To appear in: *Recent Advances in Optimization*,
Dingzhu Du, Liqun Qi and Robert Womersley, editors,
World Scientific Publishers ©1995

Second-Order Nonsmooth Analysis in Nonlinear Programming

René Poliquin[1]
Department of Mathematical Sciences, University of Alberta, Edmonton, Alberta Canada, T6G 2G1

Terry Rockafellar[2]
Department of Mathematics, University of Washington, Seattle, WA 98195 USA

Abstract

Problems of nonlinear programming are placed in a broader framework of composite optimization. This allows second-order smoothness in the data structure to be utilized despite apparent nonsmoothness in the objective. Second-order epi-derivatives are shown to exist as expressions of such underlying smoothness, and their connection with several kinds of second-order approximation is examined. Expansions of the Moreau envelope functions and proximal mappings associated with the essential objective functions for certain optimization problems in composite format are studied in particular.

1 Introduction

Problems in nonlinear programming are customarily stated in terms of a finite system of equality and inequality constraints, defining a feasible set over which a certain function is to be minimized. For most numerical work it is assumed that the constraint and objective functions are $\mathcal{C}^2$, so that second-order methodology can be utilized.

[1]This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under grant OGP41983.
[2]This work was supported in part by the National Science Foundation under grant DMS–9200303.

This is taken as the model for "smooth" optimization, and any problem whose objective function fails to enjoy such differentiability, for instance by being only piecewise $\mathcal{C}^2$, belongs then to the category of "nonsmooth" optimization. But in practice a distinction between smooth and nonsmooth optimization based on such grounds is artificial.

Many problems that start out with a nonsmooth objective, perhaps involving penalty functions and "max" expressions, can be recast with a smooth objective. On the other hand, nominally smooth problems with inequality constraints inherently exhibit nonsmoothness in their geometry. Anyway, techniques for solving those problems often veer into nonsmoothness by appealing to merit functions or dualization.

The real issue in numerical and theoretical optimization alike is how to represent and exploit to the fullest whatever degree of smoothness may be available in a problem's elements. In this respect the traditional format falls short. Its deficiency is that it places all the emphasis in problem formulation on making a list of constraints, which must be simple equations or inequalities, each associated with an explicit constraint function, and afterward merely specifying one additional function for the objective. While a vehicle is provided for working with nonsmoothness in the boundary of the feasible set, none is provided for nonsmoothness as it might be found in the graph of the function being minimized, or for that matter for any other structural features of the objective.

In contrast, the *composite format* for problems of optimization treats both constraints and objective more supportively and is able to span a wider range of situations with ease. In the composite format, a problem is set up by specifying a representation of the type

$$(\mathcal{P}) \qquad \text{minimize} \quad f(x) := g(F(x)) \text{ over } x \in I\!\!R^n,$$

where $F : I\!\!R^n \to I\!\!R^m$ is the *data mapping* and $g : I\!\!R^m \to \overline{I\!\!R}$ is the *model function*. The mapping F supplies the problem's special elements and carries its smoothness, whereas the function g provides the structural mold. Not only does g have no need to be smooth, it can even be extended-real-valued, with values in $\overline{I\!\!R} = [-\infty, \infty]$ instead of just $I\!\!R = (-\infty, \infty)$. This is central to the idea. The *feasible set* in $(\mathcal{P})$ is defined to be

$$C = \operatorname{dom} f := \{x \mid f(x) < \infty\} = \{x \mid F(x) \in D\},$$

where

$$D = \operatorname{dom} g := \{u \mid g(u) < \infty\}.$$

Here we aim at applying second-order nonsmooth analysis to the essential objective function f of problem $(\mathcal{P})$ in this format. Keeping close to the ordinary domain of nonlinear programming, if not the usual framework, we concentrate on the case in which

- F is a $\mathcal{C}^2$ mapping, and
- g is a proper, convex function that is *polyhedral*,

i.e., such that the set $\operatorname{epi} g := \{(u,\alpha) \in I\!R^m \times I\!R \mid \alpha \geq g(u)\}$ is polyhedral convex, cf. [17, Section 19]. The set D is polyhedral then as well.

The nature and extent of the problem class covered under these restrictions is explored in Section 2 along with the relationship to "amenable" functions, which by definition have composite expressions $f = g \circ F$ with smooth F and convex g satisfying a certain constraint qualification. For amenable functions a highly developed theory of first- and second-order generalized derivatives is now in place and ready for application under the circumstances described here. Formulas for such derivatives are worked out in Section 3 and incorporated into optimality conditions in the composite format, in particular second-order conditions related to epigraphical approximation. In Section 4, second-order expansions in terms of uniform convergence instead of epigraphical convergence are studied, and the question of Hessian matrices in a standard or generalized sense is taken up.

Finally, Section 5 analyzes the Moreau envelope functions

$$e_\lambda(x) \;:=\; \min_{x'} \Big\{ f(x') + \frac{1}{2\lambda}|x'-x|^2 \Big\} \;\text{ for }\; \lambda > 0,$$

which relate to epigraphical approximation of f because $e_\lambda(x)$ increases to $f(x)$ as $\lambda \searrow 0$. These functions not only approximate but provide a kind of *regularization* of f. While f may be extended-real-valued and have discontinuities (in particular, jumps to ∞), e_λ is finite and locally Lipschitz continuous and has one-sided directional derivatives at all points. Moreover, the minimizing sets agree: $\operatorname{argmin} e_\lambda = \operatorname{argmin} f$ for all $\lambda > 0$. We investigate the degree to which second-order properties of e_λ at minimizing points $\bar{x}$ correspond to such properties of f at these points. Second-order properties of e_λ have a bearing on numerical techniques like the proximal point algorithm in the minimization of f, since they inevitably depend on the proximal mapping

$$P_\lambda(x) \;:=\; \operatorname*{argmin}_{x'} \Big\{ f(x') + \frac{1}{2\lambda}|x'-x|^2 \Big\} \;\text{ for }\; \lambda > 0.$$

This phase of our effort owes its inspiration to recent work of Lemaréchal and Sagagastizábal [6], followed by Qi [16], who were motivated by the goals just mentioned. These authors have concentrated on finite, convex functions f, not necessarily of the composite form adopted here, whereas we relinquish convexity and welcome infinite values in order to obtain results that deal with constraints. On the other hand, Qi [16] takes up the topic of semi-smoothness of ∇e_λ, which is not addressed here.

2 Problem characteristics and amenability

To understand better the class of optimization problems $(\mathcal{P})$ covered by the composite format through some choice of a $\mathcal{C}^2$ mapping F and polyhedral function g, it helps first to see how problems that are stated in the traditional manner can be accommodated.

Example 2.1. *For $\mathcal{C}^2$ functions $f_0, f_1, \ldots, f_m$ on $I\!R^n$, consider the minimization of $f_0(x)$ subject to*

$$f_i(x) \begin{cases} \leq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m. \end{cases}$$

This fits the composite format of minimizing $f = g \circ F$ over $I\!R^n$ for the $\mathcal{C}^2$ mapping $F : I\!R^n \to I\!R^{m+1}$ defined by $F(x) = (f_0(x), f_1(x), \ldots, f_m(x))$ and the polyhedral function $g : I\!R^{m+1} \to \overline{I\!R}$ defined by

$$g(u) = g(u_0, u_1, \ldots, u_m) = \begin{cases} u_0 & \text{if } u_i \begin{cases} \leq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m, \end{cases} \\ \infty & \text{otherwise.} \end{cases}$$

Next we look at an apparently very different model, which illustrates accommodations that can be made to nonsmoothness.

Example 2.2. *For $\mathcal{C}^2$ functions $f_1, \ldots, f_m$ on $I\!R^n$, consider the minimization of*

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

over all $x \in I\!R^n$ (no constraints). This fits the composite format $f = g \circ F$ with $F(x) = (f_1(x), \ldots, f_m(x))$ and $g(u) = g(u_1, \ldots, u_m) = \max\{u_1, \ldots, u_m\}$. The mapping F is $\mathcal{C}^2$ and the function g is polyhedral.

It is well known that this kind of problem, although nominally concerned with unconstrained minimization of a nonsmooth function, can be posed instead in terms of minimizing a linear function subject to smooth inequality constraints. Indeed, in the notation $\tilde{x} = (x, \alpha) \in I\!R^{n+1}$ it corresponds to minimizing $\tilde{f}_0(\tilde{x})$ subject to $\tilde{f}_i(\tilde{x}) \leq 0$ for $i = 0, 1, \ldots, m$, where $\tilde{f}_0(\tilde{x}) = \alpha$ and $\tilde{f}_i(\tilde{x}) = f_i(x) - \alpha$ for $i = 1, \ldots, m$. Thus it surely deserves to be treated on a par with other problems where smoothness dominates the numerical methodology, at least as long as the dimension n is not unduly large.

Another sort of flexibility in the composite model comes to light in the way constraints can be handled in patterns deviating from the standard one in Example 2.1. Simple equations and inequalities can be supplemented by conditions that restrict a function's values to lie in a certain interval. Box constraints on x do not have to be written with explicit constraint functions at all.

Example 2.3. *For $\mathcal{C}^2$ functions $f_0, f_1, \ldots, f_m$ on $I\!R^n$, nonempty closed intervals $I_1, \ldots, I_m$ in $I\!R$ and a nonempty polyhedral set $X \subset I\!R^n$, consider the problem of minimizing $f_0(x)$ over the set*

$$C := \{x \in X \mid f_i(x) \in I_i,\ i = 1 \ldots m\},$$

or equivalently, minimizing $f(x)$ over all $x \in I\!R^n$ in the case of

$$f(x) = f_0(x) + \delta_C(x) = \begin{cases} f_0(x) & \text{if } x \in C, \\ \infty & \text{if } x \notin C. \end{cases}$$

This concerns $f = g \circ F$ *for the* $\mathcal{C}^2$ *mapping* $F : I\!R^n \to I\!R^{m+n+1}$ *defined by*

$$F(x) \;=\; (f_0(x), f_1(x), \ldots, f_m(x), x)$$

and the polyhedral function $g : I\!R^{m+n+1} \to \overline{I\!R}$ *defined by*

$$g(u) = g(u_0, u_1, \ldots, u_m, u_{m+1}, \ldots, u_{m+n}) = \begin{cases} u_0 & \text{if } \begin{cases} u_i \in I_i \text{ for } i = 1, \ldots, m, \\ (u_{m+1}, \ldots, u_{m+n}) \in X, \end{cases} \\ \infty & \text{otherwise.} \end{cases}$$

Example 2.3 encompasses Example 2.1 as the special case where $X = I\!R^n$ and $I_i = (-\infty, 0]$ for $i = 1, \ldots, s$ but $I_i = [0, 0]$ for $i = s+1, \ldots, m$. On the other hand, Example 2.3 could be extended by taking f_0 to be a max function as in Example 2.2, $f_0(x) = \max\{f_{01}(x), \ldots, f_{0r}(x)\}$. Then the $\mathcal{C}^2$ functions f_{0k} would become additional components of F, and the u_0 part of u would turn into a vector $(u_{01}, \ldots, u_{0r})$, with $\max\{u_{01}, \ldots, u_{0r}\}$ entering the formula for $g(u)$.

An alternative way of arriving at nonsmoothness in the objective is illustrated by the following model.

Example 2.4. *For* $\mathcal{C}^2$ *functions* $f_0, f_1, \ldots, f_m$ *on* $I\!R^n$ *and proper polyhedral functions* $g_i : I\!R \to \overline{I\!R}$ *for* $i = 1, \ldots, m$*, the problem of minimizing*

$$f_0(x) + g_1(f_1(x)) + \cdots + g_m(f_m(x))$$

over all $x \in I\!R^n$ *corresponds to* $f = g \circ F$ *for the* $\mathcal{C}^2$ *mapping* F *with* $F(x) = (f_0(x), f_1(x), \ldots, f_m(x))$ *and the polyhedral function* g *with*

$$g(u) = g(u_0, u_1, \ldots, u_m) = u_0 + g_1(u_1) + \cdots + g_m(u_m).$$

Polyhedral functions g_i of a single real variable as in Example 2.4 are piecewise linear convex functions in the obvious sense, except that they could have the value ∞ outside of a some closed interval I_i. As a special case, such a function could have just one "piece," being affine on I_i, or even just 0 on I_i (with the term $g_i(f_i(x))$ just representing then a constraint $f_i(x) \in I_i$). Piecewise linear functions with multiple slopes arise in a setting like Example 2.4 when constraints are relaxed by linear penalty expressions. Of course, a geometric constraint $x \in X$ with X polyhedral (e. g. a box—a product of closed intervals, not necessarily bounded) could be built into Example 2.4 as in Example 2.3.

Within nonsmooth analysis, the composite format in optimization is closely associated with concept of "amenability." For simplicity in stating the definition and working with it in the rest of the paper, we introduce the following notation. For any mapping $F : I\!R^n \to I\!R^m$ and any vector $y \in I\!R^m$ we simply write yF for the scalar function defined by $(yF)(x) = \langle y, F(x) \rangle$. Thus,

$$(yF)(x) = y_1 f_1(x) + \cdots + y_m f_m(x) \text{ when}$$
$$F = (f_1, \ldots, f_m) \text{ and } y = (y_1, \ldots, y_m),$$

and if F is $\mathcal{C}^1$ with Jacobian $\nabla F(x)$ one has further that

$$\nabla(yF)(x) = y_1\nabla f_1(x) + \cdots + y_m\nabla f_m(x) = \nabla F(x)^T y.$$

Definition 2.5. *A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is amenable at $\bar{x}$ if $f(\bar{x})$ is finite and, at least locally around $\bar{x}$, there is a representation $f = g\circ F$ in which the mapping F is $\mathcal{C}^1$, the function g is proper, lsc (lower semicontinuous) and convex, and the following condition, an abstract constraint qualification, is satisfied by the normal cone $N_D(F(\bar{x}))$ to the convex set $D = \operatorname{dom} g$ at $F(\bar{x})$:*

$$\text{there is no vector } y \neq 0 \text{ in } N_D(F(\bar{x})) \text{ with } \nabla(yF)(\bar{x}) = 0. \qquad (CQ)$$

It is strongly amenable if F is $\mathcal{C}^2$ rather than just $\mathcal{C}^1$, and fully amenable if, in addition, g is piecewise linear-quadratic.

To say that g is *piecewise linear-quadratic* is to say that its effective domain D is the union of finitely many polyhedral sets, on each of which the formula for g is linear-quadratic, i.e., a polynomial of degree at most 2. When no quadratic terms are involved, g is just *piecewise linear* (piecewise affine might be a better term). The convex functions that are piecewise linear are precisely the polyhedral functions of convex analysis we have been referring to so far. This leads to the following observation, which paves the way for us to applying the theory of amenable functions, cf. [11]–[15], to the class of problems under consideration.

Proposition 2.6. *For problem $(\mathcal{P})$ in the composite format with F of class $\mathcal{C}^2$ and g polyhedral, let $\bar{x}$ be a point of the feasible set C at which constraint qualification (CQ) is satisfied. Then the essential objective function f is fully amenable at all points $x \in C$ in some neighborhood of $\bar{x}$.*

Proof. This merely records the import for problem $(\mathcal{P})$ of the observations just made, utilizing the fact that if (CQ) holds at $\bar{x}$ it must hold for all $x \in C$ in some neighborhood of $\bar{x}$ (cf. [13]). □

The constraint qualification (CQ) is satisfied trivially when $F(\bar{x}) \in \operatorname{int} D$, since $N_D(\bar{u}) = \{0\}$ at all points $\bar{u} \in \operatorname{int} D$. To see what it means in other situations, we inspect the preceding examples one by one.

Example 2.1′. *In Example 2.1, the constraint qualification (CQ) reduces to the Mangasarian-Fromovitz condition (written in its equivalent dual form): unless all the coefficients $y_1, \ldots, y_m$ are taken to be 0, it is impossible to have the equation*

$$y_1\nabla f_1(\bar{x}) + \cdots + y_m\nabla f_m(\bar{x}) = 0$$

with $y_i \geq 0$ for indices $i \in \{1, \ldots, s\}$ such that $f_i(\bar{x}) = 0$, and $y_i = 0$ for indices $i \in \{1, \ldots, s\}$ such that $f_i(\bar{x}) < 0$ (but y_i unrestricted for indices $i \in \{s+1, \ldots, m\}$).

Detail. The set D in this case consists of all vectors $u = (u_1, \ldots, u_m)$ such that $u_i \leq 0$ for $i = 1, \ldots, s$ and $u_i = 0$ for $i = s+1, \ldots, m$. For any $\bar{u} \in D$, therefore, the normal cone $N_D(\bar{u})$ consists of the vectors y with $y_i \geq 0$ for $i \in \{1, \ldots, s\}$ such that $\bar{u}_i = 0$, whereas $y_i = 0$ for $i \in \{1, \ldots, s\}$ such that $\bar{u}_i < 0$. □

Example 2.2′. *In Example 2.2, condition (CQ) reduces to triviality; it is satisfied automatically at every point $\bar{x} \in \mathbb{R}^n$.*

Detail. In this case $D = \mathbb{R}^n$, hence $F(\bar{x}) \in \operatorname{int} D$ always. □

Example 2.3′. *In Example 2.3, the constraint qualification (CQ) at a feasible point $\bar{x}$ means that*

$$\begin{gathered}\text{the only multipliers } y_i \in N_{I_i}(f_i(\bar{x})) \text{ satisfying} \\ -\textstyle\sum_{i=1}^m y_i \nabla f_i(\bar{x}) \in N_X(\bar{x}) \text{ are } y_1 = 0, \dots, y_m = 0.\end{gathered}$$

Here I_i is a closed interval with lower bound a_i and upper bound b_i (these bounds possibly being infinite, with $a_i \le b_i$), and the relation $y_i \in N_{I_i}(f_i(\bar{x}))$ restricts sign of y_i in the following pattern, depending on how the constraint $f_i(\bar{x}) \in I_i$ is satisfied at $\bar{x}$ relative to these bounds:

$$y_i \in N_{I_i}(f_i(\bar{x})) \iff \begin{cases} y_i \ge 0 & \text{when } a_i < f_i(\bar{x}) = b_i, \\ y_i \le 0 & \text{when } a_i = f_i(\bar{x}) < b_i, \\ y_i = 0 & \text{when } a_i < f_i(\bar{x}) < b_i, \\ y_i \text{ free} & \text{when } a_i = f_i(\bar{x}) = b_i. \end{cases}$$

Detail. The representation $f = g \circ F$ for this case has

$$D = I_1 \times \cdots \times I_m \times X,$$

and consequently

$$N_D(F(\bar{x})) = N_{I_1}(f_1(\bar{x})) \times \cdots \times N_{I_m}(f_m(\bar{x})) \times N_X(\bar{x}).$$

The characterization of the one-dimensional relations $y_i \in N_{I_i}(u_i)$ is elementary. □

Note that the constraint qualification in Example 2.3′ reduces to the Mangasarian-Fromovitz condition in Example 2.1′ when X is the whole space, so that $N_X(\bar{x}) = \{0\}$, while $I_i = (-\infty, 0]$ for $i = 1, \dots, s$ (so that $N_{I_i}(u_i)$ equals $[0, \infty)$ if $u_i = 0$ but equals $\{0\}$ if $u_i < 0$), whereas $I_i = [0, 0]$ for $i = s+1, \dots, m$ (so that $N_{I_i}(u_i) = (-\infty, \infty)$ as long as $u_i = 0$).

Example 2.4′. *In Example 2.4 with the closed intervals* $\operatorname{dom} g_i$ *denoted by I_i (these possibly being all of $\mathbb{R}$ for some indices i), the constraint qualification (CQ) takes the same form as it does in Example 2.3′, except that $N_X(\bar{x})$ is replaced by $\{0\}$.*

The examples have indicated the advantages of the composite format in allowing optimization problems to be expressed in a variety of ways. But just how general is the class of problems the composite format covers under our restrictions? This question is answered by the next result.

Theorem 2.7. *The optimization problems that can be placed in the composite format as $(\mathcal{P})$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g are precisely the ones which, in principle, concern the minimization over a set C, specifiable by a finite system of $\mathcal{C}^2$ equality and inequality constraints, of a function f_0 that is either $\mathcal{C}^2$ itself or expressible as the pointwise max of a finite collection of $\mathcal{C}^2$ functions.*

Moreover, the representation can always be set up in such a way that a point $\bar{x} \in C$ satisfies the constraint qualification (CQ) for $(\mathcal{P})$ if and only if it satisfies the Mangasarian-Fromovitz condition relative to the equality and inequality constraints utilized in representing C.

Proof. If an optimization problem has a representation of the kind described, it fits into the composite format in the manner of Example 2.1 as supplemented by the device explained after Example 2.2. Then (CQ) reduces to the Mangasarian-Fromovitz constraint qualification just as in Examples 2.1′.

Conversely, suppose $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. The epigraph set $\operatorname{epi} f$ consists then of the points (x, α) such that $(F(x), \alpha) \in \operatorname{epi} g$. To say that g is polyhedral is to say that $\operatorname{epi} g$ can be represented by a finite system of linear constraints, say

$$(u, \alpha) \in \operatorname{epi} g \iff l_k(u, \alpha) \begin{cases} \leq 0 & \text{for } k = 1, \dots, q, \\ = 0 & \text{for } k = q+1, \dots, r, \end{cases}$$

where each function l_k is affine on $\mathbb{R}^{m+1}$. Without loss of generality this system can be set up so that the Mangasarian-Fromovitz condition is satisfied at all points of $\operatorname{epi} g$. (Proceeding from an arbitrary system, one can rewrite as equality constraints any inequalities that never hold strictly, and then pare down the list of equality constraints until none is redundant.)

The equality constraint functions l_k must have the form $l_k(u, \alpha) = \langle a_k, u \rangle - b_k$ for some vector $a_k \in \mathbb{R}^m$ and scalar $b_k \in \mathbb{R}$, since otherwise the hyperplane defined by $l_k(u, \alpha) = 0$ could not contain $\operatorname{epi} g$. The same form may be present for some of the inequality constraint functions. We can suppose that for a certain $p \leq q$ all of the functions l_k for $k = p+1, \dots, r$ have this special form, whereas for $k = 1, \dots, p$ none of them has it. In the latter case we can rescale l_k to write it $l_k(u, \alpha) = \langle a_k, u \rangle - b_k - \alpha$ for some $a_k \in \mathbb{R}^m$ and $b_k \in \mathbb{R}$, since otherwise, again, the half-space defined by $l_k(u, \alpha) \leq 0$ could not contain $\operatorname{epi} g$. The set $D = \operatorname{dom} g$ is given then by

$$u \in D \iff \langle a_k, u \rangle - b_k \begin{cases} \leq 0 & \text{for } k = p+1, \dots, q, \\ = 0 & \text{for } k = q+1, \dots, r. \end{cases}$$

We have $\nabla l_k(u, \alpha) \equiv (a_k, -1)$ for $k = 1, \dots, p$, but $\nabla l_k(u, \alpha) \equiv (a_k, 0)$ for $k = p+1, \dots, r$. The fact that the Mangasarian-Fromovitz condition holds everywhere for the system representing $\operatorname{epi} g$ implies that it holds everywhere for this system representing D.

Because $\operatorname{epi} f$ consists of all pairs (x, α) such that $(F(x), \alpha) \in \operatorname{epi} g$, it is specified by $l_k(F(x), \alpha) \leq 0$ for $k = 1, \dots, q$ and $l_k(F(x), \alpha) = 0$ for $k = q+1, \dots, r$. Let

$h_k(x) = \langle a_k, F(x)\rangle - b_k$ for $i = 1, \ldots, r$. Thus, according to what we have arranged,

$$(x, \alpha) \in \operatorname{epi} f \quad \Longleftrightarrow \quad h_k(x) \begin{cases} \leq \alpha & \text{for } k = 1, \ldots, p, \\ \leq 0 & \text{for } k = p+1, \ldots, q, \\ = 0 & \text{for } k = q+1, \ldots, r. \end{cases}$$

In other words, the set $C = \operatorname{dom} f$ is specified by $h_k(x) \leq 0$ for $k = p+1, \ldots, q$ and $h_k(x) = 0$ for $k = q+1, \ldots, r$, and the problem of minimizing f over $I\!R^n$ corresponds to minimizing over this set C the function $f_0(x) = \max\{h_1(x), \ldots, h_p(x)\}$.

How do the constraint qualifications correspond in this framework? Consider any $\bar{x} \in C$. Condition (CQ) forbids the existence of a nonzero vector $y \in N_D(F(\bar{x}))$ such that $\nabla(yF)(\bar{x}) = 0$. We know from the representation given to D that $N_D(F(\bar{x}))$ consists of all $y = \sum_{k=p+1}^{r} \lambda_k a_k$ such that

$$\lambda_k \begin{cases} \geq 0 & \text{for } k \in \{p+1, \ldots, q\} \text{ with } \langle a_k, F(\bar{x})\rangle - b_k = 0, \\ = 0 & \text{for } k \in \{p+1, \ldots, q\} \text{ with } \langle a_k, F(\bar{x})\rangle - b_k < 0, \\ \text{free} & \text{for } k \in \{q+1, \ldots, r\}, \end{cases}$$

where furthermore (because the Mangasarian-Fromovitz condition is satisfied universally in the representation of D) the vector $y = \sum_{k=p+1}^{r} \lambda_k a_k$ cannot be 0 unless all the coefficients λ_k vanish. It follows that the vectors of the form $\nabla(yF)(\bar{x})$ for some $y \in N_D(F(\bar{x}))$ are precisely those of the form $\sum_{k=p+1}^{r} \lambda_k \nabla h_k(\bar{x})$, and that (CQ) requires, under the restrictions listed for λ_k, that the zero vector cannot be expressed in this form except by taking every $\lambda_k = 0$. Thus, (CQ) at $\bar{x}$ comes out as identical to the Mangasarian-Fromovitz constraint qualification at $\bar{x}$ relative to the specification of C by the functions h_k. □

In the statement of Theorem 2.7, the words "in principle," "specifiable," and "expressible" warn that although it may be possible to reduce a problem to the special form described, this may be neither easy nor expedient. The advantage of the composite format is that it bypasses such reformulation and allows one to move ahead without it, if that is preferred.

3 Subgradients, epi-derivatives and optimality

Our task in analyzing problem $(\mathcal{P})$ is greatly assisted by Proposition 2.6. When a function $f : I\!R^n \to \overline{I\!R}$ is amenable at $\bar{x}$, it is Clarke regular at $\bar{x}$ in particular; cf. [2] and [12]. In consequence, all the various definitions of "subgradient" that might in general be invoked lead to the same set $\partial f(\bar{x})$.

Derivatives simplify as well. First-order one-sided derivatives arise from considering difference quotient functions

$$\Delta_{x,t} f : \xi \mapsto [f(x + t\xi) - f(x)]/t \quad \text{for } t > 0.$$

Classical differentiability of f at $\bar{x}$ can be identified with the case where, as $t \searrow 0$, the functions $\Delta_{\bar{x},t} f$ converge pointwise, uniformly on all bounded sets, to some linear function. Such uniform convergence, even if to a possibly nonlinear function, is too narrow an idea, though, to serve when f is extended-real-valued, as we wish it to be here in harmony with our mode of handling constraints. A substitute notion with many interesting ramifications can be based instead on *epi-convergence* of functions, which expresses set convergence of their epigraphs.

We say that f is *epi-differentiable* at $\bar{x}$ if, as $t \searrow 0$, the functions $\Delta_{\bar{x},t} f$ epi-converge to a proper function h; such a limit function need not be linear but must of necessity be lsc and positively homogeneous. Then h is the first-order *epi-derivative function* for f at $\bar{x}$ and is denoted by $f'_{\bar{x}}$. The property of epi-convergence translates into having, for each choice of a sequence $t^\nu \searrow 0$ and a vector ξ, that

$$\begin{cases} \liminf_\nu \Delta_{\bar{x},t^\nu}(\xi^\nu) \geq f'_{\bar{x}}(\xi) & \text{for every sequence } \xi^\nu \to \xi, \\ \limsup_\nu \Delta_{\bar{x},t^\nu}(\xi^\nu) \leq f'_{\bar{x}}(\xi) & \text{for some sequence } \xi^\nu \to \xi. \end{cases}$$

We say further that f is *strictly epi-differentiable* at $\bar{x}$ if, not only as $t \searrow 0$ but as $x \to \bar{x}$ with $f(x) \to f(\bar{x})$, the functions $\Delta_{x,t} f$ epi-converge (the limit in this wider sense necessarily still being the function $f'_{\bar{x},t}$).

Theorem 3.1. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any feasible solution to $(\mathcal{P})$ at which condition (CQ) holds. Then at all feasible solutions x in some neighborhood of $\bar{x}$, f is epi-differentiable at x and has at least one subgradient there as well, the subgradients being characterized as the vectors v such that*

$$f(x') \;\geq\; f(x) + \langle v, x' - x \rangle + o(|x' - x|).$$

The epi-derivative function f'_x is convex and positively homogeneous, the subgradient set $\partial f(x)$ is convex and closed, and the two are related by

$$\begin{aligned} f'_x(\xi) &= \sup_{v \in \partial f(x)} \langle v, \xi \rangle, \\ \partial f(x) &= \{ v \mid f'_x(\xi) \geq \langle v, \xi \rangle \ \text{ for all } \xi \}. \end{aligned}$$

Furthermore, these epi-derivative functions and subgradient sets are obtained from those for g by the formulas

$$\begin{aligned} f'_x(\xi) &= g'_{F(x)}(\nabla F(x)\xi), \\ \partial f(x) &= \{ \nabla (yF)(x) \mid y \in \partial g(F(x)) \}. \end{aligned}$$

In addition, there is a neighborhood U of $\bar{x}$ such that, relative to $U \times \mathbb{R}^n$, the set of points (x, v) with $x \in U$ and $v \in \partial f(x)$ is closed, and relative to this set, the mapping

$$(x, v) \mapsto Y(x, v) \;:=\; \{ y \mid \nabla(yF)(x) = v, \ y \in \partial g(F(x)) \},$$

is locally bounded with closed graph, hence in particular compact-valued, while the function $(x, v) \mapsto f(x)$ is continuous.

Proof. From Proposition 2.6 we know that f is fully amenable at every point $x \in C$ near enough to $\bar{x}$. All these properties, except for the very last (concerning continuity of f), are already understood to hold for any amenable function; see [19], [22]. Really, they only need F to be $\mathcal{C}^1$ and g to be lsc, proper, convex. The last property has been established in [14, Prop. 2.5] in the name of strongly amenable functions, but again the proof only requires amenability. □

Moving to second-order concepts, we work with second-order difference quotient functions which depend not only on a point x where f is finite but also on the choice of a subgradient $v \in \partial f(x)$, namely the functions

$$\Delta^2_{x,v,t} f : \xi \mapsto [f(x + t\xi) - f(x) - t\langle v, \xi\rangle]/\tfrac{1}{2}t^2 \text{ for } t > 0.$$

We say that f is *twice epi-differentiable* at $\bar{x}$ for a vector $\bar{v}$ if $f(\bar{x})$ is finite, $\bar{v} \in \partial f(\bar{x})$, and the functions $\Delta^2_{\bar{x},\bar{v},t} f$ epi-converge to a proper function as $t \searrow 0$. The limit is then the *second epi-derivative* function $f''_{\bar{x},\bar{v}} : I\!R^n \to \overline{I\!R}$; see [12], [19] and [21]. When $\partial f(\bar{x})$ is a singleton consisting of $\bar{v}$ alone, the notation $f''_{\bar{x},\bar{v}}$ can be simplified to $f''_{\bar{x}}$. The second epi-derivative function, when it exists, has to be lsc and positively homogeneous of degree 2, although not necessarily quadratic. Further, we call f *strictly* twice epi-differentiable at $\bar{x}$ for $\bar{v}$ if the stronger property holds that the functions $\Delta^2_{x,v,t} f$ epi-converge as $t \searrow 0$, $x \to \bar{x}$ with $f(x) \to f(\bar{x})$, and $v \to \bar{v}$ with $v \in \partial f(x)$.

It is important to appreciate that, because it is defined in terms of epi-convergence, second-order epi-differentiability is essentially a *geometric property of approximation of epigraphs*. This kind of approximation differs in general from the classical kind of approximation expressed by uniform convergence of functions on bounded sets, although key relationships can be detected in special situations. Such uniform convergence is not a viable concept for broad use in an environment like ours here. Circumstances where it does nicely come into play will be identified in Sections 4 and 5, where second-order "expansions" of f and its envelopes e_λ will be considered. For now, $f''_{\bar{x},\bar{v}}$ has to be thought of as providing a second-order approximation

$$f(\bar{x} + t\xi) \;\approx\; f(\bar{x}) + t\langle \bar{v}, \xi\rangle + \tfrac{1}{2}t^2 f''_{\bar{x},\bar{v}}(\xi),$$

not in the usual sense of local uniformity, but the closeness of $\operatorname{epi} \Delta^2_{\bar{x},\bar{v},t} f$ to $\operatorname{epi} f''_{\bar{x},\bar{v}}$.

A rather remarkable fact about second-order epi-differentiability was established in [19]: when f is fully amenable at $\bar{x}$, it is twice epi-differentiable there for every $\bar{v} \in \partial f(\bar{x})$. The widespread availability of this property in the context of optimization is what makes it especially interesting. Before looking at what the theory of second epi-derivatives tells us about the class of problems under consideration, we look at a parallel concept which turns out to be closely connected with this.

Second-order differentiation of f can be contemplated also in the framework of first-order differentiation of the subgradient mapping $\partial f : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ (where $\partial f(x)$ is always regarded as the empty set when $f(x)$ is not finite). For any set-valued mapping $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, one can work with difference quotient mappings $\Delta_{x,v,t}T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ associated with pairs (x, v) in the graph of T, namely

$$\Delta_{x,v,t}T : \xi \mapsto [T(x + t\xi) - v]/t \quad \text{for } t > 0.$$

The mapping T is said to be *proto-differentiable* at $\bar{x}$ for $\bar{v}$ if $\bar{v} \in T(\bar{x})$ and the mappings $\Delta_{\bar{x},\bar{v},t}T$ converge graphically to a mapping Δ as $t\searrow 0$, in which event the limit mapping is denoted by $T'_{\bar{x},\bar{v}}$ and called the *proto-derivative* of T at $\bar{x}$ for $\bar{v}$; see [13], [20], [22]. (Graph convergence of these mappings refers to the convergence of their graphs as subsets of $\mathbb{R}^n \times \mathbb{R}^n$.) We say that T is *strictly* proto-differentiable at $\bar{x}$ for $\bar{v}$ if in fact the mappings $\Delta_{x,v,t}T$ converge graphically to $T'_{\bar{x},\bar{v}}$ as $t\searrow 0$ and $(x, v) \to (\bar{x}, \bar{v})$ with $v \in T(x)$.

Again, a geometric notion of approximation is invoked. We have

$$T(\bar{x} + t\xi) \;\approx\; T(\bar{x}) + tT'_{\bar{x},\bar{v}}(\xi),$$

not with respect to some kind of uniform local bound on the difference, but in the sense that the set $\operatorname{epi} \Delta_{\bar{x},\bar{v},t}T$ can be made arbitrarily close to $\operatorname{epi} T'_{\bar{x},\bar{v}}$ (relative to the concepts of set convergence appropriate for unbounded sets) by taking the parameter $t > 0$ sufficiently small. The mapping $T'_{\bar{x},\bar{v}}$ assigns to each $\xi \in \mathbb{R}^n$ a subset $T'_{\bar{x},\bar{v}}(\xi)$ of $\mathbb{R}^n$, which could be empty for some choices of ξ. When $T(\bar{x})$ is a singleton consisting of $\bar{v}$ only (as for instance in the case where T is actually single-valued everywhere), the notation $T'_{\bar{x},\bar{v}}(\xi)$ can be simplified to $T'_{\bar{x}}(\xi)$.

In stating the next theorem, we continue the notation introduced in advance of Definition 2.5 by writing $\nabla^2(yF)$ for the matrix of second partial derivatives of the function $yF : x \mapsto \langle y, F(x)\rangle$. Then

$$\begin{aligned}\nabla^2(yF)(x) = y_1\nabla^2 f_1(x) + \cdots + y_m\nabla^2 f_m(x)\\ \text{when } F = (f_1, \dots, f_m) \text{ and } y = (y_1, \dots, y_m).\end{aligned}$$

Theorem 3.2. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g{\circ}F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any feasible solution to $(\mathcal{P})$ at which condition (CQ) holds. Then at all feasible solutions x in some neighborhood of $\bar{x}$, and for all subgradients $v \in \partial f(x)$,*

(a) *f is twice epi-differentiable at x for v,*

(b) *∂f is proto-differentiable at x for v,*

and the second epi-derivative function $f''_{x,v}$ and proto-derivative mapping $(\partial f)'_{x,v}$ are related to each other by

$$(\partial f)'_{x,v} \;=\; \partial(\tfrac{1}{2} f''_{x,v}).$$

Furthermore, one has the formula

$$f''_{x,v}(\xi) = \begin{cases} \max_{y \in Y(x,v)} \langle \xi, \nabla^2(yF)(x)\xi \rangle & \text{if } \xi \in \Xi(x,v), \\ \infty & \text{if } \xi \notin \Xi(x,v), \end{cases}$$

where $Y(x,v)$ is the compact subset of $\mathbb{R}^m$ defined in Theorem 3.1 and $\Xi(x,v)$ is the closed cone in $\mathbb{R}^n$ defined by

$$\Xi(x,v) = N_{\partial f(x)}(v) = \{\xi \mid f'_x(\xi) = \langle v, \xi \rangle\} = \{\xi \mid f'_x(\xi) \le \langle v, \xi \rangle\}.$$

In fact $Y(x,v)$ and $\Xi(x,v)$ are polyhedral, and in terms of the finite set $Y_{\text{ext}}(x,v)$ consisting of the extreme points of $Y(x,v)$ the second-order epi-derivative formula can be written as

$$f''_{x,v}(\xi) = \begin{cases} \max_{y \in Y_{\text{ext}}(x,v)} \langle \xi, \nabla^2(yF)(x)\xi \rangle & \text{if } \langle \nabla(yF)(x) - v, \xi \rangle \le 0 \text{ for all } y \in Y(x,v), \\ \infty & \text{otherwise.} \end{cases}$$

Proof. Once more we appeal to Proposition 2.6 for the observation that our hypothesis implies f is fully amenable at points x near enough to $\bar{x}$ with $f(x)$ finite. Then we apply the twice epi-differentiability result and formula of [19] with the proto-differentiability result and formula of [10]. This, in combination with the results in Theorem 3.1, takes care of all the assertions except those at the end relying on the polyhedral nature of $Y(x,v)$.

The fact that $Y(x,v)$ is polyhedral is obvious from its definition in Theorem 3.1 as the set of vectors $y \in \partial g(F(x))$ satisfying the linear equation $\nabla F(x)^T y = v$, since the subgradient set $\partial g(F(x))$ is itself polyhedral (due to g being polyhedral). Indeed, this has previously been observed in [11], [13]. For any fixed vector ξ the function $y \mapsto \langle \xi, \nabla^2(yF)(x)\xi \rangle$ is linear, so its maximum over $Y(x,v)$ has to be attained at one of the finitely many points of $Y_{\text{ext}}(x,v)$.

Because the set $\partial f(x)$ is the image of the polyhedral set $\partial g(F(x))$ under the linear mapping $y \mapsto \nabla(yF)(x) = \nabla F(x)^T y$, it is polyhedral as well. Then $\Xi(x,v)$ must be polyhedral, since it is the normal cone to $\partial f(x)$ at v. The definition of this normal cone characterizes $\Xi(x,v)$ as consisting of the vectors ξ such that $\langle v' - v, \xi \rangle \le 0$ for all $v' \in \partial f(x)$. Hence it consist of all ξ such that $\langle \nabla(yF)(x) - v, \xi \rangle \le 0$ for all $y \in \partial g(F(x))$. □

The last part of Theorem 3.2 reveals interestingly enough that the second epi-derivative function $f''_{x,v}$ has the same character as that ascribed to f itself in Theorem 2.7, although simpler. It is the max of finitely many $\mathcal{C}^2$ (actually quadratic) functions plus the indicator of a set defined by finitely many $\mathcal{C}^2$ (actually linear) constraints. Note again that just because we know that a set can in principle be expressed in terms of such constraints, this does not mean we can readily make use of such an expression. To write $\Xi(x,v)$ in terms of a *finite* system of linear constraints we would have to identify all extreme points and extreme rays of $\partial g(F(x))$. Depending on the circumstances, this might or might not be easy.

Additional formulas for the proto-derivative mapping $(\partial f)'_{x,v}$ can be developed from this description of $f''_{x,v}$ by following the lines in [13].

To see more closely what the results in Theorems 3.1 and 3.2 mean in common situations, we focus on two key cases, the ones in Examples 2.2 and 2.3 (as extended in Examples 2.2′ and 2.3′).

Example 3.3 [13, Thm. 2]**.** *In the problem of Example 2.2, consider any $x \in I\!R^n$ and let $I(x)$ denote the set of indices i such that $f_i(x) = f(x)$. Then f is epi-differentiable at x and has at least one subgradient there, with*

$$\partial f(x) = \operatorname{co}\{\nabla f_i(x) \,|\, i \in I(x)\}, \qquad f'_x(\xi) = \max_{i \in I(x)} \langle \nabla f_i(x), \xi \rangle.$$

Moreover, f is twice epi-differentiable at x for any subgradient $v \in \partial f(x)$, with the second-order epi-derivative function given by

$$f''_{x,v}(\xi) = \begin{cases} \displaystyle\max_{y \in Y_{\rm ext}(x,v)} \sum_{i=1}^m y_i \langle \xi, \nabla^2 f_i(x)\xi \rangle & \text{if } \langle \nabla f_i(x) - v, \xi \rangle \le 0 \text{ for all } i \in I(x), \\ \infty & \text{otherwise.} \end{cases}$$

where $Y_{\rm ext}(x,v)$ is the finite set of extreme points of the compact polyhedral set

$$Y(x,v) \;:=\; \Big\{ y \,\Big|\, y_i \ge 0 \text{ if } i \in I(x),\ y_i = 0 \text{ if } i \notin I(x), \\ \textstyle\sum_{i=1}^m y_i = 1,\ \sum_{i=1}^m y_i \nabla f_i(x) = v \Big\}.$$

Moving on now to the problem in Example 2.3, which subsumes the one in Example 2.1, we denote by $T_C(x)$ the *tangent cone* to C at a point $x \in C$, and similarly by $T_X(x)$ the tangent cone to the polyhedral set X at x. These tangent cones are polar to the normal cones $N_C(x)$ and $N_X(x)$ (because we are dealing with convex sets or more generally sets that are Clarke regular, for which the various definitions in use for tangent cones all agree). The tangent cone notation will be useful also in handling constraints: we denote by $T_{I_i}(u_i)$ the tangent cone to the closed interval $I_i \subset I\!R$ at $u_i \in I_i$, which simply indicates the directions in which one can move from u_i without leaving I_i. Specifically, in parallel with the formulas for the normal cones, in the case where I_i has lower bound a_i and upper bound b_i (these possibly being infinite, with $a_i \le b_i$), one has

$$T_{I_i}(f_i(x)) = \begin{cases} (-\infty, 0] & \text{when } a_i < f_i(x) = b_i, \\ [0, \infty) & \text{when } a_i = f_i(x) < b_i, \\ (-\infty, \infty) & \text{when } a_i < f_i(x) < b_i, \\ [0, 0] & \text{when } a_i = f_i(x) = b_i. \end{cases}$$

Example 3.4 [13, Thm. 4]**.** *For the problem of Example 2.3, consider any $\bar{x} \in C$ satisfying the constraint qualification described in Example 2.3′. Let*

$$L(x,y) = f_0(x) + y_1 f_1(x) + \cdots + y_m f_m(x).$$

For all $x \in C$ in some neighborhood of $\bar{x}$, f is epi-differentiable at x and has at least one subgradient there, with

$$\begin{aligned}
\partial f(x) &= \nabla f_0(x) + N_C(x) = \Big\{\nabla_x L(x,y) \,\Big|\, y_i \in N_{I_i}(f_i(x))\Big\} + N_X(x), \\
f'_x(\xi) &= \begin{cases} \langle \nabla f_0(x), \xi\rangle & \text{if } \xi \in T_X(x) \text{ and } \langle \nabla f_i(x), \xi\rangle \in T_{I_i}(f_i(x)) \text{ for all } i, \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Moreover f it twice epi-differentiable at x for every subgradient $v \in \partial f(x)$, with the second-order epi-derivative function given in terms of the Lagrangian L by

$$f''_{x,v}(\xi) = \max_{y \in Y(x,v)} \langle \xi, \nabla^2_{xx} L(x,y)\xi\rangle \;+\; \delta_{\Xi(x,v)}(\xi),$$

where $Y(x,v)$ is a compact polyhedral set and $\Xi(x,v)$ is a polyhedral cone, namely

$$\begin{aligned}
Y(x,v) &= \Big\{y \,\Big|\, y_i \in N_{I_i}(f_i(x)),\;\; v - \nabla_x L(x,y) \in N_X(x)\Big\}, \\
\Xi(x,v) &= \Big\{\xi \in T_C(x) \,\Big|\, \langle v - \nabla f_0(x), \xi\rangle = 0\Big\} \\
&= \Big\{\xi \in T_X(x) \,\Big|\, \langle \nabla f_i(x), \xi\rangle \in T_{I_i}(f_i(x)) \text{ for all } i,\; \langle v - \nabla f_0(x), \xi\rangle = 0\Big\}.
\end{aligned}$$

Here $Y(x,v)$ can be replaced in the max expression by its finite set of extreme points.

In Example 3.4 the function f_0 has been assumed to be $\mathcal{C}^2$, but the methodology is not limited to that case. We could easily go further by taking $f = f_0 + \delta_C$ with the set C chosen according to the specifications in Example 2.3, but with f_0 taken to be any fully amenable function. In particular, f_0 could be a max function of the kind in Examples 2.1 and 3.3, hence nonsmooth. This generality is attained through the calculus we have developed in [12], which provides formulas for $f''_{x,v}(\xi)$ and $(\partial f)'_{x,v}(\xi)$ when f is expressed as the sum of two fully amenable functions under an associated "constraint qualification" on the domains of the functions. For $f = f_0 + \delta_C$ this constraint qualification is satisfied in particular when f_0 is finite everywhere, as in the max function case. Then $\partial f(x) = \partial f_0(x) + N_C(x)$, and for any $v \in \partial f(x)$ one has in terms of the set

$$V(x,v) := \{(v_0, v_1) \,|\, v_0 \in \partial f_0(x),\; v_1 \in N_C(x),\; v_0 + v_1 = v\}$$

the expressions

$$\begin{aligned}
f''_{x,v}(\xi) &= \max_{(v_0,v_1)\in V(x,v)} \Big\{(f_0)''_{x,v_0}(\xi) + (\delta_C)''_{x,v_1}(\xi)\Big\}, \\
(\partial f)'_{x,v}(\xi) &= \bigcup_{(v_0,v_1)\in V_{\max}(x,v,\xi)} \Big\{(\partial f_0)'_{x,v_0}(\xi) + (\partial \delta_C)'_{x,v_1}(\xi)\Big\},
\end{aligned}$$

where $V_{\max}(x,v,\xi)$ is the set of vectors (v_0, v_1) that achieve the maximum.

The problem in Example 2.4 could likewise be handled by such calculus or tackled directly through Theorems 3.1 and 3.2.

The first- and second-order epi-derivatives that have been shown to exist for the general problems in composite format we have been considering can be used employed in particular in the statement of optimality conditions.

Theorem 3.5. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any feasible solution at which the condition (CQ) is satisfied. Let $Y(\bar{x}, 0) := \{y \in \partial g(F(\bar{x})) \mid \nabla(yF)(\bar{x}) = 0\}$, this being a compact, polyhedral convex set (possibly empty), and let $Y_{\text{ext}}(\bar{x}, 0)$ be its finite set of extreme points.*

(a) *If $\bar{x}$ is locally optimal, then $Y(\bar{x}, 0)$ must be nonempty, and*

$$\max_{y \in Y_{\text{ext}}(\bar{x})} \langle \xi, \nabla^2 (yF)(\bar{x}) \xi \rangle \geq 0 \ \text{ for } \xi \text{ satisfying}$$
$$\langle \nabla (yF)(\bar{x}), \xi \rangle \leq 0 \ \text{ for all } y \in \partial g(F(\bar{x})).$$

(b) *If $Y(\bar{x}, 0)$ is nonempty and*

$$\max_{y \in Y_{\text{ext}}(\bar{x},0)} \langle \xi, \nabla^2 (yF)(\bar{x}) \xi \rangle > 0 \ \text{ for } \xi \neq 0 \text{ satisfying}$$
$$\langle \nabla (yF)(\bar{x}), \xi \rangle \leq 0 \ \text{ for all } y \in \partial g(F(\bar{x})),$$

then $\bar{x}$ is locally optimal.

Proof. This applies the formulas of Theorems 3.1 and 3.2 to the general characterization of local optimality in terms of first- and second-order epi-derivatives in [19]. □

4 Hessians and second-order expansions

Pursuing second-order properties to a greater depth, we turn to the question of the existence of second-order expansions for f in the sense of locally uniform convergence of difference quotient functions rather than the epi-convergence employed so far. In this endeavor we draw on results from our paper [15]. Two definitions from this paper set the stage.

Definition 4.1. *A single-valued mapping G from an open neighborhood of $\bar{x} \in \mathbb{R}^n$ into $\mathbb{R}^m$ has a first-order expansion at a point $\bar{x} \in O$ if there is a continuous mapping D such the difference quotient mappings*

$$\Delta_{\bar{x},t} G : [G(\bar{x} + t\xi) - G(\bar{x})]/t \ \text{ for } t > 0$$

converge to D uniformly on bounded sets as $t \searrow 0$. The expansion is strict if actually the mappings

$$\Delta_{x,t} G : [G(x + t\xi) - G(x)]/t \ \text{ for } t > 0$$

converge to D uniformly on bounded sets as $t \searrow 0$ and $x \to \bar{x}$.

The existence of a first-order expansion means that G is *directionally differentiable* at $\bar{x}$: for every vector $\bar{\xi} \in \mathbb{R}^n$, the *directional derivative* limit

$$\lim_{\substack{\xi \to \bar{\xi} \\ t \searrow 0}} \frac{G(\bar{x} + t\xi) - G(\bar{x})}{t}$$

exists. The existence of a strict first-order expansion means that G is *strict* directional differentiable at $\bar{x}$; it corresponds to the existence for every $\bar{\xi}$ of the more complicated limit where $\bar{x}$ is replaced by x, and $x \to \bar{x}$ along with $\xi \to \bar{\xi}$ and $t \searrow 0$. In both cases the mapping D in Definition 4.1 gives for each $\bar{\xi}$ the directional derivative $D(\bar{\xi})$.

Definition 4.2. *Consider a function g on $\mathbb{R}^n$ and a point $\bar{x}$ where g is finite and differentiable.*

(a) *g has a second-order expansion at $\bar{x}$ if there is a finite, continuous function h such that the second-order difference quotient functions*

$$\Delta^2_{\bar{x},t} g(\xi) := [g(\bar{x}+t\xi) - g(\bar{x}) - t\langle \nabla g(\bar{x}), \xi\rangle]/\tfrac{1}{2}t^2$$

converge to h uniformly on bounded sets as $t \searrow 0$. The expansion is strict if g is differentiable not only at $\bar{x}$ but on a neighborhood of $\bar{x}$, and the functions

$$\Delta^2_{x,t} g(\xi) := [g(x+t\xi) - g(x) - t\langle \nabla g(x), \xi\rangle]/\tfrac{1}{2}t^2$$

converge to h uniformly on bounded sets as $t \searrow 0$ and $x \to \bar{x}$.

(b) *g has a Hessian matrix H at $\bar{x}$, this being a symmetric $n \times n$ matrix, if g has a second-order expansion with $h(\xi) = \langle \xi, H\xi\rangle$. The Hessian is strict if the expansion is strict.*

(c) *g is twice differentiable at $\bar{x}$ if its first partial derivatives exist on a neighborhood of $\bar{x}$ and are themselves differentiable at $\bar{x}$, i.e., the second partial derivatives of g exist at $\bar{x}$. Then $\nabla^2 g(\bar{x})$ denotes the matrix formed by these second partial derivatives.*

A second-order expansion in the sense of Definition 4.2 automatically requires the function h also to be positively homogeneous of degree 2: $h(\lambda\xi) = \lambda^2 h(\xi)$ for $\lambda > 0$, and in particular, $h(0) = 0$. It means that

$$g(\bar{x}+t\xi) \;=\; g(\bar{x}) + t\langle \nabla g(\bar{x}), \xi\rangle + \tfrac{1}{2}t^2 h(\xi) + o(t^2|\xi|^2)$$

for such a function h that is finite and continuous. The existence of a Hessian corresponds to h actually being quadratic.

The existence of a second-order expansion for an essential function f can be settled in a definitive manner on the basis of the second-order epi-derivative formula in Theorem 3.2 and a general result in our paper [14]. It is crucial for this purpose that strongly amenable functions f, such as we know we are dealing with now by virtue of Proposition 2.6, have a property called "prox-regularity," which we introduced in [14] (cf. Prop. 2.5 of that paper). This property is a typical hypothesis for most of the results of [14] and [15] that will be applied in what follows. Here we leave all discussion of it aside, jumping directly to the conclusions it supports.

Theorem 4.3. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any point of the feasible set $C = \operatorname{dom} f$ at which the condition (CQ) holds. Then for all x sufficiently close to $\bar{x}$ with $f(x)$ finite, the following properties are equivalent:*

(a) *f has a second-order expansion at x;*
(b) *f is differentiable at x;*
(c) *$\partial f(x)$ contains a solitary vector v;*
(d) *$\nabla(yF)(x)$ is the same vector v for all $y \in \partial g(F(x))$;*
(e) *$(\partial f)'_{x,v}(0) = \{0\}$ for some v.*

Under these circumstances necessarily $x \in \operatorname{int} C$ and $\nabla f(x) = v$, and the expansion of f takes the form

$$f(x + t\xi) \;=\; f(x) + \langle v, \xi\rangle + \tfrac{1}{2}t^2 \max_{y \in Y_{\rm ext}(x)} \langle \xi, \nabla^2(yF)(x)\xi\rangle + o(t^2|\xi|^2),$$

where $Y_{\rm ext}(x)$ is the set of extreme points of the compact, polyhedral convex set $\partial g(F(x))$, and the max expression also equals $f''_{x,v}(\xi)$.

Proof. In view of Proposition 2.6, condition (CQ) extends from $\bar{x}$ to all points x sufficiently near to $\bar{x}$ with $f(x)$ finite. It suffices therefore to argue the equivalences just at $\bar{x}$ itself.

If a second-order expansion exists at $\bar{x}$, f must in particular be differentiable at $\bar{x}$ and the function h expressing the second-order term must be the second epi-derivative function $f''_{\bar{x}}$, inasmuch as locally uniform convergence of difference quotient functions implies their epi-convergence. Conversely, if for any $\bar{v} \in \partial f(\bar{x})$ the function $f''_{\bar{x},\bar{v}}$ is finite, we obtain from [14, Thm. 6.7] (through the prox-regularity of f mentioned prior to the statement of the present theorem) that (b) and (c) hold with $\nabla f(\bar{x})$, and moreover that (a) holds with the second-order term in the expansion dictated by $h = f''_{\bar{x},\bar{v}}$. At this juncture we can apply the formula for $f''_{\bar{x},\bar{v}}$ in Theorem 3.2, which yields all the rest. In particular, (e) is obtained as an equivalent condition because $(\partial f)'_{\bar{x},\bar{v}}(0)$ consists of the subgradients of $\frac{1}{2}f''_{\bar{x},\bar{v}}$ at 0. The subgradient formula for this function (cf. [13]) indicates that the unique subgradient at the origin is 0 if and only if the cone $\Xi(\bar{x},\bar{v})$, which is the effective domain of $f''_{\bar{x},\bar{v}}$, has the origin in its interior, i.e., this cone is the whole space. □

When does the expansion in Theorem 4.3 correspond actually to a Hessian for f at $\bar{x}$? The following lemma will help answer this and a subsequent question as well.

Lemma 4.4. *Let Q_i, $i = 0, 1 \ldots, m$ be symmetric matrices in $\mathbb{R}^{n\times n}$, and let M be any subspace of $\mathbb{R}^n$ (perhaps $\mathbb{R}^n$ itself). Then in order to have the property*

$$\max_{i=1,\ldots,m} \langle \xi, Q_i\xi\rangle = \langle \xi, Q_0\xi\rangle \;\text{ for all } \xi \in M,$$

there must actually be an index $i_0 \in \{1, \ldots, m\}$ such that the quadratic forms associated with Q_{i_0} and Q_0 agree on M. In other words, there must exist i_0 such that

$$i_0 \in \operatorname*{argmax}_{i=1,\ldots,m} \langle \xi, Q_i\xi\rangle \;\text{ for all } \xi \in M.$$

Proof. We may assume without loss of generality that $M = \mathbb{R}^n$, since otherwise a change of coordinates can be employed to bring about a reduction to a space $\mathbb{R}^d$ with

$d < n$. For each $i \in \{1, \ldots, m\}$ let C_i denote the closed subset of $I\!R^n$ consisting of the points x where index i gives the max, i.e., where the quadratic function q_i associated with Q_i agrees with the quadratic function q_0 associated with Q_0. The union of these sets C_i is $I\!R^n$. By suppressing indices one by one as needed, we can come up with a collection indexed by $i \in I \subset \{1, \ldots, m\}$ such that the union of the C_i's for $i \in I$ is all of $I\!R^n$, but no subcollection has this property. Then every C_i for $i \in I$ must have nonempty interior, because it covers the complement of the (closed) union of all the other sets in this collection. The fact that q_i agrees with q_0 on the nonempty, open set $\operatorname{int} C_i$ implies $Q_i = Q_0$ (e.g., because the two functions q_i and q_0 have the same second derivatives there). Hence we have $Q_i = Q_0$ for all $i \in I$. □

Theorem 4.5. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any point of the feasible set $C = \operatorname{dom} f$ at which the condition (CQ) holds. Then for all x sufficiently close to $\bar{x}$ with $f(x)$ finite, the following properties are equivalent:*

(a) *f has a Hessian at x;*

(b) *f is differentiable at x, and the function f''_x is quadratic;*

(c) *$\partial f(x)$ is a singleton, and $(\partial f)'_x$ is single-valued everywhere and linear;*

(d) *there is a vector $\hat{y} \in Y_{\rm ext}(x)$ such that, for every $y \in Y_{\rm ext}(x)$, one has both $\nabla([y - \hat{y}]F)(x) = 0$ and $\nabla^2([y - \hat{y}]F)(x)$ positive semidefinite.*

Proof. The equivalence between (a), (b) and (d) is immediate from Theorem 4.3 and Lemma 4.4. Condition (c) comes into the picture because $(\partial f)'_x$ is the subgradient mapping for $\frac{1}{2} f''_x$ by Theorem 3.2, so it is linear if and only if f''_x is quadratic. □

These results make clear that the existence of a Hessian for f is quite a special property in our context. It corresponds to $f''_{x,v}$ being quadratic with v the unique element of $\partial f(x)$, and that only shows up in cases where constraints and first-order discontinuities are out of the immediate picture. However, there is an interesting concept to fall back on, which operates in wider territory.

Recall that a function $h : I\!R^n \to \overline{I\!R}$ is a *generalized (purely) quadratic function* if it is expressible in the form

$$h(\xi) = \begin{cases} \frac{1}{2}\langle \xi, Q\xi \rangle & \text{if } \xi \in M, \\ \infty & \text{if } \xi \notin M, \end{cases}$$

where M is a linear subspace of $I\!R^n$ and Q is a symmetric matrix in $I\!R^{n \times n}$. On the other hand, a possibly set-valued mapping $D : I\!R^n \rightrightarrows I\!R^m$ is a *generalized linear mapping* if its graph is a linear subspace of $I\!R^n \times I\!R^m$. The generalized quadratic functions are known to be precisely (up to an additive constant) the functions whose subgradient mappings are generalized linear mappings.

Let us think of f as having a *generalized Hessian* at x relative to a subgradient $v \in \partial f(x)$ if the second-order epi-derivative function $f''_{x,v}$ exists and is a generalized quadratic function. We do not want to push this terminology too far, since the concept reverts to approximation in the sense of epi-convergence rather than locally uniform convergence, but a certain case can be made for it, especially in view of the results that will be obtained in the next section in connection with envelope functions. The idea is that a generalized quadratic function h can be regarded as associated with a "generalized matrix for which some of the eigenvalues may be ∞," this being identified with a subspace M and an equivalence class of symmetric $n \times n$ matrices Q with respect to inducing the same quadratic form on M. These matrices all have the same eigenvalues relative to M; by an isometric change of coordinates that preserves the orthogonal decomposition of $\mathbb{R}^n$ into the sum of the subspaces M and $M^\perp$, they can simultaneously be reduced to the same diagonal matrix whose entries are these eigenvalues. We can simply regard $M^\perp$ as the eigenspace associated with the eigenvalue ∞.

These remarks are chiefly intended to be motivational, but the question of when $f''_{x,v}$ is a generalized quadratic function turns out to be important for a number of reasons. We proceed with putting together an answer. In this we denote by $\operatorname{ri} B$ the relative interior of a convex set B (in the sense of convex analysis [17]).

Theorem 4.6. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any point of the feasible set $C = \operatorname{dom} f$ at which the condition (CQ) holds. Then for all x sufficiently close to $\bar{x}$ with $f(x)$ finite, and all $v \in \partial f(x)$, the following properties are equivalent:*

(a) *$f''_{x,v}$ is generalized quadratic;*

(b) *$(\partial f)'_{x,v}$ is generalized linear;*

(c) *there exists $y \in \operatorname{ri} \partial g(F(x))$ such that $\nabla(yF)(x) = v$; further, there exists $\hat{y} \in Y_{\rm ext}(x, v)$ such that*

$$\langle \xi, \nabla^2([y - \hat{y}]F)(x)\xi \rangle \geq 0 \text{ for all } y \in Y_{\rm ext}(x, v) \text{ and } \xi \in \Xi(x, v),$$

where the notation is that of Theorem 3.2.

Proof. The equivalence between (a) and (b) is assured by the relation between $f''_{x,v}$ and $(\partial f)'_{x,v}$ in Theorem 3.2. For the equivalence between (a) and (c), we recall from Theorem 3.2 that, for all x in a neighborhood of $\bar{x}$, the domain of $f''_{x,v}$ is the normal cone to the convex set $\partial f(x)$ at v. The normal cone to a convex set is a subspace precisely when the point under consideration belongs to the interior of the set. Because $\partial f(x)$ is the image of the convex set $\partial g(F(x))$ under the linear transformation $y \mapsto \nabla F(x)^T y$ (by Theorem 3.1), its relative interior is the image under $\operatorname{ri} \partial g(F(x))$ under this transformation (cf. [17, Sec. 6]). Thus, the cone $\Xi(x, v)$ is a subspace if and only if $v = \nabla(yF)(x)$ for some $y \in \operatorname{ri} \partial g(F(x))$. It remains only to apply Lemma 4.4. □

The "generalized Hessian" case also arises in connection with strict second-order epi-differentiability of f.

Theorem 4.7. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any point of the feasible* *set $C = \operatorname{dom} f$ at which the condition (CQ) holds. Then for all x sufficiently close to $\bar{x}$ with $f(x)$ finite, and for all $v \in \partial f(x)$, the following properties are equivalent and imply in particular that $f''_{x,v}$ is a generalized quadratic function:*

(a) *f is strictly twice epi-differentiable at x for v;*

(b) *$f''_{x',v'}$ epi-converges (to something) as $(x',v') \to (x,v)$ in the set of pairs (x',v') with $v' \in \partial f(x')$ for which $f''_{x',v'}$ is generalized quadratic.*

Proof. This comes out of [15, Cor. 4.3] because of Theorem 3.2 and the prox-regularity of f consequent to the strong amenability in Proposition 2.6. □

A test of sorts for the case in Theorem 4.7, albeit a stringent one, is the following.

Proposition 4.8. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any point of the feasible set $C = \operatorname{dom} f$ at which the condition (CQ) holds. Suppose the function $f''_{\bar{x},\bar{v}}$ is generalized quadratic for a certain $\bar{v} \in \partial f(\bar{x})$, and for all points (x,v) near $(\bar{x},\bar{v})$ such that $f''_{x,v}$ is generalized quadratic denote by $Y_{\max}(x,v)$ the set of vectors $\hat{y}$ satisfying the associated condition in Theorem 4.7(c).*

Then a sufficient condition for f to be strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$ is that both $\Xi_{x,v} \to \Xi_{\bar{x},\bar{v}}$ and $Y_{\max}(x,v) \to Y_{\max}(\bar{x},\bar{v})$ as $x \to \bar{x}$ and $v \to \bar{v}$ in the set of pairs (x,v) with $v \in \partial f(x)$ for which $f''_{x,v}$ is generalized quadratic.

Proof. All we need to do, according to Theorem 4.7, is to show that $f''_{x,v}$ epi-converges to $f''_{\bar{x},\bar{v}}$ as $(x,v) \to (\bar{x},\bar{v})$ in the set of pairs (x,v) with $v \in \partial f(x)$ for which $f''_{x,v}$ is generalized quadratic. We first need to show that for all ξ

$$f''_{\bar{x},\bar{v}}(\xi) \le \liminf_{k\to\infty} f''_{x_k,v_k}(\xi_k)$$

whenever $\xi_k \to \xi$, $x_k \to \bar{x}$ and $v_k \to \bar{v}$ in the set of pairs (x_k,v_k) with $v_k \in \partial f(x_k)$ for which f_{x_k,v_k} is generalized quadratic. If $\xi_{x_k,v_k} \notin \Xi(x_k,v_k)$ for all k sufficiently large there is nothing to show. Assume not, then $\xi \in \Xi(\bar{x},\bar{v})$. Now consider $y \in Y_{\max}(\bar{x},\bar{v})$. Because $Y_{\max}(x,v) \to Y_{\max}(\bar{x},\bar{v})$, there exists $y_k \in Y_{\max}(x_k,v_k)$ with $y_k \to y$. It follows that $f''_{x_k,v_k}(\xi_k) = \langle \xi_k, \nabla^2(y_k F)(x_k)\xi_k\rangle$, and in the limit we get the desired inequality.

Finally we show that for all ξ there exist $\xi_{x,v} \to \xi$, as $x \to \bar{x}$ and $v \to \bar{v}$ in the set of pairs (x,v) with $v \in \partial f(x)$ for which $f_{x,v}$ is generalized quadratic, with

$$\limsup_{x\to\bar{x},\ v\to\bar{v}} \ f''_{x,v}(\xi_{x,v}) \le f''_{\bar{x},\bar{v}}(\xi).$$

If $\xi \notin \Xi(\bar{x},\bar{v})$ there is nothing to show. When $\xi \in \Xi(\bar{x},\bar{v})$ there exists $\xi_{x,v} \in \Xi(x,v)$ with $\xi_{x,v} \to \xi$. We have $f''_{x,v}(\xi_{x,v}) = \langle \xi_{x,v}, \nabla^2(y_{x,v}F)(x_{x,v})\xi_{x,v}\rangle$ for some $y_{x,v} \in Y_{\max}(x,v)$. We may assume that $y_{x,v} \to y$ with $y \in Y_{\max}(\bar{x},\bar{v})$; this is due to (CQ). In the limit we get the desired inequality. □

To what extent are these various properties realized in our examples? The case of a max function furnishes some good insights.

Proposition 4.9. *In the case of a function $f = \max\{f_1, \ldots, f_m\}$ in Example 2.2 (as continued in Examples 2.2′ and 3.3), consider any $x \in I\!R^n$ and any $v \in \partial f(x) = \operatorname{co}\{\nabla f_i(x) \mid i \in I(x)\}$.*

(a) *f has a second-order expansion at x if and only if the vectors $\nabla f_i(x)$ for $i \in I(x)$ coincide (or $I(x)$ is just a singleton). It has a Hessian at x if and only if, in addition, the matrices $\nabla^2 f_i(x)$ for $i \in I(x)$ coincide, this common matrix then being the Hessian matrix.*

(b) *$f''_{x,v}$ has a subspace for its effective domain $\Xi(x,v)$ if and only if one actually has $v \in \operatorname{ri}[\operatorname{co}\{\nabla f_i(x) \mid i \in I(x)\}]$, in which event*

$$\Xi(x,v) := \Big\{\xi \,\Big|\, \langle \nabla f_i(x) - v, \xi\rangle = 0 \text{ for all } i \in I(x)\Big\}.$$

For $f''_{x,v}$ to be generalized quadratic, it is necessary and sufficient to have, in addition, the existence of some $\hat{y}$ in the set

$$Y(x,v) = \Big\{y \,\Big|\, y_i \geq 0 \text{ if } i \in I(x),\ y_i = 0 \text{ if } i \notin I(x),\ \textstyle\sum_{i=1}^m y_i = 1,\ \sum_{i=1}^m y_i \nabla f_i(x) = v\Big\}$$

such that

$$\sum_{i=1}^m y_i \langle \xi, \nabla^2 f_i(x)\xi\rangle \geq \sum_{i=1}^m y'_i \langle \xi, \nabla^2 f_i(x)\xi\rangle \text{ for all } y' \in Y(x,v) \text{ and } \xi \in \Xi(x,v).$$

Proof. These results follow from Theorem 4.6 via Theorem 3.2. □

Strict twice epi-differentiability is harder to pin down in this example, but an elementary sufficient condition for it can readily be developed. Recall that a set of vectors $v_0, v_1 \ldots, v_s$ is affinely independent if the set $\{v_1 - v_0, \ldots, v_s - v_0\}$ is linearly independent.

Proposition 4.10. *For the max function in Proposition 4.8, suppose that*

(a) *the vectors $\nabla f_i(\bar{x})$ for $i \in I(\bar{x})$ are affinely independent, and*

(b) $\bar{v} \in \operatorname{ri}[\operatorname{co}\{\nabla f_i(\bar{x}) \mid i \in I(\bar{x})\}]$.

Then f is strictly twice differentiable at $\bar{x}$ for $\bar{v}$. Indeed in this case, for all (x,v) sufficiently close to $(\bar{x},\bar{v})$ with $v \in \partial f(x)$, the function $f''_{x,v}$ is generalized quadratic and depends epi-continuously on (x,v).

Proof. Let $\operatorname{gph} \partial f$ denote the graph of the mapping ∂f, i.e., the set of pairs (x,v) with $v \in \partial f(x)$. We first show that under our assumptions there is a neighborhood U of $(\bar{x},\bar{v})$ such that for all $(x,v) \in U \cap \operatorname{gph} \partial f$, we have $I(x) = I(\bar{x})$. Consider $x_k \to \bar{x}$ and $v_k \to \bar{v}$ with $v_k \in \partial f(x_k)$. We have $\sum_{i\in I(x_k)} (y_k)_i \nabla f_i(x_k) = v_k$ for some vector $y_k \in Y(x_k, v_k)$. Because $\sum_{i\in I(x_k)} (y_k)_i = 1$ and $(y_k)_i \geq 0$, we may assume that $(y_k)_i \to y_i$ (as $k \to \infty$). We may also assume (by taking a subsequence if necessary) that $I(x_k) = I^*$ for some subset I^* of $\{1, \ldots, m\}$. In the limit we have $\sum_{i\in I^*} y_i \nabla f_i(\bar{x}) = \bar{v}$. Then it follows from our assumptions that $I^* = I(\bar{x})$.

We next show that $Y(x, v)$ consists of only one vector when $\{\nabla f_i(x) \,|\, i \in I(x)\}$ is affinely independent. To see this, assume that

$$\sum_{i\in I(x)} y_i \nabla f_i(x) = v = \sum_{i\in I(x)} y_i' \nabla f_i(x)$$

for y and y' in $Y(x, v)$. This in turn means that

$$\sum_{i\in I(x)} (y_i - y_i')\nabla f_i(x) = 0 = \sum_{i\in I(x)} (y_i - y_i')\nabla f_1(x),$$

because $\sum_{i\in I(x)} y_i = 1 = \sum_{i\in I(x)} y_i'$. Therefore

$$\sum_{i\in I(x)} (y_i - y_i')(\nabla f_i(x) - \nabla f_1(x)) = 0,$$

which shows that $y_i = y_i'$ for all i.

It follows easily from the preceding observations that (a) and (b) are satisfied at $(x, v) \in U \cap \operatorname{gph} \partial f$. Also note that the arguments we have furnished show that for all $(x, v) \in U \cap \operatorname{gph} \partial f$ we have $y' \to y$ as $x' \to x$ and $v' \to v$ where $y' = Y(x', v')$ and $y = Y(x, v)$. We know then from Proposition 4.8 that $f''_{x,v}$ is a generalized quadratic for all $(x, v) \in U \cap \operatorname{gph} \partial f$ (inasmuch as $Y(x, v)$ is a singleton).

Finally we demonstrate that for all $(x, v) \in \operatorname{gph} \partial f$ in a neighborhood of $(\bar{x}, \bar{v})$ the function f is strictly twice epi-differentiable at x for v. We know that $I(x') = I(\bar{x})$ for all $(x', v') \in U \cap \operatorname{gph} \partial f$. Fix $(x, v) \in U \cap \operatorname{gph} \partial f$. Because the set $\{(\nabla f_i(x') - v') \,|\, i \in I(\bar{x})\}$ is affinely independent, we have $\Xi(x', v') \to \Xi(x, v)$ as $x' \to x$ and $v' \to v$ with $v' \in \partial f(x')$. Recall that $Y_{\max}(x', v') \to Y_{\max}(x, v)$. We now apply Proposition 4.8, and this completes the proof. □

The condition in Proposition 4.10 is so powerful that it guarantees not only the strict second-order epi-differentiability of f at $\bar{x}$ for $\bar{v}$ but the same also for all (x, v) near $(\bar{x}, \bar{v})$ in the graph of ∂f. It is hard to come up with a tractable condition for strict second-order epi-differentiability that is more modest in its consequences. The following example does show, however, that a max of finitely many $\mathcal{C}^2$ functions can be strictly twice epi-differentiable at a point $\bar{x}$ (actually here a point of global minimum) without necessarily being strictly twice epi-differentiable at nearby points.

Example 4.11. *Let $f_1(x_1, x_2) := {x_1}^3{x_2}^2$ and $f_2(x_1, x_2) := -f_1(x_1, x_2)$. Consider*

$$f(x_1, x_2) := |f_1(x_1, x_2)| = \max\{f_1(x_1, x_2), f_2(x_1, x_2)\}.$$

This function f is $\mathcal{C}^1$ (in fact it is both $\mathcal{C}^{1+}$ (differentiable with locally Lipschitz continuous gradient mapping) and lower-C^2), and it is strictly twice epi-differentiable at $\bar{x} = (0, 0)$, yet it does not have this property at points of the x_1-axis away from the origin.

Detail. The functions f_1 and f_2 agree on the x_1- and x_2-axes, with $\nabla f_i(x_1, x_2) = (0, 0)$ there for $i = 1, 2$. This shows that f is $\mathcal{C}^{1+}$ as well as lower-$\mathcal{C}^2$, and in particular

$\mathcal{C}^1$. Furthermore, f has a global minimum at $\bar{x} = (0,0)$, where both f_1 and f_2 the null matrix as their Hessian. We therefore have $f''_{(0,0),(0,0)}(\xi) = 0$ for all ξ by Theorem 3.2, so the function $f''_{(0,0),(0,0)}$ is quadratic. At a general point x not on the x_1- or x_2-axes, $f''_{x,v}$ is the quadratic associated with the Hessian of f_1 or f_2. For points with $x_1 = 0$, the second-order epi-derivative likewise has the property that $f''_{(0,x_2),(0,0)}(\xi) = 0$ for all ξ. But when $x_2 = 0$ we have $\langle \xi, \nabla^2 f_1(x_1,0)\xi\rangle = 2x_1^3{\xi_2}^2$ and $\langle \xi, \nabla^2 f_2(x_1,0)\xi\rangle = -2x_1^3{\xi_2}^2$, so that except for the origin, f is not twice differentiable at such a point nor strictly twice epi-differentiable there. Instead, $f''_{(x_1,0),(0,0)}(\xi) = \max\{2x_1^3{\xi_2}^2, -2x_1^3{\xi_2}^2\} = |2x_1^3{\xi_2}^2|$ for all $\xi = (\xi_1,\xi_2)$. The formulas we have identified for the second-order epi-derivative show that $f''_{x,\nabla f(x)}$ converges uniformly on bounded sets to $f''_{(0,0),(0,0)}$ as $x \to 0$; in particular they epi-converge. Hence by Theorem 4.7, f is strictly twice epi-differentiable at $(0,0)$ for $(0,0)$. □

We now turn our attention to Example 2.3, where $f(x) = f_0(x) + \delta_C(x)$ with f_0 smooth. Adopting the terminology of [1], we say in this setting that a pair (x,v) for $v \in \partial f(x)$ furnishes a *nondegenerate stationary point* (relative to the problem of minimizing $f - \langle v, \cdot\rangle$ in $\mathbb{R}^n$) if $v - \nabla f_0(x) \in \operatorname{ri} N_C(x)$.

Proposition 4.12. *In Example 2.3, consider any point $x \in C$ where the constraint qualification is satisfied (as characterized in Example 2.3′), and let $v \in \partial f(x)$, which is equivalent to $v - \nabla f_0(x) \in N_C(x)$. Then*

(a) *the effective domain $\Xi(x,v)$ of $f''_{x,v}$ is a subspace if and only if (x,v) furnishes a nondegenerate stationary point, in which event*

$$\begin{aligned}\Xi(x,v) &= \Big\{\xi \in T_C(x) \,\Big|\, \langle v - \nabla f_0(x), \xi\rangle = 0\Big\}\\ &= \Big\{\xi \in T_X(x) \,\Big|\, \langle \nabla f_i(x), \xi\rangle \in T_{I_i}(f_i(x)) \text{ for all } i,\ \langle v - \nabla f_0(x), \xi\rangle = 0\Big\};\end{aligned}$$

(b) *$f''_{x,v}$ is a generalized quadratic function if and only if, in addition, there is a multiplier vector $\hat{y}$ in the set*

$$Y(x,v) = \Big\{y \,\Big|\, y_i \in N_{I_i}(f_i(x)),\ v - \nabla_x L(x,y) \in N_X(x)\Big\}$$

with the property that

$$\Big\langle \xi, \nabla^2_{xx} L(x,y)\xi\Big\rangle \le \Big\langle \xi, \nabla^2_{xx} L(x,\hat{y})\xi\Big\rangle \text{ for all } y' \in Y(x,v) \text{ and } \xi \in \Xi(x,v).$$

Proof. This result follows from Example 3.4 and Theorem 4.6. Note that from Example 3.4 we do have $\partial f(x) = \nabla f_0(x) + N_C(x)$, and therefore $v \in \operatorname{ri} \partial f(x)$ if and only if $v - \nabla f_0(x) \in \operatorname{ri} N_C(x)$, i.e., (x,v) is a nondegenerate stationary point. □

Proposition 4.13. *In Example 2.3, consider any $\bar{x} \in C$ with $\bar{v} \in \nabla f_0(\bar{x}) + N_C(\bar{x})$. Assume that*

(a) *$(\bar{x},\bar{v})$ furnishes a nondegenerate stationary point,*

(b) *$\{\nabla f_i(\bar{x}) \,|\, f_i(\bar{x}) \notin \operatorname{ri} I_i\}$ is linearly independent,*

(c) $X = I\!R^n$.

Then for all (x, v) in a neighborhood of $(\bar{x}, \bar{v})$ with $v \in \partial f(x)$ the function f is strictly twice epi-differentiable at x for v, and in particular $f''_{x,v}$ is generalized quadratic.

Proof. The line of proof is very similar to that of Proposition 4.10. First notice that there exists a neighborhood U of $(\bar{x}, \bar{v})$ such that for all $(x, v) \in U \cap \operatorname{gph} \partial f$ we must have $\{\nabla f_i(x) \mid f_i(x) \notin \operatorname{ri} I_i\}$ linearly independent. Next notice that we may also assume that

$$\{i \mid f_i(x) \in \operatorname{ri} I_i\} = \{i \mid f_i(\bar{x}) \in \operatorname{ri} I_i\}$$

when $(x, v) \in U \cap \operatorname{gph} \partial f$. This is because $\bar{v} - \nabla f_0(\bar{x}) \in \operatorname{ri} N_C(\bar{x})$, where

$$N_C(\bar{x}) = \{\nabla_x L(\bar{x}, y) \mid y_i \in N_{I_i}(f_i(\bar{x}))\}$$

(recall that $L(x, y) = f_0(x) + \sum y_i f_i(x)$). From this it follows that $Y(x, v)$ is a singleton for all $(x, v) \in U \cap \operatorname{gph} \partial f$. We now easily conclude that $Y_{\max}(x, v) \to Y_{\max}(\bar{x}, \bar{v})$ and $\Xi(x, v) \to \Xi(\bar{x}, \bar{v})$ when $x \to \bar{x}$ and $v \to \bar{v}$ with $v \in \partial f(x)$. To finish off, we apply Proposition 4.8. □

5 Proximal mappings and envelopes

From now on we concentrate on the envelope functions e_λ and proximal mappings P_λ defined at the end of Section 2 in association with a function f. We continue to take f to be the essential objective function for problem in composite format. Mainly we concentrate henceforth on the case of minimizing points $\bar{x} \in \operatorname{argmin} f$. Such points have $\bar{v} = 0$ as a subgradient: $0 \in \partial f(\bar{x})$ by Theorem 3.5.

First on the agenda is the specialization to this context of a selection of facts from [14] and [15]. (The interested reader should consult these papers for many other results.)

Theorem 5.1. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any optimal solution at which the condition (CQ) is satisfied.*

Then for each $\lambda > 0$ sufficiently small, there is a neighborhood of $\bar{x}$ on which the function e_λ is $\mathcal{C}^{1+}$ and lower-$\mathcal{C}^2$, the mapping P_λ is single-valued and Lipschitz continuous, and

$$\begin{aligned} \nabla e_\lambda &= \lambda^{-1}[I - P_\lambda] = [\lambda I + (\partial f)^{-1}]^{-1}, \\ P_\lambda &= (I + \lambda \partial f)^{-1} \text{ with } P_\lambda(\bar{x}) = \bar{x}. \end{aligned}$$

Proof. We invoke [14, Thms. 4.4, 4.6, 5.2], making the observation, as above, that our assumptions entail through Proposition 2.6 that f has the prox-regularity demanded in those theorems. □

Functions that are $\mathcal{C}^{1+}$ have been the focus of much research recently. The reader interested in the study of generalized second-order directional derivatives and Hessians of these functions will surely want to consult the work of Cominetti and Correa [3], Hiriart-Urruty [4], Jeyakumar and Yang [5], Páles and Zeidan [9], and Yang and Jeyakumar [23]. Note that here the function e_λ is not only $\mathcal{C}^{1+}$ but also lower-$\mathcal{C}^2$.

Theorem 5.2. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g\circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any optimal solution at which condition (CQ) is satisfied (so that $0 \in \partial f(\bar{x})$ in particular), and for $\lambda > 0$ define*

$$d_\lambda(\xi) \;=\; \min_{\xi'}\left\{\frac{1}{2}f''_{\bar{x},0}(\xi') + \frac{1}{2\lambda}|\xi' - \xi|^2\right\} \quad \text{for all } \xi.$$

Then for all λ sufficiently small the function d_λ is both $\mathcal{C}^{1+}$ and lower-$\mathcal{C}^2$, the gradient mapping ∇d_λ being Lipschitz continuous globally, and the following properties hold:

(a) *e_λ has a second-order expansion at $\bar{x}$, given by*

$$e_\lambda(\bar{x} + t\xi) \;=\; e_\lambda(\bar{x}) + t^2 d_\lambda(\xi) + o(|t\xi|^2),$$

(b) *∇e_λ has a first-order expansion at $\bar{x}$, given by*

$$\nabla e_\lambda(\bar{x} + t\xi) \;=\; t\nabla d_\lambda(\xi) + o(|t\xi|),$$

(c) *P_λ has a first-order expansion at $\bar{x}$, given by*

$$P_\lambda(\bar{x} + t\xi) \;=\; \bar{x} + t[I - \lambda\nabla d_\lambda(\xi)] + o(|t\xi|).$$

Proof. This time we apply [15, Thm. 3.5], again utilizing the prox-regularity of f furnished through Proposition 2.6. □

Theorem 5.3. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g\circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any optimal solution at which the condition (CQ) is satisfied. Then for every $\lambda > 0$ sufficiently small, the following properties are equivalent and necessarily involve the same matrix H_λ:*

(a) *e_λ has a Hessian matrix H_λ at $\bar{x}$;*
(b) *∇e_λ is differentiable at $\bar{x}$ with Jacobian matrix H_λ;*
(c) *e_λ is twice differentiable at $\bar{x}$, with $H_\lambda = \nabla^2 e_\lambda(\bar{x})$;*
(d) *P_λ is differentiable at $\bar{x}$ with Jacobian matrix $I - \lambda H_\lambda$;*
(e) *$f''_{\bar{x},0}$ is generalized quadratic.*

Proof. This goes back to [15, Thm. 3.8], once more under the prox-regularity that our hypothesis guarantees. □

Theorem 5.4. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any optimal solution at which the condition (CQ) is satisfied. Then for every $\lambda > 0$ sufficiently small, the following properties are equivalent:*

(a) *f is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$;*

(b) *e_λ has a strict Hessian at $\bar{x}$;*

(c) *∇e_λ is strictly differentiable at $\bar{x}$;*

(d) *e_λ is twice differentiable at $\bar{x}$, and $\nabla^2 e_\lambda(x) \to \nabla^2 e_\lambda(\bar{x})$ as $x \to \bar{x}$ in the set of points x where e_λ is twice differentiable;*

(e) *e_λ is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$;*

(f) *∇e_λ is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$;*

(g) *P_λ is strictly differentiable at $\bar{x}$;*

(h) *P_λ is strictly proto-differentiable at $\bar{x}$;*

Proof. This quotes [15, Thms. 4.1,4.2] in the environment of the prox-regularity of f that comes from Proposition 2.6. □

Theorem 5.5. *Let f be the essential objective function in problem $(\mathcal{P})$, with $f = g \circ F$ for a $\mathcal{C}^2$ mapping F and a polyhedral function g. Let $\bar{x}$ be any optimal solution at which the condition (CQ) is satisfied. Then for every $\lambda > 0$ sufficiently small, the following properties are equivalent:*

(a) *e_λ is $\mathcal{C}^2$ on a neighborhood of $\bar{x}$;*

(b) *P_λ is $\mathcal{C}^1$ on a neighborhood of $\bar{x}$;*

(c) *For all (x, v) near to $(\bar{x}, \bar{v})$ in the graph of ∂f, f is twice epi-differentiable, $f''_{x,v}$ is generalized quadratic, and $f''_{x,v}$ depends epi-continuously on (x, v), i.e., $f''_{x',v'}$ epi-converges to $f''_{x,v}$ as $(x', v') \to (x, v)$ with $v' \in \partial f(x')$.*

Proof. We appeal here to [15, Thm. 4.4]. □

Corollary 5.6. *In the case of Theorem 5.5 where f happens to be differentiable at $\bar{x}$, or merely if it satisfies a local growth condition of type $f(x) \leq f(\bar{x}) + s|x - \bar{x}|^2$, properties (a) and (b) hold if and only if f is itself $\mathcal{C}^2$ on a neighborhood of $\bar{x}$.*

Proof. The additional assumption forces $f''_{\bar{x},\bar{v}}$ to be finite (cf. Theorem 4.3), and the property in (c) of Theorem 5.5 reduces then to f being $\mathcal{C}^2$; see also [15](Cor. 4.5). □

Example 5.7. *For the function f of Example 2.2, the assumptions of Proposition 4.8 and Theorem 5.5 ensure the presence of properties (a) and (b) of Theorem 5.5.*

Example 5.8. *For the function f of Example 2.3, the assumptions of Proposition 4.13 and Theorem 5.5 ensure the presence of properties (a) and (b) of Theorem 5.5.*

References

[1] J. V. Burke, On the identification of active constraints II: the nonconvex case, *SIAM J. Numer. Anal.* 27 (1990), 1081–1102.

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

[3] R. Cominetti and R. Correa, A generalized second-order derivative in nonsmooth optimization, *SIAM J. Control and Optimization* 28 (1990), 789–809.

[4] J.-B. Hiriart-Urruty, Characterization of the plenary hull of the generalized Jacobian matrix, *Math. Prog. Study* 17 (1982), 1–12.

[5] V. Jeyakumar and X. Q. Yang, Second-order analysis of $C^{1,1}$ functions and convex composite minimization, preprint 1992.

[6] C. Lemaréchal and C. Sagastizábal, Practical aspects of the Moreau-Yosida regularization I: Theoretical properties, preprint 1994.

[7] A. Levy and R. T. Rockafellar, Variational conditions and the proto-differentiation of partial subgradient mappings, *Nonlinear Anal. Th. Meth. Appl.* submitted.

[8] P. D. Loewen, *Optimal Control via Nonsmooth Analysis*, CRM Proceedings & Lecture Notes 2, AMS, 1993.

[9] Z. Páles and V. Zeidan, Generalized Hessian for $\mathcal{C}^{1+}$ functions in infinite dimensional normed spaces, preprint 1994.

[10] R. A. Poliquin, An extension of Attouch's Theorem and its application to second-order epi-differentiation of convexly composite functions, *Trans. Amer. Math. Soc.* 332 (1992), 861–874.

[11] R. A. Poliquin and R. T. Rockafellar, Amenable functions in optimization, *Nonsmooth Optimization Methods and Applications*, F. Giannessi (ed.), Gordon & Breach, Philadelphia, 1992, 338–353.

[12] R. A. Poliquin and R. T. Rockafellar, A calculus of epi-derivatives applicable to optimization, *Canadian J. Math.* 45 (4) (1993), 879–896.

[13] R. A. Poliquin and R. T. Rockafellar, Proto-derivative formulas for basic subgradient mappings in mathematical programming, *Set-Valued Analysis* 2 (1994), 275–290.

[14] R. A. Poliquin and R. T. Rockafellar, Prox-regular functions in variational analysis, preprint October 1994.

[15] R. A. Poliquin and R. T. Rockafellar, Generalized Hessian properties of regularized nonsmooth functions, preprint November 1994.

[16] L. Qi, Second-order analysis of the Moreau-Yosida approximation of a convex function, preprint 1994.

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

[18] R. T. Rockafellar, Maximal monotone relations and the second derivatives of nonsmooth functions, *Ann. Inst. H. Poincaré: Analyse Non Linéaire* 2 (1985), 167–184.

[19] R. T. Rockafellar, First- and second-order epi-differentiability in nonlinear programming, *Trans. Amer. Math. Soc.* 307 (1988), 75–107.

[20] R. T. Rockafellar, Proto-differentiability of set-valued mappings and its applications in optimization, *Analyse Non Linéaire*, H. Attouch et al. (eds.), Gauthier-Villars, Paris (1989), 449–482.

[21] R. T. Rockafellar, Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives, *Math. of Oper. Research* 14 (1989), 462–484.

[22] R. T. Rockafellar, Generalized second derivatives of convex functions and saddle functions, *Trans. Amer. Math. Soc.* 320 (1990), 810–822.

[23] X. Q. Yang and V. Jeyakumar, Generalized second-order directional derivatives and optimization with $C^{1,1}$ functions, *Optimization* 26 (1992), 165–185.