FIRST- AND SECOND-ORDER EPI-DIFFERENTIABILITY IN NONLINEAR PROGRAMMING

R. T. ROCKAFELLAR

ABSTRACT. Problems are considered in which an objective function express-
ible as a max of finitely many C^2 functions, or more generally as the composi-
tion of a piecewise linear-quadratic function with a C^2 mapping, is minimized
subject to finitely many C^2 constraints. The essential objective function in
such a problem, which is the sum of the given objective and the indicator
of the constraints, is shown to be twice epi-differentiable at any point where
the active constraints (if any) satisfy the Mangasarian-Fromovitz qualification.
The epi-derivatives are defined by taking epigraphical limits of classical first-
and second-order difference quotients instead of pointwise limits, and they re-
veal properties of local geometric approximation that have not previously been
observed.

1. Introduction. Nonlinear programming used to be viewed, at least for com-
putational purposes, as the minimization of a smooth (i.e. continuously differen-
tiable) objective function subject to finitely many equality or inequality constraints
given by other smooth functions. Many applications of optimization, however,
concern objective functions that are not necessarily smooth but of "max" type,
expressible as the pointwise maximum of certain other functions which are them-
selves smooth. Penalty representations of constraints, whether in terms of l_1 or
l_2 penalty functions or augmented Lagrangians, likewise have focused attention on
nonsmoothness. Such representations are now the rule of the day in the develop-
ment of numerical methods and can be used even in mathematical modeling itself.

A problem form that is becoming recognized as fundamental for theory and
computation in nonlinear programming and for its much greater versatility than
the traditional form is:

$$(P) \qquad \text{minimize } g(F(x)) \text{ subject to } F(x) \in D,$$

where g is a convex function on $\mathbb{R}^d$ with nonempty effective domain D (a convex
subset of $\mathbb{R}^d$), and F is a smooth mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. The present paper focuses
on the case where F is actually of class C^2 and g is "piecewise linear-quadratic"
(see Definition 1.1). This case is broad enough to cover all of the most common
types of problems seen in practice, as will be clear from examples.

Problem (P) can be identified abstractly with minimizing the function $f(x) = g(F(x))$ over all $x \in \mathbb{R}^n$. (Since D is the effective domain of g, one has $g(u) = \infty$ when $u \notin G$ and consequently $f(x) = \infty$ when $F(x) \notin D$.) We call f the *essential objective function* for (P). Our aim is to study f by methods of nonsmooth analysis. introducing certain first and second derivatives called "epi-derivatives" and developing formulas for them in terms of g and F. The idea is that by doing this in a "neoclassical" manner, with close parallels to the example of smooth unconstrained optimization where f is itself a $\mathcal{C}^2$ function, a unified foundation can be laid for the treatment of many questions of theoretical and computational interest in optimization.

The main difference between our work and that of others who have explored such an approach is in the choice of concepts and their level of generality. A number of authors have defined first and second derivatives in the framework of nonsmooth analysis that are viable for various kinds of functions f outside the class considered here, for instance Lipschitz continuous functions. and have used them to formulate necessary or sufficient conditions for a local minimum or in sensitivity analysis. Especially to be cited in this respect are Clarke [1, 2] (first derivatives only), Ioffe [3–5], Hiriart-Urruty [6–8], Chaney [9–12], Auslender [13]. Aubin [14], and Seeger [15]. We forego such breadth and concentrate instead on notions that only are appropriate in a more limited context—although it should be noted that we do, on the other hand. allow for extended-real-valuedness, in contrast to most of the authors cited.

From the technical standpoint, the use of epi-convergence of difference quotients in defining the "epi-derivatives" introduced in this paper deserves emphasis. Epi-convergence of functions, which refers to convergence of the epigraphs of the functions as sets. is coming to the fore as the correct concept for many situations in optimization. Until now, however, the only instance of epi-convergence being invoked in connection with derivatives was in a special analysis of "two-sided" second derivatives of convex functions in Rockafellar [16]. The strong feature of epi-convergence is that it corresponds to a geometric concept of approximation much like the one used in classical differential analysis. Derivatives defined in terms of it therefore have a certain "robustness" that can be advantageous. One of the principal objectives here may be seen as the identification of a central class of functions for which such derivatives do always exist.

In the context of mathematical programming the project we take on may best be compared with the efforts of Ben-Tal and Zowe in [17–20]. Those authors too introduce a kind of second derivative and investigate special classes of functions for which it exists. They use it to derive necessary and sufficient conditions for some of the most common types of problems in smooth and nonsmooth programming. A major difference between our results and theirs is that their second derivative does not correspond to a geometric notion of "uniform" approximation and therefore cannot be the basis of a truly abstract sufficient condition for optimality. They only get sufficiency by introducing a very particular structure for f. (Without such structure they have to assume that f is continuously differentiable with Lipschitz continuous gradient [20].) Our second-order "epi-derivative", when it exists (as for the functions f of the type in this paper) does support an abstract theory of sufficiency, as we shall demonstrate in [21].

Actually, as part of our work on "epi-derivatives" we also investigate certain "parabolic derivatives" that resemble the ones of Ben-Tal and Zowe, but involve epi-convergence. These reveal further connections between our approach and theirs.

To elucidate the nature of problem (P) under our assumptions, the following concept is required.

DEFINITION 1.1. A function $g: \mathbb{R}^d \to \bar{\mathbb{R}}$ with effective domain $D = \{u | g(u) < \infty\}$ will be called *piecewise linear-quadratic* if D can be expressed as the union of finitely many sets D_j (for $j \in J$, a finite index set), such that D_j is a convex polyhedron and the restriction of g to D_j is a quadratic (or affine) function. (Then in particular g must be continuous relative to the sets D_j, which are all closed, and consequently continuous relative to D.) If the restriction of g to D_j is affine for every j, then g will be called *piecewise linear*.

Note that when g is convex, as assumed in (P), the set D must itself be a convex polyhedron according to this definition. The class of piecewise linear convex functions is identical to the class of polyhedral convex functions defined in convex analysis [22]. Incidentally, a function representable as a maximum of finitely many quadratic (or affine) functions need *not* be piecewise linear-quadratic in the sense of Definition 1.1, because the joins between different "pieces" could be quadratic surfaces not subject to a polyhedral representation.

EXAMPLE 1.2. Let

$$(1.1) \qquad F(x) = (x, f_0(x), f_1(x), \ldots, f_m(x)) \in \mathbb{R}^n \times \mathbb{R}^{m+1}$$

where each f_i is of class $\mathcal{C}^2$,

$$(1.2) \qquad D = X \times \mathbb{R} \times I_1 \times \cdots \times I_m \subset \mathbb{R}^n \times \mathbb{R}^{m+1}$$

where X is a convex polyhedron in $\mathbb{R}^n$ and each I_i is a closed interval in $\mathbb{R}$ (bounded or unbounded),

$$(1.3) \qquad g(u) = g(x, \alpha_0, \alpha_1, \ldots, \alpha_m) = \begin{cases} \alpha_0 & \text{if } u \in D, \\ \infty & \text{if } u \notin D, \end{cases}$$

where $u = (x, \alpha_0, \alpha_1, \ldots, \alpha_m) \in \mathbb{R}^n \times \mathbb{R}^{m+1}$. The condition $F(x) \in D$ corresponds then to the general constraint system

$$(1.4) \qquad x \in X \text{ and } f_i(x) \in I_i \quad \text{for } i = 1, \ldots, m,$$

and for points x satisfying this system one has $g(F(x)) = f_0(x)$. Thus (P) is the problem

$$(1.5) \qquad \text{minimize } f_0(x) \text{ subject to } (1.4),$$

which is a traditional nonlinear programming problem with several refinements: the condition $x \in X$ is available for representing some underlying linear constraints (such as nonnegativity or bounds on variables) that are not conveniently expressed by constraint functions, and the conditions $f_i(x) \in I_i$ allow for general upper and/or lower bounds on the constraint functions f_i. (Obviously an inequality $f_i(x) \leq c_i$ corresponds to $I_i = (-\infty, c_i]$, while an equation $f_i(x) = c_i$ corresponds to $I_i = [c_i, c_i]$, the degenerate interval consisting of the single number c_i.)

EXAMPLE 1.3. This extends the preceding example of (P) to the case where the objective in (1.5) is of max type:

$$(1.6) \qquad f_0(x) = \max\{f_{01}(x), \ldots, f_{0s}(x)\}.$$

The component $f_0(x)$ in (1.1) is replaced by the vector $(f_{01}(x), \ldots, f_{0s}(x))$ and the component $\mathbb{R}$ in (1.2) by $\mathbb{R}^r$. (It is assumed that $f_{0k} \in \mathcal{C}^2$.) Instead of (1.3) one takes

$$(1.7) \quad g(u) = g(x, \alpha)_{01}, \ldots, \alpha_{0s}, \alpha_1, \ldots, \alpha_m) = \begin{cases} \max\{\alpha_{01}, \ldots, \alpha_{0s}\} & \text{if } u \in D, \\ \infty & \text{if } u \notin D. \end{cases}$$

Here $u = (x, \alpha_{01}, \ldots, \alpha_{0s}, \alpha_1, \ldots, \alpha_m) \in \mathbb{R}^n \times \mathbb{R}^r \times \mathbb{R}^m$.

EXAMPLE 1.4. The general l_1-penalty representation of the traditional problem (1.5) in Example 1.2 is to

$$(1.8) \qquad \text{minimize } f_0(x) + \sum_{i=1}^m p_i d_{I_i}(f_i(x)) \text{ over all } x \in X,$$

where the coefficients $p_i > 0$ are penalty parameters and

$$(1.9) \qquad d_{I_i}(\alpha_i) = [\text{distance of } \alpha_i \text{ from the interval } I_i].$$

(If $I_i = [0,0]$, then $d_{I_i}(\alpha_i) = |\alpha_i|$. If $I_i(-\infty, 0]$, then $d_{I_i}(\alpha_i) = [\alpha_i]_+$.) This penalty problem is the case of (P) where F is given by (1.1).

$$(1.10) \qquad\qquad D = X \times \mathbb{R}^{m+1}.$$

$$(1.11) \qquad g(u) = g(x, \alpha_0, \alpha_1, \ldots, \alpha_m) = \begin{cases} \alpha_0 + \sum_{i=1}^m p_i d_{I_i}(\alpha_i) & \text{if } x \in X, \\ \infty & \text{otherwise.} \end{cases}$$

Each of the functions d_{I_i} is piecewise linear on $\mathbb{R}$, so g is piecewise linear.

EXAMPLE 1.5. The basic l_2-penalty representation of problem (1.5) is the simple modification of Example 1.4 where d_{I_i} is replaced by $d_{I_i}^2$ in (1.8) and (1.11). Then g is piecewise linear-quadratic (and convex), not just piecewise linear. More generally one can look at representations of the form

$$(1.12) \qquad \text{minimize } f_0(x) + \sum_{i=1}^m \rho_i(d_{I_i}(f_i(x))) \text{ over } x \in X,$$

where $\rho_i : \mathbb{R}_+ \to \mathbb{R}_+$ is a nondecreasing, convex, piecewise linear-quadratic penalty function. Then the term $p_i d_{I_i}(\alpha_i)$ is replaced by $\rho_i(d_{I_i}(\alpha_i))$ in (1.11), and g is again piecewise linear-quadratic and convex. In stochastic programming, for instance, the case where ρ_i is quadratic initially but affine for high values is of interest; see Rockafellar and Wets [23].

EXAMPLE 1.6. Augmented Lagrangians of the standard quadratic-based kind lead to further cases of (P) that are important in computation. The general augmented Lagrangian expression associated with problem (1.5) can be written as

$$(1.13) \qquad f_0(x) + \sum_{i=1}^m \tfrac{1}{2} r_i [d_{I_i}^2(f_i(x) + \lambda_i/r_i) - (\lambda_i/r_i)^2],$$

where $r_i > 0$ and $\lambda_i \in \mathbb{R}$ are parameters. In $I_i = [0,0]$ one has

$$\tfrac{1}{2} r_i [d_{I_i}^2(f_i(x) + \lambda_i/r_i) - (\lambda_i/r_i)^2] = \lambda_i f_i(x) + \tfrac{1}{2} r_i f_i(x)^2$$

(the term introduced by Hestenes [24] and Powell [25] for equality constraints). whereas if $I_i = (-\infty, 0]$ one has

$$\tfrac{1}{2}r_i[d^2_{I_i}(f_i(x) + \lambda_i/r_i) - (\lambda_i/r_i)^2] = \begin{cases} \lambda_i f_i(x) + \tfrac{1}{2}r_i f_i(x)^2 & \text{when } f_i(x) \geq -\lambda_i/r_i. \\ -\tfrac{1}{2}\lambda_i^2/r_i & \text{when } f_i(x) \leq -\lambda_i/r_i \end{cases}$$

(the term introduced by Rockafellar [26] for inequality constraints). Minimizing the expression (1.13) over $x \in X$ is the case of (P) where F is given by (1.1), D is given by (1.10), and
(1.14)

$$g(x, \alpha_0, \alpha_1, \ldots, \alpha_r) = \begin{cases} \alpha_0 + \sum_{i=1}^{m} \dfrac{1}{2} r_i[d^2_{I_i}(\alpha_i + \lambda_i/r_i) - (\lambda_i/r_i)^2] & \text{if } x \in X, \\ \infty & \text{otherwise.} \end{cases}$$

Clearly g is a piecewise linear-quadratic convex function in this case also.

EXAMPLE 1.7. Unconstrained problems (P) where the objective function f has the following form have been explored by Ben-Tal and Zowe [17, 18]:

$$(1.15) \qquad f(x) = \sum_{i=1}^{m} h_i(f_i(x)), \quad \text{where } f_i(x) = \max_{j=1,\ldots,s_i} f_{ij}(x).$$

The functions $h_i \colon \mathbb{R} \to \mathbb{R}$ and $f_{ij} \colon \mathbb{R}^n \to \mathbb{R}$ are assumed to be of class $\mathcal{C}^2$, and differential properties of f are studied at a point $\bar{x}$ such that $h'_i(f_i(\bar{x})) \geq 0$ for $i = 1, \ldots, m$. If the slightly stronger assumption is made that h'_i is nonnegative on a *neighborhood* of $f_i(\bar{x})$ for every i, the situation fits the framework in the present paper, because f can then be written locally around $\bar{x}$ as $f(x) = g(F(x))$ with

$$(1.16) \qquad F(x) = (\ldots, h_i(f_{ij}(x)), \ldots),$$

$$(1.17) \qquad g(u) = \sum_{i=1}^{m} \max_{j=1,\ldots,s_i} u_{ij} \quad \text{for } u = (\ldots, u_{ij}, \ldots).$$

Clearly F is a $\mathcal{C}^2$ mapping from $\mathbb{R}^n$ to $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{s_m}$. and g is a piecewise linear convex function from $\mathbb{R}^{s_1} \times \cdots \times \mathbb{R}^{s_m}$ to $\mathbb{R}$.

Generalized second derivatives of functions f of type (1.15) have specifically been treated also by Chaney [12] as examples within his theory for classes of Lipschitz nonsmooth continuous functions. Such results will be important to us later in putting the results of this paper in perspective.

We would like to point out that the general problem (P) can be seen as arising from Lagrangian

$$(1.18) \qquad L(x, y) = y \cdot F(x) - h(y) \quad \text{and } y \in Y,$$

where h is the convex function conjugate to g and Y is its effective domain:

$$(1.19) \qquad h(y) = \sup_{u \in \mathbb{R}^d} \{y \cdot u - g(y)\},$$

$$(1.20) \qquad Y = \{y \in \mathbb{R}^d | h(y) < \infty\}.$$

Since g is in turn the conjugate of h, i.e.

$$(1.21) \qquad g(u) = \sup_{y \in \mathbb{R}^n} \{y \cdot u - h(y)\} = \sup_{y \in Y} \{y \cdot u - h(y)\},$$

one has

(1.22) $f(x) = g(F(x)) = \sup_{y \in Y} L(x,y)$ for all x.

Sun [27] in his recent dissertation has demonstrated that g is piecewise linear-quadratic if and only if h is piecewise linear-quadratic (the set Y then being, of course, a polyhedron). Our problems (P) are therefore precisely the ones that concern the minimization of a function f expressible as in (1.22) by a Lagrangian (1.18) where $F \in C^2$ and h is a piecewise linear-quadratic convex function with dom $h = Y$. This Lagrangian representation of (P) is a natural one for computational purposes and can also play a useful role in the statement of optimality conditions. It may be used alternatively as the basis for developing such conditions. For a complementary theory along such lines, in the more general setting where the convex function h need not be piecewise linear-quadratic (and epi-derivatives are not applicable), we refer to Burke [28].

The plan of the paper is first to define epi-derivatives in §2 and look at some of their elementary properties. The next task is to analyze in §3 the epi-derivatives of piecewise linear-quadratic convex functions and relate them to parabolic derivatives similar to the derivatives of Ben-Tal and Zowe. A constraint qualification is devised in §4 to handle the condition $F(x) \in D$ in (P) when x might be such that $F(x)$ is a boundary point of D. This is invoked along with the results in §3 to establish the existence of first- and second-order epi-derivatives of functions $f(x) = g(F(x))$ with g piecewise linear-quadratic convex, as in (P). (The centerpiece is Theorem 4.5.) Finally, a duality between second-order epi-derivatives and parabolic second derivatives is demonstrated.

2. Epi-derivatives. Our basic problem (P) corresponds to the essential objective function

(2.1) $f(x) = g(F(x))$ ($= \infty$ when $F(x) \notin D$),

under our stated assumptions about F, g and D. Although the specific structure inherent in any given case, such as illustrated by Examples 1.2–1.6, must ultimately be accommodated in analyzing (P) or computing its solutions, the viewpoint of the essential objective function enables us to draw parallels between different cases and to focus on the aspects that are the most fundamental. For the purposes of this section therefore, we do not assume that f is necessarily given by (2.1) but proceed more generally.

A brief mention of the classical ideas when f happens itself to be a function of class C^2 will put us in the right frame of mind for taking a "neoclassical" approach when f is not of such type. Classical first-order differentiability of f at x means the existence of a vector $v \in \mathbb{R}^n$ (which will be the gradient $\nabla f(x)$) such that

(2.2) $\sum_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \frac{f(x + t\xi') - f(x)}{t} = \xi \cdot v$ for all $\xi \in \mathbb{R}^n$.

Second-order differentiability, in the sense of a Taylor expansion of degree 2, means further the existence of a symmetric matrix H (which will be the hessian $\nabla^2 f(x)$)

such that

$$(2.3) \qquad \lim_{\substack{t\downarrow 0 \\ \xi'\to\xi}} \frac{f(x+t\xi') - f(x) - t\xi'\cdot v}{t^2/2} = \xi\cdot H\xi \quad\text{for all } \xi\in\mathbf{R}^n.$$

These concepts can be interpreted geometrically in terms of the *graph* of f. The limits in (2.2) and (2.3) describe local approximations to this graph at the point $(x, f(x))$ by the graphs of other functions, namely the ones giving the first- and second-order Taylor expansions of f around x.

Such ideas are not adequate to the task of handling more general functions f like the essential objective in (2.1). A lesson which has been learned in convex analysis and carried over to other forms of nonsmooth analysis is that the geometrical thinking that used to be directed towards the graph of f needs, for the sake of achieving a more versatile theory, to be directed instead towards the *epigraph* of f, which is the set

$$(2.4) \qquad \operatorname{epi} f = \{(x,\alpha)\in\mathbf{R}^n\times\mathbf{R}\,|\,\alpha\ge f(x)\}.$$

This is closed in $\mathbf{R}^n\times\mathbf{R}$ if and only if f is *lower semicontinuous* on $\mathbf{R}^n$, a property that holds certainly for the essential objective function (2.1) and is more appropriate anyway for many contexts in optimization than ordinary continuity.

A "neoclassical" approach to the local study of f can be characterized as an attempt to follow the classical approach, even when f is not differentiable, by working systematically with epigraphs instead of graphs. Such an approach calls for the replacement of the "graphical" limits (2.2) and (2.3) by "epigraphical" limits that yield first- and second-order approximations to f at x in a more general sense. What we refer to here as "epigraphical" limits are limits expressed by the notion of "epi-convergence", which was first introduced for convex functions by Wijsman [29, 30] (although not under that designation) and has in recent years come to be recognized as an analytical tool of great promise. We mention in particular the work as Mosco [31], DeGiorgi [32], Attouch [33], Wets [34]. Attouch and Wets [35]. An exposition of some of the main ideas and their motivation is contained in Rockafellar and Wets [36].

The notion of epi-convergence of functions, which we explain in a manner attuned to our purposes. depends on the notion of set convergence that has often been associated with Kuratowski [37] but really has a much longer history starting with Painlevé and his students. Rather than focusing on sequences. let us look at a family of subsets $S_t\subset\mathbf{R}^n$ parameterized (or indexed) by $t>0$. One says that S_t *converges* to a subset S as $t\downarrow 0$, written

$$(2.5) \qquad S = \lim_{t\downarrow 0} S_t.$$

if

$$(2.6) \qquad S = \limsup_{t\downarrow 0} S_t = \liminf_{t\downarrow 0} S_t,$$

where

$$(2.7) \quad \limsup_{t\downarrow 0} S_t := \{\xi\,|\,\exists \text{ sequences } t^\nu\downarrow 0,\ \xi^\nu\to\xi,$$

$$\text{with } \xi^\nu\in S_{t^\nu} \text{ for all } \nu=1,2,\dots\},$$

(2.8)
$$\liminf_{t\downarrow 0} S_t := \{\xi | \forall \text{ sequences } t^\nu \downarrow 0, \exists \xi^\nu \to \xi \text{ with } \xi^\nu \in S_{t^\nu}$$

$$\text{for all } \nu \text{ sufficiently high}\}.$$

An equivalent statement of (2.5) is this: S is a *closed* set such that for every $\varepsilon > 0$ and bounded set A there exists $\tau > 0$ for which

(2.9) $S_t \cap A \subset S + \varepsilon B$ and $S \cap A \subset S_t + \varepsilon B.$ $\forall t \in (0, \tau).$

(Here B denotes the closed unit ball in $\mathbb{R}^n$ with respect to the euclidean norm.) Still another way of expressing the concept is in terms of the distance functions

(2.10) $d_{S_t}(\xi) = \text{dist}(\xi, S_t) = \inf_{\xi' \in S_t} |\xi - \xi'|.$

One has (2.5) if and only if S is a *closed* set such that

(2.11) $\lim_{t\downarrow 0} d_{S_t}(\xi) = d_S(t),$ $\forall \xi \in \mathbb{R}^n.$

Consider now a family of functions $\varphi_t: \mathbb{R}^n \to \bar{\mathbb{R}}.$ where $\bar{\mathbb{R}} = [-\infty, \infty]$ (extended real line). One says that φ_t *epi-converges* to a function $\varphi: \mathbb{R}^n \to \bar{\mathbb{R}}$ as $t \downarrow 0$, written

(2.12) $\varphi = \text{epi-}\lim_{t\downarrow 0} \varphi_t.$

if the sets $\text{epi}\,\varphi_t$ converge to $\text{epi}\,\varphi$ in $\mathbb{R}^n \times \mathbb{R}$ as $t \downarrow 0$. Note that φ must in this case be a lower semicontinuous function. because the limit set $\text{epi}\,\varphi$ is necessarily closed. This concept is in general distinct from classical *pointwise* convergence. where $\varphi_t(\xi) \to \varphi(\xi)$ for each fixed ξ. The latter may or may not hold in a particular instance of epi-convergence. For example. when φ_t and φ are the indicators of sets S_t and S in $\mathbb{R}^n$ (vanishing on these sets but taking the value ∞ outside). the notion of epi-convergence in (2.12) is equivalent to that of set convergence (2.5): but depending on the way the sets S_t "move". the values of the functions φ_t can pop back and forth between 0 and ∞ at any fixed ξ and therefore will not necessarily converge to $\varphi(\xi)$.

One way of expressing the epi-convergence (2.12). in parallel with (2.6). is to user "semi-limits" of various kinds such as may be seen in the original papers of Wijsman [30. 31]. In the notation introduced by Rockafellar [38] one can write (2.12) equivalently as

(2.13) $\varphi(\xi) = \limsup_{t\downarrow 0} \inf_{\xi' \to \xi} \varphi_t(\xi') = \liminf_{t\downarrow 0} \inf_{\xi' \to \xi} \varphi_t(\xi').$

where

(2.14) $\limsup_{t\downarrow 0} \inf_{\xi' \to \xi} \varphi_t(\xi') := \lim_{\varepsilon \downarrow 0} \lim_{\tau \downarrow 0} \sup_{t \in (0,\tau)} \inf_{\xi' \in \xi + \varepsilon B} \varphi_t(\xi'),$

(2.15) $\liminf_{t\downarrow} \inf_{\xi' \to \xi} \varphi_t(\xi') := \lim_{\varepsilon \downarrow 0} \lim_{\tau \downarrow 0} \inf_{t \in (0,\tau)} \inf_{\xi' \in \xi + \varepsilon B} \varphi_t(\xi') = \lim_{\substack{t\downarrow 0 \\ \xi' \to \xi}} \inf \varphi_t(\xi').$

We turn now to the study of a function $f: \mathbb{R}^n \to \bar{\mathbb{R}}$ and a point x where f is finite. By applying epi-convergence to the usual difference quotients for f at x. in place of pointwise (or locally uniform pointwise) convergence, we obtain new concepts of "epi-differentiation" that will provide us with tools we need for a more general analysis of optimality.

DEFINITION 2.1. The function f is *epi-differentiable* at x if the first-order difference quotient functions

$$(2.16) \qquad \varphi_{x,t}(\xi) = [f(x + t\xi) - f(x)]/t \quad \text{for } t > 0$$

have the property that the limit function

$$(2.17) \qquad f'_x := \text{epi-}\lim_{t \downarrow 0} \varphi_{x,t}$$

exists and $f'_x(0) > -\infty$. Then the values $f'_x(\xi)$ are called *first-order (directional) epi-derivatives* of f at x. A vector $v \in \mathbf{R}^n$ is a *epi-gradient* of f at x if

$$(2.18) \qquad f'_x(\xi) \geq \xi \cdot v \quad \text{for all } \xi \in \mathbf{R}^n.$$

DEFINITION 2.2. The function f is *twice epi-differentiable at x relative to a vector v* if it is (once) epi-differentiable at x in the sense of the preceding definition and the second-order difference quotient functions

$$(2.19) \qquad \varphi_{x,v,t}(\xi) = [f(x + t\xi) - f(x) - t\xi \cdot v]/\tfrac{1}{2}t^2$$

have the property that the limit function

$$(2.20) \qquad f''_{x,v} := \text{epi-}\lim_{t \downarrow 0} \varphi_{x,v,t}$$

exists and $f''_{x,v}(0) > -\infty$. Then the values $f''_{x,v}(\xi)$ are called *second-order (directional) epi-derivatives* of f at x relative to v. A symmetric matrix $H \in \mathbf{R}^{n \times n}$ is called a *epi-hessian* of f at x relative to v if

$$(2.21) \qquad f''_{x,v}(\xi) \geq \xi \cdot H\xi \quad \text{for all } \xi \in \mathbf{R}^n.$$

Some elementary properties entailed by these definitions are explored in the following propositions.

PROPOSITION 2.3. *The first-order epi-derivative function f'_x, if it exists, is lower semicontinuous and positively homogeneous:*

$$(2.22) \qquad f'_x(\lambda\xi) = \lambda f'_x(\xi) \quad \text{for all } \lambda > 0, \ \xi \in \mathbf{R}^n.$$

The property $f'_x(0) > -\infty$ is equivalent to

$$(2.23) \qquad f'_x(\xi) > -\infty \quad \text{for all } \xi, \qquad f'_x(0) = 0.$$

PROOF. Any function expressible as an epi-limit is lower semicontinuous, as already noted. The positive homogeneity of f'_x is immediate from the form of the functions $\varphi_{x,t}$ in (2.16). Lower semicontinuity and positive homogeneity imply that if $f'_x(\xi) = -\infty$ for any ξ, then $f'_x(\lambda\xi) = -\infty$ for all $\lambda > 0$ and $f'_x(0) = -\infty$. On the other hand one has $f'_x(0) \leq 0$ trivially from the definition, so the property $f'_x(0) > -\infty$ must be equivalent to (2.23) as claimed. $\square$

PROPOSITION 2.4. *The function f is epi-differentiable at x if and only if the contingent cone to epi f at $(x, f(x))$, which is defined as the set*

$$(2.24) \qquad \limsup_{t \downarrow 0}[\text{epi } f - (x, f(x))]/t.$$

exists actually as a limit and does not contain the "downward pointing" vector $(0, -1)$. This cone is then the epigraph of the function f'_x:

$$(2.25) \qquad \text{epi } f'_x = \lim_{t \downarrow 0}[\text{epi } f - (x, f(x))]/t.$$

PROOF. This is immediate from the relation

(2.26) $$\text{epi } \varphi_{x,t} = [\text{epi } f - (x, f(x))]/t$$

and the definition of epi-convergence.   □

PROPOSITION 2.5. *Suppose that f is epi-differentiable at x. Then the epi-gradients of f at v, if any, are the vectors v such that*

(2.27) $$f(x') \geq f(x) + v \cdot (x' - x) + o(|x' - x|).$$

PROOF. Under epi-differentiability we have in particular from the epi-convergence expressions (2.13)–(2.15), as applied to the functions $\varphi_{x,t}$ that

$$\liminf_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \frac{f(x + t\xi') - f(x)}{t} = f'_x(\xi).$$

The defining inequality (2.18) for v to be a epi-gradient can therefore be written as

$$\liminf_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \frac{f(x + t\xi') - f(x) - t\xi' \cdot v}{t} \geq 0,$$

which is the same as (2.27) according to the meaning of the notation.   □

PROPOSITION 2.6. *The function f is epi-differentiable at x in particular when it is subdifferentially regular at x (in the sense of Clarke) and the subgradient set $\partial f(x)$ (in the sense of Clarke) is nonempty. Then $\partial f(x)$ consists exactly of the epi-gradients v of f at x, and f'_x is a convex function satisfying*

(2.28) $$f'_x(\xi) = \sup_{v \in \partial f(x)} \xi \cdot v \quad \text{for all } \xi.$$

PROOF. Subdifferential regularity is the case where the contingent cone to epi f at $(x, f(x))$ coincides with the Clarke tangent cone at $(x, f(x))$; see Clarke [1]. The Clarke tangent cone is by definition the set

$$\lim_{\substack{t \downarrow 0 \\ (x', \alpha) \to (x, f(x)) \\ (x', \alpha) \in \text{epi } f}} \inf \frac{\text{epi } f - (x', \alpha)}{t},$$

which is always contained in

(2.29) $$\liminf_{t \downarrow 0} \frac{\text{epi } f - (x, f(x))}{t}.$$

Subdifferential regularity thus guarantees the equality of the sets (2.24) and (2.29) and gives us the existence of the limit (2.25), as required. The Clarke tangent cone is moreover convex always, so that f'_x is then a convex function. The remaining assertions merely restate well-known facts of nonsmooth analysis in this case, but in the terminology and notation of Definition 2.1.   □

The significance of Proposition 2.6 lies in our ability to identify, by means of the well-developed subdifferential calculus for nonsmooth functions, a large number of instances where the property of subdifferential regularity does hold; cf. Clarke [1]. Convex functions in particular are subdifferentially regular, as are general "max" functions.

PROPOSITION 2.7. *The second-order epi-derivative function $f''_{x,v}$, if it exists, is lower semicontinuous and positively homogeneous of degree 2:*

$$(2.30) \qquad f''_{x,v}(\lambda \xi) = \lambda^2 f''_{x,v}(\xi) \quad \text{for all } \lambda > 0, \ \xi \in \mathbf{R}^n.$$

The property $f''_{x,v}(0) > -\infty$ implies

$$(2.31) \qquad f''_{x,v}(\xi) > -\infty \quad \text{for all } \xi, \qquad f''_{x,v}(0) = 0.$$

PROOF. As with Proposition 2.3, these properties are immediate from the form of the difference quotient functions involved in the definition. $\square$

PROPOSITION 2.8. *Suppose that f is twice epi-differentiable at x relative to a vector v. Then $f'_x(\xi) \geq \xi \cdot v$ for all ξ, i.e. v is a epi-gradient of f at x, and in addition*

$$(2.32) \qquad \{\xi | f''_{x,v}(\xi) < \infty\} \subset \{\xi | f'_x(\xi) = \xi \cdot v\}.$$

Furthermore there exists at least one epi-hessian of f at x relative to v. Indeed, H is such a epi-hessian if and only if H is a symmetric matrix satisfying

$$(2.33) \qquad f(x') \geq f(x) + v \cdot (x' - x) + \tfrac{1}{2}(x' - x) \cdot H(x' - x) + o(|x' - x|^2).$$

PROOF. The assumption of twice epi-differentiability implies by Proposition 2.7 that the function $f''_{x,v}$ is lower semicontinuous on $\mathbf{R}^n$ and nowhere $-\infty$. The value

$$(2.34) \qquad \bar{\mu} := \min_{|\xi|=1} f''_{x,v}(\xi)$$

therefore exists and is not $-\infty$. Properties (2.30) and (2.31) yield for any real $\mu \leq \bar{\mu}$ that

$$f''_{x,v}(\xi) \geq \mu |\xi|^2 \quad \text{for all } \xi \in \mathbf{R}^n.$$

In particular this inequality says that the matrix $H = \mu I$ is a epi-hessian, i.e. satisfies (2.21).

Recall now from the definition of $f''_{x,v}$ and the epi-convergence expression (2.13)–(2.15) as applied to the functions $\varphi_{x,v,t}$ that

$$(2.35) \qquad \liminf_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \frac{f(x + t\xi') - f(x) - t\xi' \cdot v}{\tfrac{1}{2}t^2} = f''_{x,v}(\xi).$$

The epi-hessian property (2.21) can therefore be written as

$$\liminf_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \frac{f(x + t\xi) - f(x) - t\xi' \cdot v - \tfrac{1}{2}t^2 \xi' \cdot H\xi'}{\tfrac{1}{2}t^2} \geq 0,$$

which is the meaning of (2.33). Of course (2.33) implies (2.27) and consequently by Proposition 2.5 that v must be a epi-gradient of f at x. The definition of $f''_{x,v}$ also asserts together with (2.35) that

$$(2.36) \qquad \limsup_{t \downarrow 0} \inf_{\xi' \to \xi} \frac{f(x + t\xi') - f(x)_t \xi' \cdot v}{\tfrac{1}{2}t^2} = f''_{x,v}(\xi),$$

and if $f''_{x,v}(\xi) < \infty$ this implies the existence of $\beta \in \mathbf{R}$ and $\xi_t \to \xi$ satisfying

$$\frac{f(x + t\xi_t) - f(x) - t\xi_t \cdot v}{\tfrac{1}{2}t^2} < \beta \quad \text{for } t \in (o, \tau)$$

(some $\tau > 0$). Then

$$[f(x + t\xi_t) - f(x)]/t \le \xi_t \cdot v + \tfrac{1}{2}t\beta \quad \text{for } t \in (0, \tau)$$

and accordingly

$$\limsup_{t \downarrow 0} \inf_{\xi' \to \xi} \frac{f(x + t\xi') - f(x)}{t} \le \xi \cdot v,$$

where the left side is $f'_x(\xi)$ by the definition of first-order epi-differentiability. Thus $f'_{x,v}(\xi) \le \xi \cdot v$ when $f''_{x,v}(\xi) < \infty$. This property in combination with v being a epi-gradient of f at x yields (2.32). $\Box$

The fact in Proposition 2.8 that v must be a epi-gradient leads us to the following simplified terminology.

DEFINITION 2.9. The function f will be called *twice epi-differentiable* at x (without mention of any particular vector v) if f is (once) epi-differentiable at x, at least one epi-gradient exists, and relative to *every* such epi-gradient v one has f twice epi-differentiable at x relative to v.

PROPOSITION 2.10. *Suppose that f is twice epi-differentiable at x. Let k be any C^2 function. Then the function $h = f + k$ is twice epi-differentiable at x. The epi-gradients of h at x are the vectors of the form $u = v + \nabla k(x)$ such that v is a epi-gradient of f at x, and for any such u one has*

$$(2.37) \qquad h''_{x,u}(\xi) = f''_{x,v}(\xi) + \xi \cdot \nabla^2 k(x)\xi.$$

PROOF. Clearly

$$(2.38) \qquad \frac{h(x + t\xi') - h(x) - t\xi' \cdot u}{\tfrac{1}{2}t^2} = \frac{f(x + t\xi') - f(x) - t\xi' \cdot v}{\tfrac{1}{2}t^2}$$
$$+ \frac{k(x + t\xi') - k(x) - t\xi' \cdot \nabla k(x)}{\tfrac{1}{2}t^2}$$

where

$$\lim_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \frac{k(x + t\xi') - k(x) - t\xi' \cdot \nabla k(x)}{\tfrac{1}{2}t^2} = \xi \cdot \nabla^2 k(x)\xi.$$

It is evident because of the strict convergence of the k quotient that equality is preserved in (2.38) when one takes either "lim sup inf" or "lim inf inf" on both sides of (2.38). This immediately gives the result. $\Box$

To explore comparisons with other derivative concepts, we need another definition, which will anyway turn out to be of importance later in this paper.

DEFINITION 2.11. Suppose that f is (first-order) epi-differentiable at x, and let ξ be a vector such that $f'_x(\xi) < \infty$. If the difference quotients

$$(2.39) \qquad \psi_{x,\xi,t}(\eta) = \frac{f(x + t\xi + \tfrac{1}{2}t^2\eta) - f(x) - tf'_x(\xi)}{\tfrac{1}{2}t^2}$$

epi-converge as $t \downarrow 0$, then the limit function will be denoted by $f''_x(\xi, \cdot)$, and the value $f''_x(\xi, \eta)$ will be called the *parabolic second derivative* of f at x relative to ξ and η. (Thus

$$(2.40) \qquad f''_x(\xi, \eta) = \limsup_{t \downarrow 0} \inf_{\eta' \to \eta} \psi_{x,\xi,t}(\eta') = \liminf_{t \downarrow 0} \inf_{\eta' \to \eta} \psi_{x,\xi,t}(\eta'),$$

where the assumption behind the symbol $f''_x(\xi, \eta)$ is that the two semi-limits are indeed equal.)

The parabolic second derivative f'' is closely related to the second derivative of

The function g'_u is convex and piecewise linear with effective domain

$$(3.3) \qquad\qquad\qquad \operatorname{dom} g'_u = T_D(u).$$

It is the support function of $\partial g(u)$, which is a nonempty convex polyhedron and coincides with the set of all epi-gradients y of g at u.

For any $y \in \partial g(u)$ the second epi-derivative function $g''_{u,y}$ is likewise expressed simply by taking limits along rays:

$$(3.4) \qquad\qquad g''_{u,y}(\omega) = \lim_{t\downarrow 0} \frac{g(u+t\omega) - g(u) - t\omega \cdot y}{\frac{1}{2}t^2}.$$

The function $g''_{u,y}$ is convex and piecewise linear-quadratic with effective domain

$$(3.5) \qquad\qquad \operatorname{dom} g''_{u,y} = \{\omega \in \mathbf{R}^n \mid g'_u(\omega) = \omega \cdot y\} = N_{\partial g(u)}(y).$$

Thus for $y \in \partial g(u)$ one has

$$(3.6) \qquad\qquad g''_{u,y}(\omega) = \begin{cases} \gamma_u(\omega) & \text{if } \omega \cdot y = g'_u(\omega), \\ \infty & \text{if } \omega \cdot y < g'_u(\omega), \end{cases}$$

where for $\omega \in \operatorname{dom} g'_u$ one defines

$$(3.7) \qquad\qquad \gamma_u(\omega) = \lim_{t\downarrow 0} \frac{g(u+t\omega) - g(u) - tg'_u(\omega)}{\frac{1}{2}t^2} < \infty$$

$$[= 0 \text{ if } g \text{ is actually piecewise linear}].$$

PROOF. Consider a representation of g as in Definition 1.1 in terms of polyhedral sets $D_j (j \in J)$. Fix $u \in D$ and let

$$(3.8) \qquad\qquad J_u = \{j \in J \mid u \in D_j\}.$$

For each $j \in J_u$ write

$$(3.9) \qquad g(u') = g(u) + q_j \cdot (u' - u) + \tfrac{1}{2}(u' - u) \cdot Q_j(u' - u) \quad \text{for } u' \in D_j$$

for some $q_j \in \mathbf{R}^d$ and symmetric $Q_j \in \mathbf{R}^{d\times d}$. Because D_j is polyhedral. there exists for each $j \in J_u$ an $\varepsilon_j > 0$ such that $[D_j - u] \cap \varepsilon_j B = T_D(u) \cap \varepsilon_j B$. In particular one has

$$(3.10) \qquad\qquad T_D(u) = \bigcup_{j \in J_u} T_{D_j}(u).$$

Let $\varepsilon = \min\{\varepsilon_j \mid j \in J_u\}$. Then for arbitrary $\rho > 0$ one has for all $\omega \in \rho B$ and $t \in (0, \varepsilon/\rho)$:

$$(3.11) \qquad \frac{g(u + t\omega) - g(u)}{t} = \begin{cases} q_j \cdot \omega + \tfrac{1}{2}t\omega \cdot Q_j\omega & \text{when } \omega \in T_{D_j}(u), \ j \in J_u, \\ \infty & \text{when } \omega \notin T_D(u). \end{cases}$$

Let $\varphi_{u,t}(\omega)$ and $\sigma_{u,t}(\omega)$ denote the left and right sides of (3.11), respectively; (3.11) asserts that $\varphi_{u,t}$ and $\sigma_{u,t}$ agree on the ball ρB when $t \in (0, \varepsilon/\rho)$. The epi-limit of $\sigma_{u,t}$ (as $t \downarrow 0$) obviously exists and equals

$$(3.12) \qquad\qquad \sigma(\omega) = \begin{cases} q_j \cdot \omega & \text{when } \omega \in T_{D_j}(u), \ j \in J_u, \\ \infty & \text{when } \omega \notin T_D(u), \end{cases}$$

since $\sigma_{u,t}$ differs from σ only by a function on $T_D(u)$ that converges to zero uniformly on bounded sets as $t \downarrow 0$. The epi-limit of $\varphi_{u,t}$ (as $t \downarrow 0$) therefore exists and equals σ too. This means that g'_u exists and

$$(3.13) \qquad g'_u(\omega) = \sigma(\omega) \quad \text{for all } \omega.$$

In particular then by formula (3.11), g'_u is a piecewise linear function, and by (3.12) its values satisfy (3.2). Thus g'_u agrees with the ordinary directional derivative function for g at u, which therefore must be a polyhedral convex function, in particular closed. In general the closure of the ordinary directional derivative function is conjugate to the indicator of $\partial g(u)$ [**22**, Theorem 23.2], and polyhedral convexity is preserved under conjugacy [**22**, Theorem 19.2]. Therefore $\partial g(u)$ is a nonempty polyhedral convex set whose support function (the conjugate of its indicator) is g'_u. Then too $\partial g(u)$ consists of the vectors y satisfying $g'_u(\omega) \geq \omega \cdot y$ for all ω, which are by definition the epi-gradients of g at u.

Passing now to second-order concerns, we transform (3.11) into the assertion that

$$(3.14) \qquad \frac{g(u+t\omega) - g(u) - t\omega \cdot y}{\frac{1}{2}t^2}$$
$$= \begin{cases} \omega \cdot Q_j \omega + 2[q_j - y] \cdot \omega/t & \text{when } \omega \in T_{D_j}(u),\ j \in J_u, \\ \infty & \text{when } \omega \notin T_D(u), \end{cases}$$

this being true for all $\omega \in \rho B$ when $t \in (0, \rho/\varepsilon)$. Let

$$(3.15) \qquad \psi(\omega) = \begin{cases} \omega \cdot Q_j \omega & \text{when } \omega \in T_{D_j}(u),\ j \in J_u, \\ \infty & \text{if } \omega \notin T_D(u), \end{cases}$$

and observe that

$$(3.16) \qquad \psi(\omega) = \lim_{t \downarrow 0} \frac{g(u+t\omega) - g(u) - t g'_u(\omega)}{\frac{1}{2}t^2} \quad \text{for all } \omega \in \mathbf{R}^d$$

by virtue of (3.11)–(3.13). Denote the difference quotient in (3.14) by $\varphi_{u,y,t}(\omega)$. Recalling that (3.12) gives $g'_u(\omega)$, we can write (3.14) as

$$(3.17) \qquad \varphi_{u,y,t}(\omega) = \psi(\omega) + 2[g'_u(\omega) - \omega \cdot y]/t,$$

an equation that holds for all $\omega \in \rho B$ as long as $t \in (0, \varepsilon/\rho)$. Under the assumption that y is a epi-gradient of g at y, we have $g'_u(\omega) - \omega \cdot y \geq 0$ for all ω, and the right side of (3.17) therefore epi-converges as $t \downarrow 0$ to

$$\psi_0(\omega) = \begin{cases} \psi(\omega) & \text{if } g'_u(\omega) - \omega \cdot y = 0, \\ \infty & \text{if } g'_u(\omega) - \omega \cdot y > 0. \end{cases}$$

It follows that $\varphi_{u,y,t}$ epi-converges likewise to ψ_0 as $t \downarrow 0$. Thus $g''_{u,y}$ exists and equals ψ_0. We can therefore write

$$(3.18) \qquad g''_{u,y} = \psi + \delta_K,$$

where δ_K is the indicator of $K = \{\omega | g'_u(\omega) = \omega \cdot y\}$. The fact already established, namely that g'_u is the support function $\partial g(u)$, i.e. satisfies

$$g'_u(\omega) = \sup_{y' \in \partial g(u)} \omega \cdot y',$$

implies

(3.19) $K = \{\omega | \omega \cdot (y' - y) \leq 0 \text{ for all } y' \in \partial g(u)\} = N_{\partial g(u)}(y)$

and consequently that K is a *polyhedral* convex cone, because $\partial g(u)$ is a polyhedral convex set. When (3.18) is combined with (3.16) and (3.19), we obtain justification for the claim that (3.5) holds, and that on the set in (3.5), formula (3.4) holds.

The fact that $g''_{u,y}$ is piecewise linear-quadratic follows from (3.18) and (3.15) because the cones $T_{D_j}(u)$ and K are polyhedral: one can write

$$\text{dom } g''_{u,y} = \bigcup_{j \in J_u} [K \cap T_{D_j}(u)].$$

The convexity of $g''_{u,y}$ follows from its definition as an epi-limit of functions $\varphi_{u,y,t}$, each of which is convex: in the context of epigraphs one observes that the limit of convex sets S_t as $t \downarrow 0$, if it exists, must be convex. $\square$

Next on our agenda are some results about the parabolic second derivative in Definition 2.11 and its relationship to second-order epi-derivatives of g.

PROPOSITION 3.2. *For any* $u \in D$ *and* $\omega \in T_D(u)$ *the parabolic second-derivative function* $g''_u(\omega, \cdot)$ *exists and has the formula*

(3.20) $g''_u(\omega, \varsigma) = g''_u(\omega, 0) + (g'_u)'_\omega(\varsigma) \quad \text{for all } \varsigma \in \mathbf{R}^d,$

with

(3.21) $g''_u(\omega, 0) = \lim_{t \downarrow 0} \dfrac{g(u + t\omega) - g(u) - t g'_u(\omega)}{\frac{1}{2} t^2}.$

Furthermore, $g''_u(\omega, \varsigma)$ *is a proper convex piecewise linear function of* ς *with effective domain equal to*

(3.22) $T'_D(u, \omega) := [\text{tangent cone to } T_D(u) \text{ at } \omega].$

PROOF. We continue with the notation and setting that led in the proof of the preceding theorem to (3.11): we have for any $\rho > 0$ that

(3.23) $g(u + t\omega + \frac{1}{2}t^2\varsigma) = g(u) + q_j(t\omega + \frac{1}{2}t^2\varsigma) + \frac{1}{2}(t\omega + \frac{1}{2}t^2\varsigma) \cdot Q_j(t\omega + \frac{1}{2}t^2\varsigma)$
$$\text{when } \omega + \tfrac{1}{2}t\varsigma \in T_{D_j}(u) \cap \rho B \text{ and } t \in (0, \varepsilon/\rho).$$

Here ω is fixed as well as u, and our concern is centered on behavior with respect to ς. Accordingly we take $\rho > |\omega|$ and define

(3.24) $J_{u,\omega} = \{j \in J_u | \omega \in T_{D_j}(u)\},$

(3.25) $T'_{D_j}(u, \omega) = [\text{tangent cone to } T_{D_j}(u) \text{ at } \omega],$

so that

(3.26) $T'_D(u, \omega) = \bigcup_{j \in J_{u,\omega}} T'_{D_j}(u, \omega).$

The cone $T'_{D_j}(u, \omega)$ agrees with $T_{D_j}(u) - \omega$ on some neighborhood of 0 because $T_{D_j}(u)$ is polyhedral. The condition $\omega + \frac{1}{2}t\varsigma \in T_{D_j}(u) \cap \rho B$ in (3.23) is therefore equivalent to $\frac{1}{2}t\varsigma \in T'_{D_j}(u, \omega) \cap (\rho B - \omega)$ when $|t\varsigma|$ is smaller than some $\delta_j > 0$. Take $\delta \in (0, 1)$ to be smaller than these δ_j's and ε. Then as long as $\rho \geq |\omega| + 1$, say,

the assumptions in (3.23) will be satisfied when $\varsigma \in T'_{D_j}(u, \omega) \cap \rho B$ and $t \in (0, \delta/\rho)$: one will have

$$(3.27) \quad g(u + t\omega + \tfrac{1}{2}t^2\varsigma) = g(u) + tq_j \cdot (\omega + \tfrac{1}{2}t\varsigma) + \tfrac{1}{2}t^2(\omega + \tfrac{1}{2}t\varsigma) \cdot Q_j(\omega + \tfrac{1}{2}t\varsigma)$$
$$\text{when } \varsigma \in T'_{D_j}(u, \omega) \cap \rho B \text{ and } t \in (0, \delta/\rho)$$

as long as ρ is sufficiently high. Remembering at the same time that

$$(3.28) \qquad g'_u(\omega) = g_j \cdot \omega \quad \text{for any } j \in J_{u,\omega}$$

we can write (3.27) in terms of the difference quotients

$$(3.29) \qquad \psi_{u,\omega,t}(\varsigma) = \frac{g(u + t\omega + \tfrac{1}{2}t^2\varsigma) - g(u) - tg'_u(\omega)}{\tfrac{1}{2}t^2}$$

as saying that the formula
(3.30)

$$\psi_{u,\omega,t}(\varsigma) = \begin{cases} q_j \cdot \varsigma + (\omega + \tfrac{1}{2}t\varsigma) \cdot Q_j(\omega + \tfrac{1}{2}t\varsigma) & \text{when } \varsigma \in T'_{D_j}(u, \omega), \ j \in J_{u,\omega}, \\ \infty & \text{when } \varsigma \notin T'_D(u, \omega), \end{cases}$$

holds when $\varsigma \in \rho B$ and $t \in (0, \delta/\rho)$, as long as $\rho > |\omega| + 1$. The function of ς given by the right side of (3.30) obviously epi-converges to

$$(3.31) \qquad \begin{cases} g_j \cdot \varsigma + \omega \cdot Q_j \omega & \text{when } \varsigma \in T'_{D_j}(u, \omega), \ j \in J_{u,\omega}, \\ \infty & \text{when } T'_D(u, \omega), \end{cases}$$

and therefore $\psi_{u,\omega,t}$ epi-converges to this too as $t \downarrow 0$. Thus $g''_u(\omega, \cdot)$, which is defined to be the epi-limit of $\psi_{u,\omega,t}$, exists and is given as a function of ς by (3.31). From the formula for g'_u derived in the proof of Theorem 3.1 as (3.19), applied in turn to the function g'_u in place of g so as to get $(g'_u)'_\omega$, we obtain

$$(3.32) \qquad (g'_u)'_\omega(\varsigma) = \begin{cases} q_j \cdot \varsigma & \text{when } \varsigma \in T'_{D_j}(u, \omega), \ j \in J_{u,\omega}, \\ \infty & \text{when } \varsigma \notin T'_D(u, \omega). \end{cases}$$

From (3.28) and (3.29), on the other hand, we have

$$(3.33) \qquad \lim_{t \downarrow 0} \frac{g(u + t\omega) - g(u) - tg'_u(\omega)}{\tfrac{1}{2}t^2} = \omega \cdot Q_j \omega \quad \text{for any } j \in J_{u,\omega}.$$

When (3.32) and (3.33) are plugged into formula (3.30) for $g''_u(\omega, \varsigma)$, the result is (3.20)-(3.21). Finally we observe that the functions $\psi_{u,\omega,t}$ are convex because g is convex. and their epi-limit $g''_u(\omega, \cdot)$ is therefore convex also. The formula (3.21) obtained for $g''_u(\omega, \varsigma)$ shows that this expression is piecewise linear in ς with effective domain $T'_D(u, \omega)$. $\square$

To help in stating the next result, we introduce some minor notation and terminology.

DEFINITION 3.3 For a mapping $\theta: [0, \tau) \to \mathbb{R}^d$, we shall write

$$(3.34) \qquad \dot\theta(0) = \lim_{t \downarrow 0} \frac{\theta(t) - \theta(0)}{t} \quad \text{(if the limit exists)},$$

$$(3.35) \qquad \ddot\theta(0) = \lim_{t \downarrow 0} \frac{\theta(t) - \theta(0) - t\theta'(0)}{\tfrac{1}{2}t^2} \quad \text{(if the limit exists)}.$$

If at least the first limit exists, we shall call θ a *first-order arc* emanating from the point $u = \theta(0)$. If both limits exist, θ is a *second-order arc* from $u = \theta(0)$. The notation (3.34) and (3.35) will also be used for real-valued functions of t, despite the one-sided nature of the limits.

PROPOSITION 3.4. *For any $u \in D$, the vector pairs (ω, ς) such that $g_u''(\omega, \varsigma) < \infty$ are the ones for which there exists a second-order arc $\theta \colon [0, \tau) \to D$ (in the sense of Definition 3.3) with*

$$(3.36) \qquad\qquad \theta(0) = u, \quad \dot{\theta}(0) = \omega, \quad \ddot{\theta}(0) = \varsigma.$$

For any such arc θ the function $\gamma(t) = g(\theta(t))$ satisfies

$$(3.37) \qquad\qquad \gamma(0) = g(u), \quad \dot{\gamma}(0) = g_u'(\omega), \quad \ddot{\gamma}(0) = g_u''(\omega, \varsigma).$$

PROOF. Any second-order arc $\theta \colon [0, \tau] \to \mathbb{R}^n$ satisfying (3.36) can be expressed by

$$(3.38) \qquad\qquad \theta(t) = u + t\omega + \tfrac{1}{2} t^2 \varsigma_t \quad \text{with } \lim_{t \downarrow 0} \varsigma_t = \varsigma,$$

namely by taking

$$(3.39) \qquad\qquad \varsigma_t = \frac{\theta(t) - \theta(0) - t\theta'(0)}{\tfrac{1}{2} t^2}.$$

The analysis in the proof of Proposition 3.2 shows that if $\theta(t) \in D$ for $t \in [0, \tau)$ then $\omega \in T_D(u)$ and $\varsigma \in T_D'(u, \omega)$ (the tangent cone to $T_D(u)$ at ω), which are the conditions in Proposition 3.2 for having $g_u''(\omega, \varsigma) < \infty$. Conversely, if ω and ς satisfy these conditions, then for t sufficiently small one has $\omega + \tfrac{1}{2} t \varsigma \in T_D(u)$ and then also, for still smaller t if necessary, one has $u + t(\omega + \tfrac{1}{2} t \varsigma) \in D$. Thus by taking $\theta(t) = u + t\omega + \tfrac{1}{2} t^2 \varsigma$ over a small enough interval $[0, \tau)$ one will have a second-order arc satisfying (3.36) with $\theta(t) \in D$. This proves our first assertion.

In turning to the second assertion we use the fact developed in the proof of Theorem 3.1 that (3.11) holds when $\omega \in \rho B$ and $t \in (0, \varepsilon/\rho)$, except we apply it to $\omega_t = \omega + \tfrac{1}{2} t \varsigma_t$ in place of ω, obtaining

$$[\gamma(t) - \gamma(0)]/t = g_j \cdot \omega_t + \tfrac{1}{2} t \omega_t \cdot Q_j \omega_t \quad \text{when } \omega_t \in T_{D_j}(u), \; j \in J_u.$$

The limit of this expression is $q_j \cdot \omega$ for any $j \in J_u$, which we know from the proof of Proposition 3.1 to be $g_u'(\omega)$. Thus $\dot{\gamma}(0) = g_u'(\omega)$.

The argument proceeds next with the formulas (3.27)–(3.28) in the proof of Proposition 3.2, which we apply to ς_t instead of ς. (The formula holds when t is sufficiently small.) This yields

$$\frac{\gamma(t) - \gamma(0) - t\dot{\gamma}(0)}{\tfrac{1}{2} t^2} = q_j \cdot \varsigma + \omega_t \cdot Q_j \omega_t \quad \text{when } \varsigma \in T_{D_j}'(u, \omega), \; j \in J_{u,\omega}.$$

The limit this time as $t \downarrow 0$ is

$$\ddot{\gamma}(0) = q_j \cdot \varsigma + \omega Q_j \omega \quad \text{for any } j \in J_{u,\omega} \text{ with } \varsigma \in T_{D_j}'(u, \omega)$$

and this expression has been shown in Proposition 3.2 to be $g_u''(\omega, \varsigma)$. $\square$

Proposition 3.4 tells us in particular that for piecewise linear-quadratic convex functions the parabolic second derivative in Definition 2.11 agrees with the second derivative of Ben-Tal and Zowe defined by (2.42): $g_u''(\omega, \varsigma) = \ddot{g}_u(\omega, \varsigma)$ (and also $g_u'(\omega) = \dot{g}_u(\omega)$). (But for more general functions they need not agree.)

The final result of this section reveals a duality between second-order epi-derivatives and parabolic second derivatives that will lead later (Theorem 4.7) to a deeper one for our general functions f.

PROPOSITION 3.5. *Let $u \in D$ and $\omega \in T_D(u)$, and define*

$$(3.40) \quad \begin{aligned} \partial g(u)_\omega &= \{y \in \partial g(u) | \omega \in N_{\partial g(u)}(y)\}, \\ &= \{y \in \partial g(u) | g'_u(\omega) = \omega \cdot y\}. \end{aligned}$$

Then

$$(3.41) \quad \sup_{y \in \partial g(u)_\omega} \{g''_{u,y}(\omega) + \varsigma \cdot y\} = g''_u(\omega, \varsigma) \quad \textit{for all } \varsigma \in \mathbf{R}^d,$$

and dually

$$(3.42) \quad \inf_{\varsigma \in \mathbf{R}^d} \{g''_u(\omega, \varsigma) - \varsigma \cdot y\} = \begin{cases} g''_{u,y}(\omega) & \textit{when } y \in \partial g(u)_\omega, \\ -\infty & \textit{otherwise.} \end{cases}$$

PROOF. Let

$$(3.43) \quad \varphi(y) = \begin{cases} -g''_{u,y}(\omega) & \text{for } y \in \partial g(u)_\omega, \\ -\infty & \text{otherwise.} \end{cases}$$

Our claim is that φ and the function $g''_u(\omega, \cdot)$ are conjugate to each other. Theorem 3.1 tells us in fact that $g''_{u,y}(\omega)$ is constant in $y \in \partial g(u)_\omega$, the constant having the value $\gamma_u(\omega)$ there, as defined by (3.7). This value is the same as $g''_u(\omega, 0)$ by (3.21) in Proposition 3.2. Thus

$$(3.44) \quad \varphi = -c + \delta_G \quad \text{where } c = g''_u(\omega, 0) \text{ and } G = \partial g(u)_\omega.$$

On the other hand the function $(g'_u)'_\omega$ in Proposition 3.2 is by Theorem 3.1 the support function of $\partial g'_u(\omega)$, and since g'_u is in turn by Theorem 3.1 the support function of $\partial g(u)$, the set $\partial g'_u(\omega)$ is just $\partial g(u)_\omega$. Thus (3.20) can be written as

$$(3.45) \quad g''_u(\omega, \cdot) = c + \delta_G^* \quad \text{where } c = g''_u(\omega, 0) \text{ and } G = \partial g(u)_\omega.$$

The conjugacy between (3.44) and (3.45) is apparent. $\square$

4. Constraint qualification and the general derivative formulas. We pass now to the setting of problem (P) itself, where $f(x) = g(F(x))$ for a $\mathcal{C}^2$ mapping $F: \mathbf{R}^n \to \mathbf{R}^d$ and a piecewise linear-quadratic proper convex function $g: \mathbf{R}^d \to \bar{\mathbf{R}}$. The effective domain of f is

$$(4.1) \quad C = \{x \in \mathbf{R}^n | F(x) \in D\},$$

where D is the effective domain of g and is a nonempty convex polyhedron in $\mathbf{R}^d$. The condition $F(x) \in D$ represents the constraint system in (P), so it will come as no surprise that a constraint qualification must be introduced before we can proceed.

The $d \times n$ matrix of first derivatives of F at x will be denoted by $\nabla F(x)$; the rows of $\nabla F(x)$ are thus the gradients of the components of the vector $F(x)$. For any $y \in \mathbf{R}^n$ one then has (In writing $y\nabla F(x)$ we think of y as a row vector.) For any first-order arc

$$(4.3) \quad \psi: [0, \tau] \to \mathbf{R}^n \quad \text{with } \psi(0) = x, \ \dot{\psi}(0) = \xi$$

in the sense of Definition 3.3, one has

$$(4.4) \quad \theta(0) = F(x) \quad \text{and} \quad \dot{\theta}(0) = \nabla F(x)\xi \quad \text{for } \theta: [0, \tau] \to F(\psi(t)).$$

The $d \times n \times n$ array of second derivatives of F at x will be denoted by $\nabla^2 F(x)$. This three-dimensional array can be viewed as a stack of $n \times n$ hessian matrices, one for each of the d components of the vector $F(x)$. In parallel with (4.2) we write

$$(4.5) \qquad \nabla^2(yF)(x) = y\nabla^2 F(x) \quad \text{for the function } yF \colon x \mapsto y \cdot F(x).$$

Thus $y\nabla^2 F(x)$ denotes for us the $n \times n$ matrix obtained by multiplying each of the Hessians in the stack $\nabla^2 F(x)$ by the corresponding scalar component of y and adding them up. On the other hand we introduce the notation $\xi\nabla^2 F(x)\xi$ to denote the vector in $\mathbb{R}^d$ obtained by multiplying each of the hessians both on the left and the right by ξ (as a row vector and as a column vector, respectively). Then

$$(4.6) \qquad y \cdot (\xi\nabla^2 F(x)\xi) = \xi \cdot y\nabla^2 F(x)\xi.$$

(Inasmuch as $y \in \mathbb{R}^d$ but $\xi \in \mathbb{R}^n$, this notation ought not to lead to any confusion.) When ψ in (4.3) is a second-order arc with $\ddot{\psi}(0) = \eta$, then θ in (4.4) is a second-order arc with

$$(4.7) \qquad \ddot{\theta}(0) = \xi\nabla^2 F(x)\xi + \nabla F(x)\eta.$$

DEFINITION 4.1. At a point $x \in C$ in (4.1) the *basic constraint qualification* will be said to hold if the only vector $y \in N_D(F(x))$ satisfying $y\nabla F(x) = 0$ is $y = 0$.

EXAMPLE 4.2. Let F and D have the form specified in Example 1.2, which corresponds to the constraint system

$$x \in X \text{ (polyhedron) and } f_i(x) \in I_i \text{ (closed interval)} \quad \text{for } i = 1, \ldots, m.$$

The basic constraint qualification at a point x satisfying this system is then the following: the only vector $y = (y_1, \ldots, y_m)$ satisfying

$$(4.8) \qquad y_i \in N_{I_i}(f_i(x)) \quad \text{for } i = 1, \ldots, m \text{ and } -\sum_{i=1}^{m} y_i \nabla f_i(x) \in N_X(x)$$

is $y = (0, \ldots, 0)$. To elucidate this further, let us note that if we write

$$(4.9) \qquad I_i = [c_i^-, c_i^+] \quad \text{for } i = 1, \ldots, m$$

(a slight abuse of notation when $c_i^+ = \infty$ or $c_i^- = -\infty$, since $I_i \subset \mathbb{R}$), we have

$$f_i(x) \in I_i \Leftrightarrow c_i^- \leq f_i(x) \leq c_i^+$$

and
(4.10)

$$N_{I_u}(f_i(x)) = \begin{cases} [0, \infty) & \text{if } f_i(x) = c_i^+ > c_i^- \text{ (active inequality constraint)}, \\ (-\infty, 0] & \text{if } f_i(x) = c_i^- < c_i^+ \text{ (active inequality constraint)}, \\ (-\infty, \infty) & \text{if } f_i(x) = c_i^- = c_i^+ \text{ (equality constraint)}, \\ [0, 0] & \text{if } c_i^- < f_i(x) < c_i^+ \text{ (inactive inequality constraint)}. \end{cases}$$

The first conditions in (4.8) are therefore just sign requirements on y in the classical mode. If $x \in \text{int}\, X$ (as when $X = \mathbb{R}^n$, for instance), the normal cone $N_X(x)$ reduces to the vector 0, and (4.8) turns into the well known dual statement of the Mangasarian-Fromovitz constraint qualification [40].

In Example 1.3, where the objective of Example 1.2 is replaced by a function of max type, the basic constraint qualification reduces to exactly the same condition

as just stated. In Examples 1.4–1.7 the basic constraint qualification trivializes and is *always* satisfied.

A second-order generalization of the classical theorem on the representation of tangent cones to sets defined by smooth constraints can now be stated. In proving this theorem we rely at a crucial point on an argument of Ben-Tal [41] concerning the existence of second-order arcs under the Mangasarian-Fromovitz constraint qualification.

THEOREM 4.3. *Suppose that the basic constraint qualification is satisfied at the point $x \in C$. Then the vectors ξ such that there exists a first-order arc*

$$(4.11) \qquad \psi \colon [0, \tau) \to C \quad with \ \psi(0) = x, \ \dot{\psi}(0) = \xi,$$

are precisely the ones such that

$$(4.12) \qquad \nabla F(x)\xi \in T_D(F(x)).$$

Moreover the vector pairs (ξ, η) such that there exists a second-order arc

$$(4.13) \qquad \psi \colon [0, \tau) \to C \quad with \ \psi(0) = x, \ \dot{\psi}(0) = \xi, \ \ddot{\psi}(0) = \eta,$$

are precisely the ones such that

$$(4.14) \quad \nabla F(x)\xi \in T_D(F(x)) \quad and \quad \xi\nabla^2 F(x)\xi + \nabla F(x)\eta \in T_D'(F(x), \nabla F(x)\xi),$$

where $T_D'(F(x), \nabla F(x)\xi)$ denotes the tangent cone to the cone $T_D(F(x))$ at $\nabla F(x)\xi$.

PROOF. An arc $\psi \colon [0, \tau) \to C$ has associated with it an arc $\theta \colon [0, \tau) \to D$, where $\theta(t) = F(\psi(t))$. From the observations preceding Definition 4.1, we know that

$$(4.15) \qquad \theta(0) = F(x), \quad \dot{\theta}(0) = F(x)\xi, \quad \ddot{\theta}(0) = \xi\nabla F(x)\xi + \nabla F(x)\eta.$$

Propositions 3.2 and 3.4 convey the information that an arc $\theta \colon [0, \tau) \to D$ has $\dot{\theta}(0) \in T_D(\theta(0))$ and $\ddot{\theta}(0) \in T_D'(\theta(0), \dot{\theta}(0))$. The necessity of (4.12) in the first-order case is therefore clear, regardless of any constraint qualification, and likewise for (4.14) in the second-order case.

The sufficiency of (4.12) and (4.14) in the respective cases will be established by converting our situation to the classical one. where our basic constraint qualification can be identified with the Mangasarian-Fromovitz constraint qualification. Let $u = F(x)$. The set D being a convex polyhedron. it coincides in some neighborhood u with the set $u + T_D(u)$. The polyhedral cone $T_D(u)$ can be represented in the form

$$(4.16) \quad T_D(u) = \{\omega \mid a_i \cdot \omega \leq 0 \text{ for } i = 1, \ldots, q; \ a_i \cdot \omega = 0 \text{ for } i = q+1, \ldots, r\}$$

for some choice of vectors $a_i \in \mathbb{R}^d$, and then from some $\varepsilon > 0$ we will have

$$D \cap (u + \varepsilon B) = \{u' \mid a_i \cdot u' \leq \alpha_i \text{ for } i = 1, \ldots, q; \ a_i \cdot u' = \alpha_i \text{ for } i = q+1, \ldots, r\},$$

where $\alpha_i = a_i \cdot u$. This implies that for some $\delta > 0$ we will have

$$(4.17)$$
$$C \cap (x + \delta B) = \{x' \mid f_i(x') \leq 0 \text{ for } i = 1, \ldots, q; \ f_i(x') = 0 \text{ for } i = q+1, \ldots, r\},$$

where

$$(4.18) \qquad f_i(x') = a_i \cdot F(x') - \alpha_i \quad \text{for } i = 1, \ldots, r.$$

In selecting the representation (4.16), it is always possible to choose the a_i's in such a way that

(4.19)   the only $\lambda = (\lambda_1, \ldots, \lambda_r)$ with $\lambda_i \geq 0$ for $i = 1, \ldots, q$ and $\sum_{i=1}^r \lambda_i a_i$
           $= 0$ is $\lambda = (0, \ldots, 0)$.

To see this, let $M = N_D(u) \cap (-N_D(u))$ (a subspace of $\mathbb{R}^d$) and let $K = N_D(u) \cap M^\perp$ (a polyhedral cone with $K \cap (-K) = \{0\}$). Then

(4.20) $N_D(u) = K + M$ and $T_D(u) = K^\circ \cap M^\perp$,

where K° is the polar of K. Let $a_1, \ldots, a_q$ be nonzero vectors that generate K as the set of all their nonnegative linear combinations, and let $a_{q+1}, \ldots, a_r$ be a basis for M. Then (4.19) holds as desired, and the representation (4.16) is obtained from the second part of (4.18).

With this refinement we have in (4.17)–(4.18) a local constraint representation of C around x in terms of $\mathcal{C}^2$ constraint functions f_i that are all active at x and satisfy, we claim, the dual statement of the Mangasarian-Fromovitz constraint qualification:

(4.21)   the only vector $\lambda = (\lambda_1, \ldots, \lambda_r)$ satisfying $\lambda_i \geq 0$ for $i = 1, \ldots, q$
           and $\sum_{i=1}^r \lambda_i \nabla f_i(x) = 0$ is $\lambda = (0, \ldots, 0)$.

Indeed, the gradients in (4.18) are

(4.22) $\nabla f_i(x) = a_i \nabla F(x)$ for $i = 1, \ldots, r$

so that if λ were a vector satisfying the conditions in (4.19) the vector $y = \sum_{i=1}^r \lambda_i a_i$ would belong to $N_D(x)$ (because $a_i \in N_D(x)$ for $i = 1, \ldots, q$ and $\pm a_i \in N_D(x)$ for $i = q+1, \ldots, r$) and give us

$$y \nabla F(x) = \sum_{i=1}^r \lambda_i a_i \nabla F(x) = \sum_{i=1}^r \lambda_i \nabla f_i(x) = 0.$$

Then $y = 0$ by our basic constraint qualification, and this implies $\lambda_i = 0$ for $i = 1, \ldots, r$ by (4.19).

The original statement of the Mangasarian-Fromovitz constraint qualification, equivalent to the property we have just derived, is that the gradients $\nabla f_i(x)$ for $i = q+1, \ldots, r$ are linearly independent and there exists at least one vector $\bar{\xi}$ with
(4.23)
$\nabla f_i(x) \cdot \bar{\xi} < 0$ for $i = 1, \ldots, q$ and $\nabla f_i(x) \cdot \bar{\xi} = 0$ for $i = q+1, \ldots, r$.

It is well known (see [**40**]) that when this holds there exists for every vector ξ satisfying
(4.24)
$\nabla f_i(x) \cdot \xi \leq 0$ for $i = 1, \ldots, q$ and $\nabla f_i(x) \cdot \xi = 0$ for $i = q+1, \ldots, r$,

a first-order arc

(4.25) $\psi : [0, \tau_0) \to \mathbb{R}^n$ with $\psi(0) = x$, $\dot{\psi}(0) = \xi$,

(4.26) $f_i(\psi(t)) \leq 0$ for $i = 1, \ldots, q$, $f_i(\psi(t)) = 0$ for $i = q+1, \ldots, r$.

In view of (4.15) we have $\psi(t) \in C$ for t sufficiently small, say for $t \in [0, \tau]$. In our local representation the condition $\nabla F(x)\xi \in T_D(F(x))$ is equivalent to

$a_i \cdot \nabla F(x)\xi \leq 0$ for $i = 1, \ldots, q$,   and $a_i \cdot \nabla F(x)\xi = 0$ for $i = q+1, \ldots, r$,

which is the same as (4.24) by (4.22). It follows therefore that if (4.12) is satisfied, an arc ψ as in (4.11) does exist. Sufficiency in the first-order case is thereby demonstrated.

For the second-order case we make use of the fact established by Ben-Tal [**41**, proof of Theorem 3.1]: if ξ and ς satisfy (4.24) and

$$(4.27) \quad \xi \cdot \nabla^2 f_i(x)\xi + \nabla f_i(x)\eta \begin{cases} < 0 & \text{for all } i = 1,\ldots,q \text{ having } \nabla f_i(x) \cdot \xi < 0, \\ = 0 & \text{for } i = q+1,\ldots,r, \end{cases}$$

and if the gradients $\nabla f_i(x)$ for $i = q+1\ldots,r$ are linearly independent, then there exists a second-order arc

$$(4.28) \quad \psi\colon [0,\tau_0) \to \mathbf{R}^n \quad \text{with } \psi(0) = x, \ \dot\psi(0) = \xi, \ \ddot\psi(0) = \eta$$

and such that (4.26) holds. (The argument of Ben-Tal essentially uses the implicit function theorem to change the coordinate system and convert the equality constraints into simple linear constraints where everything is easy.) In preparation for applying this fact we need to see what condition (4.14) means in our local representation. The first part of (4.14) has already been translated into (4.24). Observing from (4.16) that the tangent cone $T'_D(u,\omega)$ to $T_D(u)$ at u is expressed by

$$T'_D(u,\omega) = \{\varsigma | a_i \cdot \varsigma \le 0 \text{ for } i = 1,\ldots,q \text{ having } a_i \cdot \omega = 0,$$
$$\text{and } a_i \cdot \varsigma = 0 \text{ for } i = q+1,\ldots,r\},$$

we see that the second condition in (4.14) is

$$a_i \cdot [\xi\nabla^2 F(x)\xi + \nabla F(x)\eta] \begin{cases} \le 0 & \text{for } i = 1,\ldots,q \text{ with } a_i \cdot \nabla F(x)\xi \le 0, \\ = 0 & \text{for } i = q+1,\ldots,r. \end{cases}$$

But

$$a_i \cdot [\xi\nabla^2 F(x)\xi + \nabla F(x)\eta] = \xi \cdot \nabla^2(a_iF)(x)\xi + \nabla(a_iF)(x)\eta$$
$$= \xi \cdot \nabla^2 f_i(x)\xi + \nabla f_i(x)\xi.$$

Therefore (4.14) is equivalent to the combination of (4.24) and

$$(4.29) \quad \xi \cdot \nabla f_i(x)\xi + \nabla f_i(x)\eta \begin{cases} \le 0 & \text{for } i = 1,\ldots,q \text{ with } \nabla f_i(x)\xi = 0, \\ = 0 & \text{for } i = q+1,\ldots,r. \end{cases}$$

What we must do to establish the sufficiency in the second-order part of the theorem is to show that for any pair (ξ,η) satisfying (4.24) and (4.29) there is an arc ψ satisfying (4.28) and (4.26).

Consider therefore a pair (ξ,η) satisfying (4.24) and (4.29). For $k = 1,2,\ldots$ let $\eta k = \eta + (1/k)\bar\xi$, where $\bar\xi$ is a vector satisfying (4.23), which we have shown to exist because of the constraint qualification. The pair (ξ,η_k) then satisfies (4.24) and (4.27), so by the cited result of Ben-Tal there exists

$$\psi_k\colon [0,\tau_k) \to C \quad \text{with } \psi_k(0) = x, \ \dot\psi_k(0) = \xi, \ \ddot\psi_k(0) = \eta_k.$$

Then for all k

$$\lim_{t\downarrow 0} \frac{\psi_k(t) - x}{t} = \xi \quad \text{and} \quad \lim_{t\downarrow 0} \frac{\psi_k(t) - x - t\xi}{\frac{1}{2}t^2} = \eta_k.$$

Lowering the values τ_k consecutively if necessary, we can arrange that $\tau_1 > \tau_2 > \cdots > 0$ and $[\psi_k(t) - x]/t$ differs from ξ by at most $1/k$ on $(0, \tau_k)$, $[\psi_k(t) - x - t\xi]/\frac{1}{2}t^2$ differs from η_k by at most $1/k$ on $(0, \tau_k)$. Define

$$\psi(t) = \psi_k(t) \text{ on } [\tau_{k+1}, \tau_k) \quad \text{for } k = 1, 2, \ldots \text{ and } \psi(0) = x.$$

Then $\psi(t) \in D$ for all $t \in [0, \tau_1)$, and one has

$$\lim_{t \downarrow 0} \frac{\psi(t) - x}{t} = \xi,$$

$$\lim_{t \downarrow 0} \frac{\psi(t) - x - t\xi}{\frac{1}{2}t^2} = \lim_{k \to \infty} \eta_k = \eta.$$

Thus ψ is a second-order arc satisfying (4.13), which is what we had to come up with. □

The proof of Theorem 4.3 reveals another fact about first-order properties of the set C. Recall that C is said to be *tangentially regular* if the Clarke tangent cone $T_C(x)$ coincides with the contingent cone $K_C(x)$ to C at x (see Clarke [2, p. 55]). This is a property of considerable interest in nonsmooth analysis, because it is known to imply for instance that

$$K_C(x) = \lim_{\substack{x' \to x \\ x' \in C}} \inf K_C(x').$$

PROPOSITION 4.4. *Suppose that the basic constraint qualification is satisfied at the point $x \in C$. Then C is tangentially regular at x with its Clarke tangent cone expressed by*

$$(4.30) \qquad T_C(x) = \{\xi \in \mathbf{R}^n | \nabla F(x)\xi \in T_D(F(x))\}.$$

PROOF. As in the proof of Theorem 4.3, give C a local representation at x of the form (4.17)-(4.18) in such a way that the Mangasarian-Fromovitz constraint qualification is satisfied and the set on the right side of (4.30) is characterized by the linear system (4.24). Use the linear independence of the equality constraint gradients $\nabla f_i(x)$, $r = q + 1, \ldots, m$, to pass by way of the implicit function theorem to a lower dimensional format in which C is identified with a set represented by inequality constraints only, still with the Mangasarian-Fromovitz constraint qualification satisfied. For this set one has tangential regularity, as proved by Clarke [2, Corollary 2 to Theorem 2.4.7]. The tangential regularity of C then follows, the cone $T_C(x)$ being given by (4.24) and therefore by (4.33). □

The next theorem is our main result on epi-derivatives of first and second order. It demonstrates their existence for a large and significant class of functions.

THEOREM 4.5. *Let x be a point of $C = \text{dom } f$ where the basic constraint qualification is satisfied. Then f is twice epi-differentiable at x and also subdifferentially regular at x in the sense of Clarke. The first epi-derivative function is given by*

$$(4.31) \qquad f'_x(\xi) = g'_{F(x)}(\nabla F(x)\xi),$$

$$(4.32) \qquad \text{dom } f'_x = \{\xi | \nabla F(x)\xi \in T_D(F(x))\} = T_C(x).$$

It is the support function of the generalized subgradient set

$$(4.33) \qquad \partial f(x) = \partial g(F(x))\nabla F(x) = \{y\nabla F(x) | y \in \partial g(F(x))\},$$

which is therefore the same as the set of all epi-gradients of f at x; this set is a nonempty convex polyhedron. The second epi-derivative function relative to any $v \in \partial f(x)$ is given by

$$(4.34) \quad f''_{x,v}(\xi) = \begin{cases} \gamma_{F(x)}(\nabla F(x)\xi) + \max_{y \in Y_v(x)} y \cdot [\xi \nabla^2 F(x)\xi] & \text{if } \xi \in \Xi_v(x), \\ \infty & \text{for } \xi \notin \Xi_v(x), \end{cases}$$

$$(4.35) \qquad \qquad \qquad \operatorname{dom} f''_{x,v} = \Xi_v(x),$$

where

$$(4.36) \qquad \qquad Y_v(x) := \{y \in \partial g(F(x)) \,|\, y \nabla F(x) = v\}$$

is a nonempty. bounded, polyhedral convex set,

$$(4.37) \qquad \begin{aligned} \Xi_v(x) &:= N_{\partial f(x)}(v) = \{\xi \,|\, f'_x(\xi) = v \cdot \xi\} \\ &= \{\xi \,|\, g'_{F(x)}(\nabla F(x)\xi) = v \cdot \xi\} \end{aligned}$$

is a polyhedral convex cone, and $\gamma_{F(x)}(\nabla F(x)\xi)$ is the expression defined from g by (3.7).

PROOF. Our first task is to show that the difference quotient functions

$$(4.38) \qquad \varphi_{x,t}(\xi) = [f(x + t\xi) - f(x)]/t = [g(F(x + t\xi)) - g(F(x))]/t$$

epi-converge as $t \downarrow 0$ to the function given by the right side of (4.31). or in other words to show that

$$(4.39) \qquad \qquad \liminf_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \varphi_{x,t}(\xi') \geq g'_{F(x)}(\nabla F(x)\xi),$$

$$(4.40) \qquad \qquad \limsup_{t \downarrow 0} \inf_{\xi' \to \xi} \varphi_{x,t}(\xi') \leq g'_{F(x)}(\nabla F(\xi)).$$

Let $u = F(x)$. We can write

$$\varphi_{x,t}(\xi') = [g(u + t\omega') - g(u)]/t \quad \text{for } \omega' = [F(x + t\xi') - F(x)]/t$$

and note that $\omega' \to \nabla F(x)\xi$ when $\xi' \to \xi$. This makes it clear that

$$\liminf_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \varphi_{x,t}(\xi') \geq \liminf_{\substack{t \downarrow 0 \\ \omega' \to \nabla F(x)\xi}} \frac{g(u + t\omega') - g(u)}{t}.$$

Since g is known to be epi-differentiable at u (Theorem 3.1), the limit on the right is $g'_u(\nabla F(x)\xi)$. This proves (4.39).

To obtain the complementary inequality (4.40), it is enough to consider a ξ with $g'_u(\nabla F(x)\xi) < \infty$, i.e. ξ belonging to the set on the right side of (4.32). For any such ξ there exists by Theorem 4.3 a first-order arc ψ as in (4.11). Then

$$(4.41) \qquad \qquad \limsup_{t \downarrow 0} \inf_{\xi' \to \xi} \varphi_{x,t}(\xi') \leq \limsup_{t \downarrow 0} \varphi_{x,t}(\xi_t)$$

where $\xi_t = [\psi(t) - \psi(0)]/t$ so that

$$\varphi_{x,t}(\xi_t) = [f(\psi(t)) - f(\psi(0))]/t = [g(\theta(t)) - g(\theta(0))]/t$$

for $\theta(t) = F(\psi(t))$, $\theta(t) \in C$ for $t \in [0, \tau)$. Then

$$\lim_{t \downarrow 0} \varphi_{x,t}(\xi_t) = g'_{\theta(0)}(\dot{\theta}(0)) = g'_{F(x)}(\nabla F(x)\xi)$$

by Proposition 3.4. This and (4.41) yield (4.40) and finish the justification of (4.31) (and therefore (4.32)).

The subdifferential regularity of f at x will be tackled next. The function g, being piecewise linear-quadratic, is locally Lipschitzian relative to its effective domain C. Let $\mu > 0$ be a Lipschitz constant that works for a neighborhood of $u = F(x)$, and define

$$(4.42) \qquad \tilde{g}(u') = \inf_{w \in \mathbf{R}^n} \{g(w) + \tilde{\mu}|u' - w|\}. \quad \text{where } \tilde{\mu} > \mu.$$

This formula says that $\tilde{g}$ is the infimal convolute of g and $\tilde{\mu}|\cdot|$; since g is convex, so is $\tilde{g}$ [**22**, Theorem 5.4]. Because of the choice of $\tilde{\mu}$ one will have

$$(4.43) \qquad \tilde{g}(u') = g(u') \quad \text{for all } u' \text{ in some neighborhood of } u$$

(namely any neighborhood where μ acts as a Lipschitz constant). In particular $\tilde{g}$ is finite at certain points, but also $\tilde{g}(u') \leq g(u) + \tilde{\mu}|u' - u| < \infty$ for all u' from (4.42). so by convexity $\tilde{g}$ is finite everywhere on $\mathbf{R}^d$ [**22**, Theorem 7.4]. Hence $\tilde{g}$ is locally Lipschitzian [**22**, Theoerm 10.4] and everywhere subdifferentially regular (Clarke [**39**]).

In view of (4.43) we have

$$(4.44) \qquad f(x') = \tilde{g}(F(x')) + \delta_C(x') \quad \text{for all } x' \text{ near } x.$$

The function $\tilde{f}(x') = \tilde{g}(F(x'))$ is subdifferentially regular, because composition of a subdifferentially regular locally Lipschitzian function with a smooth mapping preserves subdifferential regularity (Clarke [**39**]). The set C is tangentially regular at x by Proposition 4.4. The indicator function δ_C is thus subdifferentially regular at x. Applying Rockafellar [**42**, Corollary 2 of Theorem 2] we are able to conclude that the sum function in (4.42) is subdifferentially regular at x and consequentially that f has this asserted property.

Because f is subdifferentially regular, f'_x is the support function of the set $\partial f(x)$ (Proposition 2.6). But $g'_{F(x)}$ is the support function of $\partial g(F(x))$ (Theorem 3.1). so formula (4.31) says

(4.45)
$$f'_x(\xi) = \sup_{y \in \partial g(F(x))} y \cdot \nabla F(x)\xi = \sup_{y \in \partial g(F(x))} (y\nabla F(x)) \cdot \xi = \sup_{v \in \partial g(F(x))\nabla F(x)} v \cdot \xi.$$

Thus f'_x is the support function $\partial g(F(x))\nabla F(x)$. The latter is nonempty convex polyhedron. because it is the image under the linear transformation $y \mapsto y\nabla F(x)$ of the set $\partial g(F(x))$, which is a nonempty convex polyhedron by Theorem 3.1. (The image of a convex polyhedron under a linear transformation is a convex polyhedron [**22**, Theorem 19.3].) In particular $\partial g(F(x))\nabla F(x)$ is a closed convex set. The correspondence between closed convex sets and their support functions is one-to-one, so the fact that f'_x is the support function of both $\partial f(x)$ and $\partial g(F(x))\nabla F(x)$ implies these two sets are the same, as claimed in (4.33).

We pass now to the second-order formula in the theorem, where ξ is fixed as well as x. Let $\omega = \nabla F(x)\xi$. Select $v \in \partial f(x) = \partial g(F(x))\nabla F(x)$ and consider $Y_v(x)$ as

in (4.36). It must be shown that the difference quotient functions

$$(4.46) \quad \varphi_{x,v,t}(\xi) = \frac{f(x+t\xi) - f(x) - t\xi \cdot v}{\frac{1}{2}t^2} = \frac{g(F(x+t\xi)) - g(F(x)) - tv \cdot \xi}{\frac{1}{2}t^2}$$

epi-converge at $t \downarrow 0$ to the function given by the right side of (4.34). For this it will suffice by the formulas for $g''_{u,y}(\omega)$ in Theorem 3.1 that we prove two facts: first that

$$(4.47) \quad \forall y \in Y_v(x), \quad \lim_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \inf \varphi_{x,v,t}(\xi') \geq g''_{u,y}(\omega) + y \cdot [\xi \nabla^2 F(x)\xi];$$

and second that for any ξ in the set $\Xi_v(x)$,

$$(4.48) \quad \exists \bar{y} \in Y_v(x), \quad \limsup_{t \downarrow 0} \inf_{\xi' \to \xi} \varphi_{x,v,t}(\xi') \leq g''_{u,\bar{y}}(\omega) + \bar{y} \cdot [\xi \nabla^2 F(x)\xi] < \infty.$$

The inclusion $\subset$ in (4.35) is already known to hold from Proposition 2.8, so, this second property will also serve to establish the equality in (4.35). (For ξ such that $f'_x(\xi) > \xi \cdot v$ one has for all $y \in Y_v$ that $g'_{F(x)}(\nabla F(x)\xi) > y \cdot \nabla F(x)\xi$ and consequently $g''_{F(x),y}(\nabla F(x)\xi) = \infty$ by Theorem 3.1. Then the maximum in (4.34) is attained trivially at every $y \in Y_v(x)$.)

We begin with the verification of (4.47). For any $y \in Y_v(x)$ we can write $y\nabla F(x)$ in place of v in (4.46). Then if we take $\omega' = [F(x+t\xi') - F(x)]/t$, which converges to $\omega = \nabla F(x)\xi$ as $t \downarrow 0$ and $\xi' \to \xi$, we have

$$\varphi_{x,v,t}(\xi') = \frac{g(u+t\omega') - g(u) - ty \cdot \omega'}{\frac{1}{2}t^2}$$
$$+ \frac{(yF)(x+t\xi') - (yF)(x) - t\nabla(yF)(x) \cdot \xi'}{\frac{1}{2}t^2},$$

where the second term converges as $t \downarrow 0$ and $\xi' \to \xi$ to

$$\xi \cdot \nabla^2(yF)(x)\xi = y \cdot [\xi \nabla^2 F(x)\xi].$$

It follows that

$$\lim_{\substack{t \downarrow 0 \\ \xi' \to \xi}} \inf \varphi_{x,v,t}(\xi') \geq \lim_{\substack{t \downarrow 0 \\ \omega' \to \omega}} \inf \frac{g(u+t\omega') - g(u) - ty \cdot \omega'}{\frac{1}{2}t^2} + y \cdot [\xi \nabla^2 F(x)\xi],$$

where the "lim inf" on the right is $g''_{u,y}(\omega)$ since $g'_{u,y}$ exists by Theorem 3.1. Thus (4.46) is true.

The argument that (4.48) holds, under the assumption that $f'_x(\xi) = v \cdot \xi$, is more involved. Let us observe that the set $Y_v(x)$ is a (nonempty) polyhedron by its definition in (4.36), inasmuch as the set $\partial g(F(x)) = \partial g(u)$ is a polyhedron (Theorem 3.1). It is bounded (hence compact), because if it were not it would have a recession vector $\eta \neq 0$:

$$(4.49) \quad y + t\eta \in Y_v(x) \quad \text{for any } y \in Y_v(x) \text{ and } t \geq 0$$

[22, Theorem 8.4]. Such a vector η would belong to the recession cone of $\partial g(u)$ and satisfy $\eta \nabla F(x) = 0$. To say that η belongs to the recession cone of $\partial g(u)$ means that for any $y \in \partial g(u)$ one has $y + t\eta \in \partial g(u)$ for all $t \geq 0$, so that

$$g(u') \geq g(u) + (y + t\eta) \cdot (u' - u) \quad \text{for all } t \geq 0, \ u' \in \mathbf{R}^d.$$

Then in particular one must have $\eta \cdot (u' - u) \leq 0$ for all $u' \in D = \operatorname{dom} g$, so that $\eta \in N_D(u)$. Thus if $Y_v(x)$ were not bounded there would exist a nonzero vector $\eta \in N_D(F(x))$ such that $\eta \nabla F(x) = 0$, which would be contrary to the basic constraint qualification.

For any $y \in Y_v(x)$ the assumption $\xi \in \Xi_v(x)$ implies $f'_x(\xi) = y \cdot \nabla F(x)\xi = y \cdot \omega$. The supremum in (4.45), which equals $g'_u(\omega)$, must therefore be attained at y: one has $g'_u(\omega) = y \cdot \omega$ and in consequence y belongs to the set $\partial g(u)_\omega$ defined in (3.40), which is a convex polyhedron included in $\partial g(u)$. This allows us to write

$$(4.50) \qquad Y_v(x) = \{y \in \partial g(u)_\omega \,|\, y\nabla F(x) = v\}.$$

For each $y \in \partial g(u)_\omega$ we have $g''_{u,y}(\omega) < \infty$ by formula (3.5) in Theorem 3.1; in fact the function $y \mapsto g''_{u,y}(\omega)$ is constant on $\partial g(u)_\omega$ with the constant being given the value $\gamma_u(\omega)$ defined by (3.7). Denote this constant for the moment by γ and let $\sigma = \xi \nabla^2 F(x)\xi$. With (4.50) being the same as (4.36) we can express the right side of (4.34) as

$$(4.51) \qquad \sup_{\substack{y \in G \\ y\nabla F(x)=v}} \{\gamma + y \cdot \sigma\}, \quad \text{where } G = \partial g(u)_\omega.$$

This maximum can be viewed as the optimal value in a certain linear programming problem, since $\partial g(u)_\omega$ is polyhedral. Hence it is attained by some $\bar{y}$. By duality theory (either linear programming duality in terms of a constraint representation for G or the more general polyhedral convex programming duality in [22, Theorem 29.2]) there is a Kuhn-Tucker multiplier vector for the constraint $y\nabla F(x) = v$ in (4.51), i.e. a vector η such that

$$(4.52) \qquad \sup_{\substack{y \in G \\ y\nabla F(x)=v}} \{\gamma + y \cdot \sigma\} = \sup_{y \in G}\{\gamma + y \cdot \sigma + [y\nabla F(x) - v] \cdot \eta\}.$$

The left side of (4.52) is attained at the vector $\bar{y} \in Y_v(x)$ and equals

$$(4.53) \qquad \gamma + \bar{y} \cdot \sigma = \gamma_u(\omega) + \bar{y} \cdot [\xi\nabla^2 F(x)\xi] < \infty.$$

The right side can be worked out from formula (3.41) of Proposition 3.5 as
$$(4.54)$$
$$\sup_{y \in \partial g(u)_\omega} \{g''_{u,y}(\omega) + y \cdot [\xi\nabla^2 F(x)\xi + \nabla F(x)\eta] - v \cdot \eta\} = g''_u(\omega, \varsigma) - v \cdot \eta$$

$$\text{for } \varsigma = \xi\nabla^2 F(x)\xi + \nabla F(x)\eta.$$

Because of the equality between (4.53) and (4.54), our task in demonstrating (4.48) is reduced now to the verification that

$$(4.55) \qquad \limsup_{t \downarrow 0} \inf_{\xi' \to \xi} \varphi_{x,v,t}(\xi') \leq g''_u(\omega, \varsigma) - v \cdot \eta$$

for $u = F(x)$, $\omega = \nabla F(x)\xi$, $\varsigma = \xi\nabla^2 F(x)\xi + \nabla F(x)\eta$, when $g''_u(\omega, \varsigma) < \infty$.

The condition $g''_u(\omega, \varsigma) < \infty$ has been shown in Proposition 3.2 to mean $\omega \in T_D(u)$, $\varsigma \in T'_D(u, \omega)$. Applying Theorem 4.3 we get the existence of a second-order arc ψ as in (4.13) whose corresponding arc $\theta : [0, \tau] \to D$ given by $\theta(t) = F(\psi(t))$ satisfies

$$\theta(0) = u, \quad \dot{\theta}(0) = \omega, \quad \ddot{\theta}(0) = \varsigma.$$

Then for $\dot{\gamma}(t) = f(\psi(t)) = g(F(\psi(t)))$ we have by Proposition 3.4 that $\gamma(0) = g(u) = f(x)$,

$$\dot{\gamma}(0) = g'_u(\omega) = g'_{F(x)}(F(x)\xi) = f'_x(\xi) = \xi \cdot v,$$

and $\ddot{\gamma}(0) = g''_u(\omega, \varsigma)$. It follows that

$$g''_u(\omega, \varsigma) = \lim_{t \downarrow 0} \frac{f(\psi(t)) - f(x) - t\xi \cdot v}{\frac{1}{2}t^2}.$$

This in turn yields in terms of $\xi_t = [\psi(t) - \psi(0)]/t$ that

$$g''_u(\omega, \varsigma) = \lim_{t \downarrow 0} \left(\frac{f(x + t\xi_t) - f(x) - t\xi_t \cdot v}{\frac{1}{2}t^2} + \frac{(\xi_t - \xi) \cdot v}{\frac{1}{2}t^2} \right)$$

where

$$\lim_{t \downarrow 0} \frac{\xi_t - \xi}{\frac{1}{2}t^2} = \lim_{t \downarrow 0} \frac{\psi(t) - \psi(0) - t\dot{\psi}(t)}{\frac{1}{2}t^2} = \ddot{\psi}(0) = \eta.$$

Thus

$$g''_u(\omega, \eta) - v \cdot \eta = \lim_{t \downarrow 0} \varphi_{x,v,t}(\xi_t), \quad \text{where} \lim_{t \downarrow 0} \xi_t = \xi.$$

This gives us 4.55 and finishes the proof of the theorem. □

Some complementary results will now be obtained for the parabolic second-derivative in Definition 2.12.

THEOREM 4.6. *At any point $x \in C = \mathrm{dom}\, f$ where the basic constraint qualification is satisfied and for any $\xi \in T_C(x) = \mathrm{dom}\, f'_x$ (cf. (4.32)), the parabolic second-derivative function $f''_x(\xi, \cdot)$ is well defined and given by*

$$(4.56) \qquad f''_x(\xi, \eta) = g''_{F(x)}(\nabla F(x)\xi, \; \xi\nabla^2 F(x)\xi + \nabla F(x)\eta) \quad \text{for all } \eta.$$

One has $f''_x(\xi, \eta) < \infty$ if and only if there exists a second-order arc

$$(4.57) \qquad \psi: [0, \tau) \to C \quad \text{with } \psi(0) = x, \; \dot{\psi}(0) = \xi, \; \ddot{\psi}(0) = \eta$$

(cf. Theorem 4.3), in which case the function $\gamma(t) = f(\psi(t))$ satisfies

$$(4.58) \qquad \gamma(0) = f(x), \quad \dot{\gamma}(0) = f'_x(\xi), \quad \ddot{\gamma}(0) = f''_x(\xi, \eta).$$

Moreover $f''_x(\xi, \eta)$ as a function of η is proper, convex and piecewise linear.

PROOF. Let

$$(4.59) \qquad \varphi_{x,\xi,t}(\eta) = \frac{f(x + t\xi + \frac{1}{2}t^2\eta) - f(x) - tf'_x(\xi)}{\frac{1}{2}t^2}.$$

The first question is whether $\varphi_{x,\xi,t}$ epi-converges to something as $t \downarrow 0$. In terms of $u = F(x)$, $\omega = \nabla F(x)\xi$ and $\varsigma = \xi\nabla^2 F(x)\xi + \nabla F(x)\eta$, we have $f'_x(\xi) = g'_u(\omega)$ by Theorem 4.5 and can express

$$(4.60) \qquad \varphi_{x,\psi,t}(\eta) = \frac{g(u + t\omega + \frac{1}{2}t^2\varsigma_t(\eta)) - g(u) - tg'_u(\omega)}{\frac{1}{2}t^2}.$$

with

$$(4.61) \qquad \varsigma_t(\eta) = \frac{F(x + t\xi + \frac{1}{2}t^2\eta) - F(x) - t\nabla F(x)\xi}{\frac{1}{2}t^2},$$

$$(4.62) \qquad \lim_{\substack{t \downarrow 0 \\ \eta' \to \eta}} \varsigma_t(\eta') = \varsigma.$$

Therefore

$$\liminf_{\substack{t \downarrow 0 \\ \eta' \to \eta}} \varphi_{x,\xi,t}(\eta') \geq \liminf_{\substack{t \downarrow 0 \\ \varsigma' \to \varsigma}} \frac{g(u + t\omega + \frac{1}{2}t^2\varsigma') - g(u) - tg'_u(\omega)}{\frac{1}{2}t^2}$$

where the limit on the right is $g''_u(\omega, \varsigma)$ because $g''_u(\omega, \cdot)$ exists by Proposition 3.2.

The complementary inequality

$$(4.63) \qquad \limsup_{t \downarrow 0} \inf_{\eta' \to \eta} \varphi_{x,\xi,t}(\eta') \leq g''_u(\omega, \varsigma)$$

must be argued next. Only the case where $g''_u(\omega, \varsigma) < \infty$ needs to be addressed. In this case $\omega \in T_D(u)$ and $\varsigma \in T_D(u, \omega)$ by Proposition 3.2, so the condition in Theorem 4.3 is satisfied and there exists an arc ψ as in (4.57). Then $\theta(t) = F(\psi(t))$ is a second-order arc

$$\theta : [0, \tau) \to D \quad \text{with } \theta(0) = u, \ \dot{\theta}(0) = \omega, \ \ddot{\theta}(0) = \varsigma$$

and $\gamma(t) = g(\theta(t))$. Proposition 3.4 asserts that

$$(4.64) \qquad \gamma(0) = g(u), \quad \dot{\gamma}(0) = g'_u(\omega), \quad \ddot{\gamma}(0) = g''_u(\omega, \varsigma).$$

If we substitute

$$\eta_t = \frac{\psi(t) - \psi(0) - t\dot{\psi}(0)}{\frac{1}{2}t^2} = \frac{\psi(t) - x - t\xi}{\frac{1}{2}t^2}$$

for η in (4.60)-(4.61) we get

$$\varphi_{x,\xi,t}(\eta_t) = \frac{\gamma(t) - \gamma(0) - t\dot{\gamma}(0)}{\frac{1}{2}t^2}.$$

It follows that

$$\limsup_{t \downarrow 0} \inf_{\eta' \to \eta} \varphi_{x,\xi,t}(\eta') \leq \limsup_{t \downarrow 0} \varphi_{x,\xi,t}(\eta_t)$$

$$= \lim_{t \downarrow 0} \frac{\gamma(t) - \gamma(0) - t\dot{\gamma}(0)}{\frac{1}{2}t^2} = \ddot{\gamma}(0).$$

This gives (4.63) by the last equation in (4.64) and brings us to the point where only the final assertion of the theorem is left to prove. In fact $g''_u(\omega, \cdot)$ is known from Proposition 3.2 to be a proper convex, piecewise linear function with effective domain $T'_D(u, \omega)$, so this assertion is immediate from (4.56). $\square$

A fundamental duality between second-order epi-derivatives and parabolic second-order derivatives exists for functions f of the class we have been treating, as can now be stated.

THEOREM 4.7. *At any point $x \in C = \mathrm{dom}\, f$ where the basic constraint qualification is satisfied and for any $\xi \in T_C(x) = \mathrm{dom}\, f'_x$, one has in terms of the set*

$$(4.65) \qquad \begin{aligned} \partial f(x)_\xi &= \{v \in \partial f(x) | f'_x(\xi) = \xi \cdot v\} \\ &= \{v \in \partial f(x) | \xi \in N_{\partial f(x)}(v)\} \end{aligned}$$

the following duality between the two kinds of second derivatives:

$$(4.66) \qquad f''_x(\xi, \eta) = \sup_{v \in \partial f(x)_\xi} \{f''_{x,v}(\xi) + v \cdot \eta\} \quad \text{for all } \eta \in \mathbf{R}^d,$$

$$(4.67) \qquad \inf_{\eta \in \mathbf{R}^n} \{f''_x(\xi, \eta) - v \cdot \eta\} = \begin{cases} f''_{x,v}(\xi) & \text{if } v \in \partial f(x)_\xi, \\ \infty & \text{otherwise.} \end{cases}$$

Moreover the supremum and the infimum are attained when finite.

PROOF. Recall that f'_x is the support function of $\partial f(x)$ (Theorem 4.5), and $g'_{F(x)}$ is the support function of $\partial g(F(x))$ (Theorem 3.1). In view of the formulas for f'_x and $\partial f(x)$ in (4.31) and (4.33), one has $v \in \partial f(x)_\xi$ if and only if there exists a vector $y \in \partial g(F(x))_{\nabla F(x)\xi}$ satisfying $y\nabla F(x) = v$. The right side of (4.66) can be written using (4.34) as

$$\sup_{y \in \partial g(F(x))_{\nabla F(x)\xi}} \{g''_{F(x),y}(\nabla F(x)\xi) + y \cdot [\xi\nabla^2 F(x)\xi + \nabla F(x)\eta]\}.$$

But this is $g''_{F(x)}(\nabla F(x)\xi, \xi\nabla^2 F(x)\xi + \nabla F(x)\eta)$ by Proposition 3.5 and equals $f''_x(\xi, \eta)$ by Theorem 4.6. Equation (4.66) is therefore correct. It can be interpreted as saying that $f''_x(\xi, \cdot)$ is the convex function conjugate to the function

$$(4.68) \qquad h(v) = -f''_{x,v}(\xi) + \delta_{\partial f(x)_\xi}(v).$$

We shall argue that h is a closed convex function and therefore is in turn the conjugate of $f''_x(\xi, \cdot)$, i.e. (4.67) holds. Certainly the indicator term in (4.68) is a closed convex function, because $\partial f(x)_\xi$ is a closed convex set. We have observed in the second-order part of the proof of Theorem 4.5 that the formula (4.6) for $f''_{x,v}(\xi)$ takes the alternative form (4.51), namely

$$f''_{x,v}(\xi) = \sup_{\substack{y \in G \\ y\nabla F(x) = v}} \{\gamma + y \cdot [\xi\nabla^2 F(x)\xi]\}, \quad \text{where } G = \partial g(F(x))_{\nabla F(x)\xi}.$$

The constant γ does not depend on v. Thus for a certain polyhedral convex function k one has

$$(4.69) \qquad -f''_{x,v}(\xi) = \inf_y \{k(u) | y\nabla F(x) = v\} \quad \text{when } v \in \partial f(x)_\xi.$$

Let $h_0(v)$ denote the infimum in this formula. The function h_0 is the so-called image of k under the linear transformation $y \mapsto y \nabla F(x)$, and it inherits from k the property of polyhedral convexity [**22**, Corollary 19.3.1] and in particular closedness, i.e. lower semicontinuity. We may conclude from (4.69) that $f_{x,v}''(\xi)$ is lower semicontinuous as a function of $v \in \partial f(x)_\xi$. Hence h is closed, as we needed to show.

It has been verified that the functions $f_x''(\xi, \cdot)$ and h are conjugate to each other. Also, though, $f_x''(\xi, \cdot)$ is polyhedral convex by Theorem 4.6. This property is preserved under conjugation, so h is polyhedral convex. The supremum in (4.66) and the infimum in (4.67) both involve polyhedral functions, therefore, and are attained when finite. $\square$

The duality in Theorem 4.7 can be compared with a similar duality discovered by Chaney [**12**] for finite functions f of the type (1.15) in Example 1.7. For functions of this type the parabolic second derivatives $f_x''(\xi, \eta)$ given by (4.56) coincide with the ones of Ben-Tal and Zowe [**17**] (except for a factor of $\frac{1}{2}$), as can readily be verified. Chaney demonstrates in effect in [**12**, Theorem 4.4] that in terms of such a formula for $f_x''(\xi, \varsigma)$ the expression $\inf_{\eta \in \mathbb{R}^n} \{f_x''(\xi, \eta) - v \cdot \eta\}$ gives his own second-order derivative (except for a factor of $\frac{1}{2}$); let us denote it by $\tilde{f}_{x,v}''(\xi)$; like $f_{x,v}''(\xi)$, this is ∞ when $v \notin \partial f(x)_\xi$. It follows then from Theorem 4.7 that $\tilde{f}_{x,v}'' = f_{x,v}''$ for this class of functions.

We shall not go into the details of Chaney's derivative here, but we note that it is applicable to other functions beyond the ones in Example 1.7 and even beyond the class $f(x) = g(F(x))$ chosen in the present paper. Whether it coincides with our derivative in other situations is an open question. Chaney's derivative is in any case only defined for *finite* functions of certain kinds, whereas ours allows f to be extended-real-valued as a way of incorporating constraints.

The following consequence of Theorem 4.7 deserves to be recorded.

COROLLARY 4.8. *If the basic constraint qualification is satisfied at the point* $x \in C$ *then* $f_x''(\xi, \eta)$ *is lower semicontinuous jointly in* ξ *and* η *relative to* $T_C(x) \times \mathbb{R}^n$.

PROOF. Since $\partial f(x)$ is a convex polyhedron, there are only finitely many sets $\partial F(x)_\xi$ that arise as ξ varies over $T_C(x)$, the closures of the faces of $\partial f(x)$. Denote the faces of $\partial f(x)$ by V_k, $k \in K$ (a finite index set). These are relatively open convex sets (cf. [**22**, §18]), and the normal cone $N_{\partial f(x)}(v)$ is the same for all $v \in V_k$; denote it by N_k. Then

$$(4.70) \qquad \bigcup_{k \in K} N_k = \{\xi | f_x'(\xi) < \infty\} = T_C(x).$$

In (4.66) we have

$$(4.71) \qquad f_x''(\xi, \eta) = \sup_{v \in V_k} \{f_{x,v}''(\xi) + v \cdot \eta\} \quad \text{when } \xi \in N_k.$$

Since $f_{x,v}''(\xi)$ is lower semicontinuous in ξ (by virtue of the definition of $f_{x,v}''$ as an epi-limit), the supremum in (4.71) gives a lower semicontinuous function of $(\xi, \eta) \in N_k \times \mathbb{R}^n$. Since K is a finite index set and (4.70) holds, this implies $f_x''(\xi, \eta)$ is lower semicontinuous on $T_C(x) \times \mathbb{R}^n$.

A final observation is that the basic constraint qualification is a stable condition that yields more than just properties at the particular x where it is assumed.

PROPOSITION 4.9. *If the basic constraint qualification holds at a point $x \in C$, then it also holds at all points $x' \in C$ in some neighborhood of x. The preceding theorems are thus applicable at such points x' as well.*

PROOF. If this were false, we could find a sequence of points $x^\nu \in C$ ($\nu = 1, 2, \ldots$) and vectors $y^\nu \in N_D(F(x^\nu))$ with $y^\nu \nabla F(x^\nu) = 0$, such that $x^\nu \to x$. The vectors y^ν could be normalized so that $|y^\nu| = 1$, and by passing to a subsequence if necessary they could be assumed to converge to some y with $|y| = 1$. Then $y \nabla F(x) = 0$ by the continuity of ∇F, and $y \in N_D(F(x))$ by the closedness of the multifunction $u \mapsto N_D(u) = \partial \delta_D(u)$ (since D is convex) [**22**, Theorem 24.4]. The basic constraint qualification at x would then be violated by y. □

REFERENCES

1. F. H. Clarke, *A new approach to Lagrange multipliers*, Math. Oper. Res. **1** (1976), 165–174.
2. ____, *Nonsmooth analysis and optimization*, Wiley, 1983.
3. A. D. Ioffe, *Necessary and sufficient conditions for a local minimum. I: A reduction theorem and first order conditions*, SIAM J. Control Optim. **17** (1979), 245–250.
4. ____, *Necessary and sufficient conditions for a local minimum. II: Conditions of Levilin-Miljutin-Osmolovskii type*, SIAM J. Control Optim. **17** (1979), 251–265.
5. ____, *Necessary and sufficient conditions for a local minimum. III: Second order conditions and augmented duality*, SIAM J. Control Optim. **17** (1979), 266–288.
6. J.-B. Hiriart-Urruty, *Approximating a second order directional derivative for nonsmooth convex functions*, SIAM J. Control Optim. **20** (1982), 783–807.
7. ____, *Limiting behavior of the approximate first and second order directional derivatives for a convex function*, Nonlinear Anal. **6** (1982), 1309–1326.
8. ____, *Calculus rules on the approximate second order directional derivative of a convex function*, SIAM J. Control Optim. **22** (1984), 381–404.
9. R. W. Chaney, *On sufficient conditions in nonsmooth optimization*, Math. Oper. Res. **7** (1982), 463–475.
10. ____, *Second-order sufficient conditions for nondifferentiable programming problems*, SIAM J. Control Optim. **20** (1982), 20–33.
11. ____, *A general sufficiency theorem for nonsmooth nonlinear programming*, Trans. Amer. Math. Soc. **276** (1983), 235–245.
12. ____, *Second-order directional derivatives for nonsmooth functions*, J. Math. Anal. Appl. **128** (1987), 495–511.
13. A. Auslender, *Stability in mathematical programming with nondifferentiable data*, SIAM J. Control Optim. **22** (1984), 239–254.
14. J. P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res. **9** (1984), 87–111.
15. A. Seeger, *Second-order directional derivatives in parametric optimization problems*, Math. Oper. Res. (to appear).
16. R. T. Rockafellar, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Ann. Inst. H. Poincaré **2** (1985), 167–184.
17. A. Ben-Tal, *Second order theory for extremum problems*, Systems Analysis and External Methods (A. V. Fiacco and K. Kostaneta, eds.), Lecture Notes in Economics and Math. Sciences, Springer-Verlag, 1980, pp. 336–356.
18. A. Ben-Tal and J. Zowe, *Necessary and sufficient conditions for a class of nonsmooth minimization problems*, Math. Programming **24** (1982), 70–91.
19. ____, *A unified theory of first and second order conditions for extremum problems in topological vector spaces*, Math. Programming Stud. **19** (1982), 39–76.
20. ____, *Directional derivatives in nonsmooth optimization*, J. Optim. Theory Appl.
21. R. T. Rockafellar, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. Oper. Res. (forthcoming).
22. ____, *Convex analysis*, Princeton Univ. Press, 1970.

23. R. T. Rockafellar and R. J.-B. Wets, *Linear-quadratic programming with stochastic penalties: the finite generation algorithm*, Stochastic Optimization (V. Arkin, A. Shiraeu and R. J.-B. Wets, eds.), Lecture Notes in Control and Information Sciences, No. 81, Springer-Verlag, 1986, pp. 545–560.

24. M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl. **4** (1969), 303–320.

25. M. J. D. Powell, *A method for nonlinear optimization in minimization problems*, Optimization (R. Fletcher, ed.), Academic Press, 1969.

26. R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control **12** (1974), 268–285.

27. J. Sun, *On monotropic piecewise quadratic programming*, Ph.D. dissertation, Dept. of Appl. Math., Univ. of Washington, Seattle, 1986.

28. J. Burke, *Second order necessary and sufficient conditions for convex composite NDO*, Math. Prog. **38** (1987), 287–302.

29. R. A. Wijsman, *Convergence of sequences of convex sets, cones and functions. I*, Bull. Amer. Math. Soc. **70** (1964), 186–188.

30. ____, *Convergence of sequences of convex sets, cones and functions. II*, Trans. Amer. Math. Soc. **123** (1966), 32–45.

31. U. Mosco, *Convergence of convex sets and solutions to variational inequalities*, Adv. in Math. **3** (1969), 510–585.

32. E. DeGiorgi, *Convergence problems for functionals and operators*, Recent Methods in Non-Linear Analysis (E. DeGiorgi, E. Magenes and U. Mosco, eds.), Pitagoro Editrice, Bologna, 1980.

33. H. Attouch, *Variational convergence for functions and operators*, Pitman, London, 1984.

34. R. J.-B. Wets, *Convergence of convex functions, variational inequalities and convex optimization problems*, Variational Inequalities and Complementary Problems (R. Cottle, F. Gianessi and J.-L. Lions, eds.), Wiley, 1980, pp. 405–419.

35. H. Attouch and R. J.-B. Wets, *Approximation and convergence in nonlinear optimization*, Non-linear Programming 4, (O. L. Mangasarian et al., eds.), Academic Press, 1981, pp. 367–394.

36. R. T. Rockafellar and R. J.-B. Wets, *Variational systems, an introduction*, Multifunctions and Integrands (G. Salinetti, ed.), Lecture Notes in Math., vol. 1091, Springer-Verlag, 1984, pp. 1–54.

37. C. Kuratowski, *Topologie*, PWN, Warsaw, 1958.

38. R. T. Rockafellar, *Generalized directional derivatives and subgradients of nonconvex functions*, Canad. J. Math. **37** (1980), 257–280.

39. F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc. **205** (1975), 247–262.

40. O. L. Mangasarian and S. Fromovitz, *The Fritz John conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl. **17** (1967), 73–74.

41. A. Ben-Tal, *Second-order and related extremality conditions in nonlinear programming*, J. Optim. Theory Appl. **31** (1980), 143–165.

42. R. T. Rockafellar, *Directionally Lipschitzian functions and subdifferential calculus*, Proc. London Math. Soc. **39** (1979), 339–355.

DEPARTMENT OF MATHEMATICS AND APPLIED MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195