NONSMOOTH ANALYSIS AND PARAMETRIC OPTIMIZATION

R. T. Rockafellar*

Abstract. In an optimization problem that depends on parameters, an important issue is the effect that perturbations of the parameters can have on solutions to the problem and their associated multipliers. Under quite broad conditions the possibly multi-valued mapping that gives these elements in terms of the parameters turns out to enjoy a property of "proto-differentiability." Generalized derivatives can then be calculated by solving an auxiliary optimization problem with auxiliary parameters. This is constructed from the original problem by taking second-order epi-derivatives of an essential objective function.

1. Solutions to Optimization Problems with Parameters.

From an abstract point of view, a general optimization problem relative to elements x of a Banach space $\mathcal{X}$ can be seen in terms of minimizing an expression $f(x)$ over all $x \in \mathcal{X}$, where f is a function on $\mathcal{X}$ with values in $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. The effective domain $\operatorname{dom} f := \{x \in \mathcal{X} \mid f(x) < \infty\}$ gives the "feasible" or "admissible" elements x. Under the assumption that f is lower semicontinuous and proper (the latter meaning that $f(x) < \infty$ for at least one x, but $f(x) > -\infty$ for all x), a solution $\bar{x}$ to the problem must satisfy $0 \in \partial f(\bar{x})$, where ∂f denotes subgradients in the sense of Clarke [1] (see also Rockafellar [2]). When f is convex, such subgradients coincide with the ones of convex analysis, and the condition $0 \in \partial f(\bar{x})$ is not only necessary for optimality but sufficient.

A substantial calculus, part of a broader subject called nonsmooth analysis, has been built up for determining the set $\partial f(\bar{x})$ in the case of particular structure of f. Dual elements such as Lagrange multipliers are often involved, and under convexity assumptions these typically solve a dual problem of optimization.

It has long been known that in order to derive and interpret the dual elements appearing in optimality conditions, it is important to study optimization problems not in isolation but in parametrized form. Only recently, however, have the tools of analysis reached the stage where it is possible to analyze in a general and effective manner the dependence of

* This work was supported in part by the National Science Foundation at the University of Washington, Seattle

solutions and multipliers on parameters and to obtain results on generalized differentiability with respect to parameters. Such results are the topic here. They are closely tied in with a newly developed theory of second derivatives of nonsmooth functions.

If a single problem of optimization consists of minimizing an expression $f(x)$ over all $x \in \mathcal{X}$, then a parameterized problem, depending on an element u in a Banach space $\mathcal{U}$, consists of minimizing $f(u,x)$ over all $x \in \mathcal{X}$, where f is a function on $\mathcal{U} \times \mathcal{X}$. All aspects of dependence, including variable constraints, can be represented in principle in this very simple form. For technical reasons that will later emerge, it is advantageous however to proceed with slightly more structure: parameterization not only by an abstract element $u \in \mathcal{U}$ but also by a "tilting" term $v \in \mathcal{X}^*$, where $\mathcal{X}^*$ is the Banach space dual to $\mathcal{X}$. Thus we adopt as our basic model the parameterized problem

$$\big(\mathcal{P}(u,v)\big) \qquad \text{minimize } f(u,x) + \langle x, v\rangle \text{ over all } x \in \mathcal{X},$$

where the function $f : \mathcal{U} \times \mathcal{X} \to \overline{I\!R}$ is lsc and proper. To each choice of $(u,v) \in \mathcal{U} \times \mathcal{X}^*$ we assign a set $M(u,v) \subset \mathcal{X} \times \mathcal{U}^*$ consisting of the primal-dual pairs (x,y) associated with an optimality condition for $\big(\mathcal{P}(u,v)\big)$. We focus on the possible "differentiability" properties of the mapping $M : \mathcal{U} \times \mathcal{X}^* \to \mathcal{X} \times \mathcal{U}^*$ and their relationship to such properties of f.

The first task is to derive the specific form of M from the calculus of first-order optimality conditions in terms of subgradients. This will provide a bridge between M and the subgradient mapping $\partial f : \mathcal{U} \times \mathcal{X} \to \mathcal{U}^* \times \mathcal{X}^*$. Next we must consider the kind of generalized differentiation to apply to M and ∂f. This will be "proto-differentiation," as defined in [3] by graphical convergence of certain difference quotient mappings. The concept requires a brief review of notions of set convergence, which are then utilized also in defining first and second-order "epi-derivatives" of f by way of epigraphical convergence of first and second-order difference quotient functions [4].

A fundamental connection between second-order epi-derivatives of f and proto-derivatives of ∂f is the key to obtaining proto-derivatives of M. It turns out that in many important cases these derivatives can be determined by solving a "derivative problem," which is a parameterized optimization problem in the same mold as $\big(\mathcal{P}(u,v)\big)$. Differentiation formulas in [4] make it possible to work out the details quite explicitly for a wide range of situations.

2. Set Convergence and Subgradients.

For a sequence $\{C^\nu\}_{\nu\in\mathbb{N}}$ of nonempty subsets of the Banach space $\mathcal{W}$, the *inner limit* set $\liminf_\nu C^\nu$ consists of all limit points of sequences $\{w^\nu\}_{\nu\in\mathbb{N}}$ selected with $w^\nu \in C^\nu$, while the *outer limit* set $\limsup_\nu C^\nu$ consists of all cluster points of such sequences. One says that $\{C^\nu\}_{\nu\in\mathbb{N}}$ *converges* if $\liminf_\nu C^\nu = \limsup_\nu C^\nu$, the latter set then being called the *limit* and denoted by $\lim_\nu C^\nu$.

This is set convergence in the Painlevé-Kuratowski sense. Set convergence in the Mosco sense [5] means that the limit exists also with respect to the weak topology on $\mathcal{W}$, and the weak limit coincides with the strong limit. On the other hand, set convergence in the Attouch-Wets sense [6] refers to the uniform convergence of the distance functions $w \mapsto d(w, C^\nu)$ on all bounded subsets of $\mathcal{W}$. When $\mathcal{W}$ is finite-dimensional, the three notions are equivalent; in general, Attouch-Wets convergence implies Painlevé-Kuratowski convergence. Mosco convergence is really appropriate only for convex sets, where it lies between the other two. Although it has been the most studied concept of convergence in the theory of variational problems, its good properties are limited inherently to reflexive Banach spaces, as shown recently by Beer and Borwein [7].

A sequence $\{f^\nu\}_{\nu\in\mathbb{N}}$ of functions from $\mathcal{W}$ to $\overline{\mathbb{R}}$ is said to *epi-converge* if the epigraphs $\operatorname{epi} f^\nu = \{(w,\alpha)\in\mathcal{W}\times\overline{\mathbb{R}} \,\big|\, \alpha \geq f^\nu(w)\}$ converge as sets. The epi-limit function f, when it exists, is characterized by the fact that for every $w \in \mathcal{W}$ both of the following properties hold:

$$\begin{aligned} &\forall w^\nu \to w, \quad \liminf_\nu f^\nu(w^\nu) \geq f(w),\\ &\exists w^\nu \to w, \quad \limsup_\nu f^\nu(w^\nu) \leq f(w). \end{aligned} \tag{2.1}$$

This type of function convergence, which in general neither implies nor is implied by pointwise convergence, it was first considered by Wijsman [8], who proved that for convex functions in a finite-dimensional setting it made the Legendre-Fenchel transform ($f \mapsto f^*$) be continuous. Continuity of the Legendre-Fenchel transform holds in reflexive Banach spaces under Mosco epi-convergence, where the epigraphs converge not just in the Painlevé sense but the Mosco sense [9]. Without reflexivity it fails for Mosco epi-convergence [7], but it holds in all normed linear spaces with respect to epi-convergence in the Attouch-Wets sense, cf. [6].

In what follows, set convergence and epi-convergence are to be interpreted in the underlying Painlevé-Kuratowski sense, unless otherwise specified.

Let C be a closed subset of $\mathcal{W}$ and let $\bar{w} \in C$. The *basic tangent cone* (or *contingent cone*) to C at $\bar{w}$ is the closed set

$$T_C(\bar{w}) := \limsup_{\tau \searrow 0} \tau^{-1}(C - \bar{w}). \tag{2.2}$$

We say that C is *derivable* at $\bar{w}$ if the sets $\tau^{-1}(C-\bar{w})$ actually converge as $\tau \searrow 0$, so that $T_C(\bar{w})$ is obtained not just as the outer limit but the limit. This certainly holds under the stronger condition that

$$\limsup_{\tau \searrow 0} \tau^{-1}(C-\bar{w}) = \liminf_{\substack{w(\in C)\to\bar{w} \\ \tau \searrow 0}} \tau^{-1}(C-\bar{w}), \tag{2.3}$$

which is the *Clarke regularity* of C at $\bar{w}$. Clarke regularity is an important property commonly satisfied in connection with the derivation of first-order optimality conditions but not, for instance, in the study of graphs of nonsmooth mappings. It is for the latter purpose that we shall find derivability to be of independent interest. The property of derivability of a set was first utilized in parametric optimizaton by Shapiro [10].

Clarke regularity holds at every point $\bar{w} \in C$ when C is convex, or when C is defined by smooth constraints for which a natural constraint qualification is satisfied [1, pp. 51–57]. When $\mathcal{W}$ is finite-dimensional, it is known to be equivalent to the condition that

$$T_C(\bar{w}) = \liminf_{w(\in C)\to\bar{w}} T_C(w). \tag{2.4}$$

The *Clarke normal cone* to C at $\bar{w}$ is the closed convex cone $N_C(\bar{w})$ in $\mathcal{W}^*$ defined in general as the polar of the limit set on the right side of (2.3), which is the *Clarke tangent cone* to C at $\bar{w}$ and is known always to be convex [1, p. 51], [2, p. 263]. Here we shall mainly use this cone in the presence of Clarke regularity. Then $N_C(\bar{w})$ can be viewed as the polar of the basic tangent cone $T_C(\bar{w})$, and vice versa:

$$\begin{aligned} N_C(\bar{w}) &= \{z \in \mathcal{W}^* \mid \langle w', v\rangle \le 0 \text{ for all } w' \in T_C(\bar{w})\}, \\ T_C(w) &= \{w' \in \mathcal{W} \mid \langle w', v\rangle \le 0 \text{ for all } z \in N_C(\bar{w})\}. \end{aligned} \tag{2.5}$$

Consider now a proper, lsc function $f : \mathcal{W} \to \overline{I\!R}$ and a point $\bar{w} \in \operatorname{dom} f$. For each $\tau > 0$ define the difference quotient function $\Delta_\tau f(\bar{w}) : \mathcal{W} \to \overline{I\!R}$ by

$$\Delta_\tau f(\bar{w})(w') = [f(\bar{w}+\tau w') - f(\bar{w})]/\tau. \tag{2.6}$$

Define the *(lower) subderivative* function $Df(\bar{x}) : \mathcal{X} \to \overline{I\!R}$ by

$$Df(\bar{w})(\bar{w}') = \liminf_{\substack{w'\to\bar{w}' \\ \tau \searrow 0}} \Delta_\tau f(\bar{w})(w'). \tag{2.7}$$

If the functions $\Delta_\tau f(\bar{w})$ epi-converge as $\tau \searrow 0$, the limit function necessarily then being $Df(\bar{w})$, we say that f is *epi-differentiable* at $\bar{w}$—*properly* if in addition $Df(\bar{w})(0) > -\infty$

(which is equivalent to $Df(\bar{w})$ being proper). The epi-differentiability of f at $\bar{w}$ is equivalent to the derivability of epi f at $\big(\bar{w}, f(\bar{w})\big)$, because the epigraph of $\Delta_\tau f(\bar{w})$ is the set $(\operatorname{epi} f) - \big(\bar{w}, f(\bar{w})\big)$, while the epigraph of $Df(\bar{w})$ is the cone $T_{\operatorname{epi} f}\big(\bar{w}, f(\bar{w})\big)$.

Epi-differentiability of f at $\bar{w}$ is implied by *Clarke regularity* of f at $\bar{w}$, defined as the Clarke regularity of epi f at $\big(\bar{w}, f(\bar{w})\big)$. This property is present at every $\bar{w} \in \operatorname{dom} f$ when f is convex. A substantial calculus is available for establishing Clarke regularity in the case of a function obtained through various constructions, cf. Clarke [1], Rockafellar [13], Borwein and Ward [14].

The set of generalized *subgradients* of f at $\bar{w}$ is defined in terms of the Clarke normal cone to epi f at $\big(\bar{w}, f(\bar{w})\big)$ by

$$\partial f(\bar{w}) = \{z \in \mathcal{W}^* \,\big|\, (z, -1) \in N_{\operatorname{epi} f}\big(\bar{w}, f(\bar{w})\big)\}, \tag{2.8}$$

and the set of *singular (or recessional) subgradients* by

$$\partial^\infty f(\bar{w}) = \{z \in \mathcal{W}^* \,\big|\, (z, 0) \in N_{\operatorname{epi} f}\big(\bar{w}, f(\bar{w})\big)\}, \tag{2.9}$$

the latter being the same as the recession cone of $\partial f(\bar{w})$ when $\partial f(\bar{w}) \neq \emptyset$. If f is Clarke regular at $\bar{w}$, so that the cones $N_{\operatorname{epi} f}\big(\bar{w}, f(\bar{w})\big)$ and $T_{\operatorname{epi} f}\big(\bar{w}, f(\bar{w})\big)$ are polar to each other, one has from (2.8) that

$$\begin{aligned} \partial f(\bar{w}) &= \{z \in \mathcal{W}^* \,\big|\, \langle w', z\rangle \leq Df(\bar{w})(w') \text{ for all } w' \in \mathcal{W}\}, \\ \partial^\infty f(\bar{w}) &= \{z \in \mathcal{W}^* \,\big|\, \langle w', z\rangle \leq 0 \text{ for all } w' \in \operatorname{dom} Df(\bar{w})\}, \end{aligned} \tag{2.10}$$

and moreover

$$Df(\bar{w})(w') = \sup_{z \in \partial f(\bar{w})} \langle w', z\rangle \text{ for all } w' \in \mathcal{W} \text{ when } \partial f(\bar{w}) \neq \emptyset, \tag{2.11}$$

the latter case then being precisely the one where f is epi-differentiable at $\bar{w}$. Many formulas for calculating subgradients are provided in [1], [13], [14], [15].

3. First-Order Optimality Conditions.

Returning to our general model of parametric optimization for a function $f : \mathcal{U} \times \mathcal{X} \to \overline{I\!R}$ that is lsc and proper, we now seek first-order conditions on a solution $\bar{x}$ to the problem $\big(\mathcal{P}(\bar{u}, \bar{v})\big)$ associated with a particular parameter choice $\bar{u}$. Our ultimate goal is the analysis of the effects on such conditions of perturbations in $\bar{u}$. In the circumstances under which we are able, for now, to obtain good results, f will be Clarke regular. We then are in a context where subgradients of f can be handled in terms of formulas (2.10) and (2.11), so the various alternative approaches to first-order conditions that might otherwise be considered turn out to coincide with the Clarke approach, and in the convex case, with the one of convex analysis. It will therefore suffice, for the background we require, to record the main, general facts about first-order conditions that are obtainable by with Clarke subgradients. The detailed development of these conditions in specific cases, such as infinite-dimensional problems involving integral functionals, lies obviously beyond our scope but can be found for example in [1].

A point $\bar{x} \in \mathcal{X}$ is said to be *locally optimal* for a problem $\big(\mathcal{P}(\bar{u}, \bar{v})\big)$ if $\bar{x}$ furnishes a local minimum to the objective function $\varphi(x) = f(\bar{u}, x) + \langle x, \bar{v} \rangle$ in this problem and $\varphi(\bar{x}) < \infty$ (i.e., the objective in $\big(\mathcal{P}(\bar{u}, \bar{v})\big)$ is not identically ∞, which would indicate the absence of any feasible or admissible points). It is *globally optimal* if the local minimum is a global minimum, as must always hold for instance in the convex case.

The *basic constraint qualification* is said to hold for $\big(\mathcal{P}(\bar{u}, \bar{v})\big)$ at $\bar{x}$ if

$$\text{the only } \bar{y} \in \mathcal{U}^* \text{ with } (\bar{y}, 0) \in \partial^\infty f(\bar{u}, \bar{x}) \text{ is } \bar{y} = 0. \tag{3.1}$$

This abstract condition on singular subgradients of f reduces in numerous special cases to the assumptions on constraint structure that are needed in deriving "multiplier rules." Such is true for instance in nonlinear programming (cf. [4]) and in the theory of convex Bolza problems in optimal control and the calculus of variations [16]. It may well hold in other cases not yet fully studied.

It is valuable to consider along with the basic constraint qualification a related condition called *calmness*. This is said to hold for $\big(\mathcal{P}(\bar{u}, \bar{x})\big)$ at $\bar{x}$ if, for some neighborhood X of $\bar{x}$ relative to which the objective function in $\big(\mathcal{P}(\bar{u}, \bar{x})\big)$ is minimized at $\bar{x}$, the local optimal value function

$$p_X(u, v) := \inf_{x \in X} \{ f(u, x) + \langle x, \bar{v} \rangle \} \tag{3.2}$$

satisfies

$$\liminf_{u \to \bar{u}} \big[p_X(u, \bar{v}) - p_X(\bar{u}, \bar{v}) \big] / \|u - \bar{u}\| > -\infty. \tag{3.3}$$

Calmness is a minimal kind of stability property. Without it, the value $p_X(\bar u, \bar v)$ drops off at an infinite rate relative when $\bar u$ is perturbed in certain directions.

The term "calmness" was introduced by Clarke in obtaining a Lagrange multiplier rule for mathematical programming problems [17] and optimality conditions for variational problems [18]. The property itself has a longer history of being used for similar purposes in convex optimization, cf. Rockafellar [19], [20]. When f is convex, one can take $X = \mathcal{X}$. The value $p(u,v) := p_{\mathcal{X}}(u,v)$ is then convex in its dependence on u, so that calmness reduces to the existence of subgradients of $p(\cdot, \bar v)$ at $\bar u$. Guaranteeing the existence of such subgradients is exactly the role of the various conditions in convex optimization that concern the intersection of effective domains, most of which reduce in our formulation here to the existence of some $\widetilde{x} \in \mathcal{X}$ such that the function $u \mapsto f(u, \widetilde{x})$ is finite on some neighborhood of $\bar u$. For more on this topic we refer to [21]. In the case where $\mathcal{U}$ and $\mathcal{X}$ are finite-dimensional, the basic constraint qualification implies calmness regardless of whether f is convex, cf. [15].

Theorem 1. *If $\bar x$ is locally optimal for problem $\big(\mathcal{P}(\bar u, \bar v)\big)$ and the calmness condition holds, or the basic constraint qualification holds with both $\mathcal{U}$ and $\mathcal{X}$ finite-dimensional, then*

$$\exists \bar y \in \mathcal{U}^* \ \text{ with } \ (\bar y, -\bar v) \in \partial f(\bar u, \bar x). \tag{3.3}$$

Conversely, if the latter is satisfied and f is convex, then $\bar x$ is globally optimal for $\big(\mathcal{P}(u,v)\big)$ and calmness holds.

Proof. Under the calmness condition there is a value $\rho > 0$ such that $p_X(u, \bar v) \geq p_X(\bar u, \bar v) - \rho \|u - \bar u\|$ for all u near $\bar u$. This means that $f(u,x) + \langle x, \bar v\rangle \geq f(\bar u, \bar x) + \langle \bar x, \bar v\rangle - \rho\|u - \bar u\|$ for all (u,x) near $(\bar u, \bar x)$. Then for $g(u,x) := \langle x, \bar v\rangle + \rho\|u - \bar u\|$ the function $f + g$ has a local minimum on $\mathcal{U} \times \mathcal{X}$ at $(\bar u, \bar x)$, which implies $(0,0) \in \partial(f+g)(\bar u, \bar x)$. Inasmuch as g is Lipschitz continuous, we have $\partial(f+g)(\bar u, \bar x) \subset \partial f(\bar u, \bar x) + \partial g(\bar u, \bar x)$ [13, p. 345]. But $\partial g(\bar u, \bar x) = \{(y, \bar v) \,\big|\, \|y\| \leq \rho\}$. Therefore $\partial f(\bar u, \bar x)$ contains an element $-(y, \bar v)$ with $\|y\| \leq \rho$. Writing y as $-\bar y$, we get (3.3).

In the finite-dimensional case with the constraint qualification satisfied, the validity of (3.3) is asserted in [15, Thm. 5.1]. When f is convex, (3.3) implies

$$f(u,x) \geq f(\bar u, \bar x) + \langle (u,x) - (\bar u, \bar x), (\bar y, -\bar v)\rangle.$$

This inequality, specialized to $u = \bar u$, yields the global optimality of $\bar x$ in $\big(\mathcal{P}(\bar u, \bar v)\big)$. It also yields

$$\inf_x \{f(u,x) + \langle x, \bar v\rangle\} \geq \inf_x \{f(\bar u, x) + \langle x, \bar v\rangle\} - \|\bar y\| \, \|u - \bar u\|,$$

which is a strong version of the calmness condition. □

On the basis of Theorem 1, we associate with the parameterization $\big(\mathcal{P}(u,v)\big)$ the set-valued mapping $M : \mathcal{U} \times \mathcal{X}^* \mapsto \mathcal{X} \times \mathcal{U}^*$ defined by

$$M(u,v) = \{(x,y) \,|\, (y,-v) \in \partial f(u,x)\}. \tag{3.4}$$

When f is convex, the pairs $(x,y) \in M(u,v)$ give the primal and dual optimal elements associated with $\big(\mathcal{P}(u,v)\big)$, but in general, of course, they only give elements that are quasi-optimal in the sense of satisfying first-order conditions.

An example in finite dimensions, covering a multitude of applications, will serve in illustrating theorem 1 and other results to follow. For this we consider the parameterized problem

$$\big(\mathcal{P}_0(u,v,w)\big) \qquad \text{minimize } g(w,x) + \langle x,v\rangle + h\big(u + G(w,x)\big) \text{ over } x \in C \subset I\!R^n$$

for parameter elements $u \in I\!R^m, v \in I\!R^n$ and $w \in I\!R^d$. This corresponds to

$$f_0(u,w,x) = g(w,x) + h\big(u + G(w,x)\big) + \delta_C(x), \tag{3.5}$$

where δ_C is the indicator of C, and the pair (u,w) now stands for what previously was denoted by just u. Rather than press for the weakest viable assumptions, we take the function $g : I\!R^p \times I\!R^n \to I\!R$ and the mapping $G : I\!R^p \times I\!R^n \to I\!R^m$ to be of class $\mathcal{C}^1$. We suppose that the set $C \subset I\!R^n$ is convex (nonempty) and closed, and that the function $h : I\!R^m \to \overline{I\!R}$ is convex (proper) and lsc. We denote by D the (convex) set $\operatorname{dom} h$, so that the implicit system of constraints in $\big(\mathcal{P}(u,v,w)\big)$ is

$$x \in C \text{ and } u + G(w,x) \in D. \tag{3.6}$$

The vector w is a catch-all for any parameters, such as coefficients or matrix entries, on which g and G may depend, and with respect to which we may wish to consider perturbations. The explicit incorporation of the u and v as parameter vectors, in addition to w, facilitates the application of various technical results.

Because of the extra structure in this example the “solution” mapping to be studied is $M_0 : I\!R^m \times I\!R^n \times I\!R^p \to I\!R^n \times I\!R^m \times I\!R^p$ as defined by

$$M_0(u,v,w) = \{(x,y,z) \,|\, (y,z,-v) \in \partial f_0(u,w,x)\}. \tag{3.7}$$

Theorem 2. *The function f_0 representing the parameterized problem $\big(\mathcal{P}_0(u,v,w)\big)$ is Clarke regular everywhere. In terms of $L(w,x,y) := g(w,x) + \langle G(w,x), y\rangle$ one has that*

$$(y,z,-v) \in \partial f_0(u,w,x) \iff \begin{cases} y \in \partial h\big(u_G(w,x)\big),\ z = \nabla_w L(w,x,y), \\ 0 \in v + \nabla_x L(w,x,y) + N_C(x). \end{cases} \tag{3.8}$$

On the other hand, in terms of $L^\infty(w,x,y) := \langle G(w,x), y\rangle$ one has that

$$(y,z,-v) \in \partial^\infty f_0(u,w,x) \iff \begin{cases} y \in N_D\big(u + G(w,x)\big),\ z = \nabla_w L^\infty(w,x,y), \\ 0 \in \nabla_x L^\infty(w,x,y) + N_C(x). \end{cases} \tag{3.9}$$

Thus the basic constraint qualification for $\big(\mathcal{P}_0(\bar u,\bar v,\bar w)\big)$ at $\bar x$ is the condition that

$$\text{the only } \bar y \in N_D\big(\bar u + G(\bar w,\bar x)\big) \text{ with } -\nabla_x L^\infty(\bar w,\bar x,\bar y) \in N_C(\bar x) \text{ is } \bar y = 0. \tag{3.10}$$

Proof. It will be helpful here and later to observe that the function f_0 can be expressed in the composite form

$$f_0 = k \circ F \text{ for } \begin{cases} F(u,w,x) = \big(g(w,x), u + G(w,x), x\big), \\ k(\alpha,u,x) = \alpha + h(u) + \delta_C(x). \end{cases} \tag{3.11}$$

The mapping $F : I\!R^m \times I\!R^p \times I\!R^n \to I\!R \times I\!R^m \times I\!R^n$ is smooth, while the function $k : I\!R \times I\!R^m \times I\!R^n \to \overline{I\!R}$ is lsc, proper and convex. The formulas asserted for ∂f and $\partial^\infty f$ follow from this representation by the calculus rules in [15], and the Clarke regularity of f_0 is then obtainable from Poliquin [22]. □

The version of the basic constraint qualification that emerges in (3.10) reduces to the constraint qualification of Mangasarian-Fromovitz in the classical case of nonlinear programming where $C = I\!R^n$ and $h = \delta_K$ for the case $K = I\!R^x_+ \times \{0\}^{m-s}$, cf. [4, p. 94].

4. Proto-differentiation of Set-valued Mappings.

Our aim is to determine the vector pairs (x',y') giving the possible rates of perturbation of a (quasi-)optimal pair $(\bar x,\bar y)$ for $\big(\mathcal{P}(\bar u,\bar v)\big)$, relative to rates of perturbation in the parameter pair $(\bar u,\bar v)$ as specified by a pair (u',v'). In other words, we wish to differentiate the mapping M in some sense. But there is no classical guide to doing this, since M is not necessarily single-valued, and even if it were it would not submit itself to standard definitions of differentiability.

In analogy with the notion that a derivative should be a limit of difference quotients, we may consider for each $\tau > 0$ the difference quotient mapping

$$\begin{aligned} (u',v') \mapsto{}& \tau^{-1}[M(\bar u + \tau u', \bar v + \tau v') - (\bar x,\bar y)] \\ ={}& \{(x',y') \,|\, (\bar x + \tau x', \bar y + \tau y') \in M(\bar u + \tau u', \bar v + \tau v')\} \end{aligned} \tag{4.1}$$

and look for an appropriate limit as $\tau \searrow 0$. It turns out that graphical limits in terms of set convergence lead to the results we require. We proceed therefore to develop this idea from a general perspective.

A sequence of (set-valued) mappings $S^\nu : \mathcal{W} \rightrightarrows \mathcal{Z}$, where $\mathcal{W}$ and $\mathcal{Z}$ are Banach spaces, is said to *converge graphically* to a mapping $S : \mathcal{W} \rightrightarrows \mathcal{Z}$ if the sets $gphS^\nu = \{(w,z) \,|\, z \in S^\nu(w)\}$ converge in $\mathcal{W} \times \mathcal{Z}$ to $\operatorname{gph} S$:

$$\operatorname{gph} S = \limsup_{\nu\to\infty} \operatorname{gph} S^\nu = \liminf_{\nu\to\infty} \operatorname{gph} S^\nu. \tag{4.2}$$

In contrast, the mappings S^ν are said to *converge pointwise* to S if for each fixed u the sets $S^\nu(u)$ converge to $S(u)$. Graphical convergence and pointwise convergence can be seen to coincide when the effective domains $\operatorname{dom} S^\nu := \{u \,|\, S^\nu(u) \neq \emptyset\}$ are identical and the mappings uniformly enjoy some Lipschitz property, for instance, but in general neither type of convergence implies the other.

Consider now a mapping $S : \mathcal{W} \rightrightarrows \mathcal{Z}$ with closed graph. For fixed $\bar{w}$ and $\bar{z} \in S(\bar{w})$, define the quotient mapping $\Delta_\tau S(\bar{w}|\,\bar{z}) : \mathcal{W} \rightrightarrows \mathcal{Z}$ by

$$\Delta_\tau S(\bar{w}|\,\bar{z})(w') = \tau^{-1}\big[S(\bar{w}+\tau w') - \bar{z}\big] = \{z' \,|\, \bar{z} + \tau z' \in S(\bar{w}+\tau w')\}. \tag{4.3}$$

Define the *(outer) subderivative* mapping $DS(\bar{w}|\,\bar{z}) : \mathcal{W} \rightrightarrows \mathcal{Z}$ by

$$DS(\bar{w}|\,\bar{z})(\bar{w}') = \limsup_{\substack{w'\to\bar{w} \\ \tau\searrow 0}} \Delta_\tau S(\bar{w}|\,\bar{z})(w'). \tag{4.4}$$

If the mappings $\Delta_\tau S(\bar{w}|\,\bar{z})$ converge graphically as $\tau \searrow 0$, the limit mapping necessarily being $DS(\bar{w}|\,\bar{z})$, we say that S is *proto-differentiable* at $\bar{w}$ relative to $\bar{z}$. This holds if and only if $\operatorname{gph} S$ is derivable at $(\bar{w},\bar{z})$, inasmuch as the graph of the mapping $\Delta_\tau S(\bar{w}|\,\bar{z})$ is $\tau^{-1}\big[\operatorname{gph} S - (\bar{w},\bar{z})\big]$ and the graph of the mapping $DS(\bar{w}|\,\bar{z})$ is $T_{\operatorname{gph} S}(\bar{w},\bar{z})$. A stronger property is the *semi-differentiability* of S at $\bar{w}$ relative to $\bar{z}$, by which we mean the property that

$$DS(\bar{w}|\,\bar{z})(\bar{w}') = \lim_{\substack{w'\to\bar{w}' \\ \tau\searrow 0}} \Delta_\tau S(\bar{w}|\,\bar{z})(w') \text{ for all } \bar{w}'. \tag{4.5}$$

It should carefully be noted that the derivatives in all cases depend not only on the choice of a point $\bar{w}$ but also on the particular element $\bar{z}$ selected from $S(\bar{w})$, and this is indeed essential to the geometric interpretation. When $S(\bar{w})$ consists of a unique element $\bar{z}$, we of course write $DS(\bar{w})$ in place of $DS(\bar{w}|\,\bar{z})$. When S is a single-valued differentiable mapping in the finite-dimensional setting, one gets $DS(\bar{w})(w') = \nabla S(\bar{w})w'$, where $\nabla S(\bar{w})$ denotes the Jacobian matrix of S at $\bar{w}$.

Proto-differentiability was introduced in Rockafellar [3], although the idea of differentiating a set-valued mapping by constructing an appropriate tangent cone to its graph was first developed in detail in the book of Aubin and Ekeland [24]. Crucial to the usefulness of proto-differentiability is the geometric fact that the graph of S is quite often derivable at points where it may fail to be Clarke regular. For instance in the case of $S : I\!R \rightrightarrows I\!R$ where the graph of S is a broken curve composed of finitely many smooth segments, S is proto-differentiable everywhere even though its graph fails to be Clarke regular at the breakpoints.

Various properties of proto-differentiability and its connection with semi-differentiability are established in [3]. Semi-differentiability in special forms has been considered by Thibault [25], Shapiro [26] [27] and Robinson [28] [29], among others.

We adopt proto-differentiability as the fundamental concept to apply in the sensitivity analysis of the mapping M in parametric optimization. Thus we look for circumstances in which the difference quotient mappings in (4.1), denoted by $\Delta_\tau M(\bar{u},\bar{v}|\,\bar{x},\bar{y})$, converge graphically as $\tau \searrow 0$ and, when they do, have as their limit $DM(\bar{u},\bar{v}|\,\bar{x},\bar{y})$ a mapping for which a usable formula is available. A simple observation enables us to connect this up with the higher-order subdifferential analysis of f.

Theorem 3. *The solution mapping M is proto-differentiable at $(\bar{u},\bar{v})$ relative to $(\bar{x},\bar{y})$ if and only if the subgradient mapping ∂f is proto-differentiable at $(\bar{u},\bar{x})$ relative to $(\bar{y},-\bar{v})$. One has*

$$DM(\bar{u},\bar{v}|\,\bar{x},\bar{y})(u',v') = \{(x',y') \,|\, (y',-v') \in D(\partial f)(\bar{u},\bar{x}|\,\bar{y},-\bar{v})(u',x')\}. \tag{4.6}$$

Proof. This is immediate from the definitions via (3.4). □

The issue can now be seen in terms of generalized second derivatives of f: we wish to differentiate the mapping ∂f, which is already a type of first derivative of f. The classical framework provides us here with some possible guidelines. For the sake of insights, temporarily consider f simply as a function on $I\!R^n$ and suppose it is of class $\mathcal{C}^2$. The subgradient mapping ∂f reduces then to the usual gradient mapping $\nabla f : I\!R^n \to I\!R^n$, and proto-differentiation of ∂f comes down to ordinary differentiation of ∇f:

$$D(\nabla f)(\bar{w})(\bar{w}') = \lim_{\substack{w' \to \bar{w}' \\ \tau \searrow 0}} \frac{\nabla f(\bar{w}+\tau w') - \nabla f(\bar{w})}{\tau} = \nabla^2 f(\bar{w})\bar{w}', \tag{4.7}$$

where $\nabla^2 f(\bar{w})$ is the matrix of second-derivatives of f at $\bar{w}$. The matrix $\nabla^2 f(\bar{w})$ can also be obtained by another route, through a limit of second-order difference quotients:

$$\lim_{\substack{w' \to \bar{w}' \\ \tau \searrow 0}} \frac{f(\bar{w}+\tau w') - f(\bar{w}) - \tau\langle w', \nabla f(\bar{w})\rangle}{\frac{1}{2}\tau^2} = \langle \bar{w}', \nabla^2 f(\bar{w})\bar{w}'\rangle. \tag{4.8}$$

A key fact is this: if the function $\bar{w}' \mapsto \langle \bar{w}', \nabla^2 f(\bar{w})\bar{w}' \rangle$ in (4.8) is denoted by $D^2 f(\bar{w})$, the gradient mapping associated with it is twice the mapping $\bar{w}' \mapsto \nabla^2 f(\bar{w})\bar{w}'$ in (4.7), so that symbolically we have

$$\nabla\big(D^2 f(\bar{w})\big) = 2D(\nabla f)(\bar{w}) \text{ for all } \bar{w}. \tag{4.9}$$

The task presents itself of generalizing this formula from $\mathcal{C}^2$ functions to more general, possibly nonsmooth functions which, like the f in our perturbation model, may even be extended-real-valued. The attraction is that if derivatives of ∂f can be calculated in terms of a second-derivative function associated with f, it will be possible to give a *variational interpretation* to the formula in Theorem 3. In fact the derivatives of M will then be obtainable by solving a "derivative" optimization problem. To bring this to reality, we must work next on the general definition of second derivatives.

5. Second-Order Epi-Derivatives.

Let $f : \mathcal{W} \to \overline{I\!R}$ be any lsc, proper function, and let $\bar{w} \in \operatorname{dom} f$. Fix any $\bar{z} \in \partial f(\bar{w})$. For each $\tau > 0$ define the second-order difference quotient function $\Delta^2_\tau f(\bar{w}|\,\bar{z}) : \mathcal{W} \to \overline{I\!R}$ by

$$\Delta^2_\tau f(\bar{w}|\,\bar{z})(w') = \frac{f(\bar{w} + \tau w') - f(\bar{w}) - \tau\langle w', \bar{z}\rangle}{\frac{1}{2}\tau^2}. \tag{5.1}$$

Define the *(lower) second subderivative* function $D^2 f(\bar{w}|\,\bar{z}) : \mathcal{W} \to \overline{I\!R}$ by

$$D^2 f(\bar{w}|\,\bar{z})(\bar{w}') = \liminf_{\substack{w' \to \bar{w}' \\ \tau \searrow 0}} \Delta^2_\tau f(\bar{w}|\,\bar{z})(w'). \tag{5.2}$$

If the functions $\Delta^2_\tau f(\bar{w}|\,\bar{z})$ epi-converge as $\tau \searrow 0$, the limit necessarily being the function $D^2 f(\bar{w}|\,\bar{z})$, we say that f is *twice epi-differentiable* at $\bar{w}$ relative to $\bar{z}$—*properly* if, in addition, $D^2 f(\bar{w}|\,\bar{z})(0) > -\infty$ (which is equivalent to $D^2 f(\bar{w}|\,\bar{z})$ being proper).

When $\bar{z}$ is the sole element of $\partial f(\bar{w})$, we can just write $D^2 f(\bar{w})$ in place of $D^2 f(\bar{w}|\,\bar{z})$, but since the multi-valuedness of the mapping ∂f is an inherent feature of nonsmooth analysis, the dependence of second-order epi-derivatives at $\bar{w}$ on the choice of a subgradient $\bar{z} \in \partial f(\bar{w})$ must be faced in general.

A central question is whether there are useful functions f that are twice epi-differentiable and for which the second derivatives can readily be calculated. A surprisingly strong answer can be given in finite dimensions.

Theorem 4. *Suppose that a function $f : I\!R^n \to \overline{I\!R}$ can be represented in the form $f(w) = \varphi\big(F(w)\big)$, where $F : I\!R^n \to I\!R^m$ is a mapping of class $\mathcal{C}^2$ and $\varphi : I\!R^m \to \overline{I\!R}$ is a* *convex function that is piecewise linear-quadratic and such that the following condition is satisfied at a given point* $\widetilde{w} \in \operatorname{dom} f$*:*

$$\textit{the only } \widetilde{y} \in \partial^\infty \varphi\big(F(\widetilde{w})\big) \textit{ satisfying } \widetilde{y}\,\nabla F(\widetilde{w}) = 0 \textit{ is } \widetilde{y} = 0. \tag{5.3}$$

Then at all $\bar{w} \in \operatorname{dom} f$ *in some neighborhood of* $\widetilde{w}$*,* f *is Clarke regular at* $\bar{w}$ *with* $\partial f(\bar{w}) = \partial\varphi\big(F(\bar{w})\big)\nabla F(\bar{w})$*, and furthermore,* f *is twice epi-differentiable at* $\bar{w}$ *relative to every* $\bar{z} \in \partial f(\bar{w})$*.*

Proof. See [4, Thm. 4.5]. □

A complete formula for the second-order subderivative functions $D^2 f(\bar{w}|\,\bar{z})$ in Theorem 4 is furnished in [4, Thm. 4.5] as well, but we shall not go into the details here. The class of functions representable in this form is much broader than one might at first imagine. In saying that the convex function φ is *piecewise linear-quadratic* we mean that the convex set $\operatorname{dom} \varphi$ is a polyhedron decomposable into finitely many cells, relative to each of which there is a quadratic—or as a special case affine—expression for $\varphi(w)$. If the expressions are all affine, φ is *piecewise linear*, and this is equivalent to the epigraph of φ being polyhedral. Even with φ just piecewise linear one has large array of possibilities.

For example, a kind of function at the top of the list in most efforts at devising generalized second derivatives in nonsmooth analysis is that given as the pointwise maximum of a finite collection of $\mathcal{C}^2$ functions. Let us imagine more generally the sum of such a function and the indicator of a set described by finitely many $\mathcal{C}^2$ constraints: in notation,

$$f(w) = \max\{f_1(w), \cdots, f_r(w)\} + \delta_C(w) \tag{5.4}$$

where

$$C = \{w \,\big|\, f_i(w) \le 0 \text{ for } i = r+1, \cdots, s;\ f_i(w) = 0 \text{ for } i = s+1, \cdots, m\}, \tag{5.5}$$

all the functions f_i being of class $\mathcal{C}^2$. We then have the representation $f(w) = \varphi\big(F(w)\big)$ for the $\mathcal{C}^2$ mapping $F : w \mapsto \big(f_1(w), \cdots, f_m(w)\big)$ and the piecewise linear function

$$\varphi(\alpha_1, \cdots, \alpha_m) = \begin{cases} \max\limits_{i=1,\cdots,r} \alpha_i & \text{if } \alpha_i \begin{cases} \le 0 & \text{for } i = r+1, \ldots, s \\ = 0 & \text{for } i = s+1, \ldots, m \end{cases} \\ \infty & \text{otherwise.} \end{cases}$$

Moreover, condition (5.3) then reduces to an equivalent form of the Mangasarian-Fromovitz constraint qualification being satisfied in (5.5) at a point $\widetilde{w} \in C$. The conclusion then is that *a function of form (5.4)-(5.5) is twice epi-differentiable whenever the active constraints that may be involved satisfy the standard constraint qualification.* Incidentally, even if the term δ_C were dropped from (5.4) it would not necessarily be true that $D^2 f(\bar{w}|\,\bar{z})$ is finite everywhere.

This example alone covers the situations most commonly treated in nonlinear programming, but the general representation in Theorem 4 also encompasses penalty terms on constraints, augmented Lagrangians, and more. An explanation of such possibilities is given in [30]. It is also shown in [30] that a full theory of second-order optimality conditions in nonlinear programming can be derived from second-order epi-differentiation of the functions in Theorem 4.

A further result about second-order epi-differentiability has been provided by Do in his recent dissertation [31]. He has shown roughly that an integral functional of the form

$$I_f(w) = \int_\Omega f\big(\omega, w(\omega)\big)\mu(d\omega) \tag{5.6}$$

on $\mathcal{L}^p(\Omega, \mathcal{A}, \mu; \mathbb{R}^n), 1 < p < \infty$, is everywhere twice epi-differentiable when for each ω the function $f(\omega, \cdot)$ is convex and twice epi-differentiable everywhere. Here the twice epi-differentiability is of I_f can be interpreted in the Mosco sense: the second-order difference quotients epi-converge in the Mosco sense.

A general connection between second-order epi-differentiability of f and proto-differentiability of ∂f may be stated next as a far-reaching generalization of the classical formula in (4.9).

Theorem 5. *Let $\mathcal{W}$ be a reflexive Banach space and let $f : \mathcal{W} \to \overline{\mathbb{R}}$ be lsc, proper and convex. Let $\bar{z} \in \partial f(\bar{w})$. Then f is twice epi-differentiable (Mosco sense) at $\bar{w}$ relative to $\bar{z}$ if and only if ∂f is proto-differentiable at $\bar{w}$ relative to $\bar{z}$, in which case*

$$\partial\big[D^2 f(\bar{w}|\bar{z})\big] = 2D(\partial f)(\bar{w}|\,\bar{z}). \tag{5.7}$$

This result was obtained in finite dimensions in Rockafellar [32] and generalized to infinite dimensions by Do [31]. Its proof is based on Attouch's theorem [33], which connects the epi-convergence of convex functions with the graphical convergence of their subgradient mappings.

Although Theorem 5 is stated for convex functions only, it extends immediately to functions representable as $f = f_1 + f_2$ where f_1 is convex and f_2 is of class $\mathcal{C}^2$. Such functions still have convex effective domain, however, so this is not a way of managing to handle nonconvex constraints. For nonconvex constraints in finite dimensions, however, there is a powerful alternative.

Theorem 6 (Poliquin [34]). *For functions f of the form in Theorem 4 and satisfying condition (5.3), one likewise has formula (5.7).*

6. Application to Perturbations of Solutions.

The results in Theorems 5 and 6, when applied to the framework given in Theorem 3, provide a prescription for calculating the derivatives of the mapping M associated with the parameterization $\big(\mathcal{P}(u,v)\big)$. We content ourselves here with stating the version that follows from Theorem 6.

Theorem 7. *Let the function f describing $\big(\mathcal{P}(u,v)\big)$ be of the form in Theorem 4. For particular $(\bar{u},\bar{v})$ and $(\bar{x},\bar{y}) \in M(\bar{u},\bar{v})$, suppose that f satisfies condition (5.3) at $(\bar{u},\bar{x})$. Then M is proto-differentiable at $(\bar{u},\bar{v})$ relative to $(\bar{x},\bar{y})$. Moreover, in the notation $\widehat{M} := DM(\bar{u},\bar{v}|\,\bar{x},\bar{v})$ and $\widehat{f} = \frac{1}{2}D^2 f(\bar{u},\bar{x}|\,\bar{y},-\bar{v})$ one has*

$$\widehat{M}(u',v') = \{(x',y')\,\big|\,(y',-v') \in \partial\widehat{f}(u',x')\}. \tag{6.1}$$

The interpretation of formula (6.1) is quite appealing. Because this formula has the same structure as (3.4), except that M is replaced by $\widehat{M}$ and f by $\widehat{f}$, it is the formula associated with the optimality conditions for an auxiliary optimization problem in parametric form, namely

$$\big(\widehat{\mathcal{P}}(u',v')\big) \qquad \text{minimize } \widehat{f}(u',x') + \langle x',v'\rangle \text{ over all } x'.$$

In other words, *given $(\bar{u},\bar{v})$ and a choice of $(\bar{x},\bar{y}) \in M(\bar{u},\bar{v})$, we form the second-derivative function $\widehat{f} = \frac{1}{2}D^2 f(\bar{u},\bar{x}|\,\bar{y},-\bar{v})$ and are then able to calculate the proto-derivatives of M at $(\bar{u},\bar{v})$ relative to $(\bar{x},\bar{y})$ as follows. For each possible perturbation pair (u',v') of $(\bar{u},\bar{v})$, the elements (x',y') of $DM(\bar{u},\bar{v}|\,\bar{x},\bar{y})(u',v')$, which describe the corresponding perturbations of $(\bar{x},\bar{y})$, are the primal-dual pairs obtainable by solving $\big(\widehat{\mathcal{P}}(u',v')\big)$.* In the convex case, where not only f but $\widehat{f}$ will be convex, the interpretation is even stronger, because primal-dual pairs (x',y') are truly optimal than for the primal problem and a dual problem.

The more structured model $\big(\mathcal{P}_0(u,w,v)\big)$ described in Section 3 satisfies all the needed conditions as long as g and G are of class $\mathcal{C}^2$, h is piecewise linear-quadratic and C is polyhedral. This provides a wide spectrum of examples. There is a great deal to say about this, for which there is no space in these notes. Details in the setting of Theorem 7 will appear in Poliquin and Rockafellar [35]. An infinite-dimensional application to fully convex optimal control is developed in the dissertation of Do [31]. For infinite-dimensional problems with nonconvex constraints there are hardly any results at present, and even when such problems are convex much needs to be done to tie the general theory to the specifics of the integral functions and ordinary or partial differential equations that may be involved. Potential applications exist also to the sensitivity analysis carried out by different means in King and Rockafellar [36] and the Lagrangian form of perturbations in Rockafellar [37].

References

1. F. H. Clarke, *Nonsmooth Analysis and Optimization*, Wiley, 1983.
2. R. T. Rockafellar, "Generalized directional derivatives and subgradients of nonconvex functions," *Canadian J. Math.* **32** (1980), 157–180.
3. R. T. Rockafellar, "Proto-differentiability of set-valued mappings and its applications in optimization," *Annales de l'Institut H. Poincaré, Analyse Non Linéaire* (1989).
4. R. T. Rockafellar, "First- and second-order epi-differentiability in nonlinear programming," *Trans. Amer. Math. Soc.* **307** (1988), 75–107.
5. U. Mosco, "Convergence of convex sets and solutions to variational inequalities," *Advances in Math.* **3** (1969), 510–585.
6. H. Attouch and R. Wets, "Quantitative stability of variational systems: I. The epigraphical distance," preprint.
7. J. Beer and J. Borwein, "Mosco convergence and duality," preprint.
8. R. Wijsman, "Convergence of sequences of convex sets, cones, and functions, II," *Trans. Amer. Math. Soc.* **123** (1966), 32–45.
9. U. Mosco, "On the continuity of the Young-Fenchel transform," *J. Math. Anal. Appl.* **35** (1971), 518–535.
10. A. Shapiro, "Second-order sensitivity analysis and asymptotic theory of parameterized nonlinear programs," *Math. Prog.* **33** (1985), 280–299.
11. J.-P. Penot, "A characterization of tangential regularity," *Nonlin. Analysis: Theory, Meth. Appl.* **5** (1981), 625–643.
12. J. S. Treiman, "Characterization of Clarke's tangent and normal cones in finite and infinite dimensions," *Nonlinear Analysis: Theory Meth. Appl.* **7** (1983), 771–783.
13. R. T. Rockafellar, "Directionally Lipschitzian functions and subdifferential calculus," *Proc. London Math. Soc.* **39** (1979), 331–355.
14. J. Borwein and D. Ward, "Nonsmooth calculus in finite dimensions," *SIAM J. Control Opt.* **25** (1987), 1312–1340.
15. R. T. Rockafellar, "Extensions of subgradient calculus with applications to optimization," *Nonlin. Anal.: Theory Math. Appl.* **9** (1985), 665–698.

16. R. T. Rockafellar, “Existence and duality theorems for convex problems of Bolza,” *Trans. Amer. Math. Soc.* **159** (1971), 1–40.
17. F. H. Clarke, “A new approach to Lagrange multipliers,” *Math. of Oper. Research* **1** (1976), 165–174.
18. F. H. Clarke, “The generalized problem of Bolza,” *SIAM J. Control Opt.* **14** (1976), 682–699.
19. R. T. Rockafellar, “Duality and stability in extremum problems involving convex functions,” *Pacific J. Math.* **21** (1967), 167–187.
20. R. T. Rockafellar, “Conjugate convex functions in optimal control and the calculus of variations,” *J. Math. Anal. Appl.* **32** (1970), 174–222.
21. R. T. Rockafellar, *Conjugate Duality and Optimization*, SIAM Publications, 1974.
22. R. Poliquin, *Proto-differentiation and Integration of Proximal Subgradients*, dissertation, University of Washington, 1988.
23. J.-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, Wiley, 1984.
24. F. Mignot, “Côntrole dans les inéquations variationelles elliptiques,” *J. Functional Analysis* **22** (1976), 130–185.
25. L. Thibault, “Tangent cones and quasi-interiorly tangent cones to multifunctions,” *Trans. Amer. Math. Soc.* **277** (1983), 601–621.
26. A. Shapiro, “Directional differentiability of metric projections onto moving sets at boundary points,” *J. Math. Anal. Appl.* **131** (1988).
27. A. Shapiro, “On concepts of directional differentiability,” preprint.
28. S. M. Robinson, “Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity,” *Math. Programming Study* **30** (1987), 45–66.
29. S. M. Robinson, “An implicit function theorem for B-differentiable functions,” Working Paper WP-88-67, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1988.
30. R. T. Rockafellar, “Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives,” *Math. of Oper. Research* (1989).
31. C. Do, *Second-order Nonsmooth Analysis and Sensitivity in Optimization Problems Involving Convex Integral Functionals*, dissertation, University of Washington, 1989.
32. R. T. Rockafellar, “Generalized second derivatives of convex functions and saddle functions,” preprint.
33. H. Attouch, “Convergence de fonctions convexes, des sous-différentials et semi-groupes associés,” C.R. Acad. Sci. Paris **284** (1977), 539–542.

34. R. Poliquin, "Proto-differentiation of subgradient set-valued mappings," *Canadian J. Math.*, to appear.
35. R. Poliquin and R. T. Rockafellar, "Proto-derivatives of solution mappings in parametric optimization," preprint.
36. A. J. King and R. T. Rockafellar, "Sensitivity analysis for nonsmooth generalized equations," preprint.
37. R. T. Rockafellar, "Perturbations of generalized Kuhn-Tucker points in finite-dimensional optimization," *Nonsmooth Optimization and Related Topics* (F. Giannessi, ed.), Plenum, 1989.