SECOND-ORDER VARIATIONAL ANALYSIS AND ITS ROLE IN OPTIMIZATION

R. Tyrrell Rockafellar[1]

Abstract

Basic tools in variational analysis, such as geometric approaches to convergence of sequences of extended-real-valued functions or set-valued mappings, support a theory of generalized differentiation with second-order capabilities. The resulting second-derivative objects are fundamental to the understanding of how solutions to parameterized problems of optimization behave when parameter values are shifted.

Keywords: *variational analysis, parameterized optimization, solution perturbations, graphical derivatives, second-order epi-derivatives, Attouch's theorem, Wijsman's theorem*

March 7, 2014

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: http://www.math.washington.edu/~rtr/mypage.html

1 Introduction

Variational analysis is built on convex analysis but bridges beyond it to nonconvexity, as motivated especially by applications in optimization. It relies on many of the same ideas that are fundamental to convex analysis in contrast to classical calculus, like epigraphs of extended-real-valued functions, but extends the definitions of concepts like tangent cones, normal cones and subgradients so that they can work effectively with nonconvex sets and functions as well.

Those concepts are essential tools in generalized *first-order* differentiation. However, variational analysis is capable also of developing generalized *second-order* differentiation, which was not fully possible in the confines of convex analysis even for convex functions. The purpose of this article is to provide an introductory picture of such developments in the hopes of stimulating further research.

A central theme of variational analysis with deep connections to second-order theory is the study of perturbations and approximations of solutions to problems of optimization. Understanding perturbations is important to validating stability in mathematical models. Approximations are of course the key to computations. In standard nonlinear programming, for instance, second-order conditions for optimality are employed in determining rates of convergence of solution algorithms.

A major difference between variational analysis and classical analysis in dealing with convergence of approximations is that the approach of variational analysis is far more geometric in its core. Much of classical analysis focuses on sequences of functions that converge to a limit pointwise uniformly over some domain. Variational analysis starts instead with sequences of subsets of a vector space that converge to a limit set with respect to associated distance functions, and it applies that to functions through their epigraphs to arrive at *epiconvergence* of sequences of extended-real-valued functions.

In optimization, extended-real-valued functions can represent entire problems of minimization by the device of assigning ∞ as the penalty for violating contraints, and therefore epiconvergence of functions can stand for a kind of convergence of problems of optimization. It is known in fact that this is exactly the right notion to employ when looking for convergence of optimal values and optimal solutions [7, Chapter 7].

To explore this framework further, consider a function

$$f : I\!R^n \to (-\infty, \infty], \quad f \not\equiv \infty,$$

its effective domain

$$\operatorname{dom} f = \{\, x \in I\!R^n \mid f(x) < \infty\} \neq \emptyset$$

and its epigraph

$$\operatorname{epi} f = \{\, (x, \alpha) \in I\!R^n \times I\!R \mid f(x) \leq \alpha\},$$

along with the associated optimization problem

$$\text{minimize } \; f(x) \; \text{ over all } \; x \in I\!R^n. \tag{1}$$

In this problem no x with $f(x) = \infty$ can be a candidate for giving the minimum, inasmuch as there exists at least one x with $f(x) < \infty$. Thus the minimization effectively takes place over the set $\operatorname{dom} f$ instead of all of $I\!R^n$, and the problem therefore has

$$\begin{array}{ll} \text{feasible solution set:} & \operatorname{dom} f \subset I\!R^n, \\ \text{optimal solution set:} & \operatorname{argmin} f \subset I\!R^n, \\ \text{optimal value:} & \inf f \in [-\infty, \infty). \end{array} \tag{2}$$

An approximation to this problem would correspond to specifying a different function

$$g : I\!R^n \to (-\infty, \infty], \quad g \not\equiv \infty,$$

which would replace the elements in (2) by $\operatorname{dom} g$, $\operatorname{argmin} g$ and $\inf g$. Certainly we would be interested in knowing that $\operatorname{argmin} g$ and $\inf g$ are not too far away from $\operatorname{argmin} f$ and $\inf f$, but for feasibility it may not matter whether the set $\operatorname{dom} g$ is nearly the same as $\operatorname{dom} f$. For instance, the infinite penalty associated with points outside of $\operatorname{dom} f$ could be changed to a very high but finite penalty in the case of the approximate problem, so that $\operatorname{dom} g$ would all of $I\!R^n$. Indeed, such a penalty approach to constraints is common and practical.

Therefore, in assessing whether the g-problem is "close" to the f-problem it would make no sense to rely on classical expressions for comparing f and g like $\sup\{\,|f(x) - g(x)|\,|\, x \in X\}$ over various subsets $X \subset I\!R$. Those expressions would in particular trivialize to ∞ unless $X \subset \operatorname{dom} f \cap \operatorname{dom} g$ and thus could not capture relaxations of constraints. The right approach instead is to compare the epigraphs $\operatorname{epi} g$ and $\operatorname{epi} f$ as subsets of $I\!R^n \times I\!R$.

Suppose now that in addition to $x = (x_1, \ldots, x_n)$ there is a parameter vector $p = (p_1, \ldots, p_d)$ on which the optimization problem depends. This can be accommodated by thinking of a function

$$f : I\!R^d \times I\!R^n \to (-\infty, \infty], \quad f \not\equiv \infty,$$

and the family of optimization problems

$$\text{minimize } f(p, x) \text{ over all } x \in I\!R^n, \text{ parameterized by } p \in I\!R^d. \tag{3}$$

with associated

$$\begin{array}{ll} \text{feasible solution sets:} & \operatorname{dom} f(p, \cdot), \\ \text{optimal solution sets:} & \operatorname{argmin} f(p, \cdot), \\ \text{optimal values:} & \inf f(p, \cdot). \end{array} \tag{4}$$

The behavior of these objects with respect to shifts in p then becomes important. Is it somehow "continuous" or even "differentiable"?

These are the kinds of questions which variational analysis is well suited to answer, but classical analysis cannot, and for which convex analysis falls short even when $f(p, x)$ is convex in x. The essential platform has to be how the epigraph set $\operatorname{epi} f(p, \cdot) \subset I\!R^n \times I\!R$ behaves with respect to p. This pushes to the forefront the challenge of understanding the generalized senses in which a "set-valued mapping" from p to $\operatorname{epi} f(p, \cdot)$ may be deemed continuous or differentiable.

The same kinds of questions are central to second-order theory, too, as is easy to see. One avenue towards second-order differentiation of a function f would be try to differentiate its subgradient mapping ∂f. But that is a set-valued mapping, in general. Another avenue would be to investigate function expressions corresponding to generalized second-order difference quotients depending on an ε and determine whether they converge to something as ε tends to 0. But that leads back to the issue of what would be the appropriate form of convergence for sequences of extended-real-valued functions.

2 Geometric Approaches to Convergence and Approximation

The first step toward progress in gaining fundamental insights into approximations of the kind important in optimization is getting a good idea of when two closed subsets C and D of $I\!R^n$ should be considered "close" to each other.

There is a classical concept, known as the Hausdorff distance between sets, which very well known, but it can only do a good job when C and D are bounded, hence compact. Something much broader is needed for our purposes because we will eventually want to apply the ideas to see whether two sets $\operatorname{epi} f$ and $\operatorname{epi} g$ are "close" to each other in $\mathbb{R}^n \times \mathbb{R}$, and epigraph sets are intrinsically unbounded. (Recall that the closedness of $\operatorname{epi} f$ and $\operatorname{epi} g$ corresponds to the functions f and g being *lower semicontinuous*, which is a natural property to ask for in the representation of problems of minimization.)

The *distance function* on $\mathbb{R}^n$ for a nonempty closed set C in $\mathbb{R}^n$ is defined by

$$\operatorname{dist}(x, C) = \inf \{ |y - x| \mid y \in C \}. \tag{5}$$

It is nonnegative and takes the value 0 at x if and only if $x \in C$ (since C is closed). Thus the correspondence between closed sets and their distance functions is one-to-one. An elementary property of distance functions is that they are Lipschitz continuous with constant 1:

$$|\operatorname{dist}(x, C) - \operatorname{dist}(y, C)| \leq |y - x| \text{ (Euclidean norm).}$$

Definition (set convergence)**.** *A sequence of closed subsets $C^\nu \subset \mathbb{R}^n$ for $\nu = 1, 2, \ldots$ converges to a closed subset $C \subset \mathbb{R}^n$ when*

$$\operatorname{dist}(x, C^\nu) \to \operatorname{dist}(x, C) \quad \textit{for all } x \in \mathbb{R}^n \textit{ as } \nu \to \infty,$$

In this case, because of Lipschitz continuity, the functions $\operatorname{dist}(\cdot, C^\nu)$ actually converge to $\operatorname{dist}(\cdot, C)$ not just pointwise, but uniformly over all bounded subsets of $\mathbb{R}^n$. Classical Hausdorff convergence would correspond to the distance function convergence being uniform over $\mathbb{R}^n$ itself, which for many purposes is far too restrictive.

This geometric approach can be applied as follows to obtain a distinctly nonclassical concept in the convergence of functions.

Definition (epi-convergence)**.** *A sequence of lower semicontinous functions $f^\nu : \mathbb{R}^n \to (-\infty, \infty]$ converges to a lower semicontinous function $f : \mathbb{R}^n \to (-\infty, \infty]$ when the sets $\operatorname{epi} f^\nu$ in $\mathbb{R}^n \times \mathbb{R}$ converge to the set $\operatorname{epi} f$.*

However, there is more. The same approach can be applied also to set-valued mappings. Recall that a set-valued mapping

$$S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n \text{ with graph set } \operatorname{gph} S = \{ (p, x) \mid x \in S(p) \}$$

is said to be *outer semicontinuous* if $\operatorname{gph} S$ is a closed subset of $\mathbb{R}^d \times \mathbb{R}^n$. The effective domain of S is $\operatorname{dom} S = \{ p \mid S(p) \neq \emptyset$, which is the projection of $\operatorname{gph} S$ on $\mathbb{R}^d$, while the effective range of S is $\operatorname{rge} S = \{ x \mid x \in S(p) \text{ for some } p \}$. The inverse S^{-1} has $\operatorname{gph} S^{-1} = \{ (x, p) \mid (p, x) \in \operatorname{gph} S \}$, so that $\operatorname{dom} S^{-1} = \operatorname{rge} S$ and $\operatorname{rge} S^{-1} = \operatorname{dom} S$.

Definition (graphical convergence)**.** *A sequence of outer semicontinous set-valued mappings $S^\nu : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ converges graphically to an outer semicontinous mapping $S : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ when the sets $\operatorname{gph} S^\nu$ in $\mathbb{R}^d \times \mathbb{R}^n$ converge to the set $\operatorname{gph} S$.*

Many rules and criteria are available for working with set convergence (see [7, Chapter 4]), epi-convergence (see [7, Chapter 7]), and graphical convergence (see [7, Chapter 5]). Rather going into more details here, our aim is to explain some interesting ways that the ideas can be utilized.

An important class of set-valued mappings for which approximation in the sense of graphical convergence will be helpful is subgradient mappings. For a function

$$f : \mathbb{R}^n \to (-\infty, \infty], \quad f \not\equiv \infty, \tag{6}$$

a *subgradient* at x (in the "regular" sense (cf. [7, Chapter 8]) is a vector v such that

$$f(y) \geq f(x) + v{\cdot}[y-x] + o(|y-x|), \tag{7}$$

where $o(|y-x|)$ is the traditional notation for an expression whose ratio to $|y-x|$ tends to 0 as y tends to x. When f is convex, the error term is superfluous and can be omitted. Thus, this definition equally covers subgradients as introduced originally in convex analysis.

If f is actually differentiable at x, the unique subgradient v is the gradient vector $\nabla f(x)$. In general, though, we have for each x a (possibly empty) *set* $\partial f(x)$ of subgradients and therefore a *set-valued subgradient mapping*

$$\partial f : x \mapsto \partial f(x) = \text{ set of all subgradients } v \text{ of } f \text{ at } x, \text{ if any.} \tag{8}$$

Whenever we have a sequence of functions f^ν on $I\!R^n$ we also have an associated sequence of subgradient mappings $\partial f^\nu : I\!R^n \rightrightarrows I\!R^n$. Is there some connection between epi-convergence of the functions f^ν and graphical convergence of the mappings ∂f^ν? The answer is yes, and it is answered most dramatically in the case of convex functions f^ν, as follows.

Attouch's Theorem for convex functions. *A sequence of lower semicontinuous convex functions f^ν epi-converges to a function f as in (6) if and only if the sequence of subgradient mappings ∂f^ν converges graphically to ∂f and, in addition,*

$$\exists\, (x^\nu, v^\nu) \to (x,v) \;\text{ with }\; v^\nu \in \partial f^\nu(x^\nu),\; v \in \partial f(x),\; f^\nu(x^\nu) \to f(x). \tag{9}$$

The extra condition (9) has to come in because ∂f determines f only up to some "constant of integration." (Passing from the epi-convergence to the graphical convergence automatically yields (9); the need is only in the other direction.)

A special situation in which (9) can be counted on to be satisfied is when

$$0 = f^\nu(0) = \inf f^\nu, \;\text{ implying also that }\; 0 = f(0) = \inf f.$$

Through convexity, that entails $0 \in \partial f^\nu(0)$ and $0 \in \partial f(0)$, so that (9) holds for $(x^\nu, v^\nu) = (0,0) = (x,v)$. We will invoke this case shortly.

Attouch's theorem [1] is remarkable for its presentation of a beautiful and valuable result which sharply departs from traditional modes of thinking in mathematics. An extension to a class of non-convex functions was obtained by Poliquin [4]. Connections with generalized second-derivatives of extended-real-valued functions will be seen below.

Another surprising fact which underscores the importance of epi-convergence in dealing with sequences of convex functions addresses duality. Recall that for a lower semicontinuous convex function f as in (6) the conjugate function f^*, defined by

$$f^*(v) = \sup_x \{\, v{\cdot}x - f(x)\}, \tag{10}$$

belongs to the same category and yields f back as its own conjugate, $f^{**} = f$, i.e.,

$$f(x) = \sup_v \{\, v{\cdot}x - f^*(v)\}. \tag{11}$$

Moreover $x \in \partial f^*(v)$ if and only if $v \in \partial f(x)$; in other words the subgradient mappings for these functions are inverse to each other:

$$\partial f^* = (\partial f)^{-1}, \qquad \partial f = (\partial f^*)^{-1}. \tag{12}$$

Wijsman's theorem for convex functions. *A sequence of lower semicontinuous convex functions f^ν epi-converges to a function f as in (6) if and only if the sequence of conjugate convex functions $f^{\nu *}$ epi-converges to the conjugate function f^*.*

In other words, the operation of passing from f to f^*, which is known as the Legendre-Fenchel transform in convex analysis, is continuous with respect to the topology of epi-convergence. For more about about Wijsman's theorem [8] and what it covers, see [7, Chapter 11G].

3 Generalized differentiation via set convergence

Consider now a "problem" of some sort, parameterized by a vector $p \in I\!R^d$, which looks for "solutions" $x \in I\!R^n$. This could take the form of the optimization in (3)–(4) with $S(p) = \operatorname{argmin} f(p, \cdot)$, but it might be something else; its formulation does not matter for now. All that matters is that we have a set-valued *solution mapping* $S : p \mapsto S(p)$.

Our focus is on a particular pair $(\bar{p}, \bar{x})$ with $\bar{x} \in S(\bar{p})$. We would like to "quantify" the way that shifts of $\bar{p}$ to other parameter vectors p induce shifts from $\bar{x}$ to other solutions $x \in S(p)$. Specifically, is there a way to think of "differentiating" S in order to get a handle on this?

It is important to realize that, unless S reduces to being single-valued at $\bar{p}$, the particular $\bar{x}$ under consideration in the set $S(\bar{p})$ must play a role. Generalized differentiation must operate in a manner that depends only on the local geometry of $\operatorname{gph} S$ around the pair $(\bar{p}, \bar{x}) \in \operatorname{gph} S$. For this purpose we look at the *second-order difference quotients mappings*

$$\Delta_\varepsilon S(\bar{p}\,|\,\bar{x}) : p' \mapsto \frac{1}{\varepsilon}\Big[S(\bar{p} + \varepsilon p') - \bar{x}\Big] \;\text{ for }\; \varepsilon > 0. \tag{13}$$

Definition (proto-differentiability of set-valued mappings)**.** *The mapping $S : I\!R^d \rightrightarrows I\!R^n$ is proto-differentiable at $\bar{p}$ for $\bar{x}$ if the second-order difference quotient mappings $\Delta_\varepsilon S(\bar{p}\,|\,\bar{x})$ converge graphically as $\varepsilon \to 0$. The limit mapping, denoted by $DS(\bar{p}\,|\,\bar{x})$, is then called the proto-derivative of S at $\bar{p}$ for the element $\bar{x} \in S(\bar{p})$.*

In geometric terms, the graph of the mapping $\Delta_\varepsilon S(\bar{p}\,|\,\bar{x})$ is obtained from the graph of S as

$$\operatorname{gph} \Delta_\varepsilon S(\bar{p}\,|\,\bar{x}) = \frac{1}{\varepsilon}\Big[\operatorname{gph} S - (\bar{p}, \bar{x})\Big] \quad\text{for}\;\; (\bar{p}, \bar{x}) \in \operatorname{gph} S, \tag{14}$$

so that the graph of $DS(\bar{p}\,|\,\bar{x})$ in the limit will be the tangent cone to the graph of S at $(\bar{p}, \bar{x})$. (The tangent cone is defined in general as merely an "outer limit," but here we are insisting on a limit in the full sense of set-convergence, which requires the "outer limit" to coincide with an associated "inner limit"; see [7, Chapter 6].)

Criteria for when proto-derivatives are available emerge, as in elementary calculus, from basic examples combined with rules for dealing with sums, compositions, and the like.

A valuable observation is that proto-differentiability of S at $\bar{p}$ for $\bar{x} \in S(\bar{p})$ corresponds to proto-differentiability of the inverse mapping S^{-1} at $\bar{x}$ for $\bar{p} \in S^{-1}(\bar{x})$ with

$$D[S^{-1}](\bar{x}\,|\,\bar{p}) = DS(\bar{p}\,|\,\bar{x})^{-1}. \tag{15}$$

This holds through the graphical geometry. Reversing the pairs in the graph of a mapping to get its inverse has no effect on the existence of the graphical limits that underlie proto-differentiability.

The proto-derivative mapping $DS(\bar{p}\,|\,\bar{x})$ does provide key information about perturbations in the case of a S being a solution mapping, and ultimately that can be worked out for the optimization in

(3)–(4). To make the connection, though, some kind of generalized differentiation associated with the function f in that model must of course be involved, and that has to be a *second-order* kind.

The immediate agenda is to approach second-order differentiation of a function f, not necessarily convex, in two ways. The first is through graphical differentiation of the subgradient mapping ∂f and the second is through epi-convergence of second-order difference quotients of f itself. In both cases the concepts are tied to a choice of $\bar{x}$ and a subgradient $\bar{v}$ of f at x.

Proto-differentiability in the sense of the definition above can be investigated for the set-valued mapping ∂f, for which the first-order difference quotient mappings are

$$\Delta_\varepsilon[\partial f](\bar{x}\,|\,\bar{v}) : x' \mapsto \frac{1}{\varepsilon}\Big[\partial f(\bar{x}+\varepsilon x') - \bar{v}\Big] \quad \text{with } \ \bar{v} \in \partial f(\bar{x}). \tag{16}$$

When these mappings converge graphically — to something — that limit mapping is the proto-derivative of ∂f at $\bar{x}$ for $\bar{v}$,

$$D[\partial f](\bar{x}\,|\,\bar{v}) : I\!R^n \rightrightarrows I\!R^n. \tag{17}$$

The combination of "subdifferentiating" f to get ∂f and then "proto-differentiating" ∂f to get $D[\partial f](\bar{x}\,|\,\bar{v})$ constitutes some version of second-order differentiation of f, but how does it relate to ordinary calculus? An example to look at is that of a $\mathcal{C}^2$ function f at a point $\bar{x}$ with $\bar{v} = \nabla f(\bar{x})$. It's easy to see that the proto-derivative mapping then exists and is simply the linear mapping $x' \to \nabla^2 f(\bar{x})x'$ given by the Hessian matrix $\nabla^2 f(\bar{x})$ (formed from the second partial derivatives of f at $\bar{x}$). From this angle, we can think of (17) as a "generalized Hessian mapping."

Turning now to the alternative approach to second-order differentiation, we look at *second-order difference quotients* of f having the form

$$\Delta_\varepsilon^2 f(\bar{x}\,|\,\bar{v}) : x' \mapsto \frac{1}{\varepsilon^2}\Big[f(\bar{x}+\varepsilon x') - f(\bar{x}) - \varepsilon x' {\cdot} \bar{v}\Big] \quad \text{with } \ \bar{v} \in \partial f(\bar{x}). \tag{18}$$

Definition (second-order epi-derivatives). *If the second-order difference quotient functions (18) epi-converge as $\varepsilon \to 0$, the limit function, denoted by $d^2 f(\bar{x}\,|\,\bar{v})$, is called the second-order epi-derivative of f at $\bar{x}$ for $\bar{v}$.*

Again it is good to refer to the example of a $\mathcal{C}^2$ function f for insights. There, with $\bar{v} = \nabla f(\bar{x})$, we get $d^2 f(\bar{x}\,|\,\bar{v})(x') = x'{\cdot}\nabla^2 f(\bar{x})x'$, the quadratic function associated with the Hessian matrix. We observe that its derivative mapping, $x' \mapsto 2\nabla^2 f(\bar{x})x'$, is, apart from the factor 2, the same as the linear mapping that turned out to be $D[\partial f](\bar{x}\,|\,\bar{v})$.

This powerfully suggests the possibility of a tight relationship holding quite generally between the two approaches to generalized second-order differation. Supporting evidence comes from the observation that

$$\partial[\Delta_\varepsilon^2 f(\bar{x}\,|\,\bar{v})](x') = 2\Delta_\varepsilon[[\partial f](\bar{x}\,|\,\bar{v})](x') \quad \text{for all } \ x', \tag{19}$$

as follows from the elementary rules for calculating subgradients. Does this relationship persist in the limit as $\varepsilon \to 0$? The issue can be complicated in general, but for convex functions Attouch's theorem comes immediately to the rescue in making use of the fact that then the functions $\Delta_\varepsilon^2 f(\bar{x}\,|\,\bar{v})$ are convex with $0 = \Delta_\varepsilon^2 f(\bar{x}\,|\,\bar{v})(0) = \inf \Delta_\varepsilon^2 f(\bar{x}\,|\,\bar{v})$.

Theorem (second-order equivalence for convex functions). *For a lower semicontinuous convex function $f : I\!R^n \to (-\infty,\infty]$ and a pair $(\bar{x},\bar{v})$ with $\bar{v} \in \partial f(\bar{x})$, the following are equivalent:*

(a) The second-order epi-derivative $d^2 f(\bar{x}\,|\,\bar{v})$ exists.

(b) The proto-derivative $D[\partial f](\bar{x}\,|\,\bar{v})$ exists.

Moreover these derivative objects are related then by

$$\partial[d^2 f(\bar{x}\,|\,\bar{v})](x') = 2D[[\partial f](\bar{x}\,|\,\bar{v})](x') \quad \textit{for all } x', \tag{20}$$

A corresponding result is known for nonconvex f having the property, locally with respect to $(\bar{x}, \bar{v})$, of being *prox-regular*, as developed in [5]. That paper also brings to light a large class of functions, nonconvex as well as convex, for which the second-order epi-derivatives and subgradient proto-derivatives do surely exist. See also [7, Chapter 13].

Alongside of Attouch's theorem we can appeal to Wijsman's theorem in this context through the fact that, in the case of f convex, the second-order difference quotient function $\Delta^2_\varepsilon f(\bar{x}\,|\,\bar{v})$ is convex as well, and the conjugate of $\frac{1}{2}\Delta^2_\varepsilon f(\bar{x}\,|\,\bar{v})$ calculates out to be $\frac{1}{2}\Delta^2_\varepsilon f^*(\bar{v}\,|\,\bar{x})$. This leads to an important conclusion in the limit as $\varepsilon \to 0$.

Theorem (second-order duality for convex functions)**.** *For a lower semicontinuous convex function $f : \mathbb{R}^n \to (-\infty, \infty]$ with conjugate f^* and a pair $(\bar{x}, \bar{v})$ with $\bar{v} \in \partial f(\bar{x})$, hence also $\bar{x} \in \partial f^*(\bar{v})$, the following are equivalent:*

(a) The second-order epi-derivative $d^2 f(\bar{x}\,|\,\bar{v})$ exists.

(b) The second-order epi-derivative $d^2 f^*(\bar{v}\,|\,\bar{x})$ exists.

The convex functions $\frac{1}{2}d^2 f(\bar{x}\,|\,\bar{v})$ and $\frac{1}{2}d^2 f^*(\bar{v}\,|\,\bar{x})$ are then conjugate to each other. Moreover the proto-derivative mappings $D[\partial f](\bar{x}\,|\,\bar{v})$ and $D[\partial f^*](\bar{v}\,|\,\bar{x})$ both exist and are inverse to each other.

The final fact is supported by formula (20) and its counterpart for f^*.

4 Application to shifts of optimal solutions

Let us look now at a special but fundamental set-up in parameterized optimization involving a lower semicontinuous function

$$g : \mathbb{R}^n \to (-\infty, \infty], \quad g \not\equiv \infty, \tag{21}$$

and a "tilting" vector $v \in \mathbb{R}^n$:

$$\text{minimize } \; g(x) - x{\cdot}v \; \text{ in } x \text{ with } v \text{ as parameter.} \tag{22}$$

This model fits into the general parameterization scheme in (4) with $p = v$ and $f(p, x) = g(x) - x{\cdot}v$.

For x to be a locally optimal solution in (22) it is necessary to have $v \in \partial g(x)$. On the other hand, this is sufficient for x to be a globally optimal solution when g is convex. More broadly, in the possible absence of convexity, points x having $v \in \partial g(x)$ can be termed *stationary points* in the optimization. From that perspective, we can think of the mapping

$$S : v \mapsto \{\, x \,|\, v \in \partial g(x)\} \tag{23}$$

as the solution mapping for the *stationarity* problem in (21), keeping in mind that when g is convex this is the same the solution mapping with respect to global optimality.

We choose a reference "tilt" $\bar{v}$ (for instance $\bar{v} = 0$), a corresponding "solution" $\bar{x} \in S(\bar{v})$, and contemplate gaining information about solution perturbations through the proto-derivative mapping $DS(\bar{v}\,|\,\bar{x}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ if it exists.

The tools that have been laid out can be put to work right away in this situation. Obviously S is the inverse of the subgradient mapping ∂g:

$$x \in S(v) \quad \Longleftrightarrow \quad v \in \partial g(x). \tag{24}$$

Consequently, through the preservation of proto-differentiability in taking inverse, we have that S is proto-differentiable at $\bar{v}$ for $\bar{x}$ if and only if ∂g is proto-differentiable at $\bar{x}$ for $\bar{v}$, and then

$$DS(\bar{v} \,|\, \bar{x}) = [D[\partial g](\bar{x} \,|\, \bar{v})]^{-1}. \tag{25}$$

In the convex case, which benefits from the theorem presenting the relation in (20), we arrive at very strong conclusions.

Theorem (tilt perturbations of optimality with convexity). *For a lower semicontinuous convex function g in (21) and the associated solution mapping S for global optimality in (22), given by (23), and for a tilt vector $\bar{v}$ and a solution $\bar{x} \in S(\bar{v})$, the following are equivalent:*

(a) The proto-derivative mapping $DS(\bar{v} \,|\, \bar{x})$ exists.

(b) The second-order epi-derivative $d^2 g(\bar{x} \,|\, \bar{v})$ exists.

In that case there is the further equivalence between

(a′) x' is an element of the set $DS(\bar{v} \,|\, \bar{x})(v')$.

(b′) v' is a subgradient of the function $\frac{1}{2} d^2 g(\bar{x} \,|\, \bar{v})$ at x'.

In consequence, the vectors x' in (a′) are the optimal solutions to the convex optimization subproblem:

$$\text{minimize } \tfrac{1}{2} d^2 g(\bar{x} \,|\, \bar{v})(x') - x' \cdot v' \text{ with respect to } x',$$

so that

$$DS(\bar{v} \,|\, \bar{x})(v') = \operatorname{argmin}_{x'} \left\{ \tfrac{1}{2} d^2 g(\bar{x} \,|\, \bar{v})(x') - x' \cdot v' \right\}. \tag{26}$$

The remarkable description of solution pertubation vectors x' as solutions themselves to an optimization problem parameterized by perturbational tilt vectors v' comes from the observation that the solutions in (26) are characterized through convexity by $v' \in \partial[\frac{1}{2} d^2 g(\bar{x} \,|\, \bar{v})](x')$. On the other hand, this subgradient mapping can be identified with the right side of (25) through (20) as applied to g.

In line with remarks at the end of the preceding section, an extension of this theorem beyond convexity, although not with the minimization fact in (26), can be made for functions having the property of prox-regularity.

Tilt perturbations can be combined with other possibilities for perturbation by adopting the problem model:

$$\text{minimize } g(u,x) - x \cdot v \text{ in } x \text{ with } u \text{ and } v \text{ as parameters.} \tag{27}$$

This corresponds to the earlier model (4) as the case where $p = (u,v)$ and $f(p,x) = g(u,x) - x \cdot v$. It may seem unnecessary to have both u and v; why not just fall back on $g(u,x) = f(p,x)$? The reason is that in developing the theory at this level one soon sees that tilt parameters must enter indirectly, if not directly, and that it is more convenient therefore to employ them from the beginning. Results developed in the (u,v) format can anyway be reduced to results in the u format by fixing $v = 0$ at the end.

Research in this general direction, with the goal of understanding when the solution mapping S in (27) would have a localization that is single-valued and Lipschitz continuous, began in [2]. Some of the latest developments can be found in [3]. A notable feature of the subject is the important way that *coderivatives* of set-valued mappings and in particular subgradient mappings, as opposed to the graphical derivatives employed here, come into play. Coderivatives correspond, with a twist, to normal cones to the graphs of mappings in place of the tangent cones that support graphical derivatives.

References

[1] H. Attouch, "Convergence de fonctions convexes, de sous-differentiels et semi-groupes," *Comptes Rendus de l'Academie des Sciences de Paris* 284 [1977], 539–542.

[2] A.B. Levy, R.A. Poliquin, R.T. Rockafellar, "Stability of locally optimal solutions," *SIAM Journal of Optimization* 10 [2000], 580–604.

[3] B.S. Mordukhovich, T.T.A. Nghia, R.T. Rockafellar, "Full stability in finite-dimensional optimization," Mathematics of Operations Research, forthcoming.

[4] R.A. Poliquin, "An extension of Attouch's theorem and its application to second-order epi-differentation of convexly composite functions," *Transactions of the American Mathematical Society* 332 [1992], 861–874.

[5] R.A. Poliquin, R.T. Rockafellar, "Prox-regular functions in variational analysis," *Transactions of the American Mathematical Society* 348 [1996], 1805–1838.

[6] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[7] R.T. Rockafellar, R. J-B Wets, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1997.

[8] R.A. Wijsman, "Convergence of sequences of convex sets, cones and functions," *Bulletin of the American Mathematical Society* 70 [1964], 186-188.