in 5A.1(c). On the other hand, 5A.1(d) has already been shown to be equivalent to the subsequently added property (e), to which 5A.3(b) reduces when $F = A$. From 5A.3(c), though, we get yet another property which could be added to the equivalences in Theorem 5A.1 for $A \in \mathscr{L}(X,Y)$, specifically that

(f) $A^{-1}: Y \rightrightarrows X$ *has the Aubin property at every* $\bar{y} \in Y$ *for every* $\bar{x} \in A^{-1}(\bar{y})$,

where $\operatorname{lip}(A^{-1};\bar{y}\,|\,\bar{x}) = \operatorname{reg} A$ always. This goes farther than the observation in Corollary 5A.2, which covered only single-valued A^{-1}. In general, of course, the Aubin property in 5A.3(c) turns into local Lipschitz continuity when F^{-1} is single-valued.

An important feature of Theorem 5A.1, which is not represented at all in Theorem 5A.3, is the assertion that surjectivity is sufficient, as well as necessary, for all these properties to hold. An extension of that aspect to nonlinear F will be possible, in a local sense, under the restriction that $\operatorname{gph} F$ is closed and convex. This will emerge in the next section, in Theorem 5B.4.

Another result which we now wish to upgrade to infinite dimensions is the estimation for perturbed inversion which appeared in matrix form in Corollary 1E.7

with elaborations in 1E.8. It lies at the heart of the theory of implicit functions and will eventually be generalized in more than one way. We provide it here with a direct proof (compare with 1E.8(b)).

Lemma 5A.4 (estimation for perturbed inversion)**.** *Let $A \in \mathscr{L}(X,Y)$ be invertible. Then for every $B \in \mathscr{L}(X,Y)$ with $\|A^{-1}\|{\cdot}\|B\| < 1$ one has*

$$\|(A+B)^{-1}\| \leq \frac{\|A^{-1}\|}{1-\|A^{-1}\|\|B\|}. \tag{7}$$

Proof. Let $C = BA^{-1}$; then $\|C\| < 1$ and hence $\|C^n\| \leq \|C\|^n \to 0$ as $n \to \infty$. Also, the elements

$$S_n = \sum_{i=0}^{n} C^i \quad \text{for } n = 0, 1, \ldots$$

form a Cauchy sequence in the Banach space $\mathscr{L}(X,Y)$ which therefore converges to some $S \in \mathscr{L}(X,Y)$. Observe that, for each n,

$$S_n(I-C) = I - C^{n+1} = (I-C)S_n,$$

and hence, through passing to the limit, one has $S = (I-C)^{-1}$. On the other hand

$$\|S_n\| \leq \sum_{i=0}^{n} \|C^i\| \leq \sum_{i=0}^{\infty} \|C\|^i = \frac{1}{1-\|C\|}.$$

Thus, we obtain

$$\|(I-C)^{-1}\| \leq \frac{1}{1-\|C\|}.$$

All that remains is to bring in the identity $(I-C)A = A - B$ and the inequality $\|C\| \leq \|A^{-1}\|\|B\|$, and to observe that the sign of B does not matter. □

Note that, with the conventions $\infty \cdot 0 = 0$, $1/0 = \infty$ and $1/\infty = 0$, Lemma 5A.4 also covers the cases $\|A^{-1}\| = \infty$ and $\|A^{-1}\|{\cdot}\|B\| = 1$.

Exercise 5A.5. *Derive Lemma 5A.4 from the contraction mapping principle 1A.2.*

Guide. Setting $a = \|A^{-1}\|$, choose $B \in \mathscr{L}(X,Y)$ with $\|B\| < \|A^{-1}\|^{-1}$ and $y \in Y$ with $\|y\| \leq 1 - a\|B\|$. Show that the mapping $\Phi : x \mapsto A^{-1}(y - Bx)$ satisfies the conditions in 1A.2 with $\lambda = a\|B\|$ and hence, there is a unique $x \in a\mathbb{B}$ such that $x = A^{-1}(y - Bx)$, that is $(A+B)x = y$. Thus, $A+B$ is invertible. Moreover $\|x\| = \|(A+B)^{-1}(y)\| \leq a$ for every $y \in (1 - a\|B\|)\mathbb{B}$, which implies that

$$\|(A+B)^{-1}z\| \leq \frac{\|A^{-1}\|}{1-\|A^{-1}\|\|B\|} \quad \text{for every } z \in \mathbb{B}.$$

This yields (7). □

Exercise 5A.6. *Let $C \in \mathscr{L}(X,Y)$ satisfy $\|C\| < 1$. Prove that*

$$\|(I-C)^{-1}-I-C\| \leq \frac{\|C\|^2}{1-\|C\|}.$$

Guide. Use the sequence of mappings S_n in the proof of 5A.4 and observe that

$$\|S_n - I - C\| = \|C^2 + C^3 + \cdots + C^n\| \leq \frac{\|C\|^2}{1-\|C\|}.$$

□

Positive homogeneity. A mapping $H : X \rightrightarrows Y$ whose graph is a cone in $X \times Y$ is called *positively homogeneous*. In infinite dimensions such mappings have properties similar to those developed in finite dimensions in Section 4A, but with some complications. Outer and inner norms are defined for such mappings H as in Section 4A, but it is necessary to take into account the possible variety of underlying norms on X and Y in place of just the Euclidean norm earlier:

(8) $$\|H\|^+ = \sup_{\|x\|\leq 1} \sup_{y \in H(x)} \|y\|, \qquad \|H\|^- = \sup_{\|x\|\leq 1} \inf_{y \in H(x)} \|y\|.$$

When $\operatorname{dom} H = X$ and H is single-valued, these two expressions agree. For $H = A \in \mathscr{L}(X,Y)$, they reduce to $\|A\|$.

The inverse H^{-1} of a positively homogeneous mapping H is another positively homogeneous mapping, and its outer and inner norms are therefore available also. The elementary relationships in Propositions 4A.5 and 4A.6 have the following update.

Proposition 5A.7 (outer and inner norms)**.** *The inner norm of a positively homogeneous mapping $H : X \rightrightarrows Y$ satisfies*

$$\|H\|^- \; = \; \inf\big\{\, \kappa \in (0,\infty) \,\big|\, H(x) \cap \kappa I\!B \neq \emptyset \;\; \text{for all } x \in I\!B \,\big\},$$

so that, in particular,

(9) $$\|H\|^- \; < \; \infty \quad \Longrightarrow \quad \operatorname{dom} H = X.$$

In parallel, the outer norm satisfies

$$\|H\|^+ \; = \; \inf\big\{\, \kappa \in (0,\infty) \,\big|\, H(I\!B) \subset \kappa I\!B \,\big\} \; = \; \sup_{\|y\|=1} \frac{1}{d(0, H^{-1}(y))},$$

and we have

(10) $$\|H\|^+ \; < \; \infty \quad \Longrightarrow \quad H(0) = \{0\},$$

with this implication becoming an equivalence when H has closed graph and $\dim X < \infty$.

The equivalence generally fails in (10) when $\dim X = \infty$ because of the lack of compactness then (with respect to the norm) of the ball $I\!B$ in X.

An extension of Lemma 5A.4 to possibly set-valued mappings that are positively homogeneous is now possible in terms of the outer norm. Recall that for a positively homogeneous $H : X \rightrightarrows Y$ and a linear $B : X \to Y$ we have $(H+B)(x) = H(x) + Bx$ for every $x \in X$.

Theorem 5A.8 (inversion estimate for the outer norm)**.** *Let $H : X \rightrightarrows Y$ be positively homogeneous with $\|H^{-1}\|^+ < \infty$. Then for every $B \in \mathscr{L}(X,Y)$ with the property that $\|H^{-1}\|^+ \cdot \|B\| < 1$, one has*

$$\|(H+B)^{-1}\|^+ \leq \frac{\|H^{-1}\|^+}{1 - \|H^{-1}\|^+ \|B\|}. \tag{11}$$

Proof. Having $\|H^{-1}\|^+ = 0$ is equivalent to having $\operatorname{dom} H = \{0\}$; in this case, since $\emptyset + y = \emptyset$ for any y, we get $\|(H+B)^{-1}\|^+ = 0$ for any $B \in \mathscr{L}(X,Y)$ as claimed. Suppose therefore instead that $0 < \|H^{-1}\|^+ < \infty$. If the estimate (11) is false, there is some $B \in \mathscr{L}(X,Y)$ with $\|B\| < [\|H^{-1}\|^+]^{-1}$ such that $\|(H+B)^{-1}\|^+ > ([\|H^{-1}\|^+]^{-1} - \|B\|)^{-1}$. In particular $B \neq 0$ then, and by definition there must exist $y \in I\!B$ and $x \in (H+B)^{-1}(y)$ such that $\|x\| > ([\|H^{-1}\|^+]^{-1} - \|B\|)^{-1}$, which is the same as

$$\frac{1}{\|x\|^{-1} + \|B\|} > \|H^{-1}\|^+. \tag{12}$$

But then $y - Bx \in H(x)$ and

$$\|y - Bx\| \leq \|y\| + \|B\|\|x\| \leq 1 + \|B\|\|x\|. \tag{13}$$

If $y = Bx$ then $0 \in H(x)$, so (10) yields $x = 0$, a contradiction. Hence $\alpha := \|y - Bx\|^{-1} > 0$, and due to the positive homogeneity of H we have $(\alpha x, \alpha(y - Bx)) \in \operatorname{gph} H$ and $\alpha\|y - Bx\| = 1$, which implies, by definition,

$$\|H^{-1}\|^+ \geq \frac{\|x\|}{\|y - Bx\|}.$$

Combining this inequality with (12) and (13), we get

$$\|H^{-1}\|^+ \geq \frac{\|x\|}{\|y - Bx\|} \geq \frac{\|x\|}{1 + \|B\|\|x\|} = \frac{1}{\|x\|^{-1} + \|B\|} > \|H^{-1}\|^+.$$

This is impossible, and the proof is at its end. $\square$

A corresponding extension of Lemma 5A.4 in terms of the inner norm will be possible later, in Section 5C.

Normal cones and polarity. For a closed, convex cone $K \subset X$, the *polar* of K is the subset K^* of the dual space X^* defined by

$$K^* = \left\{ x^* \in X^* \,\middle|\, \langle x, x^* \rangle \le 0 \;\text{ for all }\; x \in K \right\}.$$

It is a closed convex cone in X^* from which K can be recovered as the polar $(K^*)^*$ of K^* in the sense that

$$K = \left\{ x \in X \,\middle|\, \langle x, x^* \rangle \le 0 \;\text{ for all }\; x^* \in K^* \right\}.$$

For any set C in a Banach space X and any point $x \in C$, the *tangent cone* $T_C(x)$ at a point $x \in C$ is defined as in 2A to consist of all limits v of sequences $(1/\tau^k)(x^k - x)$ with $x^k \to x$ in C and $\tau^k \searrow 0$. When C is convex, $T_C(x)$ has an equivalent description as the closure of the convex cone consisting of all vectors $\lambda(x' - x)$ with $x' \in C$ and $\lambda > 0$.

In infinite dimensions, the *normal cone* $N_C(x)$ to C at x can be introduced in various ways that extend the general definition given for finite dimensions in 4C, but we will only be concerned with the case of convex sets C. For that case, the special definition in 2A suffices with only minor changes caused by the need to work with the dual space X^* and the pairing $\langle x, x^* \rangle$ between X and X^*. Namely, $N_C(x)$, for $x \in C$, consists of all $x^* \in X^*$ such that

$$\langle x' - x, x^* \rangle \le 0 \;\text{ for all } x' \in C.$$

Equivalently, through the alternative description of $T_C(x)$ for convex C, the normal cone $N_C(x)$ is the polar $T_C(x)^*$ of the tangent cone $T_C(x)$. It follows that $T_C(x)$ is in turn the polar cone $N_C(x)^*$.

As earlier, $N_C(x)$ is taken to be the empty set when $x \notin C$ so as to get a set-valued mapping N_C defined for all x, but this *normal cone mapping* now goes from X to X^* instead of from the underlying space into itself (except in the case of a Hilbert space, where X^* can be identified with X as recalled above). A generalized equation of the form

$$f(x) + N_C(x) \ni 0 \;\text{ for a function }\; f : X \to X^*$$

is again a *variational inequality*. Such generalized equations are central, for instance, to many applications involving differential or integral operators, especially in a Hilbert space framework.

Exercise 5A.9 (normals to cones)**.** *Show that for a closed convex cone $K \subset X$ and its polar $K^* \subset X^*$, one has*

$$x^* \in N_K(x) \iff x \in K,\; x^* \in K^*,\; \langle x, x^* \rangle = 0.$$

Exercise 5A.10 (linear variational inequalities on cones)**.** *Let $H(x) = Ax + N_K(x)$ for $A \in \mathscr{L}(X, X^*)$ and a closed, convex cone $K \subset X$. Show that H is positively homogeneous with closed graph, but this graph is not convex unless K is a subspace of X in which case the graph is a subspace of $X \times X^*$.*

5B. Mappings with Closed Convex Graphs

For X and Y Banach spaces, and for any mapping $F : X \rightrightarrows Y$ with convex graph, the sets $\operatorname{dom} F$ and $\operatorname{rge} F$, as the projections of $\operatorname{gph} F$ on the spaces X and Y, are convex sets as well. When $\operatorname{gph} F$ is closed, these sets can fail to be closed (a famous example being the case where F is a "closed" linear mapping from X to Y with domain dense in X). However, if either of them has nonempty interior, or even nonempty "core" (in the sense about to be explained), there are highly significant consequences for the behavior of F. This section is dedicated to developing such consequences for properties like openness and metric regularity, but we begin with some facts that are more basic.

The *core* of a set $C \subset X$ is defined by

$$\operatorname{core} C = \left\{ x \,\middle|\, \forall w \in X \;\; \exists \varepsilon > 0 \text{ such that } x + tw \in C \text{ when } 0 \leq t \leq \varepsilon \right\}.$$

A set C is called *absorbing* if $0 \in \operatorname{core} C$. Obviously $\operatorname{core} C \supset \operatorname{int} C$ always, but there are circumstances where necessarily $\operatorname{core} C = \operatorname{int} C$. It is elementary that this holds when C is convex with $\operatorname{int} C \neq \emptyset$, but more attractive is the potential of using the purely algebraic test of whether a point x belongs to $\operatorname{core} C$ to confirm that $x \in \operatorname{int} C$ without first having to establish that $\operatorname{int} C \neq \emptyset$. Most importantly for our purposes here,

(1) *for a closed convex subset C of a Banach space,* $\operatorname{core} C = \operatorname{int} C$.

An equivalent statement, corresponding to how this fact is often recorded in functional analysis, is that if C is a closed convex set which is absorbing, then C must be a neighborhood of 0. Through the observation already made about convexity, the confirmation of this comes down to establishing that C has nonempty interior. That can be deduced from the Baire category theorem, according to which the union of a sequence of nowhere dense subsets of a complete metric space cannot cover the whole space. If $\operatorname{int} C$ were empty, the closed sets nC for $n = 1, 2, \dots$ would be nowhere dense with the entire Banach space as their union, but this is impossible.

We demonstrate now that, for some of the convex sets central to the study of closed convex graphs on which we are embarking, the core and interior coincide even without closedness.

Theorem 5B.1 (interiority criteria for domains and ranges)**.** *For any mapping $F : X \rightrightarrows Y$ with closed convex graph, one has*

(2) $\operatorname{core} \operatorname{rge} F = \operatorname{int} \operatorname{rge} F, \qquad \operatorname{core} \operatorname{dom} F = \operatorname{int} \operatorname{dom} F.$

In addition, $\operatorname{core} \operatorname{cl} \operatorname{rge} F = \operatorname{int} \operatorname{cl} \operatorname{rge} F$ *and* $\operatorname{core} \operatorname{cl} \operatorname{dom} F = \operatorname{int} \operatorname{cl} \operatorname{dom} F$*, where moreover*

(3) $$\begin{aligned} &\operatorname{int} \operatorname{cl} \operatorname{rge} F = \operatorname{int} \operatorname{rge} F \;\; \textit{when } \operatorname{dom} F \textit{ is bounded,} \\ &\operatorname{int} \operatorname{cl} \operatorname{dom} F = \operatorname{int} \operatorname{dom} F \;\; \textit{when } \operatorname{rge} F \textit{ is bounded.} \end{aligned}$$

In particular, if $\operatorname{dom} F$ *is bounded and* $\operatorname{rge} F$ *is dense in* Y, *and then* $\operatorname{rge} F = Y$. *Likewise, if* $\operatorname{rge} F$ *is bounded and* $\operatorname{dom} F$ *is dense in* X, *and then* $\operatorname{dom} F = X$.

Proof. The equations in the line following (2) merely apply (1) to the closed convex sets $\operatorname{cl}\operatorname{rge} F$ and $\operatorname{cl}\operatorname{dom} F$. Through symmetry between F and F^{-1}, the two claims in (2) are equivalent to each other, as are the two claims in (3). Also, the claims after (3) are immediate from (3). Therefore, we only have to prove one of the claims in (2) and one of the claims in (3).

We start with the first claim in (3). Assuming that $\operatorname{dom} F$ is bounded, we work toward verifying that $\operatorname{int}\operatorname{rge} F \supset \operatorname{int}\operatorname{cl}\operatorname{rge} F$; this gives equality, inasmuch as the opposite inclusion is obvious. In fact, for this we only need to show that $\operatorname{rge} F \supset \operatorname{int}\operatorname{cl}\operatorname{rge} F$.

We choose $\tilde{y} \in \operatorname{int}\operatorname{cl}\operatorname{rge} F$; then there exists $\delta > 0$ such that $\operatorname{int}\mathbb{B}_{2\delta}(\tilde{y}) \subset \operatorname{int}\operatorname{cl}\operatorname{rge} F$. We will find a point $\tilde{x}$ such that $(\tilde{x}, \tilde{y}) \in \operatorname{gph} F$, so that $\tilde{y} \in \operatorname{rge} F$. The point $\tilde{x}$ will be obtained by means of a sequence $\{(x^k, y^k)\}$ which we now construct by induction.

Pick any $(x^0, y^0) \in \operatorname{gph} F$. Suppose we have already determined $\{(x^j, y^j)\} \in \operatorname{gph} F$ for $j = 0, 1, \ldots, k$. If $y^k = \tilde{y}$, then take $\tilde{x} = x^k$ and $(x^n, y^n) = (\tilde{x}, \tilde{y})$ for all $n = k, k+1, \ldots$; that is, after the index k the sequence is constant. Otherwise, with $\alpha^k = \delta/\|y^k - \tilde{y}\|$, we let $w^k := \tilde{y} + \alpha^k(\tilde{y} - y^k)$. Then $w^k \in \mathbb{B}_\delta(\tilde{y}) \subset \operatorname{cl}\operatorname{rge} F$. Hence there exists $v^k \in \operatorname{rge} F$ such that $\|v^k - w^k\| \le \|y^k - \tilde{y}\|/2$ and also u^k with $(u^k, v^k) \in \operatorname{gph} F$. Having gotten this far, we pick

$$(x^{k+1}, y^{k+1}) = \frac{\alpha^k}{1+\alpha^k}(x^k, y^k) + \frac{1}{1+\alpha^k}(u^k, v^k).$$

Clearly, $(x^{k+1}, y^{k+1}) \in \operatorname{gph} F$ by its convexity. Also, the sequence $\{y^k\}$ satisfies

$$\|y^{k+1} - \tilde{y}\| = \frac{\|v^k - w^k\|}{1+\alpha^k} \le \frac{1}{2}\|y^k - \tilde{y}\|.$$

If $y^{k+1} = \tilde{y}$, we take $\tilde{x} = x^{k+1}$ and $(x^n, y^n) = (\tilde{x}, \tilde{y})$ for all $n = k+1, k+2, \ldots$. If not, we perform the induction step again. As a result, we generate an infinite sequence $\{(x^k, y^k)\}$, each element of which is equal to $(\tilde{x}, \tilde{y})$ after some k or has $y^k \neq \tilde{y}$ for all k and also

$$(4) \qquad \|y^k - \tilde{y}\| \le \frac{1}{2^k}\|y^0 - \tilde{y}\| \quad \text{for all } k = 1, 2, \ldots.$$

In the latter case, we have $y^k \to \tilde{y}$. Further, for the associated sequence $\{x^k\}$ we obtain

$$\|x^{k+1} - x^k\| = \frac{\|x^k - u^k\|}{1+\alpha^k} \le \frac{\|x^k\| + \|u^k\|}{\|y^k - \tilde{y}\| + \delta}\|y^k - \tilde{y}\|.$$

Both x^k and u^k are from $\operatorname{dom} F$ and thus are bounded. Therefore, from (4), $\{x^k\}$ is a Cauchy sequence, hence (because X is a complete metric space) convergent to some $\tilde{x}$. Because $\operatorname{gph} F$ is closed, we end up with $(\tilde{x}, \tilde{y}) \in \operatorname{gph} F$, as required.

Next we address the second claim in (2), where the inclusion $\operatorname{core}\operatorname{dom} F \subset \operatorname{int}\operatorname{dom} F$ suffices for establishing equality. We must show that an arbitrarily chosen point of $\operatorname{core}\operatorname{dom} F$ belongs to $\operatorname{int}\operatorname{dom} F$, but through a translation of $\operatorname{gph} F$ we can focus without loss of generality on that point in $\operatorname{core}\operatorname{dom} F$ being 0, with $F(0) \ni 0$. Let $F_0 : X \rightrightarrows Y$ be defined by $F_0(x) = F(x) \cap \mathbb{B}$. The graph of F_0, being $[X \times \mathbb{B}] \cap \operatorname{gph} F$, is closed and convex, and we have $\operatorname{dom} F_0 \subset \operatorname{dom} F$ and $\operatorname{rge} F_0 \subset \mathbb{B}$ (bounded). The relations already established in (3) tell us that $\operatorname{int}\operatorname{cl}\operatorname{dom} F_0 = \operatorname{int}\operatorname{dom} F_0$, where $\operatorname{cl}\operatorname{dom} F_0$ is a closed convex set. By demonstrating that $\operatorname{cl}\operatorname{dom} F_0$ is absorbing, we will be able to conclude from (1) that $0 \in \operatorname{int}\operatorname{dom} F_0$, hence $0 \in \operatorname{int}\operatorname{dom} F$. It is enough actually to show that $\operatorname{dom} F_0$ itself is absorbing.

Consider any $x \in X$. We have to show the existence of $\varepsilon > 0$ such that $tx \in \operatorname{dom} F_0$ for $t \in [0, \varepsilon]$. We do know, because $\operatorname{dom} F$ is absorbing, that $tx \in \operatorname{dom} F$ for all $t > 0$ sufficiently small. Fix $t_0 > 0$ and let $y_0 \in F(t_0 x)$; then for $y = y_0/t_0$ we have $t_0(x,y) \in \operatorname{gph} F$. The pair $t(x,y) = (tx, ty)$ belongs then to $\operatorname{gph} F$ for all $t \in [0, t_0]$ through the convexity of $\operatorname{gph} F$ and our arrangement that $(0,0) \in \operatorname{gph} F$. Take $\varepsilon > 0$ small enough that $\varepsilon\|y\| \le 1$. Then for $t \in [0, \varepsilon]$ we have $\|ty\| = t\|y\| \le 1$, giving us $ty \in F(tx) \cap \mathbb{B}$ and therefore $tx \in \operatorname{dom} F_0$, as required. □

Regularity properties will now be explored. The property of a mapping $F : X \rightrightarrows Y$ being open at $\bar{x}$ for $\bar{y}$, as extended to Banach spaces X and Y in 5A(1), can be restated equivalently in a manner that more closely resembles the linear openness property defined in 5A(6):

(5) *for every* $a > 0$ *there exists* $b > 0$ *such that* $F(\bar{x} + a \operatorname{int} \mathbb{B}) \supset \bar{y} + b \operatorname{int} \mathbb{B}$.

Linear openness requires a linear scaling relationship between a and b. Under positive homogeneity, such scaling is automatic. On the other hand, an intermediate type of property holds automatically without positive homogeneity when the graph of F is convex, and it will be a stepping stone toward other, stronger, consequences of convexity.

Proposition 5B.2 (openness of mappings with convex graph). *Consider a mapping $F : X \rightrightarrows Y$ with convex graph, and let $\bar{y} \in F(\bar{x})$. Then openness of F at $\bar{x}$ for $\bar{y}$ is equivalent to the simpler condition that*

(6) *there exists* $c > 0$ *with* $F(\bar{x} + \operatorname{int} \mathbb{B}) \supset \bar{y} + c \operatorname{int} \mathbb{B}$.

Proof. Clearly, (5) implies (6). For the converse, assume (6) and consider any $a > 0$. Take $b = \min\{1, a\}c$. If $a \ge 1$, the left side of (6) is contained in the left side of (5), and hence (5) holds. Suppose therefore that $a < 1$. Let $w \in \bar{y} + b \operatorname{int} \mathbb{B}$. The point $v = (w/a) - (1-a)(\bar{y}/a)$ satisfies $\|v - \bar{y}\| = \|w - \bar{y}\|/a < b/a = c$, hence $v \in \bar{y} + c \operatorname{int} \mathbb{B}$. Then from (6) there exists $u \in \bar{x} + \operatorname{int} \mathbb{B}$ with $(u, v) \in \operatorname{gph} F$. The convexity of $\operatorname{gph} F$ implies $a(u,v) + (1-a)(\bar{x}, \bar{y}) \in \operatorname{gph} F$ and yields $av + (1-a)\bar{y} \in F(au + (1-a)\bar{x}) \subset F(\bar{x} + a \operatorname{int} \mathbb{B})$. Substituting $v = (w/a) - (1-a)(\bar{y}/a)$ in this inclusion, we see that $w \in F(\bar{x} + a \operatorname{int} \mathbb{B})$, and since w was an arbitrary point in $\bar{y} + b \operatorname{int} \mathbb{B}$, we get (5). □

The following fact bridges, for set-valued mappings with convex graphs, between condition (6) and metric regularity.

Lemma 5B.3 (metric regularity estimate)**.** *Let $F : X \rightrightarrows Y$ have convex graph containing $(\bar{x},\bar{y})$, and suppose (6) is fulfilled. Then*

$$(7)\qquad d(x,F^{-1}(y)) \;\le\; \frac{1+\|x-\bar{x}\|}{c-\|y-\bar{y}\|}d(y,F(x)) \quad \text{for all } x\in X,\; y\in \bar{y}+c\,\mathrm{int}\,\mathbb{B}.$$

Proof. We may assume that $(\bar{x},\bar{y}) = (0,0)$, since this can be arranged by translating $\mathrm{gph}\,F$ to $\mathrm{gph}\,F-(\bar{x},\bar{y})$. Then condition (6) has the simpler form

$$(8)\qquad \text{there exists } c>0 \text{ with } F(\mathrm{int}\,\mathbb{B}) \supset c\,\mathrm{int}\,\mathbb{B}.$$

Let $x\in X$ and $y\in c\,\mathrm{int}\,\mathbb{B}$. Observe that (7) is automatically true when $x\notin \mathrm{dom}\,F$ or $y\in F(x)$, so assume that $x\in\mathrm{dom}\,F$ but $y\notin F(x)$. Let $\alpha := c-\|y\|$. Then $\alpha>0$. Choose $\varepsilon\in(0,\alpha)$ and find $y'\in F(x)$ such that $\|y'-y\|\le d(y,F(x))+\varepsilon$. The point $\tilde{y} := y+(\alpha-\varepsilon)\|y'-y\|^{-1}(y-y')$ satisfies $\|\tilde{y}\|\le\|y\|+\alpha-\varepsilon = c-\varepsilon<c$, hence $\tilde{y}\in c\,\mathrm{int}\,\mathbb{B}$. By (8) there exists $\tilde{x}\in\mathrm{int}\,\mathbb{B}$ with $\tilde{y}\in F(\tilde{x})$. Let $\beta := \|y-y'\|(\alpha-\varepsilon+\|y-y'\|)^{-1}$; then $\beta\in(0,1)$. From the convexity of $\mathrm{gph}\,F$ we have

$$y = (1-\beta)y'+\beta\tilde{y}\in(1-\beta)F(x)+\beta F(\tilde{x})\subset F((1-\beta)x+\beta\tilde{x}).$$

Thus $x+\beta(\tilde{x}-x)\in F^{-1}(y)$, so $d(x,F^{-1}(y))\le\beta\|x-\tilde{x}\|$. Noting that $\|x-\tilde{x}\|\le\|x\|+\|\tilde{x}\|<\|x\|+1$ and $\beta\le(\alpha-\varepsilon)^{-1}\|y-y'\|$, we obtain

$$d(x,F^{-1}(y)) < \frac{1+\|x\|}{\alpha-\varepsilon}[d(y,F(x))+\varepsilon].$$

Letting $\varepsilon\to 0$, we finish the proof. □

Condition (6) entails in particular having $\bar{y}\in\mathrm{int\,rge}\,F$. It turns out that when the graph of F is not only convex but also closed, the converse implication holds as well, that is, $\bar{y}\in\mathrm{int\,rge}\,F$ is equivalent to (6). This is a consequence of the following theorem, which furnishes a far-reaching generalization of the Banach open mapping theorem.

Theorem 5B.4 (Robinson–Ursescu theorem)**.** *Let $F : X \rightrightarrows Y$ have closed convex graph and let $\bar{y}\in F(\bar{x})$. Then the following are equivalent:*

(a) $\bar{y}\in\mathrm{int\,rge}\,F$;

(b) *F is open at $\bar{x}$ for $\bar{y}$;*

(c) *F is metrically regular at $\bar{x}$ for $\bar{y}$.*

Proof. We first demonstrate that

$$(9)\qquad \bar{y}\in\mathrm{int}\,F(\bar{x}+\mathrm{int}\,\mathbb{B}) \;\text{ when } \bar{x}\in F^{-1}(\bar{y}) \text{ and } \bar{y}\in\mathrm{int\,rge}\,F.$$

By a translation, we can reduce to the case of $(\bar{x},\bar{y}) = (0,0)$. To conclude (9) in this setting, where $F(0)\ni 0$ and $0\in\mathrm{int\,rge}\,F$, it will be enough to show, for an arbitrary $\delta \in (0,1)$, that $0 \in \operatorname{int} F(\delta \mathbb{B})$. Define the mapping $F_\delta : X \rightrightarrows Y$ by $F_\delta(x) = F(x)$ when $x \in \delta \mathbb{B}$ but $F_\delta(x) = \emptyset$ otherwise. Then F_δ has closed convex graph given by $[\delta \mathbb{B} \times Y] \cap \operatorname{gph} F$. Also $F(\delta \mathbb{B}) = \operatorname{rge} F_\delta$ and $\operatorname{dom} F_\delta \subset \delta \mathbb{B}$. We want to show that $0 \in \operatorname{int} \operatorname{rge} F_\delta$, but have Theorem 5B.1 at our disposal, according to which we only need to show that $\operatorname{rge} F_\delta$ is absorbing. For that purpose we use an argument which closely parallels one already presented in the proof of Theorem 5B.1. Consider any $y \in Y$. Because $0 \in \operatorname{int} \operatorname{rge} F$, there exists t_0 such that $ty \in \operatorname{rge} F$ when $t \in [0, t_0]$. Then there exists x_0 such that $t_0 y \in F(x_0)$. Let $x = x_0/t_0$, so that $(t_0 x, t_0 y) \in \operatorname{gph} F$. Since $\operatorname{gph} F$ is convex and contains $(0,0)$, it also then contains (tx, ty) for all $t \in [0, t_0]$. Taking $\varepsilon > 0$ for which $\varepsilon \|x\| \le \delta$, we get for all $t \in [0, \varepsilon]$ that $(tx, ty) \in \operatorname{gph} F_\delta$, hence $ty \in \operatorname{rge} F_\delta$, as desired.

Utilizing (9), we can put the argument for the equivalences in Theorem 5B.4 together. That (b) implies (a) is obvious. We work next on getting from (a) to (c). When (a) holds, we have from (9) that (6) holds for some c, in which case Lemma 5B.3 provides (7). By restricting x and y to small neighborhoods of $\bar{x}$ and $\bar{y}$ in (7), we deduce the metric regularity of F at $\bar{x}$ for $\bar{y}$ with any constant $\kappa > 1/c$. Thus, (c) holds. Finally, out of (c) and the equivalences in Theorem 5A.3 we may conclude that F is linearly open at $\bar{x}$ for $\bar{y}$, and this gets us back to (b). $\square$

The preceding argument passed through linear openness as a fourth property which could be added to the equivalences in Theorem 5B.4, but which was left out of the theorem's statement for historical reasons. We now record this fact separately.

Theorem 5B.5 (linear openness from openness and convexity)**.** *For a mapping $F : X \rightrightarrows Y$ with closed convex graph, openness at $\bar{x}$ for $\bar{y}$ always entails linear openness at $\bar{x}$ for $\bar{y}$.*

Another fact, going beyond the original versions of Theorem 5B.4, has come up as well.

Theorem 5B.6 (core criterion for regularity)**.** *Condition (a) of Theorem 5B.4 can be replaced by the criterion that $\bar{y} \in \operatorname{core} \operatorname{rge} F$.*

Proof. This calls up the core property in Theorem 5B.1. $\square$

We can finish tying up loose ends now by returning to the Banach open mapping theorem at the beginning of this chapter and tracing how it fits with the Robinson–Ursescu theorem.

Derivation of Theorem 5A.1 from Theorem 5B.4. It was already noted in the sequel to 5A.1 that condition (d) in that result was equivalent to the metric regularity of the linear mapping A, stated as condition (e). It remains only to observe that when Theorem 5B.4 is applied to $F = A \in \mathscr{L}(X,Y)$ with $\bar{x} = 0$ and $\bar{y} = 0$, the graph of A being a closed subspace of $X \times Y$ (in particular a convex set), and the positive homogeneity of A is brought in, we not only get (b) and (c) of Theorem 5A.1, but also (a). $\square$

The argument for Theorem 5B.4, in obtaining metric regularity, also revealed a relationship between that property and the openness condition in 5B.2 which can be stated in the form

$$\sup\{\, c \in (0,\infty) \,\big|\, (6) \text{ holds } \} \le [\operatorname{reg}(F;\bar{x}\,|\,\bar{y})]^{-1}. \tag{10}$$

Exercise 5B.7 (counterexample)**.** *Show that for the mapping $F : \mathbb{R} \rightrightarrows \mathbb{R}$ defined as*

$$F(x) = \begin{cases} [-x, 0.5] & \text{if } x > 0.25, \\ [-x, 2x] & \text{if } x \in [0, 0.25], \\ \emptyset & \text{if } x < 0, \end{cases}$$

which is not positively homogeneous, and for $\bar{x} = \bar{y} = 0$ the inequality (10) is strict.

Exercise 5B.8 (effective domains of convex functions)**.** *Let $g : X \to (-\infty, \infty]$ be convex and lower semicontinuous, and let $D = \{\, x \,|\, g(x) < \infty \}$. Show that D is a convex set which, although not necessarily closed in X, is sure to have* $\operatorname{core} D = \operatorname{int} D$*. Moreover, on that interior g is locally Lipschitz continuous.*

Guide. Look at the mapping $F : X \rightrightarrows \mathbb{R}$ defined by $F(x) = \{\, y \in \mathbb{R} \,|\, y \ge g(x) \}$. Apply results in this section and also 5A.3. □

5C. Sublinear Mappings

An especially interesting class of positively homogeneous mappings $H : X \rightrightarrows Y$ acting between Banach spaces X and Y consists of the ones for which $\operatorname{gph} H$ is not just a cone, but a convex cone. Such mappings are called *sublinear*, because these geometric properties of $\operatorname{gph} H$ are equivalent to the rules that

$$\begin{aligned} &0 \in H(0), \quad H(\lambda x) = \lambda H(x) \text{ for } \lambda > 0, \\ &H(x + x') \supset H(x) + H(x') \text{ for all } x, x', \end{aligned} \tag{1}$$

which resemble linearity. Since the projection of a convex cone in $X \times Y$ into X or Y is another convex cone, it is clear for a sublinear mapping H that $\operatorname{dom} H$ is a convex cone in X and $\operatorname{rge} H$ is a convex cone in Y. The inverse H^{-1} of a sublinear mapping H is another sublinear mapping.

Although sublinearity has not been mentioned as a specific property before now, sublinear mappings have already appeared many times. Obviously, every linear mapping $A : X \to Y$ is sublinear (its graph being not just a convex cone but in fact a subspace of $X \times Y$). Sublinear also, though, is any mapping $H : X \rightrightarrows Y$ with $H(x) = Ax - K$ for a convex cone K in Y. Such mappings enter the study of constraint systems, with linear equations corresponding to $K = \{0\}$. When A is continuous and K is closed, their graphs are closed.

Sublinear mappings with closed graph enjoy the properties laid out in 5B along with those concerning outer and inner norms at the end of 5A. But their properties go a lot further, as in the result stated now about metric regularity.

Theorem 5C.1 (metric regularity of sublinear mappings). *For a sublinear mapping $H : X \rightrightarrows Y$ with closed graph, and any $(x,y) \in \operatorname{gph} H$ we have*

(2) $$\operatorname{reg}(H;x\,|\,y) \leq \operatorname{reg}(H;0\,|\,0) = \inf\big\{\, \kappa > 0 \,\big|\, H(\kappa \operatorname{int} I\!B) \supset \operatorname{int} I\!B \big\} = \|H^{-1}\|^{-}.$$

Moreover, $\operatorname{reg}(H;0\,|\,0) < \infty$ *if and only if H is surjective, in which case H^{-1} is Lipschitz continuous on Y (in the sense of Pompeiu-Hausdorff distance as defined in 3A) and the infimum of the Lipschitz constant κ for this equals $\|H^{-1}\|^{-}$.*

Proof. Let $\kappa > \operatorname{reg}(H;0\,|\,0)$. Then, from 5A.3, H is linearly open at 0 for 0 with constant κ, which reduces to $H(\kappa \operatorname{int} I\!B) \supset \operatorname{int} I\!B$. On the other hand, just from knowing that $H(\kappa \operatorname{int} I\!B) \supset \operatorname{int} I\!B$, we obtain for arbitrary $(x,y) \in \operatorname{gph} H$ and $r > 0$ through the sublinearity of H that

$$H(x + \kappa r \operatorname{int} I\!B) \supset H(x) + rH(\kappa \operatorname{int} I\!B) \supset y + r \operatorname{int} I\!B.$$

This establishes that $\operatorname{reg}(H;x\,|\,y) \leq \operatorname{reg}(H;0\,|\,0)$ for all $(x,y) \in \operatorname{gph} H$. Appealing again to positive homogeneity, we get

(3) $$\operatorname{reg}(H;0\,|\,0) = \inf\big\{\, \kappa > 0 \,\big|\, H(\kappa \operatorname{int} I\!B) \supset \operatorname{int} I\!B \big\}.$$

The right side of (3) does not change if we replace the open balls with their closures, hence, by 5A.7 or just by the definition of the inner norm, it equals $\|H^{-1}\|^{-}$. This confirms (2).

The finiteness of the right side of (3) corresponds to H being surjective, by virtue of positive homogeneity. We are left now with showing that H^{-1} is Lipschitz continuous on Y with $\|H^{-1}\|^{-}$ as the infimum of the available constants κ.

If H^{-1} is Lipschitz continuous on Y with constant κ, it must in particular have the Aubin property at 0 for 0 with this constant, and then $\kappa \geq \operatorname{reg}(H;0\,|\,0)$ by 5A.3. We already know that this regularity modulus equals $\|H^{-1}\|^{-}$, so we are left with proving that, for every $\kappa > \operatorname{reg}(H;0\,|\,0)$, H^{-1} is Lipschitz continuous on Y with constant κ.

Let $c < [\|H^{-1}\|^{-}]^{-1}$ and $\kappa > 1/c$. Taking (2) into account, we apply the inequality 5B(7) stated in 5B.3 with $x = \bar{x} = 0$ and $\bar{y} = 0$, obtaining the existence of $a > 0$ such that

$$d(0, H^{-1}(y)) \leq \kappa d(y, H(0)) \leq \kappa \|y\| \quad \text{for all } y \in aI\!B.$$

(Here, without loss of generality, we replace the open ball for y by its closure.) For any $y \in Y$, we have $ay/\|y\| \in aI\!B$, and from the positive homogeneity of H we get

(4) $$d(0, H^{-1}(y)) \leq \kappa \|y\| \quad \text{for all } y \in Y.$$

If $\|H^{-1}\|^{-} = 0$ then $0 \in H^{-1}(y)$ for all $y \in Y$ (see 4A.9), hence (4) follows automatically.

Let $y, y' \in Y$ and $x' \in H^{-1}(y')$. Through the surjectivity of H again, we can find for every $\delta > 0$ an $x_\delta \in H^{-1}(y-y')$ such that $\|x_\delta\| \le d(0, H^{-1}(y-y')) + \delta$, and then from (4) we get

$$\|x_\delta\| \le \kappa\|y-y'\| + \delta. \tag{5}$$

Invoking the sublinearity of H yet once more, we obtain

$$x := x' + x_\delta \in H^{-1}(y') + H^{-1}(y-y') \subset H^{-1}(y'+y-y') = H^{-1}(y).$$

Hence $x' = x - x_\delta \in H^{-1}(y) + \|x_\delta\|\mathbb{B}$. Recalling (5), we arrive finally at the existence of $x \in H^{-1}(y)$ such that $\|x - x'\| \le \kappa\|y-y'\| + \delta$. Since δ can be arbitrarily small, this yields Lipschitz continuity of H^{-1}, and we are done. □

Corollary 5C.2 (finiteness of the inner norm). *Let $H : X \rightrightarrows Y$ be a sublinear mapping with closed graph. Then*

$$\begin{aligned} \operatorname{dom} H = X \quad &\Longleftrightarrow \quad \|H\|^- < \infty, \\ \operatorname{rge} H = Y \quad &\Longleftrightarrow \quad \|H^{-1}\|^- < \infty. \end{aligned}$$

Proof. This comes from applying Theorem 5C.1 to both H and H^{-1}. □

Exercise 5C.3 (regularity modulus at zero). *For a sublinear mapping $H : X \rightrightarrows Y$ with closed graph, prove that*

$$\operatorname{reg}(H;0\,|\,0) = \inf\big\{\, \kappa \,\big|\, H(x + \kappa r\mathbb{B}) \supset H(x) + r\mathbb{B} \text{ for all } x \in X,\ r > 0 \big\}.$$

Guide. Utilize the connections with openness properties. □

Example 5C.4 (application to linear constraints). For $A \in \mathscr{L}(X,Y)$ and a closed, convex cone $K \subset Y$, define the solution mapping $S : Y \rightrightarrows X$ by

$$S(y) = \big\{\, x \in X \,\big|\, Ax - y \in K \big\}.$$

Then S is a sublinear mapping with closed graph, and the following properties are equivalent:

(a) $S(y) \neq \emptyset$ for all $y \in Y$;
(b) there exists κ such that $d(x, S(y)) \le \kappa d(Ax - y, K)$ for all $x \in X$, $y \in Y$;
(c) there exists κ such that $h(S(y), S(y')) \le \kappa\|y - y'\|$ for all $y, y' \in Y$,

in which case the infimum of the constants κ that work in (b) coincides with the infimum of the constants κ that work in (c) and equals $\|S\|^-$.

Detail. Here $S = H^{-1}$ for $H(x) = Ax - K$, and the assertions of Theorem 5C.1 then translate into this form. □

Additional insights into the structure of sublinear mappings will emerge from applying a notion which comes out of the following fact.

Exercise 5C.5 (directions of unboundedness in convexity). *Let C be a closed, convex subset of X and let x_1 and x_2 belong to C. If $w \neq 0$ in X has the property that $x_1 + tw \in C$ for all $t \geq 0$, then it also has the property that $x_2 + tw \in C$ for all $t \geq 0$.*

Guide. Fixing any $t_2 > 0$, show that $x_2 + t_2 w$ can be approached arbitrarily closely by points on the line segment between x_2 and $x_1 + t_1 w$ by taking t_1 larger. □

On the basis of the property in 5C.5, the *recession cone* $\operatorname{rc} C$ of a closed, convex set $C \subset X$, defined by

$$\text{(6)} \qquad \operatorname{rc} C = \left\{ w \in X \,\middle|\, \forall x \in C,\ \forall t \geq 0 : x + tw \in C \right\},$$

can equally well be described by

$$\text{(7)} \qquad \operatorname{rc} C = \left\{ w \in X \,\middle|\, \exists x \in C,\ \forall t \geq 0 : x + tw \in C \right\}.$$

It is easily seen that $\operatorname{rc} C$ is a closed, convex cone. In finite dimensions, C is bounded if and only if $\operatorname{rc} C$ is just $\{0\}$, but in infinite dimensions there are unbounded sets for which that holds. For a closed, convex cone K, one just has $\operatorname{rc} K = K$, as seen from the equivalence between (6) and (7) by taking $x = 0$ in (7).

We will apply this now to the graph of a sublinear mapping H. It should be recalled that $\operatorname{dom} H$ is a convex cone, and for any convex cone K the set $K \cap [-K]$ is a subspace, in fact the largest subspace within K. On the other hand, $K - K$ is the smallest subspace that includes K.

Proposition 5C.6 (recession cones in sublinearity). *A sublinear mapping $H : X \rightrightarrows Y$ with closed graph has*

$$\text{(8)} \qquad \operatorname{rc} H(x) = H(0) \quad \textit{for all } x \in \operatorname{dom} H,$$

and on the other hand,

$$\text{(9)} \qquad x \in \operatorname{dom} H \cap [-\operatorname{dom} H] \quad \Longrightarrow \quad \begin{cases} H(x) + H(-x) \subset H(0), \\ H(x) - H(x) \subset H(0) - H(0). \end{cases}$$

Proof. Let $G = \operatorname{gph} H$, this being a closed, convex cone in $X \times Y$, therefore having $\operatorname{rc} G = G$. For any $(x,y) \in G$, the recession cone $\operatorname{rc} H(x)$ consists of the vectors w such that $y + tw \in H(x)$ for all $t \geq 0$, which are the same as the vectors w such that $(x,y) + t(0,w) \in G$ for all $t \geq 0$, i.e., the vectors w such that $(0,w) \in \operatorname{rc} G = G$. But these are the vectors $w \in H(0)$. That proves (8).

The first inclusion in (9) just reflects the rule that $H(x + [-x]) \supset H(x) + H(-x)$ by sublinearity. To obtain the second inclusion, let y_1 and y_2 belong to $H(x)$, which is the same as having $y_1 - y_2 \in H(x) - H(x)$, and let $y \in H(-x)$. Then by the first inclusion we have both $y_1 + y$ and $y_2 + y$ in $H(0)$, hence their difference lies in $H(0) - H(0)$. □

Theorem 5C.7 (single-valuedness of sublinear mappings)**.** *For a sublinear mapping $H : X \rightrightarrows Y$ with closed graph and* $\operatorname{dom} H = X$, *the following conditions are equivalent:*

(a) *H is a linear mapping from $\mathscr{L}(X,Y)$;*
(b) *H is single-valued at some point $\{x\}$;*
(c) $\|H\|^+ < \infty$.

Proof. Certainly (a) leads to (b). On the other hand, (c) necessitates $H(0) = \{0\}$ and hence (b) for $x = 0$. By 5C.6, specifically the second part of (9), if (b) holds for any x it must hold for all x. The first part of (9) reveals then that if $H(x)$ consists just of y, then $H(-x)$ consists just of $-y$. This property, along with the rules of sublinearity, implies linearity. The closedness of the graph of H implies that the linear mapping so obtained is continuous, so we have come back to (a). □

Note that, without the closedness of the graph of H in Theorem 5C.7, there would be no assurance that (b) implies (a). We would still have a linear mapping, but it might not be continuous.

Corollary 5C.8 (single-valuedness of solution mappings)**.** *In the context of Theorem 5C.1, it is impossible for H^{-1} to be single-valued at any point without actually turning out to be a continuous linear mapping from Y to X. The same holds for the solution mapping S for the linear constraint system in 5C.4 when a solution exists for every $y \in Y$.*

We next state the counterpart to Lemma 5A.4 which works for the inner norm of a positively homogeneous mapping. In contrast to the result presented in 5A.8 for the outer norm, convexity is now essential: we must limit ourselves to sublinear mappings.

Theorem 5C.9 (inversion estimate for the inner norm)**.** *Let $H : X \rightrightarrows Y$ be sublinear with closed graph and have $\|H^{-1}\|^- < \infty$. Then for every $B \in \mathscr{L}(X,Y)$ such that $\|H^{-1}\|^- \cdot \|B\| < 1$, one has*

$$\|(H+B)^{-1}\|^- \;\leq\; \frac{\|H^{-1}\|^-}{1 - \|H^{-1}\|^- \|B\|}.$$

The proof of this is postponed until 5E, where it will be deduced from the connection between these properties and metric regularity in 5C.1. Perturbations of metric regularity will be a major theme, starting in Section 5D.

Duality. A special feature of sublinear mappings, with parallels linear mappings, is the availability of “adjoints” in the framework of the duals X^* and Y^* of the Banach spaces X and Y. For a sublinear mapping $H : X \rightrightarrows Y$, the *upper adjoint* $H^{*+} : Y^* \rightrightarrows X^*$ is defined by

$$(10) \qquad (y^*, x^*) \in \operatorname{gph} H^{*+} \iff \langle x^*, x\rangle \le \langle y^*, y\rangle \text{ for all } (x,y) \in \operatorname{gph} H,$$

whereas the *lower adjoint* $H^{*-} : Y^* \rightrightarrows X^*$ is defined by

$$(11) \qquad (y^*,x^*) \in \operatorname{gph} H^{*-} \iff \langle x^*,x\rangle \geq \langle y^*,y\rangle \text{ for all } (x,y) \in \operatorname{gph} H.$$

These formulas correspond to modified polarity operations on the convex cone $\operatorname{gph} H \subset X \times Y$ with polar $[\operatorname{gph} H]^* \subset X^* \times Y^*$. They say that $\operatorname{gph} H^{*+}$ consists of the pairs (y^*,x^*) such that $(x^*,-y^*) \in [\operatorname{gph} H]^*$, while $\operatorname{gph} H^{*-}$ consists of the pairs (y^*,x^*) such that $(-x^*,y^*) \in [\operatorname{gph} H]^*$, and thus imply in particular that the graphs of these adjoints are closed, convex cones — so that both of these mappings from Y^* to X^* are sublinear with closed graph.

The switches of sign in (10) and (11) may seem a pointless distinction to make, but they are essential in capturing rules for recovering H from its adjoints through the fact that, when the spaces X and Y are reflexive, $[\operatorname{gph} H]^{**} = \operatorname{gph} H$ when the convex cone $\operatorname{gph} H$ is closed, and then we get

$$(12) \qquad [H^{*+}]^{*-} = H \text{ and } [H^{*-}]^{*+} = H \text{ for sublinear } H \text{ with closed graph.}$$

When H reduces to a linear mapping $A \in \mathscr{L}(X,Y)$, both adjoints come out as the usual adjoint $A^* \in \mathscr{L}(Y^*,X^*)$. In that setting the graphs are subspaces instead of just cones, and the difference between (10) and (11) has no effect. The fact that $\|A^*\| = \|A\|$ in this case has the following generalization.

Theorem 5C.10 (duality of inner and outer norms)**.** *For any sublinear mapping $H : X \rightrightarrows Y$ with closed graph, one has*

$$(13) \qquad \begin{aligned} \|H\|^+ &= \|H^{*-}\|^- = \|H^{*+}\|^-, \\ \|H\|^- &= \|H^{*-}\|^+ = \|H^{*+}\|^+. \end{aligned}$$

The proof requires some additional background. First, we need to update to Banach spaces the semicontinuity properties introduced in a finite-dimensional framework in Section 3B, but this only involves an extension of notation. A mapping $F : X \rightrightarrows Y$ is *inner semicontinuous* at $\bar{x} \in \operatorname{dom} F$ if for every $y \in F(\bar{x})$ and every neighborhood V of y one can find a neighborhood U of $\bar{x}$ with $U \subset F^{-1}(V)$ or, equivalently, $F(x) \cap V \neq \emptyset$ for all $x \in U$ (for more, see 3B.2). Outer semicontinuity has a parallel extension. Next, we record a standard fact in functional analysis which will be called upon.

Theorem 5C.11 (Hahn–Banach theorem)**.** *Let M be a linear subspace of a Banach space X, and let $p : X \to I\!R$ satisfy*

$$(14) \qquad p(x+y) \leq p(x) + p(y) \text{ and } p(tx) = t\,p(x) \text{ for all } x,y \in X,\ t \geq 0.$$

Let $f : M \to I\!R$ be a linear functional such that $f(x) \leq p(x)$ for all $x \in M$. Then there exists a linear functional $l : X \to I\!R$ such that $l(x) = f(x)$ for all $x \in M$, and $l(x) \leq p(x)$ for all $x \in X$.

In this formulation of the Hahn–Banach theorem, nothing is said about continuity, so X could really be any linear space — no topology is involved. But the main applications are ones in which p is continuous and it follows that l is continuous.

Another standard fact in functional analysis, which can be derived from the Hahn–Banach theorem in that manner, is the following separation theorem.

Theorem 5C.12 (separation theorem). *Let C be a nonempty, closed, convex subset of a Banach space X, and let $x_0 \in X$. Then $x_0 \notin C$ if and only if there exists $x^* \in X^*$ such that*

$$\langle x^*, x_0 \rangle > \sup_{x \in C} \langle x^*, x \rangle.$$

Essentially, this says geometrically that a closed convex set is the intersection of all the "closed half-spaces" that include it.

Proof of Theorem 5C.10. First, observe from (10) and (11) that

$$H^{*+}(y^*) = -H^{*-}(-y^*) \quad \text{for any } y^* \in Y^*,$$

so that

$$\|H^{*-}\|^- = \|H^{*+}\|^- \qquad \text{and} \qquad \|H^{*-}\|^+ = \|H^{*+}\|^+.$$

To prove that $\|H\|^+ = \|H^{*-}\|^-$ we fix any $y^* \in Y^*$ and show that

$$\sup_{x \in \mathbb{B}} \sup_{y \in H(x)} \langle y^*, y \rangle = \inf_{x^* \in H^{*-}(y^*)} \|x^*\| \quad \text{for all } y^* \in \mathbb{B}. \tag{15}$$

If $\inf_{x^* \in H^{*-}(y^*)} \|x^*\| < r$ for some $r > 0$, then there exist $x^* \in H^{*-}(y^*)$ such that $\|x^*\| < r$. For any $\tilde{x} \in \mathbb{B}$ and $\tilde{y} \in H(\tilde{x})$ we have

$$\langle y^*, \tilde{y} \rangle \leq \langle x^*, \tilde{x} \rangle \leq \sup_{x \in \mathbb{B}} \langle x^*, x \rangle = \|x^*\| < r,$$

and then of course $\sup_{x \in \mathbb{B}} \sup_{y \in H(x)} \langle y^*, y \rangle \leq r$. Hence

$$\sup_{x \in \mathbb{B}} \sup_{y \in H(x)} \langle y^*, y \rangle \leq \inf_{x^* \in H^{*-}(y^*)} \|x^*\|. \tag{16}$$

To prove the inequality opposite to (16) and hence the equality (15), assume that $\sup_{x \in \mathbb{B}} \sup_{y \in H(x)} \langle y^*, y \rangle < r$ for some $r > 0$ and pick $0 < d < r$ such that

$$\sup_{x \in \mathbb{B}} \sup_{y \in H(x)} \langle y^*, y \rangle \leq d. \tag{17}$$

Define the mapping $G : X \rightrightarrows \mathbb{R}$ by

$$G : x \mapsto \{ z \mid z = \langle y^*, y \rangle,\ y \in H(x + \mathbb{B}) \}.$$

First, observe that $\operatorname{gph} G$ is convex. Indeed, if $(x_1, z_1), (x_2, z_2) \in \operatorname{gph} G$ and $0 < \lambda < 1$, then there exist $y_i \in Y$ and $w_i \in \mathbb{B}$ with $z_i = \langle y^*, y_i \rangle$ and $y_i \in H(x_i + w_i)$, for $i = 1, 2$. Since H is sublinear, we get $\lambda y_1 + (1-\lambda) y_2 \in H(\lambda(x_1 + w_1) + (1-\lambda)(x_2 + w_2))$. Hence, $\lambda y_1 + (1-\lambda) y_2 \in H(\lambda x_1 + (1-\lambda) x_2 + \mathbb{B})$, and thus,

$$\lambda(x_1, z_1) + (1-\lambda)(x_2, z_2) = (\lambda x_1 + (1-\lambda) x_2, \langle y^*, \lambda y_1 + (1-\lambda) y_2 \rangle) \in \operatorname{gph} G.$$

We will show next that G is inner semicontinuous at 0. Take $\tilde{z} \in G(0)$ and $\varepsilon > 0$. Let $\tilde{z} = \langle y^*, \tilde{y}\rangle$ for $\tilde{y} \in H(\tilde{w})$ and $\tilde{w} \in I\!\!B$. Since $\langle y^*, \cdot\rangle$ is continuous, there is some $\gamma > 0$ such that $|\langle y^*, y\rangle - \tilde{z}| \le \varepsilon$ when $\|y - \tilde{y}\| \le \gamma$. Choose $\delta \in (0,1)$ such that $\delta\|\tilde{y}\| \le \gamma$. If $\|x\| \le \delta$, we have

$$\|(1-\delta)\tilde{w} - x\| \le \|(1-\delta)\tilde{w}\| + \|x\| \le 1,$$

and hence $(1-\delta)\tilde{w} - x \in I\!\!B$. Because H is sublinear,

$$(1-\delta)\tilde{y} \in H((1-\delta)\tilde{w}) = H(x + ((1-\delta)\tilde{w} - x)) \subset H(x + I\!\!B) \quad \text{whenever } \|x\| \le \delta.$$

Moreover, $\|(1-\delta)\tilde{y} - \tilde{y}\| = \delta\|\tilde{y}\| \le \gamma$, and then $|\langle y^*, (1-\delta)\tilde{y}\rangle - \tilde{z}| \le \varepsilon$. Therefore, for all $x \in \delta I\!\!B$, we have $\langle y^*, (1-\delta)\tilde{y}\rangle \in G(x) \cap I\!\!B_\varepsilon(\tilde{z})$, and hence G is inner semicontinuous at 0 as desired.

Let us now define a mapping $K : X \rightrightarrows I\!\!R$ whose graph is the *conical hull* of $\operatorname{gph}(d - G)$ where d is as in (17); that is, its graph is the set of points λh for $h \in \operatorname{gph}(d - G)$ and $\lambda \ge 0$. The conical hull of a convex set is again convex, so K is another sublinear mapping. Since G is inner semicontinuous at 0, there is some neighborhood U of 0 with $U \subset \operatorname{dom} G$, and therefore $\operatorname{dom} K = X$. Consider the functional

$$k : x \mapsto \inf\left\{ z \,\middle|\, z \in K(x)\right\} \quad \text{for } x \in X.$$

Because K is sublinear and $d - H(0) \subset I\!\!R_+$, we have

$$K(x) + K(-x) \subset K(0) \subset I\!\!R_+. \tag{18}$$

This inclusion implies in particular that any point in $-K(-x)$ furnishes a lower bound in $I\!\!R$ for the set of values $K(x)$, for any $x \in X$. Indeed, let $x \in X$ and $y \in -K(-x)$. Then (18) yields $K(x) - y \subset I\!\!R_+$, and consequently $y \le z$ for all $z \in K(x)$. Therefore $k(x)$ is finite for all $x \in X$; we have $\operatorname{dom} k = X$. Also, from the sublinearity of K and the properties of the infimum, we have

$$k(x+y) \le k(x) + k(y) \quad \text{and} \quad k(\alpha x) = \alpha k(x) \;\; \text{for all } x, y \in X \text{ and } \alpha \ge 0.$$

Consider the subspace $M = \{0\} \subset X$ and define $f : M \to I\!\!R$ simply by $f(0) := k(0) = 0$. Applying Hahn–Banach theorem 5C.11 to f, we get a linear functional $l : X \to I\!\!R$ such that $l(0) = 0$ and $l(x) \le k(x)$ for all $x \in X$. We will show now that l is continuous at 0 and hence continuous on the whole X.

Continuity at 0 means that for every $\varepsilon > 0$ there is $\delta > 0$ such that $(l(x) + I\!\!R_+) \cap \varepsilon I\!\!B \ne \emptyset$ whenever $x \in \delta I\!\!B$. Let $z \in d - G(0)$ and take $0 < \lambda < 1$ along with a neighborhood V of z such that $\lambda V \subset \varepsilon I\!\!B$. Since G is inner semicontinuous at 0, there is some $\delta > 0$ such that

$$(d - G(x)) \cap V \ne \emptyset, \text{ for all } x \in (\delta/\lambda)I\!\!B.$$

Since $d - G(x) \subset k(x) + I\!\!R_+$ and $k(x) \ge l(x)$, we have $d - G(x) \subset l(x) + I\!\!R_+$ and $(l(x) + I\!\!R_+) \cap V \ne \emptyset$ for all $x \in (\delta/\lambda)I\!\!B$, so that $(l(\lambda x) + I\!\!R_+) \cap \lambda V \ne \emptyset$ for all $x \in$

$(\delta/\lambda)\mathbb{B}$. This yields

$$(l(x)+\mathbb{R}_+)\cap\varepsilon\mathbb{B}\neq\emptyset \ \text{ for all } x\in\delta\mathbb{B},$$

which means that for all $x\in\delta\mathbb{B}$ there exists some $z\geq l(x)$ with $|z|\leq\varepsilon$. The linearity of l makes $l(x)=-l(-x)$, and therefore $|l(x)|\leq\varepsilon$ for all $x\in\delta\mathbb{B}$. This confirms the continuity of l.

The inclusion $d-G(x)-l(x)\subset\mathbb{R}_+$ is by definition equivalent to having $d-\langle y^*,y\rangle-l(x)\geq 0$ whenever $x\in H^{-1}(y)-\mathbb{B}$. Let $x^*\in X^*$ be such that $\langle x^*,x\rangle=-l(x)$ for all $x\in X$. Then

$$\langle y^*,y\rangle-\langle x^*,x\rangle\leq d \ \text{ for all } y\in Y \ \text{ and all } x\in H^{-1}(y)-\mathbb{B}.$$

Pick any $y\in H(x)$ and $\lambda>0$. Then $\lambda y\in H(\lambda x)$ and $\langle y^*,\lambda y\rangle-\langle x^*,\lambda x\rangle\leq d$, or equivalently,

$$\langle y^*,y\rangle-\langle x^*,x\rangle\leq d/\lambda.$$

Passing to the limit with $\lambda\to\infty$, we obtain $x^*\in H^{*-}(y^*)$. Let now $x\in\mathbb{B}$. Since $0\in H(0)$, we have $0\in H(-x+\mathbb{B})$ and hence $\langle y^*,0\rangle-\langle x^*,-x\rangle\leq d$. Therefore $\|x^*\|\leq d<r$, so that $\inf_{x^*\in H^{*-}(y^*)}\|x^*\|<r$. This, combined with (16), gives us the equality in (15) and hence the equalities in the first line of (13).

We will now confirm the equality in the second line of (13). Suppose $\|H\|^-<r$ for some $r>0$. Then for any $\tilde{x}\in\mathbb{B}$ there is some $\tilde{y}\in H(\tilde{x})$ such that $\|\tilde{y}\|<r$. Given $y^*\in\mathbb{B}$ and $x^*\in H^{*+}(y^*)$, we have

$$\langle x^*,\tilde{x}\rangle\leq\langle y^*,\tilde{y}\rangle\leq\|\tilde{y}\|<r.$$

This being valid for arbitrary $\tilde{x}\in\mathbb{B}$, we conclude that $\|x^*\|\leq r$, and therefore $\|H^{*+}\|^+\leq r$.

Suppose now that $\|H^{*+}\|^+<r$ and pick $s>0$ with

$$\sup_{x^*\in H^{*+}(\mathbb{B})}\|x^*\|=\|H^{*+}\|^+\leq s<r,$$

in which case $H^{*+}(\mathbb{B})\subset s\mathbb{B}$. We will show that

$$\text{(19)} \qquad \langle x^*,x\rangle\leq 1 \ \text{ for all } \ x\in H^{-1}(\mathbb{B}) \quad\Longrightarrow\quad \|x^*\|\leq s.$$

The condition on the left of (19) can be written as $\sup_{y\in\mathbb{B}}\sup_{x\in H^{-1}(y)}\langle x^*,x\rangle\leq 1$, which in turn is completely analogous to (17), with $d=1$ and H replaced by H^{-1} and with y and y^* replaced by x and x^*, respectively. By repeating the argument in the first part of the proof after (17), we obtain $y^*\in(H^{-1})^{*-}(x^*)=(H^{*+})^{-1}(x^*)$ with $\|y^*\|\leq 1$. But then $x^*\in H^{*+}(\mathbb{B})$, and since $H^{*+}(\mathbb{B})\subset s\mathbb{B}$ we have (19).

Now we will show that (19) implies

$$\text{(20)} \qquad s^{-1}\mathbb{B}\subset\operatorname{cl}H^{-1}(\mathbb{B}).$$

If $u \notin \operatorname{cl} H^{-1}(\mathbb{B})$, then from 5C.12 there exists $\tilde{x}^* \in X^*$ with

$$\langle \tilde{x}^*, u \rangle > \sup_{x \in \operatorname{cl} H^{-1}(\boldsymbol{B})} \langle \tilde{x}^*, x \rangle \geq \langle \tilde{x}^*, 0 \rangle = 0.$$

Choose $\lambda > 0$ such that

$$\sup_{x \in \operatorname{cl} H^{-1}(\boldsymbol{B})} \langle \tilde{x}^*, x \rangle < \lambda^{-1} < \langle \tilde{x}^*, u \rangle.$$

Then

$$\langle \lambda \tilde{x}^*, u \rangle > 1 > \sup_{x \in \operatorname{cl} H^{-1}(\boldsymbol{B})} \langle \lambda \tilde{x}^*, x \rangle.$$

According to (19) this implies that $\lambda \tilde{x}^* \in s\mathbb{B}$. Thus,

$$s \geq \|\lambda \tilde{x}^*\| \geq \langle \lambda \tilde{x}^*, u/\|u\| \rangle > \frac{1}{\|u\|},$$

and therefore $u \notin s^{-1}\mathbb{B}$, so (20) holds.

Our next task is to demonstrate that

$$\operatorname{int} s^{-1}\mathbb{B} \subset \operatorname{int} H^{-1}(\mathbb{B}). \tag{21}$$

Define the mapping

$$x \mapsto H_0(x) = \begin{cases} H(x) & \text{for } x \in \mathbb{B}, \\ \emptyset & \text{otherwise.} \end{cases}$$

Then $\operatorname{gph} H_0 = \operatorname{gph} H \cap (X \times \mathbb{B})$ is a closed convex set and $\operatorname{rge} H_0 \subset \mathbb{B}$. By 5B.1 we have

$$\operatorname{int} \operatorname{cl} H^{-1}(\mathbb{B}) = \operatorname{int} \operatorname{cl} \operatorname{dom} H_0 = \operatorname{int} \operatorname{dom} H_0 = \operatorname{int} H^{-1}(\mathbb{B}).$$

This equality combined with the inclusion (20) gives us (21). But then $r^{-1}\mathbb{B} \subset \operatorname{int} s^{-1}\mathbb{B} \subset H^{-1}(\mathbb{B})$, ensuring $\|H\|^- \leq r$. This completes the proof of the second line in (13). $\square$

The above proof can be shortened considerably in the case when X and Y are reflexive Banach spaces, by utilizing the equality (12).

Exercise 5C.13 (more norm duality). *For a sublinear mapping $H : X \rightrightarrows Y$ with closed graph show that*

$$\|(H^{*+})^{-1}\|^+ = \|H^{-1}\|^-.$$

Exercise 5C.14 (adjoint of a sum). *For a sublinear mapping $G : X \rightrightarrows Y$ and $B \in \mathscr{L}(X,Y)$ prove that*

$$(H+B)^{*+} = H^{*+} + B^* \quad \textit{and} \quad (H+B)^{*-} = H^{*-} + B^*.$$

5D. The Theorems of Lyusternik and Graves

We start with the observation that inequality 5A(7) in Lemma 5A.4, giving an estimate for inverting a perturbed linear mapping A, can also be written in the form

$$(\operatorname{reg} A)\cdot\|B\| \leq 1 \quad \Longrightarrow \quad \operatorname{reg}(A+B) \leq \frac{\operatorname{reg} A}{1-(\operatorname{reg} A)\cdot\|B\|},$$

since for an invertible mapping $A \in \mathscr{L}(X,Y)$ one has $\operatorname{reg} A = \|A^{-1}\|$. This alternative formulation shows the way to extending the estimate to nonlinear and even set-valued mappings.

First, we recall a basic definition of differentiability in infinite dimensions, which is just an update of the definition employed in the preceding chapters in finite dimensions. With differentiability as well as Lipschitz continuity and calmness, the only difference is that the Euclidean norm is now replaced by the norms of the Banach spaces X and Y that we work with.

Fréchet differentiability and strict differentiability. *A function $f : X \to Y$ is said to be Fréchet differentiable at $\bar{x}$ if $\bar{x} \in \operatorname{int} \operatorname{dom} f$ and there is a mapping $M \in \mathscr{L}(X,Y)$ such that $\operatorname{clm}(f-M;\bar{x}) = 0$. When such a mapping M exists, it is unique; it is called the Fréchet derivative of f at $\bar{x}$ and denoted by $Df(\bar{x})$, so that*

$$\operatorname{clm}(f - Df(\bar{x});\bar{x}) = 0.$$

If actually

$$\operatorname{lip}(f - Df(\bar{x});\bar{x}) = 0,$$

then f is said to be strictly differentiable at $\bar{x}$.

Partial Fréchet differentiability and partial strict differentiability can be introduced as well on the basis of the partial Lipschitz moduli, by updating the definitions in Section 1D to infinite dimensions. Building on the formulas for the calmness and Lipschitz moduli, we could alternatively express these definitions in an epsilon-delta mode as at the beginning of Chapter 1. If a function f is Fréchet differentiable at every point x of an open set O and the mapping $x \mapsto Df(x)$ is continuous from O to the Banach space $\mathscr{L}(X,Y)$, then f is said to be *continuously* Fréchet differentiable on O. Most of the assertions in Section 1D about functions acting in finite dimensions remain valid in Banach spaces, e.g., continuous Fréchet differentiability around a point implies strict differentiability at this point.

The extension of the Banach open mapping theorem to nonlinear and set-valued mappings goes back to the works of Lyusternik and Graves. In 1934, L. A. Lyusternik published a result saying that if a function $f : X \to Y$ is continuously Fréchet differentiable in a neighborhood of a point $\bar{x}$ where $f(\bar{x}) = 0$ and its derivative mapping $Df(\bar{x})$ is surjective, then the tangent manifold to $f^{-1}(0)$ at $\bar{x}$ is the set $\bar{x}+\ker Df(\bar{x})$. In the current setting we adopt the following statement[1] of Lyusternik theorem:

Theorem 5D.1 (Lyusternik theorem). *Consider a function $f: X\to Y$ that is continuously Fréchet differentiable in a neighborhood of a point $\bar{x}$ with the derivative mapping $Df(\bar{x})$ surjective. Then, in terms of $\bar{y}:=f(\bar{x})$, for every $\varepsilon>0$ there exists $\delta>0$ such that*

$$d(x,f^{-1}(\bar{y}))\le\varepsilon\|x-\bar{x}\| \text{ whenever } x\in\bar{x}+\ker Df(\bar{x}) \text{ and } \|x-\bar{x}\|\le\delta.$$

In 1950 L. M. Graves published a result whose formulation and proof we present here in full, up to some minor adjustments in notation:

Theorem 5D.2 (Graves theorem). *Consider a function $f: X\to Y$ and a point $\bar{x}\in$ int dom f and let f be continuous in $\mathbb{B}_\varepsilon(\bar{x})$ for some $\varepsilon>0$. Let $A\in\mathscr{L}(X,Y)$ be surjective and let $\kappa>\operatorname{reg}A$. Suppose there is a nonnegative μ such that $\mu\kappa<1$ and*

$$(1)\qquad \|f(x)-f(x')-A(x-x')\|\le\mu\|x-x'\| \quad \text{whenever } x,x'\in\mathbb{B}_\varepsilon(\bar{x}).$$

Then, in terms of $\bar{y}:=f(\bar{x})$ and $c=\kappa^{-1}-\mu$, if y is such that $\|y-\bar{y}\|\le c\varepsilon$, then the equation $y=f(x)$ has a solution $x\in\mathbb{B}_\varepsilon(\bar{x})$.

Proof. Without loss of generality, let $\bar{x}=0$ and $\bar{y}=f(\bar{x})=0$. Note that $\kappa^{-1}>\mu$, hence $0<c<\infty$. Take $y\in Y$ with $\|y\|\le c\varepsilon$. Starting from $x^0=0$ we use induction to construct an infinite sequence $\{x^k\}$, the elements of which satisfy for all $k=1,2,\ldots$ the following three conditions:

$$(2a)\qquad A(x^k-x^{k-1})=y-f(x^{k-1}),$$

$$(2b)\qquad \|x^k-x^{k-1}\|\le\kappa(\kappa\mu)^{k-1}\|y\|$$

and

$$(2c)\qquad \|x^k\|\le\|y\|/c.$$

By (d) in the Banach open mapping theorem 5A.1 there exists $x^1\in X$ such that

$$Ax^1=y \quad\text{and}\quad \|x^1\|\le\kappa\|y\|\le\|y\|/c.$$

That is, x^1 satisfies all three conditions (2a), (2b) and (2c). In particular, by the choice of y and the constant c we have $\|x^1\|\le\varepsilon$.

Suppose now that for some $j\ge1$ we have obtained points x^k satisfying (2a), (2b) and (2c) for $k=1,\ldots,j$. Then, since $\|y\|/c\le\varepsilon$, we have from (2c) that all the points x^k satisfy $\|x^k\|\le\varepsilon$. Again using (d) in 5A.1, we can find x^{j+1} such that

[1] In his paper of 1934 Lyusternik did not state his result as a theorem; the statement in 5D.1 is from Dmitruk, Milyutin and Osmolovskiĭ [1980].

(3) $\qquad A(x^{j+1}-x^j) = y - f(x^j) \quad$ and $\quad \|x^{j+1}-x^j\| \le \kappa\|y - f(x^j)\|.$

If we plug $y = Ax^j - Ax^{j-1} + f(x^{j-1})$ into the second relation in (3) and use (1) for $x = x^j$ and $x' = x^{j-1}$ which, as we already know, are from $\varepsilon I\!B$, we obtain

$$\|x^{j+1}-x^j\| \le \kappa\|f(x^j) - f(x^{j-1}) - A(x^j - x^{j-1})\| \le \kappa\mu\|x^j - x^{j-1}\|.$$

Then, by the induction hypothesis,

$$\|x^{j+1}-x^j\| \le \kappa(\kappa\mu)^j\|y\|.$$

Furthermore,

$$\|x^{j+1}\| \le \|x^1\| + \sum_{i=1}^{j}\|x^{i+1}-x^i\| \le \sum_{i=0}^{j}(\kappa\mu)^i\kappa\|y\| \le \frac{\kappa\|y\|}{1-\kappa\mu} = \|y\|/c.$$

The induction step is complete: we obtain an infinite sequence of points x^k satisfying (2a), (2b) and (2c). For any k and j with $k > j > 1$ we have

$$\|x^k - x^j\| \le \sum_{i=j}^{k-1}\|x^{i+1}-x^i\| \le \sum_{i=j}^{k-1}(\kappa\mu)^i\kappa\|y\| \le (\kappa\mu)^j\kappa\|y\|\sum_{i=0}^{\infty}(\kappa\mu)^i \le \frac{\kappa\|y\|}{1-\kappa\mu}(\kappa\mu)^j.$$

Thus, $\{x^k\}$ is a Cauchy sequence, hence convergent to some x, and then, passing to the limit with $k \to \infty$ in (2a) and (2c), this x satisfies $y = f(x)$ and $\|x\| \le \|y\|/c$. The final inequality gives us $\|x\| \le \varepsilon$ and the proof is finished. $\square$

Observe that in the Graves theorem no differentiability of the function f is required, but only "approximate differentiability" as in the theorem of Hildebrand and Graves; see the commentary to Chapter 1. If we suppose that *for every* $\mu > 0$ there exists $\varepsilon > 0$ such that (1) holds for every $x, x' \in I\!B_\varepsilon(\bar{x})$, then A is, by definition, the strict derivative of f at $\bar{x}$, $A = Df(\bar{x})$. That is, the Graves theorem encompasses the following special case: if f is strictly differentiable at $\bar{x}$ and its derivative $Df(\bar{x})$ is onto, then there exist $\varepsilon > 0$ and $c > 0$ such that for every $y \in Y$ with $\|y - \bar{y}\| \le c\varepsilon$ there is an $x \in X$ such that $\|x - \bar{x}\| \le \varepsilon$ and $y = f(x)$.

The statement of the Graves theorem above does not reflect all the information that can be extracted from its proof. In particular, the solution x of $f(x) = y$ whose existence is claimed is not only in the ball $I\!B_\varepsilon(\bar{x})$ but also satisfies $\|x - \bar{x}\| \le \|y - \bar{y}\|/c$. Taking into account that $x \in f^{-1}(y)$, which yields $d(\bar{x}, f^{-1}(y)) \le \|x - \bar{x}\|$, along with the form of the constant c, we get

$$d(\bar{x}, f^{-1}(y)) \le \frac{\kappa}{1-\kappa\mu}\|y - f(\bar{x})\|.$$

Furthermore, this inequality actually holds not only at $\bar{x}$ but also for all x close to $\bar{x}$, and this important extension is hidden in the proof of the theorem.

Indeed, let (1) hold for $x, x' \in I\!B_\varepsilon(\bar{x})$ and choose a positive $\tau < \varepsilon$. Then there is a neighborhood U of $\bar{x}$ such that $I\!B_\tau(x) \subset I\!B_\varepsilon(\bar{x})$ for all $x \in U$. Make U smaller if necessary so that $\|f(x)-f(\bar{x})\| < c\tau$ for $x \in U$. Pick $x \in U$ and a neighborhood V of $\bar{y}$ such that $\|y-f(x)\| \le c\tau$ for $y \in V$. Then, remembering that in the proof $\bar{x}=0$, modify the first induction step in the following way: there exists $x^1 \in X$ such that

$$Ax^1 = y - f(x) + Ax \quad \text{and} \quad \|x^1 - x\| \le \kappa \|y - f(x)\|.$$

Then, construct a sequence $\{x^k\}$ with $x^0 = x$ satisfying (3), thereby obtaining

$$\|x^k - x^{k-1}\| \le \kappa(\kappa\mu)^{k-1}\|y - f(x)\|$$

and then

$$(4) \qquad \|x^k - x\| \le \sum_{i=1}^{k} \|x^i - x^{i-1}\| \le \kappa\|y - f(x)\| \sum_{i=1}^{\infty} (\kappa\mu)^{i-1} \le \frac{\kappa}{1-\kappa\mu}\|y - f(x)\|.$$

Thus,

$$\|x^k - x\| \le \|y - f(x)\|/c \le \tau.$$

The sequence $\{x^k\}$ is a Cauchy sequence, and therefore convergent to some $\tilde{x}$. In passing to the limit in (4), we get

$$\|\tilde{x} - x\| \le \frac{\kappa}{1-\kappa\mu}\|y - f(x)\|.$$

Since $\tilde{x} \in f^{-1}(y)$, we see that, under the conditions of the Graves theorem, there exist neighborhoods U of $\bar{x}$ and V of $f(\bar{x})$ such that

$$(5) \qquad d(x, f^{-1}(y)) \le \frac{\kappa}{1-\kappa\mu}\|y - f(x)\| \quad \text{for } (x,y) \in U \times V.$$

We defined the property described in (5) in Chapter 3 and introduced again in infinite dimensions in 5A: this is metric regularity of the function f at $\bar{x}$ for $\bar{y}$. Noting that the μ in (1) satisfies $\mu \ge \operatorname{lip}(f-A)(\bar{x})$ and it is sufficient to have $\kappa \ge \operatorname{reg} A$ in (5), we arrive at the following result:

Theorem 5D.3 (updated Graves theorem). *Let $f : X \to Y$ be continuous in a neighborhood of $\bar{x}$, let $A \in \mathscr{L}(X,Y)$ satisfy $\operatorname{reg} A \le \kappa$, and suppose $\operatorname{lip}(f-A)(\bar{x}) \le \mu$ for some μ with $\mu\kappa < 1$. Then*

$$(6) \qquad \operatorname{reg}(f;\bar{x}\,|\,\bar{y}) \le \frac{\kappa}{1-\kappa\mu} \quad \textit{for } \bar{y} = f(\bar{x}).$$

For $f = A + B$ we obtain from this result the estimation for perturbed inversion of linear mappings in 5A.4.

The version of the Lyusternik theorem[2] stated as Theorem 5D.1 can be derived from the updated Graves theorem 5D.3. Indeed, the assumptions of 5D.1 are clearly stronger than those of Theorem 5D.3. From (5) with $y = f(\bar{x})$ we get

$$d(x, f^{-1}(\bar{y})) \leq \frac{\kappa}{1-\kappa\mu}\|f(x) - f(\bar{x})\| \tag{7}$$

for all x sufficiently close to $\bar{x}$. Let $\varepsilon > 0$ and choose $\delta > 0$ such that

$$\|f(x) - f(\bar{x}) + Df(\bar{x})(x-\bar{x})\| \leq \frac{(1-\kappa\mu)\varepsilon}{\kappa}\|x-\bar{x}\| \text{ whenever } x \in \mathbb{B}_\delta(\bar{x}). \tag{8}$$

But then for any $x \in (\bar{x} + \ker Df(\bar{x})) \cap \mathbb{B}_\delta(\bar{x})$, from (7) and (8) we obtain

$$d(x, f^{-1}(y)) \leq \frac{\kappa}{1-\kappa\mu}\|f(x) - f(\bar{x})\| \leq \varepsilon\|x-\bar{x}\|,$$

which is the conclusion of 5D.1.

Exercise 5D.4 (correction function version of Graves theorem)**.** *Show that on the conditions of Theorem 5D.3, for $\bar{y} = f(\bar{x})$ there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that for every $y \in V$ and $x \in U$ there exists ξ with the property*

$$f(\xi + x) = y \quad \text{and} \quad \|\xi\| \leq \frac{\kappa}{1-\kappa\mu}\|f(x) - y\|.$$

Guide. From Theorem 5D.3 we see that there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that for every $x \in U$ and $y \in V$

$$d(x, f^{-1}(y)) \leq \frac{\kappa}{1-\kappa\mu}\|y - f(x)\|.$$

Without loss of generality, let $y \neq f(x)$; then we can slightly increase μ so that the latter inequality becomes strict. Then there exists $\eta \in f^{-1}(y)$ such that

$$\|x - \eta\| \leq \frac{\kappa}{1-\kappa\mu}\|y - f(x)\|.$$

Next take $\xi = \eta - x$. □

If the function f in 5D.3 is strictly differentiable at $\bar{x}$, we can choose $A = Df(\bar{x})$, and then $\mu = 0$. In this case (6) reduces to

$$\operatorname{reg}(f;\bar{x}\,|\,\bar{y}) \leq \operatorname{reg} Df(\bar{x}) \text{ for } \bar{y} = f(\bar{x}). \tag{9}$$

[2] The iteration (3), which is a key step in the proof of Graves, is also present in the original proof of Lyusternik [1934], see also Lyusternik and Sobolev [1965]. In the case when A is invertible, it goes back to Goursat [1903], see the commentary to Chapter 1.

In the following section we will show that this inequality actually holds as equality. We call this result the basic Lyusternik–Graves theorem and devote the following several sections to extending it in various directions.

Theorem 5D.5 (basic Lyusternik–Graves theorem)**.** *For a function $f : X \to Y$ which is strictly differentiable at $\bar{x}$, one has*

$$\operatorname{reg}(f;\bar{x}\,|\,\bar{y}) = \operatorname{reg} Df(\bar{x}).$$

We terminate this section with two major observations. For the first, assume that the mapping A in Theorem 5D.2 is not only surjective but also invertible. Then the iteration procedure used in the proof of the Graves theorem becomes the iteration used by Goursat (see the commentary to Chapter 1), namely

$$x^{j+1} = x^j - A^{-1}(f(x^j) - y).$$

In that case, one obtains the existence of a single-valued graphical localization of the inverse f^{-1} around $f(\bar{x})$ for $\bar{x}$, as in Theorem 1A.1. If the derivative mapping $Df(\bar{x})$ is merely surjective, as assumed in the Graves theorem, the inverse f^{-1} may not have a single-valued graphical localization around $\bar{y}$ for $\bar{x}$ but, still, this inverse, being a set-valued mapping, has the Aubin property at $\bar{y}$ for $\bar{x}$.

Our second observation is that in the proof of Theorem 5D.2 we use the linearity of the mapping A only to apply the Banach open mapping theorem. But we can employ the regularity modulus for any nonlinear, even set-valued, mapping. After this somewhat historical section, we will explore this idea further in the sections which follow.

5E. Extending the Lyusternik–Graves Theorem

In this section we show that the updated Graves theorem 5D.3, in company with the stability of metric regularity under perturbations, demonstrated in Theorem 3F.1, can be extended to a much broader framework of set-valued mappings acting in abstract spaces. Specifically, we consider a set-valued mapping F acting from a metric space X to another metric space Y, where both metrics are denoted by ρ but may be defined differently. In such spaces the standard definitions, e.g. of the ball in X with center x and radius r and the distance from a point x to a set C in Y, need only be adapted to metric notation:

$$\mathbb{B}_r(\bar{x}) = \left\{ x \in X \,\middle|\, \rho(x,\bar{x}) \le r \right\}, \qquad d(x,C) = \inf_{x' \in C} \rho(x,x').$$

Recall that a subset C of a complete metric space is closed if $d(x,C) = 0 \Leftrightarrow x \in C$. Also recall that a set C is locally closed at a point $x \in C$ if there is a neighborhood U of x such that the intersection $C \cap U$ is closed.

In metric spaces X and Y, the definition of the Lipschitz modulus of a function $g : X \to Y$ is extended in a natural way, with attention paid to the metric notation of distance:

$$\operatorname{lip}(g;\bar{x}) = \limsup_{\substack{x,x' \to \bar{x},\\ x \neq x'}} \frac{\rho(g(x), g(x'))}{\rho(x,x')}.$$

For set-valued mappings F acting in such spaces, the definitions of metric regularity and the Aubin property persist in the same manner while the linear openness of a mapping F at $\bar{x}$ for $\bar{y}$ translates as the existence of a constant $\kappa > 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$F(\operatorname{int} \mathbb{B}_{\kappa r}(x)) \supset \operatorname{int} \mathbb{B}_r(y) \quad \text{for all} \quad (x,y) \in \operatorname{gph} F \cap (U \times V) \quad \text{and} \quad r > 0.$$

The equivalence of metric regularity with the Aubin property of the inverse and the linear openness (3E.7 and 3E.9) with the same constant remains valid as well. Recall that all three definitions require the graph of the mapping be locally closed at the reference point.

It will be important for our efforts to take the metric space X to be *complete* and to suppose that Y is a *linear* space equipped with a *shift-invariant* metric ρ. Shift invariance means that

$$\rho(y+z, y'+z) = \rho(y,y') \text{ for all } y, y', z \in Y.$$

Of course, any Banach space meets these requirements.

The result stated next is just a reformulation of Theorem 3F.1 for mappings acting in metric spaces. Since this result has its roots in the theorems of Lyusternik and Graves, we call it the extended Lyusternik–Graves theorem.

Theorem 5E.1 (extended Lyusternik–Graves theorem)**.** *Let X be a complete metric space and let Y be a linear space with shift-invariant metric. Consider a mapping $F : X \rightrightarrows Y$, a point $(\bar{x},\bar{y}) \in \operatorname{gph} F$ and a function $g : X \to Y$ with $\bar{x} \in \operatorname{int} \operatorname{dom} g$. Let κ and μ be nonnegative constants such that*

$$\kappa\mu < 1, \quad \operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \le \kappa \quad \textit{and} \quad \operatorname{lip}(g;\bar{x}) \le \mu.$$

Then

$$\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) \le \frac{\kappa}{1-\kappa\mu}. \tag{1}$$

Before arguing this, we note that it immediately allows us to supply 5C.9 and 5D.5 with proofs.

Proof of 5C.9. We apply 5E.1 with X and Y Banach spaces, $F = H$, $\bar{x} = 0$ and $\bar{y} = 0$. According to 5C.1, $\operatorname{reg}(H;0\,|\,0) = \|H^{-1}\|^{-}$, so 5E.1 tells us that for any

$\lambda > \|H^{-1}\|^{-}$, any $B \in \mathscr{L}(X,Y)$ with $\|B\| < 1/\lambda$, and any μ with $\|B\| < \mu < 1/\lambda$ one has from (1) that $\|(H+B)^{-1}\|^{-} \le \lambda/(1-\lambda\mu)$. It remains only to pass to the limit as $\lambda \to \|H^{-1}\|^{-}$ and $\mu \to \|B\|$. □

Proof of 5D.5. To obtain the inequality opposite to 5D(9), choose $F = f$ and $g = Df(\bar{x}) - f$ and apply 5E.1, in this case with $\mu = 0$. □

We proceed now with presenting a proof of Theorem 5E.1 which echoes on a more abstract level the way we proved the classical inverse function theorem 1A.1 in Chapter 1, by using an iteration in line with the original argument in the proof of the Graves theorem 5D.2. Another way of proving 5E.1 will be demonstrated on a more general implicit function version of it, based on a contraction mapping principle for set-valued mappings. Then we present yet another proof of 5E.1.

Proof I of Theorem 5E.1. Let κ and μ be as in the statement of the theorem and choose a function $g : X \to Y$ with $\operatorname{lip}(g;\bar{x}) \le \mu$. Without loss of generality, suppose $g(\bar{x}) = 0$. Let $\lambda > \kappa$ and $\nu > \mu$ satisfy $\lambda\nu < 1$. Choose $\alpha > 0$ small enough such that the set $\operatorname{gph} F \cap (\mathbb{B}_\alpha(\bar{x}) \times \mathbb{B}_\alpha(\bar{y}))$ is closed, g is Lipschitz continuous with constant ν on $\mathbb{B}_\alpha(\bar{x})$, and

$$(2) \qquad d(x, F^{-1}(y)) \le \lambda d(y, F(x)) \quad \text{for all } (x,y) \in \mathbb{B}_\alpha(\bar{x}) \times \mathbb{B}_\alpha(\bar{y}).$$

From (2) with $x = \bar{x}$, it follows that

$$(3) \qquad F^{-1}(y) \neq \emptyset \quad \text{for all } y \in \mathbb{B}_\alpha(\bar{y}).$$

Having fixed λ, α and ν, consider the following system of inequalities:

$$(4) \qquad \begin{cases} \lambda\nu + \varepsilon < 1, \\ \frac{1}{1-(\lambda\nu+\varepsilon)}[(1+\lambda\nu)a + \lambda b + \varepsilon] + a \le \alpha, \\ b + \nu\left(\frac{1}{1-(\lambda\nu+\varepsilon)}[(1+\lambda\nu)a + \lambda b + \varepsilon] + a\right) \le \alpha. \end{cases}$$

It is not difficult to see that there are positive a, b and ε that satisfy this system. Indeed, first fix ε such that these inequalities hold strictly for $a = b = 0$; then pick sufficiently small a and b so that both the second and the third inequality are not violated.

Let $x \in \mathbb{B}_a(\bar{x})$ and $y \in \mathbb{B}_b(\bar{y})$. According to the choice of a and of b in (4), we have

$$(5) \qquad \rho(y - g(x), \bar{y}) \le \nu\rho(x,\bar{x}) + \rho(y,\bar{y}) \le \nu a + b \le \alpha.$$

Let $(x_k, y_k) \in \operatorname{gph}(g+F) \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(g(\bar{x}) + \bar{y}))$ and let $(x_k, y_k) \to (x,y)$ as $k \to \infty$. Then from (4) and (5) we have that $x_k \in F^{-1}(y_k - g(x_k)) \cap \mathbb{B}_\alpha(\bar{x})$ and $y_k - g(x_k) \in \mathbb{B}_\alpha(\bar{y})$. Passing to the limit and using the local closedness of $\operatorname{gph} F$ we conclude that $(x,y) \in \operatorname{gph}(g+F) \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(\bar{y}))$, hence $\operatorname{gph}(g+F)$ is locally closed at $(\bar{x}, g(\bar{x}) + \bar{y})$.

We will show next that

$$(6) \qquad d(x,(g+F)^{-1}(y)) \leq \frac{\lambda}{1-\lambda\nu} d(y,(g+F)(x)).$$

Since x and y are arbitrarily chosen in the corresponding balls $\mathbb{B}_a(\bar{x})$ and $\mathbb{B}_b(\bar{y})$, and λ and ν are arbitrarily close to κ and μ, respectively, (6) implies (1).

Through (3) and (5), there exists $z^1 \in F^{-1}(y-g(x))$ such that

$$(7) \qquad \rho(z^1,x) \leq d(x,F^{-1}(y-g(x))) + \varepsilon \leq \lambda d(y,(g+F)(x)) + \varepsilon.$$

If $z^1 = x$, then $x \in F^{-1}(y-g(x))$, which is the same as $x \in (g+F)^{-1}(y)$. Then (6) holds automatically, since its left side is 0. Let $z^1 \neq x$. In this case, using (2), we obtain

$$(8) \qquad \begin{aligned} \rho(z^1,x) &\leq \rho(x,\bar{x}) + d(\bar{x},F^{-1}(y-g(x))) + \varepsilon \\ &\leq \rho(x,\bar{x}) + \lambda d(y-g(x),F(\bar{x})) + \varepsilon \\ &\leq \rho(x,\bar{x}) + \lambda\,\rho(y,\bar{y}) + \lambda\,\rho(g(x),g(\bar{x})) + \varepsilon \\ &\leq \rho(x,\bar{x}) + \lambda\,\rho(y,\bar{y}) + \lambda\nu\,\rho(x,\bar{x}) + \varepsilon \\ &\leq (1+\lambda\nu)a + \lambda b + \varepsilon. \end{aligned}$$

Hence, by (4),

$$(9) \qquad \rho(z^1,\bar{x}) \leq \rho(z^1,x) + \rho(x,\bar{x}) \leq (1+\lambda\nu)a + \lambda b + \varepsilon + a \leq \alpha.$$

By induction, we construct a sequence of vectors $z^k \in \mathbb{B}_\alpha(\bar{x})$, with $z^0 = x$, such that, for $k = 0,1,\ldots,$

$$(10) \qquad z^{k+1} \in F^{-1}(y-g(z^k)) \quad \text{and} \quad \rho(z^{k+1},z^k) \leq (\lambda\nu+\varepsilon)^k \rho(z^1,x).$$

We already found z^1 which gives us (10) for $k=0$. Suppose that for some $n \geq 1$ we have generated $z^1, z^2, \ldots, z^n$ satisfying (10). If $z^n = z^{n-1}$ then $z^n \in F^{-1}(y-g(z^n))$ and hence $z^n \in (g+F)^{-1}(y)$. Then, by using (2), (7) and (10), we get

$$\begin{aligned} d(x,(g+F)^{-1}(y)) &\leq \rho(z^n,x) \leq \sum_{i=0}^{n-1} \rho(z^{i+1},z^i) \\ &\leq \sum_{i=0}^{n-1} (\lambda\nu+\varepsilon)^i \rho(z^1,x) \leq \frac{1}{1-(\lambda\nu+\varepsilon)} \rho(z^1,x) \\ &\leq \frac{\lambda}{1-(\lambda\nu+\varepsilon)} \left(d(y,(g+F)(x)) + \frac{\varepsilon}{\lambda} \right). \end{aligned}$$

Since the left side of this inequality does not depend on the ε on the right, we are able to obtain (6) by letting ε go to 0.

Assume $z^n \neq z^{n-1}$. We will first show that $z^i \in \mathbb{B}_\alpha(\bar{x})$ for all $i = 2,3,\ldots,n$. Utilizing (10), for such an i we have

$$\rho(z^i,x) \leq \sum_{j=0}^{i-1} \rho(z^{j+1},z^j) \leq \sum_{j=0}^{i-1} (\lambda\nu+\varepsilon)^j \rho(z^1,x) \leq \frac{1}{1-(\lambda\nu+\varepsilon)} \rho(z^1,x)$$

and therefore, through (8) and (4),

$$(11)\quad \rho(z^i,\bar{x}) \le \rho(z^i,x)+\rho(x,\bar{x}) \le \frac{1}{1-(\lambda\nu+\varepsilon)}[(1+\lambda\nu)a+\lambda b+\varepsilon]+a \le \alpha.$$

Thus, we have $z^i \in \mathbb{B}_\alpha(\bar{x})$ for all $i=1,\dots,n$.

Taking into account the estimate in (11) for $i=n$ and the third inequality in (4), we get

$$\begin{aligned}\rho(y-g(z^n),\bar{y}) &\le \rho(y,\bar{y})+\nu\,\rho(z^n,\bar{x})\\ &\le b+\nu\left(\frac{1}{1-(\lambda\nu+\varepsilon)}[(1+\lambda\nu)a+\lambda b+\varepsilon]+a\right) \le \alpha.\end{aligned}$$

Since $\rho(z^n,z^{n-1})>0$, from (3) there exists $z^{n+1}\in F^{-1}(y-g(z^n))$ such that

$$\rho(z^{n+1},z^n) \le d\big(z^n,F^{-1}(y-g(z^n))\big)+\varepsilon\rho(z^n,z^{n-1}),$$

and then (2) yields

$$\rho(z^{n+1},z^n) \le \lambda d(y-g(z^n),F(z^n))+\varepsilon\rho(z^n,z^{n-1}).$$

Since $z^n\in F^{-1}(y-g(z^{n-1}))$ and hence $y-g(z^{n-1})\in F(z^n)$, by invoking the induction hypothesis, we obtain

$$\begin{aligned}\rho(z^{n+1},z^n) &\le \lambda\,\rho(g(z^n),g(z^{n-1}))+\varepsilon\,\rho(z^n,z^{n-1})\\ &\le (\lambda\nu+\varepsilon)\,\rho(z^n,z^{n-1}) \le (\lambda\nu+\varepsilon)^n\rho(z^1,x).\end{aligned}$$

The induction is complete, and therefore (10) holds for all k.

Right after (10) we showed that when $z^k=z^{k-1}$ for some k then (6) holds. Suppose now that $z^{k+1}\ne z^k$ for all k. By virtue of the second condition in (10), we see for any natural n and m with $m<n$ that

$$\rho(z^n,z^m) \le \sum_{k=m}^{n-1}\rho(z^{k+1},z^k) \le \sum_{k=m}^{n-1}(\lambda\nu+\varepsilon)^k\rho(z^1,x) \le \frac{\rho(z^1,x)}{1-(\lambda\nu+\varepsilon)}(\lambda\nu+\varepsilon)^m.$$

We conclude that the sequence $\{z^k\}$ satisfies the Cauchy condition, and all its elements are in $\mathbb{B}_\alpha(\bar{x})$. Hence this sequence converges to some $z\in\mathbb{B}_\alpha(\bar{x})$ which, from (10) and the local closedness of $\operatorname{gph} F$, satisfies $z\in F^{-1}(y-g(z))$, that is, $z\in(g+F)^{-1}(y)$. Moreover,

$$\begin{aligned}d(x,(g+F)^{-1}(y)) \le \rho(z,x) = \lim_{k\to\infty}\rho(z^k,x) &\le \lim_{k\to\infty}\sum_{i=0}^{k}\rho(z^{i+1},z^i)\\ &\le \lim_{k\to\infty}\sum_{i=0}^{k}(\lambda\nu+\varepsilon)^i\rho(z^1,x) \le \frac{1}{1-(\lambda\nu+\varepsilon)}\rho(z^1,x)\end{aligned}$$

$$\leq \frac{1}{1-(\lambda\nu+\varepsilon)}\bigg(\lambda d(y,(g+F)(x))+\varepsilon\bigg),$$

the final inequality being obtained from (2) and (7). Taking the limit as $\varepsilon \to 0$ we obtain (6), and the proof is finished. □

Theorem 5E.1 can be also deduced from a more general implicit function theorem (5E.5 below) which we supply with a proof based on the following extension of the contraction mapping principle 1A.2 for set-valued mappings, furnished with a proof the idea of which goes back to Banach [1922], if not earlier.

Theorem 5E.2 (contraction mapping principle for set-valued mappings)**.** *Let X be a complete metric space with metric ρ, and consider a set-valued mapping $\Phi : X \rightrightarrows X$ and a point $\bar{x} \in X$. Suppose that there exist scalars $a > 0$ and $\lambda \in (0,1)$ such that the set* gph $\Phi \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_a(\bar{x}))$ *is closed and*

(a) $d(\bar{x},\Phi(\bar{x})) < a(1-\lambda)$*;*

(b) $e(\Phi(u)\bigcap \mathbb{B}_a(\bar{x}),\Phi(v)) \leq \lambda\,\rho(u,v)$ *for all* $u,v \in \mathbb{B}_a(\bar{x})$.

Then Φ has a fixed point in $\mathbb{B}_a(\bar{x})$; that is, there exists $x \in \mathbb{B}_a(\bar{x})$ such that $x \in \Phi(x)$.

Proof. By assumption (a) there exists $x^1 \in \Phi(\bar{x})$ such that $\rho(x^1,\bar{x}) < a(1-\lambda)$. Proceeding by induction, let $x^0 = \bar{x}$ and suppose that there exists $x^{k+1} \in \Phi(x^k) \cap \mathbb{B}_a(\bar{x})$ for $k = 0,1,\ldots,j-1$ with

$$\rho(x^{k+1},x^k) < a(1-\lambda)\lambda^k.$$

By assumption (b),

$$d(x^j,\Phi(x^j)) \leq e(\Phi(x^{j-1})\cap \mathbb{B}_a(\bar{x}),\Phi(x^j)) \leq \lambda\,\rho(x^j,x^{j-1}) < a(1-\lambda)\lambda^j.$$

This implies that there is an $x^{j+1} \in \Phi(x^j)$ such that

$$\rho(x^{j+1},x^j) < a(1-\lambda)\lambda^j.$$

By the triangle inequality,

$$\rho(x^{j+1},\bar{x}) \leq \sum_{i=0}^{j} \rho(x^{i+1},x^i) < a(1-\lambda)\sum_{i=0}^{j}\lambda^i < a.$$

Hence $x^{j+1} \in \Phi(x^j) \cap \mathbb{B}_a(\bar{x})$ and the induction step is complete.

For any $k > m > 1$ we then have

$$\rho(x^k,x^m) \leq \sum_{i=m}^{k-1} \rho(x^{i+1},x^i) < a(1-\lambda)\sum_{i=m}^{k-1}\lambda^i < a\lambda^m.$$

Thus, $\{x^k\}$ is a Cauchy sequence and consequently converges to some $x \in \mathbb{B}_a(\bar{x})$. Since $(x^{k-1},x^k) \in$ gph $\Phi \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_a(\bar{x}))$ which is a closed set, we conclude that $x \in \Phi(x)$. □

For completeness, we now supply with a proof the (standard) contraction mapping principle 1A.2.

Proof of Theorem 1A.2. Let Φ be a function which is Lipschitz continuous on $\mathbb{B}_a(\bar{x})$ with constant $\lambda \in [0,1)$ and let $\rho(\bar{x},\Phi(\bar{x})) \leq a(1-\lambda)$. If $\lambda = 0$ then Φ is a constant function on $\mathbb{B}_a(\bar{x})$ whose value $\Phi(x)$, the same for all $x \in \mathbb{B}_a(\bar{x})$, is a point in $\mathbb{B}_a(\bar{x})$. This value is the unique fixed point of Φ. Let $\lambda > 0$. By repeating the argument in the proof of 5E.2 for the sequence of points x^k satisfying $x^{k+1} = \Phi(x^k)$, $k = 0,1,\ldots$, for $x^0 = \bar{x}$, with all strict inequalities replaced by non-strict ones, we obtain that Φ has a fixed point in $\mathbb{B}_a(\bar{x})$. Suppose that Φ has two fixed points in $\mathbb{B}_a(\bar{x})$, that is, there are $x, x' \in \mathbb{B}_a(\bar{x})$, $x \neq x'$, with $x = \Phi(x)$ and $x' = \Phi(x')$. Then we have

$$0 < \rho(x,x') = \rho(\Phi(x),\Phi(x')) \leq \lambda\, \rho(x,x') < \rho(x,x'),$$

which is absurd. Hence, Φ has a unique fixed point. □

Theorem 5E.2 is a generalization of the following theorem due to Nadler [1969], which is most known in the literature.

Theorem 5E.3 (Nadler fixed point theorem)**.** *Let X be a complete metric space and suppose that Φ maps X into the set of closed subsets of X and is Lipschitz continuous in the sense of Pompeiu-Hausdorff distance on X with Lipschitz constant $\lambda \in (0,1)$. Then Φ has a fixed point.*

Proof. We will first show that Φ has closed graph. Indeed, let $(x_k,y_k) \in \operatorname{gph}\Phi$ and $(x_k,y_k) \to (x,y)$. Then

$$\begin{aligned} d(y,\Phi(x)) &\leq \rho(y,y_k) + d(y_k,\Phi(x)) \\ &\leq \rho(y,y_k) + h(\Phi(x_k),\Phi(x)) \\ &\leq \rho(y,y_k) + \lambda\rho(x_k,x) \to 0 \text{ as } k \to \infty. \end{aligned}$$

Hence $d(y,\Phi(x)) = 0$ and since $\Phi(x)$ is closed we have $(x,y) \in \operatorname{gph}\Phi$, and therefore $\operatorname{gph}\Phi$ is closed as claimed.

Let $\bar{x} \in X$ and choose $a > d(\bar{x},\Phi(\bar{x}))/(1-\lambda)$. Clearly, the set $\operatorname{gph}\Phi \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_a(\bar{x}))$ is closed. Furthermore, for every $u, v \in \mathbb{B}_a(\bar{x})$ we obtain

$$e(\Phi(u) \cap \mathbb{B}_a(\bar{x}),\Phi(v)) \leq e(\Phi(u),\Phi(v)) \leq h(\Phi(u),\Phi(v)) \leq \lambda\rho(u,v).$$

Hence, by Theorem 5E.2 there exists $x \in X$ such that $x \in \Phi(x)$. □

Exercise 5E.4. *Prove Nadler's theorem 5E.3 by using Ekeland's principle 4B.5.*

Guide. Consider the function $f(x) = d(x,\Phi(x))$ for $x \in X$. First show that this function is lower semicontinuous continuous on X by using the assumption that Φ is Lipschitz continuous (use 1D.4 and the proof of 3A.3). Also note that $f(x) \geq 0$ for all $x \in X$. Choose $\delta \in (0,1-\lambda)$. Ekeland's principle yields that there exists $u_\delta \in X$ such that

$$d(u_\delta, \Phi(u_\delta)) \le d(x, \Phi(x)) + \delta\rho(u_\delta, x) \text{ for every } x \in X.$$

Let $d(u_\delta, \Phi(u_\delta)) > 0$. From the last inequality and the Lipschitz continuity of Φ, for any $x \in \Phi(u_\delta)$ we have

$$d(u_\delta, \Phi(u_\delta)) \le h(\Phi(u_\delta), \Phi(x)) + \delta\rho(u_\delta, x) \le (\lambda + \delta)\rho(u_\delta, x).$$

Passing to the infimum with $x \in \Phi(u_\delta)$, we obtain

$$0 < d(u_\delta, \Phi(u_\delta)) \le (\lambda + \delta) d(u_\delta, \Phi(u_\delta)) < d(u_\delta, \Phi(u_\delta)),$$

a contradiction. Hence $d(u_\delta, \Phi(u_\delta)) = 0$. Since Φ is closed valued, we conclude that $u_\delta \in \Phi(u_\delta)$. $\square$

We are now ready to state and prove an implicit mapping version of Theorem 5E.1.

Theorem 5E.5 (extended Lyusternik–Graves theorem in implicit form). *Let X be a complete metric space, let Y be a linear metric space with shift-invariant metric, and let P be a metric space, with all metrics denoted by ρ. For $f : P \times X \to Y$ and $F : X \rightrightarrows Y$, consider the generalized equation $f(p,x) + F(x) \ni 0$ with solution mapping*

$$S(p) = \left\{ x \,\middle|\, f(p,x) + F(x) \ni 0 \right\} \text{ having } \bar{x} \in S(\bar{p}).$$

Let $h : X \to Y$ be a strict estimator of f with respect to x uniformly in p at $(\bar{p},\bar{x})$ with constant μ, and suppose that $h + F$ is metrically regular at $\bar{x}$ for 0 with $\operatorname{reg}(h + F;\bar{x}\,|\,0) \le \kappa$. Assume

$$\kappa\mu < 1 \text{ and } \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})) \le \gamma < \infty.$$

Then S has the Aubin property at $\bar{p}$ for $\bar{x}$, and moreover

$$\operatorname{lip}(S;\bar{p}\,|\,\bar{x}) \le \frac{\kappa\gamma}{1 - \kappa\mu}. \tag{12}$$

Proof. Choose $\lambda > \kappa$ and $\nu > \mu$ such that $\lambda\nu < 1$. Also, let $\beta > \gamma$. Then there exist positive scalars α and τ such that

$$\text{the set } \operatorname{gph}(h + F) \cap (\mathbb{B}_\alpha(\bar{x}) \times \mathbb{B}_\alpha(0)) \text{ is closed,} \tag{13}$$

and

$$e\left((h + F)^{-1}(y') \cap \mathbb{B}_\alpha(\bar{x}), (h + F)^{-1}(y) \right) \le \lambda\, \rho(y', y) \text{ for all } y', y \in \mathbb{B}_\alpha(0). \tag{14}$$

Furthermore, for $r(p,x) = f(p,x) - h(x)$,

$$\rho(r(p,x'), r(p,x)) \le \nu\, \rho(x', x) \text{ for all } x', x \in \mathbb{B}_\alpha(\bar{x}) \text{ and } p \in \mathbb{B}_\tau(\bar{p}) \tag{15}$$

and also

$$\rho(f(p',x),f(p,x)) \le \beta\,\rho(p',p) \quad \text{for all } p',p \in \mathbb{B}_\tau(\bar p) \text{ and } x \in \mathbb{B}_\alpha(\bar x). \tag{16}$$

Fix λ^+ such that

$$\frac{2\lambda\beta}{1-\lambda\nu} \ge \lambda^+ > \frac{\lambda\beta}{1-\lambda\nu}. \tag{17}$$

Now, choose positive $a < \alpha\min\{1,1/\nu\}$ and then $q \le \tau$ such that

$$\nu a + \beta q \le \alpha \quad \text{and} \quad 2\lambda^+ q + a \le \alpha. \tag{18}$$

Then, from (15) and (16), for every $x \in \mathbb{B}_a(\bar x)$ and $p \in \mathbb{B}_q(\bar p)$ we have

$$\begin{aligned}\rho(r(p,x),0) &\le \rho(r(p,x),r(p,\bar x)) + \rho(r(p,\bar x),r(\bar p,\bar x))\\ &\le \nu\,\rho(x,\bar x) + \beta\,\rho(p,\bar p) \le \nu a + \beta q \le \alpha.\end{aligned} \tag{19}$$

Fix $p \in \mathbb{B}_q(\bar p)$ and consider the mapping $\Phi_p : X \rightrightarrows X$ defined as

$$\Phi_p(x) = (h+F)^{-1}(-r(p,x)).$$

Let $p',p \in \mathbb{B}_q(\bar p)$ with $p \ne p'$ and let $x' \in S(p') \cap \mathbb{B}_a(\bar x)$. Then $x' \in \Phi_{p'}(x') \cap \mathbb{B}_a(\bar x)$. Let $\varepsilon := \lambda^+\rho(p',p)$; then $\varepsilon \le \lambda^+(2q)$. Thus, from (14), where we use (19), and from (15), (16), and (17) we deduce that

$$\begin{aligned}d(x',\Phi_p(x')) &\le e\Big((h+F)^{-1}(-r(p',x')) \cap \mathbb{B}_\alpha(\bar x), (h+F)^{-1}(-r(p,x'))\Big)\\ &\le \lambda\,\rho(f(p',x'),f(p,x')) \le \lambda\beta\,\rho(p',p) < \lambda^+\rho(p',p)(1-\lambda\nu) = \varepsilon(1-\lambda\nu).\end{aligned}$$

Since $x' \in \mathbb{B}_a(\bar x)$ and $\varepsilon + a \le \alpha$ from (18), we get $\mathbb{B}_\varepsilon(x') \subset \mathbb{B}_\alpha(\bar x)$. Then, the set $\operatorname{gph}\Phi_p \cap (\mathbb{B}_\varepsilon(x') \times \mathbb{B}_\varepsilon(x'))$ is closed. Further, for any $u,v \in \mathbb{B}_\varepsilon(x')$ using again (14) (with (19)) and (15), we see that

$$\begin{aligned}&e(\Phi_p(u) \cap \mathbb{B}_\varepsilon(x'), \Phi_p(v))\\ &\qquad \le e\Big((h+F)^{-1}(-r(p,u)) \cap \mathbb{B}_\alpha(\bar x), (h+F)^{-1}(-r(p,v))\Big)\\ &\qquad \le \lambda\,\rho(r(p,u),r(p,v)) \le \lambda\nu\,\rho(u,v).\end{aligned}$$

The contraction mapping principle in Theorem 5E.2 then applies, with the λ there taken to be the $\lambda\nu$ here, and it follows that there exists $x \in \Phi_p(x) \cap \mathbb{B}_\varepsilon(x')$ and hence $x \in S(p) \cap \mathbb{B}_\varepsilon(x')$. Thus,

$$d(x',S(p)) \le \rho(x',x) \le \varepsilon = \lambda^+\rho(p',p).$$

Since this inequality holds for every $x' \in S(p') \cap \mathbb{B}_a(\bar x)$ and evey λ^+ fulfilling (17), we arrive at

$$e(S(p') \cap \mathbb{B}_a(\bar x), S(p)) \le \lambda^+\rho(p',p).$$

Noting that $(p,x) \in \operatorname{gph} S \cap (\mathbb{B}_q(\bar{p}) \times \mathbb{B}_a(\bar{x}))$ is the same as $x \in (h+F)^{-1}(-r(p,x)) \cap \mathbb{B}_a(\bar{x})$, $p \in \mathbb{B}_q(\bar{p})$, and using (13) and (19), we conclude that gph S is locally closed at $(\bar{p},\bar{x})$. Hence, S has the Aubin property at $\bar{p}$ for $\bar{x}$ with modulus not greater than λ^+. Since λ^+ can be arbitrarily close to $\lambda\beta/(1-\lambda\nu)$, and λ, ν and β can be arbitrarily close to κ, μ, and γ, respectively, we achieve the estimate (12). □

Clearly, Theorem 5E.1 is the particular case of 5E.5 for $P = Y$, $h = 0$ and $f(p,x) = -p + g(x)$.

A Banach space version of Theorem 3F.9 is given next whose statement and proof need only minor adjustments in notation and terminology.

Theorem 5E.6 (using strict differentiability and ample parameterization)**.** *Let X, Y and P be Banach spaces. For $f : P \times X \to Y$ and $F : X \rightrightarrows Y$, consider the generalized equation $f(p,x) + F(x) \ni 0$ with solution mapping S as in 5E.5 and a pair $(\bar{p},\bar{x})$ with $\bar{x} \in S(\bar{p})$. Suppose that f is strictly differentiable at $(\bar{p},\bar{x})$. If the mapping*

$$h + F \text{ for } h(x) = f(\bar{p},\bar{x}) + D_x f(\bar{p},\bar{x})(x - \bar{x})$$

is metrically regular at $\bar{x}$ for 0, then S has the Aubin property at $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(S;\bar{p}\,|\,\bar{x}) \leq \operatorname{reg}(h+F;\bar{x}\,|\,0) \cdot \|D_p f(\bar{p},\bar{x})\|.$$

Furthermore, when f satisfies the ample parameterization condition

$$\text{the mapping } D_p f(\bar{p},\bar{x}) \text{ is surjective,}$$

then the converse implication holds as well: the mapping $h+F$ is metrically regular at $\bar{x}$ for 0 provided that S has the Aubin property at $\bar{p}$ for $\bar{x}$.

The single-valued version of 5E.1 is commonly known as Milyutin's theorem.

Theorem 5E.7 (Milyutin)**.** *Let X be a complete metric space and Y be a linear normed space. Consider functions $f : X \to Y$ and $g : X \to Y$ and let κ and μ be nonnegative constants such that $\kappa\mu < 1$. Suppose that f is linearly open at $\bar{x}$ with a constant κ and g is Lipschitz continuous around $\bar{x}$ with constant μ. Then $f + g$ is linearly open at $\bar{x}$ with constant $\kappa/(1-\kappa\mu)$.*

Exercise 5E.8. *Prove that, under the conditions of Theorem 5E.1,*

$$\inf_{g:X\to Y} \left\{ \operatorname{lip}(g;\bar{x}) \,\middle|\, F + g \text{ is not metrically regular at } \bar{x} \text{ for } \bar{y} + g(\bar{x}) \right\} \geq \frac{1}{\operatorname{reg}(F;\bar{x}\,|\,\bar{y})}.$$

5F. Strong Metric Regularity and Implicit Functions

Here we first present a strong regularity analogue of Theorem 5E.1 that provides a sharper view of the interplay among the constants and neighborhoods of a mapping and its perturbation. As in the preceding section, we consider mappings acting in metric spaces, to which the concept of strong regularity can be extended in an obvious way. The following result extends Theorem 3G.3 to metric spaces and can be proved in the same way, by utilizing Theorem 5E.1 instead of 3F.1 together with the metric space versions of Propositions 3G.1 and 3G.2. It could be also proved directly, by using the standard (single-valued) version of the contracting mapping principle, 1A.2. We will use this kind of a direct proof in the more general implicit function Theorem 5F.4 and then will derive 5F.1 from it.

Theorem 5F.1 (inverse functions and strong metric regularity in metric spaces)**.** *Let X be a complete metric space and let Y be a linear metric space with shift-invariant metric. Let κ and μ be nonnegative constants such that $\kappa\mu < 1$. Consider a mapping $F : X \rightrightarrows Y$ and any $(\bar{x},\bar{y}) \in \operatorname{gph} F$ such that F is strongly metrically regular at $\bar{x}$ for $\bar{y}$ with $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \leq \kappa$ and a function $g : X \to Y$ with $\bar{x} \in \operatorname{int}\operatorname{dom} g$ and $\operatorname{lip}(g;\bar{x}) \leq \mu$. Then the mapping $g+F$ is strongly metrically regular at $\bar{x}$ for $g(\bar{x})+\bar{y}$. Moreover,*

$$\operatorname{reg}(g+F;\bar{x}\,|\,g(\bar{x})+\bar{y}) \leq \frac{\kappa}{1-\kappa\mu}.$$

Exercise 5F.2. *Derive 5F.1 from 5E.1.*

Guide. Follow in the way 3G.3 is derived from 3F.1. □

Exercise 5F.3. *Let X be a complete metric space and let Y be a linear metric space with shift-invariant metric, let κ and μ be nonnegative constants such that $\kappa\mu < 1$. Let $\bar{x} \in X$ and $\bar{y} \in Y$, and let g be Lipschitz continuous around $\bar{x}$ with Lipschitz constant μ and f be Lipschitz continuous around $\bar{y}+g(\bar{x})$ with Lipschitz constant κ. Prove that the mapping $(f^{-1}+g)^{-1}$ has a Lipschitz continuous single-valued localization around $\bar{y}+g(\bar{x})$ for $\bar{x}$ with Lipschitz constant $\kappa/(1-\kappa\mu)$.*

Guide. Apply 5F.1 with $F = f^{-1}$. □

We present next a strong regularity version of Theorem 5E.5 which has already appeared in various forms in the preceding chapters. We proved a weaker form of this result in 2B.5 via 2B.6 and stated it again in Theorem 3G.4, which we left unproved. Here we treat a general case, and since the result is central in this book, we supply it with an unabbreviated direct proof.

Theorem 5F.4 (implicit functions with strong metric regularity in metric spaces)**.** *Let X be a complete metric space, let Y be a linear metric space with shift-invariant metric, and let P be a metric space. For $f : P \times X \to Y$ and $F : X \rightrightarrows Y$, consider the generalized equation $f(p,x)+F(x) \ni 0$ with solution mapping*

$$S(p) = \{ x \mid f(p,x) + F(x) \ni 0 \} \quad \textit{having } \bar{x} \in S(\bar{p}).$$

Let $f(\cdot,\bar{x})$ *be continuous at* $\bar{p}$ *and let* $h : X \to Y$ *be a strict estimator of* f *with respect to* x *uniformly in* p *at* $(\bar{p},\bar{x})$ *with constant* μ. *Suppose that* $h+F$ *is strongly metrically regular at* $\bar{x}$ *for* 0 *or, equivalently, the inverse* $(h+F)^{-1}$ *has a Lipschitz continuous single-valued localization* ω *around* 0 *for* $\bar{x}$ *such that there exists* $\kappa \geq \operatorname{reg}(h+F;\bar{x}|0) = \operatorname{lip}(\omega;0)$ *with* $\kappa\mu < 1$.

Then the solution mapping S *has a single-valued localization* s *around* $\bar{p}$ *for* $\bar{x}$. *Moreover, for every* $\varepsilon > 0$ *there exists a neighborhood* Q *of* $\bar{p}$ *such that*

$$(1) \qquad \rho(s(p'),s(p)) \leq \frac{\kappa+\varepsilon}{1-\kappa\mu}\, \rho(f(p',s(p)), f(p,s(p))) \quad \textit{for all } p',p \in Q.$$

In particular, s *is continuous at* $\bar{p}$. *In addition, if*

$$(2) \qquad \operatorname{clm}_p(f;(\bar{x},\bar{p})) < \infty,$$

then the solution mapping S *has a single-valued graphical localization* s *around* $\bar{p}$ *for* $\bar{x}$ *which is calm at* $\bar{p}$ *with*

$$(3) \qquad \operatorname{clm}(s;\bar{p}) \leq \frac{\kappa}{1-\kappa\mu} \operatorname{clm}_p(f;(\bar{p},\bar{x})).$$

If (2) is replaced by the stronger condition

$$(4) \qquad \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})) < \infty,$$

then the graphical localization s *of* S *around* $\bar{p}$ *for* $\bar{x}$ *is Lipschitz continuous near* $\bar{p}$ *with*

$$(5) \qquad \operatorname{lip}(s;\bar{p}) \leq \frac{\kappa}{1-\kappa\mu} \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})).$$

If $h : X \to Y$ *is not only a strict estimator of* f, *but also a strict first-order approximation of* f *with respect to* x *uniformly in* p *at* $(\bar{p},\bar{x})$, *then, under (2), we have*

$$\operatorname{clm}(s;\bar{p}) \leq \operatorname{lip}(\omega;0) \operatorname{clm}_p(f;(\bar{p},\bar{x})),$$

and under (4),

$$\operatorname{lip}(s;\bar{p}) \leq \operatorname{lip}(\omega;0) \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})).$$

Proof. Let $\varepsilon > 0$ and choose $\lambda > \kappa$ and $\nu > \mu$ such that

$$(6) \qquad \lambda\nu < 1 \quad \text{and} \quad \frac{\lambda}{1-\lambda\nu} \leq \frac{\kappa+\varepsilon}{1-\kappa\mu}.$$

Then there exist positive scalars α and τ such that for each $y \in \mathbb{B}_\alpha(0)$ the set $(h+F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$ is a singleton, equal to the value $\omega(y)$ of the single-valued localization of $(h+F)^{-1}$, and this localization ω is Lipschitz continuous with

Lipschitz constant λ on $I\!B_\alpha(0)$. We adjust α and τ to also have, for $e(p,x) = f(p,x) - h(x)$,

(7) $\quad \rho(e(p,x'),e(p,x)) \le \nu\,\rho(x',x)$ for all $x',x \in I\!B_\alpha(\bar{x})$ and $p \in I\!B_\tau(\bar{p})$.

Choose a positive $a \le \alpha$ satisfying

(8)
$$\nu a + \frac{a(1-\lambda\nu)}{\lambda} \le \alpha$$

and then a positive $r \le \tau$ such that

(9)
$$\rho(f(p,\bar{x}), f(\bar{p},\bar{x})) \le \frac{a(1-\lambda\nu)}{\lambda} \quad \text{for all } p \in I\!B_r(\bar{p}).$$

Then for every $x \in I\!B_a(\bar{x})$ and $p \in I\!B_r(\bar{p})$, from (7)–(9) we have

$$\begin{aligned}\rho(e(p,x),0) &\le \rho(e(p,x),e(p,\bar{x})) + \rho(e(p,\bar{x}),e(\bar{p},\bar{x}))\\ &\le \nu\rho(x,\bar{x}) + \rho(f(p,\bar{x}),f(\bar{p},\bar{x})) \le \nu a + a(1-\lambda\nu)/\lambda \le \alpha.\end{aligned}$$

Hence, for such x and p, $e(p,x) \in \operatorname{dom}\omega$.

Fix an arbitrary $p \in I\!B_r(\bar{p})$ and consider the mapping

$$\Phi_p : x \mapsto \omega(-e(p,x)) \text{ for } x \in I\!B_a(\bar{x}).$$

Observe that for any $x \in I\!B_a(\bar{x})$ having $x = \Phi_p(x)$ implies $x \in S(p) \cap I\!B_a(\bar{x})$, and conversely. Noting that $\bar{x} = \omega(0)$ and using (9) we obtain

$$\rho(\bar{x},\Phi_p(\bar{x})) = \rho(\omega(0),\omega(-e(p,\bar{x}))) \le \lambda\,\rho(f(\bar{p},\bar{x}),f(p,\bar{x})) \le a(1-\lambda\nu).$$

Further, for any $u,v \in I\!B_a(\bar{x})$, using (7) we see that

$$\begin{aligned}\rho(\Phi_p(u),\Phi_p(v)) &= \rho(\omega(-e(p,u)),\omega(-e(p,v)))\\ &\le \lambda\,\rho(e(p,u),e(p,v)) \le \lambda\nu\,\rho(u,v).\end{aligned}$$

The contraction mapping principle 1A.2 applies, with the λ there taken to be the $\lambda\nu$ here, hence for each $p \in I\!B_r(\bar{p})$ there exists exactly one $s(p)$ in $I\!B_a(\bar{x})$ such that $s(p) \in S(p)$; thus

(10)
$$s(p) = \omega(-e(p,s(p))).$$

The function $p \mapsto s(p)$ is therefore a single-valued localization of S around $\bar{p}$ for $\bar{x}$. Moreover, from (10), for each $p',p \in I\!B_r(\bar{p})$ we have

$$\begin{aligned}\rho(s(p'),s(p)) &= \rho(\omega(-e(p',s(p'))),\omega(-e(p,s(p))))\\ &\le \rho(\omega(-e(p',s(p'))),\omega(-e(p',s(p))))\\ &\quad + \rho(\omega(-e(p',s(p))),\omega(-e(p,s(p))))\\ &\le \lambda\,\rho(-e(p',s(p')),-e(p',s(p)))\end{aligned}$$

$$+\lambda\,\rho(-e(p',s(p)),-e(p,s(p)))$$
$$\leq \lambda\nu\,\rho(s(p'),s(p))+\lambda\,\rho(f(p',s(p)),f(p,s(p))).$$

Hence,

$$\rho(s(p'),s(p)) \leq \frac{\lambda}{1-\lambda\nu}\rho(f(p',s(p)),f(p,s(p))).$$

Taking into account (6), we obtain (1). In particular, for $p=\bar{p}$, from the continuity of $f(\cdot,\bar{x})$ at $\bar{p}$ we get that s is continuous at $\bar{p}$. Under (2), the estimate (3) directly follows from (1) by passing to zero with ε, and the same for (5) under (4). If h is a strict first-order approximation of f, then μ could be arbitrarily small, and by passing to $\operatorname{lip}(\omega;0)$ with κ and to 0 with μ we obtain from (3) and (5) the last two estimates in the statement. ☐

Proof of 5F.1. Apply 5F.2 with $P=Y$, $h=0$ and $f(p,x)=-p+g(x)$; then pass to zero with ε in (1). ☐

Utilizing strict differentiability and ample parameterization we come to the following infinite-dimensional implicit function theorem which parallels 5E.6.

Theorem 5F.5 (using strict differentiability and ample parameterization)**.** *Let X, Y and P be Banach spaces. For $f: P\times X\to Y$ and $F: X\rightrightarrows Y$, consider the generalized equation $f(p,x)+F(x)\ni 0$ with solution mapping S and a pair $(\bar{p},\bar{x})$ with $\bar{x}\in S(\bar{p})$ and suppose that f is strictly differentiable at $(\bar{p},\bar{x})$. If the mapping*

$$h+F \text{ for } h(x)=f(\bar{p},\bar{x})+D_xf(\bar{p},\bar{x})(x-\bar{x})$$

is strongly metrically regular at $\bar{x}$ for 0, then S has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$ with

$$\operatorname{lip}(s;\bar{p}) \leq \operatorname{reg}(h+F;\bar{x}\,|\,0)\cdot\|D_pf(\bar{p},\bar{x})\|.$$

Furthermore, when f satisfies the ample parameterization condition:

$$\text{the mapping } D_pf(\bar{p},\bar{x}) \text{ is surjective,}$$

then the converse implication holds as well: the mapping $h+F$ is strongly metrically regular at $\bar{x}$ for 0 provided that S has a Lipschitz continuous single-valued localization around $\bar{p}$ for $\bar{x}$.

5G. Parametric Inverse Function Theorems

In this section we first present an extension of Theorem 5E.1 in which the function g, added to the metrically regular mapping F, depends on a parameter. This result shows in particular that the regularity constant of F and the neighborhoods of metric regularity of the perturbed mapping $g+F$ depend only on the regularity modulus of the underlying mapping F and the Lipschitz modulus of the function g, and not on the actual value of the parameter in a neighborhood of the reference value. The proof is parallel to the proof of 5E.5 with some subtle differences; therefore we present it in full detail. For better transparency, we consider mappings acting in Banach spaces, but the extension to metric spaces as in 5E is straightforward.

Theorem 5G.1 (parametric Lyusternik–Graves theorem). *Let X, Y and P be Banach spaces and consider a mapping $F : X \rightrightarrows Y$ and a point $(\bar{x},\bar{y}) \in \operatorname{gph} F$. Consider also a function $g : P \times X \to Y$ with $(\bar{q},\bar{x}) \in \operatorname{int}\operatorname{dom} g$ and suppose that there exist nonnegative constants κ and μ such that*

$$\kappa\mu < 1 \quad \operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \leq \kappa \quad \textit{and} \quad \widehat{\operatorname{lip}}_x(g;(\bar{q},\bar{x})) \leq \mu.$$

Then for every $\kappa' > \kappa/(1-\kappa\mu)$ there exist neighborhoods Q' of $\bar{q}$, U' of $\bar{x}$ and V' of $\bar{y}$ such that for each $q \in Q'$ the mapping $g(q,\cdot)+F(\cdot)$ is metrically regular in x at $\bar{x}$ for $g(q,\bar{x})+\bar{y}$ with constant κ' and neighborhoods U' of $\bar{x}$ and $g(q,\bar{x})+V'$ of $g(q,\bar{x})+\bar{y}$.

Proof. Pick the constants κ and μ as in the statement of the theorem and then any $\kappa' > \kappa/(1-\kappa\mu)$. Let $\lambda > \kappa$ and $\nu > \mu$ be such that $\lambda\nu < 1$ and $\lambda/(1-\lambda\nu) < \kappa'$. Then there exist positive constants a and b such that

$$(1) \qquad d(x,F^{-1}(y)) \leq \lambda d(y,F(x)) \quad \text{for all} \quad (x,y) \in \mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(\bar{y}).$$

Adjust a and b if necessary to obtain

$$(2) \qquad \text{the set } \operatorname{gph} F \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(\bar{y})) \text{ is closed.}$$

Then choose $c > 0$ and make a smaller if necessary such that $\nu a < b/2$ and

$$(3) \qquad \|g(q,x) - g(q,x')\| \leq \nu\|x - x'\| \quad \text{for all } x,x' \in \mathbb{B}_a(\bar{x}) \text{ and } q \in \mathbb{B}_c(\bar{q}).$$

Choose positive constants α and β to satisfy

$$(4) \qquad \alpha + 5\kappa'\beta \leq a, \quad 8\beta \leq b, \quad \alpha \leq 2\kappa'\beta, \quad \text{and} \quad \nu(\alpha + 5\kappa'\beta) + \beta \leq b.$$

Fix $q \in \mathbb{B}_c(\bar{q})$. We will first prove that for every $x \in \mathbb{B}_\alpha(\bar{x})$, $y \in \mathbb{B}_\beta(g(q,\bar{x})+\bar{y})$ and $y' \in (g(q,x)+F(x)) \cap \mathbb{B}_{4\beta}(g(q,\bar{x})+\bar{y})$ one has

$$(5) \qquad d(x,(g(q,\cdot)+F(\cdot))^{-1}(y)) \leq \kappa'\|y - y'\|.$$

Let $y' \in (g(q,x)+F(x)) \cap \mathbb{B}_{4\beta}(g(q,\bar{x})+\bar{y})$. If $y = y'$ then $x \in (g(q,\cdot)+F(\cdot))^{-1}(y)$ and (5) holds since both the left side and the right side are zero. Suppose $y' \neq y$ and consider the mapping

$$\Phi : u \mapsto F^{-1}(-g(q,u)+y) \ \text{ for } u \in \mathbb{B}_a(\bar{x}).$$

Using (3) and (4), for any $u \in \mathbb{B}_a(\bar{x})$ we have

$$\|-g(q,u)+y'-\bar{y}\| \leq \|-g(q,u)+g(q,\bar{x})\| + \|y'-\bar{y}-g(q,\bar{x})\| \leq \nu a + 4\beta \leq b.$$

The same estimate holds of course with y' replaced by y because y was chosen in $\mathbb{B}_\beta(g(q,\bar{x})+\bar{y})$. Hence, both $-g(q,x)+y'$ and $-g(q,x)+y$ are in $\mathbb{B}_b(\bar{y})$.

Let $r := \kappa'\|y-y'\|$. We will show next that the set gph $\Phi \cap (\mathbb{B}_r(x) \times \mathbb{B}_r(x))$ is closed. Let $(u_n,z_n) \in$ gph $\Phi \cap (\mathbb{B}_r(x) \times \mathbb{B}_r(x))$ and $(u_n,z_n) \to (\tilde{u},\tilde{z})$. Then

$$(z_n, -g(q,u_n)+y) \in \text{gph}\, F$$

and also, from (4),

$$\|z_n - \bar{x}\| \leq \|z_n - x\| + \|x - \bar{x}\| \leq r + \alpha = \kappa'\|y-y'\| + \alpha \leq 5\kappa'\beta + \alpha \leq a$$

and

$$\begin{aligned}\|-g(q,u_n)+y-\bar{y}\| &\leq \|-g(q,u_n)+g(q,\bar{x})\| + \|y-\bar{y}-g(q,\bar{x})\| \\ &\leq \nu\|u_n-\bar{x}\| + \beta \leq \nu(\|u_n-x\|+\|x-\bar{x}\|)+\beta \\ &\leq \nu(r+\alpha)+\beta \leq \nu(5\kappa'\beta+\alpha)+\beta \leq b.\end{aligned}$$

Thus $(z_n, -g(q,u_n)+y) \in$ gph $F \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(\bar{y}))$ which is closed by (2). Note that $r \leq \kappa'(4\beta+\beta)$ and hence, from the first relation in (4), $\mathbb{B}_r(x) \subset \mathbb{B}_a(\bar{x})$. Since $g(q,\cdot)$ is continuous in $\mathbb{B}_a(\bar{x})$ (even Lipschitz, from (3)) and $u_n \in \mathbb{B}_r(x) \subset \mathbb{B}_a(\bar{x})$, we get that $(\tilde{z}, -g(q,\tilde{u})+y) \in$ gph F. Clearly, both $\tilde{u}$ and $\tilde{z}$ are from $\mathbb{B}_r(x)$, thus $(\tilde{u},\tilde{z}) \in$ gph $\Phi \cap (\mathbb{B}_r(x) \times \mathbb{B}_r(x))$. Hence, the set gph $\Phi \cap (\mathbb{B}_r(x) \times \mathbb{B}_r(x))$ is closed, which implies that gph$(g(q,\cdot)+F(\cdot))$ is locally closed at $(\bar{x}, g(q,\bar{x})+\bar{y})$.

Since $x \in (g(q,\cdot)+F(\cdot))^{-1}(y') \cap \mathbb{B}_a(\bar{x})$, utilizing the metric regularity of F we obtain

$$\begin{aligned}d(x,\Phi(x)) &= d(x, F^{-1}(-g(q,x)+y)) \leq \lambda d(-g(q,x)+y, F(x)) \\ &\leq \lambda\|-g(q,x)+y-(y'-g(q,x))\| = \lambda\|y-y'\| \\ &< \kappa'\|y-y'\|(1-\lambda\nu) = r(1-\lambda\nu).\end{aligned}$$

Then (1), combined with (3) and the observation above that $\mathbb{B}_r(x) \subset \mathbb{B}_a(\bar{x})$, implies that for any $u, v \in \mathbb{B}_r(x)$,

$$\begin{aligned}&e(\Phi(u) \cap \mathbb{B}_r(x), \Phi(v)) \\ &\quad \leq \sup\{d(z, F^{-1}(-g(q,v)+y)) : z \in F^{-1}(-g(q,u)+y) \cap \mathbb{B}_a(\bar{x})\} \\ &\quad \leq \sup\{\lambda d(-g(q,v)+y, F(z)) : z \in F^{-1}(-g(q,u)+y) \cap \mathbb{B}_a(\bar{x})\}\end{aligned}$$

$$\leq \lambda \| -g(q,u)+g(q,v)\| \leq \lambda \nu \|u-v\|.$$

Theorem 5E.2 then yields the existence of a point $\hat{x} \in \Phi(\hat{x}) \cap \mathbb{B}_r(x)$; that is,

$$y \in g(q,\hat{x})+F(\hat{x}) \quad \text{and} \quad \|\hat{x}-x\| \leq \kappa' \|y-y'\|.$$

Thus, since $\hat{x} \in (g(q,\cdot)+F(\cdot))^{-1}(y)$ we obtain (5).

To complete the proof it remains to show that for every $x \in \mathbb{B}_\alpha(\bar{x})$ and $y \in \mathbb{B}_\beta(g(q,\bar{x})+\bar{y})$ one has

$$(6) \qquad d(x,(g(q,\cdot)+F(\cdot))^{-1}(y)) \leq \kappa' d(y, g(q,x)+F(x)),$$

which gives us the desired property of $g(q,\cdot)+F(\cdot)$. First, note that if $g(q,x)+F(x)=\emptyset$ the right side of (6) is ∞ and we are done. Let $\varepsilon > 0$ and $w \in g(q,x)+F(x)$ be such that

$$(7) \qquad \|w-y\| \leq d(y,g(q,x)+F(x))+\varepsilon.$$

If $w \in \mathbb{B}_{4\beta}(g(q,\bar{x})+\bar{y})$ then from (5) and (7) we have that

$$d(x,(g(q,\cdot)+F(\cdot))^{-1}(y)) \leq \kappa' \|y-w\| \leq \kappa'(d(y,g(q,x)+F(x))+\varepsilon)$$

and since the left side of this inequality does not depend on ε, we obtain (6). If $w \notin \mathbb{B}_{4\beta}(g(q,\bar{x})+\bar{y})$ then

$$\|w-y\| \geq \|w-g(q,\bar{x})-\bar{y}\| - \|y-g(q,\bar{x})-\bar{y}\| \geq 3\beta.$$

Using (5) with $x=\bar{x}$ and $y'=g(q,\bar{x})+\bar{y})$ and the Lipschitz continuity of the distance function with constant 1, see 1D.4(b), we obtain

$$\begin{aligned} d(x,(g(q,\cdot)+F(\cdot))^{-1}(y)) &\leq \|x-\bar{x}\| + d(\bar{x},(g(q,\cdot)+F(\cdot))^{-1}(y)) \\ &\leq \alpha + \kappa' \|\bar{y}+g(q,\bar{x})-y\| \leq 3\kappa'\beta \\ &\leq \kappa' \|w-y\| \leq \kappa'(d(y,g(q,x)+F(x))+\varepsilon). \end{aligned}$$

This again implies (6) and the proof is complete. □

Our next theorem is a parametric version of 5F.1 which easily follows from 5G.1 and (the Banach space version of) 3G.1.

Theorem 5G.2 (parametric strong metric regularity)**.** *Let X, Y and P be Banach spaces and consider a mapping $F: X \rightrightarrows Y$ and $(\bar{x},\bar{y}) \in \operatorname{gph} F$ such that F is strongly metrically regular at $\bar{x}$ for $\bar{y}$. Let κ and μ be nonnegative constants such that $\kappa \geq \operatorname{reg}(F;\bar{x}|\bar{y})$ and $\kappa\mu < 1$ and consider a function $g: P \times X \to Y$ which satisfies $\widehat{\operatorname{lip}}_x(g;(\bar{q},\bar{x})) \leq \mu$. Then for every $\kappa' > \kappa/(1-\kappa\mu)$ there exist neighborhoods Q of $\bar{q}$, U of $\bar{x}$ and V of $\bar{y}$ such that for each $q \in Q$ the mapping $g(q,\cdot)+F(\cdot)$ is strongly* *metrically regular in* x *at* $\bar{x}$ *for* $g(q,\bar{x})+\bar{y}$ *with constant* κ' *and neighborhoods* U *of* $\bar{x}$ *and* $g(q,\bar{x})+V$ *of* $g(q,\bar{x})+\bar{y}$.

At the end of this section we present a theorem which combines perturbed versions of 5E.1 and 5F.1. It complements 5G.1 and 5G.2 and is more convenient to use in certain cases. We will apply all three results in Chapter 6.

Theorem 5G.3 (perturbed [strong] metric regularity)**.** *Let* X,Y *be Banach spaces. Consider a mapping* $F: X \rightrightarrows Y$ *and a point* $(\bar{x},\bar{y}) \in \operatorname{gph} F$ *at which* F *is metrically regular, that is, there exist positive constants* a, b, *and a nonnegative* κ *such that*

$$\text{the set } \operatorname{gph} F \cap (\mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(\bar{y})) \text{ is closed} \tag{8}$$

and

$$d(x,F^{-1}(y)) \le \kappa d(y,F(x)) \quad \text{for all } (x,y) \in \mathbb{B}_a(\bar{x}) \times \mathbb{B}_b(\bar{y}). \tag{9}$$

Let $\mu > 0$ *be such that* $\kappa\mu < 1$ *and let* $\kappa' > \kappa/(1-\kappa\mu)$. *Then for every positive* α *and* β *such that*

$$\alpha \le a/2, \quad \mu\alpha + 2\beta \le b \quad \text{and} \quad 2\kappa'\beta \le \alpha \tag{10}$$

and for every function $g: X \to Y$ *satisfying*

$$\|g(\bar{x})\| \le \beta \tag{11}$$

and

$$\|g(x)-g(x')\| \le \mu\|x-x'\| \quad \text{for every} \quad x,x' \in \mathbb{B}_\alpha(\bar{x}), \tag{12}$$

the mapping $g+F$ *has the following property: for every* $y,y' \in \mathbb{B}_\beta(\bar{y})$ *and every* $x \in (g+F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$ *there exists* $x' \in (g+F)^{-1}(y')$ *such that*

$$\|x-x'\| \le \kappa'\|y-y'\|. \tag{13}$$

In addition, if the mapping F *is strongly metrically regular at* $\bar{x}$ *for* $\bar{y}$*; that is, the mapping* $y \mapsto F^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ *is single-valued and Lipschitz continuous on* $\mathbb{B}_b(\bar{y})$ *with a Lipschitz constant* κ, *then for* μ, κ', α *and* β *as above and any function* g *satisfying (11) and (12), the mapping* $y \mapsto (g+F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$ *is a Lipschitz continuous function on* $\mathbb{B}_\beta(\bar{y})$ *with a Lipschitz constant* κ'.

Observe that if we assume $g(\bar{x}) = 0$ then this theorem reduces to the Banach space versions of 5E.1 and 5F.1. However, if $g(\bar{x}) \neq 0$ then $(\bar{x},\bar{y})$ may be not in the graph of $g+F$ and we cannot claim that $g+F$ is (strongly) metrically regular at $\bar{x}$ for $\bar{y}$. This could be handled easily by choosing a new function $\tilde{g}$ with $\tilde{g}(x) = g(x) - g(\bar{x})$. We prefer however to have explicit bounds on the constants involved, which is important in applications.

Proof. Choose μ and κ' as required and then α and β to satisfy (10). For any $x \in \mathbb{B}_\alpha(\bar{x})$ and $y \in \mathbb{B}_\beta(\bar{y})$, using (11), (12) and the triangle inequality, we obtain

$$\begin{aligned} \|-g(x)+y-\bar{y}\| &\le \|g(\bar{x})\| + \|g(\bar{x})-g(x)\| + \|y-\bar{y}\| \\ &\le \beta + \mu\|x-\bar{x}\| + \beta \le 2\beta + \mu\alpha \le b, \end{aligned} \tag{14}$$

where the last inequality follows from the second inequality in (10). Fix $y' \in \mathbb{B}_\beta(\bar{y})$ and consider the mapping

$$\mathbb{B}_\alpha(\bar{x}) \ni x \mapsto \Phi_{y'}(x) := F^{-1}(-g(x)+y') \cap \mathbb{B}_\alpha(\bar{x}).$$

Let $y \in \mathbb{B}_\beta(\bar{y})$, $y \neq y'$ and let $x \in (g+F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$. We will apply Theorem 5E.2 with the complete metric space X identified with the closed ball $\mathbb{B}_a(\bar{x})$ to show that there is a fixed point $x' \in \Phi_{y'}(x')$ in the closed ball centered at x with radius

$$r := \kappa'\|y-y'\|. \tag{15}$$

From the third inequality in (10), we obtain

$$r \le \kappa'(2\beta) \le \alpha.$$

Hence, from the first inequality in (10) we get $\mathbb{B}_r(x) \subset \mathbb{B}_a(\bar{x})$. Let $(x_n, z_n) \in \operatorname{gph} \Phi_{y'} \cap (\mathbb{B}_r(x) \times \mathbb{B}_r(x))$ and $(x_n, z_n) \to (\tilde{x}, \tilde{z})$. From (14), $\|-g(x_n)+y'-\bar{y}\| \le b$; also note that $\|z_n - \bar{x}\| \le a$. Using (8) and passing to the limit we obtain that $(\tilde{x}, \tilde{z}) \in \operatorname{gph} \Phi_{y'} \cap (\mathbb{B}_r(x) \times \mathbb{B}_r(x))$, hence this set is closed.

Since $y \in g(x)+F(x)$ and (x,y) satisfies (14), from the assumed metric regularity of F we have

$$\begin{aligned} d(x, \Phi_{y'}(x)) &= d(x, F^{-1}(-g(x)+y')) \le \kappa d(-g(x)+y', F(x)) \\ &= \kappa d(y', g(x)+F(x)) \le \kappa\|y-y'\| \\ &< \kappa'\|y-y'\|(1-\kappa\mu) = r(1-\kappa\mu). \end{aligned}$$

For any $u, v \in \mathbb{B}_r(x)$, using (12), we have

$$\begin{aligned} e(\Phi_{y'}(u) \cap \mathbb{B}_r(x), \Phi_{y'}(v)) &\le e(F^{-1}(-g(u)+y') \cap \mathbb{B}_a(\bar{x}), F^{-1}(-g(v)+y')) \\ &\le \kappa\|g(u)-g(v)\| \le \kappa\mu\,\|u-v\|. \end{aligned}$$

Applying 5E.2 to the mapping $\Phi_{y'}$, with $\bar{x}$ identified with x and constants $a = r$ and $\lambda = \kappa\mu$, we obtain the existence of a fixed point $x' \in \Phi_{y'}(x')$, which is equivalent to $x' \in (g+F)^{-1}(y')$, within distance r given by (15) from x. This proves (13).

For the second part of the theorem, suppose that $y \mapsto s(y) := F^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ is a Lipschitz continuous function on $\mathbb{B}_b(\bar{y})$ with a Lipschitz constant κ. Choose μ, κ', α and β as in the statement and let g satisfy (11) and (12). For any $y \in \mathbb{B}_\beta(\bar{y})$, since $\bar{x} \in (g+F)^{-1}(\bar{y}+g(\bar{x})) \cap \mathbb{B}_\alpha(\bar{x})$, from (13) we obtain that there exists $x \in (g+F)^{-1}(y)$ such that

$$\|x-\bar{x}\| \le \kappa'\|y-\bar{y}-g(\bar{x})\|.$$

Since $\|y-\bar{y}-g(\bar{x})\| \le 2\beta$, by (10) we get $\|x-\bar{x}\| \le \alpha$, that is, $(g+F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x}) \neq \emptyset$. Hence the domain of the mapping $(g+F)^{-1} \cap \mathbb{B}_\alpha(\bar{x})$ contains $\mathbb{B}_\beta(\bar{y})$.

If $x \in (g+F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$, then $x \in F^{-1}(y-g(x)) \cap \mathbb{B}_\alpha(\bar{x}) \subset F^{-1}(y-g(x)) \cap \mathbb{B}_a(\bar{x}) = s(y-g(x))$ since $y-g(x) \in \mathbb{B}_b(\bar{y})$ according to (14). Hence,

$$F^{-1}(y-g(x)) \cap \mathbb{B}_\alpha(\bar{x}) = s(y-g(x)) = x. \tag{16}$$

Let $y, y' \in \mathbb{B}_\beta(\bar{y})$. Utilizing the equality $\sigma(y) = s(-g(\sigma(y))+y)$ which comes from (16), we have

$$\begin{aligned} \|\sigma(y)-\sigma(y')\| &= \|s(-g(\sigma(y))+y) - s(-g(\sigma(y'))+y')\| \\ &\le \kappa\|g(\sigma(y)) - g(\sigma(y'))\| + \kappa\|y-y'\| \le \kappa\mu\|\sigma(y)-\sigma(y')\| + \kappa\|y-y'\|. \end{aligned} \tag{17}$$

If $y = y'$, taking into account that $\kappa\mu < 1$ we obtain that $\sigma(y)$ must be equal to $\sigma(y')$. Hence, the mapping $y \mapsto \sigma(y) := (g+F)^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$ is single-valued, that is, a function, defined on $\mathbb{B}_\beta(\bar{y})$. From (17) this fucntion satisfies

$$\|\sigma(y)-\sigma(y')\| \le \kappa'\|y-y'\|.$$

The proof is complete. □

5H. Further Extensions in Metric Spaces

In this section we extend the Lyusternik–Graves theorem 5E.1 to nonlinear metric spaces. The key to obtaining such extensions is to employ regularity properties on a set rather than just at a point. In this section X and Y are metric spaces with both metrics denoted by ρ, as in 5E.

Metric regularity on a set. *Let U, V be nonempty subsets of X and Y, respectively. A set-valued mapping F from X to Y is said to be metrically regular on U for V when the set* $\operatorname{gph} F \cap (U \times V)$ *is closed and there is a constant $\kappa \ge 0$ such that*

$$d(x, F^{-1}(y)) \le \kappa d(y, F(x) \cap V) \quad \textit{for all } (x,y) \in U \times V. \tag{1}$$

A particular case of metric regularity on a set was displayed in Proposition 3C.1, where it was shown for the inverse F^{-1} of F and $U = X$, $V \subset \operatorname{dom} F^{-1}$ that (1) is equivalent to the Lipschitz continuity of F^{-1} on V with respect to the Pompeiu-Hausdorff distance.

Observe that on the right side of (1) the value $F(x)$ is intersected with the set V; this makes metric regularity on a set more general than the metric regularity at a point introduced in Section 3E and studied in detail earlier in the current chapter.

Recall that a mapping $F : X \rightrightarrows Y$ is said to be metrically regular at $\bar{x}$ for $\bar{y}$ when $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ and there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ with $\operatorname{gph} F \cap (U \times V)$ closed and a constant $\kappa > 0$ such that

$$(2) \qquad d(x, F^{-1}(y)) \leq \kappa d(y, F(x)) \quad \text{for all } (x, y) \in U \times V.$$

The infimum of κ over all combinations of κ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ in (2) is the modulus of metric regularity $\operatorname{reg}(F; \bar{x}|\bar{y})$; then the presence of metric regularity of F at $\bar{x}$ for $\bar{y}$ is identified with $\operatorname{reg}(F; \bar{x}|\bar{y}) < \infty$.

If F is metrically regular at a point $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ with neighborhoods U of $\bar{x}$ and V of $\bar{y}$ with constant κ, then F is clearly metrically regular on U for V with the same constant κ. Conversely, when the sets U and V are neighborhoods of points $\bar{x}$ and $\bar{y}$ with $\bar{y} \in F(\bar{x})$, then metric regularity on U for V becomes equivalent to metric regularity at $\bar{x}$ for $\bar{y}$ but perhaps with *different* neighborhoods. We will supply the latter statement with a proof.

Proposition 5H.1 (metric regularity on a set implies metric regularity at a point)**.** *For positive scalars a, b and κ, and points $\bar{x} \in X$, $\bar{y} \in Y$ consider a mapping $F : X \rightrightarrows Y$ with $\bar{y} \in F(\bar{x})$ and assume that F is metrically regular on $\mathbb{B}_a(\bar{x})$ for $\mathbb{B}_b(\bar{y})$ with constant κ. Then F is metrically regular at $\bar{x}$ for $\bar{y}$ with constant κ.*

Proof. Set $\beta = b/4$ and $\alpha = \min\{a, \kappa b/4\}$. Let $x \in \mathbb{B}_\alpha(\bar{x})$ and $y \in \mathbb{B}_\beta(\bar{y})$. If $F(x) = \emptyset$ the right side of (2) is ∞. If not, let $y' \in F(x)$. We consider two cases. First, if $\rho(y, y') > b/2$, then we have

$$\begin{aligned} d(x, F^{-1}(y)) &\leq \rho(x, \bar{x}) + d(\bar{x}, F^{-1}(y)) \leq \rho(x, \bar{x}) + \kappa d(y, F(\bar{x}) \cap \mathbb{B}_b(\bar{y})) \\ &\leq \rho(x, \bar{x}) + \kappa \rho(y, \bar{y}) \leq \kappa b/4 + \kappa b/4 = \kappa b/2 < \kappa \rho(y, y'). \end{aligned}$$

Further, if $\rho(y, y') \leq b/2$, then

$$\rho(y', \bar{y}) \leq \rho(y', y) + \rho(y, \bar{y}) \leq b/2 + b/4 = 3b/4,$$

hence

$$d(x, F^{-1}(y)) \leq \kappa d(y, F(x) \cap \mathbb{B}_b(\bar{y})) \leq \kappa \rho(y, y').$$

Thus, we obtain that for any $y' \in F(x)$ we have $d(x, F^{-1}(y)) \leq \kappa \rho(y, y')$. Taking infimum of the right side over $y' \in F(x)$ we obtain (2). $\square$

As an example showing the difference between metric regularity on a set and metric regularity at a point, consider the mapping $F : \mathbb{R} \to \mathbb{R}$ whose graph is $\{(x, y) \mid y = 0\}$. Then F is metrically regular on any closed set U for $V = \{0\}$ with any constant $\kappa > 0$ but it is not metrically regular at any $\bar{x}$ for 0.

We focus next on the equivalence of the metric regularity with the Aubin property of the inverse and linear openness, where the differences between these properties on a set and at a point become more visible. But first, let us introduce the Aubin property and linear openness on a set.

Aubin property on a set. *Let U and V be nonempty subsets of X and Y respectively. A mapping $S : Y \rightrightarrows X$ is said to have the Aubin property on V for U when the set* $\operatorname{gph} S \cap (V \times U)$ *is closed and there exists a constant $\kappa \geq 0$ such that*

$$(3) \qquad e(S(y) \cap U, S(y')) \leq \kappa \rho(y, y') \quad \text{for all} \quad y, y' \in V.$$

Clearly, when U and V are assumed to be neighborhoods of reference points $\bar{x}$ and $\bar{y}$, respectively, with $(\bar{y}, \bar{x}) \in \operatorname{gph} S$, we obtain the Aubin property at the point which we introduced in Section 3E. If $U = X$, then the Aubin property becomes the usual Lipschitz continuity on V with respect to the Pompeiu-Hausdorff distance. Note that in that case S must have closed graph.

Next comes a definition of openness with linear rate on a set.

Linear openness on a set. *Let U and V be two nonempty subsets of X and Y respectively. A mapping $F : X \rightrightarrows Y$ is said to be open with linear rate (or linearly open) on U for V when the set* $\operatorname{gph} F \cap (U \times V)$ *is closed and there exists a constant $\kappa > 0$ such that*

$$(4) \quad \operatorname{int} \mathbb{B}_r(y) \cap V \subset F(\operatorname{int} \mathbb{B}_{\kappa r}(x)) \quad \text{for all} \quad r \in (0, \infty] \quad \text{and} \quad (x, y) \in \operatorname{gph} F \cap (U \times V).$$

In Section 3 we established equivalence between metric regularity, linear openness, and the inverse mapping having the Aubin property, all *at a point* in the graph of the mapping; this equivalence was stated in 5A.3 in the setting of Banach spaces. For completeness we state this result once again, now in metric spaces.

Theorem 5H.2 (metric regularity, linear openness and Aubin property of the inverse at a point)**.** *For metric spaces X and Y, a mapping $F : X \rightrightarrows Y$, and a constant $\kappa > 0$, the following properties with respect to a pair $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ are equivalent:*

(a) *F is linearly open at $\bar{x}$ for $\bar{y}$ with constant κ;*

(b) *F is metrically regular at $\bar{x}$ for $\bar{y}$ with constant κ;*

(c) *F^{-1} has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant κ.*

Moreover $\operatorname{reg}(F; \bar{x} | \bar{y}) = \operatorname{lip}(F^{-1}; \bar{y} | \bar{x})$.

Taking into account 5H.1, it is now clear that the properties in (a)(b)(c) are equivalent with the same constant κ but perhaps with different neighborhoods of the points $\bar{x}$ and $\bar{y}$.

In the following theorem we show that, when we switch to definitions on a set as given above rather than at a point, metric regularity on a set of a mapping F is equivalent to the Aubin property on a set of the inverse F^{-1} and also to linear openness on a set of F with the same sets U and V and the same constant κ.

Theorem 5H.3 (metric regularity, linear openness and Aubin property of the inverse on a set)**.** *Let U and V be nonempty subsets of X and Y respectively, let $\kappa > 0$, and consider a mapping $F : X \rightrightarrows Y$ such that*

$$(5) \qquad \operatorname{gph} F \cap (U \times V) \neq \emptyset.$$

Then the following are equivalent:

(a) F *is metrically regular on* U *for* V *with constant* κ*;*
(b) F^{-1} *has the Aubin property on* V *for* U *with constant* κ*;*
(c) F *is linearly open on* U *for* V *with constant* κ*.*

Proof. Each of (a), (b) and (c) requires the set $\operatorname{gph} F \cap (U \times V)$ be closed. Let (a) hold. First, note that by (5) there exists $\bar{x} \in U$ such that $F(\bar{x}) \cap V \neq \emptyset$. Then from (1) it follows that $d(\bar{x}, F^{-1}(y)) < \infty$ for any $y \in V$. Thus, $V \subset \operatorname{dom} F^{-1}$. Now, fix $y, y' \in V$. Since $V \subset \operatorname{dom} F^{-1}$, we have $F^{-1}(y') \neq \emptyset$. If $F^{-1}(y) \cap U = \emptyset$, then the left side of (3) is zero and hence (3) is automatically satisfied. Let $x \in F^{-1}(y) \cap U$. Then, from (1),

$$d(x, F^{-1}(y')) \leq \kappa d(y', F(x) \cap V) \leq \kappa \rho(y', y)$$

since $y \in F(x) \cap V$. Taking the supremum of the left side with respect to $x \in F^{-1}(y) \cap U$ we obtain (3), that is, (b).

Assume (b). By (5) there exists $(x, y) \in \operatorname{gph} F \cap (U \times V)$. Let $r > 0$. If $\operatorname{int} \mathbb{B}_r(y) \cap V = \emptyset$ then (4) holds automatically. For any $w \in V$ from (3) we have that $e(F^{-1}(y) \cap U, F^{-1}(w)) \leq \kappa \rho(y, w) < \infty$ and since $x \in F^{-1}(y) \cap U \neq \emptyset$, we get that $F^{-1}(w) \neq \emptyset$. But then $V \subset \operatorname{dom} F^{-1}$. Let $y' \in \operatorname{int} \mathbb{B}_r(y) \cap V$; then $y' \in \operatorname{rge} F = \operatorname{dom} F^{-1}$ and hence $F^{-1}(y') \neq \emptyset$. We have

$$d(x, F^{-1}(y')) \leq e(F^{-1}(y) \cap U, F^{-1}(y')) \leq \kappa \rho(y, y') < \kappa r.$$

Hence, there exists $x' \in F^{-1}(y')$ with $\rho(x, x') < \kappa r$, that is, $x' \in \operatorname{int} \mathbb{B}_{\kappa r}(x)$. Thus $y' \in F(x') \subset F(\operatorname{int} \mathbb{B}_{\kappa r}(x))$ and we obtain (4), that is, (c) is satisfied.

Now, assume (c). Let $x \in U$, $y \in V$ and let $y' \in F(x) \cap V$; if there is no such y' the right side in (1) is ∞ and we are done. If $y = y'$ then (1) holds since both the left and the right sides are zero. Let $r := \rho(y, y') > 0$ and let $\varepsilon > 0$. Then of course $y \in \operatorname{int} \mathbb{B}_{r(1+\varepsilon)}(y') \cap V$. From (c) there exists $x' \in \operatorname{int} \mathbb{B}_{\kappa r(1+\varepsilon)}(x) \cap F^{-1}(y)$. Then

$$d(x, F^{-1}(y)) \leq \rho(x, x') \leq \kappa r(1+\varepsilon) = \kappa(1+\varepsilon)\rho(y, y'). \tag{6}$$

Taking infimum in the right side of (6) with respect to $\varepsilon > 0$ and $y' \in F(x) \cap V$ we obtain (1) and hence (a). The proof is complete. □

Note that if condition (5) is violated, then (a) and (c) hold automatically, whereas (b) holds if and only if $V \subset \operatorname{rge} F$. Also note that for metric regularity at a point the condition (5) is always satisfied.

It turns out that if instead of (1) we choose (2) for a (stronger) definition of metric regularity for the sets U and V, then, in order to have a kind of equivalence displayed in 5H.3, we have to change the definitions of the Aubin property and the linear openness. Specifically, we have the following theorem whose proof is similar to that of 5H.3; for completeness and because of some subtle differences we present it in full.

Theorem 5H.4 (equivalence of alternative definitions)**.** *Let U and V be nonempty sets in X and Y respectively, let $\kappa > 0$ and consider a mapping $F : X \rightrightarrows Y$ such that condition (5) is fulfilled. Then the following are equivalent:*

(a) $d(x, F^{-1}(y)) \leq \kappa d(y, F(x))$ *for all* $(x, y) \in U \times V$;
(b) $e(F^{-1}(y') \cap U, F^{-1}(y)) \leq \kappa\rho(y, y')$ *for all* $y' \in \operatorname{rge} F$ *and* $y \in V$;
(c) $\operatorname{int} \mathbb{B}_r(F(x)) \cap V \subset F(\operatorname{int} \mathbb{B}_{\kappa r}(x))$ *for all* $r \in (0, \infty]$ *and* $x \in U$.

Proof. Let (a) hold. By (5) there exists $\bar{x} \in U$ such that $F(\bar{x}) \neq \emptyset$. Then from (a) $F^{-1}(y) \neq \emptyset$ for any $y \in V$, hence $V \subset \operatorname{dom} F^{-1}$. Now, let $y' \in \operatorname{rge} F$ and $y \in V$. Then $F^{-1}(y) \neq \emptyset$ and if $F^{-1}(y') \cap U = \emptyset$ then the left side of the inequality in (b) is zero, hence (b) holds automatically. If not, let $x \in U$ be such that $y' \in F(x)$. Applying (a) with so chosen x and y and taking supremum on the left with respect to $x \in F^{-1}(y') \cap U$ we obtain (b).

Assume (b). Let $x \in U$ and $r > 0$. If $\operatorname{int} \mathbb{B}_r(F(x)) \cap V = \emptyset$ then (c) holds automatically. If not, let $y' \in V$ and $y \in F(x)$ be such that $\rho(y', y) < r$. Then, since $y \in \operatorname{rge} F$ we have from (b) that

$$d(x, F^{-1}(y')) \leq e(F^{-1}(y) \cap U, F^{-1}(y')) \leq \kappa\rho(y, y') < \kappa r.$$

Then there exists $x' \in F^{-1}(y') \cap \operatorname{int} \mathbb{B}_{\kappa r}(x)$, that is, $y' \in F(x') \subset F(\operatorname{int} \mathbb{B}_{\kappa r}(x))$ and thus (c) holds.

Assume (c). Let $x \in U$, $y \in V$ and let $y' \in F(x)$; if there is no such y' the right side in (2) is ∞ and hence (a) holds. If $y = y'$ then (a) holds since both the left and the right sides are zero. Let $r := \rho(y, y') > 0$ and let $\varepsilon > 0$. Then $y \in \operatorname{int} \mathbb{B}_{r(1+\varepsilon)}(F(x)) \cap V$. It remains to repeat the last part of the proof of 5H.3. □

All six properties in 5H.3 and 5H.4 become equivalent when understood as properties at a point.

We will now utilize the regularity properties on a set to obtain a generalization of Theorem 5E.1 to the case when Y is not necessarily a linear space and then the perturbation is not additive. We denote by $\operatorname{fix} F$ the set of fixed points of a mapping F.

Theorem 5H.5 (extended Lysternik–Graves in metric spaces)**.** *Let X be a complete metric space, Y and P be metric spaces and let κ, μ, α and β be positive constants such that $\kappa\mu < 1$. Consider a mapping $F : X \rightrightarrows Y$ and a function $g : P \times X \to Y$, and let $(\bar{p}, \bar{x}) \in P \times X$ and $(\bar{\bar{x}}, \bar{y}) \in X \times Y$ be such that $U := \mathbb{B}_\alpha(\bar{x}) \cap \mathbb{B}_\alpha(\bar{\bar{x}}) \neq \emptyset$. Assume that the set $\operatorname{gph} F \cap (U \times \mathbb{B}_\beta(\bar{y}))$ is closed, the set $\operatorname{gph} F \cap (\mathbb{B}_\alpha(\bar{\bar{x}}) \times \mathbb{B}_\beta(\bar{y}))$ is nonempty, and F is metrically regular on $\mathbb{B}_\alpha(\bar{\bar{x}})$ for $\mathbb{B}_\beta(\bar{y})$ with constant κ. Also, assume that there exists a neighborhood Q of $\bar{p}$ such that g is continuous in $Q \times X$, Lipschitz continuous with respect to x in $\mathbb{B}_\alpha(\bar{x})$ uniformly in $p \in Q$ with constant μ, and satisfies*

$$\rho(\bar{y}, g(p, x)) \leq \beta \quad \text{for every } p \in Q \text{ and } x \in U. \tag{7}$$

Let c and ε be positive constants such that

$$c + \varepsilon \le \alpha. \tag{8}$$

Then for evey $x \in \mathbb{B}_c(\bar{x}) \cap \mathbb{B}_c(\bar{\bar{x}})$ *and* $p \in Q$ *that satisfy*

$$\frac{\kappa}{1-\kappa\mu} d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})) < \varepsilon \tag{9}$$

one has

$$d(x, \operatorname{fix}(F^{-1}(g(p,\cdot)))) \le \frac{\kappa}{1-\kappa\mu} d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})). \tag{10}$$

In particular, there exists a fixed point of the mapping $F^{-1}(g(p,\cdot))$ *which is at distance from* x *less than* ε.

Proof. From the assumed metric regularity of F and 5H.3 with condition (5) satisfied, we obtain

$$e(F^{-1}(y') \cap \mathbb{B}_\alpha(\bar{\bar{x}}), F^{-1}(y)) \le \kappa\rho(y', y) \quad \text{for all} \quad y', y \in \mathbb{B}_\beta(\bar{y}). \tag{11}$$

We also have that

$$\rho(g(p,x'), g(p,x)) \le \mu\rho(x', x) \quad \text{for all} \quad p \in Q \quad \text{and} \quad x', x \in \mathbb{B}_\alpha(\bar{x}). \tag{12}$$

Pick $c > 0$ and $\varepsilon > 0$ that satisfy (8) and then choose $x \in \mathbb{B}_c(\bar{x}) \cap \mathbb{B}_c(\bar{\bar{x}})$ and $p \in Q$ such that (9) holds. If $d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})) = 0$, then $x \in \operatorname{fix}(F^{-1}(g(p,\cdot)))$, the left side of (10) is zero and there is nothing more to prove. Let $d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})) > 0$ and let $\kappa^+ > \kappa$ be such that

$$\frac{\kappa^+}{1-\kappa\mu} d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})) \le \varepsilon.$$

Let

$$\gamma := \frac{\kappa^+}{1-\kappa\mu} d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})). \tag{13}$$

Then $\gamma \le \varepsilon$ and for every $u \in \mathbb{B}_\gamma(x)$ we have from (8) that

$$\rho(u, \bar{x}) \le \rho(u, x) + \rho(x, \bar{x}) \le \gamma + c \le \varepsilon + c \le \alpha.$$

In the same way, $\rho(u, \bar{\bar{x}}) \le \alpha$, and hence $\mathbb{B}_\gamma(x) \subset U$. We apply Theorem 5E.2 to the mapping $x \mapsto \Phi_p(x) := F^{-1}(g(p,x))$ with the following specifications: $\bar{x} = x$, $a = \gamma$ and $\lambda = \kappa\mu$. Clearly, the set $\operatorname{gph} \Phi_p \cap (\mathbb{B}_\gamma(x) \times \mathbb{B}_\gamma(x))$ is closed. Furthermore, utilizing metric regularity of F on a set, (7) and (13), we obtain

$$\begin{aligned} d(x, \Phi_p(x)) = d(x, F^{-1}(g(p,x))) &\le \kappa d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})) \\ &< \kappa^+ d(g(p,x), F(x) \cap \mathbb{B}_\beta(\bar{y})) = \gamma(1-\kappa\mu). \end{aligned}$$

Also, for any $u, v \in I\!B_\gamma(x)$, from (8), (11), (12) and (13) we have

$$\begin{aligned} e(\Phi_p(u) \cap I\!B_\gamma(x), \Phi_p(v)) &\le e(F^{-1}(g(p,u)) \cap I\!B_\alpha(\bar{\bar{x}}), F^{-1}(g(p,v))) \\ &\le \kappa\rho(g(p,u), g(p,v)) \le \kappa\mu\rho(u,v). \end{aligned}$$

Hence, from 5E.2 we obtain the existence of $\tilde{x} \in \Phi_p(\tilde{x}) \cap I\!B_\gamma(x)$, that is, $\tilde{x} \in F^{-1}(g(p,\tilde{x})) \cap I\!B_\gamma(x)$. Using (13) and noting that κ^+ can be arbitrarily close to κ, we complete the proof. □

It turns out that 5H.5 not only follows from but is actually *equivalent* to Theorem 5E.2:

Proof of Theorem 5E.2 from Theorem 5H.5. We apply Theorem 5H.5 with $X = Y = P$, $F = \Phi^{-1}$, $g(p,x) = x$, $\kappa = \lambda$, $\mu = 1$ and $\alpha = \beta = a$. Then we choose $\bar{\bar{x}}$, $\bar{x}$ and $\bar{y}$ in 5H.5 all equal to $\bar{x}$ in 5E.2, $c = a(1-\kappa)$ and $\varepsilon = a\kappa$. By assumption, $\Phi = F^{-1}$ has the Aubin property on $I\!B_a(\bar{x})$ for $I\!B_a(\bar{x})$ with constant λ; hence, $I\!B_a(\bar{x}) \subset \operatorname{dom} \Phi = \operatorname{rge} F$ and then by 5H.4, F is metrically regular on $I\!B_a(\bar{x})$ for $I\!B_a(\bar{x})$ with constant $\kappa = \lambda$. The conditions (7) and (8) hold trivially. From the assumption (a) in 5E.2, which now becomes $d(\bar{x}, F^{-1}(\bar{x})) < a(1-\kappa)$, it follows that there exists $\tilde{x} \in F^{-1}(\bar{x})$ such that $\rho(\tilde{x}, \bar{x}) < a(1-\kappa) = c$. Then $\bar{x} \in F(\tilde{x})$ and from

$$\frac{\kappa}{1-\kappa} d(\tilde{x}, F(\tilde{x}) \cap I\!B_a(\bar{x})) \le \frac{\kappa}{1-\kappa}\rho(\bar{x}, \tilde{x}) < a\kappa = \varepsilon,$$

thus condition (9) holds for $x = \tilde{x}$. Then, by the last claim in the statement of 5H.5, $\Phi = F^{-1}$ has a fixed point in $I\!B_\varepsilon(\tilde{x})$. But $I\!B_\varepsilon(\tilde{x}) \subset I\!B_a(\bar{x})$ and hence Φ has a fixed point in $I\!B_a(\bar{x})$. □

The following exercise gives a version of 5E.5 in metric spaces:

Exercise 5H.6 (metric Lyusternik–Graves theorem in implicit form)**.** *Let X be a complete metric space, Y and P be metric spaces and let κ, μ and ν be positive constants such that $\kappa\mu < 1$. Consider a mapping $F : X \rightrightarrows Y$ and a function $g : P \times X \to Y$, and let $\bar{x} \in X$, $\bar{p} \in P$ and $\bar{y} \in Y$ be such that $\bar{y} \in F(\bar{x})$ and $\bar{y} = g(\bar{p}, \bar{x})$. Assume that F is metrically regular at $\bar{x}$ for $\bar{y}$ with $\operatorname{reg}(F; \bar{x}\,|\,\bar{y}) < \kappa$ and g is Lipschitz continuous around $(\bar{p}, \bar{x})$ with $\widehat{\operatorname{lip}}_x(g; (\bar{p}, \bar{x})) < \mu$ and $\widehat{\operatorname{lip}}_p(g; (\bar{p}, \bar{x})) < \nu$. Then the mapping $p \mapsto \operatorname{fix}(F^{-1}(g(p, \cdot)))$ has the Aubin property at $\bar{p}$ for $\bar{x}$ with constant $\kappa\nu/(1-\kappa\mu)$.*

Guide. Follow the proof of 5E.5 with $h + F$ there replaced by F and $r(p,x)$ there replaced by $g(p,x)$. Apply 5E.2 to the mapping $x \mapsto \Phi_p(x) = F^{-1}(g(p,x))$ noting that $x \in \Phi_p(x)$ whenever $x \in \operatorname{fix}(F^{-1}(g(p, \cdot)))$. □

Sometimes it is more convenient to use the following version of linear openness at a point: there exist positive constants κ, δ and h_0 such that the set $\operatorname{gph} F \cap (I\!B_\delta(\bar{x}) \times I\!B_\delta(\bar{y}))$ is closed and

(14) $\mathbb{B}_h(y) \subset F(\mathbb{B}_{\kappa h}(x))$ for all $h \in [0,h_0]$ and $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_\delta(\bar{x}) \times \mathbb{B}_\delta(\bar{y}))$.

Proposition 5H.7 (equivalence to metric regularity)**.** *The property described in (14) is equivalent to metric regularity of F at $\bar{x}$ for $\bar{y}$ with constant κ.*

Proof. Let (14) be satisfied and choose $a > 0$ such that $a < \min\{\delta, h_0/2\}$. Pick $x \in \mathbb{B}_a(\bar{x})$ and $y \in \mathbb{B}_a(\bar{y})$. Let $F(x) \cap \mathbb{B}_a(\bar{y}) \neq \emptyset$. Choose $y' \in F(x) \cap \mathbb{B}_a(\bar{y})$. Since $\rho(y,y') < h_0$ there exists $\varepsilon > 0$ such that $h := \rho(y,y')(1+\varepsilon) < h_0$. Then $y \in \mathbb{B}_h(y')$ and from (14) there exists $x' \in F^{-1}(y) \cap \mathbb{B}_{\kappa h}(x)$. Thus

$$d(x,F^{-1}(y)) \leq \rho(x,x') \leq \kappa h = \kappa \rho(y,y')(1+\varepsilon),$$

and taking the limit with $\varepsilon \to 0$ and infimum with respect to $y' \in F(x) \cap \mathbb{B}_a(\bar{y})$ we obtain

$$d(x,F^{-1}(y)) \leq \kappa d(y,F(x) \cap \mathbb{B}_a(\bar{y})).$$

If $F(x) \cap \mathbb{B}_a(\bar{y}) = \emptyset$ this holds automatically. Hence F is metrically regular on $\mathbb{B}_a(\bar{x})$ for $\mathbb{B}_a(\bar{y})$ with constant κ. From 5H.1 we obtain that F is metrically regular at $\bar{x}$ for $\bar{y}$ with constant κ.

Let F^{-1} have the Aubin property with neighborhoods $\mathbb{B}_a(\bar{x})$ and $\mathbb{B}_b(\bar{y})$ and with constant κ, and make b smaller if necessary so that $F^{-1}(y) \cap \mathbb{B}_a(\bar{x}) \neq \emptyset$ for all $y \in \mathbb{B}_b(\bar{y})$ (recall 3E.1). Choose positive δ and h_0 such that $\delta + h_0 \leq b$, then pick $h \in [0,h_0]$. Let $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_\delta(\bar{x}) \times \mathbb{B}_\delta(\bar{y}))$ and let $y' \in \mathbb{B}_h(y)$. Since $y' \in \mathbb{B}_b(\bar{y}) \subset \operatorname{rge} F$ there exists $x' \in F^{-1}(y')$ such that $\rho(x,x') \leq \kappa \rho(y,y') \leq \kappa h$. Hence $y' \in F(\mathbb{B}_{\kappa h}(x))$ and we come to (14). □

At the end of this section we present a new proof of Theorem 5E.1 based on the following result:

Theorem 5H.8 (openness from relaxed openness)**.** *Let X and Y be metric spaces with X being complete and Y having a shift-invariant metric. Consider a mapping $F : X \rightrightarrows Y$ and a point $(\bar{x},\bar{y}) \in \operatorname{gph} F$ at which $\operatorname{gph} F$ is locally closed. Let τ and μ be positive constants such that $\tau > \mu$. Then the following are equivalent:*

(a) *there exists $a > 0$ and $\alpha > 0$ such that for every $h \in [0,a]$*

(15) $\mathbb{B}_{\tau h}(y) \subset \mathbb{B}_{\mu h}(F(\mathbb{B}_h(x)))$ *whenever* $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_\alpha(\bar{x}) \times \mathbb{B}_\alpha(\bar{y}))$;

(b) *there exist $b > 0$ and $\beta > 0$ such that for every $h \in [0,b]$*

(16) $\mathbb{B}_{(\tau-\mu)h}(y) \subset F(\mathbb{B}_h(x))$ *whenever* $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_\beta(\bar{x}) \times \mathbb{B}_\beta(\bar{y}))$.

Proof. Clearly, (b) implies (a). Let (a) hold. Without loss of generality, let the set $\operatorname{gph} F \cap (\mathbb{B}_\alpha(\bar{x}) \times \mathbb{B}_\alpha(\bar{y}))$ be closed. Let $\gamma := \mu/\tau \in (0,1)$ and choose positive b and β such that

(17) $$b \leq a, \quad b + \beta \leq \alpha \text{ and } (\mu+\tau)(1-\gamma)b + \beta \leq \alpha.$$

Let $(x,y) \in \operatorname{gph} F \cap (\mathbb{B}_\beta(\bar{x}) \times \mathbb{B}_\beta(\bar{y}))$ and let $0 < h \le b$. We construct next a sequence (x_n, y_n) by induction, with starting point $(x_0, y_0) = (x, y)$.

Since $(1-\gamma)h < a$, we have

$$\mathbb{B}_{\tau(1-\gamma)h}(y_0) \subset \mathbb{B}_{\mu(1-\gamma)h}(F(\mathbb{B}_{(1-\gamma)h}(x_0))).$$

Let $v \in \mathbb{B}_{\tau(1-\gamma)h}(y_0)$; then $v \in \mathbb{B}_{\mu(1-\gamma)h}(F(\mathbb{B}_{(1-\gamma)h}(x_0)))$. Hence there exists $(x_1, y_1) \in \operatorname{gph} F$ such that

$$\rho(x_1, x_0) \le (1-\gamma)h \quad \text{and} \quad \rho(y_1, v) \le \mu(1-\gamma)h. \tag{18}$$

Using (17), we obtain

$$\rho(x_1, \bar{x}) \le \rho(x_1, x) + \rho(x, \bar{x}) \le (1-\gamma)h + \beta \le \alpha \tag{19}$$

and

$$\rho(y_1, \bar{y}) \le \rho(y_1, v) + \rho(v, y) + \rho(y, \bar{y}) \le \mu(1-\gamma)h + \tau(1-\gamma)h + \beta \le \alpha. \tag{20}$$

Noting that $\gamma(1-\gamma)h < a$ and $\tau\gamma = \mu$, from (15) we get

$$\mathbb{B}_{\mu(1-\gamma)h}(y_1) = \mathbb{B}_{\tau\gamma(1-\gamma)h}(y_1) \subset \mathbb{B}_{\mu(1-\gamma)\gamma h}(F(\mathbb{B}_{(1-\gamma)\gamma h}(x_1))).$$

Thus, from the second inequality in (18) we obtain

$$v \in \mathbb{B}_{\mu(1-\gamma)h}(y_1) \subset \mathbb{B}_{\mu(1-\gamma)\gamma h}(F(\mathbb{B}_{(1-\gamma)\gamma h}(x_1))).$$

Hence, there exists $(x_2, y_2) \in \operatorname{gph} F$ such that

$$\rho(x_2, x_1) \le \gamma(1-\gamma)h \quad \text{and} \quad \rho(v, y_2) \le \gamma\mu(1-\gamma)h.$$

By induction, let $(x_i, y_i) \in \operatorname{gph} F$, $i = 1, 2, \ldots, n$ be such that, for all such i,

$$\rho(x_i, x_{i-1}) \le \gamma^{i-1}(1-\gamma)h \quad \text{and} \quad \rho(v, y_i) \le \mu(1-\gamma)\gamma^{i-1}h.$$

Then,

$$\rho(x_n, x) \le \sum_{i=1}^{n} \rho(x_i, x_{i-1}) \le \sum_{i=1}^{\infty} \gamma^{i-1}(1-\gamma)h = h. \tag{21}$$

Using the second inequality in (17), we have

$$\begin{aligned} \rho(x_n, \bar{x}) &\le \rho(x_n, x) + \rho(x, \bar{x}) \\ &\le \sum_{i=1}^{n} \rho(x_i, x_{i-1}) + \beta \le h + \beta \le b + \beta \le \alpha. \end{aligned}$$

Furthermore, by the third inequality in (17),

$$\begin{aligned}\rho(y_n,\bar{y}) &\le \rho(y_n,v)+\rho(v,y)+\rho(y,\bar{y})\\ &\le \mu(1-\gamma)\gamma^{n-1}h+\tau(1-\gamma)h+\beta\\ &\le (\mu+\tau)(1-\gamma)b+\beta\le\alpha.\end{aligned}$$

By repeating the argument used in the first step, we get

$$v\in I\!\!B_{\mu(1-\gamma)\gamma^n h}(F(I\!\!B_{(1-\gamma)\gamma^n h}(x_n))),$$

hence there exists $(x_{n+1},y_{n+1})\in \operatorname{gph} F$ such that

$$\rho(v,y_{n+1})\le\gamma^n\mu(1-\gamma)h \text{ and } \rho(x_{n+1},x_n)\le\gamma^n(1-\gamma)h. \tag{22}$$

We obtain a sequence $\{(x_n,y_n)\}$ such that $(x_n,y_n)\in\operatorname{gph} F$ and $\{x_n\}$ is a Cauchy sequence, hence convergent to some $u\in I\!\!B_h(x)$, from (21). Furthermore from (22), y_n converges to v, hence $(u,v)\in\operatorname{gph} F$. Since v is arbitrary in $I\!\!B_{\tau(1-\gamma)h}(y)$ we obtain (16). □

From the above result we can derive Theorem 5E.1 in yet another way.

Proof of Theorem 5E.1 from 5H.8. By the equivalence of metric regularity at $\bar{x}$ for $\bar{y}$ with constant $\kappa>0$ and the property (14) established in 5H.7, there exist positive constants α and h_0 such that for every $(x,y)\in\operatorname{gph} F\cap(I\!\!B_\alpha(\bar{x})\times I\!\!B_\alpha(\bar{y}))$ we have

$$I\!\!B_{\tau h}(y)\subset F(I\!\!B_h(x)) \text{ for all } h\in[0,h_0], \tag{23}$$

where $\tau=1/\kappa$. Let $a>0$ satisfy $a(1+\mu)\le\alpha$ and let $(u,v)\in\operatorname{gph}(g+F)\cap(I\!\!B_a(\bar{x})\times I\!\!B_a(\bar{y}+g(\bar{x})))$. Then

$$\rho(v-g(u),\bar{y})\le\rho(v,\bar{y}+g(\bar{x}))+\rho(g(u),g(\bar{x}))\le a+\mu a\le\alpha.$$

Also, since $v\in(g+F)(u)$, we obtain that $(u,v-g(u))\in\operatorname{gph} F$, that is, there exists $y\in F(u)$ such that $y=v-g(u)$.

Let $h\in[0,h_0]$ and let $t\in I\!\!B_{\tau h}(v)$. Then $\rho(t-g(u),y)=\rho(t,v)\le\tau h$, which is the same as $t-g(u)\in I\!\!B_{\tau h}(y)$, and hence, from (23), $t-g(u)\in F(I\!\!B_h(u))$. Thus, there exists $w\in I\!\!B_h(u)$ such that $t-g(u)\in F(w)$. Therefore,

$$t\in F(w)+g(u)=(g+F)(w)+g(u)-g(w)$$

which, by the Lipschitz continuity of g with constant μ, implies

$$I\!\!B_{\tau h}(v)\subset(g+F)(I\!\!B_h(u))+\mu h I\!\!B.$$

Theorem 5H.8 then yields the existence of $b>0$ and $\beta>0$ such that for every $h\in[0,b]$ and every $(x,y)\in\operatorname{gph}(g+F)\cap(I\!\!B_\beta(\bar{x})\times I\!\!B_\beta(\bar{y}+g(\bar{x})))$ one has

$$I\!\!B_{(\tau-\mu)h}(y)\subset(g+F)(I\!\!B_h(x)).$$

In order to apply 5H.7, it remains to observe that the graph of $g+F$ is locally closed at $(\bar{x},\bar{y}+g(\bar{x}))$. Hence, the mapping $g+F$ is metrically regular at $\bar{x}$ for $\bar{y}+g(\bar{x})$ with constant $1/(1/\kappa-\mu)=\kappa/(1-\kappa\mu)$. □

5I. Metric Regularity and Fixed Points

Is a result of the kind given in Theorem 5E.1 valid for *set-valued* perturbations? Specifically, the question is whether the function g can be replaced by a set-valued mapping G, perhaps having the Aubin property with suitable modulus, or even a Lipschitz continuous single-valued localization. The answer to this question turns out to be *no* in general, as the following example confirms.

Example 5I.1 (counterexample for set-valued perturbations)**.** Consider $F:\mathbb{R}\rightrightarrows\mathbb{R}$ and $G:\mathbb{R}\rightrightarrows\mathbb{R}$ specified by

$$F(x)=\{-2x,1\}\qquad\text{and}\qquad G(x)=\{x^2,-1\}\ \text{ for } x\in\mathbb{R}.$$

Then F is metrically regular at 0 for 0 while G has the Aubin property at 0 for 0. Moreover, $\operatorname{reg}(F;0|0)=1/2$ whereas the Lipschitz modulus of the single-valued localization of G around 0 for 0 is 0 and serves also as the infimum of all Aubin constants. The mapping

$$(F+G)(x)=\{x^2-2x,x^2+1,-2x-1,0\}\ \text{ for } x\in\mathbb{R}$$

is not metrically regular at 0 for 0. Indeed, for x and y close to zero and positive we have that $d(x,(F+G)^{-1}(y))=|x-1+\sqrt{1+y}|$, but also $d(y,(F+G)(x))=\min\{|x^2-2x-y|,y\}$. Take $x=\varepsilon>0$ and $y=\varepsilon^2$. Then, since $(\varepsilon-1+\sqrt{1+\varepsilon^2})/\varepsilon^2\to\infty$, the mapping $F+G$ is seen not to be metrically regular at 0 for 0.

It is clear from this example that adding set-valued mappings may lead to mismatching the reference points. One way to avoid such a mismatch is to consider regularity properties on sets rather than at points, what we did in the preceding section. Here we go a step further focusing on global regularity properties. Following the pattern of the preceding section, we define global metric regularity of a mapping $F:X\rightrightarrows Y$, where X and Y are metric spaces, as metric regularity of F on X for Y. This requires F have closed graph. From 5H.3, global metric regularity is equivalent to the Aubin property of F^{-1} on Y for X, which is the same as global Lipschitz continuity of F^{-1} supplemented with closedness of its graph. Recall that a mapping $G:Y\rightrightarrows X$ is globally Lipschitz continuous when it is closed valued and there exists $\mu\geq 0$ (Lipschitz constant) such that

$$h(G(y),G(y'))\leq\mu\rho(y,y')\qquad\text{for all }\ y,y'\in Y.$$

If a mapping G acting from a metric space Y to a complete metric space X is globally Lipschitz continuous, then G necessarily has closed graph. Indeed, let $(y_k,x_k) \in \operatorname{gph} G$ and $(y_k,x_k) \to (y,x)$; then

$$d(x,G(y)) \le \rho(x,x_k) + d(x_k,G(y_k)) + h(G(y_k),G(y)) \le \rho(x,x_k) + \mu\rho(y_k,y) \to 0,$$

and hence $(y,x) \in \operatorname{gph} G$. Having this in mind, we start with a global version of the Lyusternik–Graves theorem.

Theorem 5I.2 (global Lyusternik–Graves theorem)**.** *Let X be a complete metric space and Y be a Banach space. Consider mappings $\Phi : X \rightrightarrows Y$ and $\Psi : X \rightrightarrows Y$ and suppose that Φ is metrically regular with constant $\kappa > 0$ and Ψ is Lipschitz continuous with constant $\mu > 0$, both globally. If $\kappa\mu < 1$ then*

$$d(x,(\Phi+\Psi)^{-1}(y)) \le \frac{\kappa}{1-\kappa\mu} d(y,(\Phi+\Psi)(x)) \quad \text{for all } (x,y) \in X \times Y.$$

Note that the graph of $\Phi+\Psi$ in the statement of 5I.2 may not be closed, what is needed to claim that $\Phi+\Psi$ is globally metrically regular. We will derive Theorem 5I.2 from the more general Theorem 5I.3 given next. As in 5H we denote by $\operatorname{fix} F$ the set of fixed points of a mapping F. For mappings $F : Y \rightrightarrows X$ and $G : X \rightrightarrows Y$, one has $\operatorname{fix}(F \circ G) = \{x \in X \mid F^{-1}(x) \cap G(x) \neq \emptyset\}$. Also, we use the notation $\mathbf{d}(A,B) = \inf\{\rho(x,y) \mid x \in A, y \in B\}$ for the minimal distance between two sets; if either A or B is empty, we set $\mathbf{d}(A,B) = \infty$.

Theorem 5I.3 (fixed points of composition)**.** *Let X and Y be complete metric spaces. Let κ and μ be positive constants such that $\kappa\mu < 1$. Consider a mapping $F : X \rightrightarrows Y$ which is metrically regular with constant κ and a mapping $G : X \rightrightarrows Y$ which is Lipschitz continuous with constant μ, both globally. Then the following inequality holds:*

$$(1) \qquad d(x,\operatorname{fix}(F^{-1} \circ G)) \le \frac{\kappa}{1-\kappa\mu} \mathbf{d}(F(x),G(x)) \quad \text{for every } x \in X.$$

Proof. Let $x \in X$ and $y \in G(x)$. If $x \notin \operatorname{dom} F$ then (1) holds automatically. Let $F(x) \neq \emptyset$. Choose $\varepsilon > 0$. Since $\operatorname{dom} F^{-1} = Y$ there exists $u \in F^{-1}(y)$ such that $\rho(x,u) \le d(x,F^{-1}(y)) + \varepsilon$. If $u = x$ then $x \in \operatorname{fix}(F^{-1} \circ G)$ and the left side of (1) is zero, so there is nothing more to prove. If not, we can write

$$(2) \qquad d(x,F^{-1}(y)) < (1+\varepsilon)\rho(u,x).$$

Let $a := (1+\varepsilon)\rho(u,x)$; then

$$(3) \qquad a \le (1+\varepsilon)(d(x,F^{-1}(y)) + \varepsilon).$$

We will construct a sequence $\{(x^k,y^k)\}$ with the following properties:

$$(4) \qquad y^k \in F(x^{k+1}) \cap G(x^k)$$

and

$$(5) \qquad \rho(y^k, y^{k-1}) < \mu a(\kappa\mu)^{k-1}, \quad \rho(x^{k+1}, x^k) < a(\kappa\mu)^k.$$

Let $x^0 = x$ and $y^0 = y$. In the first step, observe that, from (2), $d(x^0, F^{-1}(y^0)) < a$. Then there exists $x^1 \in F^{-1}(y^0)$ with $\rho(x^1, x^0) < a(\kappa\mu)^0$. Furthermore, since

$$d(y^0, G(x^1)) \le h(G(x^0), G(x^1)) \le \mu\rho(x^0, x^1) < \mu a,$$

there exists $y^1 \in G(x^1)$ such that $\rho(y^1, y^0) < \mu a(\kappa\mu)^0$. From the Lipschitz continuity of F^{-1} we have

$$d(x^1, F^{-1}(y^1)) \le h(F^{-1}(y^0), F^{-1}(y^1)) \le \kappa\rho(y^0, y^1) < \kappa\mu a(\kappa\mu)^0 = a(\kappa\mu).$$

Hence there exists $x^2 \in F^{-1}(y^1)$ with

$$\rho(x^2, x^1) < a(\kappa\mu)^1.$$

We obtain (4) and (5) for $k = 1$.

Proceeding by induction, suppose that we have already found a sequence $\{(x^k, y^k)\}$ satisfying (4) and (5) for $k = 1, \dots, j$, for some $j > 1$. Then

$$d(y^j, G(x^{j+1})) \le h(G(x^j), G(x^{j+1})) \le \mu\rho(x^j, x^{j+1}) < \mu a(\kappa\mu)^j.$$

Hence one can find $y^{j+1} \in G(x^{j+1})$ such that

$$\rho(y^j, y^{j+1}) < \mu a(\kappa\mu)^j.$$

The Lipschitz continuity of F^{-1} gives us

$$d(x^{j+1}, F^{-1}(y^{j+1})) \le h(F^{-1}(y^j), F^{-1}(y^{j+1})) \le \kappa\rho(y^j, y^{j+1}) < \kappa\mu a(\kappa\mu)^j = a(\kappa\mu)^{j+1}.$$

Then there exists $x^{j+2} \in F^{-1}(y^{j+1})$ with

$$\rho(x^{j+2}, x^{j+1}) < a(\kappa\mu)^{j+1}.$$

The induction step is complete. For natural k and m with $k - 1 > m \ge 1$, we have

$$\rho(x^k, x^m) \le \sum_{i=m}^{k-1} \rho(x^{i+1}, x^i) < \frac{a(\kappa\mu)^m}{1 - \kappa\mu}$$

and

$$\rho(y^k, y^m) \le \sum_{i=m}^{k-1} \rho(y^{i+1}, y^i) < \frac{\mu a(\kappa\mu)^m}{1 - \kappa\mu}.$$

Thus $\{(x^k, y^k)\}$ is a Cauchy sequence. Since the space $X \times Y$ is complete, the sequence (x^k, y^k) is convergent to some $(\hat{x}, \hat{y})$. Both F and G have closed graphs, then from (4) we obtain $\hat{y} \in F(\hat{x}) \cap G(\hat{x})$, hence $\hat{x} \in \operatorname{fix}(F^{-1} \circ G)$.

To complete the proof, note that

$$\rho(x^k,x) \leq \sum_{i=0}^{k-1} \rho(x^{i+1},x^i) < a\sum_{i=0}^{k-1}(\kappa\mu)^i \leq \frac{a}{(1-\kappa\mu)}.$$

Passing to the limit with $k \to \infty$ and using (3), we obtain

$$d(x,\operatorname{fix}(F^{-1}\circ G)) \leq \rho(\hat{x},x) \leq \frac{(1+\varepsilon)(d(x,F^{-1}(y))+\varepsilon)}{1-\kappa\mu}.$$

Since ε can be arbitrarily small we get

$$d(x,\operatorname{fix}(F^{-1}\circ G)) \leq \frac{1}{1-\kappa\mu} d(x,F^{-1}(y)). \tag{6}$$

Metric regularity of F combined with (6) gives us

$$d(x,\operatorname{fix}(F^{-1}\circ G)) \leq \frac{\kappa}{1-\kappa\mu} d(y,F(x)).$$

Taking into account that y can be any point in $G(x)$ we obtain (1). □

Note that the estimate (1) is sharp in the sense that if the left side is zero, so is the right side. Possible extensions to noncomplete metric spaces can be made by adapting the proof accordingly, but we shall not go into this further.

Nadler's fixed point theorem 5E.3 easily follows from Theorem 5I.3. Indeed, let X be a complete metric space and let $\Phi : X \rightrightarrows X$ be Lipschitz continuous on X with constant $\lambda \in (0,1)$. Then in particular $\operatorname{dom}\Phi = X$ and Φ has closed graph. Apply 5I.3 with $F^{-1} = \Phi$ and G the identity, obtaining that for any $x \in X = \operatorname{rge} F$ the right side of (1) is finite, hence the set of fixed points of Φ is nonempty.

Proof of 5I.2. Choose $y \in Y$ and let $u \in \operatorname{fix}(\Phi^{-1}\circ(-\Psi+y))$. Then there exists $z \in -\Psi(u)+y$ with $u \in \Phi^{-1}(z)$, therefore $u \in (\Phi+\Psi)^{-1}(y)$. Thus, $\operatorname{fix}(\Phi^{-1}\circ(-\Psi+y)) \subset (\Phi+\Psi)^{-1}(y)$. Therefore, for any $x \in X$ we have

$$d(x,(\Phi+\Psi)^{-1}(y)) \leq d(x,\operatorname{fix}(\Phi^{-1}\circ(-\Psi+y))).$$

Applying Theorem 5I.3 with $F = \Phi$ and $G(\cdot) = -\Psi(\cdot)+y$, from (1) we obtain

$$d(x,(\Phi+\Psi)^{-1}(y)) \leq \frac{\kappa}{1-\kappa\mu}\mathbf{d}(\Phi(x),-\Psi(x)+y) \leq \frac{\kappa}{1-\kappa\mu} d(y,\Phi(x)+\Psi(x)).$$

□

As another application of 5I.3 we obtain the following result:

Theorem 5I.4 (one-sided estimate for fixed points)**.** *Let X be a complete metric space and let T_1 and T_2 map X into the family of nonempty closed subsets of X.*

Suppose that both T_1 and T_2 are globally Lipschitz continuous with the same Lipschitz constant $\mu \in (0,1)$. Then for $i, j \in \{1,2\}$ we have

$$e(\mathrm{fix}(T_i), \mathrm{fix}(T_j)) \le \frac{1}{1-\mu} \sup_{x \in X} e(T_i(x), T_j(x)).$$

Proof from 5I.3. First note that, by using the observation before the statement of 5E.2, both T_1 and T_2 have closed graphs. Then apply Theorem 5I.3 with F the identity mapping and $G = T_1$. From (1) we have that for any $x \in X$,

$$d(x, \mathrm{fix}(T_1)) \le \frac{1}{1-\mu} d(x, T_1(x)). \tag{8}$$

Then, taking supremum in (8) with respect to $x \in \mathrm{fix}T_2$ we have

$$\begin{aligned} e(\mathrm{fix}(T_2), \mathrm{fix}(T_1)) &\le \frac{1}{1-\mu} \sup_{x \in \mathrm{fix}T_2} d(x, T_1(x)) \\ &\le \frac{1}{1-\mu} \sup_{x \in \mathrm{fix}T_2} e(T_2(x), T_1(x)) \le \frac{1}{1-\mu} \sup_{x \in X} e(T_2(x), T_1(x)). \end{aligned}$$

By symmetry, the proof is complete. □

Theorem 5I.4 can be also derived from the contracting mapping principle for set-valued mappings 5E.2.

Proof from 5E.2. As already observed, T_1 has closed graph. Let

$$a = \frac{1}{1-\mu} \sup_{x \in X} e(T_1(x), T_2(x)).$$

The fixed point theorem 5E.2, via Nadler's theorem 5E.3, yields that $\mathrm{fix}T_2 \neq \emptyset$. Let $x \in \mathrm{fix}T_2$. Then

$$d(x, T_1(x)) \le e(T_2(x), T_1(x)) \le a(1-\mu).$$

Furthermore, for any $u, v \in I\!B_a(x)$

$$e(T_1(u) \cap I\!B_a(x), T_1(v)) \le h(T_1(u), T_1(v)) \le \mu\rho(u,v).$$

Hence, from 5E.2 with $\Phi = T_1$ and $\lambda = \mu$, the mapping T_1 has a fixed point $\hat{x} \in I\!B_a(x)$. Then

$$d(x, \mathrm{fixT}_1) \le \rho(x, \hat{x}) \le a = \frac{1}{1-\mu} \sup_{x \in X} e(T_1(x), T_2(x)),$$

and taking supremum with respect to $x \in \mathrm{fix}T_2$ we complete the proof. □

As a corollary of 5I.4 we obtain the following result known as Lim's lemma[3]:

[3] Cf. Lim [1985].

Corollary 5I.5 (Lipschitz estimate for fixed points). *On the assumptions of 5I.5,*

$$h(\mathrm{fix}(T_1), \mathrm{fix}(T_2)) \leq \frac{1}{1-\mu} \sup_{x \in X} h(T_1(x), T_2(x)).$$

Recall that a mapping $F : X \rightrightarrows Y$ is said to be *outer Lipschitz continuous* at $\bar{x}$ relative to a set $D \subset \mathrm{dom}\, F$ when $\bar{x} \in D$ and there exists a constant $\lambda \geq 0$ such that

$$e(F(x), F(\bar{x})) \leq \lambda \rho(x, \bar{x}) \quad \text{for all } \ x \in D.$$

When $D = X$ we say that F is outer Lipschitz continuous at $\bar{x}$. In that case, dom $F = X$. Another corollary of 5I.4 is the following.

Corollary 5I.6 (outer Lipschitz estimate for fixed points). *Let X be a complete metric spaces and Y be metric space. Consider a mapping $M : Y \times X \rightrightarrows X$ having the following properties:*

(i) $M(y, \cdot)$ *is Lipschitz continuous with a Lipschitz constant* $\mu \in (0, 1)$ *uniformly in* $y \in Y$;

(ii) $M(\cdot, x)$ *is outer Lipschitz continuous at* $\bar{y}$ *with a constant* λ *uniformly in* $x \in X$.

Then the mapping $y \mapsto \mathrm{fix}(M(y, \cdot))$ *is outer Lipschitz continuous at* $\bar{y}$ *with constant* $\lambda/(1-\mu)$.

Proof. Applying 5I.4 with $T_1(x) = M(y, x)$ and $T_2(x) = M(\bar{y}, x)$ we have

$$e(\mathrm{fix}(M(y, \cdot), \mathrm{fix}(M(\bar{y}, \cdot)) \leq \frac{1}{1-\mu} \sup_{x \in X} e(M(y, x), M(\bar{y}, x)) \leq \frac{\lambda}{1-\mu} \rho(y, \bar{y}).$$

$\square$

5J. The Bartle–Graves Theorem and Extensions

To set the stage, we begin with a Banach space version of the implication (a) $\Rightarrow$ (b) in the symmetric inverse function theorem 1D.9.

Theorem 5J.1 (inverse function theorem in infinite dimensions). *Let X be a Banach space and consider a function $f : X \to X$ and a point $\bar{x} \in \mathrm{int\,dom}\, f$ at which f is strictly (Fréchet) differentiable and the derivative mapping $Df(\bar{x})$ is invertible. Then the inverse mapping f^{-1} has a single-valued graphical localization s around $\bar{y} := f(\bar{x})$ for $\bar{x}$ which is strictly differentiable at $\bar{y}$, and moreover*

$$Ds(\bar{y}) = [Df(\bar{x})]^{-1}.$$

In Section 1F we considered what may happen (in finite dimensions) when the derivative mapping is merely surjective; by adjusting the proof of Theorem 1F.6 one obtains that when the Jacobian $\nabla f(\bar{x})$ has full rank, the inverse f^{-1} has a local selection which is strictly differentiable at $f(\bar{x})$. The claim can be easily extended to Hilbert (and even more general) spaces:

Exercise 5J.2 (differentiable inverse selections). *Let X and Y be Hilbert spaces and let $f: X \to Y$ be a function which is strictly differentiable at $\bar{x}$ and such that the derivative $A := Df(\bar{x})$ is surjective. Then the inverse f^{-1} has a local selection s around $\bar{y} := f(\bar{x})$ for $\bar{x}$ which is strictly differentiable at $\bar{y}$ with derivative $Ds(\bar{y}) = A^*(AA^*)^{-1}$, where A^* is the adjoint of A.*

Guide. Use the argument in the proof of 1F.6 with adjustments to the Hilbert space setting. Another way of proving this result is to consider the function

$$g : (x,u) \mapsto \begin{pmatrix} x + A^*u \\ f(x) \end{pmatrix} \quad \text{for } (x,u) \in X \times Y,$$

which satisfies $g(\bar{x},0) = (\bar{x},\bar{y})$ and whose Jacobian is

$$J = \begin{pmatrix} I & A^* \\ A & 0 \end{pmatrix}.$$

In the Hilbert space context, if A is surjective then the operator J is invertible. Hence, by Theorem 5J.1, the mapping g^{-1} has a single-valued graphical localization $(\xi,\eta) : (v,y) \mapsto (\xi(v,y), \eta(v,y))$ around $(\bar{x},\bar{y})$ for $(\bar{x},0)$. In particular, for some neighborhoods U of $\bar{x}$ and V of $\bar{y}$, the function $s(y) := \xi(\bar{x},y)$ satisfies $y = f(s(y))$ for $y \in V$. To obtain the formula for the strict derivative, find the inverse of J. □

In the particular case when the function f in 5J.2 is linear, the mapping $A^*(AA^*)^{-1}$ is a continuous linear selection of A^{-1}. A famous result by R. G. Bartle and L. M. Graves [1952] yields that, for arbitrary Banach spaces X and Y, the surjectivity of a mapping $A \in \mathscr{L}(X,Y)$ implies the existence of a continuous local selection of A^{-1}; this selection, however, may not be linear. The original Bartle–Graves theorem is for nonlinear mappings and says the following:

Theorem 5J.3 (Bartle–Graves theorem). *Let X and Y be Banach spaces and let $f: X \to Y$ be a function which is strictly differentiable at $\bar{x}$ and such that the derivative $Df(\bar{x})$ is surjective. Then there is a neighborhood V of $\bar{y} := f(\bar{x})$ along with a continuous function $s: V \to X$ and a constant $\gamma > 0$ such that*

$$(1) \qquad f(s(y)) = y \;\text{ and }\; \|s(y) - \bar{x}\| \le \gamma \|y - \bar{y}\| \;\text{ for every } y \in V.$$

In other words, the surjectivity of the strict derivative at $\bar{x}$ implies that f^{-1} has a local selection s which is continuous around $f(\bar{x})$ and calm at $f(\bar{x})$. It is known[4] that, in contrast to the strictly differentiable local selection in 5J.2 for Hilbert spaces,

[4] Cf. Deville, Godefroy and Zizler [1993], p. 200.

the selection in the Bartle–Graves theorem, even for a bounded linear mapping f, might be not even Lipschitz continuous around $\bar{y}$. For this case we have:

Corollary 5J.4 (inverse selection of a surjective linear mapping in Banach spaces). *For any bounded linear mapping A from X onto Y, there is a continuous (but generally nonlinear) mapping B such that $ABy = y$ for every $y \in Y$.*

Proof. Theorem 5J.3 tells us that A^{-1} has a continuous local selection at 0 for 0. Since A^{-1} is positively homogeneous, this selection is global. □

In this section we develop a generalization of the Bartle–Graves theorem for metrically regular set-valued mappings. First, recall that a mapping $F : Y \rightrightarrows X$ is (sequentially) inner semicontinuous on a set $T \subset Y$ if for every $y \in T$, every $x \in F(y)$ and every sequence of points $y^k \in T$, $y^k \to y$, there exists $x^k \in F(y^k)$ for $k = 1, 2, \ldots$ such that $x^k \to x$ as $k \to \infty$. We also need a basic result which we only state here without proof:

Theorem 5J.5 (Michael selection theorem). *Let X and Y be Banach spaces and consider a mapping $F : Y \rightrightarrows X$ which is closed-convex-valued and inner semicontinuous on* $\operatorname{dom} F \neq \emptyset$. *Then F has a continuous selection $s : \operatorname{dom} F \to X$.*

We require a lemma which connects the Aubin property of a mapping with the inner semicontinuity of a truncation of this mapping:

Lemma 5J.6 (inner semicontinuous selection from the Aubin property). *Consider a mapping $S : Y \rightrightarrows X$ and any $(\bar{y}, \bar{x}) \in \operatorname{gph} S$, and suppose that S has the Aubin property at $\bar{y}$ for $\bar{x}$ with constant κ. Suppose, for some $c > 0$, that the sets $S(y) \cap \mathbb{B}_c(\bar{x})$ are convex and closed for all $y \in \mathbb{B}_c(\bar{y})$. Then for any $\alpha > \kappa$ there exists $\beta > 0$ such that the mapping*

$$y \mapsto M_\alpha(y) = \begin{cases} S(y) \cap \mathbb{B}_{\alpha\|y-\bar{y}\|}(\bar{x}) & \text{for } y \in \mathbb{B}_\beta(\bar{y}), \\ \emptyset & \text{otherwise} \end{cases}$$

is nonempty-closed-convex-valued and inner semicontinuous on $\mathbb{B}_\beta(\bar{y})$.

Proof. Let a and b be positive numbers such that the balls $\mathbb{B}_a(\bar{x})$ and $\mathbb{B}_b(\bar{y})$ are associated with the Aubin property of S (metric regularity of S^{-1}) with a constant κ. Without loss of generality, let $\max\{a, b\} < c$. Fix $\alpha > \kappa$ and choose β such that

$$0 < \beta \leq \min\left\{\frac{a}{\alpha}, \frac{c}{3\alpha}, b, c\right\}.$$

For such a β the mapping M_α has nonempty closed convex values. It remains to show that M_α is inner semicontinuous on $\mathbb{B}_\beta(\bar{y})$.

Let $(y, x) \in \operatorname{gph} M_\alpha$ and $y^k \to y$, $y^k \in \mathbb{B}_\beta(\bar{y})$. First, let $y = \bar{y}$. Then $M_\alpha(y) = \bar{x}$, and from the Aubin property of S there exists a sequence of points $x^k \in S(y^k)$ such that $\|x^k - \bar{x}\| \leq \kappa\|y^k - \bar{y}\|$. Then $x^k \in M_\alpha(y^k)$, $x^k \to x$ as $k \to \infty$ and we are done in this case.

Now let $y \neq \bar{y}$. The Aubin property of S yields that there exists $\check{x}^k \in S(y^k)$ such that

$$\|\check{x}^k - \bar{x}\| \leq \kappa \|y^k - \bar{y}\|$$

and also there exists $\tilde{x}^k \in S(y^k)$ such that

$$\|\tilde{x}^k - x\| \leq \kappa \|y^k - y\|.$$

Then, the choice of β above yields

$$\|\check{x}^k - \bar{x}\| \leq \kappa\beta \leq \alpha \frac{c}{3\alpha} \leq c$$

and

$$\begin{aligned} \|\tilde{x}^k - \bar{x}\| &\leq \|\tilde{x}^k - x\| + \|x - \bar{x}\| \\ &\leq \kappa \|y^k - y\| + \alpha \|y - \bar{y}\| \\ &\leq 2\kappa\beta + \alpha\beta \leq 3\alpha\beta \leq c. \end{aligned}$$

Let

$$\varepsilon^k = \frac{(\alpha + \kappa)\|y^k - y\|}{(\alpha - \kappa)\|y^k - \bar{y}\| + (\alpha + \kappa)\|y^k - y\|}. \tag{2}$$

Then $0 \leq \varepsilon^k < 1$ and $\varepsilon^k \searrow 0$ as $k \to \infty$. Let $x^k = \varepsilon^k \check{x}^k + (1 - \varepsilon^k)\tilde{x}^k$. Then $x^k \in S(y^k)$. Moreover, we have

$$\begin{aligned} \|x^k - \bar{x}\| &\leq \varepsilon^k \|\check{x}^k - \bar{x}\| + (1 - \varepsilon^k)\|\tilde{x}^k - \bar{x}\| \\ &\leq \varepsilon^k \kappa \|y^k - \bar{y}\| + (1 - \varepsilon^k)(\|\tilde{x}^k - x\| + \|x - \bar{x}\|) \\ &\leq \varepsilon^k \kappa \|y^k - \bar{y}\| + (1 - \varepsilon^k)\kappa \|y^k - y\| + (1 - \varepsilon^k)\alpha \|y - \bar{y}\| \\ &\leq \varepsilon^k \kappa \|y^k - \bar{y}\| + (1 - \varepsilon^k)\kappa \|y^k - y\| \\ &\quad + (1 - \varepsilon^k)\alpha \|y^k - \bar{y}\| + (1 - \varepsilon^k)\alpha \|y^k - y\| \\ &\leq \alpha \|y^k - \bar{y}\| - \varepsilon^k(\alpha - \kappa)\|y^k - \bar{y}\| \\ &\quad + (1 - \varepsilon^k)(\alpha + \kappa)\|y^k - y\| = \alpha \|y^k - \bar{y}\|, \end{aligned}$$

where in the last inequality we take into account the expression (2) for ε^k. Thus $x^k \in M_\alpha(y^k)$, and since $x^k \to x$, we are done. □

Lemma 5J.6 allows us to apply Michael's selection theorem to the mapping M_α, obtaining the following result:

Theorem 5J.7 (continuous inverse selection from metric regularity)**.** *Consider a mapping $F : X \rightrightarrows Y$ which is metrically regular at $\bar{x}$ for $\bar{y}$. Let, for some $c > 0$, the sets $F^{-1}(y) \cap \mathbb{B}_c(\bar{x})$ be convex and closed for all $y \in \mathbb{B}_c(\bar{y})$. Then for every $\alpha > \operatorname{reg}(F;\bar{x}\,|\,\bar{y})$ the mapping F^{-1} has a continuous local selection s around $\bar{y}$ for $\bar{x}$ which is calm at $\bar{y}$ with*

$$\text{(3)} \qquad \operatorname{clm}(s;\bar{y}) \le \alpha.$$

Proof. Choose α such that $\alpha > \operatorname{reg}(F;\bar{x}\,|\,\bar{y})$, and apply Michael's theorem 5J.5 to the mapping M_α in 5J.6 for $S = F^{-1}$. By the definition of M_α, the continuous local selection obtained in this way is calm with a constant α. □

Note that the continuous local selection s in 5J.7 depends on α and therefore we cannot replace α in (3) with $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$.

In the remainder of this section we show that if a mapping F satisfies the assumptions of Theorem 5J.7, then for any function $g : X \to Y$ with $\operatorname{lip}(g;\bar{x}) < 1/\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$, the mapping $(g+F)^{-1}$ has a continuous and calm local selection around $g(\bar{x})+\bar{y}$ for $\bar{x}$. We will prove this generalization of the Bartle–Graves theorem by repeatedly using an argument similar to the proof of Lemma 5J.6, the idea of which goes back to (modified) Newton's method used to prove the theorems of Lyusternik and Graves and, in fact, to Goursat's proof on his version of the classical inverse function theorem. We put the theorem in the format of the general implicit function theorem paradigm:

Theorem 5J.8 (inverse mappings with continuous calm local selections)**.** *Consider a mapping $F : X \rightrightarrows Y$ and any $(\bar{x},\bar{y}) \in \operatorname{gph} F$ and suppose that for some $c > 0$ the mapping $I\!B_c(\bar{y}) \ni y \mapsto F^{-1}(y) \cap I\!B_c(\bar{x})$ is closed-convex-valued. Consider also a function $g : X \to Y$ with $\bar{x} \in \operatorname{int}\operatorname{dom} g$. Let κ and μ be nonnegative constants such that*

$$\kappa\mu < 1, \quad \operatorname{reg}(F;\bar{x}\,|\,\bar{y}) \le \kappa \quad \text{and} \quad \operatorname{lip}(g;\bar{x}) \le \mu.$$

Then for every γ satisfying

$$\frac{\kappa}{1-\kappa\mu} < \gamma,$$

the mapping $(g+F)^{-1}$ has a continuous local selection s around $g(\bar{x})+\bar{y}$ for $\bar{x}$, which moreover is calm at $g(\bar{x})+\bar{y}$ with

$$\text{(4)} \qquad \operatorname{clm}(s;g(\bar{x})+\bar{y}) \le \gamma.$$

Proof. The proof consists of two main steps. In the first step, we use induction to obtain a Cauchy sequence of continuous functions $z^0, z^1, \ldots$, such that z^n is a continuous and calm selection of the mapping $y \mapsto F^{-1}(y - g(z^{n-1}(y)))$. Then we show that this sequence has a limit in the space of continuous functions acting from a fixed ball around $\bar{y}$ to the space X and equipped with the supremum norm, and this limit is the selection whose existence is claimed.

Choose κ and μ as in the statement of the theorem and let $\gamma > \kappa/(1-\kappa\mu)$. Let λ, α and ν be such that $\kappa < \lambda < \alpha < 1/\nu$ and $\nu > \mu$, and also $\lambda/(1-\alpha\nu) \le \gamma$. Without loss of generality, we can assume that $g(\bar{x}) = 0$. Let $I\!B_a(\bar{x})$ and $I\!B_b(\bar{y})$ be the neighborhoods of $\bar{x}$ and $\bar{y}$, respectively, that are associated with the assumed properties of the mapping F and the function g. Specifically,

(a) For every $y, y' \in I\!B_b(\bar{y})$ and $x \in F^{-1}(y) \cap I\!B_a(\bar{x})$ there exists $x' \in F^{-1}(y')$ with

$$\|x' - x\| \le \lambda \|y' - y\|.$$

(b) For every $y \in \mathbb{B}_b(\bar{y})$ the set $F^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ is nonempty, closed and convex (that is, $\max\{a,b\} \le c$).

(c) The function g is Lipschitz continuous on $\mathbb{B}_a(\bar{x})$ with a constant ν.

According to 5J.7, there we can find a constant β, $0 < \beta \le b$, and a continuous function $z^0 : \mathbb{B}_\beta(\bar{y}) \to X$ such that

$$F(z^0(y)) \ni y \;\text{ and }\; \|z^0(y) - \bar{x}\| \le \lambda \|y - \bar{y}\| \;\text{ for all } y \in \mathbb{B}_\beta(\bar{y}).$$

Choose a positive τ such that

$$\tau \le (1 - \alpha\nu) \min\left\{a, \frac{a}{2\lambda}, \frac{\beta}{2}\right\} \tag{5}$$

and consider the mapping $y \mapsto M_1(y)$ where

$$M_1(y) = \left\{ x \in F^{-1}(y - g(z^0(y))) \,\middle|\, \|x - z^0(y)\| \le \alpha\nu \|z^0(y) - \bar{x}\| \right\}$$

for $y \in \mathbb{B}_\tau(\bar{y})$ and $M_1(y) = \emptyset$ for $y \notin \mathbb{B}_\tau(\bar{y})$. Clearly, $(\bar{y}, \bar{x}) \in \operatorname{gph} M_1$. Also, for any $y \in \mathbb{B}_\tau(\bar{y})$, we have from the choice of λ and α, using (5), that $z^0(y) \in \mathbb{B}_a(\bar{x})$ and therefore

$$\|y - g(z^0(y)) - \bar{y}\| \le \tau + \nu\|z^0(y) - \bar{x}\| \le \tau + \nu\lambda\tau \le (1 - \alpha\nu)(1 + \nu\lambda)(\beta/2) \le \beta \le b.$$

Then from the Aubin property of F^{-1} there exists $x \in F^{-1}(y - g(z^0(y)))$ with

$$\|x - z^0(y)\| \le \lambda \|g(z^0(y)) - g(\bar{x})\| \le \alpha\nu\|z^0(y) - \bar{x}\|,$$

which implies $x \in M_1(y)$. Thus M_1 is nonempty-valued. Further, if $(y, x) \in \operatorname{gph} M_1$, using (5) we have that $y \in \mathbb{B}_b(\bar{y})$ and also

$$\|x - \bar{x}\| \le \|x - z^0(y)\| + \|z^0(y) - \bar{x}\| \le (1 + \alpha\nu)\lambda\tau \le (1 - (\alpha\nu)^2)\lambda \frac{a}{2\lambda} \le \frac{a}{2}.$$

Then, from the property (b) above, since for any $y \in \operatorname{dom} M$ the set $M_1(y)$ is the intersection of a closed ball with a closed convex set, the mapping M_1 is closed-convex-valued on its domain. We will show that this mapping is inner semicontinuous on $\mathbb{B}_\tau(\bar{y})$.

Let $y \in \mathbb{B}_\tau(\bar{y})$ and $x \in M_1(y)$, and let $y^k \in \mathbb{B}_\tau(\bar{y})$, $y^k \to y$ as $k \to \infty$. If $z^0(y) = \bar{x}$, then $M_1(y) = \{\bar{x}\}$ and therefore $x = \bar{x}$. Any $x^k \in M_1(y^k)$ satisfies

$$\|x^k - z^0(y^k)\| \le \alpha\nu\|z^0(y^k) - \bar{x}\|.$$

Using the continuity of the function z^0, we see that $x^k \to z^0(y) = \bar{x} = x$; thus M_1 is inner semicontinuous.

Now let $z^0(y) \ne \bar{x}$. Since $z^0(y^k) \in F^{-1}(y^k - g(\bar{x})) \cap \mathbb{B}_a(\bar{x})$, the Aubin property of F^{-1} furnishes the existence of $\check{x}^k \in F^{-1}(y^k - g(z^0(y^k)))$ such that

$$\|\check{x}^k - z^0(y^k)\| \le \lambda \|g(z^0(y^k)) - g(\bar{x})\| \le \lambda\nu\|z^0(y^k) - \bar{x}\| \le \alpha\nu\|z^0(y^k) - \bar{x}\|. \tag{6}$$

Then $\check{x}^k \in M_1(y^k)$, and in particular, $\check{x}^k \in \mathbb{B}_a(\bar{x})$. Further, the inclusion $x \in F^{-1}(y - g(z^0(y))) \cap \mathbb{B}_a(\bar{x})$ combined with the Aubin property of F^{-1} entails the existence of $\tilde{x}^k \in F^{-1}(y^k - g(z^0(y^k)))$ such that

$$(7) \qquad \|\tilde{x}^k - x\| \leq \lambda(\|y^k - y\| + \nu\|z^0(y^k) - z^0(y)\|) \to 0 \quad \text{as} \quad k \to \infty.$$

Then $\tilde{x}^k \in \mathbb{B}_a(\bar{x})$ for large k. Let

$$\varepsilon^k := \frac{(1+\alpha\nu)\|z^0(y^k) - z^0(y)\| + \|\tilde{x}^k - x\|}{\alpha\nu\|z^0(y) - \bar{x}\| - \lambda\nu\|z^0(y^k) - \bar{x}\|}.$$

Note that, for $k \to \infty$, the numerator in the definition of ε^k goes to 0 because of the continuity of z^0 and (7), while the denominator converges to $(\alpha - \lambda)\nu\|z^0(y) - \bar{x}\| > 0$; therefore $\varepsilon^k \to 0$ as $k \to \infty$. Let

$$x^k = \varepsilon^k \check{x}^k + (1 - \varepsilon^k)\tilde{x}^k.$$

Since $\tilde{x}^k \to x$ and $\varepsilon^k \to 0$, we get $x^k \to x$ as $k \to \infty$ and also, since $y \mapsto F^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ is convex-valued around $(\bar{x}, \bar{y})$, we have $x^k \in F^{-1}(y^k - g(z^0(y^k)))$ for large k. By (6), (7), the assumption that $x \in M_1(y)$, and the choice of ε^k, we have

$$\begin{aligned}
\|x^k - z^0(y^k)\| &\leq \varepsilon^k\|\check{x}^k - z^0(y^k)\| + (1-\varepsilon^k)\|\tilde{x}^k - z^0(y^k)\| \\
&\leq \varepsilon^k\lambda\nu\|z^0(y^k) - \bar{x}\| + (1-\varepsilon^k)(\|\tilde{x}^k - x\| \\
&\quad + \|x - z^0(y)\| + \|z^0(y) - z^0(y^k)\|) \\
&\leq \varepsilon^k\lambda\nu\|z^0(y^k) - \bar{x}\| + \|\tilde{x}^k - x\| \\
&\quad + (1-\varepsilon^k)\alpha\nu\|z^0(y) - \bar{x}\| + \|z^0(y) - z^0(y^k)\| \\
&\leq \alpha\nu\|z^0(y^k) - \bar{x}\| + \alpha\nu\|z^0(y^k) - z^0(y)\| \\
&\quad + \|\tilde{x}^k - x\| + \|z^0(y) - z^0(y^k)\| \\
&\quad - \varepsilon^k\alpha\nu\|z^0(y) - \bar{x}\| + \varepsilon^k\lambda\nu\|z^0(y^k) - \bar{x}\| \\
&\leq \alpha\nu\|z^0(y^k) - \bar{x}\| + \|\tilde{x}^k - x\| + (1+\alpha\nu)\|z^0(y) - z^0(y^k)\| \\
&\quad - \varepsilon^k(\alpha\nu\|z^0(y) - \bar{x}\| - \lambda\nu\|z^0(y^k) - \bar{x}\|) \\
&= \alpha\nu\|z^0(y^k) - \bar{x}\|.
\end{aligned}$$

We obtain that $x^k \in M_1(y^k)$, and since $x^k \to x$, we conclude that the mapping M_1 is inner semicontinuous on its domain $\mathbb{B}_\tau(\bar{y})$. Hence, by Michael's selection theorem 5J.5, it has a continuous selection $z^1 : \mathbb{B}_\tau(\bar{y}) \to X$; that is, a continuous function z^1 which satisfies

$$z^1(y) \in F^{-1}(y - g(z^0(y))) \text{ and } \|z^1(y) - z^0(y)\| \leq \alpha\nu\|z^0(y) - \bar{x}\| \text{ for all } y \in \mathbb{B}_\tau(\bar{y}).$$

Then for $y \in \mathbb{B}_\tau(\bar{y})$, by the choice of γ,

$$\|z^1(y) - \bar{x}\| \leq \|z^1(y) - z^0(y)\| + \|z^0(y) - \bar{x}\| \leq (1+\alpha\nu)\lambda\|y - \bar{y}\| \leq \gamma\|y - \bar{y}\|.$$

The induction step is parallel to the first step. Let z^0 and z^1 be as above and suppose we have also found functions $z^1, z^2, \ldots, z^n$, such that each z^j, $j = 1, 2, \ldots, n$, is a continuous selection of the mapping $y \mapsto M_j(y)$, where

$$M_j(y) = \left\{ x \in F^{-1}(y - g(z^{j-1}(y))) \,\middle|\, \|x - z^{j-1}(y)\| \le \alpha\nu \|z^{j-1}(y) - z^{j-2}(y)\| \right\}$$

for $y \in \mathbb{B}_\tau(\bar{y})$ and $M_j(y) = \emptyset$ for $y \notin \mathbb{B}_\tau(\bar{y})$, where we put $z^{-1}(y) = \bar{x}$ for $y \in \mathbb{B}_\tau(\bar{y})$. Then for $y \in \mathbb{B}_\tau(\bar{y})$ we obtain

$$\|z^j(y) - z^{j-1}(y)\| \le (\alpha\nu)^{j-1} \|z^1(y) - z^0(y)\| \le (\alpha\nu)^j \|z^0(y) - \bar{x}\|, \quad j = 2, \ldots, n.$$

Therefore,

$$\begin{aligned} \|z^j(y) - \bar{x}\| &\le \sum_{i=0}^{j} \|z^i(y) - z^{i-1}(y)\| \\ &\le \sum_{i=0}^{j} (\alpha\nu)^i \|z^0(y) - \bar{x}\| \le \frac{\lambda}{1 - \alpha\nu} \|y - \bar{y}\| \le \gamma \|y - \bar{y}\|. \end{aligned}$$

Hence, from (5), for $j = 1, 2, \ldots, n$,

$$\|z^j(y) - \bar{x}\| \le a \tag{8}$$

and also

$$\|y - g(z^j(y)) - \bar{y}\| \le \tau + \nu \|z^j(y) - \bar{x}\| \le \tau + \frac{\lambda\nu\tau}{1 - \alpha\nu} \le \frac{\tau}{1 - \alpha\nu} \le \beta \le b. \tag{9}$$

Consider the mapping $y \mapsto M_{n+1}(y)$ where

$$M_{n+1}(y) = \left\{ x \in F^{-1}(y - g(z^n(y))) \,\middle|\, \|x - z^n(y)\| \le \alpha\nu \|z^n(y) - z^{n-1}(y)\| \right\}$$

for $y \in \mathbb{B}_\tau(\bar{y})$ and $M_{n+1}(y) = \emptyset$ for $y \notin \mathbb{B}_\tau(\bar{y})$. As in the first step, we find that M_{n+1} is nonempty-closed-convex-valued. Let $y \in \mathbb{B}_\tau(\bar{y})$ and $x \in M_{n+1}(y)$, and let $y^k \in \mathbb{B}_\tau(\bar{y})$, $y^k \to y$ as $k \to \infty$. If $z^{n-1}(y) = z^n(y)$, then $M_{n+1}(y) = \{z^n(y)\}$, and consequently $x = z^n(y)$; then from $z^n(y^k) \in F^{-1}(y^k - g(z^{n-1}(y^k))) \cap \mathbb{B}_a(\bar{x})$ and $y^k - g(z^{n-1}(y^k)) \in \mathbb{B}_b(\bar{y})$, we obtain, using the Aubin property of F^{-1}, that there exists $x^k \in F^{-1}(y^k - g(z^n(y^k)))$ such that

$$\|x^k - z^n(y^k)\| \le \lambda \|g(z^n(y^k)) - g(z^{n-1}(y^k))\| \le \alpha\nu \|z^n(y^k) - z^{n-1}(y^k)\|.$$

Therefore $x^k \in M_{n+1}(y^k)$, $x^k \to z^n(y) = x$ as $k \to \infty$, and hence M_{n+1} is inner semicontinuous for the case considered.

Let $z^n(y) \ne z^{n-1}(y)$. From (8) and (9) for $y = y^k$, since

$$z^n(y^k) \in F^{-1}(y^k - g(z^{n-1}(y^k))) \cap \mathbb{B}_a(\bar{x}),$$

the Aubin property of F^{-1} implies the existence of $\check{x}^k \in F^{-1}(y^k - g(z^n(y^k)))$ such that

$$\|\check{x}^k - z^n(y^k)\| \le \lambda \|g(z^n(y^k)) - g(z^{n-1}(y^k))\| \le \lambda \nu \|z^n(y^k) - z^{n-1}(y^k)\|.$$

Similarly, since $x \in F^{-1}(y - g(z^n(y))) \cap \mathbb{B}_a(\bar{x})$, there exists $\tilde{x}^k \in F^{-1}(y^k - g(z^n(y^k)))$ such that

$$\begin{aligned}\|\tilde{x}^k - x\| &\le \lambda (\|y^k - y\| + \|g(z^n(y^k)) - g(z^n(y))\|)\\ &\le \lambda (\|y^k - y\| + \nu \|z^n(y^k) - z^n(y)\|) \to 0 \text{ as } k \to \infty.\end{aligned}$$

Put

$$\varepsilon^k := \frac{\alpha\nu\|z^{n-1}(y) - z^{n-1}(y^k)\| + (1+\alpha\nu)\|z^n(y) - z^n(y^k)\| + \|\tilde{x}^k - x\|}{\alpha\nu\|z^n(y) - z^{n-1}(y)\| - \lambda\nu\|z^n(y^k) - z^{n-1}(y^k)\|}.$$

Then $\varepsilon^k \to 0$ as $k \to \infty$. Taking

$$x^k = \varepsilon^k \check{x}^k + (1-\varepsilon^k)\tilde{x}^k,$$

we obtain that $x^k \in F^{-1}(y^k - g(z^n(y^k)))$ for large k. Further, we can estimate $\|x^k - z^n(y^k)\|$ in the same way as in the first step, that is,

$$\begin{aligned}\|x^k - z^n(y^k)\| &\le \varepsilon^k \|\check{x}^k - z^n(y^k)\| + (1-\varepsilon^k)\|\tilde{x}^k - z^n(y^k)\|\\
&\le \varepsilon^k \lambda\nu \|z^n(y^k) - z^{n-1}(y^k)\|\\
&\quad + (1-\varepsilon^k)(\|\tilde{x}^k - x\| + \|x - z^n(y)\| + \|z^n(y) - z^n(y^k)\|)\\
&\le \varepsilon^k \lambda\nu \|z^n(y^k) - z^{n-1}(y^k)\| + \|\tilde{x}^k - x\|\\
&\quad + (1-\varepsilon^k)\alpha\nu\|z^n(y) - z^{n-1}(y)\| + \|z^n(y) - z^n(y^k)\|\\
&\le \alpha\nu\|z^n(y^k) - z^{n-1}(y^k)\| + \alpha\nu\|z^n(y^k) - z^n(y)\|\\
&\quad + \alpha\nu\|z^{n-1}(y^k) - z^{n-1}(y)\| + \|\tilde{x}^k - x\|\\
&\quad + \|z^n(y) - z^n(y^k)\| - \varepsilon^k\alpha\nu\|z^n(y) - z^{n-1}(y)\|\\
&\quad + \varepsilon^k\lambda\nu\|z^n(y^k) - z^{n-1}(y^k)\|\\
&\le \alpha\nu\|z^n(y^k) - z^{n-1}(y^k)\| + \|\tilde{x}^k - x\|\\
&\quad + (1+\alpha\nu)\|z^n(y) - z^n(y^k)\| + \alpha\nu\|z^{n-1}(y) - z^{n-1}(y^k)\|\\
&\quad - \varepsilon^k(\alpha\nu\|z^n(y) - z^{n-1}(y)\| - \lambda\nu\|z^n(y^k) - z^{n-1}(y^k)\|)\\
&= \alpha\nu\|z^n(y^k) - z^{n-1}(y^k)\|.\end{aligned}$$

We conclude that $x^k \in M_{n+1}(y^k)$, and since $x^k \to x$ as $k \to \infty$, the mapping M_{n+1} is inner semicontinuous on $\mathbb{B}_\tau(\bar{y})$. Hence, the mapping M_{n+1} has a continuous selection $z^{n+1} : \mathbb{B}_\tau(\bar{y}) \to X$, that is,

$$z^{n+1}(y) \in F^{-1}(y - g(z^n(y))) \quad \text{and} \quad \|z^{n+1}(y) - z^n(y)\| \le \alpha\nu\|z^n(y) - z^{n-1}(y)\|.$$

Thus

$$\|z^{n+1}(y) - z^n(y)\| \leq (\alpha\nu)^{(n+1)}\|z^0(y) - \bar{x}\|.$$

The induction step is now complete. In consequence, we have an infinite sequence of bounded continuous functions $z^0, \dots, z^n, \dots$ such that for all $y \in \mathbb{B}_\tau(\bar{y})$ and for all n,

$$\|z^n(y) - \bar{x}\| \leq \sum_{i=0}^{n}(\alpha\nu)^i\|z^0(y) - \bar{x}\| \leq \frac{\lambda}{1-\alpha\nu}\|y - \bar{y}\| \leq \gamma\|y - \bar{y}\|$$

and moreover,

$$\sup_{y \in \mathbb{B}_\tau(\bar{y})}\|z^{n+1}(y) - z^n(y)\| \leq (\alpha\nu)^n \sup_{y \in \mathbb{B}_\tau(\bar{y})}\|z^0(y) - \bar{x}\| \leq (\alpha\nu)^n\lambda\tau \quad \text{for } n \geq 1.$$

The sequence $\{z^n\}$ is a Cauchy sequence in the space of functions that are continuous and bounded on $\mathbb{B}_\tau(\bar{y})$ equipped with the supremum norm. Then this sequence has a limit s which is a continuous function on $\mathbb{B}_\tau(\bar{y})$ and satisfies

$$s(y) \in F^{-1}(y - g(s(y)))$$

and

$$\|s(y) - \bar{x}\| \leq \frac{\lambda}{1-\alpha\nu}\|y - \bar{y}\| \leq \gamma\|y - \bar{y}\| \quad \text{for all } y \in \mathbb{B}_\tau(\bar{y}).$$

Thus, s is a continuous local selection of $(g+F)^{-1}$ which has the calmness property (4). This brings the proof to its end. □

Proof of Theorem 5J.3. Apply 5J.8 with $F = Df(\bar{x})$ and $g(x) = f(x) - Df(\bar{x})x$. Metric regularity of F is equivalent to surjectivity of $Df(\bar{x})$, and F^{-1} is convex-closed-valued. The mapping g has $\operatorname{lip}(g;\bar{x}) = 0$ and finally $F + g = f$. □

Note that Theorem 5J.7 follows from 5J.8 with g being the zero function.

We present next an implicit mapping version of Theorem 5J.7.

Theorem 5J.9 (implicit mapping version)**.** *Let X, Y and P be Banach spaces. For $f : P \times X \to Y$ and $F : X \rightrightarrows Y$, consider the generalized equation $f(p,x) + F(x) \ni 0$ with solution mapping*

$$S(p) = \{\, x \mid f(p,x) + F(x) \ni 0 \,\} \quad \textit{having } \bar{x} \in S(\bar{p}).$$

Suppose that F satisfies the conditions in Theorem 5J.7 with $\bar{y} = 0$ and associate constant $\kappa \geq \operatorname{reg}(F;\bar{x}|0)$ and also that f is continuous on a neighborhood of $(\bar{p},\bar{x})$ and has $\widehat{\operatorname{lip}}_x(f;(\bar{p},\bar{x})) \leq \mu$, where μ is a nonnegative constant satisfying $\kappa\mu < 1$. Then for every γ satisfying

(10)
$$\frac{\kappa}{1-\kappa\mu} < \gamma,$$

there exist neighborhoods U of $\bar{x}$ and Q of $\bar{p}$ along with a continuous function $s : Q \to U$ such that

$$(11)\qquad s(p) \in S(p) \ \text{ and } \ \|s(p) - \bar{x}\| \le \gamma \|f(p,\bar{x}) - f(\bar{p},\bar{x})\| \ \text{ for every } \ p \in Q.$$

Proof. The proof is parallel to the proof of Theorem 5J.7. First we choose γ satisfying (10) and then λ, α and ν such that $\kappa < \lambda < \alpha < \nu^{-1}$ and $\nu > \mu$, and also

$$(12)\qquad \frac{\lambda}{1 - \alpha\nu} < \gamma.$$

There are neighborhoods U, V and Q of $\bar{x}$, 0 and $\bar{p}$, respectively, which are associated with the metric regularity of F at $\bar{x}$ for 0 with constant λ and the Lipschitz continuity of f with respect to x with constant ν uniformly in p. By appropriately choosing a sufficiently small radius τ of a ball around $\bar{p}$, we construct an infinite sequence of continuous and bounded functions $z^k : \mathbb{B}_\tau(\bar{p}) \to X$, $k = 0, 1, \ldots$, which are uniformly convergent on $\mathbb{B}_\tau(\bar{p})$ to a function s satisfying the conditions in (11). The initial z^0 satisfies

$$z^0(p) \in F^{-1}(-f(p,\bar{x})) \ \text{ and } \ \|z^0(p) - \bar{x}\| \le \lambda \|f(p,\bar{x}) - f(\bar{p},\bar{x})\|.$$

For $k = 1, 2, \ldots$, the function z^k is a continuous selection of the mapping

$$M_k : p \mapsto \left\{ x \in F^{-1}(-f(p, z^{k-1}(p))) \,\middle|\, \|x - z^{k-1}(p)\| \le \alpha\nu \|z^{k-1}(p) - z^{k-2}(p)\| \right\}$$

for $p \in \mathbb{B}_\tau(\bar{p})$, where $z^{-1}(p) = \bar{x}$. Then for all $p \in \mathbb{B}_\tau(\bar{p})$ we obtain

$$z^k(p) \in F^{-1}(-f(p, z^{k-1}(p))) \text{ and } \|z^k(p) - z^{k-1}(p)\| \le (\alpha\nu)^k \|z^0(p) - \bar{x}\|,$$

hence,

$$(13)\qquad \|z^k(y) - \bar{x}\| \le \frac{\lambda}{1 - \alpha\nu} \|f(p,\bar{x}) - f(\bar{p},\bar{x})\|.$$

The sequence $\{z^k\}$ is a Cauchy sequence of continuous and bounded function, hence it is convergent with respect to the supremum norm. In the limit with $k \to \infty$, taking into account (12) and (13), we obtain a selection s with the desired properties. □

Exercise 5J.10 (specialization for closed sublinear mappings)**.** *Let $F : X \rightrightarrows Y$ have convex and closed graph, let $f : X \to Y$ be strictly differentiable at $\bar{x}$ and let $(\bar{x}, \bar{y}) \in \operatorname{gph}(f + F)$. Suppose that*

$$\bar{y} \in \operatorname{int} \operatorname{rge}(f(\bar{x}) + Df(\bar{x})(x - \bar{x}) + F).$$

Apply the Robinson–Ursescu theorem 5B.4 to prove that there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$, a continuous function $s : V \to U$, and a constant γ, such that

$$(f + F)(s(y)) \ni y \qquad \text{and} \qquad \|s(y) - \bar{x}\| \le \gamma \|y - \bar{y}\| \ \text{ for every } \ y \in V.$$

5K. Selections from Directional Differentiability

Here, at the end of this chapter, we present a result which is in the spirit of the inverse function theorems for nonstrictly differentiable functions in finite dimensions established in 1G. We deal with functions, acting in Banach spaces, that are directionally differentiable. We show that when the inverse of the directional derivative of a such a function has a uniformly bounded selection, then the inverse of the function itself has a selection which is calm.

We start by restating the definition of directional differentiability given in Section 2D for functions acting from a Banach space X to a Banach space Y.

Directional derivative. *For a function $f : X \to Y$, a point $\bar{x} \in \operatorname{dom} f$ and a direction $w \in X$ the limit*

$$f'(\bar{x};w) = \lim_{t \searrow 0} \frac{f(\bar{x}+tw) - f(\bar{x})}{t},$$

when it exists, is said to be the (one-sided) directional derivative of f at $\bar{x}$ for w. For an open subset U of X, if $f'(x;w)$ exists for all $x \in U$ and $w \in X$, then we say that f is directionally differentiable on U.

The main result of this section is the following inverse function theorem.

Theorem 5K.1 (inverse selection from directional differentiability). *Let $f : X \to Y$ be continuous and directionally differentiable on a neighborhood U of $\bar{x} \in \operatorname{int} \operatorname{dom} f$, and assume that there exists $\kappa > 0$ such that for any $x \in U$ there exists a selection $\sigma(x;\cdot)$ for $f'(x;\cdot)^{-1}$ such that*

$$\|\sigma(x;y)\| \le \kappa\|y\| \text{ for all } y \in Y. \tag{1}$$

Then the inverse f^{-1} has a local selection s around $f(\bar{x})$ for $\bar{x}$ which is calm at $f(\bar{x})$ with $\operatorname{clm}(s; f(\bar{x})) \le \kappa$.

That $\sigma(x;\cdot)$ is a selection for $f'(x;\cdot)^{-1}$ simply means that $\sigma(x;\cdot)$ is a function from Y to X such that $f'(x,\sigma(x;y)) = y$ for all $y \in Y$. In particular, $\sigma(x;0) = 0$ for any $x \in U$.

In preparation to proving the theorem, first recall[5] that for any $y \neq 0$ there exists $y^* \in Y^*$ satisfying

$$\|y^*\| = 1 \quad \text{and} \quad \langle y^*, y\rangle = \|y\|$$

such that, for any $w \in Y$,

$$\lim_{t \searrow 0} \frac{1}{t}(\|y+tw\| - \|y\|) = \langle y^*, w\rangle. \tag{2}$$

Thus, $\langle y^*, w\rangle$ is just the derivative of the norm at y in direction w. If Y is a Hilbert space, then $y^* = \|y\|^{-1}y$. We will use (2) in the following lemma which gives the di-

[5] See Rockafellar [1974], Theorem 11. Here y^* is a subgradient of the norm mapping.

rectional derivative of the composition of the norm and a directionally differentiable function.

Lemma 5K.2. *Consider a function $f : X \to Y$ and a point $x \in \operatorname{int} \operatorname{dom} f$ for which there exists a neighborhood U of x such that $f(u) \neq 0$ for all $u \in U$ and f is directionally differentiable on U. Let $\xi \in X$ and t_0 be such that $x + t\xi \in U$ for all $t \in [0, t_0)$. Then for any $t \in [0, t_0)$ there exists y^* such that*

$$(3) \qquad \|y^*\| = 1 \qquad \langle y^*, f(x+t\xi)\rangle = \|f(x+t\xi)\|$$

and moreover

$$(4) \qquad \lim_{h \searrow 0} \frac{1}{h}\Big(\|f(x+(t+h)\xi)\| - \|f(x+t\xi)\|\Big) = \langle y^*, f'(x+t\xi;\xi)\rangle.$$

Proof. Having $t \in [0, t_0)$ and $\xi \in X$ fixed, let

$$z(h) = \frac{1}{h}(f(x+(t+h)\xi) - f(x+t\xi)).$$

Since f is directionally differentiable around x, we have

$$\lim_{h \searrow 0} z(h) = f'(x+t\xi;\xi).$$

Further, combining

$$\frac{1}{h}\Big(\|f(x+(t+h)\xi)\| - \|f(x+t\xi)\|\Big) = \frac{1}{h}\Big(\|f(x+t\xi)+hz(h)\| - \|f(x+t\xi)\|\Big)$$

with

$$\frac{1}{h}\|f(x+t\xi)+hz(h)\| - \frac{1}{h}\|f(x+t\xi)+hf'(x+t\xi;\xi)\| \leq \|z(h) - f'(x+t\xi;\xi)\|$$

and passing to the limit as $h \searrow 0$, we obtain

$$\begin{aligned} &\lim_{h \searrow 0} \frac{1}{h}\Big(\|f(x+(t+h)\xi)\| - \|f(x+t\xi)\|\Big) \\ &\qquad = \lim_{h \searrow 0} \frac{1}{h}\Big(\|f(x+t\xi)+hf'(x+t\xi;\xi)\| - \|f(x+t\xi)\|\Big). \end{aligned}$$

Applying the relation (2) stated before the lemma to the last equality, we conclude that there exists y^* with the desired properties (3) and (4). □

Proof of Theorem 5K.1. Without loss of generality, let $\bar{x} = 0$ and $f(\bar{x}) = 0$. The theorem will be proved if we show that for any $\mu > \kappa$ there exist positive a and b such that for any $y \in \mathbb{B}_b(0)$ there exists $x \in \mathbb{B}_a(0)$ satisfying

$$(5) \qquad f(x) = y \quad \text{and} \quad \|x\| \leq \mu\|y\|.$$

Let $\mu > \kappa$ and choose positive a and b such that the estimate (1) holds for all $x \in \mathbb{B}_a(0)$ and $y \in \mathbb{B}_b(0)$, and moreover, $\mu b < a$. Fix $y \in \mathbb{B}_b(0)$ and set $\varphi(x) = \|f(x) - y\|$. Note that $\varphi(0) = \|y\|$. We apply Ekeland's variational principle 4B.5, obtaining that for every $\delta > 0$ there exists a point x_δ such that

$$\varphi(x_\delta) + \delta\|x_\delta\| \le \varphi(0)$$

and

$$\varphi(x_\delta) \le \varphi(x) + \delta\|x - x_\delta\| \quad \text{for every } x \in X.$$

Take $\delta = 1/\mu$. Then we get $\varphi(x_\delta) \le \varphi(0)$, $\|x_\delta\| \le \mu\|y\|$, and

$$(6) \qquad \varphi(x) \ge \varphi(x_\delta) - \frac{1}{\mu}\|x - x_\delta\| \quad \text{for every } x \in X.$$

Note that $\|x_\delta\| \le \mu\|y\| \le \mu b < a$. The proof will be complete if we show that $f(x_\delta) = y$.

On the contrary, assume that $f(x_\delta) \ne y$, that is, $\varphi(0) \ne 0$. Let $u \in X$. Let $t_0 > 0$ be such that $f(x_\delta + tu) \ne y$ for all $t \in [0, t_0)$ and let $t \in [0, t_0)$. Taking $x = x_\delta + tu$ in (6) we obtain

$$(7) \qquad \frac{1}{t}(\|f(x_\delta + tu) - y\| - \|f(x_\delta) - y\| \ge -\frac{1}{\mu}\|u\|.$$

From 5K.2 there exists y^* satisfying (3) and (4) with $x = x_\delta$ and $\xi = u$; that is,

$$(8) \qquad \|y^*\| = 1 \quad \text{and} \quad \langle y^*, f(x_\delta) - y\rangle = \|f(x_\delta) - y\|$$

and

$$\lim_{t \searrow 0} \frac{1}{t}(\|f(x_\delta + tu) - y\| - \|f(x_\delta) - y\|) = \langle y^*, f'(x_\delta; u)\rangle.$$

Passing to the limit in (7), we obtain

$$\langle y^*, f'(x_\delta; u)\rangle \ge -\frac{1}{\mu}\|u\|.$$

Then, taking $u = \sigma(x_\delta; -f(x_\delta) + y)$ we come to

$$\langle y^*, f'(x_\delta; \sigma(x_\delta; -f(x_\delta) + y))\rangle = \langle y^*, -(f(x_\delta) - y)\rangle \ge -\frac{1}{\mu}\|u\|.$$

From the last inequality, taking into account (1) and (8), we get

$$\|f(x_\delta) - y\| \le \frac{1}{\mu}\|\sigma(x_\delta; -f(x_\delta) + y)\| \le \frac{\kappa}{\mu}\|f(x_\delta) - y\|.$$

This is a contradiction since $\kappa < \mu$. □

Exercise 5K.3. *State and prove an implicit function analogue of Theorem 5K.1.*

Commentary

The equivalence of (a) and (d) in 5A.1 for mappings whose ranges are closed was shown in Theorem 10 on p. 150 of the original treatise of S. Banach [1932]. The statements of this theorem usually include the equivalence of (a) and (b), which is called in Dunford and Schwartz [1958] the "interior mapping principle." Lemma 5A.4 is usually stated for Banach algebras, see, e.g., Theorem 10.7 in Rudin [1991]. Theorem 5A.8 is from Robinson [1972].

The generalization of the Banach open mapping theorem to set-valued mappings with convex closed graphs was obtained independently by Robinson [1976] and Ursescu [1975]; the proof of 5B.3 given here is close to the original proof in Robinson [1976]. A particular case of this result for positively homogeneous mappings was shown earlier by Ng [1973]. The Baire category theorem can be found in Dunford and Schwartz [1958], p. 20. The Robinson–Ursescu theorem is stated in various ways in the literature, see, e.g., Theorem 3.3.1 in Aubin and Ekeland [1984], Theorem 2.2.2 in Aubin and Frankowska [1990], Theorem 2.83 in Bonnans and Shapiro [2000], Theorem 9.48 in Rockafellar and Wets [1998], and Theorem 1.3.11 in Zălinescu [2002].

Sublinear mappings (under the name "convex processes") and their adjoints were introduced by Rockafellar [1967]; see also Rockafellar [1970]. Theorem 5C.9 first appeared in Lewis [1999], see also Lewis [2001]. The norm duality theorem, 5C.10, was originally proved by Borwein [1983], who later gave in Borwein [1986b] a more detailed argument. The statement of the Hahn–Banach theorem 5C.11 is from Dunford and Schwartz [1958], p. 62.

Theorems 5D.1 and 5D.2 are versions of results originally published in Lyusternik [1934] and Graves [1950], with some adjustments to the current setting. Lyusternik apparently viewed his theorem mainly as a stepping stone to obtain the Lagrange multiplier rule for abstract minimization problems, and the title of his paper from 1934 clearly says so. It is also interesting to note that, after the statement of the Lyusternik theorem as 8.10.2 in the functional analysis book by Lyusternik and Sobolev [1965], the authors say that "the proof of this theorem is a modification of the proof of the implicit function theorem, and the [Lyusternik] theorem is a direct generalization of this [implicit function] theorem."

In the context of his work on implicit functions, see Commentary to Chapter 1, it is quite likely that Graves considered his theorem 5D.2 as an extension of the Banach open mapping theorem for nonlinear mappings[6]. But there is more in its statement and proof; namely, the Graves theorem does not involve differentiation and then, as shown in 5D.3, can be easily extended to become a generalization of

[6] A little known fact is that Graves was the supervisor of the master thesis of W. Karush ["Minima of functions of several variables with inequalities as side conditions", Departament of of Mathematics, University of Chicago, 1939], where the necessary optimality conditions for nonlinear programming problems now known as the Karush-Kuhn-Tucker conditions (see 2A) were first derived; for a review of Karush's work see Cottle [2012]. It is quite possible that already in the 1930s Graves new the connection between metric regularity with the Lagrange multiplier rule.

the basic Lemma 5A.4 for nonlinear mappings. This was mentioned already in the historical remarks of Dunford and Schwartz [1958], p. 85. A further generalization in line with the present setting was revealed in Dmitruk, Milyutin and Osmolovskiĭ [1980], where the approximating linearization is replaced by a function such that the difference between the original mapping and this function is Lipschitz continuous with a sufficiently small Lipschitz constant; this is now Theorem 5E.7. Estimates for the regularity modulus of the kind given in 5D.3 are also present in Ioffe [1979], see also Ioffe [2000].

In the second part of the last century, when the development of optimality conditions was a key issue, the approach of Lyusternik was recognized for its virtues and extended to great generality. Historical remarks regarding these developments can be found in Rockafellar and Wets [1998]. The statement in 5D.4 is a modified version of the Lyusternik theorem as given in Section 0.2.4 of Ioffe and Tikhomirov [1974].

Theorem 5E.1 is given as in Dontchev, Lewis and Rockafellar [2003]; earlier results in this vein were obtained by Dontchev and Hager [1993, 1994]. The contraction mapping theorem 5E.2 is from Dontchev and Hager [1994]. Theorem 5G.1 is from Aragón Artacho et al. [2011]. Most of the results in 5H are from Dontchev and Frankowska [2011], [2012]. Theorem 5H.8 has its origins in the works of Frankowska [1990, 1992] and Ursescu [1996]. Theorem 5I.3 can be traced back to Arutyunov [2007]. Corollary 5I.5 is due to Lim [1985].

Since the publication of the first edition of this book, a large number of works have appeared dealing with various aspects of metric regularity. Among them are: Azé [2006], Schirotzek [2007], Azé and Corvellec [2009], Dmitruk and Kruger [2009], Zheng and Ng [2009], Pang [2011], Uderzo [2009, 2012a, 2012b], Ioffe [2008, 2010a, 2010b, 2011, 2013], Durea and Strugariu [2012a, 2012b], Cibulka [2011], Cibulka and Fabian [2013], Ngai, Kruger, and Théra [2012], Ngai, Nguyen and Théra [2013], Klatte, Kruger and Kummer [2012], Gfrerer [2013], Bianchi, Kassay and Pini [2013], and Apetrii, Durea and Strugariu [2013].

In this chapter we present developments centered around metric regularity in abstract spaces, but there are many other techniques and results that fall into the same category. In particular, we do not discuss the concept of *slope* introduced by De Georgi, Marino and Tosques [1980] which, as shown by Aze, Corvellec and Lucchetti [1999] and later by Ioffe [2001] can be quite instrumental in deriving criteria for metric regularity.

Theorem 5J.3 gives the original form of the Bartle–Graves theorem as contributed in Bartle and Graves [1952]. The particular form 5J.5 of Michael's selection theorem[7] is Lemma 2.1 in Deimling [1992]. Lemma 5J.6 was first given in Borwein and Dontchev [2003], while Theorem 5J.8 is from Dontchev [2004]. These two papers were largely inspired by contacts of the first author of this book with Robert G. Bartle, who was able to read them before he passed away Sept. 18, 2002. Shortly

[7] The original statement of Michael's selection theorem is for mappings acting from a paracompact space to a Banach space; by a theorem of A. H. Stone every metric space is paracompact and hence every subset of a Banach space is paracompact.

before he died he sent to the first author a letter, where he, among other things, wrote the following:

"Your results are, indeed, an impressive and far-reaching extension of the theorem that Professor Graves and I published over a half-century ago. I was a student in a class of Graves in which he presented the theorem in the case that the parameter domain is the interval $[0,1]$. He expressed the hope that it could be generalized to a more general domain, but said that he didn't see how to do so. By a stroke of luck, I had attended a seminar a few months before given by André Weil, which he titled "On a theorem by Stone." I (mis)understood that he was referring to M. H. Stone, rather than A. H. Stone, and attended. Fortunately, I listened carefully enough to learn about paracompactness and continuous partition of unity[8] (which were totally new to me) and which I found to be useful in extending Graves' proof. So the original theorem was entirely due to Graves; I only provided an extension of his proof, using methods that were not known to him. However, despite the fact that I am merely a 'middleman,' I am pleased that this result has been found to be useful."

The material in Section 5K is basically from Ekeland [2011], where the role of the local selection s_x is played by the right inverse of the Gâteaux derivative derivative of f.

In this book we present inverse/implicit function theorems related to variational analysis, but there are many other theorems that fall into the same category and are designed as tools in other areas of mathematics. A prominent example of such a result not discussed in this book is the celebrated Nash–Moser theorem used mainly in geometric analysis and partial differential equations. A rigorous introduction of this theorem and the theory around it would require a lot of space and would tip the balance of topics and ideas away from what we want to emphasize. More importantly, we have not been able to identify (as of yet) a specific, sound application of this theorem in variational analysis such as would have justified the inclusion. A rigorous and nice presentation of the Nash–Moser theorem along with the theory and applications behind it is given in Hamilton [1982]. A Nash-Moser type version of Theorem 5K.1 can be found in Ekeland [2011]. In the following lines, we only briefly point out a connection to the results in Section 5E.

The Nash–Moser theorem is about mappings acting in *Fréchet* spaces, which are more general than the Banach spaces. Consider a linear (vector) space F equipped with the collection of seminorms $\{\|\cdot\|_n | n \in \mathbb{N}\}$ (a seminorm differs from a norm in that the seminorm of a nonzero element could be zero). The topology induced by this (countable) collection of seminorms makes the space F a locally convex topological vector space. If $x = 0$ when $\|x\|_n = 0$ for all n, the space is *Hausdorff*. In a Hausdorff space, one may define a metric based the family of seminorms in the following way:

$$(1) \qquad \rho(x,y) = \sum_{n=1}^{\infty} 2^{-n} \frac{\|x-y\|_n}{1+\|x-y\|_n}.$$

[8] Michael's theorem was not known at that time.

It is not difficult to see that this metric is shift-invariant. A sequence $\{x_k\}$ is said to be Cauchy when $\|x_k - x_j\|_n \to 0$ as k and $j \to \infty$ for all n, or, equivalently, $\rho(x_k, x_j) \to 0$ as $k \to \infty$ and $j \to \infty$. As usual, a space is complete if every Cauchy sequence converges. A Fréchet space is a complete Hausdorff metrizable locally convex topological vector space.

Having two Fréchet spaces F and G, we can now introduce a metric ρ associated with their collections of seminorms as in (1) above, and define Lipschitz continuity and metric regularity accordingly. Then 5E.1 will apply of course and from it we can obtain a Graves-type theorem in Fréchet spaces. To get to the Nash–Moser theorem, however, we have a long way to go, translating the meaning of the assumptions in terms of the metric ρ for the collections of seminorms and the mappings considered. For that we will need more structure in the spaces, an ordering (grading) of the sequence of seminorms and, moreover, a certain uniform approximation property called the *tameness* condition. For the mappings, the associated tameness property means that certain growth estimates hold. The basic statement of the Nash–Moser theorem is surprisingly similar to the classical inverse function theorem, but the meaning of the concepts used is much more involved: when a smooth tame mapping f acting between Fréchet spaces has an invertible tame derivative, then f^{-1} has a smooth tame single-valued localization. The rigorous introduction of the tame spaces, mappings and derivatives is beyond the scope of this book; we only note here that extending the Nash–Moser theorem to set-valued mappings, e.g. in the setting of Section 5E, is a challenging avenue for future research.

Chapter 6
Applications in Numerical Variational Analysis

The classical implicit function theorem finds a wide range of applications in numerical analysis. For instance, it helps in deriving error estimates for approximations to differential equations and is often relied on in establishing the convergence of algorithms for solving nonlinear equations. Can the generalizations of the classical theory to which we have devoted so much of this book have comparable applications in the numerical treatment of nonclassical problems for generalized equations and beyond? In this chapter we provide positive answers in several directions.

We begin with a topic at the core of numerical work, the "conditioning" of a problem and how it extends to concepts like metric regularity. We also explain how the conditioning of a feasibility problem, like solving a system of inequalities, can be understood. Next we take up the age-old procedure known as Newton's method in several guises. We go a step further with Newton's method by showing that the mapping which assigns to an instance of a parameter the set of all sequences generated by the method obeys, in a Banach space of sequences, the Lyusternik–Graves/implicit function theorem paradigm in the same pattern as the solution mapping for the underlying generalized equation. Then we give generalized forms of the Dennis–Moré theorem and also consider inexact and nonsmooth versions of Newton's method. We deal next with a parameterized generalized equation and derive an error estimate for an extended version of the Euler-Newton path-following method. Approximations of quadratic optimization problems in Hilbert spaces are then studied. In the last two sections we apply our methodology to discrete approximations in optimal control.

6A. Radius Theorems and Conditioning

In numerical analysis, a measure of "conditioning" of a problem is typically conceived as a bound on the ratio of the size of solution (output) error to the size of data (input) error. At its simplest, this pattern is seen when solving a linear equation $Ax = y$ for x in terms of y when A is a nonsingular matrix in $\mathbb{R}^{n\times n}$. The data input then is y and the solution output is $A^{-1}y$, but for computational purposes the story cannot just be left at that. Much depends on the extent to which an input error δy leads to an output error δx. The magnitudes of the errors can be measured by the Euclidean norm, say. Then, through linearity, there is the tight bound $|\delta x| \le |A^{-1}||\delta y|$, in which $|A^{-1}|$ is the corresponding matrix (operator) norm of the mapping $y \mapsto A^{-1}y$ and in fact is the global Lipschitz constant for this mapping. In providing such a bound on the ratio of $|\delta x|$ to $|\delta y|$, $|A^{-1}|$ is called the *absolute condition number* for the problem of solving $Ax = y$. A high value of $|A^{-1}|$ is a warning flag signaling trouble in computing the solution x for a given y.

Another popular conditioning concept concerns relative errors instead of absolute errors. In solving $Ax = y$, the relative error of the input is $|\delta y|/|y|$ (with $y \neq 0$), while the relative error of the output is $|\delta x|/|x|$. It is easy to see that the best bound on the ratio of $|\delta x|/|x|$ to $|\delta y|/|y|$ is the product $|A||A^{-1}|$. Therefore, $|A||A^{-1}|$ is called the *relative condition number* for the problem of solving $Ax = y$. But absolute conditioning will be the chief interest in our present context, for several reasons.

The reciprocal of the absolute condition number $|A^{-1}|$ of a nonsingular matrix A has a geometric interpretation which will serve as an important guide to our developments. It turns out to give an exact bound on how far A can be perturbed to $A + B$ before good behavior breaks down by $A + B$ becoming singular, and thus has more significance for numerical analysis than simply comparing the size of δx to the size of δy. This property of the absolute condition number comes from a classical result about matrices which was stated and proved in Chapter 1 as 1E.9:

$$\inf\left\{ |B| \,\middle|\, A + B \text{ is singular} \right\} = \frac{1}{|A^{-1}|}, \quad \text{for any nonsingular matrix } A.$$

In this sense, $|A^{-1}|^{-1}$ gives the *radius of nonsingularity* around A. As long as B lies within that distance from A, the nonsingularity of $A + B$ is assured. Clearly from this angle as well, a large value of the condition number $|A^{-1}|$ points toward numerical difficulties.

The model provided to us by this example is that of a *radius theorem*, furnishing a bound on how far perturbations of some sort in the specification of a problem can go before some key property is lost. Radius theorems can be investigated not only for solving equations, linear and nonlinear, but also generalized equations, systems of constraints, etc.

We start down that track by stating the version of the cited matrix result that works in infinite dimensions for bounded linear mappings acting in Banach spaces.

Theorem 6A.1 (radius theorem for invertibility of bounded linear mappings)**.** *Let X and Y be Banach spaces and let $A \in \mathscr{L}(X,Y)$ be invertible*[1]*. Then*

$$\inf_{B\in\mathscr{L}(X,Y)} \left\{ \|B\| \,\middle|\, A+B \text{ is not invertible} \right\} = \frac{1}{\|A^{-1}\|}. \tag{1}$$

Moreover the infimum is the same if restricted to mappings B of rank one.

Proof. The estimation for perturbed inversion in Lemma 5A.4 gives us "$\geq$" in (1). To obtain the opposite inequality and thereby complete the proof, we take any $r > 1/\|A^{-1}\|$ and construct a mapping B of rank one such that $A+B$ is not invertible and $\|B\| < r$. There exists $\hat{x}$ with $\|A\hat{x}\| = 1$ and $\|\hat{x}\| > 1/r$. Choose an $x^* \in X^*$ such that $x^*(\hat{x}) = \|\hat{x}\|$ and $\|x^*\| = 1$. The linear and bounded mapping

$$Bx = -\frac{x^*(x)A\hat{x}}{\|\hat{x}\|} \tag{2}$$

has $\|B\| = 1/\|\hat{x}\|$ and $(A+B)\hat{x} = A\hat{x} - A\hat{x} = 0$. Then $A+B$ is not invertible and hence the infimum in (1) is $\leq r$. It remains to note that B in (2) is of rank one. $\square$

The initial step that can be taken toward generality beyond linear mappings is in the direction of positively homogeneous mappings $H : X \rightrightarrows Y$; here and further on, X and Y are Banach spaces. For such a mapping, ordinary norms can no longer be of help in conditioning, but the outer and inner norms introduced in 4A in finite dimensions and extended in 5A to Banach spaces can come into play:

$$\|H\|^+ = \sup_{\|x\|\leq 1} \sup_{y\in H(x)} \|y\| \quad \text{and} \quad \|H\|^- = \sup_{\|x\|\leq 1} \inf_{y\in H(x)} \|y\|.$$

Their counterparts for the inverse H^{-1} will have a role as well:

$$\|H^{-1}\|^+ = \sup_{\|y\|\leq 1} \sup_{x\in H^{-1}(y)} \|x\| \quad \text{and} \quad \|H^{-1}\|^- = \sup_{\|y\|\leq 1} \inf_{x\in H^{-1}(y)} \|x\|.$$

In thinking of $H^{-1}(y)$ as the set of solutions x to $H(x) \ni y$, it is clear that the outer and inner norms of H^{-1} capture two different aspects of solution behavior, roughly the distance to the farthest solution and the distance to the nearest solution (when multivaluedness is present). We are able to assert, for instance, that

$$\operatorname{dist}(0, H^{-1}(y)) \leq \|H^{-1}\|^- \|y\| \text{ for all } y.$$

From that angle, $\|H^{-1}\|^-$ could be viewed as a sort of *inner* absolute condition number—and in a similar manner, $\|H^{-1}\|^+$ could be viewed as a sort of *outer* absolute condition number. This idea falls a bit short, though, because we only have a comparison between sizes of $\|x\|$ and $\|y\|$, not the size of a shift from x to $x+\delta x$

[1] This assumption can be dropped if we identify invertibility of A with $\|A^{-1}\| < \infty$ and adopt the convention $1/\infty = 0$. Similar adjustments can be made in the remaining radius theorems in this section.

caused by a shift from y to $y+\delta y$. Without H being linear, there seems little hope of quantifying that aspect of error, not to speak of relative error. Nonetheless, it will be possible to get radius theorems in which the reciprocals of $\|H^{-1}\|^{+}$ and $\|H^{-1}\|^{-}$ are featured.

For $\|H^{-1}\|^{+}$, we can utilize the inversion estimate for the outer norm in 5A.8. A definition is needed first.

Extended nonsingularity. *A positively homogeneous mapping $H : X \rightrightarrows Y$ is said to be nonsingular if $\|H^{-1}\|^{+} < \infty$; it is said to be singular if $\|H^{-1}\|^{+} = \infty$.*

As shown in 5A.7, nonsingularity of H in this sense implies that $H^{-1}(0) = \{0\}$; moreover, when $\dim X < \infty$ and $\operatorname{gph} H$ is closed the converse is true as well.

Theorem 6A.2 (radius theorem for nonsingularity of positively homogeneous mappings)**.** *For any $H : X \rightrightarrows Y$ that is positively homogeneous and nonsingular, one has*

$$\inf_{B\in\mathscr{L}(X,Y)} \left\{ \|B\| \,\Big|\, H+B \text{ is singular} \right\} = \frac{1}{\|H^{-1}\|^{+}}. \tag{3}$$

Moreover the infimum is the same if restricted to mappings B of rank one.

Proof. The proof is parallel to that of 6A.1. From 5A.8 we get "$\geq$" in (3), and also "=" for the case $\|H^{-1}\|^{+} = 0$ under the convention $1/0 = \infty$. Let $\|H^{-1}\|^{+} > 0$ and consider any $r > 1/\|H^{-1}\|^{+}$. There exists $(\hat{x},\hat{y}) \in \operatorname{gph} H$ with $\|\hat{y}\| = 1$ and $\|\hat{x}\| > 1/r$. Let $x^* \in X^*$, $x^*(\hat{x}) = \|\hat{x}\|$ and $\|x^*\| = 1$. The linear and bounded mapping

$$Bx = -\frac{x^*(x)\hat{y}}{\|\hat{x}\|}$$

has $\|B\| = 1/\|\hat{x}\| < r$ and $(H+B)(\hat{x}) = H(\hat{x}) - \hat{y} \ni 0$. Then the nonzero vector $\hat{x}$ belongs to $(H+B)^{-1}(0)$, hence $\|(H+B)^{-1}\|^{+} = \infty$, i.e., $H+B$ is singular. The infimum in (3) must therefore be less than r. Appealing to the choice of r we conclude that the infimum in (3) cannot be more than $1/\|H^{-1}\|^{+}$, and we are done. □

To develop a radius theorem about $\|H^{-1}\|^{-}$, we have to look more narrowly at *sublinear* mappings, which are characterized by having graphs that are not just cones, as corresponds to positive homogeneity, but *convex* cones. For such a mapping H, if its graph is also closed, we have an inversion estimate for the inner norm in 5C.9. Furthermore, we know from 5C.2 that the surjectivity of H is equivalent to having $\|H^{-1}\|^{-} < \infty$. We also have available the notion of the adjoint mapping as introduced in Section 5C: the upper adjoint of $H : X \rightrightarrows Y$ is the sublinear mapping $H^{*+} : Y^* \rightrightarrows X^*$ defined by

$$(y^*,x^*) \in \operatorname{gph} H^{*+} \iff \langle x^*,x\rangle \leq \langle y^*,y\rangle \text{ for all } (x,y) \in \operatorname{gph} H.$$

Recall too, from 5C.13, that for a sublinear mapping H with closed graph,

$$\|(H^{*+})^{-1}\|^{+} = \|H^{-1}\|^{-}, \tag{4}$$

and also, from 5C.14,

$$(H+B)^{*+} = H^{*+} + B^* \quad \text{for any} \quad B \in \mathscr{L}(X,Y). \tag{5}$$

Theorem 6A.3 (radius theorem for surjectivity of sublinear mappings)**.** *For any $H : X \rightrightarrows Y$ that is sublinear, surjective, and with closed graph,*

$$\inf_{B\in\mathscr{L}(X,Y)} \left\{ \|B\| \;\middle|\; H+B \text{ is not surjective} \right\} = \frac{1}{\|H^{-1}\|^{-}}.$$

Moreover the infimum is the same if restricted to B of rank one.

Proof. For any $B \in \mathscr{L}(X,Y)$, the mapping $H+B$ is sublinear with closed graph, so that $(H+B)^{*+} = H^{*+} + B^*$ by (5). By the definition of the adjoint, $H+B$ is surjective if and only if $H^{*+} + B^*$ is nonsingular. It follows that

$$\begin{aligned} &\inf_{B\in\mathscr{L}(X,Y)} \left\{ \|B\| \;\middle|\; H+B \text{ is not surjective} \right\} \\ &\qquad = \inf_{B\in\mathscr{L}(X,Y)} \left\{ \|B^*\| \;\middle|\; H^{*+} + B^* \text{ is singular} \right\}. \end{aligned} \tag{6}$$

The right side of (6) can be identified through Theorem 6A.2 with

$$\inf_{C\in\mathscr{L}(Y^*,X^*)} \left\{ \|C\| \;\middle|\; H^{*+} + C \text{ is singular} \right\} = \frac{1}{\|(H^{*+})^{-1}\|^{+}} \tag{7}$$

by the observation that any $C \in \mathscr{L}(Y^*,X^*)$ of rank one has the form B^* for some $B \in \mathscr{L}(X,Y)$ of rank one. It remains to apply the relation in (4). In consequence of that, the left side of (7) is $1/\|H^{-1}\|^{-}$, and we get the desired equality. □

In the case of H being a bounded linear mapping $A : X \to Y$, Theorems 6A.2 and 6A.3 both furnish results which complement Theorem 6A.1, since nonsingularity just comes down to A^{-1} being single-valued on rge A, while surjectivity corresponds only to dom A^{-1} being all of Y, and neither of those properties automatically entails the other. When $X = Y = \mathbb{R}^n$, of course, all three theorems reduce to the matrix result recalled at the beginning of this section.

The surjectivity result in 6A.3 offers more than an extended insight into equation solving, however. It can be applied also to systems of inequalities. This is true even in infinite dimensions, but we are not yet prepared to speak of inequality constraints in that framework, so we limit the following illustration to solving $Ax \le y$ in the case of a matrix $A \in \mathbb{R}^{m\times n}$. It will be convenient to say that

$Ax \le y$ is *universally solvable* if it has a solution $x \in \mathbb{R}^n$ for every $y \in \mathbb{R}^m$.

We adopt for $y = (y_1, \dots, y_m) \in \mathbb{R}^m$ the maximum norm $|y|_\infty = \max_{1\le k\le m} |y_k|$ but equip $\mathbb{R}^n$ with any norm. The associated operator norm for linear mappings acting from $\mathbb{R}^n$ to $\mathbb{R}^m$ is denoted by $|\cdot|_\infty$. Also, we use the notation for $y = (y_1, \dots, y_m)$ that

$$y^+ = (y_1^+, \dots, y_m^+), \text{ where } y_k^+ = \max\{0, y_k\}.$$

Example 6A.4 (radius of universal solvability for systems of linear inequalities)**.** Suppose for a matrix $A \in I\!R^{m\times n}$ that $Ax \le y$ is universally solvable. Then

$$\inf_{B\in I\!R^{m\times n}} \left\{ |B|_\infty \,\middle|\, (A+B)x \le y \text{ is not universally solvable} \right\} \;=\; \frac{1}{\sup_{|x|\le 1} |\,[Ax]^+|_\infty}.$$

Detail. We apply Theorem 6A.3 to the mapping $F : x \mapsto \{ y \,|\, Ax \le y \}$. Then $|F^{-1}|^- = \sup_{|x|\le 1} \inf_{y\ge Ax} |y|_\infty$, where the infimum equals $|\,[Ax]^+|_\infty$. □

More will be said about constraint systems in Section 6B.

Since surjectivity of a sublinear mapping is also equivalent to its metric regularity at 0 for 0, we could restate Theorem 6A.3 in terms of metric regularity as well. Such a result is actually true for any strictly differentiable function. Specifically, the basic Lyusternik–Graves theorem 5D.5 says that, for a function $f : X \to Y$ which is strictly differentiable at $\bar{x}$, one has

$$\operatorname{reg}(f;\bar{x}\,|\,\bar{y}) = \operatorname{reg} Df(\bar{x}) \quad \text{for } \bar{y} = f(\bar{x}).$$

Since any linear mapping is sublinear, this equality combined with 6A.3 gives us yet another radius result.

Corollary 6A.5 (radius theorem for metric regularity of strictly differentiable functions)**.** *Let $f : X \to Y$ be strictly differentiable at $\bar{x}$, let $\bar{y} := f(\bar{x})$, and let $Df(\bar{x})$ be surjective. Then*

$$\inf_{B\in \mathscr{L}(X,Y)} \left\{ \|B\| \,\middle|\, f+B \text{ is not metrically regular at } \bar{x} \text{ for } \bar{y}+B\bar{x} \right\} \;=\; \frac{1}{\|Df(\bar{x})^{-1}\|^-}.$$

It should not escape attention here that in 6A.5 we are not focused any more on the origins of X and Y but on a general pair $(\bar{x},\bar{y})$ in the graph of f. This allows us to return to "conditioning" from a different perspective, if we are willing to think of such a property in a local sense only.

Suppose that a y near to $\bar{y}$ is perturbed to $y+\delta y$. The solution set $f^{-1}(y)$ to the problem of solving $f(x) \ni y$ is thereby shifted to $f^{-1}(y+\delta y)$, and we have an interest in understanding the "error" vectors δx such that $x+\delta x \in f^{-1}(y+\delta y)$. Since anyway x need not be the only element of $f^{-1}(y)$, it is appropriate to quantify the shift by looking for the smallest size of δx, or in other words at $\operatorname{dist}(x, f^{-1}(y+\delta y))$ and how it compares to $\|\delta y\|$. This ratio, in its limit as (x,y) goes to $(\bar{x},\bar{y})$ and $\|\delta y\|$ goes to 0, is precisely $\operatorname{reg}(f;\bar{x}\,|\,\bar{y})$.

In this sense, $\operatorname{reg}(f;\bar{x}\,|\,\bar{y})$ can be deemed the absolute condition number *locally* with respect to $\bar{x}$ and $\bar{y}$ for the problem of solving $f(x) \ni y$ for x in terms of y. We then have a *local, nonlinear* analog of Theorem 6A.1, tying a condition number to a radius. It provides something more even for linear f, of course, since in contrast to Theorem 6A.1, it imposes no requirement of invertibility.

Corollary 6A.5 can be stated in a more general form, which we give here as an exercise:

Exercise 6A.6. *Let $F : X \rightrightarrows Y$ be metrically regular at $\bar{x}$ for $\bar{y}$, and let $f : X \to Y$ satisfy $\bar{x} \in \operatorname{int}\operatorname{dom} f$ and $\operatorname{lip}(f;\bar{x}) = 0$. Then*

$$\inf_{B\in\mathscr{L}(X,Y)} \left\{ \|B\| \;\middle|\; F+B \text{ is not metrically regular at } \bar{x} \text{ for } \bar{y}+B\bar{x} \right\}$$
$$= \inf_{B\in\mathscr{L}(X,Y)} \left\{ \|B\| \;\middle|\; F+f+B \text{ is not metrically regular at } \bar{x} \text{ for } \bar{y}+f(\bar{x})+B\bar{x} \right\}.$$

Guide. Observe that, by the Banach space version of 3F.3 (which follows from 5E.1), the mapping $F+B$ is metrically regular at $\bar{y}+B\bar{x}$ if and only if the mapping $F+f+B$ is metrically regular at $\bar{y}+f(\bar{x})+B\bar{x}$. □

We will show next that, in *finite* dimensions at least, the radius result in 6A.5 is valid when f is replaced by *any set-valued mapping F* whose graph is locally closed around the reference pair $(\bar{x},\bar{y})$.

Theorem 6A.7 (radius theorem for metric regularity)**.** *Let X and Y be Euclidean spaces, and for $F : X \rightrightarrows Y$ and $\bar{y} \in F(\bar{x})$ let $\operatorname{gph} F$ be locally closed at $(\bar{x},\bar{y})$. Then*

$$\text{(8)}\quad \inf_{B\in\mathscr{L}(X,Y)} \left\{ |B| \;\middle|\; F+B \text{ is not metrically regular at } \bar{x} \text{ for } \bar{y}+B\bar{x} \right\} = \frac{1}{\operatorname{reg}(F;\bar{x}\,|\,\bar{y})}.$$

Moreover, the infimum is unchanged if taken with respect to linear mappings of rank 1, but also remains unchanged when the class of perturbations B is enlarged to all locally Lipschitz continuous functions g, with $|B|$ replaced by the Lipschitz modulus $\operatorname{lip}(g;\bar{x})$ of g at $\bar{x}$.

Proof. If F is not metrically regular at $\bar{x}$ for $\bar{y}$, then (8) holds under the convention $1/\infty = 0$. Let $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) < \infty$. The general perturbation inequality derived in Theorem 5E.1, see 5E.8, produces the estimate

$$\inf_{g:X\to Y} \left\{ \operatorname{lip}(g;\bar{x}) \;\middle|\; F+g \text{ is not metrically regular at } \bar{x} \text{ for } \bar{y}+g(\bar{x}) \right\} \;\geq\; \frac{1}{\operatorname{reg}(F;\bar{x}\,|\,\bar{y})},$$

which becomes the equality (8) in the case when $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = 0$ under the convention $1/0 = \infty$. To confirm the opposite inequality when $\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) > 0$, we apply Theorem 4C.6, according to which

$$\text{(9)}\qquad \operatorname{reg}(F;\bar{x}\,|\,\bar{y}) = \limsup_{\substack{(x,y)\to(\bar{x},\bar{y})\\ (x,y)\in\operatorname{gph} F}} |\tilde{D}F(x\,|\,y)^{-1}|^{-},$$

where $\tilde{D}F(x\,|\,y)$ is the convexified graphical derivative of F at x for y. Take a sequence of positive real numbers $\varepsilon_k \searrow 0$. Then for k sufficiently large, say $k > \bar{k}$, by (9) there exists $(x_k, y_k) \in \operatorname{gph} F$ with $(x_k, y_k) \to (\bar{x},\bar{y})$ and

$$\operatorname{reg}(F;\bar{x}\,|\,\bar{y}) + \varepsilon_k \geq |\tilde{D}F(x_k|y_k)^{-1}|^{-} \geq \operatorname{reg}(F;\bar{x}\,|\,\bar{y}) - \varepsilon_k > 0.$$

Let $H_k := \tilde{D}F(x_k|y_k)$ and $S_k := H_k^{*+}$; then norm duality gives us $|H_k^{-1}|^- = |S_k^{-1}|^+$, see 5C.13.

For each $k > \bar{k}$ choose a positive real r_k satisfying $|S_k^{-1}|^+ - \varepsilon_k < 1/r_k < |S_k^{-1}|^+$. From the last inequality there must exist $(\hat{y}_k, \hat{x}_k) \in \operatorname{gph} S_k$ with $|\hat{x}_k| = 1$ and $|S_k^{-1}|^+ \geq |\hat{y}_k| > 1/r_k$. Pick $y_k^* \in Y$ with $\langle \hat{y}_k, y_k^* \rangle = |\hat{y}_k|$ and $|y_k^*| = 1$, and define the rank-one mapping $\hat{G}_k \in \mathscr{L}(Y,X)$ by

$$\hat{G}_k(y) := -\frac{\langle y, y_k^* \rangle}{|\hat{y}_k|}\hat{x}_k.$$

Then $\hat{G}_k(\hat{y}_k) = -\hat{x}_k$ and hence $(S_k + \hat{G}_k)(\hat{y}_k) = S_k(\hat{y}_k) + \hat{G}_k(\hat{y}_k) = S_k(\hat{y}_k) - \hat{x}_k \ni 0$. Therefore, $\hat{y}_k \in (S_k + \hat{G}_k)^{-1}(0)$, and since $\hat{y}_k \neq 0$ and S_k is positively homogeneous with closed graph, we have by Proposition 5A.7, formula 5A(10), that

$$(10) \qquad |(S_k + \hat{G}_k)^{-1}|^+ = \infty.$$

Note that $|\hat{G}_k| = |\hat{x}_k|/|\hat{y}_k| = 1/|\hat{y}_k| < r_k$.

Since the sequences $\hat{y}_k$, $\hat{x}_k$ and y_k^* are bounded (and the spaces are *finite-dimensional*), we can extract from them subsequences converging respectively to $\hat{y}$, $\hat{x}$ and y^*. The limits then satisfy $|\hat{y}| = \operatorname{reg}(F;\bar{x}\,|\,\bar{y})$, $|\hat{x}| = 1$ and $|y^*| = 1$. Define the rank-one mapping $\hat{G} \in \mathscr{L}(Y,X)$ by

$$\hat{G}(y) := -\frac{\langle y, y^* \rangle}{|\hat{y}|}\hat{x}.$$

Then we have $|\hat{G}| = \operatorname{reg}(F;\bar{x}\,|\,\bar{y})^{-1}$ and $|\hat{G}_k - \hat{G}| \to 0$.

Let $B := (\hat{G})^*$ and suppose $F + B$ is metrically regular at $\bar{x}$ for $\bar{y} + B\bar{x}$. Theorem 4C.6 yields that there is a finite positive constant c such that for $k > \bar{k}$ sufficiently large, we have

$$c > |\tilde{D}(F+B)(x_k|y_k + Bx_k)^{-1}|^-.$$

Through 4C.7, this gives us

$$c > \|(\tilde{D}F(x_k|y_k) + B)^{-1}\|^- = \|(H_k + B)^{-1}\|^-.$$

Since B is linear, we have $B^* = ((\hat{G})^*)^* = \hat{G}$, and since $H_k + B$ is sublinear, it follows further by 5C.14 that

$$(11) \qquad c > |([H_k + B]^{*+})^{-1}|^+ = |(H_k^{*+} + B^*)^{-1}|^+ = |(S_k + \hat{G})^{-1}|^+.$$

Take $k > \bar{k}$ sufficiently large such that $|\hat{G} - \hat{G}_k| \leq 1/(2c)$. Setting $P_k := S_k + \hat{G}$ and $Q_k := \hat{G}_k - \hat{G}$, we have that

$$[|P_k^{-1}|^+]^{-1} \geq 1/c > 1/(2c) \geq |Q_k|.$$

By using the inversion estimate for the outer norm in 5A.8, we have

$$|(S_k+\hat{G}_k)^{-1}|^+ = |(P_k+Q_k)^{-1}|^+ \le \big([|P_k^{-1}|^+]^{-1} - |Q_k|\big)^{-1} \le 2c < \infty.$$

This contradicts (10). Hence, $F+B$ is not metrically regular at $\bar{x}$ for $\bar{y}+B\bar{x}$. Noting that $|B| = |\hat{G}| = 1/\mathrm{reg}\,(F;\bar{x}\,|\,\bar{y})$ and that B is of rank one, we are finished. □

In a pattern just like the one laid out after Corollary 6A.5, it is appropriate to consider $\mathrm{reg}\,(F;\bar{x}\,|\,\bar{y})$ as the *local* absolute condition number with respect to $\bar{x}$ and $\bar{y}$ for the problem of solving $F(x) \ni y$ for x in terms of y. An even grander extension of the fact in 6A.1, that the reciprocal of the absolute condition number gives the radius of perturbation for preserving an associated property, is thereby achieved.

Based on Theorem 6A.7, it is now easy to obtain a parallel radius result for strong metric regularity.

Theorem 6A.8 (radius theorem for strong metric regularity)**.** *Let X and Y be Euclidean spaces, and let $F : X \rightrightarrows Y$ be strongly metrically regular at $\bar{x}$ for $\bar{y}$. Then*

$$(12)\qquad \inf_{B\in\mathscr{L}(X,Y)}\Big\{|B|\;\Big|\;F+B \text{ is not strongly regular at } \bar{x} \text{ for } \bar{y}+B\bar{x}\Big\} = \frac{1}{\mathrm{reg}\,(F;\bar{x}\,|\,\bar{y})}.$$

Moreover, the infimum is unchanged if taken with respect to linear mappings of rank 1, but also remains unchanged when the class of perturbations B is enlarged to the class of locally Lipschitz continuous functions g with $|B|$ replaced by the Lipschitz modulus $\mathrm{lip}\,(g;\bar{x})$.

Proof. Theorem 5F.1 reveals that "$\ge$" holds in (12) when the linear perturbation is replaced by a Lipschitz perturbation, and moreover that (12) is satisfied in the limit case $\mathrm{reg}\,(F;\bar{x}\,|\,\bar{y}) = 0$ under the convention $1/0 = \infty$. The inequality becomes an equality with the observation that the assumed strong metric regularity of F implies that F is metrically regular at $\bar{x}$ for $\bar{y}$. Hence the infimum in (12) is not greater than the infimum in (8). □

Next comes a radius theorem for strong subregularity to go along with the ones for metric regularity and strong metric regularity.

Theorem 6A.9 (radius theorem for strong metric subregularity)**.** *Let X and Y be Euclidean spaces, and for for $F : X \rightrightarrows Y$ and $\bar{y} \in F(\bar{x})$ let* gph F *be locally closed at $(\bar{x},\bar{y})$. Suppose that F is strongly metrically subregular at $\bar{x}$ for $\bar{y}$. Then*

$$\inf_{B\in\mathscr{L}(X,Y)}\Big\{\,|B|\;\Big|\;F+B \text{ is not strongly subregular at } \bar{x} \text{ for } \bar{y}+B\bar{x}\,\Big\} = \frac{1}{\mathrm{subreg}\,(F;\bar{x}\,|\,\bar{y})}.$$

Moreover, the infimum is unchanged if taken with respect to mappings B of rank 1, but also remains unchanged when the class of perturbations is enlarged to the class of functions $g : X \to Y$ that are calm at $\bar{x}$ and continuous around $\bar{x}$, with $|B|$ replaced by the calmness modulus $\mathrm{clm}\,(g;\bar{x})$.

Proof. From the equivalence of the strong subregularity of a mapping F at $\bar{x}$ for $\bar{y}$ with the nonsingularity of its graphical derivative $DF(\bar{x}\,|\,\bar{y})$, as shown in Theorem 4E.1, we have

$$\begin{aligned}(13)\qquad &\inf_{B\in\mathscr{L}(X,Y)}\left\{\,|B|\,\middle|\,F+B \text{ is not strongly subregular at } \bar{x} \text{ for } \bar{y}+B\bar{x}\,\right\}\\ &\qquad = \inf_{B\in\mathscr{L}(X,Y)}\left\{\,|B|\,\middle|\,D(F+B)(\bar{x}\,|\,\bar{y}+B\bar{x}) \text{ is singular}\,\right\}.\end{aligned}$$

We know from the sum rule for graphical differentiation (see 4A.2) that $D(F+B)(\bar{x}\,|\,\bar{y}+B\bar{x}) = DF(\bar{x}\,|\,\bar{y})+B$, hence

$$\begin{aligned}(14)\qquad &\inf_{B\in\mathscr{L}(X,Y)}\left\{\,|B|\,\middle|\,D(F+B)(\bar{x}\,|\,\bar{y}+B\bar{x}) \text{ is singular}\,\right\}\\ &\qquad = \inf_{B\in\mathscr{L}(X,Y)}\left\{\,|B|\,\middle|\,DF(\bar{x}\,|\,\bar{y})+B \text{ is singular}\,\right\}.\end{aligned}$$

Since $DF(\bar{x}\,|\,\bar{y})+B$ is positively homogeneous, 6A.2 translates to

$$(15)\qquad \inf_{B\in\mathscr{L}(X,Y)}\left\{\,|B|\,\middle|\,DF(\bar{x}\,|\,\bar{y})+B \text{ is singular}\,\right\} = \frac{1}{|DF(\bar{x}\,|\,\bar{y})^{-1}|^{+}},$$

including the case $|DF(\bar{x}\,|\,\bar{y})^{-1}|^{+}=0$ with the convention $1/0=\infty$. Theorem 4E.1 tells us also that $|DF(\bar{x}\,|\,\bar{y})^{-1}|^{+} = \operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$ and then, putting together (13), (14) and (15), we get the desired equality. □

As with the preceding results, the modulus $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$ can be regarded as a sort of local absolute condition number. But in this case only the ratio of $\operatorname{dist}(\bar{x}, F^{-1}(\bar{y}+\delta y))$ to $|\delta y|$ is considered in its limsup as δy goes to 0, not the limsup of all the ratios $\operatorname{dist}(x, F^{-1}(y+\delta y))/|\delta y|$ with $(x,y)\in \operatorname{gph} F$ tending to $(\bar{x},\bar{y})$, which gives $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$. Specifically, with $\operatorname{reg}(F;\bar{x}\,|\,\bar{y})$ appropriately termed the absolute condition number for F locally with respect to $\bar{x}$ and $\bar{y}$, $\operatorname{subreg}(F;\bar{x}\,|\,\bar{y})$ is the corresponding *subcondition number*.

The radius-type theorems above could be rewritten in terms of the associated equivalent properties of the inverse mappings. For example, Theorem 6A.7 could be restated in terms of perturbations B of a mapping F whose inverse has the Aubin property.

6B. Constraints and Feasibility

Universal solvability of systems of linear inequalities in finite dimensions was already featured in Example 6A.4 as an application of one of our radius theorems, but now we will go more deeply into the subject of constraint systems and their solvability. We take as our focus problems of the very general type

$$(1)\qquad \text{find } x \text{ such that } F(x)\ni 0$$

for a set-valued mapping $F : X \rightrightarrows Y$ from one Banach space to another. Of course, the set of all solutions is just $F^{-1}(0)$, but we are thinking of F as representing a kind of constraint system and are concerned with whether the set $F^{-1}(0)$ might shift from nonempty to empty under some sort of perturbation. Mostly, we will study the case where F has convex graph.

Feasibility. *Problem (1) will be called feasible if* $F^{-1}(0) \neq \emptyset$, *i.e.,* $0 \in \operatorname{rge} F$, *and strictly feasible if* $0 \in \operatorname{int} \operatorname{rge} F$.

Two examples will point the way toward progress. Recall that any closed, convex cone $K \subset Y$ with nonempty interior induces a partial ordering "$\leq_K$" under the rule that $y_0 \leq_K y_1$ means $y_1 - y_0 \in K$. Correspondingly, $y_0 <_K y_1$ means $y_1 - y_0 \in \operatorname{int} K$.

Example 6B.1 (convex constraint systems)**.** Let $C \subset X$ be a closed convex set, let $K \subset Y$ be a closed convex cone, and let $A : C \to Y$ be a continuous and convex mapping with respect to the partial ordering in Y induced by K; that is,

$$A((1-\theta)x_0 + \theta x_1) \leq_K (1-\theta)A(x_0) + \theta A(x_1) \quad \text{for } x_0, x_1 \in C \text{ when } 0 < \theta < 1.$$

Define the mapping $F : X \rightrightarrows Y$ by

$$F(x) = \begin{cases} A(x) + K & \text{if } x \in C, \\ \emptyset & \text{if } x \notin C. \end{cases}$$

Then F has closed, convex graph, and feasibility of solving $F(x) \ni 0$ for x refers to

$$\exists \bar{x} \in C \text{ such that } A(\bar{x}) \leq_K 0.$$

On the other hand, as long as $\operatorname{int} K \neq \emptyset$, strict feasibility refers to

$$\exists \bar{x} \in C \text{ such that } A(\bar{x}) <_K 0.$$

Detail. To see this, note that, when $\operatorname{int} K \neq \emptyset$, we have $K = \operatorname{cl} \operatorname{int} K$, so that the convex set $\operatorname{rge} F = A(C) + K$ is the closure of the open set $O := A(C) + \operatorname{int} K$. Also, O is convex. It follows then that $\operatorname{int} \operatorname{rge} F = O$. □

Example 6B.2 (linear-conic constraint systems)**.** Consider the convex constraint system in Example 6B.1 under the additional assumptions that A is linear and C is a cone, so that the condition $\bar{x} \in C$ can be written equivalently as $\bar{x} \geq_C 0$. Then F is sublinear, and its adjoint $F^{*+} : Y^* \rightrightarrows X^*$ is given in terms of the adjoint A^* of A and the dual cones $K^+ = -K^*$ and $C^+ = -C^*$ (where * denotes polar) by

$$F^{*+}(y^*) = \begin{cases} A^*(y^*) - C^+ & \text{if } y^* \in K^+, \\ \emptyset & \text{if } y^* \notin K^+, \end{cases}$$

so that $F^{*+}(y^*) \ni x^*$ if and only if $y^* \geq_{K^+} 0$ and $A^*(y^*) \geq_{C^+} x^*$.

Detail. In this case the graph of F is clearly a convex cone, and that means F is sublinear. The claims about the adjoint of F follow by elementary calculation. □

Along the lines of the analysis in 6A, in dealing with the feasibility problem (1) we will be interested in perturbations in which F is replaced by $F+B$ for some $B \in \mathscr{L}(X,Y)$, and at the same time, the zero on the right is replaced by some other $b \in Y$. Such a double perturbation, the magnitude of which can be quantified by the norm

$$\|(B,b)\| = \max\{\, \|B\|, \|b\| \,\}, \tag{2}$$

transforms the condition $F(x) \ni 0$ to $(F+B)(x) \ni b$ and the solution set $F^{-1}(0)$ to $(F+B)^{-1}(b)$, creating infeasibility if $(F+B)^{-1}(b) = \emptyset$, i.e., if $b \notin \operatorname{rge}(F+B)$. We want to understand how large $\|(B,b)\|$ can be before this happens.

Distance to infeasibility. *For $F : X \rightrightarrows Y$ with convex graph and $0 \in \operatorname{rge} F$, the distance to infeasibility of the system $F(x) \ni 0$ is defined to be the value*

$$\inf_{B \in \mathscr{L}(X,Y),\, b \in Y} \left\{ \|(B,b)\| \,\middle|\, b \notin \operatorname{rge}(F+B) \right\}. \tag{3}$$

Surprisingly, perhaps, it turns out that there would be no difference if feasibility were replaced by strict feasibility in this definition. In the next pair of lemmas, it is assumed that $F : X \rightrightarrows Y$ has convex graph and $0 \in \operatorname{rge} F$.

Lemma 6B.3 (distance to infeasibility versus distance to strict infeasibility)**.** *The distance to infeasibility is the same as the distance to strict infeasibility, namely the value*

$$\inf_{B \in \mathscr{L}(X,Y),\, b \in Y} \left\{ \|(B,b)\| \,\middle|\, b \notin \operatorname{int} \operatorname{rge}(F+B) \right\}. \tag{4}$$

Proof. Let S_1 denote the set of (B,b) over which the infimum is taken in (3) and let S_2 be the corresponding set in (4). Obviously $S_1 \subset S_2$, so the first infimum cannot be less than the second. We must show that it also cannot be greater. This amounts to demonstrating that for every $(B,b) \in S_2$ and every $\varepsilon > 0$ we can find $(B',b') \in S_1$ such that $\|(B',b')\| \le \|(B,b)\| + \varepsilon$. In fact, we can get this with $B' = B$ simply by noting that when $b \notin \operatorname{int} \operatorname{rge}(F+B)$ there must exist $b' \in Y$ with $b' \notin \operatorname{rge}(F+B)$ and $\|b' - b\| \le \varepsilon$. □

By utilizing the Robinson–Ursescu theorem 5B.4, we can see furthermore that the distance to infeasibility is actually the same as the distance to metric nonregularity:

Lemma 6B.4 (distance to infeasibility equals radius of metric regularity)**.** *The distance to infeasibility in problem (1) coincides with the value*

$$\inf_{(B,b) \in \mathscr{L}(X,Y) \times Y} \left\{ \|(b,B)\| \,\middle|\, F+B \text{ is not metrically regular at any } \bar{x} \text{ for } b \right\}. \tag{5}$$

Proof. In view of the equivalence of infeasibility with strict feasibility in 6B.3, the Robinson–Ursescu theorem 5B.4 just says that problem (1) is feasible if and only if F is metrically regular at $\bar{x}$ for 0 for any $\bar{x} \in F^{-1}(0)$, hence (5). □

In order to estimate the distance to infeasibility in terms of the modulus of metric regularity, we pass from F to a special mapping $\bar{F}$ constructed as a "homogenization" of F. We will then be able to apply to $\bar{F}$ the result on distance to metric nonregularity of sublinear mappings given in 6A.3.

We use the *horizon mapping* F^∞ associated with F, the graph of F^∞ in $X \times Y$ being the recession cone of gph F in the sense of convex analysis:

$$(x', y') \in \operatorname{gph} F^\infty \iff \operatorname{gph} F + (x', y') \subset \operatorname{gph} F.$$

Homogenization. *For $F : X \rightrightarrows Y$ and $0 \in \operatorname{rge} F$, the homogenization of the constraint system $F(x) \ni 0$ in (1) is the system $\bar{F}(x,t) \ni 0$, where $\bar{F} : X \times I\!R \rightrightarrows Y$ is defined by*

$$\bar{F}(x,t) = \begin{cases} tF(t^{-1}x) & \text{if } t > 0, \\ F^\infty(x) & \text{if } t = 0, \\ \emptyset & \text{if } t < 0. \end{cases}$$

The solution sets to the two systems are related by

$$x \in F^{-1}(0) \iff (x,1) \in \bar{F}^{-1}(0).$$

Note that if F is positively homogeneous with closed graph, then $tF(t^{-1}x) = F(x) = F^\infty(x)$ for all $t > 0$, so that we simply have $\bar{F}(x,t) = F(x)$ for $t \geq 0$, but $\bar{F}(x,t) = \emptyset$ for $t < 0$.

In what follows, we adopt the norm

$$\|(x,t)\| = \|x\| + |t| \quad \text{for } (x,t) \in X \times I\!R. \tag{6}$$

We are now ready to state and prove a result which gives a quantitative expression for the magnitude of the distance to infeasibility:

Theorem 6B.5 (distance to infeasibility for the homogenized mapping)**.** *Let $F : X \rightrightarrows Y$ have closed, convex graph and let $0 \in \operatorname{rge} F$. Then in the homogenized system $\bar{F}(t,x) \ni 0$ the mapping $\bar{F}$ is sublinear with closed graph, and*

$$0 \in \operatorname{int} \operatorname{rge} F \iff 0 \in \operatorname{int} \operatorname{rge} \bar{F} \iff \bar{F} \text{ is surjective.} \tag{7}$$

Furthermore, for the given constraint system $F(x) \ni 0$ one has

$$\text{distance to infeasibility} = 1/\operatorname{reg}(\bar{F}; 0, 0 \,|\, 0). \tag{8}$$

Proof. The definition of $\bar{F}$ corresponds to gph $\bar{F}$ being the closed convex cone in $X \times I\!R \times Y$ that is generated by $\{(x,1,y) \,|\, (x,y) \in \operatorname{gph} F\}$. Hence $\bar{F}$ is sublinear, and also, rge $\bar{F}$ is a convex cone. We have $(\operatorname{rge} F) = F(X) = \bar{F}(X,1)$. So it is obvious that if $0 \in \operatorname{int} \operatorname{rge} F$, then $0 \in \operatorname{int} \operatorname{rge} \bar{F}$. Since $\operatorname{rge} \bar{F}$ is a convex cone, the latter is equivalent to having $\operatorname{rge} \bar{F} = Y$, i.e., surjectivity.

Conversely now, suppose $\bar{F}$ is surjective. Theorem 5B.4 (Robinson–Ursescu) informs us that in this case, $0 \in \operatorname{int} \bar{F}(W)$ for every neighborhood W of the origin in $\mathbb{R} \times X$. It must be verified, however, that $0 \in \operatorname{int} \operatorname{rge} F$. In terms of $C(t) = \bar{F}(\mathbb{B}, t) \subset Y$, it will suffice to show that $0 \in \operatorname{int} C(t)$ for some $t > 0$. Note that the sublinearity of $\bar{F}$ implies that

$$(9) \qquad C((1-\theta)t_0 + \theta t_1) \supset (1-\theta)C(t_0) + \theta C(t_1) \quad \text{for } 0 < \theta < 1.$$

Our assumption that $0 \in \operatorname{rge} F$ ensures having $F^{-1}(0) \neq \emptyset$. Choose $\tau \in (0, \infty)$ small enough that $1/(2\tau) > d(0, F^{-1}(0))$. Then

$$(10) \qquad 0 \in C(t) \quad \text{for all} \quad t \in [0, 2\tau],$$

whereas, because $[-2\tau, 2\tau] \times \mathbb{B}$ is a neighborhood W of the origin in $\mathbb{R} \times X$, we have

$$(11) \qquad 0 \in \operatorname{int} \bar{F}(\mathbb{B}, [-2\tau, 2\tau]) = \operatorname{int} \bigcup_{0 \le t \le 2\tau} C(t).$$

We will use this to show that actually $0 \in \operatorname{int} C(\tau)$. For $y^* \in Y^*$ define

$$\sigma(y^*, t) := \sup_{y \in C(t)} \langle y, y^* \rangle, \qquad \lambda(t) := \inf_{\|y^*\|=1} \sigma(y^*, t).$$

The property in (9) makes $\sigma(y^*, t)$ concave in t, and the same then follows for $\lambda(t)$. As long as $0 \le t \le 2\tau$, we have $\sigma(y^*, t) \ge 0$ and $\lambda(t) \ge 0$ by (10). On the other hand, the union in (11) includes some ball around the origin. Therefore,

$$(12) \qquad \exists \varepsilon > 0 \text{ such that } \sup_{0 \le t \le 2\tau} \sigma(y^*, t) \ge \varepsilon \text{ for all } y^* \in Y^* \text{ with } \|y^*\| = 1.$$

We argue next that $\lambda(\tau) > 0$. If not, then since λ is a nonnegative concave function on $[0, 2\tau]$, we would have to have $\lambda(t) = 0$ for all $t \in [0, 2\tau]$. Supposing that to be the case, choose $\delta \in (0, \varepsilon/2)$ and, in the light of the definition of $\lambda(\tau)$, an element $\hat{y}^*$ with $\sigma(\hat{y}^*, \tau) < \delta$. The nonnegativity and concavity of $\sigma(\hat{y}^*, \cdot)$ on $[0, 2\tau]$ imply then that $\sigma(\hat{y}^*, t) \le (\delta/\tau)t$ when $\tau \le t \le 2\tau$ and $\sigma(\hat{y}^*, t) \le 2\delta - (\delta/\tau)t$ when $0 \le t \le \tau$. But that gives us $\sigma(\hat{y}^*, t) \le 2\delta < \varepsilon$ for all $t \in [0, 2\tau]$, in contradiction to the property of ε in (12). Therefore, $\lambda(\tau) > 0$, as claimed.

We have $\sigma(y^*, \tau) \ge \lambda(\tau)$ when $\|y^*\| = 1$, and hence by positive homogeneity $\sigma(y^*, \tau) \ge \lambda(\tau)\|y^*\|$ for all $y^* \in Y^*$. In this inequality, $\sigma(\cdot, \tau)$ is the support function of the convex set $C(\tau)$, or equivalently of $\operatorname{cl} C(\tau)$, whereas $\lambda(\tau)\|\cdot\|$ is the support function of $\lambda(\tau)\mathbb{B}$. It follows therefore that $\operatorname{cl} C(\tau) \supset \lambda(\tau)\mathbb{B}$, so that at least $0 \in \operatorname{int} \operatorname{cl} C(\tau)$.

Now, remembering that $C(\tau) = \tau F(\tau^{-1}\mathbb{B})$, we obtain $0 \in \operatorname{int} \operatorname{cl} F(\tau^{-1}\mathbb{B})$. Consider the mapping

$$\tilde{F}(x) = \begin{cases} F(x) & \text{if } x \in \tau^{-1}\mathbb{B}, \\ \emptyset & \text{otherwise.} \end{cases}$$

Clearly $\operatorname{rge} \tilde{F} \subset \operatorname{rge} F$. Applying Theorem 5B.1 to the mapping $\tilde{F}$ gives us

$$0 \in \operatorname{int} \operatorname{cl} \operatorname{rge} \tilde{F} = \operatorname{int} \operatorname{rge} \tilde{F} \subset \operatorname{int} \operatorname{rge} F.$$

This completes the proof of (7).

Let us turn now to (8). The first thing to observe is that every $\bar{B} \in \mathscr{L}(X \times \mathbb{R}, Y)$ can be identified with a pair $(B,b) \in \mathscr{L}(X,Y) \times Y$ under the formula $\bar{B}(x,t) = B(x) - tb$. Moreover, under this identification we get $\|\bar{B}\|$ equal to the expression in (2), due to the choice of norm in (6). The next thing to observe is that

$$(\bar{F} + \bar{B})(x,t) = \begin{cases} t(F+B)(t^{-1}x) - tb & \text{if } t > 0, \\ (F+B)^{\infty}(x) & \text{if } t = 0, \\ \emptyset & \text{if } t < 0, \end{cases}$$

so that $\bar{F} + \bar{B}$ gives the homogenization of the perturbed system $(F+B)(x) \ni b$. Therefore, on the basis of what has so far been proved, we have

$$b \in \operatorname{int} \operatorname{rge}(F+B) \iff \bar{F} + \bar{B} \text{ is surjective.}$$

Hence, through Lemma 6B.4, the distance to infeasibility for the system $F(x) \ni 0$ is the infimum of $\|\bar{B}\|$ over all $\bar{B} \in \mathscr{L}(X \times \mathbb{R}, Y)$ such that $\bar{F} + \bar{B}$ is not surjective. Theorem 6A.3 then furnishes the conclusion in (8). □

Passing to the adjoint mapping, we can obtain a "dual" formula for the distance to infeasibility:

Corollary 6B.6 (distance to infeasibility for closed convex processes)**.** *Let $F : X \rightrightarrows Y$ have closed, convex graph, and let $0 \in \operatorname{rge} F$. Define the convex, positively homogeneous function $h : X^* \times Y^* \to (-\infty, \infty]$ by*

$$h(x^*, y^*) = \sup_{x,y} \{ \langle x, x^* \rangle - \langle y, y^* \rangle \mid y \in F(x) \}.$$

Then for the system $F(x) \ni 0$,

$$\text{(13)} \qquad \textit{distance to infeasibility} = \inf_{\|y^*\|=1,\ x^*} \max \Big\{ \|x^*\|,\ h(x^*, y^*) \Big\}.$$

Proof. By Theorem 6B.5, the distance to infeasibility is $1/\operatorname{reg}(\bar{F}; 0, 0 \,|\, 0)$. On the other hand, $\operatorname{reg}(\bar{F}; 0, 0 \,|\, 0) = \|(\bar{F}^{*+})^{-1}\|^{+}$ for the adjoint mapping $\bar{F}^{*+} : Y^* \rightrightarrows X^* \times \mathbb{R}$. By definition, $(x^*, s) \in \bar{F}^{*+}(y^*)$ if and only if $(x^*, s, -y^*)$ belongs to the polar cone $(\operatorname{gph} \bar{F})^*$. Because $\operatorname{gph} \bar{F}$ is the closed convex cone generated by $\{ (x,1,y) \mid (x,y) \in \operatorname{gph} F \}$, this condition is the same as

$$s + \langle x, x^* \rangle - \langle y, y^* \rangle \le 0 \text{ for all } (x,y) \in \operatorname{gph} F$$

and can be expressed as $s + h(x^*, y^*) \le 0$. Hence

(14) $$\|(F^{*+})^{-1}\|^{+} = \sup\Big\{ \|y^*\| \,\Big|\, \|(x^*,s)\| \le 1,\ s + h(x^*,y^*) \le 0 \Big\},$$

where the norm on $X^* \times I\!R$ dual to the one in (6) is $\|(x^*,s)\| = \max\{\|x^*\|,|s|\}$. The distance to infeasibility, being the reciprocal of the quantity in (14), can be expressed therefore (through the positive homogeneity of h) as

(15) $$\inf_{\|y^*\|=1,\ x^*,s} \Big\{ \max\{\|x^*\|,|s|\} \,\Big|\, s + h(x^*,y^*) \le 0 \Big\}.$$

(In converting from (14) to an infimum restricted to $\|y^*\| = 1$ in (15), we need to be cautious about the possibility that there might be no elements $(x^*,s,y^*) \in \operatorname{gph}(\bar{F}^{*+})^{-1}$ with $y^* \neq 0$, in which case the infimum in (15) is ∞. But then the expression in (10) is 0, so the statement remains correct under the convention $1/\infty = 0$.) Observe next that, in the infimum in (15), s will be taken to be as near to 0 as possible while maintaining $-s \ge h(x^*,y^*)$. Thus, $|s|$ will be the max of 0 and $h(x^*,y^*)$, and $\max\{\|x\|,|s|\}$ will be the max of these two quantities and $\|x\|$ —but then the 0 is superfluous, and we end up with (15) equaling the expression on the right side of (3). □

We can now present our result for homogeneous systems:

Corollary 6B.7 (distance to infeasibility for sublinear mappings)**.** *Let $F : X \rightrightarrows Y$ be sublinear with closed graph and let $0 \in \operatorname{rge} F$. Then for the inclusion $F(x) \ni 0$,*

$$\textit{distance to infeasibility} = \inf_{\|y^*\|=1} d\big(0, F^{*+}(y^*)\big).$$

Proof. In this case the function h in 6B.6 has $h(x^*,y^*) = 0$ when $x^* \in F^{*+}(y^*)$, but $h(x^*,y^*) = \infty$ otherwise. □

In particular, for a linear-conic constraint system of type $x \ge_C 0, A(x) \le_K 0$, with respect to a continuous linear mapping $A : X \to Y$ and closed, convex cones $C \subset X$ and $K \subset Y$, we obtain

$$\text{distance to infeasibility} = \inf_{y^* \in K^+,\ \|y^*\|=1} d(A^*(y^*), C^+).$$

6C. Iterative Processes for Generalized Equations

Our occupation with numerical matters turns even more serious in this section, where we consider computational methods for solving generalized equations. The problem is to

$$\text{find } x \text{ such that } f(x)+F(x)\ni 0, \tag{1}$$

where $f : X \to Y$ is a continuously Fréchet differentiable function and $F : X \rightrightarrows Y$ is a set-valued mapping with closed graph; both X and Y are Banach spaces. As we already know, the model of a generalized equation covers huge territory. The classical case of nonlinear equations corresponds to having $F = 0$, whereas by taking $F \equiv -K$ for a fixed set K one gets various constraint systems. When F is the normal cone mapping N_C associated with a closed, convex set $C \subset X$, and $Y = X^*$, we have a variational inequality. In Section 6I we will deal with a generalized equation representing optimality conditions for an optimal control problem.

With the aim of approximating a solution to the generalized equation (1), we consider first the following version of *Newton's method*:

$$f(x_k)+Df(x_k)(x_{k+1}-x_k)+F(x_{k+1})\ni 0, \quad \text{for } k=0,1,\ldots. \tag{2}$$

This approach uses "partial linearization," in which we linearize f at the current point but leave F intact. It reduces to the standard version of Newton's method for solving the nonlinear equation $f(x) = 0$ when F is the zero mapping. We used this method to prove the classical inverse function theorem 1A.1. Although one might imagine that a "true" Newton-type method for (1) ought to involve some kind of approximation to F as well as f, such an extension runs into technical difficulties, in particular for infinite-dimensional variational problems.

In the case when (1) represents the optimality systems for a nonlinear programming problem, the iteration (2) becomes the popular sequential quadratic programming (SQP) algorithm for optimization. We will briefly describe the SQP algorithm later in the section.

We shall not discuss stopping criteria in this book; we assume that Newton's method (2), and any other iterative method considered later in the book, generates an infinite sequence $\{x_k\}$; if this sequence is convergent to a solution $\bar{x}$, then we say that the method is convergent. There are various modes of convergence; in this book we utilize the terms of *superlinear* and *quadratic* convergence for which it is convenient to use the following definitions. A sequence $\{x_k\}$ with $x_k \neq \bar{x}$ is *superlinearly convergent* to $\bar{x}$ when

$$\lim_{k\to\infty} \frac{\|x_{k+1}-\bar{x}\|}{\|x_k-\bar{x}\|} = 0.$$

A sequence $\{x_k\}$ is *quadratically convergent* to $\bar{x}$ when there exist $\gamma > 0$ and a natural k_0 such that

$$\|x_{k+1}-\bar{x}\| \le \gamma\|x_k-\bar{x}\|^2 \quad \text{for all } k \ge k_0.$$

In Theorem 6C.1 we show superlinear convergence of Newton's method (2) on the assumption of strong metric regularity of the mapping $f+F$ at a solution $\bar{x}$ of (1) for 0. Specifically, we show that when the starting point x_0 is chosen in a sufficiently small neighborhood O of the solution $\bar{x}$, then Newton's method (2) generates a unique in O sequence which converges superlinearly to $\bar{x}$. The proof of this convergence result is postponed until 6F where we present a much stronger version of it for a nonsmooth function f. In the case of an equation, that is, when F in (1) is the zero mapping, strong metric regularity is equivalent to the invertibility of the derivative mapping $Df(\bar{x})$, and we arrive at a well known result present in most textbooks on numerical analysis. In Section 6D we will show that when the derivative mapping Df is not only continuous but also Lipschitz continuous around $\bar{x}$, then the convergence is quadratic.

In Section 6D we focus on the stability properties of the set of sequences generated by the method under metric regularity. Then in Section 6E we discuss an approximate version of the Newton method (2) under strong metric subregularity. Section 6F is devoted to nonsmooth Newton's method in finite dimensions.

Theorem 6C.1 (superlinear convergence of Newton method)**.** *Consider Newton's method (2) applied to (1) and assume that the mapping $f+F$ is strongly metrically regular at $\bar{x}$ for 0. Then there exists a neighborhood O of $\bar{x}$ such that, for every $x_0 \in O$, there is a unique sequence $\{x_k\}$ generated by the method (2) all elements of which are contained in O; moreover, this sequence is superlinearly convergent to $\bar{x}$.*

Proof. Choose $\kappa > \operatorname{reg}(f+F;\bar{x}|0)$ and $\mu > 0$ such that $2\kappa\mu < 1$. From 2B.10 and later in the book we know that

$$\operatorname{reg}(f+F;\bar{x}|0) = \operatorname{reg}(f(\bar{x}) + Df(\bar{x})(\cdot - \bar{x}) + F(\cdot);\bar{x}|0).$$

Choose $a > 0$ and $b > 0$ so that the mapping

$$\mathbb{B}_b(0) \ni y \mapsto s(y) := (f(\bar{x}) + Df(\bar{x})(\cdot - \bar{x}) + F(\cdot))^{-1}(y) \cap \mathbb{B}_a(\bar{x})$$

is a Lipschitz continuous function with Lipschitz constant κ. Make $a > 0$ smaller if necessary to have

$$\|Df(x') - Df(x)\| \leq \mu \quad \text{for } x',\ x \in \mathbb{B}_a(\bar{x}) \tag{3}$$

and also

$$3\mu a \leq b. \tag{4}$$

From the standard equality

$$f(u) - f(v) = \int_0^1 Df(v + t(u-v))(u-v)dt,$$

we have through (3) that, for all $u, v \in \mathbb{B}_a(\bar{x})$,

$$\text{(5)} \qquad \begin{aligned} &\|f(u)-f(v)-Df(v)(u-v)\| \\ &\quad = \|\textstyle\int_0^1 Df(v+t(u-v))(u-v)dt - Df(v)(u-v)\| \leq \mu\|u-v\|. \end{aligned}$$

For the function

$$I\!B_a(\bar{x}) \times I\!B_a(\bar{x}) \ni (w,x) \mapsto g(w,x) = f(\bar{x}) + Df(\bar{x})(x-\bar{x}) - f(w) - Df(w)(x-w),$$

using (4) and (5) we get

$$\text{(6)} \qquad \begin{aligned} \|g(w,x)\| &\leq \|f(\bar{x}) + Df(\bar{x})(x-\bar{x}) - f(x)\| \\ &\quad + \|f(x)-f(w)-Df(w)(x-w)\| \\ &\leq \mu\|x-\bar{x}\| + \mu\|x-w\| \leq \mu a + \mu 2a \leq b. \end{aligned}$$

Pick $x_0 \in I\!B_a(\bar{x})$, $x_0 \neq \bar{x}$, and consider the mapping

$$I\!B_a(\bar{x}) \ni x \mapsto \Phi_0(x) = (f(\bar{x}) + Df(\bar{x})(\cdot - \bar{x}) + F(\cdot))^{-1}(g(x_0,x)) \cap I\!B_a(\bar{x}).$$

Let $c := \kappa\mu/(1-\kappa\mu)$ and $r_0 := c\|x_0 - \bar{x}\|$; then $c < 1$ and $r_0 < a$. Noting that $\Phi_0(\bar{x}) = s(g(x_0,\bar{x}))$ and $\bar{x} = s(0)$, and using the Lipschitz continuity of s together with (4), (5) and (6), we obtain

$$\begin{aligned} \|\bar{x} - \Phi_0(\bar{x})\| &= \|s(0) - s(g(x_0,\bar{x}))\| \leq \kappa\|g(x_0,\bar{x})\| \\ &\leq \kappa\|f(\bar{x}) - f(x_0) - Df(x_0)(\bar{x}-x_0)\| \leq \kappa\mu\|x_0-\bar{x}\| < r_0(1-\kappa\mu). \end{aligned}$$

Moreover, for any $u,v \in I\!B_{r_0}(\bar{x})$,

$$\begin{aligned} \|\Phi_0(u) - \Phi_0(v)\| &\leq \kappa\|g(x_0,u) - g(x_0,v)\| \\ &= \kappa\|(Df(\bar{x}) - Df(x_0))(u-v)\| \leq \kappa\mu\|u-v\|. \end{aligned}$$

Hence, by the standard contraction mapping principle 1A.2 there exists a unique $x_1 = \Phi_0(x_1) \cap I\!B_{r_0}(\bar{x})$, which translates to having x_1 obtained from x_0 as a first iterate of Newton's method (2) and satisfying

$$\|x_1 - \bar{x}\| \leq c\|x_0 - \bar{x}\|.$$

The induction step is completely analogous. Namely, for $r_k := c\|x_k - \bar{x}\|$ at the $(k+1)$-step we show that the function

$$I\!B_a(\bar{x}) \ni x \mapsto \Phi_k(x) = (f(\bar{x}) + Df(\bar{x})(\cdot - \bar{x}) + F(\cdot))^{-1}(g(x_k,x)) \cap I\!B_a(\bar{x})$$

has a unique fixed point $x_{k+1} = s(g(x_k,x_{k+1}))$ in $I\!B_{r_k}(\bar{x})$, which means that x_{k+1} is obtained from (2) and satisfies

$$\|x_{k+1} - \bar{x}\| \leq c\|x_k - \bar{x}\|.$$

Since $c < 1$, the sequence $\{x_k\}$ generated in this way is convergent to $\bar{x}$. Also note that this sequence is unique in $I\!B_a(\bar{x})$. Indeed, if for some k we obtain from x_k two different x_{k+1} and x'_{k+1}, then, since $x_{k+1} = s(g(x_k, x_{k+1}))$ and $x'_{k+1} = s(g(x_k, x'_{k+1}))$, from the Lipschitz continuity of s with constant κ and (3) we get

$$0 < \|x_{k+1} - x'_{k+1}\| \leq \kappa\|(Df(\bar{x}) - Df(x_k))(x_{k+1} - x'_{k+1})\| \leq \kappa\mu < \frac{1}{2}\|x_{k+1} - x'_{k+1}\|,$$

which is absurd.

To show superlinear convergence of the sequence $\{x_k\}$, take any $\varepsilon \in (0, 1/(2\kappa)$ and choose $\alpha \in (0,a)$ such that (3) holds with $\mu = \varepsilon$ for all $x \in \mathbb{B}_\alpha(\bar{x})$. Let k_0 be so large that $x_k \in \mathbb{B}_\alpha(\bar{x})$ for all $k \geq k_0$ and let $x_k \neq \bar{x}$ for all $k \geq k_0$. Since $x_{k+1} = s(g(x_k, x_{k+1}))$, isung the Lipschitz continuity of s and the first inequality in the third line of (6), we obtain

$$\begin{aligned}
\|x_{k+1} - \bar{x}\| &= \|s(g(x_k, x_{k+1})) - s(0)\| \leq \kappa\|g(x_k, x_{k+1}))\| \\
&\leq \kappa\varepsilon\|x_{k+1} - \bar{x}\| + \kappa\varepsilon\|x_{k+1} - x_k\| \\
&\leq 2\kappa\varepsilon\|x_{k+1} - \bar{x}\| + \kappa\varepsilon\|x_k - \bar{x}\|.
\end{aligned}$$

Hence,

$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} \leq \frac{\kappa\varepsilon}{1 - 2\kappa\varepsilon} \qquad \text{for all} \quad k \geq k_0.$$

Since the sequence $\{x_k\}$ does not depend on ε and ε can be arbitrarily small, passing to the limit with $k \to \infty$ we obtain superlinear convergence. □

It is important to note that the map F in (1) could be quite arbitrary, and then Newton's iteration (2) may generate in general more than one sequence. However, as 6C.1 says, there is only one such sequence which stays near the solution $\bar{x}$ and this sequence is moreover superlinearly convergent to $\bar{x}$. In other words, if the method generates two or more sequences, only one of them will converge to the solution $\bar{x}$ and the other sequences will be not convergent at all or at least not convergent to the same solution. This is in contrast with the case of an equation where, in the case when the Jacobian at the solution $\bar{x}$ is nonsingular, the method generates exactly one sequence when started close to $\bar{x}$ and this sequence converges superlinearly to $\bar{x}$.

As an application, consider the nonlinear programming problem studied in sections 2A and 2G:

$$\text{(7)} \qquad \text{minimize } g_0(x) \text{ over all } x \text{ satisfying } g_i(x) \begin{cases} \leq 0 & \text{for } i \in [1,s], \\ = 0 & \text{for } i \in [s+1,m], \end{cases}$$

where the functions $g_i : \mathbb{R}^n \to \mathbb{R}$ are twice continuously differentiable. In terms of the Lagrangian function

$$L(x,y) = g_0(x) + y_1 g_1(x) + \cdots + y_m g_m(x),$$

the first-order optimality (Karush–Kuhn–Tucker) condition takes the form

$$\text{(8)} \qquad \begin{cases} \nabla_x L(x,y) = 0, \\ g(x) \in N_{\mathbb{R}^s_+ \times \mathbb{R}^{m-s}}(y), \end{cases}$$

where we denote by $g(x)$ the vector with components $g_1(x), \ldots, g_m(x)$. Let $\bar{x}$ be a local minimum for (7) satisfying the standard constraint qualification condition 2A.9(14), and let $\bar{y}$ be an associated Lagrange multiplier vector. As applied to the variational inequality (8), Newton's method (2) consists in generating a sequence $\{(x_k, y_k)\}$ starting from a point (x_0, y_0), close enough to $(\bar{x}, \bar{y})$, according to the iteration

$$\text{(9)} \qquad \begin{cases} \nabla_x L(x_k, y_k) + \nabla^2_{xx} L(x_k, y_k)(x_{k+1} - x_k) + \nabla g(x_k)^{\mathsf{T}}(y_{k+1} - y_k) = 0, \\ g(x_k) + \nabla g(x_k)(x_{k+1} - x_k) \in N_{\mathbb{R}^s_+ \times \mathbb{R}^{m-s}}(y_{k+1}). \end{cases}$$

Theorem 2G.8 (with suppressed dependence on the parameter p in its statement) provides conditions under which the mapping appearing in the variational inequality (8) is strongly metrically regular at the reference point: linear independence of the gradients of the active constraints and strong second-order sufficient optimality; we recall these conditions in the next Example 6C.2. Under these conditions, we can find (x_{k+1}, y_{k+1}) which satisfies (9) by solving the quadratic programming problem

$$\text{(10)} \qquad \begin{aligned} &\text{minimize } \Big[\frac{1}{2} \langle x - x_k, \nabla^2_{xx} L(x_k, y_k)(x - x_k) \rangle \\ &\qquad\qquad + \langle \nabla_x L(x_k, y_k) - \nabla g(x_k)^{\mathsf{T}} y_k, (x - x_k) \rangle \Big] \\ &\text{subject to } \; g_i(x_k) + \nabla g_i(x_k)(x - x_k) \begin{cases} \le 0 & \text{for } i \in [1, s], \\ = 0 & \text{for } i \in [s+1, m]. \end{cases} \end{aligned}$$

Thus, in the circumstances of (7) under strong metric regularity of the mapping in (8), Newton's method (2) comes down to sequentially solving quadratic programs of the form (10). This specific application of Newton's method is therefore called the *sequential quadratic programming* (SQP) method.

We summarize the conclusions obtained so far about the SQP method as an illustration of the power of the theory developed in this section.

Example 6C.2 (convergence of SQP)**.** Consider the nonlinear programming problem (7) with the associated Karush–Kuhn–Tucker condition (8) and let $\bar{x}$ be a solution with an associated Lagrange multiplier vector $\bar{y}$. In the notation

$$\begin{aligned} I &= \{ i \in [1, m] \mid g_i(\bar{x}) = 0 \} \supset \{s+1, \ldots, m\}, \\ I_0 &= \{ i \in [1, s] \mid g_i(\bar{x}) = 0 \text{ and } \bar{y}_i = 0 \} \subset I \end{aligned}$$

and

$$\begin{aligned} M^+ &= \{ w \in \mathbb{R}^n \mid w \perp \nabla_x g_i(\bar{x}) \text{ for all } i \in I \backslash I_0 \}, \\ M^- &= \{ w \in \mathbb{R}^n \mid w \perp \nabla_x g_i(\bar{x}) \text{ for all } i \in I \}, \end{aligned}$$

suppose that the following conditions are both fulfilled:

(a) the gradients $\nabla_x g_i(\bar{x})$ for $i \in I$ are linearly independent;

(b) $\langle w, \nabla^2_{xx} L(\bar{x},\bar{y})w\rangle > 0$ for every nonzero $w \in M^+$ with $\nabla^2_{xx} L(\bar{x},\bar{y})w \perp M^-$.

Then there exists a neighborhood O of $(\bar{x},\bar{y})$ such that, for every starting point $(x_0,y_0) \in O$, the SQP method (9) generates a unique in O sequence which converges superlinearly to $(\bar{x},\bar{y})$.

There are various numerical issues related to implementation of the SQP method that have been investigated in the last several decades, and various enhancements are available as commercial software, but we shall not go into this further.

We consider next the following version of the *proximal point method* applied to the generalized equation (1):

$$\lambda_k(x_{k+1}-x_k) + f(x_{k+1}) + F(x_{k+1}) \ni 0, \quad \text{for } k = 0,1,\ldots, \tag{11}$$

where λ_k is a sequence of positive numbers convergent to zero. From now on $X = Y$. First, we prove superlinear convergence of this method under strong metric regularity.

Theorem 6C.3 (convergence of proximal point method). *Consider the generalized equation (1) where the function f is continuous and F has closed graph. Assume that the mapping $f+F$ is strongly metrically regular at $\bar{x}$ for 0 and let μ be a positive scalar such that*

$$2\mu \operatorname{reg}(f+F;\bar{x}\,|\,0) < 1. \tag{12}$$

Consider the proximal point method (11) for a sequence λ_k convergent to zero and such that $\lambda_k \le \mu$ for all k. Then there exists a neighborhood O such that for every $x_0 \in O$ the method (11) generates a unique in O sequence $\{x_k\}$; moreover, this sequence is superlinearly convergent to $\bar{x}$.

Proof. Let $\kappa > \operatorname{reg}(f+F;\bar{x}\,|\,0)$ be such that, by (12), $2\mu\kappa < 1$. Then there exist $a > 0$ and $b > 0$ such that the mapping

$$I\!B_b(0) \ni y \mapsto s(y) := (f+F)^{-1}(y) \cap I\!B_a(\bar{x})$$

is a Lipschitz continuous function with Lipschitz constant κ, and this remains true if we replace a by any positive $a' \le a$. Make a smaller if necessary so that $2a\mu \le b$. Then for any $x, x' \in I\!B_a(\bar{x})$ we have

$$\|\lambda_k(x-x')\| \le 2\mu a \le b. \tag{13}$$

Pick $x_0 \in I\!B_a(\bar{x})$, $x_0 \ne \bar{x}$ and consider the mapping

$$I\!B_a(\bar{x}) \ni x \mapsto \Phi_0(x) = (f+F)^{-1}(-\lambda_0(x-x_0)) \cap I\!B_a(\bar{x}) = s(-\lambda_0(x-x_0)).$$

Since $\bar{x} = s(0)$ and $\Phi_0(\bar{x}) = s(-\lambda_0(\bar{x}-x_0))$, from (13) we have

$$\|\bar{x} - \Phi_0(\bar{x})\| = \|s(0) - s(-\lambda_0(\bar{x}-x_0))\| \le \kappa\lambda_0\|\bar{x}-x_0\| \le r_0(1-\kappa\lambda_0),$$

where

$$r_0 = \frac{\kappa\lambda_0}{1-\kappa\lambda_0}\|x_0 - \bar{x}\|.$$

Noting that $r_0 \leq a$ and using (13), for any $u, v \in \mathbb{B}_{r_0}(\bar{x})$ we obtain

$$\|\Phi_0(u) - \Phi_0(v)\| = \|s(-\lambda_0(u-x_0)) - s(-\lambda_0(v-x_0))\| \leq \kappa\lambda_0\|u-v\|.$$

Hence, there exists a unique fixed point $x_1 = \Phi_0(x_1)$ in $\mathbb{B}_{r_0}(\bar{x})$, i.e., x_1 satisfies (11) for $k = 1$ and also

$$\|x_1 - \bar{x}\| \leq \frac{\kappa\lambda_0}{1-\kappa\lambda_0}\|x_0 - \bar{x}\|.$$

Clearly $x_1 \in \mathbb{B}_a(\bar{x})$. It turns out that x_1 is the only iterate in $\mathbb{B}_a(\bar{x})$. Assume on the contrary that it is not; that is, there exists $x_1' = \Phi_0(x_1') \cap \mathbb{B}_a(\bar{x})$. Then, since $x_1 = s(-\lambda_0(x_1 - x_0))$ and $x_1' = s(-\lambda_0(x_1' - x_0))$ we get

$$\|x_1 - x_1'\| = \|s(-\lambda_0(x_1-x_0)) - s(-\lambda_0(x_1'-x_0))\| \leq \kappa\lambda_0\|x_1 - x_1'\| < \|x_1 - x_1'\|,$$

a contradiction.

If $x_1 = \bar{x}$ there is nothing more to prove. Assume $x_1 \neq \bar{x}$ and consider the mapping

$$\mathbb{B}_a(\bar{x}) \ni x \mapsto \Phi_1(x) = s(-\lambda_1(x - x_1)).$$

Then

$$\|\bar{x} - \Phi_1(\bar{x})\| = \|s(0) - s(-\lambda_1(\bar{x} - x_1))\| \leq \kappa\lambda_1\|\bar{x} - x_1\| \leq r_1(1-\kappa\lambda_1),$$

where

$$r_1 := \frac{\kappa\lambda_1}{1-\kappa\lambda_1}\|x_1 - \bar{x}\|.$$

Also, for any $u, v \in \mathbb{B}_{r_1}(\bar{x})$ we have

$$\|\Phi_1(u) - \Phi_1(v)\| = \|s(-\lambda_1(u-x_1)) - s(-\lambda_1(v-x_1))\| \leq \kappa\lambda_1\|u-v\|.$$

Therefore, there exists a unique $x_2 = \Phi_1(x_2)$ in $\mathbb{B}_{r_1}(\bar{x})$ which turns out to be unique also in $\mathbb{B}_a(\bar{x})$, that is, x_2 satisfies (11) for $k = 2$ and also

$$\|x_2 - \bar{x}\| \leq \frac{\kappa\lambda_1}{1-\kappa\lambda_1}\|x_1 - \bar{x}\|.$$

The induction step is completely analogous. Thus, we obtain a sequence $\{x_k\}$ satisfying (11) the elements of which are unique in $O = \mathbb{B}_a(\bar{x})$ and such that

$$\|x_{k+1} - \bar{x}\| \leq \frac{\kappa\lambda_k}{1-\kappa\lambda_k}\|x_k - \bar{x}\|.$$

If $x_k \neq \bar{x}$ for all k, since $\lambda_k \searrow 0$, this sequence is superlinearly convergent to $\bar{x}$. $\square$

Lastly, we will discuss a bit more the proximal point method in the context of monotone mappings. First, note that the iterative process (11) can be equally well written as

$$(14) \qquad x_{k+1} \in [I + \lambda_k^{-1} T]^{-1}(x_k) \quad \text{for } k = 0, 1, \ldots, \text{ where } T = f + F.$$

It has been extensively studied under the additional assumption that X is a Hilbert space (e.g., consider $I\!\!R^n$ under the Euclidean norm) and T is a *maximal monotone* mapping from X to X. Monotonicity, which we introduced in 1H for functions, and a local version of it in 3G for set-valued mappings, refers here in the case of a set-valued mapping $T : X \rightrightarrows X$ to the property of having

$$(15) \qquad \langle y' - y, x' - x \rangle \geq 0 \quad \text{for all } (x,y), (x',y') \in \operatorname{gph} T.$$

It is called maximal when no more points can be added to $\operatorname{gph} T$ without running into a violation of (15). (A localized monotonicity for set-valued mappings was introduced at the end of 3G, but again only in finite dimensions.)

The following fact about maximal monotone mappings, recalled here without its proof, underlies much of the literature on the proximal point method in basic form and indicates its fixed-point motivation.

Theorem 6C.4 (resolvents of maximal monotone mappings)**.** *Let X be a Hilbert space, and let $T : X \rightrightarrows X$ be maximal monotone. Then for any $c > 0$ the mapping $P_c = (I + cT)^{-1}$ is single-valued with all of X as its domain and moreover is non-expansive; in other words, it is globally Lipschitz continuous from X into X with Lipschitz constant 1. The fixed points of P_c are the points $\bar{x}$ such that $T(\bar{x}) \ni 0$ (if any), and they form a closed, convex set.*

According to this, x_{k+1} always exists and is uniquely determined from x_k in the proximal point iterations (11) as expressed in (14), when $f + F$ is maximal monotone. Here is an important example of that circumstance, which we again state without bringing out its proof:

Theorem 6C.5 (maximal monotonicity in a variational inequality)**.** *Let X be a Hilbert space, and let $F = N_C$ for a nonempty, closed, convex set C in X. Let $f : C \to X$ be continuous and monotone. Then $f + F$ is maximal monotone.*

The "proximal point" terminology comes out of this framework through an application to optimization, as now explained. For a real-valued function g on a Hilbert space X with derivative $Dg(x)$, we denote by $\nabla g(x)$, as in the case of $X = I\!\!R^n$, the unique element of X such that $Dg(x)w = \langle \nabla g(x), w \rangle$ for all $w \in X$.

Example 6C.6 (connections with minimization)**.** Let X be a Hilbert space, let C be a nonempty, closed, convex subset of X, and let $h : X \to I\!\!R$ be convex and continuously (Fréchet) differentiable. Let $f(x) = \nabla h(x)$. Then f is continuous and monotone, and the variational inequality

$$f(x) + N_C(x) \ni 0,$$

as an instance of the generalized equation (1), describes the points x (if any) which minimize h over C. In comparison, in the iterations for this case of the proximal point method in the basic form (11), the point x_{k+1} determined from x_k is the unique minimizer of $h(x)+(\lambda_k/2)||x-x_k||^2$ over C.

Detail. This invokes the gradient monotonicity property associated with convexity in 2A.6, along with the optimality condition in 2A.7, both of which are equally valid in infinite dimensions. The addition of the quadratic expression $(\lambda_k/2)||x-x_k||^2$ to h creates a function h_k which is strongly convex with constant λ_k and thus attains its minimum, moreover uniquely. □

The expression $(\lambda_k/2)||x-x_k||^2$ in 6C.6 is called a *proximal term* because it helps to keep x near to the current point x_k. Its effect is to stabilize the procedure while inducing technically desirable properties like strong convexity in place of plain convexity. It's from this that the algorithm got its name.

Instead of adding a quadratic term to h, the strategy in Example 6C.6 could be generalized to adding a term $r_k(x-x_k)$ for some other convex function r_k having its minimum at the origin, and adjusting the algorithm accordingly.

6D. Metric Regularity of Newton's Iteration

In this section we put Newton's iteration in the context of metric regularity, by considering the parameterized generalized equation

$$(1) \qquad f(p,x)+F(x)\ni 0, \text{ or equivalently } -f(p,x)\in F(x),$$

for a function $f:P\times X\to Y$ and a set-valued mapping $F:X\rightrightarrows Y$, where $p\in P$ is a parameter and P, X and Y are Banach spaces. Associated with the generalized equation (1) as usual is its solution mapping

$$S: p\mapsto \left\{x \,\middle|\, f(p,x)+F(x)\ni 0\right\} \quad \text{for } p\in P.$$

We focus on a neighborhood of a given reference solution $\bar{x}$ of (1) for $\bar{p}$. In this section our standing assumption is that the function f is continuously differentiable in a neighborhood of $(\bar{p},\bar{x})$ with strict partial derivatives denoted by $D_xf(\bar{p},\bar{x})$ and $D_pf(\bar{p},\bar{x})$ such that

$$\operatorname{lip}(D_xf;(\bar{p},\bar{x}))<\infty,$$

and that the mapping F has closed graph.

As in the preceding section, we consider the following Newton's method for solving (1), now for a fixed value of the parameter p:

$$(2) \qquad f(p,x_k)+D_xf(p,x_k)(x_{k+1}-x_k)+F(x_{k+1})\ni 0, \quad \text{for } k=0,1,\ldots,$$

with a given starting point x_0. Under strong metric regularity of the mapping $f(p,\cdot)+F(\cdot)$ we proved there that when the starting point is in a neighborhood of the reference solution, the method generates a sequence which is unique in this neighborhood and converges quadratically to the reference solution. In this section we focus on the case when the underlying mapping is merely metrically regular.

We start with a technical result which follows from 5G.1 and 5G.2.

Theorem 6D.1 (parametrized linearization)**.** *Consider the parameterized mapping*

$$\text{(3)}\quad X \ni x \mapsto G_{p,u}(x) = f(p,u) + D_x f(p,u)(x-u) + F(x), \qquad \text{for } p \in P,\ u \in X,$$

and suppose that the mapping

$$\text{(4)}\qquad x \mapsto G(x) = f(\bar{p},\bar{x}) + D_x f(\bar{p},\bar{x})(x-\bar{x}) + F(x)$$

is metrically regular at $\bar{x}$ for 0. Then for every $\lambda > \operatorname{reg}(G;\bar{x}|0)$ there exist positive numbers a, b and c such that

$$d(x, G_{p,u}^{-1}(y)) \le \lambda d(y, G_{p,u}(x)) \quad \text{for every } u, x \in \mathbb{B}_a(\bar{x}),\ y \in \mathbb{B}_b(0),\ p \in \mathbb{B}_c(\bar{p}).$$

If in addition the mapping G is strongly metrically regular, then for every $\lambda > \operatorname{reg}(G;\bar{x}|0)$ there exist positive numbers a, b and c such that for every $u \in \mathbb{B}_a(\bar{x})$ and $p \in \mathbb{B}_c(\bar{p})$ the mapping $y \mapsto G_{p,u}^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ is a Lipschitz continuous function on $\mathbb{B}_b(0)$ with a Lipschitz constant λ.

Proof. To prove the first claim, we apply Theorem 5G.1 with the following specifications: $F(x) = G(x)$, $\bar{y} = 0$, $q = (p,u)$, $\bar{q} = (\bar{p},\bar{x})$, and

$$g(q,x) = f(p,u) + D_x f(p,u)(x-u) - f(\bar{p},\bar{x}) - D_x f(\bar{p},\bar{x})(x-\bar{x}).$$

Let $\lambda > \kappa \ge \operatorname{reg}(G;\bar{x}|0)$. Pick any $\mu > 0$ such that $\mu\kappa < 1$ and $\lambda > \kappa/(1-\kappa\mu)$. Choose a positive L such that

$$L > \max\{\widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x})), \widehat{\operatorname{lip}}_x(D_x f;(\bar{p},\bar{x}))\}.$$

Then there exist positive constants α and β such that

$$\text{(5)}\qquad \|f(p,x) - f(p',x)\| \le L\|p-p'\| \quad \text{for every } p,p' \in \mathbb{B}_\beta(\bar{p}),\quad x \in \mathbb{B}_\alpha(\bar{x}),$$

$$\text{(6)}\quad \|D_x f(p,x) - D_x f(p,x')\| \le L\|x-x'\| \quad \text{for every } x,x' \in \mathbb{B}_\alpha(\bar{x}),\quad p \in \mathbb{B}_\beta(\bar{p}),$$

and

$$\text{(7)}\qquad \|D_x f(p,u) - D_x f(\bar{p},\bar{x})\| \le \mu \quad \text{for every } p \in \mathbb{B}_\beta(\bar{p}),\quad u \in \mathbb{B}_\alpha(\bar{x}).$$

Observe that for any $x,x' \in X$ and any $q = (p,u) \in \mathbb{B}_\beta(\bar{p}) \times \mathbb{B}_\alpha(\bar{x})$, from (7),

$$\|g(q,x) - g(q,x')\| \le \|D_x f(p,u) - D_x f(\bar{p},\bar{x})\|\|x-x'\| \le \mu\|x-x'\|,$$

that is, $\widehat{\mathrm{lip}}_x(g;(\bar{q},\bar{x})) \le \mu$. Thus, the assumptions of Theorem 5G.1 are satisfied and hence there exist positive constants $a' \le \alpha$, b' and $c' \le \beta$ such that for any $q \in \mathbb{B}_{c'}(\bar{p}) \times \mathbb{B}_{a'}(\bar{x})$ the mapping $G_{p,u}(x) = g(q,x) + G(x)$ is metrically regular at $\bar{x}$ for $g(q,\bar{x}) = f(p,u) + D_x f(p,u)(\bar{x}-u) - f(\bar{p},\bar{x})$ with constant λ and neighborhoods $\mathbb{B}_{a'}(\bar{x})$ and $\mathbb{B}_{b'}(g(q,\bar{x}))$. Now choose positive scalars a, b and c such that

$$a \le a', \quad c \le c', \quad \text{and} \quad La^2/2 + Lc + b \le b'.$$

Fix any $q = (p,u) \in \mathbb{B}_c(\bar{p}) \times \mathbb{B}_a(\bar{x})$. Using (6) in the standard estimation

$$\begin{aligned}(8) \qquad &\|f(p,u) + D_x f(p,u)(\bar{x}-u) - f(p,\bar{x})\| \\ &\quad = \left\| \int_0^1 D_x f(p,\bar{x}+t(u-\bar{x}))(u-\bar{x})dt - D_x f(p,u)(u-\bar{x}) \right\| \\ &\quad \le L \int_0^1 (1-t)dt \, \|u-\bar{x}\|^2 = \tfrac{L}{2}\|u-\bar{x}\|^2,\end{aligned}$$

and applying (5) and (8), we obtain that, for $y \in \mathbb{B}_b(0)$,

$$\begin{aligned}\|g(q,\bar{x}) - y\| &\le \|f(p,u) + D_x f(p,u)(\bar{x}-u) - f(\bar{p},\bar{x})\| + \|y\| \\ &\le \|f(p,u) + D_x f(p,u)(\bar{x}-u) - f(p,\bar{x})\| + \|f(p,\bar{x}) - f(\bar{p},\bar{x})\| + \|y\| \\ &\le \frac{L}{2}\|u-\bar{x}\|^2 + L\|p-\bar{p}\| + b \le La^2/2 + Lc + b \le b'.\end{aligned}$$

Thus, $\mathbb{B}_b(0) \subset \mathbb{B}_{b'}(g(q,\bar{x}))$ and the proof of the first part of the theorem is complete. The proof of the second part regarding strong metric regularity is completely analogous, using Theorem 5G.2 instead of Theorem 5G.1. □

Our first result reveals quadratic convergence, uniform in the parameter, under metric regularity.

Theorem 6D.2 (quadratic convergence of Newton method). *Suppose that the mapping G defined in (4) is metrically regular at $\bar{x}$ for 0. Then for every*

$$(9) \qquad \gamma > \frac{1}{2}\,\mathrm{reg}\,(G;\bar{x}|0) \cdot \widehat{\mathrm{lip}}_x(D_x f;(\bar{p},\bar{x}))$$

there are positive constants $\bar{a}$ and $\bar{c}$ such that for every $p \in \mathbb{B}_{\bar{c}}(\bar{p})$ the set $S(p) \cap \mathbb{B}_{\bar{a}/2}(\bar{x})$ is nonempty and for every $\sigma \in S(p) \cap \mathbb{B}_{\bar{a}/2}(\bar{x})$ and every $x_0 \in \mathbb{B}_{\bar{a}}(\bar{x})$ there exists a sequence $\{x_k\}$ in $\mathbb{B}_{\bar{a}}(\bar{x})$ satisfying (2) for p with starting point x_0 and this sequence converges to σ; moreover, the convergence is quadratic, namely

$$(10) \qquad \|x_{k+1} - \sigma\| \le \gamma \|x_k - \sigma\|^2 \quad \text{for all} \quad k = 0,1,\ldots.$$

If the mapping G in (4) is not only metrically regular but also strongly metrically regular at $\bar{x}$ for 0, which is the same as having $f+F$ strongly metrically regular at $\bar{x}$ for 0, then there exist neighborhoods Q of $\bar{p}$ and U of $\bar{x}$ such that, for every $p \in Q$ and $u \in U$, there is exactly one sequence $\{x_k\}$ in U generated by Newton's iteration (2) for p with x_0 as a starting point. This sequence converges to the value $s(p)$ of *the Lipschitz continuous localization s of the solution mapping S around $\bar{p}$ for $\bar{x}$; moreover, the convergence is quadratic with constant γ, as in (10).*

Proof. Choose γ as in (9) and let $\lambda > \operatorname{reg}(G;\bar{x}\,|\,0)$ and $L > \widehat{\operatorname{lip}}_x(D_x f;(\bar{p},\bar{x}))$ be such that

$$\gamma > \frac{1}{2}\lambda L.$$

According to 6D.1 there exist positive a and c such that

$$d(x, G_{p,u}^{-1}(0)) \le \lambda d(0, G_{p,u}(x)) \quad \text{for every } u, x \in \mathbb{B}_a(\bar{x}),\ p \in \mathbb{B}_c(\bar{p}).$$

The Aubin property of the mapping S established in Theorem 5E.5 implies that for every $d > \operatorname{lip}(S;\bar{p}\,|\,\bar{x})$ there exists $c' > 0$ such that $\bar{x} \in S(p) + d\|p - \bar{p}\|\mathbb{B}$ for any $p \in \mathbb{B}_{c'}(\bar{p})$. Then

$$S(p) \cap \mathbb{B}_{d\|p-\bar{p}\|}(\bar{x}) \neq \emptyset \text{ for } p \in \mathbb{B}_{c'}(\bar{p}).$$

We choose next positive constants $\bar{a}$ and $\bar{c}$ such that the following inequalities are satisfied:

$$\text{(11)} \qquad \bar{a} < a, \quad \bar{c} < \min\{\frac{\bar{a}}{2d}, c, c'\} \ \text{ and } \ \frac{9}{2}\gamma\bar{a} \le 1.$$

Then, for every $p \in \mathbb{B}_{\bar{c}}(\bar{p})$ the set $S(p) \cap \mathbb{B}_{\bar{a}/2}(\bar{x})$ is nonempty. Moreover, for every $s \in S(p) \cap \mathbb{B}_{\bar{a}/2}(\bar{x})$ and $u \in \mathbb{B}_{\bar{a}}(\bar{x})$ we have

$$\text{(12)} \qquad d(s, G_{p,u}^{-1}(0)) \le \lambda d(0, G_{p,u}(s)).$$

Fix arbitrary $p \in \mathbb{B}_{\bar{c}}(\bar{p})$, $\sigma \in S(p) \cap \mathbb{B}_{\bar{a}/2}(\bar{x})$ and $u \in \mathbb{B}_{\bar{a}}(\bar{x})$. In the next lines we will show the existence of x_1 such that

$$\text{(13)} \qquad G_{p,u}(x_1) \ni 0, \quad \|x_1 - \sigma\| \le \gamma\|u - \sigma\|^2 \ \text{ and } \ x_1 \in \mathbb{B}_{\bar{a}}(\bar{x}).$$

If $d(0, G_{p,u}(\sigma)) = 0$ we set $x_1 = \sigma$. Since F is closed-valued, (12) implies the first relation in (13), while the second one is obvious and the third one follows from the fact that $\sigma \in \mathbb{B}_{\bar{a}/2}(\bar{x})$.

If $d(0, G_{p,u}(\sigma)) > 0$, then

$$d(\sigma, G_{p,u}^{-1}(0)) \le \lambda d(0, G_{p,u}(\sigma)) < \frac{2\gamma}{L} d(0, G_{p,u}(\sigma)),$$

and hence there exists $x_1 \in G_{p,u}^{-1}(0)$ such that

$$\text{(14)} \qquad \|\sigma - x_1\| < \frac{2\gamma}{L} d(0, G_{p,u}(\sigma)).$$

Since $f(p,\sigma) + F(\sigma) \ni 0$, using the estimate (8) with $\bar{x}$ replaced by σ we obtain

$$d(0, G_{p,u}(\sigma)) \le \|f(p,u) + D_x f(p,u)(\sigma - u) - f(p,\sigma)\| \le \frac{L}{2}\|u - \sigma\|^2.$$

Then (14) implies the inequality in (13). To complete the proof of (13) we estimate

$$\|x_1 - \bar{x}\| \le \|x_1 - \sigma\| + \|\sigma - \bar{x}\| < \gamma\|u - \sigma\|^2 + \bar{a}/2 \le \gamma(3\bar{a}/2)^2 + \bar{a}/2 < \bar{a},$$

where we use (11).

Due to the inequality in (13), the same argument can be applied with $u = x_1$, to obtain the existence of x_2 such that $\|x_2 - \sigma\| \le \gamma\|x_1 - \sigma\|^2$, and in the same way we get the existence of a sequence $\{x_k\}$ satisfying (10) for all k.

We will now show that this sequence is convergent. From the third inequality in (11),

$$\theta := \gamma\|x_0 - \sigma\| \le \gamma(\|x_0 - \bar{x}\| + \|\sigma - \bar{x}\|) \le \gamma(\bar{a} + \bar{a}/2) < 1,$$

and then from (10) we obtain

$$\|x_{k+1} - \sigma\| \le \theta^{2^{k+1}-1}\|x_0 - \sigma\|. \tag{15}$$

Thus the sequence $\{x_k\}$ is convergent to σ. The quadratic rate of convergence is already established.

The second part of the proof is completely analogous, using the strong regularity of the mapping $G_{p,u}$ claimed in the second part of 6D.1. □

We now present a Lyusternik–Graves type theorem connecting metric regularity of the linearized mapping (4) with metric regularity of a mapping whose values are the sets of all convergent sequences generated by Newton's method (2). This result shows that Newton's iteration is, roughly, as "stable" as the mapping of the inclusion to be solved. This conclusion may have important implications in the analysis of the effect of various errors, including the errors of approximations of the problem in hand, on the complexity of the method.

Specifically, we consider the method (2) more broadly by reconceiving Newton's iteration as an inclusion, the solution of which gives a *whole sequence* instead of just an element in X. Let $l_\infty(X)$ be the Banach space consisting of all infinite sequences $\xi = \{x_1, x_2, \ldots, x_k, \ldots\}$ with elements $x_k \in X$, $k = 1, 2, \ldots$ equipped with the supremum norm

$$\|\xi\|_\infty = \sup_{k\ge 1}\|x_k\|$$

and let $l_\infty^c(X)$ be the closed subspace of $l_\infty(X)$ consisting of all infinite sequences $\xi = \{x_1, x_2, \ldots, x_k, \ldots\}$ that are *convergent*. Define the mapping $\Xi : X \times P \rightrightarrows l_\infty^c(X)$

$$\begin{aligned}\Xi : (p,u) \mapsto \Big\{\xi = \{x_1, x_2, \ldots\} \in l_\infty^c(X) \,\Big|\, \xi \ \text{ is such that} \\ f(p,x_k) + D_x f(p,x_k)(x_{k+1} - x_k) + F(x_{k+1}) \ni 0 \\ \text{for } \ k = 0, 1, \ldots, \ \text{with } x_0 = u\Big\}.\end{aligned} \tag{16}$$

That is, for a given (p,u) the value of $\Xi(p,u)$ is the set of all convergent sequences $\{x_k\}_{k=1}^\infty$ generated by Newton's iteration (2) for p that start from u. If $\bar{x}$ is a solution to (1) for $\bar{p}$, then the constant sequence $\bar{\xi}$ whose components are all $\bar{x}$ satisfies $\bar{\xi} \in \Xi(\bar{p},\bar{x})$. Further, note that if $s \in S(p)$ then the constant sequence whose components are all s belongs to $\Xi(p,s)$. Also note that if $\xi \in \Xi(p,u)$ for some (p,u) close enough to $(\bar{p},\bar{x})$, then by definition ξ is convergent and since F has closed graph the limit of ξ is from $S(p)$, that is a solution of (1) for p.

By using the parameterized mapping $G_{p,u}(x)$ in (3) we can define equivalently Ξ as

$$\Xi : (p,u) \mapsto \left\{ \xi \in l_\infty^c(X) \,\middle|\, x_0 = u \text{ and } G_{p,x_k}(x_{k+1}) \ni 0 \ \text{ for } k = 0,1,\ldots \right\}.$$

Along with the mapping Ξ defined in (16) we introduce a mapping $\widehat{\Xi}$ in exactly the same way but with $l_\infty^c(X)$ replaced by $l_\infty(X)$. Clearly, $\Xi(p,u) \subset \widehat{\Xi}(p,u)$ for all $(p,u) \in P \times X$. Theorem 6D.2 says that for each constant γ satisfying (9) there are neighborhood Q of $\bar{p}$ and O of $\bar{x}$ such that the mapping $\widehat{\Xi}$ has a *selection* $Q \times O \ni p \mapsto \xi(p,u) \in \Xi(p,u)$ with the property that for each $(p,u) \in Q \times O$ the sequence $\xi(p,u)$ is quadratically convergent to a solution $s \in S(p)$, hence it is an element of $\Xi(p,u)$. If the mapping G is *strongly metrically regular* at $\bar{x}$ for 0, then this selection is locally unique, that is, locally, there is exactly one sequence generated by Newton's method. Specifically, in the case of strong metric regularity the mapping $\widehat{\Xi}$ has a single-valued graphical localization ξ around $(\bar{p},\bar{x})$ for $\bar{\xi}$; moreover, this localization has property that for u close to $\bar{x}$ and p close to $\bar{p}$ the value $\xi(p,u)$ of the localization is a sequence which converges quadratically to the associated solution $s(p)$ for p. Hence, the mapping $\widehat{\Xi}$ locally coincides with the mapping Ξ and both mappings are one and the same function whose values are quadratically convergent sequences.

To proceed we need the following lemma, which extends Theorem 6D.1.

Lemma 6D.3. *Suppose that the mapping G defined in (4) is metrically regular at $\bar{x}$ for 0 and let γ, γ_1 and γ_2 be positive constants such that*

$$\gamma > \frac{1}{2}\operatorname{reg}(G;\bar{x}\,|\,0)\cdot\widehat{\operatorname{lip}}_x(D_x f;(\bar{p},\bar{x})),\ \gamma_1 > \operatorname{reg}(G;\bar{x}\,|\,0)\cdot\|D_p f(\bar{p},\bar{x})\|,$$

and

$$\gamma_2 > \operatorname{reg}(G;\bar{x}\,|\,0)\cdot\operatorname{lip}(D_x f;(\bar{p},\bar{x})).$$

Then there exist positive ζ and α such that for every $p,\, p' \in \mathbb{B}_\zeta(\bar{p})$, $u,\, u' \in \mathbb{B}_\alpha(\bar{x})$ and $x \in G_{p,u}^{-1}(0) \cap \mathbb{B}_\alpha(\bar{x})$ there exists $x' \in G_{p',u'}^{-1}(0)$ satisfying

$$\|x - x'\| \le \gamma\|u - u'\|^2 + \gamma_1\|p - p'\| + \gamma_2(\|p - p'\| + \|u - u'\|)\|x - u\|. \tag{17}$$

Proof. Let $\lambda' > \lambda > \operatorname{reg}(G;\bar{x}\,|\,0)$, $L > \widehat{\operatorname{lip}}_x(D_x f;(\bar{p},\bar{x}))$, $L_1 > \|D_p f(\bar{p},\bar{x})\| = \widehat{\operatorname{lip}}_p(f;(\bar{p},\bar{x}))$ and $L_2 > \operatorname{lip}(D_x f;(\bar{p},\bar{x}))$ be such that

$$\gamma > \frac{\lambda'}{2}L, \quad \gamma_1 > \lambda' L_1, \quad \gamma_2 > \lambda' L_2.$$

Now we chose positive α and ζ which are smaller than the numbers a and c in the claim of 6D.1 and such that $D_x f$ is Lipschitz with respect to $x \in \mathbb{B}_\alpha(\bar{x})$ with constant L uniformly in $p \in \mathbb{B}_\zeta(\bar{p})$, f is Lipschitz with constant L_1 with respect to $p \in \mathbb{B}_\zeta(\bar{p})$ uniformly in $x \in \mathbb{B}_\alpha(\bar{x})$, and D_x is Lipschitz with constant L_2 in $\mathbb{B}_\zeta(\bar{p}) \times \mathbb{B}_\alpha(\bar{x})$.

Let p, p', u, u', x be as in the statement of the lemma. If $d(0, G_{p',u'}(x)) = 0$, by the closedness of $G^{-1}_{p',u'}(0)$ we obtain that $x \in G^{-1}_{p',u'}(0)$ and there is nothing more to prove. If not, then, from 6D.1 we get

$$d(x, G^{-1}_{p',u'}(0)) \leq \lambda d(0, G_{p',u'}(x)),$$

hence there exists $x' \in G^{-1}_{p',u'}(0)$ such that

$$\|x - x'\| < \lambda' d(0, G_{p',u'}(x)). \tag{18}$$

Since

$$\begin{aligned} 0 \in G_{p,u}(x) &= f(p,u) + D_x f(p,u)(x-u) + F(x) \\ &= G_{p',u'}(x) + f(p,u) + D_x f(p,u)(x-u) \\ &\qquad - f(p',u') - D_x f(p',u')(x-u'), \end{aligned}$$

the estimate (18) implies

$$\|x - x'\| < \lambda' \|f(p,u) + D_x f(p,u)(x-u) - f(p',u') - D_x f(p',u')(x-u')\|.$$

By the choice of the constants γ, γ_1 and γ_2 we obtain

$$\begin{aligned} \|x - x'\| &< \lambda' \big[\|f(p',u) + D_x f(p',u')(x-u) - f(p',u') - D_x f(p',u')(x-u')\| \\ &\quad + L_1\|p-p'\| + L_2(\|p-p'\| + \|u-u'\|)\|x-u\|\big] \\ &< \lambda' \|f(p',u) + D_x f(p',u')(u'-u) - f(p',u')\| \\ &\quad + \gamma_1\|p-p'\| + \gamma_2(\|p-p'\| + \|u-u'\|)\|x-u\| \\ &\leq \gamma\|u-u'\|^2 + \gamma_1\|p-p'\| + \gamma_2(\|p-p'\| + \|u-u'\|)\|x-u\|. \end{aligned}$$

This gives us (17). □

The result presented next is stated in two ways: the first exhibits the qualitative side of it while the second one gives quantitative estimates.

Theorem 6D.4 (Lyusternik–Graves theorem for Newton method). *If the mapping G defined in (4) is metrically regular at $\bar{x}$ for 0 then the mapping Ξ defined in (16) has the partial Aubin property with respect to p uniformly in x at $(\bar{p},\bar{x})$ for $\bar{\xi}$. In fact, we have the following: if the mapping G is metrically regular at $\bar{x}$ for 0 then*

$$\widehat{\operatorname{lip}}_u(\Xi;(\bar{p},\bar{x})\,|\,\bar{\xi}) = 0 \quad \textit{and} \quad \widehat{\operatorname{lip}}_p(\Xi;(\bar{p},\bar{x})\,|\,\bar{\xi}) \leq \operatorname{reg}(G;\bar{x}\,|\,0)\cdot\|D_p f(\bar{p},\bar{x})\|. \tag{19}$$

Furthermore, if the function f satisfies the ample parameterization condition

$$\text{(20)} \qquad \textit{the mapping } D_p f(\bar{p},\bar{x}) \textit{ is surjective,}$$

then the converse implication holds as well: if the mapping Ξ *has the partial Aubin property with respect to* p *uniformly in* x *at* $(\bar{p},\bar{x})$ *for* $\bar{\xi}$*, then the mapping* G *is metrically regular at* $\bar{x}$ *for* 0.

Proof. Fix γ, γ_1, γ_2 as in Lemma 6D.3 and let α, ζ be the corresponding constants whose existence is claimed in this lemma, while $\bar{a}$ and $\bar{c}$ are the constants in the statement of Theorem 6D.2. Choose positive reals ε and d satisfying the inequalities

$$\text{(21)} \qquad \varepsilon \le \bar{a}/2, \quad \varepsilon \le \alpha, \quad \tau := 2(\gamma+\gamma_2)\varepsilon < \frac{1}{8},$$

$$\text{(22)} \qquad d \le \bar{c}, \quad d \le \zeta, \quad \frac{1}{1-\tau}(\gamma_1+\tau)d < \frac{\varepsilon}{8},$$

and

$$\text{(23)} \qquad e(S(p)\cap \mathbb{B}_{\varepsilon/2}(\bar{x}), S(p')) < \gamma_1\|p-p'\| \ \text{ for } p,\, p' \in \mathbb{B}_d(\bar{p}),\ p \ne p'.$$

The existence of ε and d such that the last relation (23) holds is implied by the Aubin property of S claimed in Theorem 5E.5.

Let p, $p' \in \mathbb{B}_d(\bar{p})$, $p \ne p'$ and $u, u' \in \mathbb{B}_\varepsilon(\bar{x})$, $u \ne u'$. Let $\xi = \{x_k\} \in \Xi(p,u) \cap \mathbb{B}_{\varepsilon/2}(\bar{\xi})$. Then ξ is convergent and its limit is an element of $S(p)$. Let

$$\delta_k := \tau^k\|u-u'\| + \frac{1-\tau^k}{1-\tau}(\gamma_1+\tau)\|p-p'\|, \quad k = 0,1,\ldots.$$

The last inequalities in (21) and (22) imply $\delta_k < \varepsilon/2$.

First we define a sequence $\xi' = \{x'_k\} \in \Xi(p',u')$ with the additional property that

$$\text{(24)} \qquad \|x_k - x'_k\| \le \delta_k, \quad \|x'_k - \bar{x}\| \le \varepsilon.$$

Since p, $p' \in \mathbb{B}_d(\bar{p}) \subset \mathbb{B}_\zeta(\bar{p})$, u, u', $x_1 \in \mathbb{B}_\varepsilon(\bar{x}) \subset \mathbb{B}_\alpha(\bar{x})$ and $x_1 \in G^{-1}_{p,u}(0)$, according to Lemma 6D.3 there exists $x'_1 \in G^{-1}_{p',u'}(0)$ such that

$$\|x_1 - x'_1\| \le \gamma\|u-u'\|^2 + \gamma_1\|p-p'\| + \gamma_2(\|p-p'\| + \|u-u'\|)\|u-x_1\|.$$

Using (21) and (22) we obtain

$$\begin{aligned}\|x_1 - x'_1\| &\le 2\gamma\varepsilon\|u-u'\| + \gamma_1\|p-p'\| + \gamma_2(\|p-p'\| + \|u-u'\|)2\varepsilon \\ &\le \tau\|u-u'\| + (\gamma_1+\tau)\|p-p'\| = \delta_1.\end{aligned}$$

In addition, we have

$$\text{(25)} \qquad \|x'_1 - \bar{x}\| \le \|x'_1 - x_1\| + \|x_1 - \bar{x}\| \le \delta_1 + \frac{\varepsilon}{2} < \varepsilon.$$

Now assume that x'_k is already defined so that (24) holds. Applying Lemma 6D.3 for p, p', x_k, x'_k and $x_{k+1} \in G^{-1}_{p,x_k}(0) \cap \mathbb{B}_{\varepsilon/2}(\bar{x})$ (instead of (p, p', u, u', x_1)) we obtain that there exists $x'_{k+1} \in G^{-1}_{p',x'_k}(0)$ such that

$$\|x_{k+1} - x'_{k+1}\| \le \gamma\|x_k - x'_k\|^2 + \gamma_1\|p - p'\| + \gamma_2(\|p - p'\| + \|x_k - x'_k\|)\|x_k - x_{k+1}\|.$$

In a similar way,

$$\begin{aligned}
\|x_{k+1} - x'_{k+1}\| &\le 2\gamma\varepsilon\|x_k - x'_k\| + \gamma_1\|p - p'\| + \gamma_2(\|p - p'\| + \|x_k - x'_k\|)2\varepsilon \\
&\le 2(\gamma + \gamma_2)\varepsilon\|x_k - x'_k\| + (\gamma_1 + 2\gamma_2\varepsilon)\|p - p'\| \\
&\le \tau\delta_k + (\gamma_1 + \tau)\|p - p'\| \\
&= \tau\left(\tau^k\|u - u'\| + \frac{1 - \tau^k}{1 - \tau}(\gamma_1 + \tau)\|p - p'\|\right) + (\gamma_1 + \tau)\|p - p'\| \\
&= \tau^{k+1}\|u - u'\| + \frac{1 - \tau^{k+1}}{1 - \tau}(\gamma_1 + \tau)\|p - p'\| = \delta_{k+1}.
\end{aligned}$$

To complete the induction, it remains to note that $\|x'_{k+1} - \bar{x}\| \le \varepsilon$ follows from the last estimate in exactly the same way as in (25). In particular, from the last inequality we obtain that

$$(26) \qquad \|x_k - x'_k\| \le \tau\|u - u'\| + \frac{\gamma_1 + \tau}{1 - \tau}\|p - p'\| \qquad \text{for all } k \ge 1.$$

Since the sequence ξ is convergent to some $s \in S(p)$, there exists a natural N such that

$$\|x_k - s\| \le \tau(\|u - u'\| + \|p - p'\|) \ \text{ for all } \ k \ge N.$$

We will now take the finite sequence $x'_1, \ldots, x'_N$ and extend it to an infinite sequence which belongs to $\Xi(p', u')$ by replacing the existing elements $x'_{N+1}, x'_{N+2}, \cdots$ by new ones. With some abuse of notation the new elements are denoted again by $x'_{N+1}, x'_{N+2}, \cdots$ and the new sequence is again denoted by ξ'. The Aubin property of the solution map S implies that there exists $s' \in S(p')$ such that $\|s' - s\| \le \gamma_1\|p - p'\|$. We also have

$$\|s' - \bar{x}\| \le \|s' - s\| + \|s - \bar{x}\| \le \gamma_1\|p - p'\| + \varepsilon/2 \le 2d\gamma_1 + \varepsilon/2 < \bar{a}/2,$$

by (21) and (22). Thus, for $k > N$ we determine x'_k as the elements of a sequence generated by the Newton iteration for p' and initial point x'_N, which is quadratically convergent to s'; indeed, Theorem 6D.2 claims that such a sequence exists. Observe that $p' \in \mathbb{B}_d(\bar{p}) \subset \mathbb{B}_{\bar{c}}(\bar{p})$ and $x'_N \in \mathbb{B}_{\bar{a}}(\bar{x})$ by the second inequality in (24) and since $\varepsilon \le \bar{a}/2$ as assumed in (21). Using (21) and (22) we also have

$$\begin{aligned}
\|x'_N - s'\| &\le \|x'_N - x_N\| + \|x_N - s\| + \|s - s'\| \\
&\le \delta_N + \tau(\|u - u'\| + \|p - p'\|) + \gamma_1\|p - p'\|
\end{aligned}$$

$$\leq 2\tau\|u-u'\| + \left(\frac{\gamma_1+\tau}{1-\tau} + \gamma_1 + \tau\right)\|p-p'\|,$$

hence

$$\|x'_N - s'\| < \frac{1}{8}2\varepsilon + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon. \tag{27}$$

According to Theorem 6D.2 there is a sequence $x'_{N+1},\ldots,x'_k,\ldots$ such that

$$\|x'_{k+1} - s'\| \leq \gamma\|x'_k - s'\|^2 \quad \text{for all } k \geq N. \tag{28}$$

Then, for $k > N$,

$$\|x'_k - s'\| \leq \gamma^{2^{k-N}-1}\|x'_N - s'\|^{2^{k-N}} \leq \gamma^{2^{k-N}-1}\varepsilon^{2^{k-N}} < \varepsilon,$$

where we apply (27) and use the inequality $\gamma\varepsilon < 1$ coming from (21). Therefore, applying the last estimate to (28), we get

$$\|x'_{k+1} - s'\| \leq \gamma\varepsilon\|x'_k - s'\| \quad \text{for all } k \geq N.$$

Further, since $\gamma\varepsilon < 1$, we come to

$$\|x'_k - s'\| \leq \gamma\varepsilon\|x'_N - s'\| \quad \text{for all } k > N.$$

From the estimate just before (27) we obtain that for $k \geq N+1$

$$\begin{aligned}
\|x_k - x'_k\| &\leq \|x_k - s\| + \|s - s'\| + \|s' - x'_k\| \\
&\leq \tau(\|u-u'\| + \|p-p'\|) + \gamma_1\|p-p'\| + \varepsilon\gamma\|x'_N - s'\| \\
&\leq (\tau + 2\tau\varepsilon\gamma)\|u-u'\| + \left[\tau + \gamma_1 + \varepsilon\gamma\left(\frac{\gamma_1+\tau}{1-\tau} + \gamma_1 + \tau\right)\right]\|p-p'\| \\
&\leq (\tau + 2\tau\varepsilon\gamma)\|u-u'\| + \left[\frac{\gamma_1+\tau}{1-\tau} + \varepsilon\gamma\left(\frac{\gamma_1+\tau}{1-\tau} + \gamma_1 + \tau\right)\right]\|p-p'\|.
\end{aligned}$$

By comparing with (26) we conclude that $\|x_k - x'_k\|$ is bounded by the same expression as on the last line above not only for $k > N$ but also for $k \leq N$, hence for all k. Thus the distance $d(\xi, \Xi(p',u'))$ is bounded by the same expression. This holds for every $p, p' \in \mathbb{B}_d(\bar{p})$, $u, u' \in \mathbb{B}_\varepsilon(\bar{x})$ and every $\xi \in \Xi(p,u) \cap \mathbb{B}_{\varepsilon/2}(\bar{\xi})$ and ε can be arbitrarily small. Observe that when ε is small, then τ is also small, hence the constant multiplying $\|u-u'\|$ can be arbitrarily close to zero, and that the constant multiplying $\|p-p'\|$ can be arbitrarily close to γ_1. This yields (19) and completes the proof of the first part of the theorem.

Now, let the ample parameterization condition (20) be satisfied. Let κ, c and a be positive constants such that

$$e(\Xi(p,u) \cap \Omega, \Xi(p',u)) \leq \kappa\|p-p'\| \quad \text{whenever} \quad p, p' \in \mathbb{B}_c(\bar{p}), u \in \mathbb{B}_a(\bar{x}),$$

where Ω is a neighborhood of $\bar{\xi}$. Make a smaller if necessary so that $\mathbb{B}_a(\bar{\xi}) \subset \Omega$ and then take c smaller so that $\kappa c < a/2$. Since gph F is closed, it follows that for any $p \in \mathbb{B}_c(\bar{p})$ and any sequence with components $x_k \in \mathbb{B}_a(\bar{x})$ convergent to x and satisfying

$$(29) \qquad f(p,x_k) + D_x f(p,x_k)(x_{k+1} - x_k) + F(x_{k+1}) \ni 0 \quad \text{for all} \quad k = 1, 2, \ldots,$$

one has $f(p,x) + F(x) \ni 0$, that is, $x \in S(p)$.

We will prove that S has the Aubin property at $\bar{p}$ for $\bar{x}$, and then we will apply Theorem 5E.6 to show the metric regularity of G at $\bar{x}$ for 0. Pick $p, p' \in \mathbb{B}_{c/2}(\bar{p})$ with $p \neq p'$ and $x \in S(p) \cap \mathbb{B}_{a/2}(\bar{x})$ (if there is no such x we are done). Let $\chi := \{x, x, \ldots\}$. Since $\|\chi - \bar{\xi}\|_\infty = \|x - \bar{x}\| \leq a/2$, we have $\chi \in \Xi(p,x) \cap \Omega$. Hence

$$d(\chi, \Xi(p',x)) \leq \kappa \|p - p'\|.$$

Take $\varepsilon > 0$ such that $(\kappa + \varepsilon)c \leq a/2$. Then there is some $\Psi \in \Xi(p',x)$ such that

$$\|\chi - \Psi\|_\infty < (\kappa + \varepsilon)\|p - p'\|,$$

with $\Psi = \{x'_1, x'_2, \ldots\}$ and $x'_k \to x' \in X$. For all k we have

$$\|x'_k - \bar{x}\| \leq \|x'_k - x\| + \|x - \bar{x}\| \leq \|\Psi - \chi\|_\infty + a/2 \leq (\kappa + \varepsilon)c + a/2 \leq a.$$

Hence from (29) we obtain $x' \in S(p') \cap \mathbb{B}_a(\bar{x})$. Moreover,

$$\|x - x'\| \leq \|x - x'_k\| + \|x'_k - x'\| \leq \|\chi - \Psi\|_\infty + \|x'_k - x'\| \leq (\kappa + \varepsilon)\|p - p'\| + \|x'_k - x'\|.$$

Making $k \to \infty$, we get $\|x - x'\| \leq (\kappa + \varepsilon)\|p - p'\|$. Thus,

$$d(x, S(p')) \leq \|x - x'\| \leq (\kappa + \varepsilon)\|p - p'\|.$$

Taking $\varepsilon \searrow 0$, we obtain that S has the Aubin property at $\bar{p}$ for $\bar{x}$ with constant κ, as claimed. From Theorem 5E.6, G is metrically regular at $\bar{x}$ for 0 and we obtain the desired result. $\square$

To shed more light on the kind of result we just proved, consider the special case when $f(p,x)$ has the form $f(x) - p$ and take, for simplicity, $\bar{p} = 0$. Then the Newton iteration (2) becomes

$$f(x_k) + Df(x_k)(x_{k+1} - x_k) + F(x_{k+1}) \ni p, \quad \text{for } k = 0, 1, \ldots,$$

where the ample parameterization condition (20) holds automatically. Define the mappings

$$X \ni x \mapsto G_u(x) = f(u) + Df(u)(x - u) + F(x), \qquad \text{for} \quad u \in X.$$

and $\Gamma : l^c_\infty(X) \rightrightarrows X \times P$ as

$$(30)\qquad \Gamma : \xi \mapsto \left\{ \begin{pmatrix} u \\ p \end{pmatrix} \mid u = x_0 \ \text{ and } \ G_{x_k}(x_{k+1}) \ni p \ \text{for every} \ k = 0,1,\dots \right\}.$$

Then Theorem 6D.4 becomes the following characterization result:

Corollary 6D.5 (symmetric Lyusternik–Graves theorem for Newton method)**.** *The following are equivalent:*

(a) *The mapping $f+F$ is metrically regular at $\bar{x}$ for 0;*
(b) *The mapping Γ defined in (30) is metrically regular at $\bar{\xi}$ for $(\bar{x},0)$.*

Next comes a statement analogous to Theorem 6D.4 for the case when the mapping G is strongly metrically regular.

Theorem 6D.6 (implicit function theorem for Newton iteration)**.** *Suppose that the mapping G defined in (4) is strongly metrically regular at $\bar{x}$ for 0. Then the mapping Ξ in (16) has a Lipschitz single-valued localization ξ around $(\bar{p},\bar{x})$ for $\bar{\xi}$, with*

$$(31)\qquad \widehat{\mathrm{lip}}_u(\xi;(\bar{p},\bar{x})) = 0 \quad \textit{and} \quad \widehat{\mathrm{lip}}_p(\xi;(\bar{p},\bar{x})) \le \mathrm{reg}\,(G;\bar{x}\,|\,0)\cdot\widehat{\mathrm{lip}}_p(f;(\bar{p},\bar{x})).$$

Moreover, for (p,u) close to $(\bar{p},\bar{x})$, $\xi(p,u)$ is a quadratically convergent sequence to a locally unique solution. Also, the same conclusion holds if we replace the space $l^c_\infty(X)$ in the definition of Ξ by the space of all sequences with elements in X, not necessarily convergent, equipped with the $l_\infty(X)$ norm; that is, we consider the mapping $\widehat{\Xi}$ introduced after Theorem 6D.2.

If the function f satisfies the ample parameterization condition (20), then the converse implication holds as well: the mapping G is strongly metrically regular at $\bar{x}$ for 0 provided that Ξ has a Lipschitz continuous single-valued localization ξ around $(\bar{p},\bar{x})$ for $\bar{\xi}$.

Proof. Assume that the mapping G is strongly metrically regular at $\bar{x}$ for 0. Then, by 5F.4, the solution mapping $p \mapsto S(p)$ is strongly metrically regular; let $x(p)$ be the locally unique solution associated with p near $\bar{p}$. Consider the mapping $\widehat{\Xi}$ defined in the same way as Ξ in (16) but with $l^c_\infty(X)$ replaced by $l_\infty(X)$. According to Theorem 6D.2, the mapping $\widehat{\Xi}$ has a single valued graphical localization ξ around $(\bar{p},\bar{x})$ for $\bar{\xi}$ satisfying (31). Moreover, from Theorem 6D.2, for (p,u) close to $(\bar{p},\bar{x})$, the values $\xi(p,u)$ of this localization are quadratically convergent sequences to the locally unique solution $x(p)$. Thus, any graphical localization of the mapping $\widehat{\Xi}$ with sufficiently small neighborhoods agrees with the corresponding graphical localization of the mapping Ξ, which gives us the first claim of the theorem.

Assume that the ample parameterization condition (20) holds and let Ξ have a Lipschitz localization ξ around $(\bar{p},\bar{x})$ for $\bar{\xi}$; that is, $(p,u) \mapsto \Xi(p,u) \cap \mathbb{B}_\beta(\bar{\xi})$ is a singleton $\xi(p,u)$ for any $p \in \mathbb{B}_\alpha(\bar{p})$ and $u \in \mathbb{B}_\alpha(\bar{x})$. Then in particular, Ξ has the Aubin property with respect to p uniformly in x at $(\bar{p},\bar{x})$ for $\bar{\xi}$ and hence, by Theorem 6D.4, G is metrically regular at $\bar{x}$ for 0. Take $\bar{a}$ as in (11) and make it smaller if necessary so that $\bar{a} \le \beta$. Since by Theorem 5E.5 the solution mapping S has the Aubin property at $\bar{p}$ for $\bar{x}$, it remains to show that S is locally nowhere multivalued.

Take $a := \min\{\bar{a}/2, \alpha, \beta\}$ and $c := \min\{\bar{c}, \alpha\}$ and let $p \in \mathbb{B}_c(\bar{p})$ and $x, x' \in S(p) \cap \mathbb{B}_a(\bar{x})$, $x \neq x'$. Clearly, $\{x, x, \ldots\} = \Xi(p,x) \cap \mathbb{B}_\beta(\bar{\xi}) = \xi(p,x)$. Further, according to Theorem 6D.2 there exists a Newton sequence ξ' for p starting again from x and each element of which is in $\mathbb{B}_{\bar{a}}(\bar{x})$, which converges to x', thus $\xi' \in \Xi(p,x) \cap \mathbb{B}_\beta(\bar{\xi})$. But this contradicts the assumption that $\Xi(p,x) \cap \mathbb{B}_\beta(\bar{\xi})$ is a singleton. Thus, S has a single-valued graphical localization around $\bar{p}$ for $\bar{x}$ and since it has the Aubin property it is in fact Lipschitz continuous, by Proposition 3G.1, whose extension to Banach spaces is straightforward. It remains to apply Theorem 5F.5 which asserts that the latter property is equivalent to the strong metric regularity of G at $\bar{x}$ for 0. □

As an illustration of possible applications of the results in Theorem 6D.6 in studying complexity of Newton's iteration, we will derive an estimate for the number of iterations needed to achieve a particular accuracy of the method, which is the same for all values of the parameter p in some neighborhood of the reference point $\bar{p}$. Consider Newton's method (2) in the context of 6D.6. Given an accuracy measure ρ, suppose that the method is to be terminated at the k-th step if

$$d(0, f(p,x_k) + F(x_k)) \le \rho. \tag{32}$$

Let a and c be positive constants such that for any $p \in \mathbb{B}_c(\bar{p})$ the locally unique solution $s(p) \in \mathbb{B}_{a/2}(\bar{x})$ and also all elements of the sequence $\{x_k\}$ generated by (2) for p are in $\mathbb{B}_a(\bar{x})$. Since x_k is a Newton's iterate from x_{k-1}, we have that

$$f(p,x_k) - f(p,x_{k-1}) - D_x f(p,x_{k-1})(x_k - x_{k-1}) \in f(p,x_k) + F(x_k).$$

Let L be a Lipschitz constant of $D_x f$ with respect to x uniformly in p around $(\bar{p}, \bar{x})$. Using (8), as in the proof of 6D.1, we have

$$\begin{aligned} &d(0, f(p,x_k) + F(x_k)) \\ &\qquad \le \|f(p,x_k) - f(p,x_{k-1}) - D_x f(p,x_{k-1})(x_k - x_{k-1})\| \\ &\qquad \le L\|x_k - x_{k-1}\|^2/2. \end{aligned} \tag{33}$$

Let k_ρ be the first iteration at which (32) holds; then for $k < k_\rho$ from (33) we obtain

$$\rho < \frac{L}{2}\|x_k - x_{k-1}\|^2. \tag{34}$$

Further, utilizing (15) with $\theta = \gamma\|x_0 - s(p)\|$ we get

$$\|x_k - x_{k-1}\| \le \|x_k - s(p)\| + \|x_{k-1} - s(p)\| \le \theta^{2^k - 2}(1+\theta)(\|x_0 - \bar{x}\| + \|s(p) - \bar{x}\|),$$

and from the choice of x_0 we have

$$\|x_k - x_{k-1}\| \le \theta^{2^k - 2}(1+\theta)\frac{3a}{2}.$$

But then, taking into account (34), we obtain

$$\rho < \frac{1}{2} L\theta^{2^{k+1}} \frac{9a^2(1+\theta)^2}{4\theta^4}.$$

Therefore k_ρ satisfies

$$k_\rho \leq \log_2 \left(\log_\theta \left(\frac{8\theta^4 \rho}{9a^2 L(1+\theta)^2} \right) \right) - 1.$$

Thus, we have obtained an upper bound of the number of iterations needed to achieve a particular accuracy, which, most importantly, is *the same for all values of the parameter p in some neighborhood of the reference value $\bar{p}$.* This tells us, for example, that small changes of parameters in a problem don't affect the performance of Newton's method as applied to this problem.

6E. Inexact Newton's Methods under Strong Metric Subregularity

In Section 3I we introduced the property of strong metric subregularity and provided an equivalent definition of it in 3I(5), which we now restate in an infinite-dimensional space setting. Throughout, X and Y are Banach spaces.

Strong metric subregularity. *A mapping $F : X \rightrightarrows Y$ is said to be strongly metrically subregular at $\bar{x}$ for $\bar{y}$ when $\bar{y} \in F(\bar{x})$ and there is a constant $\kappa \geq 0$ together with a neighborhood U of $\bar{x}$ such that*

$$\|x - \bar{x}\| \leq \kappa d(\bar{y}, F(x)) \quad \text{for all } x \in U.$$

From Theorem 3I.3 and the discussion after it, the strong metric subregularity of F at $\bar{x}$ for $\bar{y}$ is equivalent to the isolated calmness property of the inverse mapping F^{-1} with the same constant $\kappa > 0$; namely, there exists a neighborhood U of $\bar{x}$ such that

$$F^{-1}(y) \cap U \subset \bar{x} + \kappa \|y - \bar{y}\| \mathbb{B} \quad \text{when} \quad y \in Y.$$

As shown in Theorem 3I.7, strong metric subregularity obeys the general paradigm of the inverse function theorem. For example, if f is Fréchet differentiable at $\bar{x}$ with derivative $Df(\bar{x})$, then the strong metric subregularity of f at $\bar{x}$ is equivalent to the following property of the derivative $Df(\bar{x})$: there exists $\kappa > 0$ such that $\|w\| \leq \kappa \|Df(\bar{x})w\|$ for all $w \in X$. If $X = Y = \mathbb{R}^n$ this reduces to nonsingularity of the Jacobian $\nabla f(\bar{x})$ and then strong metric subregularity becomes equivalent to strong metric regularity. Note that a strongly subregular at $\bar{x}$ mapping F could be empty-valued at some points in every neighborhood of $\bar{x}$.

Recall that a function $f: X \to Y$ is said to be *calm* at $\bar{x} \in \operatorname{dom} f$ when there exists a constant $L > 0$ and a neighborhood U of $\bar{x}$ such that

$$\|f(x) - f(\bar{x})\| \le L\|x - \bar{x}\| \quad \text{for all } x \in U \cap \operatorname{dom} f.$$

The following proposition shows that under calmness and strong metric subregularity of a function f one can characterize superlinear convergence of a sequence $\{x_k\}$ through convergence of the function values $f(x_k)$.

Proposition 6E.1 (characterization of superlinear convergence)**.** *Let $f: X \to X$ be a function which is both calm and strongly metrically subregular at $\bar{x} \in \operatorname{int} \operatorname{dom} f$; that is, there exist positive constants κ and L, and a neighborhood U of $\bar{x}$ such that*

$$\frac{1}{\kappa}\|x - \bar{x}\| \le \|f(x) - f(\bar{x})\| \le L\|x - \bar{x}\| \quad \textit{for all } x \in U.$$

Consider any sequence $\{x_k\}$ the elements of which satisfy $x_k \ne \bar{x}$ for all k. Then $x_k \to \bar{x}$ superlinearly if and only if

$$(1) \qquad x_k \in U \quad \textit{for all } k \textit{ sufficiently large} \quad \textit{and} \quad \lim_{k \to \infty} \frac{\|f(x_{k+1}) - f(\bar{x})\|}{\|s_k\|} = 0.$$

Proof. Consider an infinite sequence $\{x_k\}$ such that $x_k \ne \bar{x}$ for all k. Let $x_k \to \bar{x}$ superlinearly. Let $\varepsilon > 0$ and choose k_0 large enough so that $x_k \in U$ for all $k \ge k_0$ and, by the superlinear convergence, $\|e_{k+1}\|/\|e_k\| < \varepsilon$ for all $k \ge k_0$. Observing that

$$\left| \frac{\|s_k\| - \|e_k\|}{\|e_k\|} \right| \le \frac{\|s_k + e_k\|}{\|e_k\|} = \frac{\|e_{k+1}\|}{\|e_k\|}$$

we obtain

$$(2) \qquad \frac{\|s_k\|}{\|e_k\|} \to 1 \quad \text{as} \quad k \to \infty.$$

Then we can take k_0 even larger if necessary so that

$$(3) \qquad \|e_{k+1}\| \le \varepsilon \|s_k\| \qquad \text{for all} \quad k \ge k_0,$$

in which case

$$\frac{\|f(x_{k+1}) - f(\bar{x})\|}{\|s_k\|} \le \frac{L\|x_{k+1} - \bar{x}\|}{\|s_k\|} = \frac{L\|e_{k+1}\|}{\|s_k\|} \le L\varepsilon.$$

Since ε can be arbitrarily small, this proves (1).

For the other direction, let (1) be satisfied for a sequence $\{x_k\}$. Choose any $\varepsilon \in (0, 1/\kappa)$ and let k_1 be so large that $x_k \in U$ for all $k \ge k_1$ and, from (1),

$$\|f(x_{k+1}) - f(\bar{x})\| \le \varepsilon \|s_k\| \qquad \text{for all } k \ge k_1.$$

The assumed strong metric subregularity yields $\|x_{k+1}-\bar{x}\| \le \kappa\|f(x_{k+1})-f(\bar{x})\|$ for all $k \ge k_1$, and hence, $\|e_{k+1}\| \le \kappa\varepsilon\|s_k\|$ for all $k \ge k_1$. But then, for such k,

$$\|e_{k+1}\| \le \kappa\varepsilon\|s_k\| \le \kappa\varepsilon(\|e_k\| + \|e_{k+1}\|),$$

therefore,

$$\frac{\|e_{k+1}\|}{\|e_k\|} \le \frac{\kappa\varepsilon}{1-\kappa\varepsilon}$$

for all $k \ge k_1$. Since ε can be arbitrarily small, we conclude that $x_k \to \bar{x}$ superlinearly as claimed. The proof is complete. □

We will now use 6E.1 to prove superlinear convergence of Newton's method for solving an equation $f(x) = 0$.

Theorem 6E.2 (superlinear convergence from strong subregularity)**.** *Consider a function $f : X \to X$ with a zero at $\bar{x}$ and suppose that there exists $a > 0$ such that f is both continuously Fréchet differentiable in the ball $\mathbb{B}_a(\bar{x})$ and strongly metrically subregular at $\bar{x}$ for 0 with neighborhood $\mathbb{B}_a(\bar{x})$. Consider the Newton method*

$$(4) \qquad f(x_k) + Df(x_k)(x_{k+1}-x_k) = 0 \quad \text{for } k = 0,1,\ldots.$$

Then every sequence $\{x_k\}$ generated by the iteration (4) which is convergent to $\bar{x}$ is in fact superlinearly convergent to $\bar{x}$.

Proof. By elementary calculus, the boundedness of Df on $\mathbb{B}_a(\bar{x})$ implies that f is calm at $\bar{x}$. Let $\varepsilon > 0$; then there exists $a' \in (0,a]$ such that

$$(5) \qquad \|Df(x) - Df(x')\| \le \varepsilon \quad \text{for all } \; x,\, x' \in \mathbb{B}_{a'}(\bar{x}).$$

Suppose that the method (4) generates a sequence $\{x_k\}$ convergent to $\bar{x}$; then $x_k \in \mathbb{B}_{a'}(\bar{x})$ for all sufficiently large k. Using (4), we have

$$f(x_{k+1}) = f(x_k) + \int_0^1 Df(x_k + \tau s_k)s_k d\tau = \int_0^1 (Df(x_k + \tau s_k) - Df(x_k))s_k d\tau.$$

Hence, from (5), for every $\varepsilon > 0$ there exists k_0 such that for all $k > k_0$ one has

$$\|f(x_{k+1}) - f(\bar{x})\| = \|f(x_{k+1})\| \le \varepsilon\|s_k\|.$$

Since ε can be arbitrarily small, this yields (1). Applying 6E.1 we complete the proof. □

In the 1960s it was discovered that in order to have fast convergence of Newton's iteration (4) for solving an equation $f(x) = 0$ with $f : \mathbb{R}^n \to \mathbb{R}^n$, it is sufficient to use an "approximation" of the derivative Df at each iteration. This has resulted in the rapid development of *quasi-Newton* methods having the form

$$(6) \qquad f(x_k) + B_k(x_{k+1}-x_k) = 0 \quad \text{for } k = 0,1,\ldots.$$

where B_k is a sequence of nonsingular matrices. The particular way B_k is constructed determines the quasi-Newton method, e.g., the Broyden class, BFGS, SR1, etc., which are now considered as methods of choice for solving nonlinear equations. We shall not discuss here specific quasi-Newton methods but rather turn our attention to a general result, the Dennis–Moré theorem, which shows exactly how the derivative has to be approximated in order to obtain superlinear convergence.

The Dennis–Moré theorem was initially proved for equations; here we will present an extended version of it for the generalized equation

$$f(x)+F(x) \ni 0, \tag{7}$$

where $f : X \to Y$ is a continuously differentiable function everywhere and $F : X \rightrightarrows Y$ is a set-valued mapping with closed graph. We consider the following class of quasi-Newton methods for solving (7):

$$f(x_k)+B_k(x_{k+1}-x_k)+F(x_{k+1}) \ni 0 \quad \text{for } k=0,1,\dots, \tag{8}$$

where B_k is a sequence of linear and bounded mappings acting from X to Y. When (7) describes the Karush-Kuhn-Tucker optimality system for a nonlinear programming problem, the method (8) may be viewed as a combination of the SQP method with a quasi-Newton method approximating the second derivative of the Lagrangian.

Theorem 6E.3 (Dennis–Moré theorem for generalized equations)**.** *Suppose that f is Fréchet differentiable in an open and convex neighborhood U of $\bar{x}$, where $\bar{x}$ is a solution of (7), and that the derivative mapping Df is continuous at $\bar{x}$. For some starting point x_0 in U consider a sequence $\{x_k\}$ generated by (8) which remains in U for all k and satisfies $x_k \neq \bar{x}$ for all k. Let $E_k = B_k - Df(\bar{x})$.*

If $x_k \to \bar{x}$ superlinearly, then

$$\lim_{k\to\infty} \frac{d(0, f(\bar{x})+E_k s_k+F(x_{k+1}))}{\|s_k\|}=0. \tag{9}$$

Conversely, if the mapping $x \mapsto H(x)=f(\bar{x})+Df(\bar{x})(x-\bar{x})+F(x)$ is strongly metrically subregular at $\bar{x}$ for 0 and the sequence $\{x_k\}$ satisfies

$$x_k \to \bar{x} \quad \textit{and} \quad \lim_{k\to\infty} \frac{\|E_k s_k\|}{\|s_k\|}=0, \tag{10}$$

then $x_k \to \bar{x}$ superlinearly.

Proof. Suppose that the method (8) generates an infinite sequence $\{x_k\}$ with elements in U such that $x_k \neq \bar{x}$ for all k and $x_k \to \bar{x}$ superlinearly. Then

$$\begin{aligned} -f(x_k)+f(\bar{x})-Df(\bar{x})(x_{k+1}-x_k) \\ \in f(\bar{x})+(B_k-Df(\bar{x}))(x_{k+1}-x_k)+F(x_{k+1}). \end{aligned} \tag{11}$$

Thus, to obtain (9) it is sufficient to show that for any $\delta > 0$

$$(12) \qquad \|f(x_k) - f(\bar{x}) + Df(\bar{x})s_k\| \le \delta \|s_k\| \qquad \text{for all sufficiently large } k.$$

Let $\varepsilon > 0$. Then there exists k_0 such that

$$(13) \qquad \|Df(\bar{x} + \tau e_k) - Df(\bar{x})\| < \varepsilon \qquad \text{for all } \tau \in [0,1] \text{ and } k \ge k_0.$$

Choose k_0 larger if necessary so that, by the superlinear convergence,

$$(14) \qquad \frac{\|e_{k+1}\|}{\|e_k\|} < \varepsilon \qquad \text{for all } k \ge k_0.$$

Relying on (2), we can take k_0 even larger if necessary to obtain

$$(15) \qquad \|e_k\| < 2\|s_k\| \qquad \text{for all } k \ge k_0.$$

Using the equality

$$\begin{aligned} f(x_k) - f(\bar{x}) + Df(\bar{x})s_k &= f(x_k) - f(\bar{x}) - Df(\bar{x})e_k + Df(\bar{x})e_{k+1} \\ &= \int_0^1 [Df(\bar{x} + \tau e_k) - Df(\bar{x})]e_k d\tau + Df(\bar{x})e_{k+1} \end{aligned}$$

together with (13), (14) and (15), we obtain that, for $k \ge k_0$,

$$(16) \qquad \begin{aligned} &\|f(x_k) - f(\bar{x}) + Df(\bar{x})s_k\| \\ &\quad < \varepsilon\|e_k\| + \|Df(\bar{x})\|\|e_{k+1}\| < 2\varepsilon(1 + \|Df(\bar{x})\|)\|s_k\|. \end{aligned}$$

Hence, for every $\delta > 0$ one can find a sufficiently small ε such that (12) holds. Then (9) is satisfied, too.

To prove the second part of the theorem, let the sequence $\{x_k\}$ be generated by (8) for some x_0 in U, remain in U and satisfy $x_k \neq \bar{x}$ for all k, and let condition (10) be satisfied. By the assumption of strong metric subregularity, there exist a positive scalar κ and a neighborhood $U' \subset U$ such that

$$(17) \qquad \|x - \bar{x}\| \le \kappa d(0, H(x)) \qquad \text{for all } x \in U'.$$

Clearly, for all sufficiently large k we have $x_k \in U'$. Since

$$f(\bar{x}) + Df(\bar{x})e_k - f(x_k) - E_k s_k \in H(x_{k+1}),$$

we obtain from (17) that, for all large k,

$$(18) \qquad \|x_{k+1} - \bar{x}\| \le \kappa\|f(\bar{x}) + Df(\bar{x})e_k - f(x_k) - E_k s_k\|.$$

Let $\varepsilon \in (0, 1/\kappa)$ and choose k_0 so large that (13) holds for $k \ge k_0$ and $\tau \in [0,1]$, and (18) is satisfied for all $k \ge k_0$. Then,

$$(19) \qquad \|f(x_k) - f(\bar{x}) - Df(\bar{x})e_k\| = \|\int_0^1 [Df(\bar{x} + \tau e_k) - Df(\bar{x})]e_k d\tau\| \le \varepsilon\|e_k\|.$$

Make k_0 even larger if necessary so that, by condition (10),

$$(20) \qquad \|E_k s_k\| < \varepsilon \|s_k\| \qquad \text{for all} \quad k \geq k_0.$$

Then, from (18), taking into account (19) and (20), for $k \geq k_0$ we obtain

$$\begin{aligned}\|x_{k+1} - \bar{x}\| &\leq \kappa \|f(\bar{x}) + Df(\bar{x})e_k - f(x_k)\| + \kappa \|E_k s_k\| \\ &< \kappa\varepsilon \|e_k\| + \kappa\varepsilon \|s_k\| < \kappa\varepsilon(2\|e_k\| + \|e_{k+1}\|).\end{aligned}$$

Thus,

$$\frac{\|e_{k+1}\|}{\|e_k\|} < \frac{2\kappa\varepsilon}{1 - \kappa\varepsilon}.$$

Since ε can be arbitrarily small, this implies that $x_k \to \bar{x}$ superlinearly. □

For the case of an equation, that is, for $F \equiv 0$ and $f(\bar{x}) = 0$, we obtain from 6E.3 the well-known Dennis–Moré theorem for equations; it was originally stated in $\mathbb{R}^n$, now we have it in Banach spaces:

Theorem 6E.4 (Dennis–Moré theorem for equations)**.** *Suppose that $f : X \to X$ is Fréchet differentiable in an open convex set D in X containing $\bar{x}$, a zero of f, that the derivative mapping Df is continuous at $\bar{x}$ and that there exists a constant $\alpha > 0$ such that $\|Df(\bar{x})w\| \geq \alpha \|w\|$ for all $w \in X$. Let $\{B_k\}$ be a sequence of linear and bounded mappings from X to X and let for some starting point x_0 in D the sequence $\{x_k\}$ be generated by (6), remain in D for all k and satisfy $x_k \neq \bar{x}$ for all k. Then $x_k \to \bar{x}$ superlinearly if and only if*

$$x_k \to \bar{x} \quad \text{and} \quad \lim_{k\to\infty} \frac{\|E_k s_k\|}{\|s_k\|} = 0,$$

where $E_k = B_k - Df(\bar{x})$.

The quasi-Newton method (6) can be viewed as an *inexact version* of the Newton method (4) which clearly has computational advantages. Another way to introduce inexactness is to terminate the iteration (4) when a certain level of accuracy is reached, determined by the *residual* $f(x_k)$. Specifically, consider the following inexact Newton method: given a sequence of positive scalars η_k and a starting point x_0, the $(k+1)$st iterate is chosen to satisfy the condition

$$(21) \qquad \|f(x_k) + Df(x_k)(x_{k+1} - x_k)\| \leq \eta_k \|f(x_k)\|.$$

For example, if the linear equation in (4) is solved by an iterative method, say SOR, then the iteration is terminated when the condition (21) is satisfied for some k. Note that (21) can be also written as

$$\big(f(x_k) + Df(x_k)(x_{k+1} - x_k)\big) \cap \mathbb{B}_{\eta_k \|f(x_k)\|}(0) \neq \emptyset.$$

In sections 6C and 6D we studied the exact Newton iteration for solving (7), but now we will focus on the following inexact version of it. Given x_0 compute x_{k+1} to satisfy

$$(22)\qquad (f(x_k)+Df(x_k)(x_{k+1}-x_k)+F(x_{k+1}))\cap R_k(x_k)\neq\emptyset \quad \text{for } k=0,1,\ldots,$$

where $R_k : X \rightrightarrows Y$ is a sequence of set-valued mappings with closed graphs which represent the inexactness. In the case when F is the zero mapping and $R_k(x_k) = \mathbb{B}_{\eta_k\|f(x_k)\|}(0)$, the iteration (22) reduces to (21).

Theorem 6E.5 (convergence of inexact Newton method)**.** *Suppose that the function f is continuously Fréchet differentiable in a neighborhood of a solution $\bar{x}$ and the mapping $f+F$ is strongly metrically subregular at $\bar{x}$ for 0.*

(a) *Let there exist a sequences of positive scalars $\gamma_k \searrow 0$ and $\beta>0$ such that*

$$(23)\qquad \sup_{x\in R_k(u)}\|x\|\leq \gamma_k\|u-\bar{x}\| \quad \text{for all } u\in\mathbb{B}_\beta(\bar{x}),\quad k=0,1,\ldots.$$

Then every sequence $\{x_k\}$ generated by the Newton method (22) which is convergent to $\bar{x}$ is in fact superlinearly convergent to $\bar{x}$.

(b) *Suppose that the derivative mapping Df is Lipschitz continuous near $\bar{x}$ with Lipschitz constant L and let there exist positive scalars γ and β such that*

$$(24)\qquad \sup_{x\in R_k(u)}\|x\|\leq \gamma\|u-\bar{x}\|^2 \quad \text{for all } u\in\mathbb{B}_\beta(\bar{x}),\quad k=0,1,\ldots.$$

Then every sequence $\{x_k\}$ generated by the Newton method (22) which is convergent to $\bar{x}$ is in fact quadratically convergent to $\bar{x}$.

In the proof we employ the following corollary of Theorem 3I.7, stated as an exercise.

Exercise 6E.6. *Suppose that the mapping $f+F$ is strongly metrically subregular at $\bar{x}$ for 0 with constant λ. For any $u\in X$ consider the mapping*

$$X\ni x\mapsto G_u(x)=f(u)+Df(u)(x-u)+F(x).$$

Prove that for every $\kappa>\lambda$ there exists $a>0$ such that

$$(25)\qquad \|x-\bar{x}\|\leq\kappa d(f(u)+Df(u)(\bar{x}-u)-f(\bar{x}),G_u(x)) \quad \text{for every } x\in\mathbb{B}_a(\bar{x}).$$

Guide. Let $\kappa>\lambda$ and let $\mu>0$ be such that $\lambda\mu<1$ and $\kappa>\lambda/(1-\lambda\mu)$. There exists $a>0$ such that

$$(26)\qquad \|Df(x)-Df(x')\|\leq\mu \quad \text{for every } x,x'\in\mathbb{B}_a(\bar{x}).$$

Fix $u\in X$ and consider the function

$$x\mapsto g_u(x)=f(u)+Df(u)(x-u)-f(x).$$

For every $x \in \mathbb{B}_a(\bar{x})$, using (26), we have

$$\begin{aligned}\|g_u(x) - g_u(\bar{x})\| &= \|f(\bar{x}) - f(x) - Df(u)(\bar{x} - x)\| \\ &= \|\int_0^1 [Df(x + t(\bar{x} - x))(\bar{x} - x) - Df(u)(\bar{x} - x)]dt\| \leq \mu\|x - \bar{x}\|,\end{aligned}$$

that is, $\operatorname{clm}(g;\bar{x}) \leq \mu$. Apply Theorem 3I.7 to the mapping $g_u + f + F = G_u$ to obtain (25). □

Proof of 6E.5(a). Let the mapping $f + F$ be strongly metrically subregular at $\bar{x}$ for 0 with constant λ and consider any sequence $\{x_k\}$ generated by the Newton method (22) which is convergent to $\bar{x}$. Let $\kappa > \lambda$ and choose $a > 0$ such that (25) holds. Pick any $\mu > 0$ and adjust a if necessary so that $a \leq \beta$ and (26) is satisfied with that μ. Then $x_k \in \mathbb{B}_a(\bar{x})$ for k sufficiently large. For all such k there exists $y_k \in R_k(x_k) \cap G_{x_k}(x_{k+1})$ and from (25) we have

$$\begin{aligned}\|x_{k+1} - \bar{x}\| &\leq \kappa\|f(x_k) + Df(x_k)(\bar{x} - x_k) - f(\bar{x}) - y_k\| \\ &\leq \kappa\mu\|x_k - \bar{x}\| + \kappa\|y_k\| \leq \kappa(\mu + \gamma_k)\|x_k - \bar{x}\|.\end{aligned}$$

If $x_k \neq \bar{x}$ for all k, passing to the limit as $k \to \infty$ we obtain

$$\lim_{k\to\infty} \frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} \leq \kappa\mu. \tag{27}$$

Since μ can be arbitrary small and the limit on the left side of (27) does not depend on it, we are done.

Proof of 6E.5(b). As in the preceding argument, pick $\kappa > \lambda$ and choose $a > 0$ such that (25) holds. Let L be a Lipschitz constant of Df and make a smaller if necessary so that $a \leq \beta$ and

$$\|Df(u) - Df(v)\| \leq L\|u - v\| \quad \text{for all} \quad u, v \in \mathbb{B}_a(\bar{x}).$$

Then, for any $x \in \mathbb{B}_a(\bar{x})$ we have the standard estimate

$$\begin{aligned}&\|f(x) + Df(x)(\bar{x} - x) - f(\bar{x})\| \\ &\quad = \left\|\int_0^1 Df(\bar{x} + t(x - \bar{x}))(x - \bar{x})dt - Df(x)(x - \bar{x})\right\| \leq \tfrac{L}{2}\|x - \bar{x}\|^2.\end{aligned} \tag{28}$$

Consider a sequence $\{x_k\}$ generated by Newton method (22) which is convergent to $\bar{x}$. By repeating the argument of case (a) and employing (28) in place of (26), we obtain

$$\|x_{k+1} - \bar{x}\| \leq \kappa(\gamma + L/2)\|x_k - \bar{x}\|^2.$$

Hence, $\{x_k\}$ converges quadratically to $\bar{x}$. □

6F. Nonsmooth Newton's Method

In this section we continue our study of Newton's method for solving the generalized equation

$$\text{find } x \text{ such that } f(x)+F(x) \ni 0, \tag{1}$$

now in finite dimensions, where the set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ has closed graph but, in contrast to the preceding section, the function $f : \mathbb{R}^n \to \mathbb{R}^n$ is not necessarily differentiable. Let $\bar{x}$ be a solution of (1). To introduce a Newton-type iteration for (1) we use a "linearization" at $\bar{x}$ of the form

$$G_A : x \mapsto f(\bar{x})+A(x-\bar{x})+F(x), \tag{2}$$

where the matrix A is an element of Clarke's generalized Jacobian $\bar{\partial} f(\bar{x})$ defined in Section 4D. We focus on a class of nonsmooth functions defined next.

Semismooth functions. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be semismooth at $\bar{x} \in \operatorname{int} \operatorname{dom} f$ when it is Lipschitz continuous around $\bar{x}$, directionally differentiable in every direction, and for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$|f(x)-f(\bar{x})-A(x-\bar{x})| \le \varepsilon|x-\bar{x}| \quad \text{for every } x \in \mathbb{B}_\delta(\bar{x}) \text{ and every } \ A \in \bar{\partial} f(x).$$

Examples.

1) The function $f(x)=|x|$ is semismooth at 0 and smooth everywhere else.

2) The function $f : \mathbb{R}^2 \to \mathbb{R}$ defined as $f(x)=x_1+x_2-|x|$ is semismooth at zero. This function is known as the Fischer-Burmeister function and is used to numerically handle complementarity problems.

3) Every piecewise smooth function is semismooth in the interior of its domain.

4) A Lipschitz continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$ is semismooth at $\bar{x}$ if and only if for any $A \in \bar{\partial} f(\bar{x}+h)$ one has $Ah-f'(\bar{x};h)=o(|h|)$.

We consider the following version of Newton's method for solving (1):

$$f(x_k)+A_k(x_{k+1}-x_k)+F(x_{k+1}) \ni 0 \qquad \text{for } A_k \in \bar{\partial} f(x_k), \quad k=0,1,\ldots. \tag{3}$$

When the function f in (1) is semismooth, the method (3) is usually referred to as *semismooth* Newton's method.

Theorem 6F.1 (superlinear convergence of semismooth Newton method)**.** *Consider the method (3) applied to (1) with a solution $\bar{x}$ for a function f which is semismooth at $\bar{x}$ and assume that*

$$\begin{array}{c}\text{for every } A \in \bar{\partial} f(\bar{x}) \text{ the mapping } G_A \text{ defined in (2)} \\ \text{is strongly metrically regular at } \bar{x} \text{ for } 0.\end{array} \tag{4}$$

Then there exists a neighborhood O of $\bar{x}$ such that for every $x_0 \in O$ and $k=0,1,\ldots$ and for every $A_k \in \bar{\partial} f(x_k)$ there is a unique in O sequence $\{x_k\}$ generated by the *method (3) and this sequence converges to $\bar{x}$. In fact, the sequence generated in such a way is superlinearly convergent to $\bar{x}$.*

In preparation to proving Theorem 6F.1 we present a proposition, which shows that G_A^{-1} has a Lipschitz continuous single-valued localization *uniformly* in $A \in \bar{\partial} f(\bar{x})$, in the sense that the neighborhoods and the Lipschitz constant associated with the localization of G_A^{-1} are the same for all $A \in \bar{\partial} f(\bar{x})$.

Proposition 6F.2 (strong regularity for generalized Jacobian)**.** *For a solution $\bar{x}$ of (1) suppose that f is Lipschitz continuous around $\bar{x}$ and condition (4) is fulfilled. Then there exist constants α, β and ℓ such that for every $A \in \bar{\partial} f(\bar{x})$, the mapping $\mathbb{B}_\beta(0) \ni y \mapsto G_A^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$ is a Lipschitz continuous function with Lipschitz constant ℓ.*

Proof. Choose any $A \in \bar{\partial} f(\bar{x})$ as well as positive constants κ and κ' such that $\kappa' < \kappa$ and the mapping G_A^{-1} has a Lipschitz continuous localization around 0 for $\bar{x}$ with a Lipschitz constant κ' and neighborhoods $\mathbb{B}_a(\bar{x})$ and $\mathbb{B}_b(0)$ for some positive a and b. Make $b > 0$ smaller if necessary so that

$$b < 2a/\kappa'.$$

Let $\varepsilon > 0$ satisfy

$$\varepsilon < 1/\kappa', \quad \kappa'/(1-\kappa'\varepsilon) \le \kappa, \quad \varepsilon a \le b/2 \text{ and } \kappa' b/2 \le a(1-\kappa'\varepsilon).$$

Choose any $A' \in \bar{\partial} f(\bar{x})$ such that $|A - A'| < \varepsilon$. Now consider the mapping $G_{A'}$ associated with A' as in (2). For every $y \in \mathbb{B}_{b/2}(0)$ and every $x \in \mathbb{B}_a(\bar{x})$ we have

$$|y + (A - A')(x - \bar{x})| \le b/2 + \varepsilon a \le b.$$

Clearly,

$$x \in G_{A'}^{-1}(y) \cap \mathbb{B}_a(\bar{x}) \iff x \in \xi_y(x) := G_A^{-1}(y + (A - A')(x - \bar{x})) \cap \mathbb{B}_a(\bar{x}).$$

Let $y \in \mathbb{B}_{b/2}(0)$. We have

$$|\xi_y(\bar{x}) - \bar{x}| = |G_A^{-1}(y) \cap \mathbb{B}_a(\bar{x}) - G_A^{-1}(0) \cap \mathbb{B}_a(\bar{x})| \le \kappa' b/2 \le a(1-\kappa'\varepsilon).$$

Further, for any $x, x' \in \mathbb{B}_a(\bar{x})$ we obtain

$$|\xi_y(x) - \xi_y(x')| \le \kappa' |(A - A')(x - x')| \le \kappa' \varepsilon |x - x'|.$$

Thus, from the standard contraction principle 1A.2 the function $\mathbb{B}_a(\bar{x}) \ni x \mapsto \xi_y(x)$ has a unique fixed point $x(y)$ in $\mathbb{B}_a(\bar{x})$. Since $x(y) = G_{A'}^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ for every $y \in \mathbb{B}_{b/2}(0)$, we conclude that the mapping $\mathbb{B}_{b/2}(0) \ni y \mapsto x(y) = G_{A'}^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ is single-valued. Furthermore, for any $y, y' \in \mathbb{B}_{b/2}(0)$ we have

$$|x(y) - x(y')| = |\xi_y(x(y)) - \xi_{y'}(x(y'))|$$

$$\leq \kappa'|y-y'| + \kappa'|(A-A')(x(y)-x(y'))|$$
$$\leq \kappa'|y-y'| + \kappa'\varepsilon|x(y)-x(y')|,$$

hence $x(\cdot)$ is Lipschitz continuous on $\mathbb{B}_{b/2}(0)$ with Lipschitz constant $\kappa'/(1-\kappa'\varepsilon) \leq \kappa$. Thus, for a, b, ε, κ and κ' as above, and for any $A' \in \bar{\partial} f(\bar{x})$ with $A' \in \operatorname{int}\mathbb{B}_\varepsilon(A)$ the mapping $G_{A'}^{-1}$ has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$ with neighborhoods $\mathbb{B}_a(\bar{x})$ and $\mathbb{B}_{b/2}(0)$ and with Lipschitz constant κ. In other words, for any matrix A' which is sufficiently close to A the sizes of the neighborhoods and the Lipschitz constant associated with the localization of $G_{A'}^{-1}$ remain the same.

Pick any $A \in \bar{\partial} f(\bar{x})$ and corresponding a, b, κ and κ', and then $\varepsilon^A > 0$ to obtain that for every $A' \in \operatorname{int}\mathbb{B}_{\varepsilon^A}(A)$ the mapping $\mathbb{B}_{b/2}(0) \ni y \mapsto G_{A'}^{-1}(y) \cap \mathbb{B}_a(\bar{x})$ is a Lipschitz continuous function with Lipschitz constant κ. Since $\bar{\partial} f(\bar{x})$ is compact, from the open covering $\cup_{A\in\bar{\partial} f(\bar{x})} \operatorname{int}\mathbb{B}_{\varepsilon^A}(A)$ of $\bar{\partial} f(\bar{x})$ we can choose a finite subcovering with open balls $\operatorname{int}\mathbb{B}_{\varepsilon^{A_i}}(A_i)$; let a_i, b_i and κ_i be the constants associated with the Lipschitz localizations for $G_{A_i}^{-1}$. Taking $\alpha = \min_i a_i$, $\beta = \min_i b_i/2$ and $\ell = \max_i \kappa_i$ we terminate the proof. □

Proof of Theorem 6F.1. According to Proposition 6F.2, under condition (4) there exist constants α, β and ℓ such that for every $A \in \bar{\partial} f(\bar{x})$, the mapping $\mathbb{B}_\beta(0) \ni y \mapsto G_A^{-1}(y) \cap \mathbb{B}_\alpha(\bar{x})$ is a Lipschitz continuous function with Lipschitz constant ℓ. Pick positive ν and a such that

$$(5) \qquad \nu < \frac{1}{2}, \quad \ell\nu < \frac{1}{2} \quad \text{and} \quad a \leq \min\{\alpha, \beta\}.$$

Make a smaller if necessary so that, from the outer semicontinuity of the generalized Jacobian $\bar{\partial} f$, for any $x \in \mathbb{B}_a(\bar{x})$ and any $A \in \bar{\partial} f(x)$ there exists $\bar{A} \in \bar{\partial} f(\bar{x})$ such that

$$(6) \qquad |A - \bar{A}| \leq \nu,$$

and also, from the semismoothness of f, for any $x \in \mathbb{B}_a(\bar{x})$ and any $A \in \bar{\partial} f(x)$,

$$(7) \qquad |f(x) - f(\bar{x}) - A(x-\bar{x})| \leq \nu|x-\bar{x}|.$$

Let $x_0 \in \mathbb{B}_a(\bar{x})$ and let $A_0 \in \bar{\partial} f(x_0)$. Choose $\bar{A}_0 \in \bar{\partial} f(\bar{x})$ such that, according to (6), $|A_0 - \bar{A}_0| \leq \nu$. Then, for any $x \in \mathbb{B}_a(\bar{x})$, from (5), (6) and (7) we have

$$|f(\bar{x}) - f(x_0) + A_0(x_0-\bar{x}) + (\bar{A}_0 - A_0)(x-\bar{x})| \leq \nu a + \nu a < a \leq \beta.$$

Consider the function

$$\mathbb{B}_a(\bar{x}) \ni x \mapsto \xi_0(x) := G_{\bar{A}_0}^{-1}(f(\bar{x}) - f(x_0) + A_0(x_0-\bar{x}) + (\bar{A}_0 - A_0)(x-\bar{x})) \cap \mathbb{B}_a(\bar{x}).$$

Since

$$\xi_0(\bar{x}) = G_{\bar{A}_0}^{-1}(f(\bar{x}) - f(x_0) + A_0(x_0-\bar{x})) \cap \mathbb{B}_a(\bar{x}) \quad \text{and} \quad \bar{x} = G_{\bar{A}_0}^{-1}(0) \cap \mathbb{B}_a(\bar{x}),$$

we have from 6F.2, (5) and (7) that

$$|\xi_0(\bar{x}) - \bar{x}| \le \ell|f(\bar{x}) - f(x_0) + A_0(x_0 - \bar{x})| \le \ell\nu a < a(1-\ell\nu).$$

For any $x, x' \in \mathbb{B}_a(\bar{x})$ we obtain

$$|\xi_0(x) - \xi_0(x')| \le \ell|(\bar{A}_0 - A_0)(x - x')| \le \ell\nu|x - x'|.$$

Hence, by the standard contraction mapping theorem 1A.2, there exists a unique $x_1 \in \mathbb{B}_a(\bar{x})$ such that $x_1 = \xi_0(x_1)$. That is, there exists a unique $x_1 \in \mathbb{B}_a(\bar{x})$ which satisfies the iteration (3) for $k = 0$. Furthermore, using (7),

$$\begin{aligned} |x_1 - \bar{x}| &= |\xi_0(x_1) - \bar{x}| \\ &= |G_{\bar{A}_0}^{-1}(f(\bar{x}) - f(x_0) + A_0(x_0 - \bar{x}) + (\bar{A}_0 - A_0)(x_1 - \bar{x})) \cap \mathbb{B}_a(\bar{x}) \\ &\qquad - G_{\bar{A}_0}^{-1}(0) \cap \mathbb{B}_a(\bar{x})| \\ &\le \ell|f(\bar{x}) - f(x_0) + A_0(x_0 - \bar{x})| + \ell\nu|x_1 - \bar{x}|, \end{aligned}$$

which gives us

$$|x_1 - \bar{x}| \le \frac{\ell|f(\bar{x}) - f(x_0) + A_0(x_0 - \bar{x})|}{1 - \ell\nu} \le \frac{\ell\nu}{1-\ell\nu}|x_0 - \bar{x}|.$$

The induction step is completely analogous. Given $x_k \in \mathbb{B}_a(\bar{x})$ and $A_k \in \bar{\partial} f(x_k)$ we choose $\bar{A}_k \in \bar{\partial} f(\bar{x})$ such that $|A_k - \bar{A}_k| \le \nu$. Then we consider the function

$$\mathbb{B}_a(\bar{x}) \ni x \mapsto \xi_k(x) := G_{\bar{A}_k}^{-1}(f(\bar{x}) - f(x_k) + A_k(x_k - \bar{x}) + (\bar{A}_k - A_k)(x - \bar{x})) \cap \mathbb{B}_a(\bar{x}),$$

for which we show the existence of a unique x_{k+1} in $\mathbb{B}_a(\bar{x})$ satisfying the iteration (3) and

$$|x_{k+1} - \bar{x}| \le \frac{\ell|f(\bar{x}) - f(x_k) + A_k(x_k - \bar{x})|}{1 - \ell\nu} \le \frac{\ell\nu}{1-\ell\nu}|x_k - \bar{x}|.$$

Since $\ell\nu/(1-\ell\nu) < 1$ the sequence $\{x_k\}$ converges to $\bar{x}$. Let $O = \mathbb{B}_a(\bar{x})$.

We will show now that the sequence $\{x_k\}$ constructed in such a way is convergent superlinearly. Let $\varepsilon \in (0, \nu)$ and choose $\alpha \in (0, a)$ such that (6) and (7) hold with $\nu = \varepsilon$. Let k_0 be so large that $x_k \in \mathbb{B}_\alpha(\bar{x})$ and $x_k \ne \bar{x}$ for all $k \ge k_0$. Then

$$\begin{aligned} |x_{k+1} - \bar{x}| &= |G_{\bar{A}_k}^{-1}(f(\bar{x}) - f(x_k) + A_k(x_k - \bar{x}) + (\bar{A}_k - A_k)(x_{k+1} - \bar{x})) \cap \mathbb{B}_a(\bar{x}) \\ &\qquad - G_{\bar{A}_k}^{-1}(0) \cap \mathbb{B}_a(\bar{x})| \\ &\le \ell|f(\bar{x}) - f(x_k) + A_k(x_k - \bar{x})| + \ell\varepsilon|x_{k+1} - \bar{x}| \\ &\le \ell\varepsilon|x_k - \bar{x}| + \ell\varepsilon|x_{k+1} - \bar{x}|. \end{aligned}$$

Hence

$$\frac{|x_{k+1}-\bar{x}|}{|x_k-\bar{x}|} \le \frac{\ell\varepsilon}{1-\ell\varepsilon}.$$

Since ε can be arbitrarily small, we have superlinear convergence. □

Since every smooth function is semismooth and condition (4) reduces to strong metric regularity of the mapping $x \mapsto f(\bar{x})+Df(\bar{x})(x-\bar{x})+F(x)$ at $\bar{x}$ for 0, which is in turn equivalent to strong metric regularity of $f+F$ at $\bar{x}$ for 0, in finite dimensions Theorem 6C.1 is a particular case of 6F.1.

In further lines we utilize the following property of the generalized Jacobian.

Proposition 6F.3 (selection of generalized Jacobian)**.** *Let $f : I\!\!R^n \to I\!\!R^m$ be Lipschitz continuous around $\bar{x}$. Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x, x' \in I\!\!B_\delta(\bar{x})$ there exists $A \in \bar{\partial} f(\bar{x})$ with the property*

$$|f(x)-f(x')-A(x-x')| \le \varepsilon|x-x'|.$$

Proof. Let $\varepsilon > 0$. From the outer semicontinuity of $\bar{\partial} f$ there exists $\delta > 0$ such that

$$\bar{\partial} f(x) \subset \bar{\partial} f(\bar{x}) + \varepsilon I\!\!B_{m\times n} \quad \text{for all} \;\; x \in I\!\!B_\delta(\bar{x}).$$

Thus, for any $x, x' \in I\!\!B_\delta(\bar{x})$,

$$\bar{\partial} f(tx+(1-t)x') \subset \bar{\partial} f(\bar{x}) + \varepsilon I\!\!B_{m\times n}.$$

The set on the right side of this inclusion is convex and does not depend on t, hence

$$\text{co} \bigcup_{t\in[0,1]} \bar{\partial} f(tx+(1-t)x') \subset \bar{\partial} f(\bar{x}) + \varepsilon I\!\!B_{m\times n}.$$

But then, from the mean value theorem stated before 4D.3, there exists $A \in \bar{\partial} f(\bar{x})$ with the desired property. □

We are now also ready to present a proof of the inverse function theorem for nonsmooth generalized equations, which we left unproved in Section 4D.

Proof of Theorem 4D.4. Without loss of generality, let $\bar{y} = 0$. Then the assumption of 4D.4 becomes condition (4). According to 6F.2, there exist constants β and ℓ such that for every $A \in \bar{\partial} f(\bar{x})$, the mapping

$$I\!\!B_{2\beta}(0) \ni y \mapsto s_A := G_A^{-1}(y) \cap I\!\!B_{2\ell\beta}(\bar{x})$$

is a Lipschitz continuous function with Lipschitz constant ℓ. Let $L > \text{lip}\,(f;\bar{x})$. Adjust β if necessary so that f is Lipschitz continuous on $I\!\!B_{2\ell\beta}(\bar{x})$ with Lipschitz constant L and, from Proposition 3G.6, the set gph $F \cap (I\!\!B_{\ell\beta}(\bar{x}) \times I\!\!B_\beta(-f(\bar{x})))$ is closed. Since $\bar{\partial} f(\bar{x})$ is compact, there exists μ such that

$$\sup_{A\in\bar{\partial} f(\bar{x})} |A| \le \mu. \tag{8}$$

Define the function

$$\bar{\partial} f(\bar{x}) \times \mathbb{B}_\beta(0) \ni (A,z) \mapsto \varphi(A,z) := s_A(z) \cap \mathbb{B}_{\ell\beta}(\bar{x}).$$

Then φ has the following properties:

(a) dom $\varphi = \bar{\partial} f(\bar{x}) \times \mathbb{B}_\beta(0)$ thanks to the Lipschitz continuity of s_A with Lipschitz constant ℓ;

(b) For each $A \in \bar{\partial} f(\bar{x})$ the function $\varphi(A,\cdot) = s_A$ is Lipschitz continuous on $\mathbb{B}_\beta(0)$ with Lipschitz constant ℓ;

(c) For each $A \in \bar{\partial} f(\bar{x})$ one has $\varphi(A,0) = \bar{x} = s_A(0)$;

(d) φ is continuous in its domain.

Only (c) needs to be proved. Let $\{A_n\}$ be a sequence of matrices from $\bar{\partial} f(\bar{x})$ which is convergent to $\bar{A}$ and $\{z_n\}$ be a sequence of points from $\mathbb{B}_\beta(0)$ convergent to $\bar{z}$; then $\bar{A} \in \bar{\partial} f(\bar{x})$ and $\bar{z} \in \mathbb{B}_\beta(0)$. Set $\bar{u} = \varphi(\bar{A},\bar{z})$ and $u_n = \varphi(A_n,z_n)$ for each natural n. We have

$$z_n \in f(\bar{x}) + A_n(u_n - \bar{x}) + F(u_n),$$

that is,

$$f(\bar{x}) + \bar{A}(u_n - \bar{x}) + F(u_n) \ni z_n + (\bar{A} - A_n)(u_n - \bar{x}).$$

Since each $u_n \in \mathbb{B}_{\ell\beta}(\bar{x})$ and $A_n \to \bar{A}$ as $n \to \infty$, we obtain that

$$z_n + (\bar{A} - A_n)(u_n - \bar{x}) \in \mathbb{B}_{2\beta}(0) \quad \text{ for all } n \text{ sufficiently large.}$$

Then, using the definitions of u_n and $\bar{u}$, and property (b),

$$\begin{aligned} |u_n - \bar{u}| &= |s_{\bar{A}}(z_n + (\bar{A} - A_n)(u_n - \bar{x})) - s_{\bar{A}}(\bar{z})| \\ &\leq l|z_n - \bar{z}| + l^2\beta|\bar{A} - A_n| \to 0 \text{ as } n \to \infty. \end{aligned}$$

Thus (c) is established.

Our next step is to show that

$$\text{(9)} \qquad \begin{array}{c} \text{the mapping } (f+F)^{-1} \text{ has a nonempty-valued} \\ \text{graphical localization around 0 for } \bar{x}. \end{array}$$

In preparation for that, fix $\varepsilon \in (0, 1/\ell)$ and then apply Proposition 6F.3 to find δ such that for each two distinct points u and v in $\mathbb{B}_{3\delta}(\bar{x})$ there exists $A \in \bar{\partial} f(\bar{x})$ satisfying

$$\text{(10)} \qquad |f(v) - f(u) - A(v-u)| < \varepsilon|v-u|.$$

Adjust δ if necessary to satisfy

$$\text{(11)} \qquad 0 < 3\delta < \frac{\beta}{(1/\ell + L + \mu)}.$$

Clearly, $\delta < \ell\beta$. Set

$$\text{(12)} \qquad b := (1 - \ell\varepsilon)\delta, \quad \text{then} \quad b < \delta.$$

For any $y \in \mathbb{B}_{\varepsilon b}(0)$, $w \in \mathbb{B}_\delta(\bar{x})$, $\tilde{u} \in \mathbb{B}_\delta(\bar{x})$ and $A \in \bar{\partial} f(\bar{x})$ the relations (8), (11) and (12) yield the estimate

$$\begin{aligned}|y - f(w) + f(\bar{x}) + A(w - \tilde{u})| &\le |y| + L|w - \bar{x}| + \mu|w - \bar{x}| + \mu|\tilde{u} - \bar{x}| \\ &\le \varepsilon b + L\delta + 2\mu\delta < 2\delta(1/\ell + L + \mu) < 2\beta/3,\end{aligned}$$

hence,

$$\begin{aligned}(13)\qquad & y - f(w) + f(\bar{x}) + A(w - \tilde{u}) \in \mathbb{B}_{2\beta/3}(0) \\ & \qquad \text{whenever } (y, w, \tilde{u}, A) \in \mathbb{B}_{\varepsilon b}(0) \times \mathbb{B}_\delta(\bar{x}) \times \mathbb{B}_\delta(\bar{x}) \times \bar{\partial} f(\bar{x}).\end{aligned}$$

Let $y \in \mathbb{B}_{\varepsilon b}(0)$ be fixed. We will now find $x \in (f + F)^{-1}(y) \cap \mathbb{B}_\delta(\bar{x})$; this will prove (9). Denote $K = \mathbb{B}_{2\ell\varepsilon\delta}(0) \setminus \{0\}$. Fix $u \in \mathbb{B}_\delta(\bar{x})$ and define the function

$$(14)\qquad \bar{\partial} f(\bar{x}) \ni A \mapsto \Phi_u(A) = \varphi\big(A, y - f(u) + f(\bar{x}) + A(u - \bar{x})\big) - u.$$

By (13) with $\tilde{u} = \bar{x}$ and $w = u$ we obtain that $\operatorname{dom} \Phi_u = \bar{\partial} f(\bar{x})$. From the continuity of φ, for any $u \in \mathbb{B}_\delta(\bar{x})$ the function Φ_u is continuous in its domain. If there exist $A \in \bar{\partial} f(\bar{x})$ and $u \in \mathbb{B}_\delta(\bar{x})$ such that $\Phi_u(A) = 0$, then $x \in (f + F)^{-1}(y)$. If this is not the case, that is,

$$(15)\qquad \Phi_u(A) \neq 0 \text{ for all } A \in \bar{\partial} f(\bar{x}) \text{ and all } u \in \mathbb{B}_\delta(\bar{x}),$$

we will construct a sequence of points convergent to an $x \in (f + F)^{-1}(y)$. To this end, we make use of the following two lemmas:

Lemma 4D.1a. *Given $u \in \mathbb{B}_\delta(\bar{x})$, suppose that there exist $v \in \mathbb{B}_\delta(\bar{x}) \setminus \{u\}$ along with $\tilde{A} \in \bar{\partial} f(\bar{x})$ satisfying*

$$(16)\qquad |f(v) - f(u) - \tilde{A}(v - u)| \le \varepsilon|v - u| \;\text{ and }\; f(v) + \tilde{A}(u - v) + F(u) \ni y.$$

Then

$$(17)\qquad 0 < |\Phi_u(A)| \le \ell\varepsilon|u - v| \text{ whenever } A \in \bar{\partial} f(\bar{x}).$$

In particular, Φ_u maps $\bar{\partial} f(\bar{x})$ into K.

Proof. Pick any $A \in \bar{\partial} f(\bar{x})$. The first inequality follows from (15). For $z := y - f(v) + f(\bar{x}) + A(u - \bar{x}) - \tilde{A}(u - v)$, the inclusion (13) with $A = \tilde{A}$, $w = v$ and $\tilde{u} = u$ combined with (11) implies that

$$|z| \le |y - f(v) + f(\bar{x}) + \tilde{A}(v - u)| + |A(u - \bar{x})| \le 2\beta/3 + \mu\delta < 2\beta/3 + \beta/3 = \beta.$$

Rearranging the inclusion in (16) gives us

$$f(\bar{x}) + A(u - \bar{x}) + F(u) \ni y + f(\bar{x}) + A(u - \bar{x}) - f(v) - \tilde{A}(u - v) = z.$$

Therefore, $u = \varphi(A, z)$ and we conclude that

$$|\Phi_u(A)| = |\varphi\big(A, y - f(u) + f(\bar{x}) + A(u-\bar{x})\big) - \varphi(A,z)|$$
$$\leq \ell|f(v) - f(u) - \tilde{A}(v-u)| \leq \ell\varepsilon|u-v| \leq 2\ell\varepsilon\delta,$$

that is, (17) holds. ◻

Keeping $u \in \mathbb{B}_\delta(\bar{x})$ fixed define the following set-valued mapping acting from K into the subsets of $\bar{\partial} f(\bar{x})$:

(18) $$K \ni h \mapsto \Psi_u(h) = \left\{A \in \bar{\partial} f(\bar{x}) : \ |f(u+h) - f(u) - Ah| \leq \varepsilon|h|\right\}.$$

Lemma 4D.1b. *Given $u \in \mathbb{B}_\delta(\bar{x})$, suppose that Φ_u maps $\bar{\partial} f(\bar{x})$ into K. Then there exists a continuous selection ψ_u of the mapping Ψ_u such that the function defined as the composition $\psi_u \circ \Phi_u$ has a fixed point.*

Proof. Note that the values of Ψ_u are closed convex sets. Fix any $h \in K$. Since $\varepsilon < 1/\ell$, we get that $|u+h-\bar{x}| \leq |u-\bar{x}| + |h| \leq \delta + 2\ell\varepsilon\delta < 3\delta$. Hence u and $u+h$ are distinct elements of $\mathbb{B}_{3\delta}(\bar{x})$. Then (10) with $v = u+h$ implies that $\operatorname{dom}\Psi_u = K$. We show next that Ψ_u is inner semicontinuous. Fix any $h \in K$, and let Ω be an open set in $\bar{\partial} f(\bar{x})$ which meets $\Psi_u(h)$. Let $A \in \Psi_u(h) \cap \Omega$. According to (10), there is $\tilde{A} \in \bar{\partial} f(\bar{x})$ such that
$$|f(u+h) - f(u) - \tilde{A}h| < \varepsilon|h|.$$
Since Ω is open and $A \in \Omega$, there exists $\lambda \in (0,1)$ such that $A_\lambda := (1-\lambda)A + \lambda\tilde{A} \in \Omega$. Put
$$V = \{\tau \in K \mid |f(u+\tau) - f(u) - A_\lambda\tau| < \varepsilon|\tau|\}.$$
The estimate

$$\begin{aligned}|f(u+h) - f(u) - A_\lambda h| &\leq (1-\lambda)|f(u+h) - f(u) - Ah| \\ &\quad + \lambda|f(u+h) - f(u) - \tilde{A}h| \\ &< (1-\lambda)\varepsilon|h| + \lambda\varepsilon|h| = \varepsilon|h|,\end{aligned}$$

tells us that $h \in V$. Employing the continuity of f, we have that every τ sufficiently close to h belongs to V; hence V is a neighborhood of h in K. From the definitions of Ψ_u and V we get that $A_\lambda \in \Psi_u(w)$ for every $w \in V$, therefore $\Psi_u(w)$ intersects Ω for every $w \in V$. This proves that Ψ_u is inner semicontinuous. Michael's selection theorem 5J.5 yields the existence of a continuous selection ψ_u for Ψ_u, that is, a function acting from K into $\bar{\partial} f(\bar{x})$ which is continuous in its domain and has the property that if $w \in K$ and $A := \psi_u(w)$ then

(19) $$|f(u+w) - f(u) - Aw| \leq \varepsilon|w|.$$

Since $\bar{\partial} f(\bar{x})$ is a compact convex set, Φ_u is a continuous function, and by assumption Φ_u maps $\bar{\partial} f(\bar{x})$ into K, by Brouwer's fixed point theorem 1G.2 the composite mapping $\psi_u \circ \Phi_u$ acting from $\bar{\partial} f(\bar{x})$ into itself has a fixed point. ◻

In our next step, based on the above lemmas, we will construct sequences $\{x_n\}$ in $\mathbb{R}^n$ and $\{A_n\}$ in $\bar{\partial} f(\bar{x})$ whose entries have the following properties for each n:

(i) $|x_n - \bar{x}| < \delta$;
(ii) $0 < |x_{n+1} - x_n| \le (l\varepsilon)^n |x_1 - x_0|$;
(iii) $|f(x_{n+1}) - f(x_n) - A_n(x_{n+1} - x_n)| \le \varepsilon |x_{n+1} - x_n|$;
(iv) $f(x_n) + A_n(x_{n+1} - x_n) + F(x_{n+1}) \ni y$.

Clearly, $x_0 := \bar{x}$ satisfies (i) for $n = 0$. For any $A \in \bar{\partial} f(\bar{x})$, we have $\Phi_{x_0}(A) = \varphi(A, y) - x_0$. Using the properties (a) and (b) of the function φ along with (12) and (15), we moreover have that for any $A \in \bar{\partial} f(\bar{x})$,

$$0 < |\Phi_{x_0}(A)| = |\varphi(A, y) - \varphi(A, 0)| \le \ell |y| \le \ell \varepsilon b < \ell \varepsilon \delta < \delta.$$

Hence Φ_{x_0} maps $\bar{\partial} f(\bar{x})$ into K. According to Lemma 4D.1b, there exists a continuous selection ψ_{x_0} of the mapping Ψ_{x_0} such that the composite function $\psi_{x_0} \circ \Phi_{x_0}$ has a fixed point. Denote this fixed point by A_0; that is, $A_0 = \psi_{x_0}(\Phi_{x_0}(A_0)) \in \bar{\partial} f(\bar{x})$. Then

$$x_1 := x_0 + \Phi_{x_0}(A_0) = \varphi(A_0, y)$$

satisfies (i) with $n = 1$, as well as (ii) and (iv) with $n = 0$. Since $A_0 = \psi_{x_0}(x_1 - x_0)$, condition (iii) holds as well thanks to (19).

Further, we proceed by induction. Suppose that for some natural number $N > 0$ we have found x_{n+1} and A_n that satisfy conditions (i)–(iv) for all $n < N$. Set $v := x_{N-1}$ and $\tilde{A} = A_{N-1}$. Conditions (ii)–(iv) with $n = N - 1$ imply that the mapping Φ_{x_N} satisfies the assumption (16) of Lemma 4D.1a. Hence, by using lemmas 4D.1a and 4D.1b, we obtain that there exists $A_N \in \bar{\partial} f(\bar{x})$ such that $A_N = \psi_{x_N}(\Phi_{x_N}(A_N))$. Let

$$x_{N+1} = x_N + \Phi_{x_N}(A_N) = \varphi\big(A_N, y - f(x_N) + f(\bar{x}) + A_N(x_N - \bar{x})\big). \tag{20}$$

Then

$$f(\bar{x}) + A_N(x_{N+1} - \bar{x}) + F(x_{N+1}) \ni y - f(x_N) + f(\bar{x}) + A_N(x_N - \bar{x}).$$

This is (iv) for $n = N$. Since $A_N = \psi_{x_N}(x_{N+1} - x_N)$, (iii) holds for $n = N$. Combining (17), (20), and (ii) for $n = N - 1$, gives us (ii) for $n = N$. Furthermore, since

$$|x_1 - x_0| = |x_1 - \bar{x}| = |\Phi_{x_0}(A_0)| \le \ell \varepsilon b,$$

using (ii) and (12), we conclude that

$$|x_{N+1} - \bar{x}| \le \sum_{n=0}^{N} |x_{n+1} - x_n| < \frac{|x_1 - \bar{x}|}{1 - \ell\varepsilon} \le \frac{\ell \varepsilon b}{1 - \ell\varepsilon} = \ell\varepsilon\delta < \delta.$$

We arrive at (i) for $n = N + 1$. The induction step is complete.

By (ii), the sequence $\{x_n\}$ is a Cauchy sequence, hence it converges to some $x \in \mathbb{B}_\delta(\bar{x})$. For any n, from (i), both x_n and x_{n+1} are in $\mathbb{B}_\delta(\bar{x}) \subset \mathbb{B}_{\ell\beta}(\bar{x})$. Moreover, from (13) for $\tilde{u} = x_n$, $w = x_{n+1}$ and $A = A_n$, combined with (ii), we get

$$\begin{aligned}
&|y - f(x_n) + f(\bar{x}) + A_n(x_{n+1} - x_n)| \\
&\quad \le |y - f(x_{n+1}) + f(\bar{x}) + A_n(x_{n+1} - x_n)| + |f(x_{n+1}) - f(x_n)| \\
&\quad < 2\beta/3 + L|x_1 - x_0| < 2\beta/3 + L\delta < 2\beta/3 + \beta/3 = \beta.
\end{aligned}$$

Using (iv) we obtain

$$\big(x_{n+1}, y - f(x_n) - A_n(x_{n+1} - x_n)\big) \in \operatorname{gph} F \cap \big(\mathbb{B}_{\ell\beta}(\bar{x}) \times \mathbb{B}_\beta(-f(\bar{x}))\big).$$

Passing to the limit and remembering the set on the right is closed, we conclude that $f(x) + F(x) \ni y$, that is, $x \in (f+F)^{-1}(y) \cap \mathbb{B}_\delta(\bar{x})$. Since $y \in \mathbb{B}_{\varepsilon b}(0)$ was chosen arbitrarily, the mapping

$$\mathbb{B}_{\varepsilon b}(0) \ni y \mapsto \sigma(y) := (f+F)^{-1}(y) \cap \mathbb{B}_\delta(\bar{x})$$

is a nonempty-valued localization of $(f+F)^{-1}$. Thus, (9) is established.

It remains to show that σ is a Lipschitz continuous function. Choose any $y', y'' \in \mathbb{B}_{\varepsilon b}(0)$. Pick any $x' \in \sigma(y')$ and $x'' \in \sigma(y'')$. Then there exists $A \in \bar{\partial} f(\bar{x})$ such that

$$|f(x') - f(x'') - A(x' - x'')| \le \varepsilon |x' - x''|.$$

Using (13) we get that both

$$y' - f(x') + f(\bar{x}) + A(x' - \bar{x}) \ \text{ and } \ y'' - f(x'') + f(\bar{x}) + A(x'' - \bar{x})$$

are in $\mathbb{B}_\beta(0)$. Moreover, we have

$$x' = s_A(y' - f(x') + f(\bar{x}) + A(x' - \bar{x})) \text{ and } x'' = s_A(y'' - f(x'') + f(\bar{x}) + A(x'' - \bar{x})).$$

Taking the difference, we obtain

$$|x' - x''| \le \ell |y' - y''| + \ell |f(x') - f(x'') - A(x' - x'')| \le \ell |y' - y''| + \ell\varepsilon |x' - x''|.$$

This gives us

$$|x' - x''| \le \frac{\ell}{1 - \ell\varepsilon} |y' - y''|.$$

This shows that σ is both single-valued and Lipschitz continuous. □

In the reminder of this section we obtain a nonsmooth version of the Dennis–Moré theorem, a result we presented in the preceding Section 6E. We will use the following immediate corollary of Proposition 6F.3:

Corollary 6F.4. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be Lipschitz continuous around $\bar{x}$ and consider a sequence $\{x_k\}$ convergent to $\bar{x}$ and such that $x_k \ne \bar{x}$. Then there exists a sequence of matrices $A_k \in \bar{\partial} f(\bar{x})$ such that*

$$\lim_{k \to \infty} \frac{|f(x_{k+1}) - f(x_k) - A_k s_k|}{|s_k|} = 0. \tag{21}$$

As before, for a sequence $\{x_k\}$ convergent to $\bar{x}$, let $e_k = x_k - \bar{x}$ and $s_k = x_{k+1} - x_k$. Consider the following quasi-Newton method for solving (1):

$$f(x_k) + B_k(x_{k+1} - x_k) + F(x_{k+1}) \ni 0, \quad \text{for} \;\; k = 0, 1, \ldots, \tag{22}$$

where x_0 is given and B_k is a sequence of $n \times n$ matrices that are to be selected in a particular way. We will not discuss here how to construct a matrix B_k but, as in Section 6E, we will focus on the issue how close these matrices should be to a "derivative" of the function f to obtain superlinear convergence. In the following theorem we show that the Dennis–Moré type theorem can be extended to generalized equations involving functions that are merely Lipschitz continuous, where the derivative is replaced by a specially selected element of the generalized Jacobian.

Theorem 6F.5 (Dennis–Moré theorem for nonsmooth generalized equations)**.** *Let f in (1) be Lipschitz continuous in a neighborhood of the reference solution $\bar{x}$. Consider a sequence $\{x_k\}$ generated by the iteration (22) which is convergent to $\bar{x}$ and such that $x_k \neq \bar{x}$. Let $\{A_k\}$ be a sequence of matrices $A_k \in \bar{\partial} f(\bar{x})$ satisfying (21) whose existence is claimed in Corollary 6F.4, and let $E_k = B_k - A_k$.*

If $x_k \to \bar{x}$ superlinearly, then

$$\lim_{k\to\infty} \frac{d(0, f(\bar{x}) + E_k s_k + F(x_{k+1}))}{|s_k|} = 0. \tag{23}$$

Conversely, if the mapping $f + F$ is strongly metrically subregular at $\bar{x}$ for 0 and

$$\lim_{k\to\infty} \frac{|E_k s_k|}{|s_k|} = 0, \tag{24}$$

then $x_k \to \bar{x}$ superlinearly.

Proof. Let $x_k \to \bar{x}$ superlinearly and let $\varepsilon > 0$. In the proof of 6E.1 we showed that

$$\frac{|s_k|}{|e_k|} \to 1 \;\; \text{as} \;\; k \to \infty. \tag{25}$$

Then, for large k we get

$$|e_{k+1}| \leq \varepsilon |s_k| \quad \text{and} \quad |e_k| \leq 2|s_k|. \tag{26}$$

Adding and subtracting in (22), we have

$$f(\bar{x}) - f(x_{k+1}) + f(x_{k+1}) - f(x_k) - A_k s_k \in f(\bar{x}) + E_k s_k + F(x_{k+1}). \tag{27}$$

Let L be a Lipschitz constant of f near $\bar{x}$; then, from (26),

$$|f(\bar{x}) - f(x_{k+1})| \leq L|e_{k+1}| \leq L\varepsilon|s_k|. \tag{28}$$

From (21), for all sufficiently large k, we obtain

$$(29) \qquad |f(x_{k+1}) - f(x_k) - A_k s_k| \le \varepsilon |s_k|.$$

Using (28) and (29), we get

$$\begin{aligned}
&|f(\bar{x}) - f(x_{k+1}) + f(x_{k+1}) - f(x_k) - A_k s_k| \\
&\quad \le |f(\bar{x}) - f(x_{k+1})| + |f(x_{k+1}) - f(x_k) - A_k s_k| \le L\varepsilon |s_k| + \varepsilon |s_k|.
\end{aligned}$$

Then (27) yields

$$d(0, f(\bar{x}) + E_k s_k + F(x_{k+1})) \le (L+1)\varepsilon |s_k|.$$

Since ε can be arbitrarily small, we obtain (23).

Now, suppose that the mapping $f+F$ is strongly metrically subregular at the solution $\bar{x}$ for 0 and consider a sequence $\{x_k\}$ generated by (22) and convergent to $\bar{x}$ for a sequence of matrices $\{B_k\}$. Let $\{A_k\}$ be a sequence of matrices $A_k \in \bar{\partial} f(\bar{x})$ satisfying (21) and suppose that (24) holds. From the assumed strong subregularity, there exists a constant $\kappa > 0$ such that, for large k,

$$(30) \qquad |e_{k+1}| \le \kappa d(0, f(x_{k+1}) + F(x_{k+1})).$$

From (22) we have

$$f(x_{k+1}) - f(x_k) - A_k s_k - E_k s_k \in f(x_{k+1}) + F(x_{k+1}).$$

Hence, using (30),

$$|e_{k+1}| \le \kappa |-f(x_k) - A_k s_k - E_k s_k + f(x_{k+1})| \le \kappa |f(x_{k+1}) - f(x_k) - A_k s_k| + \kappa |E_k s_k|.$$

Let $\varepsilon \in (0, 1/2\kappa)$. From (24), for large k,

$$|E_k s_k| \le \varepsilon |s_k|.$$

Thus, from (29) and the last two estimates we obtain

$$|e_{k+1}| \le 2\kappa\varepsilon |s_k| \le 2\kappa\varepsilon |e_{k+1}| + 2\kappa\varepsilon |e_k|.$$

Hence, if $e_k \ne 0$ for all large k, we have

$$\frac{|e_{k+1}|}{|e_k|} \le \frac{2\kappa\varepsilon}{1 - 2\kappa\varepsilon}.$$

Since ε can be arbitrarily small this yields superlinear convergence. □

For the case of an equation, when $F \equiv 0$ in (1), we obtain from 6F.5 the following nonsmooth version of the Dennis–Moré theorem 6E.4:

Corollary 6F.6 (Dennis–Moré theorem for nonsmooth equations)**.** *Consider the equation $f(x) = 0$ with a solution $\bar{x}$, and let f be Lipschitz continuous in a neighborhood U of $\bar{x}$ and such that*

$$|x-\bar{x}| \le \kappa |f(x)| \qquad \text{for all } \; x \in U;$$

that is, f is strongly metrically subregular at $\bar{x}$ with constant $\kappa > 0$ and neighborhood U. Consider a sequence $\{x_k\}$ generated by the iteration

$$f(x_k) + B_k(x_{k+1} - x_k) = 0, \qquad \text{for } \; k = 0,1,\ldots,$$

which is convergent to $\bar{x}$ and such that $x_k \ne \bar{x}$. Let $\{A_k\}$ be a sequence of matrices $A_k \in \bar{\partial} f(\bar{x})$ satisfying (21) and let $E_k = B_k - A_k$. Then $x_k \to \bar{x}$ superlinearly if and only if

$$\text{(31)} \qquad \lim_{k\to\infty} \frac{|E_k s_k|}{|s_k|} = 0.$$

When the function f is continuously differentiable in a neighborhood of $\bar{x}$, 6F.6 becomes 6E.4 in finite dimensions.

If the function f is not only Lipschitz continuous but also semismooth, it turns out that the necessary condition (23) for superlinear convergence is valid for *any* choice of matrices $A_k \in \bar{\partial} f(x_k)$. The proof of this statement is left to the reader as the following exercise:

Exercise 6F.7. *Consider a function f which is semismooth at $\bar{x}$ and a sequence $\{x_k\}$ generated by (22) which converges to $\bar{x}$ superlinearly. Prove that for every sequence of matrices $\{A_k\}$ such that $A_k \in \bar{\partial} f(x_k)$ for all k, condition (23) holds with $E_k = B_k - A_k$.*

Guide. Repeat the proof of 6F.5 until formula (27) and write instead

$$\text{(32)} \qquad f(\bar{x}) - f(x_k) + A_k e_k - A_k e_{k+1} \in f(\bar{x}) + E_k s_k + F(x_{k+1}).$$

Since the generalized Jacobian $\bar{\partial} f$ is outer semicontinuous and compact-valued, the sequence $\{A_k\}$ is bounded, say, by a constant λ. Then use the semismoothness of f and (26) to obtain

$$\begin{aligned} |f(x_k) - f(\bar{x}) - A_k e_k + A_k e_{k+1}| &\le |f(x_k) - f(\bar{x}) - A_k e_k| + |A_k||e_{k+1}| \\ &\le \varepsilon |e_k| + \lambda |e_{k+1}| \le (\lambda + 2)\varepsilon |s_k|. \end{aligned}$$

The inclusion (32) then implies

$$d(0, f(\bar{x}) + E_k s_k + F(x_{k+1})) \le (\lambda + 2)\varepsilon |s_k|$$

for all sufficiently large k. □

Exercise 6F.8. *State and prove a version of 6F.1 on the assumption that each G_A is metrically regular, but not necessarily strongly metrically regular.*

Guide. Combine the arguments in the proofs of 6D.2 and 6F.1 □

6G. Uniform Strong Metric Regularity and Path-Following

In the section we consider the generalized equation

$$(1) \qquad f(t,u)+F(u) \ni 0,$$

where the function f now depends on a scalar parameter $t \in [0,1]$. Throughout, $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^m$ is twice continuously differentiable everywhere (for simplicity) and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping with closed graph. The equation case corresponds to $F \equiv 0$ while for $F = N_C$, the normal cone mapping to a convex and closed set C we obtain a variational inequality. The generally set-valued mapping

$$S : t \mapsto S(t) = \left\{ u \in \mathbb{R}^n \,\middle|\, f(t,u)+F(u) \ni 0 \right\}$$

is the solution mapping associated with (1), and a *solution trajectory* over $[0,1]$ is in this case a function $\bar{u}(\cdot)$ such that $\bar{u}(t) \in S(t)$ for all $t \in [0,1]$, that is, $\bar{u}(\cdot)$ is a *selection* for S over $[0,1]$. Clearly, the map S has closed graph.

For any given $(t,u) \in \operatorname{gph} S$, the graph of the solution mapping of (1), define the mapping

$$(2) \qquad v \mapsto G_{t,u}(v) := f(t,u)+\nabla_u f(t,u)(v-u)+F(v).$$

A point $(t,u) \in \mathbb{R}^{1+n}$ is said to be a *strongly regular point* for the generalized equation (1) when $(t,u) \in \operatorname{gph} S$ and the mapping $G_{t,u}$ is strongly metrically regular at u for 0. Then, from 2B.7 we obtain that when $(\bar{t},\bar{u})$ is a strongly regular point for (1), then there are open neighborhoods T of $\bar{t}$ and U of $\bar{u}$ such that the mapping $T \cap [0,1] \ni \tau \mapsto S(\tau) \cap U$ is single-valued and Lipschitz continuous on $T \cap [0,1]$.

In the theorem which follows we show that if each point in $\operatorname{gph} S$ is strongly regular then there are finitely many solution trajectories; moreover, each of these trajectories is Lipschitz continuous on $[0,1]$ and their graphs never intersect each other. In addition, along any such trajectory $\bar{u}(\cdot)$ the mapping $G_{t,\bar{u}(t)}$ is strongly regular *uniformly* in $t \in [0,1]$, meaning that the neighborhoods and the constant involved in the definition do not depend on t.

Theorem 6G.1 (uniform strong metric regularity). *Suppose that there exists a bounded set $C \subset \mathbb{R}^n$ such that for each $t \in [0,1]$ the set $S(t)$ is nonempty and contained in C for all $t \in [0,1]$. Also, suppose that each point in $\operatorname{gph} S$ is strongly regular. Then there are finitely many Lipschitz continuous functions $\bar{u}_j : [0,1] \to \mathbb{R}^n$, $j = 1,2,\cdots,M$ such that for each $t \in [0,1]$ one has $S(t) = \cup_{1 \le j \le M}\{\bar{u}_j(t)\}$; moreover, the graphs of the functions $\bar{u}_j$ are isolated from each other, in the sense that there exists $\delta > 0$ such that*

$$|u_{j'}(t) - u_j(t)| \ge \delta \quad \text{for every } j' \neq j \text{ and every } t \in [0,1].$$

Also, there exist positive constants a, b and λ such that for each such function $\bar{u}_i$ and for each $t \in [0,1]$ the mapping

$$(3) \qquad \mathbb{B}_b(0) \ni w \mapsto G^{-1}_{t,\bar{u}_i(t)}(w) \cap \mathbb{B}_a(\bar{u}_i(t))$$

is a Lipschitz continuous function with a Lipschitz constant λ.

Proof. From the assumed uniform boundedness of the solution mapping S and the continuity of f and its derivatives, we get, for use in what follows, the existence of a constant $K > 0$ such that

$$(4) \qquad \sup_{t\in[0,1],v\in C} (|\nabla_t f(t,v)| + |\nabla_u f(t,v)| + |\nabla^2_{uu} f(t,v)| + |\nabla^2_{ut} f(t,v)|) \le K.$$

Let $(t,v) \in \operatorname{gph} S$. Then, according to 2B.7 there exists a neighborhood $T_{t,v}$ of t which is open relative to $[0,1]$ and an open neighborhood $U_{t,v}$ of v such that the mapping $T_{t,v} \ni \tau \mapsto S(\tau) \cap U_{t,v}$ is a function, denoted $u_{t,v}(\cdot)$, which is Lipschitz continuous on $T_{t,v}$ with Lipschitz constant $L_{t,v}$. From the open covering $\{T_{t,v} \times U_{t,v}\}_{(t,v)\in \operatorname{gph} S}$ of the graph of S, which is a compact set in $\mathbb{R}^{1+n}$, we extract a finite subcovering $\{T_{t_j,v_j} \times U_{t_j,v_j}\}_{j=1}^M$. Let $L = \max_{1\le j\le M} L_{t_j,v_j}$.

Let $\tau \in [0,1]$ and choose any $\bar{u} \in S(\tau)$. Now we will prove that there exists a Lipschitz continuous function $\bar{u}(\cdot)$ with Lipschitz constant L such that $\bar{u}(t) \in S(t)$ for all $t \in [0,1]$ and also $\bar{u}(\tau) = \bar{u}$.

Assume $\tau < 1$. Then there exists $j \in \{1,\cdots,M\}$ such that $(\tau,\bar{u}) \in T_{t_j,v_j} \times U_{t_j,v_j}$. Define $\bar{u}(t) = u_{t_j,v_j}(t)$ for all $t \in (t'_j, t''_j) := T_{t_j,v_j}$. Then $\bar{u}(\tau) = \bar{u}$ and $\bar{u}(\cdot)$ is Lipschitz continuous on $[t'_j, t''_j]$. If $t''_j < 1$ then there exists some $i \in \{1,\cdots,M\}$ such that $(t''_j, \bar{u}(t''_j)) \in T_{t_i,v_i} \times U_{t_i,v_i} := (t'_i, t''_i) \times U_{t_i,v_i}$. Then of course $u_{t_i,v_i}(t''_j) = \bar{u}(t''_j)$ and we can extend $\bar{u}(\cdot)$ to $[t_j, t_{i''}]$ as $\bar{u}(t) = u_{t_i,v_i}(t)$ for $t \in [t_{j'}, t_{i''}]$. After at most M such steps we extend $\bar{u}(\cdot)$ to $[t_{j'}, 1]$. By repeating the same argument on the interval $[0,\tau]$ we extend $\bar{u}(\cdot)$ on the entire interval $[0,1]$ thus obtaining a Lipschitz continuous selection for S. It $\tau = 1$ then we repeat the same argument on $[0,1]$ starting from 1 and going to the left.

Now assume that $(\tau,\bar{u})$ and $(\theta,\tilde{u})$ are two points in gph S and let $\bar{u}(\cdot)$ and $\tilde{u}(\cdot)$ be the functions determined by the above procedure such that $\bar{u}(\tau) = \bar{u}$ and $\tilde{u}(\theta) = \tilde{u}$. Assume that $\bar{u}(0) \ne \tilde{u}(0)$ and the set $\Delta := \{t \in [0,1] \,\big|\, \bar{u}(t) = \tilde{u}(t)\}$ is nonempty. Since Δ is closed, $\inf \Delta := \nu > 0$ is attained and then we have that $\bar{u}(\nu) = \tilde{u}(\nu)$ and $\bar{u}(t) \ne \tilde{u}(t)$ for $t \in [0,\nu)$. But then $(\nu, \bar{u}(\nu)) \in \operatorname{gph} S$ cannot be a strongly regular point of S, a contradiction. Thus, the number of different Lipschitz continuous functions $\bar{u}(\cdot)$ constructed from points $(\tau,\bar{u}) \in \operatorname{gph} S$ is not more than the number of points in $S(0)$. Hence there are finitely many Lipschitz continuous functions $\bar{u}_j(\cdot)$ such that for every $t \in [0,1]$ one has $S(t) = \cup_j \{\bar{u}_j(t)\}$. This proves the first part of the theorem.

Choose a Lipschitz continuous function $\bar{u}(\cdot)$ whose values are in the set of values of S, that is, $\bar{u}(\cdot)$ is one of the functions $\bar{u}_j(\cdot)$ and its Lipschitz constant is L. Let $t \in (0,1)$ and let $G_t = G_{t,\bar{u}(t)}$, for simplicity. Let a_t, b_t and λ_t be positive constants such that the mapping

$$(5) \qquad \mathbb{B}_{b_t}(0) \ni w \mapsto G_t^{-1}(w) \cap \mathbb{B}_{a_t}(\bar{u}(t))$$

is a Lipschitz continuous function with Lipschitz constant λ_t. Make $b_t > 0$ smaller if necessary so that

$$(6) \qquad 2b_t\lambda_t < a_t.$$

Let $\rho_t > 0$ be such that $L\rho_t < a_t/2$. Then, from the Lipschitz continuity of $\bar{u}$ around t we have that $\mathbb{B}_{a_t/2}(\bar{u}(\tau)) \subset \mathbb{B}_{a_t}(\bar{u}(t))$ for all $\tau \in (t-\rho_t, t+\rho_t)$. Make $\rho_t > 0$ smaller if necessary so that

$$(7) \qquad K(L+1)\rho_t < 1/\lambda_t.$$

We will now apply (the strong regularity part of) Theorem 5G.3 to show that there exist a neighborhood O_t of t and positive constants α_t and β_t such that for each $\tau \in O_t \cap [0,1]$ the mapping

$$(8) \qquad \mathbb{B}_{\beta_t}(0) \ni w \mapsto G_\tau^{-1}(w) \cap \mathbb{B}_{\alpha_t}(\bar{u}(t))$$

is a Lipschitz continuous function. Consider the function

$$\begin{aligned} g_{t,\tau}(v) = {} & f(t,\bar{u}(t)) - f(\tau,\bar{u}(\tau)) \\ & + (\nabla_u f(t,\bar{u}(t)) - \nabla_u f(\tau,\bar{u}(\tau)))v \\ & - \nabla_u f(t,\bar{u}(t))\bar{u}(t) + \nabla_u f(\tau,\bar{u}(\tau))\bar{u}(\tau). \end{aligned}$$

Note that the Lipschitz constant of $g_{t,\tau}$ is bounded by the expression on the left of (7). For each v we have

$$G_t(v) = G_\tau(v) + g_{t,\tau}(v).$$

We apply Theorem 5G.3 with $F = G_t$, $\bar{x} = \bar{u}(t)$, $\bar{y} = 0$, $g = g_{t,\tau}$, $2a = a_t$, $b = b_t$, $\kappa = \lambda_t$, $\mu = \mu_t := K(L+1)\rho_t$, and

$$(9) \qquad \kappa' = \lambda_t' := \frac{3\lambda_t}{2(1-K(L+1)\rho_t\lambda_t)} > \frac{\lambda_t}{1-\lambda_t\mu_t}.$$

For that purpose we need to show that there exist constants α_t and β_t that satisfy the inequalities

$$(10) \qquad \alpha_t \le a_t/2, \;\; \mu_t\alpha_t + 2\beta_t \le b_t, \;\; 2\lambda_t'\beta_t \le \alpha_t(1-\lambda_t\mu_t), \;\; |g_{t,\tau}(\bar{u}(t))| \le \beta_t.$$

Elementary calculus gives us

$$\begin{aligned} g_{t,\tau}(\bar{u}(t)) &= \int_0^1 \frac{d}{ds} f(\tau + s(t-\tau), \bar{u}(\tau) + s(\bar{u}(t) - \bar{u}(\tau)))ds \\ &\quad - \nabla_u f(\tau,\bar{u}(\tau))(\bar{u}(t) - \bar{u}(\tau)) \\ &= \int_0^1 (t-\tau)\nabla_t f(\tau + s(t-\tau), \bar{u}(\tau) + s(\bar{u}(t) - \bar{u}(\tau)))ds \\ &\quad + \int_0^1 \Big(\nabla_u f(\tau + s(t-\tau), \bar{u}(\tau) + s(\bar{u}(t) - \bar{u}(\tau))) \end{aligned}$$

$$-\nabla_u f(\tau,\bar{u}(\tau))\Big)(\bar{u}(t)-\bar{u}(\tau))ds$$

which yields

$$|g_{t,\tau}(\bar{u}(t))| \leq K\rho_t + \frac{1}{2}KL\rho_t^2 + \frac{1}{2}KL^2\rho_t^2.$$

Choose ρ_t smaller if necessary such that $\frac{1}{2}L\rho_t + L^2\rho_t < 1$; then

$$|g_{t,\tau}(\bar{u}(t))| \leq 2K\rho_t.$$

Denoting $A := K(1+L)$ and $B := 2K$ we have

$$\mu_t = A\rho_t \quad \text{and} \quad |g_{t,\tau}(\bar{u}(t))| \leq B\rho_t.$$

Set $\beta_t := B\rho_t$. We will now show that there exists a positive α_t which satisfies all inequalities in (10) and also

$$\lambda_t'\beta_t < \alpha_t. \tag{11}$$

Substituting the already chosen μ_t and β_t in (10) we obtain that α_t should satisfy

$$\begin{cases} \alpha_t \leq a_t/2, \\ A\rho_t\alpha_t + 2B\rho_t \leq b_t, \\ 2\lambda_t B\rho_t \leq \alpha_t(1-\lambda_t A\rho_t). \end{cases} \tag{12}$$

System (12) has a solution $\alpha_t > 0$ provided that

$$\frac{2\lambda_t B\rho_t}{1-\lambda_t A\rho_t} \leq \frac{b_t - 2B\rho_t}{A\rho_t} \quad \text{and} \quad \frac{b_t - 2B\rho_t}{A\rho_t} \leq a_t/2.$$

Thus, everything comes down to checking whether this system of inequalities is consistent. But this system is consistent whenever

$$(2B + b_t\lambda_t A)\rho_t \leq b_t \leq (2B + Aa_t/2)\rho_t,$$

which in turn always holds because of (6). Hence, there exist α_t satisfying (12). Moreover, using (9) and the third inequality in (10) we obtain

$$\lambda_t'\beta_t = \frac{3}{2}\,\frac{\beta_t\lambda_t}{1-\lambda_t\mu_t} < \frac{2\beta_t\lambda_t}{1-\lambda_t\mu_t} \leq \alpha_t,$$

hence (11) holds.

We are now ready to apply Theorem 5G.3 from which we conclude that the mapping in (8) is a Lipschitz continuous function with Lipschitz constant λ_t'. From the open covering $\cup_{t\in[0,1]}(t-\rho_t, t+\rho_t)$ of $[0,1]$ choose a finite subcovering of open intervals $(t_i-\rho_{t_i}, t_i+\rho_{t_i})$, $i = 1,2,\ldots,m$. Let $a = \min\{\alpha_{t_i} \mid i = 1,\ldots,m\}$, $b = \min\{\beta_{t_i} \mid i = 1,\ldots,m\}$ and $\lambda = \max\{\lambda_{t_i}' \mid i = 1,\ldots,m\}$.

We will now use the following fact which is a straightforward consequence of the definition of strong regularity: Let a mapping H be strongly regular at $\bar{x}$ for $\bar{y}$ with a Lipschitz constant κ and neighborhoods $\mathbb{B}_a(\bar{x})$ and $\mathbb{B}_b(\bar{y})$. Then for every positive constants $a' \le a$ and $b' \le b$ such that $\kappa b' \le a'$, the mapping H is strongly regular with a Lipschitz constant κ and neighborhoods $\mathbb{B}_{a'}(\bar{x})$ and $\mathbb{B}_{b'}(\bar{y})$. Indeed, in this case any $y \in \mathbb{B}_{b'}(\bar{y})$ will be in the domain of $H^{-1}(\cdot) \cap \mathbb{B}_{a'}(\bar{x})$.

From (11) we get $b \le a/\lambda$; then the above observation applies, hence for each $\tau \in (t_i - \rho_{t_i}, t_i + \rho_{t_i}) \cap [0,1]$ the mapping $\mathbb{B}_b(0) \ni w \mapsto G_\tau^{-1}(w) \cap \mathbb{B}_a(\bar{u}(\tau))$ is a Lipschitz continuous function with Lipschitz constant λ. Let $t \in [0,1]$; then $t \in (t_i - \rho_{t_i}, t_i + \rho_{t_i})$ for some $i \in \{1, \dots, m\}$. Hence the mapping $\mathbb{B}_b(0) \ni w \mapsto G_t^{-1}(w) \cap \mathbb{B}_a(\bar{u}(t))$ is a Lipschitz continuous function with Lipschitz constant λ. The proof is complete. □

In the second part of this section we introduce and study, in this setting of the parameterized generalized equation (1), a method of Euler-Newton type. This method is a straightforward extension of the standard Euler-Newton continuation, or path-following, for solving equations of the form $f(t,u) = 0$ obtained from (1) by simply taking $F \equiv 0$ and $m = n$. That standard scheme is a predictor-corrector method of the following kind. For $N > 1$, let $\{t_i\}_{i=0}^N$ with $t_0 = 0$, $t_N = 1$, be a uniform (for simplicity) grid on $[0,1]$ with step size $h = t_{i+1} - t_i = 1/N$ for $i = 0, 1, \dots, N-1$. Starting from a solution u_0 to $f(0,u) = 0$, the method iterates between an Euler predictor step and a Newton corrector step:

$$(13) \qquad \begin{cases} v_{i+1} = u_i - h\nabla_u f(t_i, u_i)^{-1} \nabla_t f(t_i, u_i), \\ u_{i+1} = v_{i+1} - \nabla_u f(t_{i+1}, v_{i+1})^{-1} f(t_{i+1}, v_{i+1}). \end{cases}$$

Here we propose an extension of the Euler-Newton continuation method to the generalized equation (1), in which both the predictor and corrector steps consist of solving linearized generalized equations:

$$(14) \qquad \begin{cases} f(t_i, u_i) + h\nabla_t f(t_i, u_i) + \nabla_u f(t_i, u_i)(v_{i+1} - u_i) + F(v_{i+1}) \ni 0, \\ f(t_{i+1}, v_{i+1}) + \nabla_u f(t_{i+1}, v_{i+1})(u_{i+1} - v_{i+1}) + F(u_{i+1}) \ni 0. \end{cases}$$

Observe that the Euler step of iteration (14) does not reduce to the Euler step in (13) when $F \equiv 0$. However, as we see later, it reduces to a method which gives the same order of error.

According to Theorem 6G.1, the graph of the solution mapping S consists of the graphs of finitely many Lipschitz continuous functions which are isolated from each other; let L be a Lipschitz constant for all such functions. Choose any of these functions and call it $\bar{u}(\cdot)$.

Theorem 6G.2 (convergence of Euler-Newton path-following)**.** *Suppose that each point in* gph S *is strongly regular and let $\bar{u}$ be a solution trajectory. Let $u_0 = \bar{u}(0)$. Then there exist positive constants c and α and a natural N_0 such that for any natural $N \ge N_0$ the iteration (14) with $h = 1/N$ generates a unique sequence $\{u_i\}$ starting from u_0 and such that $u_i \in \mathbb{B}_\alpha(\bar{u}(t_i))$ for $i = 0, 1, \dots, N$. Moreover, this sequence satisfies*

$$\max_{0\le i\le N}|u_i-\bar u(t_i)|\le ch^4.$$

Proof. We already know from 6G.1 that there exist positive a, b and κ such that, for each $i=0,1,\dots,N-1$, the mapping

$$\mathbb{B}_b(0)\ni w\mapsto G_{t_i}^{-1}(w)\cap\mathbb{B}_a(\bar u(t_i))$$

is a Lipschitz continuous function with Lipschitz constant κ, where we recall that $G_{t_i}=G_{t_i,\bar u(t_i)}$ for $G_{t,u}$ given in (2). Let κ', μ, α and β be chosen according to 5G(9) in the statement of Theorem 5G.3. Let K be as in (4), let

$$c:=\frac{K^3\kappa'^3}{2}(1+L+L^2)^2, \tag{15}$$

and chose N_0 so large that for $h=1/N$ with $N\ge N_0$ the following inequalities hold:

$$ch^3(2+2L+ch^3)\le 1+L^2, \tag{16}$$

$$Kh(1+L+ch^3)\le\mu,\qquad Kh^2\left(1+L+L^2\right)\le\beta, \tag{17}$$

$$\kappa'K^2h^2(1+L+L^2)\le\mu,\qquad ch^4\le\alpha. \tag{18}$$

To prove the theorem we use induction. First, for $i=0$ we have $u_0=\bar u(t_0)$ and there is nothing more to prove. Let for $j=1,2,\cdots,i$ the iterates u_j be already generated by (14) uniquely in $\mathbb{B}_\alpha(\bar u(t_j))$ and in such a way that

$$|u_j-\bar u(t_j)|\le ch^4\ \text{ for all }\ j=1,2,\cdots,i.$$

We will prove that (14) determines a unique $u_{i+1}\in\mathbb{B}_\alpha(\bar u(t_{i+1}))$ which satisfies

$$|u_{i+1}-\bar u(t_{i+1})|\le ch^4. \tag{19}$$

We start with the Euler step. The generalized equation

$$f(t_i,u_i)+\nabla_u f(t_i,u_i)(v-u_i)+h\nabla_t f(t_i,u_i)+F(v)\ni 0 \tag{20}$$

for any $v\in\mathbb{R}^n$ can be written as

$$g(v)+G_{t_{i+1}}(v)\ni 0, \tag{21}$$

where, as before, $G_t=G_{t,\bar u(t)}$ with $G_{t,u}$ defined in (2), and

$$\begin{aligned}g(v)&=f(t_i,u_i)+\nabla_u f(t_i,u_i)(v-u_i)+h\nabla_t f(t_i,u_i)\\&\quad-[f(t_{i+1},\bar u(t_{i+1}))+\nabla_u f(t_{i+1},\bar u(t_{i+1}))(v-\bar u(t_{i+1}))].\end{aligned}$$

For any v, $v'\in\mathbb{R}^n$ we have

$$\begin{aligned}|g(v)-g(v')| &= |[\nabla_u f(t_i,u_i)-\nabla_u f(t_{i+1},\bar{u}(t_{i+1}))](v-v')|\\ &\le K\,(h+|u_i-\bar{u}(t_{i+1})|)\,|v-v'|\\ &\le K(h+|u_i-\bar{u}(t_i)|+|\bar{u}(t_i)-\bar{u}(t_{i+1})|)\,|v-v'|\\ &\le K(h+ch^4+Lh)|v-v'|\le \mu|v-v'|,\end{aligned}$$

where we use the first inequality in (17). Furthermore,

$$\begin{aligned}&|g(\bar{u}(t_{i+1}))|\\ &\le \left|\int_0^1 h\,[\nabla_t f(t_i+sh,u_i+s(\bar{u}(t_{i+1})-u_i))-\nabla_t f(t_i,u_i)]\,ds\right|\\ &\qquad+\left|\int_0^1 [\nabla_u f(t_i+sh,u_i+s(\bar{u}(t_{i+1})-u_i))-\nabla_u f(t_i,u_i)]\,(\bar{u}(t_{i+1})-u_i))\,ds\right|\\ &\le \int_0^1 K(sh^2+sh|\bar{u}(t_{i+1})-u_i|)\,ds+\int_0^1 K(sh|\bar{u}(t_{i+1})-u_i|+s|\bar{u}(t_{i+1})-u_i|^2)ds\\ &\le \frac{Kh^2}{2}+\frac{Kh}{2}|\bar{u}(t_{i+1})-u_i|+\frac{Kh}{2}|\bar{u}(t_{i+1})-u_i|+\frac{K}{2}|\bar{u}(t_{i+1})-u_i|^2\\ &\le \frac{Kh^2}{2}+Kh(|\bar{u}(t_{i+1})-\bar{u}(t_i)|+|\bar{u}(t_i)-u_i|)\\ &\qquad+\frac{K}{2}(|\bar{u}(t_{i+1})-\bar{u}(t_i)|+|\bar{u}(t_i)-u_i|)^2\\ &\le \frac{K}{2}(h^2+2h(Lh+ch^4)+(Lh+ch^4)^2)\le Kh^2(1+L+L^2),\end{aligned}$$

where in the last inequality we use (16). This implies that $|g(\bar{u}(t_{i+1}))|\le\beta$ due to the second relation in (17). Applying Theorem 5G.3 we obtain the existence of a unique in $\mathbb{B}_\alpha(\bar{u}(t_{i+1}))$ solution v_{i+1} of (21), hence of (20), and moreover the function

$$\mathbb{B}_\beta(0)\ni y\mapsto \xi(y):=(g+G_{t_{i+1}})^{-1}(y)\cap \mathbb{B}_\alpha(\bar{u}(t_{i+1}))$$

is Lipschitz continuous on $\mathbb{B}_\beta(0)$ with Lipschitz constant κ'. Observe that $v_{i+1}=\xi(0)$ and $\bar{u}(t_{i+1})=\xi(g(\bar{u}(t_{i+1})))$; then

$$\begin{aligned}|v_{i+1}-\bar{u}(t_{i+1})| &= |\xi(0)-\xi(g(\bar{u}(t_{i+1})))|\\ &\le \kappa'|g(\bar{u}(t_{i+1}))|\le \kappa' Kh^2\left(1+L+L^2\right).\end{aligned}\tag{22}$$

The Newton step solves the generalized equation

$$f(t_{i+1},v_{i+1})+\nabla_u f(t_{i+1},v_{i+1})(u-v_{i+1})+F(u)\ni 0,\tag{23}$$

which can be rewritten as

$$h(u)+G_{t_{i+1}}(u)\ni 0,$$

where

$$h(u)\,=\,f(t_{i+1},v_{i+1})+\nabla_u f(t_{i+1},v_{i+1})(u-v_{i+1})$$

$$-[f(t_{i+1},\bar{u}(t_{i+1}))+\nabla_u f(t_{i+1},\bar{u}(t_{i+1}))(u-\bar{u}(t_{i+1}))].$$

For any $u, u' \in \mathbb{R}^n$ we have

$$\begin{aligned}|h(u)-h(u')| &= |(\nabla_u f(t_{i+1},v_{i+1})-\nabla_u f(t_{i+1},\bar{u}(t_{i+1}))(u-u')| \\ &\le K|v_{i+1}-\bar{u}(t_{i+1})||u-u'| \\ &\le \kappa' K^2 h^2(1+L+L^2)|u-u'| \le \mu|u-u'|,\end{aligned}$$

where we use (22) and the first inequality in (18). Moreover,

$$\begin{aligned}&|h(\bar{u}(t_{i+1}))| \\ &= |f(t_{i+1},v_{i+1})+\nabla_u f(t_{i+1},v_{i+1})(\bar{u}(t_{i+1})-v_{i+1})-f(t_{i+1},\bar{u}(t_{i+1}))| \\ &= \left|\int_0^1 \frac{d}{ds} f(t_{i+1},v_{i+1}+s(\bar{u}(t_{i+1})-v_{i+1}))\,ds-\nabla_u f(t_{i+1},v_{i+1})(\bar{u}(t_{i+1})-v_{i+1})\right| \\ &= \left|\int_0^1 [\nabla_u f(t_{i+1},v_{i+1}+s(v_{i+1}-\bar{u}(t_{i+1})))-\nabla_u f(t_{i+1},v_{i+1})]\,(v_{i+1}-\bar{u}(t_{i+1}))\,ds\right| \\ &\le \int_0^1 sK|v_{i+1}-\bar{u}(t_{i+1})|^2\,ds = \frac{K}{2}|v_{i+1}-\bar{u}(t_{i+1})|^2 \\ &\le \frac{K}{2}\left(\kappa' K h^2\left(1+L+L^2\right)\right)^2 = (c/\kappa')h^4.\end{aligned}$$

In particular, this implies that $|h(\bar{u}(t_{i+1}))| \le \beta$ due to the second relation in (18). Applying Theorem 5G.3 with $g=h$ in the same way as for the estimate (22) we obtain that there exists a unique in $\mathbb{B}_\alpha(\bar{u}(t_{i+1}))$ solution u_{i+1} of (23) which moreover satisfies (19). This completes the induction step and the proof of the theorem. □

We should note that a solution trajectory of the parameterized generalized equation (1) cannot be expected to be smoother than Lipschitz continuous; therefore a piecewise linear (or of higher order) interpolation across (t_i,u_i) will have error of order no better than $O(h)$ if we measure it in the uniform norm over the interval $[0,1]$ rather than in the interpolation points only.

For $F \equiv 0$ and $m=n$ the Euler step for the method (14) becomes

$$v_{i+1} = u_i - \nabla_u f(t_i,u_i)^{-1}(h\nabla_t f(t_i,u_i)+f(t_i,u_i)) \tag{24}$$

which is different from the Euler step in the equation case (13). We just proved in 6G.2 that the method combining the modified Euler step (24) with the standard Newton step has error of order $O(h^4)$. It turns out that the error has the same order when we use the method (13). This could be shown in various ways; in our case the simplest is to follow the proof of 6G.2. Indeed, if instead of g in (21) we use the function

$$\begin{aligned}\bar{g}(v) = {}& \nabla_u f(t_i,u_i)(v-u_i)+h\nabla_t f(t_i,u_i) \\ &-[f(t_{i+1},\bar{u}(t_{i+1}))+\nabla_u f(t_{i+1},\bar{u}(t_{i+1}))(v-\bar{u}(t_{i+1}))],\end{aligned}$$

then, from the induction hypothesis and the fact that $f(t_i,\bar{u}(t_i)) = 0$, we get

$$|f(t_i,u_i)| = |f(t_i,u_i) - f(t_i,\bar{u}(t_i))| \le Kch^4.$$

Hence,

$$\begin{aligned}|\bar{g}(\bar{u}(t_{i+1}))| &\le |g(\bar{u}(t_{i+1}))| + |\bar{g}(\bar{u}(t_{i+1})) - g(\bar{u}(t_{i+1}))| \\ &\le |g(\bar{u}(t_{i+1}))| + |f(t_i,u_i)| \le |g(\bar{u}(t_{i+1}))| + Kch^4.\end{aligned}$$

Thus, the estimate for $|\bar{g}(\bar{u}(t_{i+1})|$ is of the same order as for $|g(\bar{u}(t_{i+1})|$ and hence the final estimate (19) is of the same order.

We end this section with an important observation. Consider the following method where we have not one but *two* corrector (Newton) steps:

$$\begin{cases} f(t_i,u_i) + h\nabla_t f(t_i,u_i) + \nabla_u f(t_i,u_i)(v_{i+1} - u_i) + F(v_{i+1}) \ni 0, \\ f(t_{i+1},v_{i+1}) + \nabla_u f(t_{i+1},v_{i+1})(w_{i+1} - v_{i+1}) + F(w_{i+1}) \ni 0, \\ f(t_{i+1},w_{i+1}) + \nabla_u f(t_{i+1},w_{i+1})(u_{i+1} - w_{i+1}) + F(u_{i+1}) \ni 0. \end{cases}$$

By repeating the argument used in the proof of Theorem 6G.2 one obtains an estimate for the error of order $O(h^8)$. A third Newton step will give $O(h^{16})$! Such a strategy would be perhaps acceptable for relatively small problems. For practical problems, however, a trade off is to be sought between theoretical accuracy and computational complexity of an algorithm. Also, one should remember that the error in the uniform norm will always be $O(h)$ in general, unless the solution has better smoothness properties than just Lipschitz continuity.

6H. Galerkin's Method for Quadratic Minimization

The topic of this section is likewise a traditional scheme in numerical analysis and its properties of convergence, again placed in a broader setting than the classical one. The problem at which this scheme will be directed is quadratic optimization in a Hilbert space setting:

$$\text{(1)} \qquad \text{minimize } \frac{1}{2}\langle x,Ax\rangle - \langle v,x\rangle \text{ over } x \in C,$$

where C is a nonempty, closed and convex set in a Hilbert space X, and $v \in X$ is a parameter. Here $\langle\cdot,\cdot\rangle$ denotes the inner product in X; the associated norm is $\|x\| = \sqrt{\langle x,x\rangle}$. We take $A : X \to X$ to be a linear and bounded mapping, entailing $\operatorname{dom} A = X$; furthermore, we take A to be self-adjoint, $\langle x,Ay\rangle = \langle y,Ax\rangle$ for all $x,y \in X$ and require that

$$\text{(2)} \qquad \langle x,Ax\rangle \ge \mu\|x\|^2 \text{ for all } x \in C - C, \text{ for a constant } \mu > 0.$$

This property of A, sometimes called *coercivity* (a term which can have conflicting manifestations), corresponds to A being strongly monotone relative to C in the sense defined in 2F, as well as to the quadratic function in (1) being strongly convex relative to C. For $X = \mathbb{R}^n$, (2) is equivalent to positive definiteness of A relative to the subspace generated by $C - C$. For any Hilbert space X in which that subspace is dense, it entails A being invertible with $\|A^{-1}\| \leq \mu^{-1}$.

In the usual framework for Galerkin's method, C would be all of X, so the targeted problem would be unconstrained. The idea is to consider an increasing sequence of finite-dimensional subspaces X_k of X, and by iteratively minimizing over X_k, to get a solution point $\hat{x}_k$, generate a sequence which, in the limit, solves the problem for X.

This approach has proven valuable in circumstances where X is a standard function space and the special functions making up the subspaces X_k are familiar tools of approximation, such as trigonometric expansions. Here, we will work more generally with convex sets C_k furnishing "inner approximations" to C, with the eventual possibility of taking $C_k = C \cap X_k$ for a subspace X_k.

In Section 2G with $X = \mathbb{R}^n$, we looked at a problem like (1) in which the function was not necessarily quadratic, and we studied the dependence of its solution on the parameter v. Before proceeding with anything else, we must update to our Hilbert space context with a quadratic function the particular facts from that development which will be called upon.

Theorem 6H.1 (optimality and its characterization)**.** *For problem (1) under condition (2), for each v there exists a unique solution x. The solution mapping $S : v \to x$ is thus single-valued with* $\operatorname{dom} S = X$. *Moreover, this mapping S is Lipschitz continuous with constant μ^{-1}, and it is characterized by a variational inequality:*

$$x = S(v) \iff -v + Ax + N_C(x) \ni 0. \tag{3}$$

Proof. The existence of a solution x for a fixed v comes from the fact that, for each sufficiently large $\alpha \in \mathbb{R}$ the set C_α of $x \in C$ for which the function being minimized in (1) has value $\leq \alpha$ is nonempty, convex, closed and bounded, with the bound coming from (2). Such a subset of X is weakly compact; hence every minimizing sequence has a convergent subsequence and each limit of such a subsequence is a solutions x. The uniqueness of such x follows however from the strong convexity of the function in question. The characterization of x in (3) is proved exactly as in the case of $X = \mathbb{R}^n$ in 2A.7. The Lipschitz property of S comes out of the same argument that was used in the second half of the proof of 2F.6, utilizing the strong monotonicity of A. $\square$

As an important consequence of Theorem 6H.1, we get a Hilbert space version of the projection result in 1D.5 for convex sets in $\mathbb{R}^n$.

Corollary 6H.2 (projections in Hilbert spaces)**.** *For a nonempty, closed, convex set C in a Hilbert space X, there exists for each $v \in X$ a unique nearest point x of*

C, called the projection of v on C and denoted by $P_C(v)$. The projection mapping $P_C : X \to C$ is Lipschitz continuous with constant 1.

Proof. Take $A = I$ in (1), noting that (2) holds then with $\mu = 1$. Problem (1) is equivalent then to minimizing $||x-v||$ over $x \in C$, because the expression being minimized differs from $\frac{1}{2}||x-v||^2$ only by the constant term $\frac{1}{2}||v||^2$. □

In Galerkin's method, when we get to it, there will be need of comparing solutions to (1) with solutions to other problems for the same v but sets different from C. In effect, we have to be able to handle the choice of C as another sort of parameter. For a start, consider just two different sets, D_1 and D_2. How might solutions to the versions of (1) with D_1 and D_2 in place of C, but with fixed v, relate to each other? To get anywhere with this we require a *joint* strong monotonicity condition extending (2):

$$(4) \qquad \langle x, Ax \rangle \geq \mu \|x\|^2 \text{ for all } x \in D_i - D_j \text{ and } i,j \in \{1,2\}, \text{ where } \mu > 0.$$

Obviously (4) holds without any fuss over different sets if we simply have A strongly monotone with constant μ on all of X.

Proposition 6H.3 (solution estimation for varying sets)**.** *Consider any nonempty, closed, convex sets D_1 and D_2 in X satisfying (4). If x_1 and x_2 are the solutions of problem (1) with constraint sets D_1 and D_2, respectively, in place of C, then*

$$(5) \quad \mu\|x_1 - x_2\|^2 \leq \langle Ax_1 - v, u_1 - x_2 \rangle + \langle Ax_2 - v, u_2 - x_1 \rangle \quad \textit{for all } u_1 \in D_1, u_2 \in D_2.$$

Proof. From (4) we have

$$(6) \qquad \mu\|x_1 - x_2\|^2 \leq \langle A(x_1 - x_2), x_1 - x_2 \rangle,$$

whereas for any $u_1 \in D_1$ and $u_2 \in D_2$, (3) gives us

$$(7) \qquad 0 \leq \langle Ax_1 - v, u_1 - x_1 \rangle, \qquad 0 \leq \langle Ax_2 - v, u_2 - x_2 \rangle.$$

Adding the inequalities in (7) to the one in (6) and rearranging the sum, we obtain

$$\begin{aligned} \mu\|x_1 - x_2\|^2 &\leq \langle A(x_1 - x_2), x_1 - x_2 \rangle + \langle Ax_1 - v, u_1 - x_1 \rangle + \langle Ax_2 - v, u_2 - x_2 \rangle \\ &= \langle Ax_1 - v, u_1 - x_2 \rangle + \langle Ax_2 - v, u_2 - x_1 \rangle, \end{aligned}$$

as claimed in (5). □

Having this background at our disposal, we are ready to make progress with our generalized version of Galerkin's method. We consider along with C a sequence of sets $C_k \subset X$ for $k = 1, 2, \ldots$ which, like C, are nonempty, closed and convex. We suppose that

$$(8) \qquad C_k \subset C_{k+1} \subset \cdots \subset C, \text{ with } \operatorname{cl}[C_1 \cup C_2 \cup \cdots] = C,$$

and let

(9) $\qquad S_k =$ the solution mapping for (1) with C_k in place of C,

as provided by Theorem 6H.1 through the observation that (2) carries over to any subset of C. By generalized Galerkin's sequence associated with (8) for a given v, we will mean the sequence $\{\hat{x}_k\}$ of solutions $\hat{x}_k = S_k(v)$, $k = 1,2,\ldots$.

Theorem 6H.4 (general rate of convergence)**.** *Let S be the solution mapping to (1) as provided by Theorem 6H.1 under condition (2), and let $\{C_k\}$ be a sequence of nonempty, closed, convex sets satisfying (8). Then for any v the associated Galerkin's sequence $\{\hat{x}_k\} = \{S_k(v)\}$ converges to $\hat{x} = S(v)$. In fact, there is a constant c such that*

$$\|\hat{x}_k - \hat{x}\| \le c\, d(\hat{x}, C_k)^{1/2} \text{ for all } k. \tag{10}$$

Proof. On the basis of (8), we have $\operatorname{dist}(\hat{x}, C_k) \to 0$. The sequence of projections $\bar{x}_k = P_{C_k}(\hat{x})$ with $||\bar{x}_k - \hat{x}|| = d(\hat{x}, C_k)$, whose existence is guaranteed by 6H.2, converges then to $\hat{x}$. From 6H.3 applied to $D_1 = C_k$ and $D_2 = C$, with $\hat{x}_k$ and $\hat{x}$ in the place of the x_1 and x_2 there, and on the other hand $u_1 = \bar{x}_k$ and $u_2 = \hat{x}_k$, we get $\mu\|\hat{x}_k - \hat{x}\|^2 \le \langle A\hat{x}_k - v, \bar{x}_k - \hat{x}\rangle$ and therefore

$$\begin{aligned} \mu\|\hat{x}_k - \hat{x}\|^2 &\le \langle A(\hat{x}_k - \hat{x}) + A\hat{x} - v, \bar{x}_k - \hat{x}\rangle \\ &\le (\|A\|\|\hat{x}_k - \hat{x}\| + \|A\|\|\hat{x}\| + \|v\|)\operatorname{dist}(\hat{x}, C_k). \end{aligned} \tag{11}$$

This quadratic inequality in $d_k = \|\hat{x}_k - \hat{x}\|$ implies that the sequence $\{d_k\}$ is bounded, say by b. Putting this b in place of $\|\hat{x}_k - \hat{x}\|$ on the right side of (11), we get a bound of the form in (10). $\square$

Is the square root describing the rate of convergence through the estimate in (10) exact? The following example shows that this is indeed the case, and no improvement is possible, in general.

Example 6H.5 (counterexample to improving the general estimate)**.** Consider problem (1) in the case of $X = \mathbb{R}^2$, $C = \{\, (x_1, x_2) \,|\, x_2 \le 0 \}$ (lower half-plane), $v = (0,1)$ and $A = I$, so that the issue revolves around projecting v on C and the solution is $\hat{x} = (0,0)$. For each $k = 1,2,\ldots$ let $a_k = (1/k, 0)$ and let C_k consist of the points $x \in C$ such that $\langle x - a_k, v - a_k\rangle \le 0$. Then the projection $\hat{x}_k$ of v on C_k is a_k, and

$$|\hat{x}_k - \hat{x}| = 1/k, \qquad d(\hat{x}, C_k) = \frac{1}{k\sqrt{1+k^2}}.$$

In this case the ratio $|\hat{x}_k - \hat{x}|/d(\hat{x}, C_k)^p$ is unbounded in k for any $p > 1/2$.

Detail. The fact that the projection of v on C_k is a_k comes from the observation that $v - a_k \in N_{C_k}(a_k)$. A similar observation confirms that the specified $\bar{x}_k$ is the projection of $\hat{x}$ on C_k. The ratio $|\hat{x}_k - \hat{x}|/d(\hat{x}, C_k)^p$ can be calculated as $k^{2p-1}(1 + 1/(k^2)^{p/2}$, and from that the conclusion is clear that it is bounded with respect to k if and only if $2 - (1/p) \le 0$, or in other words, $p \le 1/2$. $\square$

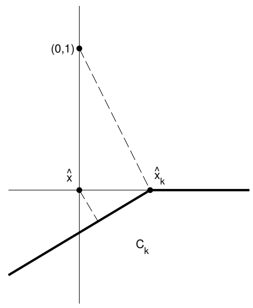

Fig. 6.1 Illustration to Example 6H.5.

There is, nevertheless, an important case in which the exponent $1/2$ in (10) can be replaced by 1. This case is featured in the following result:

Theorem 6H.6 (improved rate of convergence for subspaces). *Under the conditions of Theorem 6H.4, if the sets C and C_k are subspaces of X, then there is a constant c such that*

(12) $$\|\hat{x}_k - \hat{x}\| \leq cd(\hat{x}, C_k) \quad \textit{for all } k.$$

Proof. In this situation the variational inequality in (3) reduces to the requirement that $Ax - v \perp C$. We then have $A\hat{x} - v \in C^{\perp} \subset C_k^{\perp}$ and $A\hat{x}_k - v \in C_k^{\perp}$, so that $A(\hat{x}_k - \hat{x}) \in C_k^{\perp}$. Consider now an arbitrary $x \in C_k$, noting that since $\hat{x}_k \in C_k$ we also have $\hat{x}_k - x \in C_k$. We calculate from (2) that

$$\begin{aligned}
\mu\|\hat{x}_k - \hat{x}\|^2 &\leq \langle A(\hat{x}_k - \hat{x}), \hat{x}_k - \hat{x}\rangle \\
&= \langle A(\hat{x}_k - \hat{x}), \hat{x}_k - x\rangle + \langle A(\hat{x}_k - \hat{x}), x - \hat{x}\rangle \\
&= \langle A(\hat{x}_k - \hat{x}), x - \hat{x}\rangle \leq \|A\|\|\hat{x}_k - \hat{x}\|\|x - \hat{x}\|.
\end{aligned}$$

This gives us the estimate (12). □

The result in 6H.6 corresponds to the classical Galerkin's method, at least if C is all of X. We can combine it with the one in 6H.4 as follows.

Corollary 6H.7 (application to intersections with subspaces). *Let S be the solution mapping to (1) as provided by Theorem 6H.1 under condition (2). Let $\{X_k\}$ be an increasing sequence of closed subspaces of X such that (8) holds for the sets $C_k = C \cap X_k$. Then for any v the associated Galerkin's sequence $\{\hat{x}_k\} = \{S_k(v)\}$ converges to $\hat{x} = S(v)$ at the rate indicated in (10), but if C itself is a subspace, it converges at the rate indicated in (12).*

The closure condition in (8), in the case of $C_k = C \cap X_k$, says that $\operatorname{dist}(x, C \cap X_k) \to 0$ as $k \to \infty$ for every $x \in C$. When the Hilbert space X is separable we may choose the subspaces X_k by taking a countable dense subset $x_1, x_2, \ldots$ of X and letting X_k be the span of $x_1, \dots, x_k$. Because the subspaces are finite-dimensional, Galerkin's method in this case can be viewed as a discretization scheme. The property that $\operatorname{dist}(x, C \cap X_k) \to 0$ as $k \to \infty$ for every $x \in C$ is called the *consistency* of the discretization scheme. In the following section we will look at the discretization of a specific variational problem.

6I. Metric Regularity and Optimal Control

For an example which illustrates how the theory of solution mappings can be applied in infinite dimensions with an eye toward numerical approximations, we turn to a basic problem in optimal control, which has the form

$$\text{(1)} \qquad \text{minimize} \quad \int_0^1 \varphi(x(t), u(t))dt$$

subject to

$$\text{(2)} \qquad \dot{x}(t) = g(x(t), u(t)) \quad \text{for a.e. } t \in [0,1], \qquad x(0) = a,$$

and the constraint that

$$\text{(3)} \qquad u(t) \in U \quad \text{for a.e. } t \in [0,1].$$

This concerns the control system given by (2) in which $x(t) \in \mathbb{R}^n$ is the *state* at time t and $u(t)$ is the *control* exercised at time t. The choice of the control function $u : [0,1] \to \mathbb{R}^m$ yields from the initial state $a \in \mathbb{R}^n$ and the differential equation in (2) a corresponding state trajectory $x : [0,1] \to \mathbb{R}^n$ with derivative $\dot{x}$. The set $U \subset \mathbb{R}^m$ in (3) from which the values of the control have to be selected is assumed to be convex and closed. Any feasible control function $u : [0,1] \to U$ is required to be Lebesgue measurable ("a.e." refers as usual to "almost everywhere" with respect to Lebesgue measure) and essentially bounded, that is, u belongs to the space $L^\infty(\mathbb{R}^m, [0,1])$, which we equip with the standard norm

$$\|u\|_{L^\infty} = \operatorname{ess\,sup}_{t \in [0,1]} |u(t)|.$$

The state trajectory x, which is a solution of the initial value problem (2) for a given control function u, is regarded as an element of $W_0^{1,\infty}(\mathbb{R}^n, [0,1])$, which is the standard notation of the space of Lipschitz continuous functions x over $[0,1]$ with values in $\mathbb{R}^n$ equipped with the norm

$$\|x\|_{W^{1,\infty}} = \|x\|_{L^\infty} + \|\dot{x}\|_{L^\infty}.$$

We assume that the functions $\varphi : \mathbb{R}^{n+m} \to \mathbb{R}$ and $g : \mathbb{R}^{n+m} \to \mathbb{R}^n$ are twice continuously differentiable everywhere.

Without going into much details, we present next a standard fact from the theory of optimal control[2] which describes a set of first-order optimality conditions sometimes referred to as *Pontryagin maximum principle.* Assume that problem (1) has a solution $(\bar{x}, \bar{u})$. Then, in terms of the Hamiltonian

$$H(x,u,\psi) = \varphi(x,u) + \psi^{\mathsf{T}} g(x,u),$$

there exists an *adjoint variable* $\bar{\psi} \in W^{1,\infty}(\mathbb{R}^n, [0,1])$ such that the triple $\bar{\xi} := (\bar{x}, \bar{u}, \bar{\psi})$ is a solution of the following two-point boundary value problem coupled with a variational inequality:

$$(4) \qquad \begin{cases} \dot{x}(t) &= g(x(t), u(t)), \quad x(0) = a, \\ \dot{\psi}(t) &= -\nabla_x H(x(t), u(t), \psi(t)), \quad \psi(1) = 0, \\ 0 &\in \nabla_u H(x(t), u(t), \psi(t)) + N_U(u(t)), \quad \text{for a.e. } t \in [0,1], \end{cases}$$

where, as usually, $N_U(u)$ is the normal cone to the convex and closed set U at the point u. Introduce the spaces $X = W^{1,\infty}(\mathbb{R}^n, [0,1]) \times W^{1,\infty}(\mathbb{R}^n, [0,1]) \times L^\infty(\mathbb{R}^m, [0,1])$ and $Y = L^\infty(\mathbb{R}^n, [0,1]) \times L^\infty(\mathbb{R}^n, [0,1]) \times L^\infty(\mathbb{R}^m, [0,1])$ which correspond to the variable ξ and the mapping in (4). Further, for $\xi = (x, u, \psi)$ considered as a function on $[0,1]$, let

$$f(\xi) = \begin{pmatrix} \dot{x} - g(x,u) \\ \dot{\psi} + \nabla_x H(x,u,\psi) \\ \nabla_u H(x,u,\psi) \end{pmatrix} \quad \text{and} \quad F(\xi) = \begin{pmatrix} 0 \\ 0 \\ N_U(u) \end{pmatrix}.$$

Then, with f and F acting from X to Y, the optimality system (4) can be written as a generalized equation on a function space, of the form

$$(5) \qquad \text{find } \ \xi \in X \ \text{ such that } \ f(\xi) + F(\xi) \ni 0_Y.$$

Thus, we come to the realm of generalized equations for the analysis of which we have already developed various techniques in this book. Note that (5) is *not* a variational inequality since Y is not the space dual to X, at least. Still, we can employ regularity properties and the various reincarnations of the implicit function theorem to estimate the effect of perturbations and approximations on a solution of (5).

In this section we will focus our attention to showing that metric regularity of the mapping $f + F$ for the optimality systems (4) implies an *a priori* error estimate for a discrete approximation to the problem. First, without loss of generality we assume that in (2) we have $a = 0$ (a simple change of variables will give us this). Suppose that the optimality system (4) is solved inexactly by means of a numerical method applied to a discrete approximation provided by the standard Euler scheme. Specifically, let N be a natural number, let $h = 1/N$ be the mesh spacing, and let

[2] In the following section we derive first-order optimality conditions for a particular optimal control problem.

$t_i = ih$, $i \in \{0,1,\ldots,N\}$. Denote by $PL_0^N(\mathbb{R}^n,[0,1])$ the space of piecewise linear and continuous functions x_N over the grid $\{t_i\}$ with values in $\mathbb{R}^n$ and such that $x_N(0)=0$, by $PL_1^N(\mathbb{R}^n,[0,1])$ the space of piecewise linear and continuous functions ψ_N over the grid $\{t_i\}$ with values in $\mathbb{R}^n$ and such that $\psi_N(1)=0$, and by $PC^N(\mathbb{R}^m,[0,1])$ the space of piecewise constant and continuous from the right functions over the grid $\{t_i\}$ with values in $\mathbb{R}^m$. Clearly, $PL_0^N(\mathbb{R}^n,[0,1])$ and $PL_1^N(\mathbb{R}^n,[0,1])$ are subsets of $W^{1,\infty}(\mathbb{R}^n,[0,1])$ and $PC^N(\mathbb{R}^m,[0,1]) \subset L^\infty(\mathbb{R}^m,[0,1])$. Then introduce the products $X^N = PL_0^N(\mathbb{R}^n,[0,1]) \times PL_1^N(\mathbb{R}^n,[0,1]) \times PC^N(\mathbb{R}^m,[0,1])$ as an approximation space for the triple (x,ψ,u). We identify $x \in PL_0^N(\mathbb{R}^n,[0,1])$ with the vector $(x^0,\ldots,x^N)$ of its values at the mesh points (and similarly for ψ), and $u \in PC^N(\mathbb{R}^m,[0,1])$ – with the vector $(u^0,\ldots,u^{N-1})$ of the values of u in the mesh subintervals.

Now, suppose that, as a result of the computations, for certain natural N a function $\tilde{\xi} = (x_N,\psi_N,u_N) \in X^N$ is found that satisfies the discrete-time optimality system

$$(6) \qquad \begin{cases} \dot{x}_N^i = g(x_N^i,u^i), \quad x_N^0 = 0, \\ \dot{\psi}_N^i = \nabla_x H(x_N^i,u^i,\psi_N^{i+1}), \quad \psi_N^N = 0, \\ 0 \in \nabla_u H(x_N^i,u^i,\psi_N^i) + N_U(u^i) \end{cases}$$

for $i = 0,1,\ldots,N-1$, where, since x_N and ψ_N are piecewise linear, we have

$$\dot{x}_N^i = \frac{x_N^{i+1} - x_N^i}{h}$$

and the same for ψ_N. The system (6) represents the *Euler discretization* of the optimality system (4).

Suppose that the mapping $f+F$ in (5) is metrically regular at $\bar{\xi}$ for 0. Then in particular, there exist positive scalars a and κ such that

$$d(\tilde{\xi},(f+F)^{-1}(0)) \le \kappa d(0,f(\tilde{\xi})+F(\tilde{\xi})) \quad \text{whenever} \quad \tilde{\xi} \in \mathbb{B}_a(\bar{x}),$$

where the right side of this inequality is the residual associated with the approximate solution $\tilde{\xi}$. In our specific case, if the function $\tilde{z} \in Y$ is defined as

$$\tilde{z}(t) = \begin{pmatrix} g(x_N(t_i),u_N(t_i)) - g(x_N(t),u_N(t)) \\ \nabla_x H(x_N(t_i),u_N(t_i),\psi_N(t_{i+1})) - \nabla_x H(x_N(t),u_N(t),\psi_N(t)) \\ \nabla_u H(x_N(t_i),u_N(t_i),\psi_N(t_i)) - \nabla_u H(x_N(t),u_N(t),\psi_N(t)) \end{pmatrix}, \quad t \in [t_i,t_{i+1}),$$

then we have $\tilde{z} \in f(\tilde{\xi}) + F(\tilde{\xi})$, and therefore

$$d(0,f(\tilde{\xi})+F(\tilde{\xi})) \le \kappa \|\tilde{z}\|_Y.$$

Hence, taking into account that

$$(7) \qquad \begin{aligned} \|\tilde{z}\|_Y \le \max\nolimits_{0\le i\le N-1} \sup\nolimits_{t_i \le t \le t_{i+1}} \big[&|g(x_N(t_i),u_N(t_i)) - g(x_N(t),u_N(t_i))| \\ &+ |\nabla_x H(x_N(t_i),u_N(t_i),\psi_N(t_{i+1})) - \nabla_x H(x_N(t),u_N(t_i),\psi_N(t))| \\ &+ |\nabla_u H(x_N(t_i),u_N(t_i),\psi_N(t_i)) - \nabla_u H(x_N(t),u_N(t_i),\psi_N(t))| \big], \end{aligned}$$

in order to estimate the residual it is sufficient to find an estimate for the right side of (7).

Observe that here x_N is a piecewise linear function across the grid $\{t_i\}$ with uniformly bounded derivative, since both x_N and u_N are in some L_∞ neighborhood of $\bar{x}$ and $\bar{u}$ respectively. Hence, taking into account that the functions g, $\nabla_x H$ and $\nabla_u H$ are continuously differentiable we obtain that the norm of the residual can be estimated by a constant times the mesh size h. Thus, we come to the following result:

Theorem 6I.1 (a priori error estimate in optimal control)**.** *Assume that the mapping of the optimality system (4) is metrically regular at $\bar{\xi} = (\bar{x}, \bar{u}, \bar{\psi})$ for 0. Then there exist constants a and c such that if the distance from a solution $\xi_N = (x_N, u_N, \psi_N)$ of the discretized system (6) to $\bar{\xi}$ is not more than a, then there exists a solution $\bar{\xi}^N = (\bar{x}^N, \bar{u}^N, \bar{\psi}^N)$ of the continuous system (4) such that*

$$\|\bar{x}^N - x_N\|_{W^{1,\infty}} + \|\bar{u}^N - u_N\|_{L^\infty} + \|\bar{\psi}^N - \psi_N\|_{W^{1,\infty}} \leq ch.$$

Furthermore, if the mapping of the optimality system (4) is strongly metrically regular at $\bar{\xi}$ for 0, then the above claim holds with $\bar{\xi}^N = \bar{\xi}$; that is, if $\xi_N = (x_N, u_N, \psi_N)$ is a sequence of approximate solutions to the discretized system (6) contained in this ball, then ξ_N converges to $\bar{\xi}$ in the norm of X with rate proportional to the mesh size h.

By following the same line of reasoning, we could also obtain *a posteriori* error estimates measuring the distance from a reference solution of the continuous system to the set of solution of the discretized system, provided that the mapping of discretized system (6) is metrically regular, uniformly in N. This would be true for example when metric regularity is preserved after discrete approximation. But note that here we approximate not only the mappings involved but also the spaces X and Y; specifically, to approximate X we use spaces of piecewise linear functions for the state and adjoint variable and piecewise constant functions for the control, and to approximate Y we use spaces of piecewise constant functions. Thus, the issue comes down to the general question whether the property of metric regularity is inherited by a restriction of the mapping on a subspace. It turns out that this is not the case, as the following counterexample shows.

Example 6I.2. Let $X = I\!R^2$, $Y = I\!R$, $f(x_1, x_2) = x_2 - x_1^3$. Here

$$f^{-1}(y) = \{(x_1, x_2) \mid x_2 = y + x_1^3,\ x_1 \in I\!R\}.$$

The function f is metrically regular at $\bar{x} = (0,0)$ for $\bar{y} = 0$ with $\kappa = 1$, since

$$d(x, f^{-1}(y)) \leq |(x_1, x_2) - (x_1, y + x_1^3)| = |y - (x_2 - x_1^3)| = |y - f(x)|.$$

On the other hand, the restriction of f to $\tilde{X} := \{(x_1, x_2) \mid x_2 = 0\}$ is not metrically regular at $\bar{x}_1 = 0$ for $\bar{y} = 0$ because for $x \in \tilde{X}$ we have $f(x) = -x_1^3$, hence $x_1 = (-y)^{1/3}$, which is not Lipschitz at $\bar{y} = 0$.

6J. The Linear-Quadratic Regulator Problem

In this section we continue the theme of optimal control by presenting a more detailed analysis of the effects of perturbations and approximations on solutions of the so-called linear-quadratic regulator problem. That problem takes the form:

$$\text{(1)} \qquad \text{minimize} \quad \int_0^1 \left(\frac{1}{2}[x(t)^\mathsf{T} Q x(t) + u(t)^\mathsf{T} R u(t)] + s(t)^\mathsf{T} x(t) - r(t)^\mathsf{T} u(t) \right) dt$$

subject to

$$\text{(2)} \qquad \dot{x}(t) = Ax(t) + Bu(t) + p(t) \quad \text{for a.e. } t \in [0,1], \qquad x(0) = a,$$

and the constraint that

$$\text{(3)} \qquad u(t) \in U \quad \text{for a.e. } t \in [0,1].$$

This concerns the linear control system governed by (2) in which $x(t) \in \mathbb{R}^n$ is the state and $u(t)$ is the control. The matrices A, B, Q and R have dimensions fitting these circumstances, with Q and R symmetric and positive semidefinite so as to ensure (as will be seen) that the function being minimized in (1) is convex.

The set $U \subset \mathbb{R}^m$ from which the values of the control have to be selected from in (3) is nonempty, convex and compact[3]. We also assume that the matrix R is positive definite relative to $U - U$; in other words, there exists $\mu > 0$ such that

$$\text{(4)} \qquad u^\mathsf{T} R u \geq \mu |u|^2 \quad \text{for all } u \in U - U.$$

Any control function $u : [0,1] \to U$ is required to be Lebesgue measurable, and since it takes values in the bounded set U for a.e. $t \in [0,1]$, it is essentially bounded that is, $u \in L^\infty(\mathbb{R}^m, [0,1])$. But we take the set of feasible control functions to be a larger subset of $L^2(\mathbb{R}^m, [0,1])$ functions, in which space the inner product and the norm are

$$\langle u, v \rangle = \int_0^1 u(t)^\mathsf{T} v(t)\, dt, \qquad \|u\|_2 = \sqrt{\langle u, u \rangle}.$$

We follow that Hilbert space pattern throughout, assuming that the function r in (1) belongs to $L^2(\mathbb{R}^m, [0,1])$ while p and s belong to $L^2(\mathbb{R}^n, [0,1])$. This is a convenient compromise which will put us in the framework of quadratic optimization in 6H.

There are two ways of looking at problem (1). We can think of it in terms of minimizing over function pairs (u, x) constrained by both (2) and (3), or we can regard x as a "dependent variable" produced from u through (2) and standard facts about differential equations, so as to think of the minimization revolving only around the choice of u. For any u satisfying (3) (and therefore essentially bounded), there is a unique state trajectory x specified by (2) in the sense of x being an absolutely continuous function of t and therefore differentiable a.e. Due to the assumption

[3] We do not really need U to be bounded, but this assumption simplifies the analysis.

that $p \in L^2(\mathbb{R}^n,[0,1])$, the derivative $\dot{x}$ can then be interpreted as an element of $L^2(\mathbb{R}^n,[0,1])$ as well. Indeed, x is given by the Cauchy formula

$$(5) \qquad x(t) = e^{At}a + \int_0^t e^{A(t-\tau)}(Bu(\tau)+p(\tau))d\tau \quad \text{for all } t \in [0,1].$$

In particular, we can view it as belonging to the Banach space $C(\mathbb{R}^n,[0,1])$ of continuous functions from $[0,1]$ to $\mathbb{R}^n$ equipped with the norm

$$\|x\|_C = \max_{0\le t\le 1}|x(t)|.$$

The relation between u and x can be cast in a frame of inputs and outputs. Define the mapping $T : L^2(\mathbb{R}^n,[0,1]) \to L^2(\mathbb{R}^n,[0,1])$ as

$$(6) \qquad (Tw)(t) = \int_0^t e^{A(t-\tau)}w(\tau)d\tau \quad \text{for a.e. } t \in [0,1],$$

and, on the other hand, let $W : L^2(\mathbb{R}^n,[0,1]) \to L^2(\mathbb{R}^n,[0,1])$ be the mapping defined by

$$(7) \qquad \text{for } p \in L^2(\mathbb{R}^n,[0,1]),\ W(p) \text{ is the solution to } \dot{W} = AW + p,\ W(0) = a.$$

Finally, with a slight abuse of notation, denote by B the mapping from $L^2(\mathbb{R}^m,[0,1])$ to $L^2(\mathbb{R}^n,[0,1])$ associated with the matrix B, that is $(Bu)(t) = Bu(t)$; later we do the same for the mappings Q and R. Then the formula for x in (5) comes out as

$$(8) \qquad x = (TB)(u) + W(p),$$

where u is the input, x is the output, and p is a parameter. Note that in this case we are treating x as an element of $L^2(\mathbb{R}^n,[0,1])$ instead of $C(\mathbb{R}^n,[0,1])$. This makes no real difference but will aid in the analysis.

Exercise 6J.1 (adjoint in the Cauchy formula). *Prove that the mapping T defined by (6) is linear and bounded. Also show that the adjoint (dual) mapping T^*, satisfying $\langle x, Tu\rangle = \langle T^*x, u\rangle$, is given by*

$$(T^*x)(t) = \int_t^1 e^{A^{\mathsf{T}}(\tau-t)}x(\tau)d\tau \quad \text{for a.e. } t \in [0,1].$$

Also show $(TB)^* = B^*T^*$, where B^* is the mapping acting from $L^2(\mathbb{R}^n,[0,1])$ to $L^2(\mathbb{R}^m,[0,1])$ and associated with the transposed matrix B^{T}; that is

$$((TB)^*x)(t) = \int_t^1 B^{\mathsf{T}} e^{A^{\mathsf{T}}(\tau-t)}x(\tau)d\tau \quad \text{for a.e. } t \in [0,1].$$

Guide. Apply the rule for changing the order of integration

$$\int_0^1 x(t)^\mathsf{T} \int_0^t e^{A(t-\tau)} w(\tau) d\tau dt = \int_0^1 \int_\tau^1 x(t)^\mathsf{T} e^{A(t-\tau)} w(\tau) dt d\tau,$$

and interpret what it says. □

To shorten notation in what follows, we will just write L^2 for both $L^2(\mathbb{R}^m, [0,1])$ and $L^2(\mathbb{R}^n, [0,1])$, leaving it to the reader to keep in mind which elements lie in $\mathbb{R}^m$ and which lie in $\mathbb{R}^n$.

The change of variables $z = x - w$ with $w = W(p)$ as in (7) gives the following reformulation of (1)–(3), where the parameter p is transferred to the problem of minimizing the objective function

$$(9)\quad \int_0^1 \left(\frac{1}{2}[z(t)^\mathsf{T} Q z(t) + u(t)^\mathsf{T} R u(t)] + (s(t) + QW(p)(t))^\mathsf{T} z(t) - r(t)^\mathsf{T} u(t) \right) dt$$

subject to

$$(10)\qquad \dot{z} = Az + Bu, \quad z(0) = 0, \quad u(t) \in U \ \text{ for a.e. } t \in [0,1].$$

In (9) we have dropped the constant terms that do not affect the solution. Noting that $z = (TB)(u)$ and utilizing the adjoint T^* of the mapping T, let

$$(11)\qquad V(y) = -r + B^* T^* (s + QW(p)) \ \text{ for } y = (p, s, r),$$

and define the self-adjoint bounded linear mapping $\mathscr{A} : L^2 \to L^2$ by

$$(12)\qquad \mathscr{A} = B^* T^* Q T B + R.$$

Here, as for the mapping B, we regard Q and R as linear bounded mappings acting between L^2 spaces: for $(Ru)(t) = Ru(t)$, and so forth. Let

$$(13)\qquad C = \left\{ u \in L^2 \,\middle|\, u(t) \in U \text{ for a.e. } t \in [0,1] \right\}.$$

With this notation, problem (9)–(10) can be written in the form treated in 6H:

$$(14)\qquad \text{minimize } \frac{1}{2}\langle u, \mathscr{A} u\rangle + \langle V(y), u\rangle \ \text{ subject to } u \in C.$$

Exercise 6J.2 (coercivity in control)**.** *Prove that the set C in (13) is a closed and convex subset of L^2 and that the mapping $\mathscr{A} \in \mathscr{L}(L^2, L^2)$ in (12) satisfies the condition*

$$\langle u, \mathscr{A} u\rangle \geq \mu \|u\|_2^2 \quad \text{for all } u \in C - C,$$

where μ is the constant in (4).

Applying Theorem 6H.1 in the presence of 6J.2, we obtain a necessary and sufficient condition for the optimality of u in problem (14), namely the variational inequality

$$(15) \qquad V(y) + \mathscr{A}u + N_C(u) \ni 0.$$

For (15), or equivalently for (14) or (1)–(3), we arrive then at the following result of implicit function type:

Theorem 6J.3 (implicit function theorem for optimal control in L^2). *Under (4), the solution mapping S which goes from parameter elements $y = (p,s,r)$ to (u,x) solving (1)–(3) is single-valued and globally Lipschitz continuous from the space $L^2(\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m, [0,1])$ to the space $L^2(\mathbb{R}^m, [0,1]) \times C(\mathbb{R}^n, [0,1])$.*

Proof. Because V in (11) is an affine function of $y = (p,s,r)$, we obtain from 6H.1 that for each y problem (14) has a unique solution $u(y)$ and, moreover, the function $y \mapsto u(y)$ is globally Lipschitz continuous in the respective norms. The value $u(y)$ is the unique optimal control in problem (1)–(3) for y. Taking norms in the Cauchy formula (5), we see further that for any $y = (p,s,r)$ and $y' = (p',s',r')$, if x and x' are the corresponding solutions of (2) for $u(y)$, p, and $u(y')$, p', then, for some constants c_1 and c_2, we get

$$\begin{aligned} |x(t) - x'(t)| &\leq c_1 \int_0^t (|B||u(y)(\tau) - u(y')(\tau)| + |p(\tau) - p'(\tau)|)d\tau \\ &\leq c_2(\|u(y) - u(y')\|_2 + \|p - p'\|_2). \end{aligned}$$

Taking the supremum on the left and having in mind that $y \mapsto u(y)$ is Lipschitz continuous, we obtain that the optimal trajectory mapping $y \mapsto x(y)$ is Lipschitz continuous from the L^2 space of y to $C(\mathbb{R}^n, [0,1])$. Putting these facts together, we confirm the claim in the theorem. $\square$

The optimal control u whose existence and uniqueness for a given y is asserted in 6J.3 is actually, as an element of L^2, an equivalence class of functions differing from each other only on sets of measure zero in $[0,1]$. Thus, having specified an optimal control function u, we may change its values $u(t)$ on a t-set of measure zero without altering the value of the expression being minimized or affecting optimality. We will go on to show now that one can pick a particular function from the equivalence class which has better continuity properties with respect to both time and the parameter dependence.

For a given control u and parameter $y = (p,s,r)$, let

$$\psi = T^*(Qx + s),$$

where x solves (8). Then, through the Cauchy formula and 6J.1, ψ is given by

$$\psi(t) = \int_t^1 e^{A^\mathsf{T}(\tau - t)}(Qx(\tau) + s(\tau))\,d\tau \qquad \text{for all } t \in [0,1].$$

Hence, ψ is a continuous function which is differentiable almost everywhere in $[0,1]$ and its derivative $\dot\psi$ is in L^2. Further, taking into account (6) and (8), $\dot\psi$ satisfies

$$(16) \qquad \dot\psi(t) = -A^\mathsf{T}\psi(t) - Qx(t) - s(t) \quad \text{for a.e. } t \in [0,1], \quad \psi(1) = 0,$$

where x is the solution of (2) for the given u. The function ψ is called the *adjoint* or *dual* trajectory associated with a given control u and its corresponding state trajectory x, and (16) is called the *adjoint equation*. Bearing in mind the particular form of $V(y)$ in (11) and that, by definition,

$$\psi = T^*(Qx+s) = T^*[QTBu+s+QW(p)],$$

we can re-express the variational inequality (15) in terms of ψ as

$$(17) \qquad \langle -r+Ru+B^*\psi, v-u\rangle \geq 0 \quad \text{for all } \; v\in C,$$

where B^* stands for the linear mapping associated with the transpose of the matrix B. The boundary value problem combining (2) and (16), coupled with the variational inequality (17), fully characterizes the solution to problem (1)–(3).

We need next a standard fact from Lebesgue integration. For a function φ on $[0,1]$, a point $\hat{t}\in(0,1)$ is said to be a *Lebesgue point* of φ when

$$\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_{\hat{t}-\varepsilon}^{\hat{t}+\varepsilon}\varphi(\tau)d\tau = \varphi(\hat{t}).$$

It is known that when φ is integrable on $[0,1]$, its set of Lebesgue points is of full measure 1.

Now, let u be the optimal control for a particular parameter value y, and let x and ψ be the associated optimal trajectory and adjoint trajectory, respectively. Let $\hat{t}\in(0,1)$ be a Lebesgue point of both u and r (the set of such $\hat{t}$ is of full measure). Pick any $w\in U$, and for $0<\varepsilon<\min\{\hat{t},1-\hat{t}\}$ consider the function

$$\hat{u}_\varepsilon(t) = \begin{cases} w & \text{for } t\in(\hat{t}-\varepsilon,\hat{t}+\varepsilon),\\ u(t) & \text{otherwise.}\end{cases}$$

Then for every sufficiently small ε the function $\hat{u}_\varepsilon$ is a feasible control, i.e., belongs to the set C in (13), and from (17) we obtain

$$\int_{\hat{t}-\varepsilon}^{\hat{t}+\varepsilon}(-r(\tau)+Ru(\tau)+B^{\mathsf{T}}\psi(\tau))^{\mathsf{T}}(w-u(\tau))d\tau \geq 0.$$

Since $\hat{t}$ is a Lebesgue point of the function under the integral (we know that ψ is continuous and hence its set of Lebesgue points is the entire interval $[0,1]$), we can pass to zero with ε and by taking into account that $\hat{t}$ is an arbitrary point from a set of full measure in $[0,1]$ and that w can be any element of U, come to the following *pointwise* variational inequality which is required to hold for a.e. $t\in[0,1]$:

$$(18) \qquad (-r(t)+Ru(t)+B^{\mathsf{T}}\psi(t))^{\mathsf{T}}(w-u(t)) \geq 0 \;\text{ for every }\; w\in U.$$

As is easily seen, (18) implies (17) as well, and hence these two variational inequalities are equivalent.

Summarizing, we can now say that a feasible control u is the solution of (1)–(3) for a given $y=(p,s,r)$ with corresponding optimal trajectory x and adjoint trajectory ψ if and only if the triple (u,x,ψ) solves the following boundary value problem coupled with a pointwise variational inequality:

$$\text{(19a)}\qquad \begin{cases} \dot{x}(t)=Ax(t)+Bu(t)+p(t), & x(0)=a,\\ \dot{\psi}(t)=-A^{\mathsf{T}}\psi(t)-Qx(t)-s(t), & \psi(1)=0, \end{cases}$$

$$\text{(19b)}\qquad r(t)\in Ru(t)+B^{\mathsf{T}}\psi(t)+N_U(u(t)) \quad \text{for a.e. } t\in[0,1].$$

That is, for an optimal control u and associated optimal state and adjoint trajectories x and ψ, there exists a set of full measure in $[0,1]$ such that (19b) holds for every t in this set. Under an additional condition on the function r we obtain the following result:

Theorem 6J.4 (Lipschitz continuous optimal control)**.** *Let the parameter $y=(p,s,r)$ in (1)–(3) be such that the function r is Lipschitz continuous on $[0,1]$. Then, from the equivalence class of optimal control functions for this y, there exists an optimal control $u(y)$ for which (19b) holds for all $t\in[0,1]$ and which is Lipschitz continuous with respect to t on $[0,1]$. Moreover, the solution mapping $y\mapsto u(y)$ is Lipschitz continuous from the space $L^2(\mathbb{R}^n\times\mathbb{R}^n,[0,1])\times C(\mathbb{R}^m,[0,1])$ to the space $C(\mathbb{R}^m,[0,1])$.*

Proof. It is clear that the adjoint trajectory ψ is Lipschitz continuous in t on $[0,1]$ for any feasible control; indeed, it is the solution of the linear differential equation (16), the right side of which is a function in L^2. Let x and ψ be the optimal state and adjoint trajectories and let u be a function satisfying (19b) for all $t\in\sigma$ where σ is a set of full measure in $[0,1]$. For $t\notin\sigma$ we define $u(t)$ to be the unique solution of the following strongly monotone variational inequality in $\mathbb{R}^n$:

$$\text{(20)}\qquad q(t)\in Ru+N_U(u), \quad \text{where } q(t)=r(t)-B^{\mathsf{T}}\psi(t).$$

Then this u is within the equivalence class of optimal controls, and, moreover, the vector $u(t)$ satisfies (20) for all $t\in[0,1]$. Noting that q is a Lipschitz continuous function in t on $[0,1]$, we get from 2F.6 that for each fixed $t\in[0,1]$ the solution mapping of (20) is Lipschitz continuous with respect to $q(t)$. Since the composition of Lipschitz continuous functions is Lipschitz continuous, the particular optimal control function u which satisfies (19b) for all $t\in[0,1]$ is Lipschitz continuous in t on $[0,1]$. □

Theorem 6J.5 (implicit function theorem for optimal control in L^∞)**.** *The solution mapping $(p,s,r)\mapsto(u,x,\psi)$ of the optimality system (19a,b) is Lipschitz continuous from the space $L^\infty(\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^m,[0,1])$ to the space $L^\infty(\mathbb{R}^m,[0,1])\times C(\mathbb{R}^n\times\mathbb{R}^n,[0,1])$.*

Proof. We already know from 6J.3 that the optimal trajectory mapping $y\mapsto x(y)$ is Lipschitz continuous from L^2 into C and hence from L^∞ into C too. Using the same argument for ψ in (16) we get that $y \mapsto \psi(y)$ is Lipschitz continuous from the L^2 to $C(\mathbb{R}^n, [0,1])$. But then, according to 2F.6, for every $y, y' \in L^\infty$ and almost every $t \in [0,1]$, with the optimal control values $u(y)(t)$ and $u(y')(t)$ at t being the unique solutions of (20), we have

$$|u(y)(t) - u(y')(t)| \leq \mu^{-1}\left(|r(t) - r'(t)| + |B||\psi(y)(t) - \psi(y')(t)|\right).$$

By invoking the L^∞ norm we get the desired result. □

We focus next on the issue of solving problem (1)–(3) numerically. By this we mean determining the optimal control function u. This is a matter of recovering a function on $[0,1]$ which is only specified implicitly, in this case by a variational problem. Aside from very special cases, it means producing numerically an acceptable approximation of the desired function u. For simplicity, let us assume that $y = (p,s,r) = 0$.

For such an approximation we need to choose a finite-dimensional space of functions on $[0,1]$ within L^2. Suppose that the interval $[0,1]$ is divided into N pieces $[t_i, t_{i+1}]$ by equally spaced nodes t_i, $i = 0,1,\ldots,N$, with $t_0 = 0$ and $t_N = 1$, the fixed mesh size being $h = t_i - t_{i-1} = 1/N$. To approximate the optimal control function u that is Lipschitz continuous on $[0,1]$ according to 6J.4, we will employ *piecewise constant* functions across the grid $\{t_i\}$ that are continuous from the right at each t_i, $i = 0,1,\ldots,N-1$, and from the left at $t_N = 1$. Specifically, for a given N, we consider the subset of L^2 given by

$$\left\{ u \,\middle|\, u(t) = u(t_i) \text{ for } t \in [t_i, t_{i+1}),\ i = 0,1,\ldots,N-2,\ u(t) = u(t_{N-1}) \text{ for } t \in [t_{N-1}, t_N] \right\}.$$

In order to fully discretize problem (1)–(3) and transform it into a finite-dimensional optimization problem, we also need to use finite-dimensional approximations of the operators of integration and differentiation involved.

Rather than (1)–(3), we now invoke a discretization of the optimality system (19ab). For solving the differential equations in (19a) we use the simplest Euler scheme over the mesh $\{t_i\}$. The Euler scheme applied to (19a), forward for the state equation and backward for the adjoint equation, combined with restricting the functional variational inequality (19b) to the nodes of the scheme, results in the following discrete-time boundary value problem coupled with a finite-dimensional variational inequality:

$$(21) \qquad \begin{cases} x_{i+1} = (I + hA)x_i + hBu_i, & x_0 = a, \\ \psi_i = (I + hA^{\mathsf{T}})\psi_{i+1} + hQx_{i+1}, & \psi_N = 0, \\ 0 \in Ru_i + B^{\mathsf{T}}\psi_i + N_U(u_i) & \qquad \text{for } i = 0,1,\ldots,N-1. \end{cases}$$

There are various numerical techniques for solving problems of this form; here we shall not discuss this issue.

We will now derive an estimate for the error in approximating the solution of problem (1)–(3) by use of discretization (21) of the optimality system (19ab).

We suppose that for each given N we can solve (21) exactly, obtaining vectors $u_i^N \in U$, $i = 0,\dots,N-1$, and $x_i^N \in \mathbb{R}^n$, $\psi_i^N \in \mathbb{R}^n$, $i = 0,\dots,N$. For a given N, the solution (u^N, x^N, ψ^N) of (21) is identified with a function on $[0,1]$, where x^N and ψ^N are the piecewise linear and continuous interpolations across the grid $\{t_i\}$ over $[0,1]$ of $(a, x_1^N, \dots, x_N^N)$ and $(\psi_0^N, \psi_1^N, \dots, \psi_{N-1}^N, 0)$, respectively, and u^N is the piecewise constant interpolation of $(u_0^N, u_1^N, \dots, u_{N-1}^N)$ which is continuous from the right across the grid points $t_i = ih$, $i = 0,1,\dots,N-1$, and from the left at $t_N = 1$. The functions x^N and ψ^N are piecewise differentiable and their derivatives $\dot{x}^N$ and $\dot{\psi}^N$ are piecewise constant functions which are assumed to have the same continuity properties in t as the control u^N. Thus, (u^N, x^N, ψ^N) is a function defined in the whole interval $[0,1]$, and it belongs to L^2.

Theorem 6J.6 (error estimate for discrete approximation). *Consider problem (1)–(3) with $r = 0$, $s = 0$ and $p = 0$ under condition (4) and let, according to 6J.4, (u, x, ψ) be the solution of the equivalent optimality system (19ab) for all $t \in [0,1]$, with u Lipschitz continuous in t on $[0,1]$. Consider also the discretization (21) and, for $N = 1,2,\dots$ and mesh size $h = 1/N$, denote by (u^N, x^N, ψ^N) its solution extended by interpolation to the interval $[0,1]$ in the manner described above. Then the following estimate holds:*

$$\|u^N - u\|_{L^\infty} + \|x^N - x\|_C + \|\psi^N - \psi\|_C = O(h). \tag{22}$$

Proof. For $t \in [t_i, t_{i+1})$, $i = 0,1,\dots,N-1$, let

$$\begin{aligned}
&p^N(t) = A(x^N(t_i) - x^N(t)),\\
&s^N(t) = A^{\mathsf{T}}(\psi^N(t_{i+1}) - \psi^N(t)) + Q(x^N(t_{i+1}) - x^N(t)),\\
&r^N(t) = -B^{\mathsf{T}}(\psi^N(t_i) - \psi^N(t)).
\end{aligned}$$

By virtue of the control u^N being piecewise constant and

$$\dot{x}^N(t) = \frac{x^N(t_{i+1}) - x^N(t_i)}{h}, \quad \dot{\psi}^N(t) = \frac{\psi^N(t_{i+1}) - \psi^N(t_i)}{h} \ \text{ for } t \in [t_i, t_{i+1}),$$

for $i = 0,1,\dots,N-1$, the discretized optimality system (21) can be written as follows: for all $t \in [t_i, t_{i+1})$, $i = 0,1,\dots,N-1$, and $t = 1$,

$$\begin{cases}
\dot{x}^N(t) = Ax^N(t) + Bu^N(t) + p^N(t), & x^N(0) = a,\\
\dot{\psi}^N(t) = -A^{\mathsf{T}}\psi^N(t) - Qx^N(t) - s^N(t), & \psi^N(1) = 0,\\
r^N(t) \in Ru^N(t) + B^{\mathsf{T}}\psi^N(t) + N_U(u^N(t)).
\end{cases} \tag{23}$$

Observe that system (23) has the same form as (19ab) with a particular choice of the parameters. Specifically, (u^N, x^N, ψ^N) is the solution of (19ab) for the parameter value $y^N := (p^N, s^N, r^N)$, while (u, x, ψ) is the solution of (19ab) for $y = (p, s, r) = (0,0,0)$. Then, by the implicit function theorem 6J.5, the solution mapping of (19) is Lipschitz continuous in the respective norms, so there exists a constant c such that

$$\|u^N - u\|_{L^\infty} + \|x^N - x\|_C + \|\psi^N - \psi\|_C \le c\|y^N\|_{L^\infty}. \tag{24}$$

To finish the proof, we need to show that

$$\|y^N\|_{L^\infty} = \max\{\|p^N\|_{L^\infty}, \|s^N\|_{L^\infty}, \|r^N\|_{L^\infty}\} = O(h). \tag{25}$$

For that purpose we employ the following standard result in the theory of difference equations which we state here without proof:

Lemma 6J.7 (discrete Gronwall lemma). *Consider reals α_i, $i = 0,\ldots,N$, which satisfy*

$$0 \le \alpha_0 \le a \quad \text{and} \quad 0 \le \alpha_{i+1} \le a + b\sum_{j=0}^{i} \alpha_j \quad \text{for } i = 0,\ldots,N.$$

Then $0 \le \alpha_i \le a(1+b)^i$ *for* $i = 0,\ldots,N$. *Similarly, if*

$$0 \le \alpha_N \le a \quad \text{and} \quad 0 \le \alpha_{i+1} \le a + b\sum_{j=i+1}^{N} \alpha_j \quad \text{for } i = 0,\ldots,N,$$

then $0 \le \alpha_i \le a(1+b)^{N-i}$ *for* $i = 0,\ldots,N$.

Continuing on this basis with the proof of (25), we observe that x^N is piecewise linear across the grid $\{t_i\}$; clearly

$$|x^N(t) - x^N(t_i)| \le |x^N(t_{i+1}) - x^N(t_i)| \quad \text{for } t \in [t_i, t_{i+1}],\ i = 0,1,\ldots,N-1.$$

Then, since all u_i are from the compact set U, from the first equation in (21) we get

$$|p^N(t)| \le h(c_1|x^N(t_i)| + c_2) \quad \text{for } t \in [t_i, t_{i+1}],\ i = 0,1,\ldots,N-1,$$

with some constants c_1, c_2 independent of N. On the other hand, the first equation in (21) can be written equivalently as

$$x^N(t_{i+1}) = a + \sum_{j=1}^{i} h(Ax^N(t_j) + Bu^N(t_j)),$$

and then, by taking norms and applying the direct part of discrete Gronwall lemma 6J.6, we obtain that $\sup_{0\le i\le N} |x^N(t_i)|$ is bounded by a constant which does not depend on N. This gives us error of order $O(h)$ for p^N in the maximum norm. By repeating this argument for the discrete adjoint equation (the second equation in (21)), but now applying the backward part of 6J.7, we get the same order of magnitude for s^N and r^N. This proves (25) and hence also (22). □

Note that the order of the discretization error in (22), $O(h)$, is sharp for the Euler scheme. Using higher-order schemes may improve the order of approximation, but this may require better properties in time of the optimal control than Lipschitz continuity which, as we already know, may be hard to obtain in the presence of constraints.

In the proof of 6J.5 we used the combination of the implicit function theorem 6J.4 for the variational system involved and the estimate (25) for the *residual* $y^N = (p^N, s^N, r^N)$ of the approximation scheme. The convergence to zero of the residual comes out of the approximation scheme and the continuity properties of the solution of the original problem with respect to time t; in numerical analysis this is called the *consistency* of the problem and its approximation. The property emerging from the implicit function theorem 6J.4, that is, the Lipschitz continuity of the solution with respect to the residual, is sometimes called *stability*. Theorem 6J.5 furnishes an illustration of a well-known paradigm in numerical analysis: *stability plus consistency yields convergence*.

Having the analysis of the linear-quadratic problem as a basis, we could proceed to more general nonlinear and nonconvex optimal control problems and obtain convergence of approximations and error estimates by applying more advanced implicit function theorems using, e.g., linearization of the associated nonlinear optimality systems. However, this would involve more sophisticated techniques which go beyond the scope of this book, so here is where we stop.

Commentary

Theorem 6A.2 is from Dontchev, Lewis and Rockafellar [2003], while Theorem 6A.3 was first shown by Lewis [1999]; see also Lewis [2001]. Theorem 6A.7 was initially proved in Dontchev, Lewis and Rockafellar [2003] by using the characterization of the metric regularity of a mapping in terms of the nonsingularity of its coderivative (see Section 4H) and applying the radius theorem for nonsingularity in 6A.2. The proof given here is from Dontchev, Quincampoix and Zlateva [2006]. For extensions to infinite-dimensional spaces see Ioffe [2003a,b], Ioffe and Sekiguchi [2009] and Sekiguchi [2010]. Theorems 6A.8 and 6A.9 are from Dontchev and Rockafellar [2004].

The material in Section 6B is basically from Dontchev, Lewis and Rockafellar [2003]. The results in Sections 6C have roots in several papers; see Rockafellar [1976a,b], Robinson [1994], Dontchev [2000], and Aragón Artacho, Dontchev, and Geoffroy [2007]. Section 6D is based on Aragón Artacho et al. [2011]. Section 6E includes results from Dontchev [2013a] and Dontchev and Rockafellar [2013]. Further results regarding superlinear convergence of the Broyden update under metric regularity are presented in Aragón Artacho et al. [2011]. Note that the proof of Theorem 6F.1 also works for smaller subdifferentials, e.g. for the B-subdifferential the convex hull of which is the Clarke generalized Jacobian. Proposition 6F.3 can be traced back to Fabian [1979] if not earlier; it is also present in Ioffe [1981] and in a more general form in Fabian and Preiss [1987]. The class of semismooth functions has been introduced by Mifflin [1977]. Starting with the pioneering works by Pang [1990], Qi and Sun [1993], and Robinson [1994], in the last twenty years there have been major developments in applying Newton-type methods to nonsmooth equations and variational problems, exhibited in a large number of papers and in the books by Klatte and Kummer [2002], Facchinei and Pang [2003], Ito and Kunisch [2008], and Ulbrich [2011]. Important insights to numerics of quasi-Newton methods in nonsmooth optimization can be found in Lewis and Overton [2013]. A link between the of semismooth functions and the semialgebraic functions is established in Bolte, Daniilidis and Lewis [2009]. Section 6G is based on Dontchev, Krastanov, Rockafellar and Veliov [2013].

Most of the results in Section 6H can be found in basic texts on variational methods; for a recent such book see Attouch, Buttazzo and Michaille [2006]. The result in 6I is from Dontchev and Veliov [2009], for more on metric regularity in optimal control see Quincampoix and Veliov [2013]. Section 6J presents a simplified version of a result in Dontchev [1996]; for advanced studies in this area see Malanowski [2001] and Veliov [2006].

References

Alt, W. [1990], The Lagrange-Newton method for infinite-dimensional optimization problems, Numerical Functional Analysis and Optimization, 11, 201–224.

Alt, W. and I. Kolumbán [1993], An implicit-function theorem for a class of monotone generalized equations, Kybernetika (Prague), 29, 210–221.

Apetrii, M., M. Durea and R. Strugariu [2013], On subregularity properties of set-valued mappings: applications to solid vector optimization, Set-valued analysis, 21, 93–126.

Apostol, T. [1962], *Calculus*, Blaisdell Publ. Co.

Aragón Artacho, F. J., A. L. Dontchev and M. H. Geoffroy [2007], Convergence of the proximal point method for metrically regular mappings, ESAIM Proceedings, 17, 1–8.

Aragón Artacho, F. J., A. L. Dontchev, M. Gaydu, M. H. Geoffroy and V. M. Veliov [2011], Metric regularity of Newton's iteration, SIAM Journal on Control and Optimization, 49, 339–362.

Aragón Artacho, F. J., A. Belyakov, A. L. Dontchev, M. López [2013], Local convergence of quasi-Newton methods under metric regularity, Computational Optimization and Applications, DOI 10.1007/s10589-013-9615-y.

Aragón Artacho, F. J. and B. Mordukhovich [2010], Metric regularity and Lipschitzian stability of parametric variational systems, Nonlinear Analysis, 72, 1149–1170.

Arutyunov, A. [2007], Covering mappings in metric spaces and fixed points, Doklady Academii Nauk, 416, 151–155.

Attouch, H., G. Buttazzo and G. Michaille [2006], *Variational analysis in Sobolev and BV spaces, Applications to PDEs and optimization,* MPS/SIAM Series on Optimization, 6.

Aubin, J-P. [1981], Contingent derivatives of set-valued maps and existence of solutions to nonlinear inclusions and differential inclusions, in *Mathematical analysis and applications, Part A,* Advances in Mathematics: Supplementary Studies, 7A, 159–229, Academic Press.

Aubin, J-P. [1984], Lipschitz behavior of solutions to convex minimization problems, Mathematics of Operations Research, 9, 87–111.

Aubin, J-P. and I. Ekeland [1984], *Applied nonlinear analysis,* John Wiley & Sons, Inc.

Aubin, J-P. and H. Frankowska [1987], On inverse function theorems for set-valued maps, Journal de Mathématiques Pures et Appliquées, 66, 71–89.

Aubin, J-P. and H. Frankowska [1990], *Set-valued analysis,* Systems and Control: Foundations & Applications, 2, Birkhäuser.

Aubin, J.-P. and H. Frankowska [1995], Partial differential inclusions governing feedback controls, Journal of Convex Analysis, 2, 19–40.

Azé, D. [2006], A unified theory for metric regularity of multifunctions, Journal of Convex Analysis, 13, 225–252.

Azé, D. and J.-N. Corvellec [2009], On some variational properties of metric spaces, Journal of Fixed Point Theory and Applications, 5, 185–200.

Azé, D., J.-N. Corvellec and R. Lucchetti [2002], Variational pairs and applications to stability in nonsmooth analysis, Nonlinear Analysis, Ser. A, 49, 643–670.

Banach, S. [1922], Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales, Fundamenta Mathematicae, 3, 133–181.

Banach, S. [1932], *Théorie des Operations Linéaires*, Monografje Matematyczne, Warszawa, English translation by North Holland, 1987.

Bank, B., J. Guddat, D. Klatte, B. Kummer and K. Tammer [1983], *Nonlinear parametric optimization,* Birkhäuser.

Bartle, R. G. and L. M. Graves [1952], Mappings between function spaces, Transactions of the American Mathematical Society, 72, 400–413.

Bartle, R. G. and D. R. Sherbert [1992], *Introduction to real analysis,* Second edition, John Wiley & Sons, Inc.

Beer, G. [1993], *Topologies on closed and closed convex sets,* Kluwer.

Bessis, D. N., Yu. S. Ledyaev and R. B. Vinter [2001], Dualization of the Euler and Hamiltonian inclusions, Nonlinear Analysis, 43, 861–882.

Bianchi, M., G. Kassay and R. Pini [2013], An inverse map result and some applications to sensitivity of generalized equations, Journal of Mathematical Analysis and Applications, 399, 279–290.

Bolte, J., A. Daniilidis and A. Lewis [2009], Tame functions are semismooth, Mathematical Programming, Ser. B, 117, 5–19.

Bonnans, J. F. and A. Shapiro [2000], *Perturbation analysis of optimization problems*, Springer Series in Operations Research, Springer.

Borwein, J. M. [1983], Adjoint process duality, Mathematics of Operations Research, 8, 403–434.

Borwein, J. M. [1986a], Stability and regular points of inequality systems, Journal of Optimization Theory and Applications, 48, 9–52.

Borwein, J. M. [1986b], Norm duality for convex processes and applications, Journal of Optimization Theory and Applications, 48, 53–64.

Borwein, J. M. and A. L. Dontchev [2003], On the Bartle-Graves theorem, Proceedings of the American Mathematical Society, 131, 2553–2560.

Borwein, J. M. and A. S. Lewis [2006], *Convex analysis and nonlinear optimization. Theory and examples*, Second edition, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC 3, Springer.

Borwein, J. M. and Q. J. Zhu [2005], *Techniques of variational analysis,* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC 20, Springer.

Cauchy, A-L. [1831], Résumé d'un mémoir sur la mécanique céleste et sur un nouveau calcul appelé calcul des limites, in *Oeuvres Complétes d'Augustun Cauchy,* volume 12, 48–112, Gauthier-Villars, Paris 1916, Edited by L'Académie des Sciences. The part pertaining to series expansions was read at a meeting of the Academie de Turin, 11 October, 1831.

Chipman, J. S. [1997], "Proofs" and proofs of the Eckart-Young theorem, in *Stochastic processes and functional analysis*, 71–83, Lecture Notes in Pure and Applied Mathematics, 186, Dekker, New York.

Cibulka, R. [2011], Constrained open mapping theorem with applications, Journal of Mathematical Analysis and Applications, 379, 205–215.

Cibulka, R. and M. Fabian [2013], A note on Robinson-Ursescu and Lyusternik-Graves theorem, Mathematical Programming, Ser. A, 139, 89–101.

Clarke, F. H. [1976], On the inverse function theorem, Pacific Journal of Mathematics, 64, 97–102.

Clarke, F. H. [1983], *Optimization and nonsmooth analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts, A Wiley-Interscience Publication, John Wiley & Sons, Inc.

Clarke, F. H. [2013], *Functional analysis, calclulus of variations and optimal control*, GTM 264, Springer.

Cornet, B. and G. Laroque [1987], Lipschitz properties of solutions in mathematical programming, Journal of Optimization Theory and Applications, 53, 407–427.

Cottle, R. W., J.-S. Pang and R. E. Stone [1992], *The linear complementarity problem,* Academic Press.

Courant, R. [1988], *Differential and integral calculus,* Vol. II, Translated from the German by E. J. McShane, Reprint of the 1936 original, Wiley Classics Library, A Wiley-Interscience Publication, John Wiley & Sons, Inc.

Dal Maso, G. and H. Frankowska [2000], Uniqueness of solutions to Hamilton-Jacobi equations arising in the calculus of variations, in *Optimal Control and Partial Differential Equations,* 336–345, IOS Press.

De Giorgi, E., A. Marino and M. Tosques [1980], Problems of evolution in metric spaces and maximal decreasing curve (Italian), Atti Accad. Cl. Sci. Naz. Lincei Rend. Fis. Mat. Natur. (8) 68 (1980), 180–187.

Deimling, K. [1992], *Multivalued differential equations*, Walter de Gruyter, Berlin.

Deville, R., G. Godefroy and V. Zizler [1993], *Smoothness and renormings in Banach spaces*, Pitman Monographs and Surveys in Pure and Applied Mathematics, 64. Longman Scientific & Technical, Harlow; copublished with John Wiley & Sons, Inc.

Deutsch, F. [2001], *Best approximation in inner product spaces,* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC 7, Springer.

Dieudonné, J. [1969], *Foundations of modern analysis*, Enlarged and corrected printing, Pure and Applied Mathematics, Vol. 10-I, Academic Press.

Dini, U. [1877/78], *Analisi infinitesimale,* Lezioni dettate nella R. Universitá di Pisa.

Dmitruk, A. V., A. A. Milyutin and N. P. Osmolovskiĭ [1980], Lyusternik's theorem and the theory of extremum, Uspekhi Matematicheskikh Nauk, 35, no. 6 (216), 11–46 (Russian).

Dmitruk, A. V. and A. Y. Kruger [2009], Extensions of metric regularity, Optimization, 58, 561–584.

Dolecki, S. and G. H. Greco [2011], Tangency vis-á-vis differentiability by Peano, Severi and Guareschi, Journal of Convex Analysis, 18, 301–339.

Dontchev, A. L. [1981], Efficient estimates of the solutions of perturbed control problems, Journal of Optimization Theory and Applications, 35, 85–109.

Dontchev, A. L. [1995a], Characterizations of Lipschitz stability in optimization, in *Recent developments in well-posed variational problems,* 95–115, Kluwer.

Dontchev, A. L. [1995b], Implicit function theorems for generalized equations, Mathematical Programming, Ser. A, 70, 91–106.

Dontchev, A. L. [1996], An a priori estimate for discrete approximations in nonlinear optimal control, SIAM Journal of Control and Optimization, 34, 1315–1328.

Dontchev, A. L. [2000], Lipschitzian stability of Newton's method for variational inclusions, in *System modelling and optimization (Cambridge, 1999)*, 119–147, Kluwer.

Dontchev, A. L. [2004], A local selection theorem for metrically regular mappings, Journal of Convex Analysis, 11, 81–94.

Dontchev, A. L. [2013a] Generalizations of the Dennis-More theorem, SIAM Journal on Optimization, 22, 821–830.

Dontchev, A. L. and H. Frankowska [2011], Lyusternik-Graves theorem and fixed points, Proceedings of the American Mathematical Society, 139, 521–534.

Dontchev, A. L. and H. Frankowska [2012], Lyusternik-Graves theorem and fixed points II, Journal of Convex Analysis, 19, 975–997.

Dontchev, A. L. and H. Frankowska [2013], On derivative criteria for metric regularity, in *Computational and Analytical Mathematics* 50, 365–374, Springer.

Dontchev, A. L. and W. W. Hager [1993], Lipschitzian stability in nonlinear control and optimization, SIAM Journal of Control and Optimization, 31, 569–603.

Dontchev, A. L. and W. W. Hager [1994], An inverse mapping theorem for set-valued maps, Proceedings of the American Mathematical Society, 121, 481–489.

Dontchev, A. L., M. I. Krastanov, R. T. Rockafellar and V. M. Veliov [2013], An Euler–Newton continuation method for tracking solution trajectories of parametric variational inequalities, SIAM Journal on Control and Optimization, 51, 1823–1840.

Dontchev, A. L., A. S. Lewis and R. T. Rockafellar [2003], The radius of metric regularity, Transactions of the American Mathematical Society, 355, 493–517.

Dontchev, A. L. and R. T. Rockafellar [1996], Characterizations of strong regularity for variational inequalities over polyhedral convex sets, SIAM Journal on Optimization, 6, 1087–1105.

Dontchev, A. L. and R. T. Rockafellar [1998], Characterizations of Lipschitzian stability in nonlinear programming, in Mathematical programming with data perturbations, 65–82, Lecture Notes in Pure and Appl. Math., 195, Dekker, New York.

Dontchev, A. L. and R. T. Rockafellar [2001], Ample parameterization of variational inclusions, SIAM Journal on Optimization, 12, 170–187.

Dontchev, A. L. and R. T. Rockafellar [2004], Regularity and conditioning of solution mappings in variational analysis, Set-Valued Analysis, 12, 79–109.

Dontchev, A. L. and R. T. Rockafellar [2009], Robinson's implicit function theorem and its extensions, Mathematical Programming, Ser. B, 117, 129–147.

Dontchev, A. L. and R. T. Rockafellar [2013], Convergence of inexact Newton methods for generalized equations, Mathematical Programming, Ser. B, 139, 115–137.

Dontchev, A. L. and V. M. Veliov [2009], Metric regularity under approximations, Control and Cybernetics, 38, 1283–1303.

Dontchev, A. L. and T. Zolezzi [1993], *Well-posed optimization problems*, Lecture Notes in Mathematics 1543, Springer.

Dontchev, A. L., M. Quincampoix and N. Zlateva [2006], Aubin criterion for metric regularity, Journal of Convex Analysis, 13, 281–297.

Dunford, N. and J. T. Schwartz [1958], *Linear Operators I. General Theory,* with the assistance of W. G. Bade and R. G. Bartle. Pure and Applied Mathematics, 7, Interscience.

Durea, M. and R. Strugariu [2012a], Chain rules for linear openness in general Banach spaces, SIAM Journal on Optimization, 22, 899–913.

Durea, M. and R. Strugariu [2012b], Openness stability and implicit multifunction theorems: applications to variational systems. Nonlinear Analysis 75, 1246–1259.

Eckart, C. and G. Young [1936], The approximation of one matrix by another of lower rank, Psychometrica, 1, 211-218.

Ekeland, I. [1990], The ε-variational principle revisited, with notes by S. Terracini, in *Methods of nonconvex analysis (Varenna 1989),* 1–15, Lecture Notes in Mathematics, 1446, Springer.

Ekeland, I. [2011], An inverse function theorem in Fréchet spaces, Annales de l'Institut Henri Poincaré, Analyse Non Linéaire, 28, 91–105.

Fabian, M. [1979], Concerning interior mapping theorem, Commentationes Mathematicae Universitatis Carolinae, 20, 345–356.

Fabian, M. and D. Preiss [1987], A generalization of the interior mapping theorem of Clarke and Pourciau, Commentationes Mathematicae Universitatis Carolinae, 28, 311–324.

Facchinei, F. and J-S. Pang [2003], *Finite-dimensional variational inequalities and complementarity problems,* Springer.

Fiacco, A. V. and G. P. McCormick [1968], *Nonlinear programming: Sequential unconstrained minimization techniques,* John Wiley & Sons, Inc.

Fikhtengolts, G. M. [1962], A course on differential and integral calculus, Fifth Edition, Moscow (Russian).

Fitzpatrick, P. [2006], *Advanced calculus*, Thomson Brooks/Cole.

Frankowska, H. [1990], Some inverse mapping theorems, Ann. Inst. H. Poincaré, Anal. Non Linéaire, 7, 183–234

Frankowska, H. [1992], Conical inverse function theorems, Bulletin of the Australian Mathematical Society, 45, 53–60.

Gfrerer, H. [2013], On directional metric subregularity and second-order optimality conditions for a class of nonsmooth mathematical programs, SIAM Journal on Optimization, 23, 632–665.

Golub, G. H. and C. F. Van Loan [1996], *Matrix computations,* The John Hopkins University Press, Baltimore, MD, 3d edition.

Goursat, Ed. [1903], Sur la théorie des fonctions implicites, Bulletin de la Société Mathématiques de France, 31, 184–192.

Goursat, E. [1904], *A course in mathematical analysis*, English translation by E. R. Hedrick, Ginn Co., Boston.

Graves, L. M. [1950], Some mapping theorems, Duke Mathematical Journal, 17, 111–114.

Halkin, H. [1974], Implicit functions and optimization problems without continuous differentiability of the data, SIAM Journal on Control, 12, 229–236.

Hamilton, R. S. [1982], The inverse function theorem of Nash and Moser, Bulletin of the American Mathematical Society (N.S.), 7, 65-222.

Hausdorff, F. [1927], *Mengenlehre*, Walter de Gruyter and Co., Berlin.

Hedrick, E. R. and W. D. A. Westfall [1916], Sur l'existence des fonctions implicites, Bulletin de la Société Mathématiques de France, 44, 1–13.

Hildebrand, H. and L. M. Graves [1927], Implicit functions and their differentials in general analysis, Transactions of the American Mathematical Society, 29, 127–153.

Hoffman, A. J. [1952], On approximate solutions of systems of linear inequalities, Journal of Research of the National Bureau of Standards, 49, 263–265.

Huynh, Van Ngai, A. Y. Kruger and M. Théra [2012], Slopes of multifunctions and extensions of metric regularity, Vietnam Journal of Mathematics 40, 355–369.

Huynh, Van Ngai, Huu Tron Nguyen and M. Théra [2013], Implicit multifunction theorems in complete metric spaces, Mathematical Programming, Ser. A, 139, 301–326.

Hurwicz, L. and M. K. Richter [2003], Implicit functions and diffeomorphisms without C^1, *Advances in mathematical economics*, 5, 65–96.

Ioffe, A. D. [1979], Regular points of Lipschitz functions, Transactions of the American Mathematical Society, 251, 61–69.

Ioffe, A. D. [1981], Nonsmooth analysis: differential calculus of nondifferentiable mappings, Transactions of the American Mathematical Society, 266, 1–56.

Ioffe, A. D. [1984], Approximate subdifferentials and applications, I. The finite-dimensional theory, Transactions of the American Mathematical Society, 281, 389–416.

Ioffe, A. D. [2000], Metric regularity and subdifferential calculus, Uspekhi Mat. Nauk, 55, no. 3 (333), 103–162 (Russian), English translation in Russian Mathematical Surveys, 55, 501–558.

Ioffe, A. D. [2001], Towards metric theory of metric regularity, in Approximation, optimization and mathematical economics, 165–176, Physica, Heidelberg, 2001.

Ioffe, A. D. [2003a], On robustness of the regularity property of maps, Control and Cybernetics, 32, 543–554.

Ioffe, A. D. [2003b], On stability estimates for the regularity property of maps, in *Topological methods, variational methods and their applications*, 133–142, World Science Publishing.

Ioffe, A. D. [2008], Critical values of set-valued maps with stratifiable graphs. Extensions of Sard and Smale-Sard theorems, Proceedings of American Mathematical Society, 136, 3111–3119.

Ioffe, A. D. [2010a], Towards variational analysis in metric spaces: metric regularity and fixed points, Mathematical Programming, Ser. B, 123, 241–252.

Ioffe, A. D. [2010b], On regularity concepts in variational analysis, Journal of Fixed Point Theory and Applications, 8, 339–363.

Ioffe, A. D. [2011], Regularity on a fixed set, SIAM Journal on Optimization, 21, 1345–1370.

Ioffe, A. D. [2013], Nonlinear regularity models, Mathematical Programming, Ser. B, 139, 223–242.

Ioffe, A. D. and V. M. Tikhomirov [1974], *Theory of extremal problems*, Nauka, Moscow (Russian).

Ioffe, A. D. and Y. Sekiguchi [2009], Regularity estimates for convex multifunctions, Mathematical Programming, Ser. B, 117, 255–270.

Ito, K. and K. Kunisch [2008], *Lagrange multiplier approach to variational problems and applications,* SIAM, Philadelphia, PA.

Izmailov, A. F. [2013], Strongly regular nonsmooth generalized equations, Mathematical Programming, Ser. A, DOI 10.1007/s10107-013-0717-1.

Jongen, H. Th., T. Möbert, J. Rückmann and K. Tammer [1987], On inertia and Schur complement in optimization, Linear Algebra and Applications, 95, 97–109.

Jongen, H. Th., D. Klatte and K. Tammer [1990/91], Implicit functions and sensitivity of stationary points, Mathematical Programming, Ser. A, 49, 123–138.

Josephy, N. H. [1979], Newton's method for generalized equations and the PIES energy model, Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin-Madison.

Kahan, W. [1966], Numerical linear algebra, Canadian Mathematics Bulletin, 9, 757–801.

Kantorovich, L. V. and G. P. Akilov [1964], *Functional analysis in normed spaces*, Macmillan, New York. Translation by D. E. Brown from Russian original, Fizmatgiz, Moscow, 1959.

Kassay, G. and I. Kolumbán [1988], Implicit function theorems for monotone mappings, Seminar on Mathematical Analysis (Cluj-Napoca, 1988–1989), 7–24, Preprint, 88-7, Univ. "Babes-Bolyai", Cluj-Napoca.

Kassay, G. and I. Kolumbán [1989], Implicit functions and variational inequalities for monotone mappings, Seminar on Mathematical Analysis (Cluj-Napoca, 1988–1989), 79–92, Preprint, 89-7, Univ. "Babes-Bolyai", Cluj-Napoca.

Kenderov, P. [1975], Semi-continuity of set-valued monotone mappings, Fundamenta Mathematicae, 88, 61–69.

King, A. J. and R. T. Rockafellar [1992], Sensitivity analysis for nonsmooth generalized equations, Mathematical Programming, Ser. A, 55(2), 193–212.

Klatte, D. and B. Kummer [2002], *Nonsmooth equations in optimization. Regularity, calculus, methods and applications,* Kluwer.

Klatte, D., A. Kruger and B. Kummer [2012], From convergence principles to stability and optimality conditions, Journal of Convex Analysis, 19, 1043–1072.

Klatte, D. and B. Kummer [2013], Aubin property and uniqueness of solutions in cone constrained optimization, Mathematical Methods in Operations Research, 77, 291–304.

Kojima, M. [1980], Strongly stable stationary solutions in nonlinear programs, in *Analysis and computation of fixed points (Madison, Wisconsin, 1979)*, 93–138, Academic Press.

Krantz, S. G. and H. R. Parks [2002], *The implicit function theorem. History, theory, and applications*, Birkhäuser.

Kruger, A. [1982], On characterizing the covering property of nonsmooth operators, in *Proceedings of the School on Theory of Operators in Function Spaces, Minsk,* 94–95 (Russian).

Kummer, B. [1991], An implicit-function theorem for $C^{0,1}$-equations and parametric $C^{1,1}$-optimization, Journal of Mathematical Analysis and Applications, 158, 35–46.

Kuratowski, K. [1933], *Topologie,* I & II, Panstwowe Wydawnictwo Naukowe, Warszawa.

Lamson, K. W. [1920], A general implicit function theorem with an application to problems of relative minima, American Journal of Mathematics, 42, 243–256.

Leach, E. B. [1961], A note on inverse function theorems, Proceedings of the American Mathematical Society, 12, 694–697.

Ledyaev, Yu. S. and Q. J. Zhu [1999], Implicit multifunction theorems, Set-Valued Analysis, 7, 209–238.

Levitin, E. S. [1992], *Perturbation theory in mathematical programming and its applications*, Nauka , Moscow (Russian).

Levy, A. B. [1996], Implicit multifunction theorems for the sensitivity analysis of variational conditions, Mathematical Programming, Ser. A, 74(3), 333–350.

Levy, A. B. and R. A. Poliquin [1997], Characterizing the single-valuedness of multifunctions, Set-Valued Analysis, 5, 351–364.

Lewis, A. S. [1999], Ill-conditioned convex processes and conic linear systems, Mathematics of Operations Research, 24, 829–834.

Lewis, A. S. [2001], Ill-conditioned inclusions, Set-Valued Analysis, 9, 375–381.

Lewis A. S and M. L. Overton [2013], Nonsmooth optimization via quasi-Newton methods, Mathematical Programming, Ser. A, 141, 135–163.

Li, Wu [1994], Sharp Lipschitz constants for basic optimal solutions and basic feasible solutions of linear programs, SIAM Journal on Control and Optimization, 32, 140–153.

Lim, T.-C. [1985], On fixed-point stability for set-valued contractive mappings with applications to generalized differential equations, Journal of Mathematical Analysis and Applications, 110, 436–441.

Lu, Shu and S. M. Robinson [2008], Variational inequalities over perturbed polyhedral convex sets, Mathematics of Operations Research, 33, 689–711.

Lyusternik, L. A. [1934], On the conditional extrema of functionals, Mat. Sbornik, 41, 390–401 (Russian).

Lyusternik, L. A. and V. I. Sobolev [1965], *Elements of functional analysis,* Nauka, Moscow (Russian).

Malanowski, K. [2001], Stability and sensitivity analysis for optimal control problems with control-state constraints, Dissertationes Mathematicae, 394.

Mifflin, R [1977], Semismooth and semiconvex functions in constrained optimization, SIAM J. Control Optimization 15, 959–972.

Mordukhovich, B. S. [1984], Nonsmooth analysis with nonconvex generalized differentials and conjugate mappings, Doklady Akad. Nauk BSSR, 28, 976–979, (Russian).

Mordukhovich, B. S. [2006], *Variational analysis and generalized differentiation, I. Basic theory,* Springer.

Nadler, S. B., Jr. [1969], Multi-valued contraction mappings, Pacific Journal of Mathematics, 30, 475–488.

Ng, Kung Fu [1973], An open mapping theorem, Proceedings of Cambridge Philosophical Society, 74, 61–66.

Nijenhuis, A. [1974], Strong derivatives and inverse mappings, The American Mathematical Monthly, 81, 969–980,

Noble, B. and J. W. Daniel [1977], *Applied linear algebra*, Second edition, Prentice-Hall, Inc., Englewood Cliffs.

Pang, C. H. J. [2011], Generalized differentiation with positively homogeneous maps: applications in set-valued analysis and metric regularity, Mathematics of Operations Research, 36, 377–397.

Pang, J.-S. [1990], Newton's method for B-differentiable equations, Mathematics of Operations Research, 15, 311–341.

Páles, Z. [1997], Inverse and implicit function theorems for nonsmooth maps in Banach spaces, Journal of Mathematical Analysis and Applications, 209, 202–220.

Peano G. [1892], Sur la définition de la dérivée, Mathesis 2, 12–14.

Penot, J.-P. [2013], *Calculus without derivatives,* GTM 266, Springer.

Pompeiu, D. [1905], Fonctions de variable complexes, Annales de la Faculté des Sciences de l´ Université de Toulouse, 7, 265–345.

Qi, Li Qun and Jie Sun [1993], A nonsmooth version of Newton's method, Mathematical Programming, Ser. A, 58, 353–367.

Quincampoix, M. and V. M. Veliov [2013], Metric regularity and stability of optimal control problems for linear systems, SIAM Journal on Control and Optimization, 51, 4118–4137.

Ralph, D. [1993], A new proof of Robinson's homeomorphism theorem for PL-normal maps, Linear Algebra and Applications, 178, 249–260.

Robinson, S. M. [1972], Normed convex processes, Transactions of the American Mathematical Society, 174, 127–140.

Robinson, S. M. [1976], Regularity and stability for convex multivalued functions, Mathematics of Operations Research, 1, 130–143.

Robinson, S. M. [1979], Generalized equations and their solutions. I Basic theory. Point-to-set maps and mathematical programming, Mathematical Programming Study, 10, 128–141.

Robinson, S. M. [1980], Strongly regular generalized equations, Mathematics of Operations Research, 5, 43–62.

Robinson, S. M. [1981], Some continuity properties of polyhedral multifunctions, Mathematical Programming Study, 14, 206–214.

Robinson, S. M. [1984], Local structure of feasible sets in nonlinear programming, Part II: Nondegeneracy, Mathematical Programming Study, 22, 217–230.

Robinson, S. M. [1991], An implicit-function theorem for a class of nonsmooth functions, Mathematics of Operations Research, 16, 292–309.

Robinson, S. M. [1992], Normal maps induced by linear transformations, Mathematics of Operations Research, 17, 691–714.

Robinson, S. M. [1994], Newton's method for a class of nonsmooth functions, Set-Valued Analysis, 2, 291–305.

Robinson, S. M. [2007], Solution continuity in monotone affine variational inequalities, SIAM Journal on Optimization, 18, 1046-1060.

Robinson, S. M. [2013], Equations on monotone graphs, Mathematical Programming, Ser. A, 141, 49–101.

Rockafellar, R. T. [1967], Monotone processes of convex and concave type, Memoirs of the American Mathematical Society, 77.

Rockafellar, R. T. [1970], *Convex analysis,* Princeton University Press.

Rockafellar, R. T. [1974], Conjugate duality and optimization, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, 16, SIAM.

Rockafellar, R. T. [1976a], Monotone operators and the proximal point algorithm, SIAM Journal on Control and Optimization, 14, 877–898.

Rockafellar, R. T. [1976b], Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Mathematics of Operations Research, 1, 97–116.

Rockafellar, R. T. [1985], Lipschitzian properties of multifunctions, Nonlinear Analysis, 9, 867–885.

Rockafellar, R. T. [1989], Proto-differentiability of set-valued mappings and its applications in optimization, in *Analyse non linéaire (Perpignan 1987)*, Annales de l'Institut Henri Poincaré, Analyse Non Linéaire, 6, suppl., 448–482.

Rockafellar, R. T. and R. J-B. Wets [1998], *Variational Analysis*, Springer.

Rudin, W. [1991], *Functional analysis,* Second edition, McGraw-Hill, Inc.

Sekiguchi, Y. [2010], Exact estimates of regularity modulus for infinite programming, Mathematical Programming, Ser. B, 123, 253–263.

Schirotzek, W. [2007], *Nonsmooth analysis*, Springer.

Scholtes, S. [2012], *Introduction to piecewise differentiable functions*, Springer Briefs in Optimization, Springer.

Shvartsman, I. [2012], On stability of minimizers in convex programming, Nonlinear Analysis, 75, 1563–1571.

Spanier, E. H. [1966], *Algebraic topology*, McGraw-Hill.

Stewart, G. W. and J. G. Sun [1990], *Matrix perturbation theory*, Academic Press.

Schwartz, L. [1967], *Analyse mathématique,* I, Hermann.

Uderzo A. [2009], On some regularity properties in variational analysis, Set-Valued and Variational Analysis 17, 409–430.

Uderzo A. [2012a], On mappings covering at a nonlinear rate and their perturbation stability, Nonlinear Analysis, 75, 1602–1616.

Uderzo A. [2012b], A metric version of Milyutin theorem, Set-valued and Variational Analysis, 20, 279–306.

Ulbrich M. [2011], *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces,* SIAM, Philadelphia, PA.

Ursescu, C. [1975], Multifunctions with convex closed graph, Czechoslovak Mathematical Journal, 25 (100), 438–441.

Ursescu, C. [1996], Inherited openness, Revue Roumaine des Mathématiques Pures et Appliquées, 41, 401–416.

Veliov, V. M. [2006], Approximations with error estimates for optimal control problems for linear systems, in *Large-scale scientific computing,* 263–270, Lecture Notes in Computer Science, 3743, Springer.

Walkup, D. W. and R. J-B. Wets [1969], A Lipschitzian characterization of convex polyhedra, Proceedings of the American Mathematical Society, 23, 167–173.

Zălinescu, C. [2002], *Convex analysis in general vector spaces,* World Scientific Publishing.

Zheng X. Y. and K. F. Ng [2009], Metric regularity of composite multifunctions in Banach spaces, Taiwanese Journal of Mathematics, 13, 1723–1735.

Index