Can. J. Math., Vol. XLII, No. 3, 1990, pp. 520–532

PROTO-DIFFERENTIATION OF SUBGRADIENT SET-VALUED MAPPINGS

RENÉ A. POLIQUIN

1. Introduction. Set-valued mappings arise quite naturally in optimization and nonsmooth analysis. In optimization, typically one has a family of optimization problems that depend on some parameter. One can then associate to this family of problems the set-valued mappings that assign to the parameter the set of optimal solutions, the set of feasible solutions or the set of multipliers. Many of these set-valued mappings encountered in optimization have been shown to be "proto-differentiable" (see Rockafellar [**16**]) i.e., in some sense these set-valued mappings are "differentiable". Using estimates provided by the proto-derivatives, see Proposition 2.1, one can then obtain information on how the sets depend on the parameter. The concept of proto-differentiation is described in Section 2.

In nonsmooth analysis, functions that are not differentiable in any classical sense are studied. To replace the gradient, several types of "subgradients" have been introduced. Proximal subgradients, lower semigradients and the (Clarke) generalized subgradients are all examples of subgradients; [**2**], [**6**] and [**10**]. In all three cases the mapping that assigns to each point the set of subgradients is potentially set-valued. Other examples of set-valued mappings are given by the normal cone, the tangent cone and the contingent cone; see [**2**]. Rockafellar has shown that the subgradient mapping of a convex function (they all agree for convex functions) is proto-differentiable almost everywhere. Since subgradients provide first-order information, this result can be viewed as giving second-order information on convex functions. In fact there is an equivalence between the subgradients being proto-differentiable and the convex function having a generalized second-order directional derivative; see Section 2.

The purpose of this paper is to show, in Section 4, that the generalized subgradient set-valued mapping of an "amenable" function is also proto-differentiable (again, as in the case of convex functions, all three subgradients mentioned earlier are the same for these functions; see [**6**]). Amenable functions are described in detail in Section 3; they are the composition of a "piecewise linear-quadratic" convex function with a C^2 (twice continuously differentiable) mapping, in addition a constraint qualification must be satisfied. Amenable functions are very important in optimization, in fact most common types of problems encountered in practice can be reformulated using these amenable functions; see [**13**].

We leave to another paper [**7**], the study of applications of this proto-differentiation result to sensitivity analysis of optimal solutions. The sensitivity

Received January 4, 1989 and in revised form May 17, 1989. This work was supported in part on NSERC grant OGP41983 and on a grant from the CRF at the University of Alberta.

AMS (1980) subject classification: 49A52.

results are similar to the ones obtained in [**17**], where the composition of a C^2 convex function with a C^2 mapping is investigated. Essentially for a family of parametrized optimization problems, formulated using amenable functions, the set-valued mapping that assigns to each parameter the set of vectors satisfying the first-order condition (i.e., the primal and dual vectors) is proto-differentiable, and the proto-derivatives can be calculated by solving an auxiliary problem similar in nature to the original optimization problem.

For a simple example of how the proto-differentiability of the generalized subgradient mapping can be used to yield information on the optimal solutions, consider the following linearly perturbed optimization problems:

$$P_v : \inf_{x \in \mathbf{R}^n} \{f(x) - \langle x, v \rangle\}.$$

If we let $\Gamma(v) = \{x : v \in \partial f(x)\}$, where $\partial f(x)$ is the set of generalized subgradients to f at x, then $\Gamma(v)$ includes the set of optimal solutions to the perturbed problems P_v. If we fix an optimal solution $\bar{x}$ to the unperturbed problem, i.e.,

$$f(\bar{x}) = \inf_{x \in \mathbf{R}^n} f(x),$$

then assuming that the subgradients are proto-differentiable at $\bar{x}$ relative to 0, we have Γ proto-differentiable at 0 relative to $\bar{x}$. Using the estimates described in Proposition 2.1 one could perform a kind of sensitivity analysis on the set of optimal solutions.

2. Proto-differentiability of a set-valued mapping and second-order epi-differentiability of a function. Rockafellar has recently introduced the concept of proto-differentiability of a set-valued mapping. Let $\Gamma : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ be a set-valued mapping (i.e., $\Gamma(x) \subset \mathbf{R}^n$) and v a vector in $\Gamma(x)$. To introduce a notion of "differentiation" of this set-valued mapping, the following difference quotients are defined: for $t > 0$, let

$$\Gamma_t = \frac{\operatorname{gph}\Gamma - (x, v)}{t} \quad (\Gamma_t \in \mathbf{R}^n \times \mathbf{R}^n),$$

where $\operatorname{gph}\Gamma$ is the graph of Γ i.e., $\{(y, z) : z \in \Gamma(y)\}$.

The set-valued mapping Γ is *proto-differentiable* at x relative to v if the sets Γ_t converge in the "Painlevé-Kuratowski" sense i.e.,

$$\limsup_{t \downarrow 0} \Gamma_t = \liminf_{t \downarrow 0} \Gamma_t \tag{2.1}$$

where

$$\limsup_{t \downarrow 0} \Gamma_t = \{w : \exists t_n \downarrow 0, \quad \text{and} \quad w_n \in \Gamma_{t_n} \quad \text{with} \quad w_n \longrightarrow w\}$$

and

$$\liminf_{t \downarrow 0} \Gamma_t = \{w : \forall t_n \downarrow 0, \exists w_n \in \Gamma_{t_n} \quad \text{with} \quad w_n \longrightarrow w\}.$$

https://doi.org/10.4153/CJM-1990-027-2 Published online by Cambridge University Press

The $\limsup_{t\downarrow 0} \Gamma_t$ is called the *contingent cone* of Γ at x relative to v, and the $\liminf_{t\downarrow 0} \Gamma_t$ is referred to as the *derivative cone* of Γ at x relative to v. The *proto-derivative*, $\Gamma'_{x,v} : \mathbf{R}^n \longrightarrow \mathbf{R}^n$, is the set-valued mapping whose graph is the limiting set in (2.1). Alternative descriptions of the lim sup and lim inf can be obtained by way of the distance function to a set; for such characterizations we refer to [**16**].

Just as the derivative of a function is used to establish estimates of the function, the same can be said of proto-derivatives. In the Proposition, the set B is the usual closed unit ball in $\mathbf{R}^n$.

PROPOSITION 2.1. (Rockafellar [**16**]) *For any* $\rho > 0$ *(arbitrarily large) and any* $\epsilon > 0$ *(arbitrarily small), there exists* $\tau > 0$, *such that for all* $t \in (0, \tau)$

$$(\Gamma_t \cap \rho B) \subset (\text{gph } \Gamma'_{x,v} + \epsilon B) \quad \textit{and}$$
$$(\text{gph } \Gamma'_{x,v} \cap \rho B) \subset (\Gamma_t + \epsilon B).$$

Proto-derivatives have been used by J.L. Ndoutoume [**5**] to obtain first-order necessary optimality conditions for the following problem: minimize $\{g(y) + h(u)\}$ subject to $B(u) + f \in A(y) + \partial\varphi(y)$ where $A : V \longmapsto V'$, $B : U \longmapsto V'$ are linear and continuous, $g : H \longmapsto \mathbf{R}$ is a locally Lipschitzian function, $h : U \longmapsto \mathbf{R} \cup \{\infty\}$ and $\varphi : V \longmapsto \mathbf{R} \cup \{\infty\}$ are closed (i.e., lower semicontinuous) convex functions and U, V and H are real Hilbert spaces. Proto-derivatives have also been used by King [**3**] to establish a generalized "delta theorem" of statistical nature for random sets.

When applied to a subgradient mapping, proto-differentiation gives rise to a second-order theory. As we will see, for convex functions the proto-derivative of the subgradient mapping is related to the following second-order difference quotient: for $f : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{\infty\}$ a closed function, x in the effective domain of f i.e., $x \in \text{dom } f = \{x : f(x) < \infty\}$, and $v \in \mathbf{R}^n$ let

$$\varphi_{x,v,t}(\xi) = \frac{f(x + t\xi) - f(x) - t\langle v, \xi\rangle}{(1/2)t^2},$$

where $\langle v, \xi\rangle$ is the usual dot product on $\mathbf{R}^n$. We say that f is *twice epi-differentiable* at x relative to v if

$$\limsup_{t\downarrow 0} \text{epi } \varphi_{x,v,t} = \liminf_{t\downarrow 0} \text{epi } \varphi_{x,v,t}$$

and the set $0 \times \mathbf{R}$ is not included in the limiting set (epi $\varphi_{x,v,t}$ is the set of all points lying on or above the graph of $\varphi_{x,v,t}$). The *second-order epi-derivative* is the function $f''_{x,v} : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{\infty\}$ whose graph is the limiting set.

Rockafellar has used second-order epi-derivatives to obtain second-order necessary and sufficient conditions for optimality; see [**14**]. These conditions are quite simple in nature: if f has a local minimum at $\bar{x}$, then $f''_{\bar{x},0}(\xi) \geqq 0$ for all

ξ; if 0 is a subgradient to f at $\bar{x}$ and $f''_{\bar{x},0}(\xi) > 0$ for all ξ, then f has a local minimum at $\bar{x}$ (in the strong sense).

For convex functions, the notions of proto-differentiability of the subgradients and second-order epi-differentiability of the function are equivalent. Essentially, the proto-derivative of the subgradients is the subgradient of the second-order epi-derivative.

THEOREM 2.2. (Rockafellar [**15**]) *Let $g : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{\infty\}$ be a closed proper (i.e.,* dom$g$ *is nonempty) convex function and $v \in \partial g(x)$. The function g is twice epi-differentiable at x relative to v if and only if ∂g is proto-differentiable at x relative to v. Moreover the subdifferential of the function $(1/2)g''_{x,v}$ is the proto-derivative of ∂g at x relative to v i.e.,*

$$\partial((1/2)g''_{x,v})(\xi) = (\partial g)'_{x,v}(\xi) \quad \textit{for all } \xi.$$

3. Amenable functions. To introduce amenable functions, we need to first introduce piecewise linear-quadratic functions and the basic constraint qualification.

Definition 3.1. A function $g : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{\infty\}$ is said to be *piecewise linear-quadratic* if D, the domain of g, is the union of finitely many polyhedra and g restricted to each polyhedron is quadratic (with linear as a special case).

Definition 3.2. For $F : \mathbf{R}^n \longrightarrow \mathbf{R}^m, g : \mathbf{R}^m \longrightarrow \mathbf{R} \cup \{\infty\}$ and $\bar{x} \in \operatorname{dom} g \circ F$, we say that the *basic constraint qualification* (b.c.q.) is satisfied at $\bar{x}$ if there does not exist y in $N_{\operatorname{dom} g}(F(\bar{x}))$ (the normal cone to dom$g$ at $F(\bar{x})$), with $y \neq 0$, and $y\nabla F(\bar{x}) = 0$. Here we think of y as a row vector and $\nabla F(\bar{x})$ as the matrix of partial derivatives.

When dom$g$ is the non-negative orthant, the basic constraint qualification turns into the dual statement of the familiar Mangasarian-Fromovitz constraint qualification [**4**]. The basic constraint qualification is a local notion i.e., if the b.c.q. holds at $\bar{x}$ then it holds on a neighborhood of $\bar{x}$; see [**13**], Proposition 4.9. There are two main purposes for the basic constraint qualification. On the one hand it guarantees that the generalized subgradients to $g \circ F$ at $\bar{x}$ can be evaluated using the generalized subgradients of g at $F(\bar{x})$ and the derivative of F at $\bar{x}$, i.e.,

$$\partial(g \circ F)(\bar{x}) = \partial g(F(\bar{x}))\nabla F(\bar{x});$$

see Rockafellar [**9**] and [**12**]. On the other hand, it ensures that for all v, the set $\{w \in \partial g(F(\bar{x})) : w\nabla F(\bar{x}) = v\}$ is bounded (hence compact). (Suppose $w_n \in \partial g(F(\bar{x}))$, with $\|w_n\| \uparrow \infty$ and $w_n\nabla F(\bar{x}) = v$. Assuming that $w_n/\|w_n\| \longrightarrow w$, it follows that $w \in N_{\operatorname{dom} g}(F(\bar{x}))$, with $\|w\| = 1$ and $w\nabla F(\bar{x}) = 0$; this contradicts the b.c.q. at $\bar{x}$.)

We now define amenable functions.

https://doi.org/10.4153/CJM-1990-027-2 Published online by Cambridge University Press

Definition 3.3. A function $f : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{\infty\}$ is *amenable* at $\bar{x} \in \operatorname{dom} f$, if in a neighborhood of $\bar{x}$, we have $f = g \circ F$ where $g : \mathbf{R}^m \longrightarrow \mathbf{R} \cup \{\infty\}$ is a piecewise linear-quadratic convex function and $F : \mathbf{R}^n \longrightarrow \mathbf{R}^m$ is a C^2 mapping. In addition the basic constraint qualification is satisfied at $\bar{x}$.

An example of an amenable function is the max of finitely many C^2 functions. As we mentioned in the introduction, most common types of optimization problems can be reformulated in terms of amenable functions; see [**13**] for the details. Rockafellar has also shown, see [**13**], that an amenable function is twice epi-differentiable with respect to any subgradient. It is even possible to write down a formula for the second-order epi-derivative.

THEOREM 3.4. (Rockafellar [**13**]) *Let $F : \mathbf{R}^n \longrightarrow \mathbf{R}^m$ be a C^2 mapping, $g : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{\infty\}$ be a piecewise linear-quadratic convex proper function, $f = g \circ F$, $\bar{x} \in \operatorname{dom} f$ and the basic constraint qualification is satisfied at $\bar{x}$. Under these assumptions, f is twice epi-differentiable at $\bar{x}$ relative to any vector $\bar{v}$ in*

$$\partial g(F(\bar{x}))\nabla F(\bar{x}) = \{w\nabla F(\bar{x}) : w \in \partial g(F(\bar{x}))\}$$

and

$$f''_{\bar{x},\bar{v}}(\xi) = \begin{cases} \gamma_{F(\bar{x})}(\nabla F(\bar{x})\xi) + \max\limits_{w \in W_{\bar{v}}(\bar{x})} \langle w, \xi\nabla^2 F(\bar{x})\xi\rangle & \text{if } \xi \in \Xi_{\bar{v}}(\bar{x}) \\ \infty & \text{if } \xi \notin \Xi_{\bar{v}}(\bar{x}) \end{cases}$$

where

$$\Xi_{\bar{v}}(\bar{x}) = N_{\partial f(\bar{x})}(\bar{v}) = \{\xi : g'_{F(\bar{x})}(\nabla F(\bar{x})\xi) = \langle \bar{v}, \xi\rangle\} = \operatorname{dom} f''_{\bar{x},\bar{v}},$$
$$W_{\bar{v}}(\bar{x}) = \{w \in \partial g(F(\bar{x})) : w\nabla F(\bar{x}) = \bar{v}\},$$

and

$$\gamma_{F(\bar{x})}(\nabla F(\bar{x})\xi) = \lim_{t \downarrow 0} \frac{[g(F(\bar{x}) + t\nabla F(\bar{x})\xi) - g(F(\bar{x})) - tg'_{F(\bar{x})}(\nabla F(\bar{x})\xi)]}{(1/2)t^2}.$$

Note. The symbol $\nabla F(\bar{x})$ refers to the $(m \times n)$ matrix of first partial derivatives, and when we write $w\nabla F(\bar{x})$ we think of w as a row vector. The $(m \times n \times n)$ matrix of second partial derivatives is denoted by $\nabla^2 F(\bar{x})$, and $\xi\nabla^2 F(\bar{x})\xi$ is the vector in $\mathbf{R}^m$ obtained by multiplying each of the Hessians, both on the left and on the right, by ξ.

We know that for convex functions, proto-differentiability of the subgradient mapping is equivalent to second-order epi-differentiability; see Theorem 2.2. The obvious question is whether such a result holds for amenable functions; this is the main result of this paper and is presented in Theorem 4.6.

4. Proto-differentiability of the subgradient mapping of an amenable function. In this section f is an amenable function at $\bar{x}$ with the following representation: $g : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{\infty\}$ is a piecewise linear-quadratic proper convex function, see Definition 3.1, with $D = \operatorname{dom} g$, $F : \mathbf{R}^n \longrightarrow \mathbf{R}^m$ is a twice continuously differentiable mapping, $f = g \circ F$, $\bar{x} \in \operatorname{dom} f$ and the basic constraint qualification, see Definition 3.2, is satisfied at $\bar{x}$.

The goal is to show that ∂f is proto-differentiable at $\bar{x}$ relative to $\bar{v}$ in

$$\partial f(\bar{x}) = \partial g(F(\bar{x}))\nabla F(\bar{x}).$$

Since this is a local notion, we assume that D (the effective domain of g) is a bounded polyhedron (hence compact).

The proof is divided into two major parts. In the first part, we prove the result assuming that $F(\bar{x}) \in \operatorname{int} \operatorname{dom} g$. This is fairly straightforward. In the general case, the trick is to convert the problem to the previous case. To do so, we need to develop a technique for extending piecewise linear-quadratic convex functions, i.e., take a piecewise linear-quadratic convex function whose domain is not the whole space and modify it to have full domain. In extending the function, we also need to preserve the piecewise linear-quadratic convex nature of the function.

Hence, the first case we study is when $F(\bar{x}) \in \operatorname{int}(D)$. To prove the result in this case we need the following lemma. The lemma proves that under certain conditions the generalized subgradient mapping of a "lower-C^2" function (locally the sum of the function and a multiple of the norm square is convex) is proto-differentiable. These functions were introduced by Rockafellar; for more details see [**11**]. In the lemma $\delta_C(x)$ refers to the indicator of C i.e., 0 if x is in the set C and ∞ otherwise and $\|\cdot\|$ is the usual norm on $\mathbf{R}^n$.

LEMMA 4.1. *Let $h : \mathbf{R}^n \longrightarrow \mathbf{R} \cup \{+\infty\}$ be lower-C^2 at $\bar{x} \in \operatorname{int}(\operatorname{dom} h)$ i.e., there exist $\lambda > 0$ and $\rho > 0$, such that $h(\cdot) + (\rho/2)\|\cdot\|^2 + \delta_{\bar{x}+\lambda B}(\cdot)$ is a convex function. If h is twice epi-differentiable at $\bar{x}$ relative to $\bar{v}$ in $\partial f(\bar{x})$, then ∂h is proto-differentiable at $\bar{x}$ relative to $\bar{v}$. Moreover*

$$\partial((1/2)h''_{\bar{x},\bar{v}})(\xi) = (\partial h)'_{\bar{x},\bar{v}}(\xi) \quad \textit{for all } \xi,$$

and $h''_{\bar{x},\bar{v}}(\cdot) + \rho\|\cdot\|^2$ is a convex function.

Proof. Let

$$g(x) = h(x) + (\rho/2)\|x\|^2 + \delta_{\bar{x}+\lambda B}(x).$$

It follows that $\partial g(x) = \partial h(x) + \rho x$ on $\operatorname{int}(\bar{x} + \lambda B)$. Since h is twice epi-differentiable at $\bar{x}$ relative to $\bar{v}$, the function g is twice epi-differentiable at $\bar{x}$ relative to $\bar{v} + \rho\bar{x}$ and

$$g''_{\bar{x},\bar{v}+\rho\bar{x}}(\xi) = h''_{\bar{x},\bar{v}}(\xi) + \rho\|\xi\|^2;$$

see Rockafellar [**13**]. By Theorem 2.2, the set-valued mapping ∂g is proto-differentiable at $\bar{x}$ relative to $\bar{v} + \rho\bar{x}$ and for all ξ

$$\partial((1/2)g''_{\bar{x},\bar{v}+\rho\bar{x}})(\xi) = (\partial g)'_{\bar{x},\bar{v}+\rho\bar{x}}(\xi).$$

We now show that these two facts imply that ∂h is proto-differentiable at $\bar{x}$ relative to $\bar{v}$ and that

$$(\partial g)'_{\bar{x},\bar{v}+\rho\bar{x}}(\xi) = (\partial h)'_{\bar{x},\bar{v}}(\xi) + \rho\xi.$$

To see this, let

$$(\xi, u) \in \limsup_{t\downarrow 0} \frac{\operatorname{gph}\partial h - (\bar{x}, \bar{v})}{t},$$

i.e., there exist $t_n \downarrow 0$ and $v_n \in \partial h(x_n)$ with

$$(1/t_n)(v_n - \bar{v}) \longrightarrow u \quad \text{and} \quad (1/t_n)(x_n - \bar{x}) \longrightarrow \xi.$$

But

$$(1/t_n)[(v_n + \rho x_n) - (\bar{v} + \rho\bar{x})] \longrightarrow u + \rho\xi;$$

and eventually, $(v_n + \rho x_n) \in \partial g(x_n)$. Therefore,

$$\begin{aligned}(\xi, u + \rho\xi) &\in \limsup_{t\downarrow 0} \frac{\operatorname{gph}\partial g - (\bar{x}, \bar{v} + \rho\bar{x})}{t}\\ &= \liminf_{t\downarrow 0} \frac{\operatorname{gph}\partial g - (\bar{x}, \bar{v} + \rho\bar{x})}{t}\end{aligned}$$

i.e., $(u + \rho\xi) \in (\partial g)'_{\bar{x},\bar{v}+\rho\bar{x}}(\xi)$.

Hence, for all $\sigma_n \downarrow 0$ there exists $w'_n \in \partial g(x'_n)$, such that

$$(1/\sigma_n)(x'_n - \bar{x}) \longrightarrow \xi \quad \text{and} \quad (1/\sigma_n)[w'_n - (\bar{v} + \rho\bar{x})] \longrightarrow u + \rho\xi.$$

Eventually, $w'_n = v'_n + \rho x'_n$, where $v'_n \in \partial h(x'_n)$. Hence,

$$(1/\sigma_n)(x'_n - \bar{x}) \longrightarrow \xi \quad \text{and} \quad (1/\sigma_n)(v'_n - \bar{v}) \longrightarrow u,$$

i.e.,

$$(\xi, u) \in \liminf_{t\downarrow 0} \frac{\operatorname{gph}\partial h - (\bar{x}, \bar{v})}{t}.$$

To establish the last part of the lemma, notice that

$$\begin{aligned}(\partial h)'_{\bar{x},\bar{v}}(\xi) = [(\partial g)'_{\bar{x},\bar{v}+\rho\bar{x}}(\xi) - \rho\xi] &= \partial((1/2)g''_{\bar{x},\bar{v}+\rho\bar{x}})(\xi) - \rho\xi\\ &= \partial((1/2h)''_{\bar{x},\bar{v}})(\xi).\end{aligned}$$

We now look at the case where $F(\bar{x}) \in \operatorname{int}(D)$.

PROPOSITION 4.2. *Let $\bar{v} \in \partial f(\bar{x})$. If $F(\bar{x}) \in \operatorname{int}(D)$, then f is lower-C^2 at $\bar{x}$ and ∂f is proto-differentiable at $\bar{x}$ relative to $\bar{v}$. Moreover,*

$$\partial((1/2)f''_{\bar{x},\bar{v}})(\xi) = (\partial f)'_{\bar{x},\bar{v}}(\xi) \quad \textit{for all} \quad \xi,$$

and $f''_{\bar{x},\bar{v}}$ is a lower-C^2 function i.e., there exists $\rho > 0$, such that $f''_{\bar{x},\bar{v}}(\cdot) + \rho\|\cdot\|^2$ is a convex function.

Proof. Since f is twice epi-differentiable at $\bar{x}$ relative to $\bar{v}$, see Rockafellar [**13**], by Lemma 4.1 we need only show that f is lower-C^2 at $\bar{x}$.

Let W be a compact neighborhood of $\bar{x}$, such that $F(x) \in \operatorname{int}(D)$ for all $x \in W$. Let

$$S = \bigcup_{x \in W} \partial g(F(x)).$$

The set S is compact and for all x in W

$$f(x) = \sup_{y \in S}\{\langle F(x), y\rangle - g^*(y)\},$$

where g^* is the convex conjugate of g; see [**8**]. Since g^* is continuous on S, f is lower-C^2 at $\bar{x}$; see Rockafellar [**11**] (this is actually the original definition of lower-C^2 functions i.e., in terms of a supremum representation).

To prove our main result, the trick is to modify the function g in order to have $F(\bar{x})$ in the interior of the modified function (then we can use Proposition 4.2). To do this we need to extend the domain of g, and still preserve the piecewise linear-quadratic property; for $\alpha > 0$, let

$$g_\alpha(y) = (g \,\square\, (1/\alpha)\|\cdot\|_1)(y) = \inf_z\{(1/\alpha)\|y - z\|_1 + g(z)\}. \tag{4.1}$$

$$\left(\|z\|_1 = \sum_1^m |z_i|, \quad \text{where } z = (z_1, z_2, \ldots, z_m).\right)$$

This operation is called inf-convolution; see Rockafellar [**8**]. There is considerable reference in the literature to this type of operation. In Wets [**19**], the set of functions $\{\lambda\|\cdot\| : \lambda > 0\}$ is an example of what is called a cast. Casts have been used by Wets to characterize epi-convergence. By far the most common operation, in the literature, is inf-convolution with $(1/2)\|\cdot\|^2$; this is called the Moreau-Yosida approximation. The technique, developed in (4.1), can be used to extend any convex function; it is employed in [**1**] to generate second-order optimality conditions in nonfinite composite convex optimization.

It follows from the properties of inf-convolution that

$$g_\alpha^*(\cdot) = g^*(\cdot) + \delta_{1/\alpha B_\infty}(\cdot),$$

where $B_\infty = \{x : \|x\|_\infty \leqq 1\}$.

By a result of J. Sun [**18**], the functions g_α are piecewise linear-quadratic convex. Indeed, he proves that a function is piecewise linear-quadratic convex if and only if its conjugate function is piecewise linear-quadratic (in our case $g^*(\cdot)$ and $\delta_{1/\alpha B_\infty}(\cdot)$ are piecewise linear-quadratic convex functions, since g is a piecewise linear-quadratic function and B_∞ is a polyhedron).

Since $D\,(=\text{ dom } g)$ is a compact set, g_α is also a finite function (i.e., dom $g_\alpha = \mathbf{R}^m$). Let $y \in D$ and suppose there exists $u \in \partial g(y)$ with $\|u\|_\infty \leqq 1/\alpha$. For all z, $g(z) \geqq g(y) + \langle u, z-y\rangle$. Since

$$\|v\|_1 = \sup_{\|u\|_\infty \leqq 1} \langle u, v\rangle,$$

we have that for all z

$$g(z) + (1/\alpha)\|z-y\|_1 \geqq g(z) - \langle u, z-y\rangle \geqq g(y).$$

In other words $g_\alpha(y) = g(y)$. We summarize the previous observations in the following proposition.

PROPOSITION 4.3. (a) *If there exist $u \in \partial g(y)$, with $\|u\|_\infty \leqq 1/\alpha$, then $g_\alpha(y) = g(y)$.*

(b) *The function $g_\alpha(\cdot)$ is a piecewise linear-quadratic finite convex function and for all $\bar{y}$ in D, there exist $\bar{\lambda} > 0$ and $\bar{\alpha} > 0$, such that if $\alpha \leqq \bar{\alpha}$ and y belongs to $(\bar{y} + \bar{\lambda}B) \cap D$, then $g_\alpha(y) = g(y)$.*

Proof. We need only show the second part of (b). Since g is piecewise linear-quadratic, there exist $\bar{\lambda} > 0$ and $\bar{\alpha} > 0$, such that $\forall y \in (\bar{y} + \bar{\lambda}B) \cap D$, there exists $u \in \partial g(y)$ with $\|u\|_\infty \leqq 1/\bar{\alpha}$.

Since we are assuming that the domain of g is compact, the infimum in the definition of g_α is always attained. This is crucial in establishing a formula for the subgradients of $g_\alpha(\cdot)$.

PROPOSITION 4.4. *If $g_\alpha(\bar{y}) = g(\bar{z}) + (1/\alpha)\|\bar{y} - \bar{z}\|_1$, where $\bar{z} \in D$, then $u \in \partial g_\alpha(\bar{y})$ if and only if $u \in \partial g(\bar{z})$ and $(\bar{y} - \bar{z}) \in N_{1/\alpha B_\infty}(u)$.*

Proof. $\Rightarrow$ Since $g_\alpha^* = g^* + \delta_{1/\alpha B_\infty}$, we have

$$\bar{y} \in \partial g_\alpha^*(u) = \partial g^*(u) + N_{1/\alpha B_\infty}(u) \quad \text{and} \quad \|u\|_\infty \leqq 1/\alpha.$$

For all z in D

$$g(z) \geqq g_\alpha(z) \geqq g_\alpha(\bar{y}) + \langle u, z - \bar{y}\rangle = g(\bar{z}) + (1/\alpha)\|\bar{y} - \bar{z}\|_1 + \langle u, z - \bar{y}\rangle. \tag{4.2}$$

This implies that

$$g(z) \geqq g(\bar{z}) + \langle u, \bar{y} - \bar{z}\rangle + \langle u, z - \bar{y}\rangle = g(\bar{z}) + \langle u, z - \bar{z}\rangle$$

(because $(1/\alpha)\|\bar{y}-\bar{z}\|_1 \geqq \langle u, \bar{y}-\bar{z}\rangle$). This shows that $u \in \partial g(\bar{z})$.

In (4.2), if $z = \bar{z}$, then

$$0 \geqq (1/\alpha)\|\bar{y}-\bar{z}\|_1 + \langle u, \bar{z}-\bar{y}\rangle.$$

But

$$(1/\alpha)\|\bar{y}-\bar{z}\|_1 \geqq \langle v, \bar{y}-\bar{z}\rangle \quad \text{if } \|v\|_\infty \leqq 1/\alpha.$$

Hence, $0 \geqq \langle u-v, \bar{y}-\bar{z}\rangle$ for all v with $\|v\|_\infty \leqq 1/\alpha$, i.e.,

$$(\bar{y}-\bar{z}) \in N_{1/\alpha B_\infty}(u).$$

$\Leftarrow$ If $u \in \partial g(\bar{z})$ and $(\bar{y}-\bar{z}) \in N_{1/\alpha B_\infty}(u)$, then

$$\bar{y} = \bar{z} + (\bar{y}-\bar{z}) \in \partial g_\alpha^*(u).$$

Therefore, $u \in \partial g_\alpha(\bar{y})$.

Corollary 4.5. (a) *If* $g_\alpha(y) = g(y)$, *then*

$$\partial g_\alpha(y) = \partial g(y) \cap (1/\alpha) B_\infty.$$

(b) *If* $y \notin D$ *and* $u \in \partial g_\alpha(y)$, *then* $\|u\|_\infty = 1/\alpha$.

Proof. (a) follows from Proposition 4.4, since $g_\alpha(y) = g(y)$. In (b), if

$$g_\alpha(y) = g(z) + (1/\alpha)\|y-z\|_1 \quad \text{and} \quad \|u\|_\infty < 1/\alpha,$$

then, by Proposition 4.4, $(y-z) \in N_{1/\alpha B_\infty}(u)$. But $N_{1/\alpha B_\infty}(u) = \{0\}$, therefore $y = z \in D$.

We now state our main result.

Theorem 4.6. *Let* $\bar{v} \in \partial f(\bar{x})$. *The set-valued mapping* ∂f *is proto-differentiable at* $\bar{x}$ *relative to* $\bar{v}$ *and*

$$\partial((1/2) f''_{\bar{x},\bar{v}})(\xi) = (\partial f)'_{\bar{x},\bar{v}}(\xi) \quad \textit{for all } \xi.$$

Moreover $f''_{\bar{x},\bar{v}}$ *is a lower-*C^2 *function i.e., there exist* $\rho > 0$, *such that* $f''_{\bar{x},\bar{v}}(\cdot) + \rho\|\cdot\|^2$ *is a convex function.*

Proof. By the basic constraint qualification and Proposition 4.3, there exist $\bar{\alpha}$, $\bar{\lambda}$, and $\bar{\sigma}$ positive, such that

$$g_{\bar{\alpha}}(y) = g(y) \quad \text{for all } y \in (F(\bar{x}) + \bar{\lambda} B) \tag{4.3}$$

and

$$\text{if} \quad w \nabla F(x) \in (\bar{v} + B), \tag{4.4}$$

where $w \in \partial g(F(x))$ and $x \in \bar{x} + \bar{\sigma}B$, then

$$\|w\|_\infty < 1/\bar{\alpha}.$$

To produce $\bar{\sigma}$, $\bar{\lambda}$ and $\bar{\alpha}$, first choose $\bar{\sigma}$, so that the basic constraint qualification holds for all x in $\bar{x} + \bar{\sigma}B$; see [**13**]. If no $\bar{\alpha}$ exists, then there exist $x_n \in \bar{x} + \bar{\sigma}B$ and $w_n \in \partial g(F(x_n))$, with $w_n \nabla F(x_n) \in (\bar{v} + B)$ and $\|w_n\| \uparrow \infty$. We may assume that

$$x_n \longrightarrow \tilde{x} \in (\bar{x} + \bar{\sigma}B) \quad \text{and} \quad \frac{w_n}{\|w_n\|} \longrightarrow w \in N_D(F(\tilde{x})).$$

This implies that $w\nabla F(\tilde{x}) = 0$, with $\|w\| = 1$, contradicting the basic constraint qualification at $\tilde{x}$.

Let $f_{\bar{\alpha}}(\cdot) = g_{\bar{\alpha}}(F(\cdot))$. By Proposition 4.3, $g_{\bar{\alpha}}$ is a finite piecewise linear-quadratic convex function and by the choice of $\bar{\alpha}$, it follows that $\bar{v} \in \partial f_{\bar{\alpha}}(\bar{x})$ and that if $F(x) \in (F(\bar{x}) + \bar{\lambda}B)$, then $f_{\bar{\alpha}}(x) = f(x)$. By Proposition 4.2, $\partial f_{\bar{\alpha}}$ is proto-differentiable at $\bar{x}$ relative to $\bar{v}$ and

$$\partial((1/2)(f_{\bar{\alpha}})''_{\bar{x},\bar{v}})(\xi) = (\partial f_{\bar{\alpha}})'_{\bar{x},\bar{v}}(\xi) \quad \text{for all } \xi.$$

The proof will be complete once we show that

$$(1) \qquad (f_{\bar{\alpha}})''_{\bar{x},\bar{v}}(\xi) = f''_{\bar{x},\bar{v}}(\xi)$$

($(f_{\bar{\alpha}})''_{\bar{x},\bar{v}}$ is lower-C^2 by Proposition 4.2) and

$$(2) \qquad (\partial f_{\bar{\alpha}})'_{\bar{x},\bar{v}}(\xi) = (\partial f)'_{\bar{x},\bar{v}}(\xi).$$

To show (1) we first show that the two functions in (1) have the same domain. We know by Theorem 3.4 that

$$\operatorname{dom}(f_{\bar{\alpha}})''_{\bar{x},\bar{v}} = N_{\partial f_{\bar{\alpha}}(\bar{x})}(\bar{v}).$$

By the choice of $\bar{\alpha}$,

$$[\partial f(\bar{x}) \cap (\bar{v} + B)] = [\partial f_{\bar{\alpha}}(\bar{x}) \cap (\bar{v} + B)].$$

(Since if $v \in [\partial f(\bar{x}) \cap (\bar{v} + B)]$, then $v = w\nabla F(\bar{x})$, where $w \in \partial g(F(\bar{x}))$ and $\|w\|_\infty < 1/\bar{\alpha}$. By Corollary 4.5, $w \in \partial g_{\bar{\alpha}}(F(\bar{x}))$.) Hence,

$$N_{\partial f_{\bar{\alpha}}(\bar{x})}(\bar{v}) = N_{\partial f(\bar{x})}(\bar{v}) = \operatorname{dom} f''_{\bar{x},\bar{v}}.$$

Therefore,

$$(4.5) \qquad \operatorname{dom}(f_{\bar{\alpha}})''_{\bar{x},\bar{v}} = N_{\partial f_{\bar{\alpha}}(\bar{x})}(\bar{v}) = N_{\partial f(\bar{x})}(\bar{v}) = \operatorname{dom} f''_{\bar{x},\bar{v}}.$$

If ξ belongs to (4.5), then in particular, by Theorem 3.4,

$$\nabla F(\bar{x})\xi \in T_D(F(\bar{x})).$$

For t small,

$$g_{\bar{\alpha}}(F(\bar{x}) + t\nabla F(\bar{x})\xi) = g(F(\bar{x}) + t\nabla F(\bar{x})\xi).$$

Hence,

$$\gamma_{F(\bar{x})}(\nabla F(\bar{x})\xi) = \lim_{t\downarrow 0} \frac{g_{\bar{\alpha}}(F(\bar{x}) + t\nabla F(\bar{x})\xi) - g_{\bar{\alpha}}(F(\bar{x})) - t(g_{\bar{\alpha}})'_{F(\bar{x})}(\nabla F(\bar{x})\xi)}{(1/2)t^2}.$$

In addition,

$$\begin{aligned}&\{w : w \in \partial g_{\bar{\alpha}}(F(\bar{x})) \text{ and } w\nabla F(\bar{x}) = \bar{v}\}\\ &= \{w : w \in \partial g(F(\bar{x})) \text{ and } w\nabla F(\bar{x}) = \bar{v}\}.\end{aligned}$$

By Theorem 3.4, $(f_{\bar{\alpha}})''_{\bar{x},\bar{v}} = f''_{\bar{x},\bar{v}}$.

To show (2), let

$$(\xi, u) \in \limsup_{t\downarrow 0} \frac{(\operatorname{gph} \partial f) - (\bar{x}, \bar{v})}{t},$$

i.e., there exist $t_n \downarrow 0$, $x_n \longrightarrow \bar{x}$ and $v_n \longrightarrow \bar{v}$ with

$$v_n \in \partial f(x_n), \quad (1/t_n)(v_n - \bar{v}) \longrightarrow u \quad \text{and} \quad (1/t_n)(x_n - \bar{x}) \longrightarrow \xi.$$

For all n we have $v_n = w_n \nabla F(x_n)$, where $w_n \in \partial g(F(x_n))$. Eventually, by (4.4), $\|w_n\|_\infty < 1/\bar{\alpha}$, hence $v_n \in \partial f_{\bar{\alpha}}(x_n)$. This shows that

$$(\xi, u) \in \limsup_{t\downarrow 0} \frac{(\operatorname{gph} \partial f_{\bar{\alpha}}) - (\bar{x}, \bar{v})}{t} = \liminf_{t\downarrow 0} \frac{(\operatorname{gph} \partial f_{\bar{\alpha}}) - (\bar{x}, \bar{v})}{t}.$$

So for all $\sigma_n \downarrow 0$ there exist $x'_n \longrightarrow \bar{x}$ and $v'_n \longrightarrow \bar{v}$ with

$$v'_n \in \partial f_{\bar{\alpha}}(x'_n), \quad (1/\sigma_n)(x'_n - \bar{x}) \longrightarrow \xi \quad \text{and} \quad (1/\sigma_n)(v'_n - \bar{v}) \longrightarrow u.$$

We know that $v'_n = w'_n \nabla F(x'_n)$, where $w'_n \in \partial g_{\bar{\alpha}}(F(x'_n))$, and again, eventually, $\|w'_n\|_\infty < 1/\bar{\alpha}$. By Corollary 4.5(b), $F(x'_n) \in D$. We have shown the existence of $x'_n \longrightarrow \bar{x}$ and $v'_n \longrightarrow \bar{v}$, with

$$v'_n \in \partial f(x'_n), \quad (1/\sigma_n)(x'_n - \bar{x}) \longrightarrow \xi \quad \text{and} \quad (1/\sigma_n)(v'_n - \bar{v}) \longrightarrow u.$$

Therefore,

$$(\xi, u) \in \liminf_{t\downarrow 0} \frac{(\operatorname{gph} \partial f) - (\bar{x}, \bar{v})}{t}.$$

We conclude that ∂f is proto-differentiable at $\bar{x}$ relative to $\bar{v}$ and that

$$u \in (\partial f')_{\bar{x},\bar{v}}(\xi) \Leftrightarrow u \in (\partial f_{\bar{\alpha}})'_{\bar{x},\bar{v}}(\xi).$$

References

1. J. V. Burke and R. A. Poliquin, *Second-order optimality conditions in non-finite convex composite optimization*, preprint.
2. F. H. Clarke, *Optimization and nonsmooth analysis* (John Wiley, New York, 1983).
3. A. J. King, *Generalized delta theorems for multivalued mappings and measurable selections*, preprint.
4. O. L. Mangasarian and S. Fromovitz, *The Fritz John conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl. *17* (1967), 37–47.
5. J. L. Ndoutoume, *Condition nécessaire d'optimalité du premier ordre pour des problèmes de contrôle optimal d'inéquations variationnelles*, Publications AVAMAC, Université Perpignan, Mathématiques, 66025 cedex.
6. R. A. Poliquin, *Proto-differentiation and integration of proximal subgradients*, Ph.D. dissertation, University of Washington (1988).
7. R. A. Poliquin and R. T. Rockafellar, *Proto-derivatives of solution mappings and sensitivity analysis in optimization*, forthcoming.
8. R. T. Rockafellar, *Convex analysis* (Princeton University Press, 1970).
9. ——— *Directionally Lipschitzian functions and subdifferential calculus*, Proceedings of the London Mathematical Society (3), *34* (1979).
10. ——— *The theory of subgradients and its applications to problems of optimization: Convex and nonconvex functions* (Heldermann, Berlin, 1981).
11. ——— *Favorable classes of Lipschitz-continuous functions in subgradient optimization*, Progress in Nondifferentiable Optimization, IISA Collaborative Proceedings Series, International Institute of Applied Systems Analysis, Laxenburg, Austria (1982), 125–144.
12. ——— *Extensions of subgradient calculus with applications to optimization*, Nonlinear Analysis *9* (1985), 665–698.
13. ——— *First and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc. *307* (1988), 75–108.
14. ———*Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. of Op. Res., to appear.
15. ——— *Generalized second derivatives of convex functions and saddle functions*, forthcoming.
16. ——— *Proto-differentiability of set-valued mappings and its applications in optimization*, Ann. Inst. H. Poincaré: Analyse Non Linéaire, submitted.
17. ——— *Perturbation of generalized Kuhn-Tucker points in finite-dimensional optimization*, forthcoming.
18. J. Sun, *On monotropic piecewise quadratic programming*, Ph.D. dissertation, University of Washington (1986).
19. R.J.B. Wets, *Convergence of convex functions, variational inequalities and convex optimization problems*, in *Variational inequalities and complimentary problems* (John Wiley and Sons).

University of Alberta,
Edmonton, Alberta