Control and Cybernetics

vol. **31** (2002) No. 3

Weak sharp minima revisited
Part I: basic theory

by

James V. Burke[1] **and Sien Deng**[2]

[1] Department of Mathematics, Box # 354350, University of Washington, Seattle, Washington 98195–4350, USA
E-mail: burke@math.washington.edu

[2] Department of Mathematical Sciences, Northern Illinois University, DeKalb, Illinois 60115, USA
E-mail: deng@math.niu.edu.

Abstract: The notion of sharp minima, or strongly unique local minima, emerged in the late 1970's as an important tool in the analysis of the perturbation behavior of certain classes of optimization problems as well as in the convergence analysis of algorithms designed to solve these problems. The work of Cromme and Polyak is of particular importance in this development. In the late 1980's Ferris coined the term *weak sharp minima* to describe the extension of the notion of sharp minima to include the possibility of a non-unique solution set. This notion was later extensively studied by many authors. Of particular note in this regard is the paper by Burke and Ferris which gives an extensive exposition of the notion and its impact on convex programming and convergence analysis in finite dimensions. In this paper we build on the work of Burke and Ferris. Specifically, we generalize their work to the normed linear space setting, further dissect the normal cone inclusion characterization for weak sharp minima, study the asymptotic properties of weak sharp minima in terms of associated recession functions, and give new characterizations for local weak sharp minima and boundedly weak sharp minima. This paper is the first of a two part work on this subject. In Part II, we study the links between the notions of weak sharp minima, bounded linear regularity, linear regularity, metric regularity, and error bounds in convex programming. Along the way, we obtain both new results and reproduce many existing results from a fresh perspective.

Keywords: weak sharp minima, boundedly weak sharp minima, recession function, recession cone, duality, normal cone inclusion, error bounds

This author's research is supported by the National Science Foundation Grant No. DMS-

1. Introduction

Let X be a normed linear space, consider nonempty closed convex sets $\bar{S} \subset S \subset X$, and let $f: X \mapsto \overline{\mathbb{R}}$ be a lower semi-continuous convex function for which $S \cap \operatorname{dom}(f) \neq \emptyset$ where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ and

$$\operatorname{dom}(f) = \{x \in X \mid f(x) < \infty\}.$$

The set $\bar{S} \subset X$ is said to be a set of *weak sharp minima* for the function f over the set S with modulus $\alpha > 0$ if

$$f(x) \geq f(\bar{x}) + \alpha \operatorname{dist}(x \mid \bar{S}) \tag{1}$$

for all $\bar{x} \in \bar{S}$ and $x \in S$, where the distance function $\operatorname{dist}(\cdot \mid \bar{S})$ is defined by

$$\operatorname{dist}(x \mid S) = \inf_{\bar{x} \in S} \|x - \bar{x}\|,$$

and $\|\cdot\|$ denotes the norm on X. Note that since $S \cap \operatorname{dom}(f) \neq \emptyset$ we have

$$\bar{S} = \arg\min_S f \subset \operatorname{dom}(f),$$

where

$$\arg\min_S f = \left\{x \in S \mid f(x) = \min_{y \in S} f(y)\right\}.$$

The notion of weak sharp minima is a generalization of the notion of sharp minima due to Polyak (1979) to include the possibility of a non-unique solution set. Sharp minima are also referred to as strongly unique local minima in the independent work of Cromme (1978). Polyak's work focuses on the case where X is finite-dimensional and $\bar{S}$ is a singleton (also see Polyak, 1987). The terminology of *weak sharp minima* was introduced by Ferris (1988), where it is extensively developed. The primary motivations for this study are the impact this notion has on sensitivity analysis (Burke, Lewis and Overton, 2000, 2001, Henrion and Outrata, 2001, Jourani, 2000, Lewis and Pang, 1996, Ye, 1998, Ye and Zhu, 1995, 1997, Ye, Zhu and Zhu, 1997) and on the convergence analysis of a wide range of optimization algorithms (Burke and Ferris, 1993, 1995, Burke and Moré, 1988, Cromme, 1978, Ferris, 1990, 1991, Li and Wang, 2002). For example, many optimization algorithms exhibit finite termination at weak sharp minima (Burke, Ferris, 1993, Ferris, 1990, Ferris, 1991).

The notion of weak sharp minima defined above (1) specifies first-order growth of the objective function away from the set of optimal solutions. Weak sharp minima of higher order growth are also of interest in parametric optimization, and lead to Hölder continuity properties of the associated solution mappings. Bonnans and Ioffe (1995) studied sufficient conditions and characterizations for weak sharp minima of order two in the case where X is finite-convex functions. Studniarski and Ward (1999) obtained some sufficient conditions and characterizations for weak sharp local minimizers of order m in terms of the limiting proximal normal cone and a generalization of the contingent cone.

The inequality (1) bounds the distance between the vector x and the set $\bar{S}$ by a residual function of the form $(f(x) - f(\bar{x}))/\alpha$. In this regard, the notion of weak sharp minima can be interpreted as a type of error bound. The study of error bounds has drawn much attention during recent years, due to their importance in the treatment of convergence analysis of iterative solution methods in optimization. We refer the reader to the recent issue of *Mathematical Programming* devoted to this topic, Luo and Pang (2000), as well as the review article Pang (1997) for an introduction to the vast literature on this subject. The connections between the notion of weak sharp minima and error bounds are made explicitly in Part II of this work where error bounds for convex inclusions and systems of convex inequalities are shown to be easily derivable from results for weak sharp minima.

The work in this paper builds on the earlier work of Burke and Ferris (1993) by extending and refining their results in a number of ways while weakening some of the underlying assumptions. The first point of departure from Burke and Ferris (1993) is that we develop the theory in infinite dimensions. This requires a somewhat more subtle use of duality techniques and facilitates an understanding of the underlying geometry by stripping away all compactness arguments. Secondly, we observe that, of many characterizations for weak sharp minima derived, the one which provides the closest point of contact with applications can be further dissected into two independent conditions. This observation has important consequences for each of the applications studied in Part II. A third point of departure from Burke and Ferris (1993) is our study of the asymptotic properties of weak sharp minima in terms of associated recession functions and recession cones. Again, these global asymptotic properties have important implications for applications. Fourthly, we study local notions of weak sharp minima. In the infinite dimensional setting one finds that some of the global characterizations for weak sharp minima do not carry over to the local setting. Nonetheless, we are able to provide a number of positive results. We also introduce the notion of *boundedly weak sharp minima*. This notion is equivalent to local weak sharp minima in finite dimensions, but is distinct in infinite dimensions. This distinction is useful in the applications studied in Part II. We conclude Part I with a reduction theorem that shows how to reduce a constrained weak sharp minima problem into an unconstrained one when f possesses Lipschitzian properties.

The notation that we employ is for the most part the same as that in Aubin and Ekeland (1984), Rockafellar (1970 and 1974). A partial list is provided below for the reader's convenience.

We denote the dual space of X by X^*. When X is endowed with the weak topology and X^* with the weak* topology then the spaces X and X^* are said on $X^* \times X$, Rockafellar (1974). We denote the norm on X^* by $\|\cdot\|_\circ$:

$$\|z\|_\circ = \sup_{x \in \mathbb{B}} \langle z, x \rangle,$$

where $\mathbb{B} = \{x \in X \mid \|x\| \leq 1\}$ is the unit ball in X.

For a nonempty subset C of any normed linear space Y, we denote the norm (or strong) closure of C and weak closure of C by $\mathrm{cl}(C)$ and $\text{w-cl}(C)$, respectively, and we denote the indicator function of C and the support function of C by $\psi_C(\cdot)$ and $\psi_C^*(\cdot)$, respectively. Thus, in particular, $\|z\|_o = \psi_{\mathbb{B}}^*(z)$. We denote the norm-topology interior of C by $\mathrm{int}\,(C)$, and the boundary of C by $\mathrm{bdry}(C) = C \setminus \mathrm{int}\,(C)$. When Y is finite-dimensional, $\mathrm{ri}\,(C)$ denotes the interior of C relative to the smallest affine set containing C. The cone generated by C is denoted by $\mathrm{cone}(C) = \cup_{\lambda \geq 0}\{\lambda C\}$.

For a closed set C in X, we define the *projection* of a point $x \in X$ onto the set C, denoted $P(x \mid C)$, as the set of all points in C that are closest to x as measured by the norm $\|\cdot\|$:

$$P(x \mid C) = \{y \in C \mid \|x - y\| = \inf\{\|x - u\| \mid u \in C\}\}.$$

For nonempty sets $C \subset X$ and $S \subset X^*$, we define the polar of C and the polar of S to be the sets

$$C^\circ = \{x^* \in X^* \mid \langle x^*, x \rangle \leq 1 \, \forall x \in C\},$$
$$S^\circ = \{x \in X \mid \langle x^*, x \rangle \leq 1 \, \forall x^* \in S\},$$

respectively. Thus, in particular, $\mathbb{B}^\circ \subset X^*$ is the unit ball associated with the dual norm $\|\cdot\|_o$. For a nonempty closed convex set C in X, and $x \in C$, we define the tangent cone to the set C at x, denoted by $T_C(x)$, as follows

$$T_C(x) = \mathrm{cl}\left(\bigcup_{t>0} \frac{C - x}{t}\right).$$

The normal cone to C at x is defined dually by the relation

$$N_C(x) = T_C(x)^o.$$

It is easy to see that

$$N_C(x) = \{x^* \in X^* | \langle x^*, y - x \rangle \leq 0, \quad \text{for any } y \in C\}.$$

Let $f\colon X \mapsto \overline{\mathbb{R}}$ be a lower semi-continuous convex function. The function $f^*\colon X^* \mapsto \overline{\mathbb{R}}$ defined by

$$f^*(x^*) = \sup_{x \in X}(\langle x^*, x \rangle - f(x))$$

is called the convex conjugate of f. The subdifferential of f at x and the directional derivative of f at x in the direction d are denoted by $\partial f(x)$ and

2. Fundamental results

In this section we show how the results given in Burke and Ferris (1993) readily extend to the infinite dimensional case. In what follows we assume that X, S, $\bar{S}$, and f are as given in (1). Characterizations of the notion of weak sharp minima are intimately tied to optimality conditions for the problem

$$\mathcal{P}: \quad \begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in S. \end{array}$$

The problem $\mathcal{P}$ can equivalently be stated as the unconstrained problem

$$\mathcal{P}_0: \quad \begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & x \in X, \end{array}$$

where $f_0: X \mapsto \overline{\mathbb{R}}$ is the essential objective function for the problem $\mathcal{P}$ and is given by

$$f_0(x) = f(x) + \psi_S(x) = \begin{cases} f(x), & \text{if } x \in S, \\ +\infty & \text{otherwise.} \end{cases} \tag{2}$$

Using this reduction one can suppress the dependence on the constraint set S. Indeed, in many applications one has $S = \operatorname{dom}(f)$. However, in other applications understanding the interplay between f and the constraint region S is crucial. Therefore, we focus on results that illustrate the separate contributions of the objective function and the constraint region.

Due to the equivalence of the problems $\mathcal{P}$ and $\mathcal{P}_0$, the most basic *first-order* optimality condition for $\mathcal{P}$ has the form

$$0 \in \partial f_0(x). \tag{3}$$

In order to decouple the roles of the objective function and constraint region, one typically posits a regularity condition that yields the validity of the addition formula

$$\partial f_0(x) = \partial(f + \psi_S)(x) = \partial f(x) + N_S(x), \tag{4}$$

in which case the optimality condition (3) can be written as

$$0 \in \partial f(x) + N_S(x) \tag{5}$$

without reference to the function f_0. A standard regularity condition under which the addition formula (4) holds is that there exists a point $z \in \operatorname{dom}(f) \cap S$ at which either f is continuous or $x \in \operatorname{int}(S)$, Ekeland and Temam (1976), Proposition 5.6, page 26.

In this study, we make use of a weak form of the addition formula, namely where the notation $\mathrm{cl}^*(S)$ denotes the weak* closure of the set $S \subset X^*$. Our use of this weak form of the addition formula is another point of departure from Burke and Ferris (1993), where the analysis depends on the addition formula (4). The weak addition rule (6) arises naturally in applications (see Appendix 7.). Note that it holds trivially without the need for the weak* closure operation if $S = \mathrm{dom}(f)$. If X is reflexive, the weak* closure of a convex set in X^* equals its norm closure in X^*. If $\partial f(x)$ is weak* compact, as is the case when f is finite-valued in a neighborhood of x, then taking the weak* closure on the right hand side of (6) is superfluous. We now give a simple example where (6) is satisfied but (4) is not.

EXAMPLE 2.1 *Let K be the ice cream (or Lorenz) cone in $\mathbb{R}^3$ given by*

$$K = \left\{x \in \mathbb{R}^3 \mid x_1^2 + x_2^2 \le x_3^2,\ 0 \le x_3\right\},$$

and let f be the support functional for K. Let S be the subspace orthogonal to the vector $x = (0,1,1)^T$. Then $f_0 = f + \psi_S$ is the support functional for the set $\mathrm{cl}^*\left(K + S^\perp\right) = \mathrm{cl}\left(K + S^\perp\right)$ so that $\partial f_0(0) = \mathrm{cl}\left(K + S^\perp\right) = \mathrm{cl}\left(\partial f(0) + N_S(0)\right)$ while the set $\partial f(0) + N_S(0) = K + S^\perp$ is not closed. It should also be noted that the set $S \cap K^\circ$ is a set of weak sharp minima for f over S.

The goal is to provide a number of variational characterizations of the notion of weak sharp minima. We consider both primal and dual characterizations. Primal characterizations involve directional derivatives and tangent cones while dual characterizations involve subgradients and normal cones. The primal characterizations are more elementary in the sense that they are derived directly from the definition, whereas the dual characterizations require the application of duality results and properties of the subdifferential calculus. Understanding the connections between primal and dual characterizations requires the application of a number of elementary duality correspondences. These correspondences are given in Appendix A.

We now establish an elementary *primal* variational characterization of weak sharp minima. This characterization is the basis for all of the characterizations examined in this paper.

THEOREM 2.2 *Let f, S, and $\bar{S}$ be as in (1), let f_0 be as in (2), and let $\alpha > 0$. Then the set $\bar{S}$ is a set of weak sharp minima for the function f over the set $S \subset X$ with modulus α if and only if*

$$f_0'(x; d) \ge \alpha\,\mathrm{dist}(d \mid T_{\bar{S}}(x)) \quad \forall x \in \bar{S} \text{ and } d \in T_S(x). \tag{7}$$

Proof. Let us first assume that the set $\bar{S}$ is a set of weak sharp minima for the function f over the set $S \subset X$ with modulus α. Let $x \in \bar{S}$. The hypothesis guarantees that for all $t > 0$ and $d \in X$

which implies that

$$\frac{f_0(x+td)-f_0(x)}{t} \geq \alpha \frac{\operatorname{dist}(x+td \mid \bar{S}) - \operatorname{dist}(x \mid \bar{S})}{t}$$

By taking the limit on both sides as $t \downarrow 0$ and applying Part 6 of Theorem A.1 (Appendix A), we obtain

$$f_0'(x;d) \geq \alpha \operatorname{dist}(d \mid T_{\bar{S}}(x)) \quad \forall x \in \bar{S} \text{ and } d \in X,$$

which implies (7).

Now assume that (7) holds and let $y \in S$ and $x \in \bar{S}$. Then

$$\begin{aligned} f_0(y) &\geq f_0(x) + f_0'(x; y-x) \geq f_0(x) + \alpha \operatorname{dist}(y - x \mid T_{\bar{S}}(x)) \\ &= f_0(x) + \alpha \operatorname{dist}(y \mid x + T_{\bar{S}}(x)). \end{aligned}$$

Therefore, by Part 4 of Theorem A.1,

$$\begin{aligned} f_0(y) &\geq f_0(x) + \alpha \sup_{x \in \bar{S}} \operatorname{dist}(y \mid x + T_{\bar{S}}(x)) \\ &= f_0(x) + \alpha \operatorname{dist}(y \mid \bar{S}). \end{aligned}$$

■

The main characterization theorem now follows.

THEOREM 2.3 *Let f, S, and $\bar{S}$ be as in (1), and assume that the addition formula (6) holds for all $x \in \bar{S}$. Let $\alpha > 0$ and consider the following statements:*

1. *The set $\bar{S}$ is a set of weak sharp minima for the function f over the set $S \subset X$ with modulus α.*
2. *The normal cone inclusion*
$$\alpha \mathbb{B}^\circ \bigcap N_{\bar{S}}(x) \subset \partial f_0(x) = \operatorname{cl}^*(\partial f(x) + N_S(x))$$
holds for all $x \in \bar{S}$.
3. *For all $x \in \bar{S}$ and $d \in T_S(x)$,*
$$f'(x;d) \geq \alpha \operatorname{dist}(d \mid T_{\bar{S}}(x)).$$
4. *The inclusion*
$$\alpha \mathbb{B}^\circ \bigcap \Big(\bigcup_{x \in \bar{S}} N_{\bar{S}}(x) \Big) \subset \bigcup_{x \in \bar{S}} \operatorname{cl}^*(\partial f(x) + N_S(x))$$
holds.
5. *(X a Hilbert space) For all $x \in \bar{S}$ and $d \in T_S(x) \bigcap N_{\bar{S}}(x)$,*
$$f'(x;d) \geq \alpha \|d\|.$$
6. *(X a Hilbert space) The inclusion*
$$\widehat{\alpha} \mathbb{B}^\circ \subset \partial f(x) + \Big[T_S(x) \bigcap N_{\bar{S}}(x) \Big]^\circ$$
holds for all $0 \leq \widehat{\alpha} < \alpha$ and $x \in \bar{S}$.
7. *For all $y \in S$,*
$$f'(p; y-p) \geq \alpha \operatorname{dist}(y \mid \bar{S}),$$

Statements 1 through 4 are equivalent. If in addition X is assumed to be a Hilbert space, then these statements are equivalent to each of the statements 5, 6, and 7.

REMARKS

1. Since for any convex set $C \subset X$ one has $T_C(x) = X$ and $N_C(x) = \{0\}$ for every $x \in \text{int}(C)$, one can replace the phrase "*for all* $x \in \bar{S}$" by the phrase "*for all* $x \in \text{bdry}(\bar{S})$" at the appropriate points in each of the statements in the theorem.
2. In Statement 2, the condition that $\alpha\mathbb{B}^\circ \bigcap N_{\bar{S}}(x) \subset \text{cl}^*(\partial f(x) + N_S(x))$ is equivalent to the statement that $\alpha\mathbb{B}^\circ \bigcap N_{\bar{S}}(x) + N_S(x) \subset \text{cl}^*(\partial f(x) + N_x(S))$ for $x \in \text{bdry}(\bar{S})$ since $N_S(x) \subset \partial f_0(x)^\infty$. Therefore, Statements 2 and 4 of Theorem 2.3 can be modified accordingly.
3. In Burke and Ferris (1993), Theorem 2.6, (a), the authors claim to have established the equivalence of Statements 5 and 6 of Theorem 2.3 for $\alpha = \widehat{\alpha}$ in the finite dimensional case. However, the proof given in Burke and Ferris (1993) is incomplete. The difficulty occurs at the end of the proof where it is incorrectly stated that for two convex sets $C_1, C_2 \subset \mathbb{R}^n$ one has
$$\psi^*_{C_1}(z) \le \psi^*_{C_2}(z)\ \forall z \in \mathbb{R}^n \iff C_1 \subset C_2.$$
The correct equivalence is
$$\psi^*_{C_1}(z) \le \psi^*_{C_2}(z)\ \forall z \in \mathbb{R}^n \iff C_1 \subset \text{cl}^*(C_2),$$
which is insufficient to establish the result for $\alpha = \widehat{\alpha}$.
4. One can replace the set S by the set $S \cap \text{dom}(f)$ to obtain a slightly refined result.
5. A local version of this theorem is considered in Section 5 to follow.

Proof. [1 ⇒ 2]: Let $x \in \bar{S}$. By Theorem 2.2,

$$f_0'(x; d) \ge \alpha\,\text{dist}(d \mid T_{\bar{S}}(x)) \quad \forall\, d \in X. \tag{8}$$

By Rockafellar (1974), Theorem 11, the function $\psi^*_{\partial f_0(x)}$ is the lower semicontinuous hull of the function $f_0'(x; \cdot)$. Since the function $\text{dist}(\cdot \mid T_{\bar{S}}(x))$ is continuous, its epi-graph is a closed convex set containing the epi-graph of the function $f_0'(x; \cdot)$. Hence the relation (8) is equivalent to the relation

$$\psi^*_{\partial f_0(x)}(d) \ge \alpha\,\text{dist}(d \mid T_{\bar{S}}(x)) \quad \forall\, d \in X. \tag{9}$$

By Part 6 of Theorem A.1, we have $\alpha\,\text{dist}(d \mid T_{\bar{S}}(x)) = \alpha\psi^*_{\mathbb{B}^\circ \cap N_{\bar{S}}(x)}(d) = \psi^*_{\alpha\mathbb{B}^\circ \cap N_{\bar{S}}(x)}(d)$. Hence, by (6), inequality (9) is equivalent to the inequality

$$\psi^*_{\partial f(x) + N_S(x)}(d) \ge \psi^*_{\alpha\mathbb{B}^\circ \cap N_{\bar{S}}(x)}(d). \tag{10}$$

Therefore, the result follows from Part 8 of Theorem A.1.

[2 ⟺ 3]: By Part 8 of Theorem A.1, the inequality (9) is equivalent to the inclusion in Part 2. In the proof of the implication [1 ⇒ 2], the inequality (9)

[2 ⇒ 4]: Trivial.
[4 ⇒ 2]: This follows immediately from Part 10 of Theorem A.1 by setting $D = \alpha\mathbb{B}^\circ$ and $C = \bar{S}$.

[3 ⇒ 1]: The condition in Part 3 is equivalent to the statement that

$$f'(x;d) + \psi_{T_S(x)}(d) \geq \alpha \operatorname{dist}(d \mid T_{\bar{S}}(x)) \quad \forall x \in \bar{S} \text{ and } d \in X. \tag{11}$$

Since the function $\operatorname{dist}(\cdot \mid T_{\bar{S}}(x))$ is continuous, its epi-graph is a closed convex set containing the epi-graph of the function $f'(x;\cdot) + \psi_{T_S(x)}(\cdot)$. Hence (11) is equivalent to (10) which, as we have seen, is equivalent to the statement (8). Therefore, the implication follows from Theorem 2.2.

(X is a Hilbert space) [5 $\iff$ 6]: In the Hilbert space setting (or more generally, in the setting of reflexive spaces), recall that the weak* topology on X^* is the same as the weak topology. Moreover, since the weak closure of a convex set is the same as its norm closure, the weak* closure of a convex set in a Hilbert space is the same as the norm closure of that set.

Since the norm is continuous and, by Rockafellar (1974), Theorem 11, for $x \in \bar{S}$, the support functional for the set $\partial f(x)$ is the lower semi-continuous hull of the function $f'(x;\cdot)$, the inequality in Part 5 is equivalent to the inequality

$$\psi^*_{\alpha\mathbb{B}^\circ}(d) \leq \psi^*_{\partial f(x)}(d) \quad \forall d \in \left[T_S(x) \bigcap N_{\bar{S}}(x)\right].$$

By Part 8 of Theorem A.1, this is equivalent to the inclusion

$$\alpha\mathbb{B}^\circ \subset \operatorname{cl}\left(\partial f(x) + [T_S(x) \cap N_{\bar{S}}(x)]^\circ\right),$$

which, by the convexity of the sets involved, is equivalent to the statement

$$\operatorname{int}(\alpha\mathbb{B}^\circ) \subset \operatorname{int}\left(\partial f(x) + [T_S(x) \cap N_{\bar{S}}(x)]^\circ\right),$$

from which the result follows.

(X is a Hilbert space) [3 ⇒ 5]: By Part 6 of Theorem A.1,

$$\operatorname{dist}(d \mid T_{\bar{S}}(x)) = \psi^*_{\mathbb{B}^\circ \cap N_{\bar{S}}(x)}(d) = \|d\|$$

for all $d \in N_{\bar{S}}(x)$. Therefore, Statement 5 follows immediately from Statement 3.

(X is a Hilbert space) [5 ⇒ 1]: Let $x \in S$ and set $\bar{x} = P(x \mid \bar{S})$. Then $(x - \bar{x}) \in T_S(\bar{x}) \bigcap N_{\bar{S}}(\bar{x})$. Therefore, by hypothesis,

$$f(x) - f(\bar{x}) \geq f'(\bar{x}; x - \bar{x}) \geq \alpha\|x - \bar{x}\| = \alpha \operatorname{dist}(x \mid \bar{S}).$$

(X is a Hilbert space) [1 ⇒ 7]: Let $y \in S$ be given and define $p := P(y \mid \bar{S})$ so that $f(y) \geq f(p) + \alpha \operatorname{dist}(y \mid \bar{S}) = f(p) + \alpha\|y - p\|$. Let $z_\lambda = \lambda y + (1-\lambda)p$ for $\lambda \in [0,1]$. Then $p = P(z_\lambda \mid \bar{S})$ for all $\lambda \in [0,1]$ and which implies that

$$\frac{f(p+\lambda(y-p))-f(p)}{\lambda} \geq \alpha \|y-p\| .$$

Taking the limit as $\lambda \searrow 0$ yields the inequality

$$f'(x; y-x) \geq \alpha \operatorname{dist}(y \mid \bar{S}).$$

[(X is a Hilbert space) $7 \Rightarrow 1$]: Let $x \in S$ and set $\bar{x} = P(x \mid \bar{S})$. Then, by the subdifferential inequality, we obtain

$$f(x) \geq f(\bar{x}) + f'(\bar{x}; x-\bar{x}) \geq f(\bar{x}) + \alpha \operatorname{dist}(x \mid \bar{S}). \qquad \blacksquare$$

Although each of the characterizations for weak sharp minima is used at different points in our development, the characterization given in Statement 2 is the key to much of our work since it is the point of closest contact to applications we consider. We now further dissect this characterization.

3. Dissecting the Inclusion $\alpha\mathbb{B}^\circ \cap N_{\bar{S}}(x) \subset \mathrm{cl}^*(\partial f(x) + N_x(S))$

The condition for weak sharp minima given in Statement 2 of Theorem 2.3 can be decomposed into two independent conditions. These two conditions play a fundamental role in the applications of the notion of weak sharp minima considered in Part II of this work. The decomposition is derived from the fact that the cone generated by the subdifferential of any lower semi-continuous convex function $f: X \mapsto \overline{\mathbb{R}}$ satisfies the inclusion

$$\mathrm{cone}(\partial f(x)) \subset N_{\mathrm{lev}_f(f(x))}(x) \tag{12}$$

for every $x \in \mathrm{dom}(f)$, where for any $\gamma \in \mathbb{R}$ the set

$$\mathrm{lev}_f(\gamma) = \{x \in X \mid f(x) \leq \gamma\}$$

is the lower level set of f of height γ. The inclusion (12) follows immediately from the subdifferential inequality for f.

LEMMA 3.1 *Let the basic assumptions of Theorem 2.3 hold. Given $x \in \bar{S}$, we have*

$$\alpha\mathbb{B}^\circ \cap N_{\bar{S}}(x) \subset \mathrm{cl}^*(\partial f(x) + N_S(x)) \tag{13}$$

if and only if

$$\mathrm{cone}(\mathrm{cl}^*(\partial f(x) + N_S(x))) = N_{\bar{S}}(x) \quad \textit{and} \tag{14}$$

$$\alpha\mathbb{B}^\circ \cap [\mathrm{cone}(\mathrm{cl}^*(\partial f(x) + N_S(x)))] \subset \mathrm{cl}^*(\partial f(x) + N_S(x)). \tag{15}$$

In addition, if the set $\partial f(x) + N_S(x)$ is weak closed, then*

Proof. Multiplying (13) by $\lambda > 0$ and taking the limit as $\lambda \nearrow +\infty$ yields the inclusion $N_{\bar{S}}(x) \subset \text{cone}(\text{cl}^*(\partial f(x) + N_S(x)))$. However, by (6) and (12),

$$\text{cone}(\text{cl}^*(\partial f(x) + N_S(x))) = \text{cone}(\partial f_0(x)) \subset N_{\text{lev}_{f_0}(f(x))}(x) = N_{\bar{S}}(x).$$

Therefore, (14) holds. The relation (15) is obtained by replacing $N_{\bar{S}}(x)$ by the cone

$$\text{cone}(\text{cl}^*(\partial f(x) + N_S(x)))$$

in (13).

Conversely, using (14) to replace $\text{cone}(\text{cl}^*(\partial f(x) + N_S(x)))$ by $N_{\bar{S}}(x)$ in (15), we obtain (13).

The final statement of the lemma follows from the definition of the cone generated by a set. ■

Conditions (14) and (15) play a pivotal role in the applications of the notion of weak sharp minima. For this reason, it is important to recognize that these conditions are independent. That is, neither of these conditions implies the other. It is easy to see that (15) does not imply (14). This is illustrated by the following simple example.

EXAMPLE 3.2 *Let $f: \mathbb{R} \mapsto \mathbb{R}$ be given by $f(x) = (\max\{0, x\})^2$, and $S = \mathbb{R}$. Then $\bar{S} = (-\infty, 0]$. In this example (15) is satisfied at $x = 0$, but (14) is not.*

On the other hand, a more sophisticated example is required to show that (14) does not imply (15). Before presenting this example, we give a lemma that provides both a necessary condition and a sufficient condition under which an inclusion of the type (15) holds. These conditions make use of the notion of an extreme point of a convex set.

DEFINITION 3.3 *An* extreme point *of a closed convex subset of a linear space is any point in the convex set that cannot be represented as the convex combination of two other points in the set.*

LEMMA 3.4 *Let C be a nonempty convex subset of the real normed linear space X. Suppose C contains the origin and let $Ext(C)$ denote the set of extreme points of C.*

1. *Suppose that there is an $\alpha > 0$ such that*
   $\alpha\mathbb{B} \cap \text{cone}(C) \subset C.$
   If $Ext(C) \setminus \{0\} \neq \emptyset$, then $\inf_{x \in Ext(C)\setminus\{0\}} \|x\| \geq \alpha$.
2. *If $C = \text{co}(0, C_0)$, where $C_0 \subset X$ is a nonempty convex set with $\inf_{x\in C_0} \|x\| \geq \alpha > 0$, then $\alpha\mathbb{B} \cap \text{cone}(C) \subset C$.*

Proof 1. Suppose $x \in Ext(C) \setminus \{0\}$. Since $0 \in C$, $\lambda x \notin C$ whenever $\lambda > 1$; otherwise, x cannot be an extreme point of C. From $\alpha \frac{x}{\|x\|} \in \alpha\mathbb{B} \cap \text{cone}(C) \subset C$,

2. By the definition of the convex hull of a set and the convexity of C_0, for $x \in \operatorname{co}(0, C_0)$, there are non-negative scalars λ_1 and λ_2 with $\lambda_1 + \lambda_2 = 1$, and $x_0 \in C_0$ such that $x = \lambda_1\, 0 + \lambda_2\, x_0$, i.e. $x \in [0, x_0]$ (the line segment joining 0 and x_0). Therefore $\alpha\mathbb{B} \cap \operatorname{cone}(C) \subset C$ since $\|x_0\| \geq \alpha$. ■

The following example shows that the condition (14) may hold while (15) does not.

EXAMPLE 3.5 *Let f be the support function for the convex hull of the set*

$$\mathcal{T} = \left\{ \begin{bmatrix} t\cos 2\pi t \\ t \sin 2\pi t \\ t \end{bmatrix} \mid t \in [0,1] \right\},$$

and $S = \mathbb{R}^3$. Then $\partial f(0) = \operatorname{co}(\mathcal{T})$ and $\bar{S} = \operatorname{cone}(\mathcal{T})^{\circ}$. Therefore, $\operatorname{cone}(\partial f(0)) = \operatorname{cone}(\operatorname{co}(\mathcal{T})) = N_{\bar{S}}(0)$ so that (14) is satisfied at $x = 0$. We claim that $Ext(\operatorname{co}(\mathcal{T}))=\mathcal{T}$. Then by Part 1 of Lemma 3.4, (15) is not satisfied since $\inf_{x\in\mathcal{T}\setminus\{0\}} \|x\| = 0$. Suppose the claim does not hold. By [Rockafellar (1970), Corollary 18.3.1], $Ext(\operatorname{co}(\mathcal{T})) \subset \mathcal{T}$. It is easy to see that $0 \in Ext(\operatorname{co}(\mathcal{T}))$. Suppose that there is a $\bar{t} \in (0,1]$ such that $x_{\bar{t}} \in \mathcal{T} \setminus Ext(\operatorname{co}(\mathcal{T}))$, where $x_{\bar{t}} = [\bar{t}\cos 2\pi\bar{t}, \bar{t}\sin 2\pi\bar{t}, \bar{t}]^T$. Then $\operatorname{co}(\mathcal{T} \setminus x_{\bar{t}}) = \operatorname{co}(\mathcal{T})$ since $Ext(\operatorname{co}(\mathcal{T})) \subset \mathcal{T} \setminus x_{\bar{t}}$. By Carathéodory's Theorem, the point $x_{\bar{t}}$ can be represented as a convex combination of 4 or fewer points from $\mathcal{T} \setminus x_{\bar{t}}$:

$$x_{\bar{t}} = \sum_{i=1}^{k} \lambda_i \begin{bmatrix} t_i \cos 2\pi t_i \\ t_i \sin 2\pi t_i \\ t_i \end{bmatrix}, \tag{16}$$

$$\textit{where } 0 \leq t_i \leq 1 \textit{ with } t_i \neq \bar{t}, 0 \leq \lambda_i \leq 1,\ i = 1, \ldots, k \leq 4, \textit{ and } \sum_{i=1}^{k} \lambda_i = 1.$$

Dividing both sides of the equation in (16) by $\bar{t}$ yields the relation

$$\begin{bmatrix} \cos 2\pi\bar{t} \\ \sin 2\pi\bar{t} \end{bmatrix} = \sum_{i=1}^{k} \eta_i \begin{bmatrix} \cos 2\pi t_i \\ \sin 2\pi t_i \end{bmatrix},$$

$$\textit{where } 0 \leq \eta_i \leq 1,\ i = 1, \ldots, k, \textit{ and } \sum_{i=1}^{k} \eta_i = 1.$$

Here, $\eta_i = \frac{\lambda_i t_i}{\bar{t}} \geq 0$, $i = 1, \ldots, k$. Taking the inner product on both sides with $[\cos 2\pi\bar{t}, \sin 2\pi\bar{t}]^T$ gives

$$\begin{aligned} 1 &= \sum_{i=1}^{k} \eta_i [\cos 2\pi t_i cos 2\pi\bar{t} + \sin 2\pi t_i \sin 2\pi\bar{t}] \\ &= \sum_{i=1}^{k} \eta_i \cos 2\pi(t_i - \bar{t}). \end{aligned}$$

Now, since $\sum_{i=1}^k \eta_i = 1$, $\eta_i \geq 0$ *for* $i = 1, \dots, k$, $|\cos 2\pi(t_i - \bar{t})| \leq 1$, $-1 \leq -\bar{t} \leq t_i - \bar{t} < 1$, *this equation can hold if and only if* $t_i = \bar{t}$ *whenever* $\eta_i > 0$. *This contradicts the original choice of* $t_i \neq \bar{t}$, *and the claim is proved.*

4. Asymptotic properties of weak sharp minima

The notion of weak sharp minima defined in (1) is a global property. This property implies that the function f and the sets $\bar{S}$ and S possess certain asymptotic properties. These properties are revealed by considering the recession function of f and the recession cones of the sets $\bar{S}$ and S. Recall from Rockafellar (1966) that the *recession cone* of a nonempty closed convex subset C of the normed linear space X is the set

$$C^\infty = \{y \mid x + y \in C, \ \forall x \in C\}. \tag{17}$$

A number of equivalent representations of the recession cone can be found in Rockafellar (1966), Theorem 2A. Of particular interest to us is the representation given by Rockafellar (1966), Theorem 2A, Part (d),

$$C^\infty = [\text{bar}\,(C)]^\circ, \tag{18}$$

where bar (C) is the *barrier cone* of C. The barrier cone of C is by definition the essential domain of the support function for C: $\text{bar}\,(C) = \text{dom}\,(\psi_C^*)$. These relationships imply that C^∞ is a nonempty closed convex cone whenever C is nonempty. On the other hand, simple examples show that the convex cone bar (C) is not always closed.

The *recession function* of a proper lower semi-continuous convex function $g: X \mapsto \overline{\mathbb{R}}$ is the unique convex function $g^\infty: X \mapsto \overline{\mathbb{R}}$ satisfying

$$\text{epi}(g^\infty) = (\text{epi}\, g)^\infty. \tag{19}$$

By Rockafellar (1966), Corollary 3D, we have that

$$g^\infty = \psi^*_{\text{dom}(g^*)} = \psi^*_{\text{cl}^*(\text{dom}(g^*))}. \tag{20}$$

Thus, in particular,

$$\partial g^\infty(0) = \text{cl}^*\,(\text{dom}\,(g^*)). \tag{21}$$

Our goal is to show that if $\bar{S}$ is a set of weak sharp minima for f relative to S, then $\bar{S}^\infty$ is a set of weak sharp minima for f^∞ relative to S^∞. For this we require a number of basic facts about the recession functions f_0^∞, f^∞, and the recession cones S^∞, and $\bar{S}^\infty$. These are stated and proved in Appendix B.

The recession results of Appendix B are used in conjunction with Statement 4 in Theorem 2.3 to characterize when $\bar{S}^\infty$ is a set of weak sharp minima for

THEOREM 4.1 *The set $\bar{S}^\infty$ is a set of weak sharp minima for f^∞ relative to S^∞ with modulus $\alpha > 0$ if and only if*

$$\alpha\mathbb{B}^\circ \cap (\bar{S}^\infty)^\circ \subset \mathrm{cl}^*(\mathrm{dom}(f_0^*)) = \mathrm{cl}^*(\partial f^\infty(0) + (S^\infty)^\circ). \tag{22}$$

Proof. By (B.1), the set $\bar{S}^\infty$ is a set of weak sharp minima for f^∞ relative to S^∞ with modulus $\alpha > 0$ if and only if set $\bar{S}^\infty$ is a set of weak sharp minima for f_0^∞ with modulus $\alpha > 0$. Statement 4 of Theorem 2.3 says that $\bar{S}^\infty$ is a set of weak sharp minima for f_0^∞ with modulus $\alpha > 0$ if and only if

$$\alpha\mathbb{B}^\circ \bigcap \Big(\bigcup_{y\in\bar{S}^\infty} N_{\bar{S}^\infty}(y) \Big) \subset \bigcup_{y\in\bar{S}^\infty} \partial f_0^\infty(y). \tag{23}$$

But by (B.4), (B.5), and (B.3), the inclusion (23) is equivalent to (22). ■

The main result of this section follows.

THEOREM 4.2 *Assume that the space X is reflexive and that the addition formula (6) holds at every point of $\bar{S}$. If $\bar{S}$ is a set of weak sharp minima for f relative to S with modulus α, then*

$$\alpha\mathbb{B}^\circ \cap (\bar{S}^\infty)^\circ \subset \mathrm{cl}^*\Big(\bigcup_{x\in\bar{S}} (\partial f(x) + N_S(x))\Big) \subset \mathrm{cl}^*(\mathrm{dom}(f_0^*)). \tag{24}$$

In particular, this implies that $\bar{S}^\infty$ is a set of weak sharp minima for f^∞ relative to S^∞ with modulus α.

Proof. If $\bar{S}$ is a set of weak sharp minima for f relative to S with modulus α and the addition formula (6) holds on $\bar{S}$, then Statement 4 of Theorem 2.3 tells us that

$$\alpha\mathbb{B}^\circ \bigcap \Big(\bigcup_{x\in\bar{S}} N_{\bar{S}}(x) \Big) \subset \bigcup_{x\in\bar{S}} \mathrm{cl}^*(\partial f(x) + N_S(x)) = \bigcup_{x\in\bar{S}} \partial f_0(x) \subset \mathrm{dom}(f_0^*).$$

Taking the weak* closure on both sides of this expression yields

$$\mathrm{cl}^*\Big(\alpha\mathbb{B}^\circ \bigcap \Big(\bigcup_{x\in\bar{S}} N_{\bar{S}}(x) \Big)\Big) \subset \mathrm{cl}^*\Big(\bigcup_{x\in\bar{S}} \partial f(x) + N_S(x) \Big) \subset \mathrm{cl}^*(\mathrm{dom}(f_0^*)),$$

where we have made use of the straightforward identity

$$\mathrm{cl}^*\Big(\bigcup_{x\in\bar{S}} \mathrm{cl}^*(\partial f(x) + N_S(x)) \Big) = \mathrm{cl}^*\Big(\bigcup_{x\in\bar{S}} (\partial f(x) + N_S(x)) \Big).$$

We now claim that

$$\mathrm{cl}^*\Big(\alpha\mathbb{B}^\circ \bigcap \Big(\bigcup_{x\in\bar{S}} N_{\bar{S}}(x) \Big)\Big) = \alpha\mathbb{B}^\circ \bigcap \mathrm{cl}^*\Big(\bigcup_{x\in\bar{S}} N_{\bar{S}}(x) \Big).$$

Since the left-hand side is clearly contained in the right-hand side ($\mathbb{B}^\circ$ is weak*

the weak and weak* topologies coincide. By (B.8) in Lemma B.3, the weak and strong closures of the set $\bigcup_{x\in\bar S} N_{\bar S}(x)$ coincide. Let $\widehat{x}\in\alpha\mathbb{B}^\circ\bigcap \mathrm{cl}^*\left(\bigcup_{x\in\bar S} N_{\bar S}(x)\right)$ and let $\{x^k\}\subset\bigcup_{x\in\bar S} N_{\bar S}(x)$ be such that $\widehat{x}$ is the strong limit of $\{x^k\}$. In particular, this implies that $\|x^k\|\to\|\widehat{x}\|\le\alpha$. Set

$$\tau_k=\sup\left\{\tau \mid 0\le\tau\le 1,\ \tau x^k\in\alpha\mathbb{B}^\circ\right\}.$$

Then $\tau_k\to 1$ since $\|x^k\|\to\|\widehat{x}\|\le\alpha$, therefore $\{\tau_k x^k\}\subset\alpha\mathbb{B}^\circ\bigcap(\bigcup_{x\in\bar S} N_{\bar S}(x))$ with $\tau_k x^k\to\widehat{x}$. Consequently,

$$\widehat{x}\in \mathrm{cl}\left(\alpha\mathbb{B}^\circ\bigcap\Big(\bigcup_{x\in\bar S} N_{\bar S}(x)\Big)\right)\subset \mathrm{cl}^*\left(\alpha\mathbb{B}^\circ\bigcap\Big(\bigcup_{x\in\bar S} N_{\bar S}(x)\Big)\right).$$

Finally, by (B.8), we have

$$\mathrm{cl}^*\left(\bigcup_{x\in\bar S} N_{\bar S}(x)\right)=(\bar S^\infty)^\circ,$$

whereby (24) is established. ■

REMARKS

1. The question remains open whether or not Theorem 4.2 holds in general Banach spaces.
2. In Deng (1997), Gowda (1996), Hu and Wang (1989), recession analysis is used to study global error bounds.

5. Local weak sharp minima

Local versions of the notion of weak sharp minima can be obtained in a number of ways. However, one must be careful when extending the various characterizations of weak sharp minima given in Theorem 2.3 to the local setting. We study a particularly useful localization of these ideas, which is related to the notion of metric regularity to be discussed in Part II.

DEFINITION 5.1 *Let $S\subset X$ and let $f\colon X\mapsto\overline{\mathbb{R}}$ where $\overline{\mathbb{R}}=\mathbb{R}\cup\{+\infty\}$. The set $\bar S:=\arg\min\{f(x)\mid x\in S\}$ is said to be a set of weak sharp minima at $\bar x\in\bar S$ for f over the set S with modulus $\alpha>0$ if there exists $\epsilon>0$ such that*

$$f(x)\ge f(\bar x)+\alpha\,\mathrm{dist}(x\mid\bar S) \tag{25}$$

for all $x\in S\cap(\bar x+\epsilon\mathbb{B})$. The set $\bar S$ is said to be a set of local weak sharp minima for f over S if it is a set of weak sharp minima at $\bar x\in\bar S$ for f over the set S

The most troublesome wrinkle in this definition is that the set $\bar{S}$ is no longer a subset of the set $S \cap (\bar{x} + \epsilon\mathbb{B})$. This has important consequences for the types of characterization theorems one can obtain. In particular, local versions of the results of Theorem 2.3 do not all carry over to this new setting. In our next result, we give an indication of what is possible.

THEOREM 5.2 *Let $\bar{S}$ and S be nonempty closed convex subsets of X with $\bar{S} \subset S$ and let $\bar{x} \in \bar{S}$. Assume that $f: X \mapsto \overline{\mathbb{R}}$ is lower semi-continuous and convex with $\operatorname{dom}(f) \neq \emptyset$ and that the addition formula (6) holds in a neighborhood of $\bar{x}$. Let $\alpha > 0$ and consider the following statements:*

(A) The set $\bar{S}$ is a set of weak sharp minima at $\bar{x} \in \bar{S}$ for the function f over the set S with modulus $\alpha > 0$.

(B) There is an $\epsilon > 0$ such that

$$\alpha\mathbb{B}^{\circ} \cap N_{\bar{S}}(x) \subset \mathrm{cl}^{*}\left(\partial f(x) + N_{S}(x)\right) \quad \forall x \in \bar{S} \cap \operatorname{int}(\bar{x} + \epsilon\mathbb{B}). \quad (26)$$

We have the following relationships between these statements:

1. *Statement (A) implies statement (B).*
2. *If X is assumed to be a Hilbert space, then statements (A) and (B) are equivalent.*
3. *If X is finite dimensional, then statements (A) and (B) are equivalent but for possibly different values of α.*

REMARK In Theorem 2.3 we focus on the condition appearing in Part 3 of Theorem 2.3. However, any of the other conditions in Theorem 2.3 can be refined in a similar way.

Proof 1. Let $\epsilon > 0$ be chosen so that the addition formula (6) holds on $\bar{x} + \epsilon\mathbb{B}$ and (25) holds for all $x \in S \cap (\bar{x} + \epsilon\mathbb{B})$. Let $x \in \bar{S} \cap \operatorname{int}(\bar{x} + \epsilon\mathbb{B})$. Then, given $d \in X$ with $d \neq 0$, and $0 < t < \frac{\epsilon - (\|x - \bar{x}\|)}{\|d\|}$, we have

$$\frac{f_0(x + td) - f_0(x)}{t} \geq \alpha \frac{\operatorname{dist}(x + td \mid \bar{S}) - \operatorname{dist}(\bar{x} \mid \bar{S})}{t},$$

since $f_0(x) = f_0(\bar{x})$. By taking the limit on both sides as $t \downarrow 0$ and applying Part 6 of Theorem A.1, we obtain the inequality

$$f_0'(x; d) \geq \operatorname{dist}(d \mid T_{\bar{S}}(x)).$$

The result now follows as in the proof of Theorem 2.3.

2. By Part 8 of Theorem A.1, the hypotheses imply that

$$f'(x; y - x) \geq \alpha \operatorname{dist}(y - x \mid T_{\bar{S}}(x)) \quad \forall x \in \bar{S} \cap \operatorname{int}(\bar{x} + \epsilon\mathbb{B}) \text{ and } y \in S.$$

Next observe that for $y \in S \cap \operatorname{int}(\bar{x} + \epsilon\mathbb{B})$ we have

$$\begin{aligned}\|P(y \mid \bar{S}) - \bar{x}\| &\leq \|P(y \mid \bar{S}) - P(\bar{x} \mid \bar{S})\| \\ &\leq \|y - \bar{x}\|\end{aligned}$$

so $P(y \mid \bar{S}) \in \bar{S} \cap \operatorname{int}(\bar{x} + \epsilon\mathbb{B})$. Therefore, for all $y \in S \cap \operatorname{int}(\bar{x} + \epsilon\mathbb{B})$

$$\begin{aligned} &f'(P(y \mid \bar{S}); y - P(y \mid \bar{S})) \geq \alpha \operatorname{dist}(y \mid P(y \mid \bar{S}) + T_{\bar{S}}(P(y \mid \bar{S}))) \\ &= \alpha \left\| P(y \mid \bar{S}) - y \right\| \\ &= \alpha \operatorname{dist}(y \mid \bar{S}). \end{aligned}$$

The subdifferential inequality now implies that for all $y \in S \cap \operatorname{int}(\bar{x} + \epsilon\mathbb{B})$

$$\begin{aligned} &f(y) \geq f(P(y \mid \bar{S})) + f'(P(y \mid \bar{S}); y - P(y \mid \bar{S})) \\ &\geq f(\bar{x}) + \alpha \operatorname{dist}(y \mid \bar{S}). \end{aligned}$$

Hence, by the lower semi-continuity of f and the continuity of the distance function, this inequality must hold for $y \in S \cap (\bar{x} + \epsilon\mathbb{B})$.
3. Due to the equivalence of norms, the inequality (1) as well as the inclusion (26) holding for one norm implies that it must hold for all norms for possibly different values of α. ■

Part 3 of Theorem 5.2 yields the following characterization for the set $\bar{S}$ to be a set of local weak sharp minima for f over S in the finite dimensional case.

COROLLARY 5.3 *Let X be finite dimensional, and assume that the addition formula (6) hold at every point of $\bar{S}$. Then $\bar{S}$ is a set of local weak sharp minima for f over S if and only if for every $r > 0$ for which $r\mathbb{B} \cap \bar{S} \neq \emptyset$ there exists $\alpha(r) > 0$ such that*

$$\alpha(r)\mathbb{B}^\circ \cap \Big(\bigcup_{x \in \bar{S} \cap r\mathbb{B}} N_{\bar{S}}(x) \Big) \subset \bigcup_{x \in \bar{S} \cap r\mathbb{B}} \operatorname{cl}(\partial f(x) + N_S(x)). \tag{27}$$

In addition, the condition (27) is equivalent to the condition

$$\alpha(r)\mathbb{B}^\circ \cap N_{\bar{S}}(x) \subset \operatorname{cl}(\partial f(x) + N_S(x)) \quad \forall x \in \bar{S} \cap r\mathbb{B}. \tag{28}$$

Proof. The fact that (27) is equivalent to (28) follows immediately from Part 10 of Theorem A.1 by setting $D = \alpha(r)\mathbb{B}^\circ$ and $C = \bar{S} \cap r\mathbb{B}$.

Let $r_0 = \operatorname{dist}(0 \mid \bar{S})$. Since $\bar{S}$ is closed, $\bar{S} \cap r\mathbb{B} \neq \emptyset$ for all $r \geq r_0$. Let us first suppose that $\bar{S}$ is a set of local weak sharp minima for f over S. Choose $\bar{\alpha} > 0$ and for each $x \in \bar{S}$ define

$$\widehat{\alpha}(x) = \min\{\sup\{\alpha \mid \alpha\mathbb{B}^\circ \cap N_{\bar{S}}(x) \subset \operatorname{cl}(\partial f(x) + N_S(x))\}, \bar{\alpha}\}.$$

Since statement (A) in Theorem 5.2 holds for every $\bar{x} \in \bar{S}$, Part 1 of Theorem 5.2 implies that $\widehat{\alpha}(x) > 0$ for all $x \in \bar{S}$. Define $\alpha(r) = \inf\{\widehat{\alpha}(x) \mid x \in \bar{S} \cap r\mathbb{B}\}$ for all $r > r_0$. Let $r > r_0$. If $\alpha(r) > 0$, then (28) holds, which in turn implies (27). On the other hand, if $\alpha(r) = 0$, then, since the set $\bar{S} \cap r\mathbb{B}$ is compact, there is a sequence $\{x^k\} \subset \bar{S} \cap r\mathbb{B}$ and an $\bar{x} \in \bar{S} \cap r\mathbb{B}$ such that $x^k \to \bar{x}$ and $\widehat{\alpha}(x^k) \to 0$. Since statement (A) in Theorem 5.2 holds at $\bar{x}$, Part 1 of Theorem every $x \in \bar{x} + \epsilon\mathbb{B}$, $\widehat{\alpha}(x) \geq \alpha > 0$. This contradicts the fact that $x^k \to \bar{x}$ with $\widehat{\alpha}(x^k) \to 0$. Hence, $\alpha(r) > 0$.

Now suppose that (27), or equivalently, (28) holds. Let $\bar{x} \in \bar{S}$ and $r > \|\bar{x}\|$. Set $\epsilon = (r - \|\bar{x}\|)/2$. Then (28) and Part 3 of Theorem 5.2 imply that statement (A) in Theorem 5.2 holds with this choice of $\epsilon > 0$ and $\alpha = \alpha(r)$. Since $\bar{x} \in \bar{S}$ was chosen arbitrarily, the reverse implication is established. ■

6. Boundedly weak sharp minima

The condition (27) given in Corollary 5.3 is interesting in its own right in the infinite dimensional case. We show that this condition always holds if $\bar{S}$ is assumed to be a set of *boundedly weak sharp minima.*

DEFINITION 6.1 *Let $S \subset X$ and let $f : X \mapsto \overline{\mathbb{R}}$ where $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. The set $\bar{S} := \arg\min \{f(x) \mid x \in S\}$ is said to a set of boundedly weak sharp minima for f over the set S if for every $r > 0$ for which $\bar{S} \cap r\mathbb{B} \neq \emptyset$ there is an $\alpha_r > 0$ such that*

$$f(x) \geq f(\bar{x}) + \alpha_r \operatorname{dist}(x \mid \bar{S}) \tag{29}$$

for all $x \in S \cap r\mathbb{B}$, where $\bar{x}$ is any element of $\bar{S}$.

This definition could have been stated with the sets $r\mathbb{B}$ replaced by bounded subsets K of X. Indeed, this is the origin of the term. However, such a restatement does not increase the generality of the definition.

LEMMA 6.2 *The set $\bar{S}$ is a set of boundedly weak sharp minima for f over the set S if and only if for every bounded set $K \subset X$ there is an $\alpha_K > 0$ such that*

$$f(x) \geq f(\bar{x}) + \alpha_K \operatorname{dist}(x \mid \bar{S}) \quad \forall x \in S \cap K, \tag{30}$$

where $\bar{x}$ is any element of $\bar{S}$.

Proof. The forward implication follows by choosing $r > 0$ so that $K \subset r\mathbb{B}$, while the reverse implication follows by taking $K = r\mathbb{B}$. ■

It is clear that the notion of weak sharp minima implies that of boundedly weak sharp minima, which, in turn, implies that of local weak sharp minima. We now relate the notion of boundedly weak sharp minima to condition (27) given in Corollary 5.3.

THEOREM 6.3 *Consider the following statements:*

(a) The set $\bar{S}$ is a set of boundedly weak sharp minima for f over the set S.

(b) For every $r > 0$ for which $\bar{S} \cap r\mathbb{B} \neq \emptyset$ there is an $\alpha(r) > 0$ such that

$$\alpha(r)\mathbb{B}^\circ \cap \Big(\bigcup N_{\bar{S}}(x) \Big) \subset \bigcup \mathrm{cl}^*(\partial f(x) + N_S(x)). \tag{31}$$

(c) For every $r > 0$ for which $\bar{S} \cap r\mathbb{B} \neq \emptyset$ there is an $\alpha(r) > 0$ such that

$$\alpha(r)\mathbb{B}^{\circ} \cap N_{\bar{S}}(x) \subset \mathrm{cl}^{*}\left(\partial f(x) + N_S(x)\right) \quad \text{for all } x \in \bar{S} \cap r\mathbb{B}. \tag{32}$$

(d) The condition

$$\mathrm{cone}(\mathrm{cl}^{*}\left(\partial f(x) + N_S(x)\right)) = N_{\bar{S}}(x) \tag{33}$$

holds for all $x \in \bar{S}$ and for every $r > 0$ for which $\bar{S} \cap r\mathbb{B} \neq \emptyset$ there is an $\alpha(r) > 0$ such that

$$\alpha(r)\mathbb{B}^{\circ} \cap [\mathrm{cone}(\mathrm{cl}^{*}\left(\partial f(x) + N_S(x)\right))] \subset \mathrm{cl}^{*}\left(\partial f(x) + N_S(x)\right)$$
$$\forall\, x \in \bar{S} \cap r\mathbb{B}. \tag{34}$$

Statement (a) implies statement (b), and statements (b), (c), and (d) are equivalent. In addition, if X is either a Hilbert space or finite dimensional, then statement (b) implies statement (a).

Proof. By Part 10 of Theorem A.1, the statement (b) is equivalent to statement (c), and, by Lemma 3.1, statement (c) is equivalent to statement (d). Thus, we need only show that statement (a) implies statement (c).

Assume that $\bar{S}$ is a set of boundedly weak sharp minima for f over the set S and let $r > 0$ be such that $\bar{S} \cap r\mathbb{B} \neq \emptyset$. Let $\alpha_{r+1} > 0$ be as in Definition 6.1 so that

$$f(x) \geq f(\bar{x}) + \alpha_{r+1}\,\mathrm{dist}(x \mid \bar{S}) \quad \text{for all } x \in S \cap (r+1)\mathbb{B} \text{ and } \bar{x} \in \bar{S}.$$

Let $\bar{x} \in \bar{S} \cap r\mathbb{B}$ and define $\alpha(r) = \alpha_{r+1}$. By Part 1 of Theorem 5.2 with $\epsilon = 1$, we have

$$\alpha(r)\mathbb{B}^{\circ} \cap N_{\bar{S}}(\bar{x}) \subset \mathrm{cl}^{*}\left(\partial f(\bar{x}) + N_S(\bar{x})\right)$$

which establishes (32).

Now we prove the "converse" under the assumption that X is either a Hilbert space or X is finite dimensional. That is, we show that statement (c) implies statement (a). Suppose that $\bar{S} \cap r\mathbb{B} \neq \emptyset$. Since (32) holds on $\bar{S} \cap (5r)\mathbb{B}$, Parts 2 and 3 of Theorem 5.2 with $\epsilon = 2r$ imply that there is some $\alpha_r > 0$ such that

$$f(x) \geq f(\bar{x}) + \alpha_r\,\mathrm{dist}(x \mid \bar{S}) \quad \forall\, x \in S \cap (\bar{x} + (2r)\mathbb{B}), \tag{35}$$

for all $\bar{x} \in \bar{S}$ since $f(\bar{x}) = f(\widehat{x})$ whenever $\bar{x}, \widehat{x} \in \bar{S}$. When $x \in S \cap r\mathbb{B}$, let $\widehat{x} \in \bar{S} \cap r\mathbb{B}$, it follows from the triangle inequality property of a norm that, for

$$\bar{x} \in \left\{u \in S \mid \|x - u\| = \mathrm{dist}(x \mid \bar{S})\right\},$$

we have

$$\|x - \bar{x}\| = \mathrm{dist}(x \mid \bar{S}) \leq \|x - \widehat{x}\| \leq 2r.$$

This shows that

$$f(x) \geq f(\bar{x}) + \alpha_r\,\mathrm{dist}(x \mid \bar{S}) \quad \text{for all } x \in S \cap r\mathbb{B},$$

and the proof is complete. ■

COROLLARY 6.4 *Let X be finite dimensional. Then the set $\bar{S}$ is a set of boundedly weak sharp minima for f over the set S if and only if $\bar{S}$ is a set of local weak sharp minima for f over S*

We now examine the difference between the notions of weak sharp minima and boundedly weak sharp minima. Our approach to this is to compare (31) with Part 4 of Theorem 2.3. It may happen that in the case of boundedly weak sharp minima one has $\alpha(r) \to 0$ as $r \to \infty$ in which case the set $\bar{S}$ is not a set of weak sharp minima for f over S (see Example 6.6 below). Conditions under which $\alpha(r) \not\to 0$ as $r \to \infty$ are related to the notion of an *asymptotic constraint qualification*, see Auslender and Crouzeix (1988), Mangasarian (1985). A simple condition assuring that $\alpha(r)$ is bounded away from zero is given in the following theorem.

THEOREM 6.5 *Suppose that X is a reflexive Banach space and that the formula (6) holds at every point of $\bar{S}$. If $\bar{S}$ admits a decomposition of the form*

$$\bar{S} = \bar{S}^{\infty} + D, \tag{36}$$

where D is a bounded closed convex subset of X, then $\bar{S}$ is a set of weak sharp minima for f over S if and only if $\bar{S}$ is a set of boundedly weak sharp minima for f over S. In addition, if X is assumed to be finite dimensional, then the decomposition (36) holds if either (a) $0 \in ri(dom f_0^*)$, in which case $\bar{S}^{\infty}$ is a subspace, or (b) $\bar{S}$ is a polyhedral set.

Proof. If $\bar{S}$ is a set of weak sharp minima for f over S, then trivially $\bar{S}$ is a set of boundedly weak sharp minima for f over S. Conversely, let us suppose that $\bar{S}$ is a set of boundedly weak sharp minima for f over S. Then, for any $x^* \in (\bar{S}^{\infty})^{\circ}$ there exists $d \in D$ such that

$$\psi^*_{\bar{S}}(x^*) = \psi^*_{\bar{S}^{\infty}}(x^*) + \psi^*_{D}(x^*) = \psi^*_{D}(x^*) = \langle x^*, d\rangle,$$

since X is reflexive. Hence, by Part 1 of Theorem A.1, $(\bar{S}^{\infty})^{\circ} \subset \cup_{x\in D} N_{\bar{S}}(x)$. But then, by Lemma B.3,

$$\bigcup_{x\in\bar{S}} N_{\bar{S}}(x) \subset (\bar{S}^{\infty})^{\circ} \subset \bigcup_{x\in D} N_{\bar{S}}(x),$$

whereby

$$(\bar{S}^{\infty})^{\circ} = \bigcup_{x\in\bar{S}} N_{\bar{S}}(x) = \bigcup_{x\in D} N_{\bar{S}}(x).$$

Let $r > 0$ be such that $D \subset r\mathbb{B}$, and let $\alpha = \alpha(r) > 0$ be chosen to satisfy the inclusion (31). Then,

$$\alpha\mathbb{B}^{\circ} \bigcap \Big(\bigcup N_{\bar{S}}(x) \Big) = \alpha\mathbb{B}^{\circ} \bigcap \Big(\bigcup N_{\bar{S}}(x) \Big)$$

$$\subset \alpha\mathbb{B}^\circ \bigcap \Big(\bigcup_{x\in\bar{S}\cap r\mathbb{B}} N_{\bar{S}}(x)\Big)$$
$$\subset \bigcup_{x\in\bar{S}\cap r\mathbb{B}} \mathrm{cl}\,(\partial f(x) + N_S\,(x))$$
$$\subset \bigcup_{x\in\bar{S}} \mathrm{cl}\,(\partial f(x) + N_S\,(x))\,.$$

Therefore, by Part 4 of Theorem 2.3, $\bar{S}$ is a set of weak sharp minima for f over S.

The fact that (36) holds with $\bar{S}^\infty$ a subspace when $0 \in \mathrm{ri}\,(\mathrm{dom}\,(f_0^*))$ is proved in Auslender, Cominetti and Crouzeix (1993), Theorem 2.3. The fact that (36) holds when $\bar{S}$ is polyhedral is an immediate consequence of Rockafellar (1970), Corollary 19.1.1. ■

REMARK. The condition $0 \in \mathrm{ri}\,(\mathrm{dom}\,(f_0^*))$ is carefully examined in Auslender, Cominetti and Crouzeix (1993), where a number of important consequences of this hypothesis are presented. A special case of Theorem 6.5, where X is finite dimensional and $0 \in \mathrm{ri}\,(\mathrm{dom}\,(f_0^*))$, was proved in Deng (1998), Corollary 5. Additional examples in optimization where $\bar{S}$ exhibits the decomposition (36), can be found in Klatte (1998).

In the finite dimensional case, Theorems 6.3 and 6.5 and Corollary 6.4 indicate that the ability to jump from local to global weak sharp minima is related to the asymptotic geometry of the sets $\bar{S}$ and S. This geometry was examined in Section 4. In the following example it is shown that $\bar{S}$ being a set of boundedly weak sharp minima for f does not imply that $\bar{S}^\infty$ is a set of weak sharp minima for f^∞.

EXAMPLE 6.6 *Consider* $f(x_1, x_2) = [b(x)]_+$, *where* $b(x) = \sqrt{x_1^2 + x_2^2} - x_1 - 1$. *Let* $\bar{S}$ *be a set of optimal solutions of* f. *It is easily to see that* $f^\infty(x) = \sqrt{x_1^2 + x_2^2} - x_1$, *and* $\bar{S}^\infty = \mathbb{R}_+ \times \{0\}$. *The set* $\bar{S}^\infty$ *is not a set of weak sharp minima for* f^∞ *since* $N_{\bar{S}^\infty}(0) = \mathbb{R}_- \times \mathbb{R}$ *and* $\partial f^\infty(0) = (-1, 0) + \mathbb{B}$. *By Proposition 4.2,* $\bar{S}$ *is not a set of weak sharp minima for* f. *However,* $b(x) \le 0$ *satisfies the Slater condition. Hence* $\bar{S}$ *is a set of boundedly weak sharp minima for* f.

This example, in conjunction with Theorem 4.2, leads one to conjecture that the two assumptions (a) $\bar{S}$ is a set of boundedly weak sharp minima for f, (b) $\bar{S}^\infty$ is a set of weak sharp minima for f^∞, taken together might imply that $\bar{S}$ is a set of weak sharp minima for f. The following example shows that the aforementioned assumptions (a) and (b) do not imply $\bar{S}$ is a set of weak sharp minima. Thus, the weakest additional conditions under which a set of boundedly weak sharp minima becomes a set of weak sharp minima are still

Example 6.7 *Let $b(x)$ be given in Example 6.6, and $g(x) = (\sqrt{x_1^2 + x_2^2} + x_1 + 1)b(x) = x_2^2 - 2x_1 - 1$. Let $h(x) = \max\{b(x), g(x)\}$. Since $\sqrt{x_1^2 + x_2^2} + x_1 + 1 \geq 1$, $b(x) \geq g(x)$ if $b(x) < 0$, and $b(x) \leq g(x)$ if $b(x) \geq 0$. For any real number $t > -1$, let $f_t(x) = [h(x) - t]_+$, and $\bar{S}_t = \arg\min_{x \in \mathbb{R}^2} f_t(x)$. Since $h(0) - t = -1 - t < 0$, the system $h(x) - t \leq 0$ satisfies the Slater condition. It follows that $\bar{S}_t$ is a set of boundedly weak sharp minima for f_t when $t > -1$. We will show that $\bar{S}_t$ is a set of weak sharp minima for f_t when $t \geq 0$, and $\bar{S}_t$ is not a set of weak sharp minima for f_t when $-1 < t < 0$.*

For $t \geq 0$, if $x = (x_1, x_2)$ is such that $g(x) = t$, then $b(x) \leq t$. This means that $\bar{S}_t = \widehat{S}_t$, where $\widehat{S}_t$ is the solution set to the inequality system $g(x) \leq t$. Since $g^\infty(1,0) = -2$, we know from Deng (1997), Theorem 2.3, that $\operatorname{dist}(x \mid \widehat{S}_t) \leq 1/2[g(x) - t]_+$ for all $x \in \mathbb{R}^2$. It follows that $\operatorname{dist}(x \mid \bar{S}_t) = \operatorname{dist}(x \mid \widehat{S}_t) \leq 1/2[g(x) - t]_+ \leq 1/2[h(x) - t]_+ = 1/2 f_t(x)$ for all $x \in \mathbb{R}^2$, i.e., S_t is a set of weak sharp minima for f_t. By Rockafellar (1970), Theorem 9.4,

$$f_t^\infty(x) = \max\{b^\infty(x), g^\infty(x), 0\} = \max\{\psi^*_{\mathbb{B}+(-1,0)}(x), \psi^*_{\{-2\}\times\mathbb{R}}(x)\},$$

where this function is independent of t, and the cones $\bar{S}_t^\infty$ are the same for all $t > -1$. Hence, by Theorem 4.2, $\bar{S}_t^\infty$ is a set of weak sharp minima for f_t^∞ for $t > -1$.

Let $-1 < t < 0$, and consider $x = (x_1, x_2)$ such that $b(x) = t$. In this case we have

$$\sqrt{x_1^2 + x_2^2} = x_1 + 1 + t, \quad \text{and} \tag{37}$$

$$x_2^2 = 2x_1(1 + t) + (1 + t)^2. \tag{38}$$

It follows that $x_1 \geq -(1 + t)/2$, and $x_1 = -(1 + t)/2$ if and only if $x_2 = 0$. If $x_2 \neq 0$, we have $t = b(x) > g(x)$ since

$$\sqrt{x_1^2 + x_2^2} + x_1 + 1 = (2x_1 + 1 + t) + 1 > 1.$$

As a consequence, we have

$$\partial h(x) = \begin{cases} co\{\partial b(x), \partial g(x)\} & \text{if } x_2 = 0, \text{ and } x_1 = -(1 + t)/2, \\ \partial b(x) & \text{otherwise.} \end{cases}$$

Therefore, for x with $h(x) = t$ and $x_2 \neq 0$, the identity (37) implies that

$$\partial h(x) = \{\nabla b(x)\} = \left\{ \left(\frac{-(1+t)}{x_1 + 1 + t}, \frac{x_2}{x_1 + t + 1} \right)^T \right\}. \tag{39}$$

Since $h(x) \leq t$ satisfies the Slater condition, $N_{\bar{S}_t}(x) = \operatorname{cone}\{\partial h(x)\} = \operatorname{cone}\{\nabla b(x)\}$. On the other hand, by (39) and (38),

$$\|\nabla b(x)\|^2 = \frac{(1+t)^2 + 2x_1(1+t) + (1+t)^2}{[illegible]} \to 0 \quad \text{as } x_1 \to +\infty$$

Hence, there is no positive α *such that*

$$\alpha \mathbb{B}^\circ \cap N_{\bar{S}_t}(x) \subset co\{\partial h(x), 0\} \quad \forall x \in \bar{S}_t.$$

As a result, by Part 3 of Theorem 2.3, for $-1 < t < 0$, $\bar{S}_t$ *is not a set of weak sharp minima for* f_t. *However, as noted above* $\bar{S}_t^\infty$ *is a set of weak sharp minima for* f_t^∞.

REMARK. The functions $b(x)$ and $g(x)$ used in Examples 6.6 and 6.7 were taken from Li and Klatte (1999), where they were used to construct counter–examples in the study of global error bounds for systems of convex inequalities under the Slater constraint qualification.

7. A reduction theorem in the Lipschitzian case

In a number of applications, the underlying function f in (1) is known to possess certain Lipschitzian properties. In this case one can relate various notions of weak sharpness to a corresponding notion of weak sharpness for the function $f + K\,\mathrm{dist}(\cdot \mid S)$ for some value of $K > 0$. This reduction from a constrained to an unconstrained problem can often simplify the analysis. We use this reduction technique in our discussion of nondifferentiable systems of convex inequalities in Part II.

THEOREM 7.1 *Let* f, $\bar{S}$, *and* S *be as in Theorem 2.3.*

1. *If* f *is globally Lipschitz continuous on* X *with Lipschitz constant* L, *then* $\bar{S}$ *is a set of weak sharp minima for* f *over* S *with modulus* α *if and only if* $\bar{S}$ *is a set of weak sharp minima for the function* $f + (\alpha + L)\,\mathrm{dist}(\cdot \mid S)$ *with modulus* α.
2. *If* f *is locally Lipschitz on* X, *then* $\bar{S}$ *is a set of local weak sharp minima for* f *over* S *if and only if for every* $\bar{x} \in \bar{S}$ *there is an* $\epsilon > 0$, $\alpha > 0$, *and* $\widehat{L} > 0$ *such that*
$$f(x) + \widehat{L}\,\mathrm{dist}(x \mid S) \geq f(\bar{x}) + \alpha\,\mathrm{dist}(x \mid \bar{S}) \quad \textit{for all } x \in \bar{x} + \epsilon\mathbb{B}.$$
3. *If* f *is Lipschitz continuous on bounded subsets of* X, *then* $\bar{S}$ *is a set of boundedly weak sharp minima for* f *over* S *if and only if for every bounded subset* K *of* X *for which* $K \cap \bar{S} \neq \emptyset$ *there is an* $\widehat{L} > 0$ *and* $\alpha > 0$ *such that*
$$f(x) + \widehat{L}\,\mathrm{dist}(x \mid S) \geq f(\bar{x}) + \alpha\,\mathrm{dist}(x \mid \bar{S}) \quad \textit{for all } x \in S \cap K.$$

Proof. We only prove Part 2 of this theorem since the pattern of proof is identical for all three. Clearly, if for every $\bar{x} \in \bar{S}$ there is an $\epsilon > 0$, $\alpha > 0$, and $\widehat{L} > 0$ such that

$$f(x) + \widehat{L}\,\mathrm{dist}(x \mid S) \geq f(\bar{x}) + \alpha\,\mathrm{dist}(x \mid \bar{S}) \quad \text{for all } x \in \bar{x} + \epsilon\mathbb{B},$$

then $\bar{S}$ is a set of local weak sharp minima for f over S, so we only prove the and f is Lipschitz continuous on $\bar{x} + 3\epsilon\mathbb{B}$ with Lipschitz constant $L \geq 1$. Let $x \in \bar{x} + \epsilon\mathbb{B}$. Given $0 < \delta < \epsilon$, there is an $x_\delta \in S$ such that

$$\|x - x_\delta\| \leq \operatorname{dist}(x \mid S) + \delta \leq \epsilon + \delta < 2\epsilon.$$

This implies that

$$\|x_\delta - \bar{x}\| \leq \|x - x_\delta\| + \|x - \bar{x}\| < 3\epsilon.$$

Therefore, since the distance function is Lipschitz with Lipschitz constant 1,

$$\begin{aligned}
&\operatorname{dist}(x \mid \bar{S}) \leq \|x - x_\delta\| + \operatorname{dist}(x_\delta \mid \bar{S}) \\
&\leq \operatorname{dist}(x \mid S) + \alpha^{-1}[f(x_\delta) - f(\bar{x})] + \delta \\
&\leq \operatorname{dist}(x \mid S) + \alpha^{-1}[f(x) + L\|x - x_\delta\| - f(\bar{x})] + \delta \\
&\leq \operatorname{dist}(x \mid S) + \alpha^{-1}[f(x) + L(\operatorname{dist}(x \mid S) + \delta) - f(\bar{x})] + \delta \\
&\leq (1 + \alpha^{-1}L)[\operatorname{dist}(x \mid S) + \delta] + \alpha^{-1}[f(x) - f(\bar{x})],
\end{aligned}$$

or equivalently,

$$f(x) + (\alpha + L)[\operatorname{dist}(x \mid S) + \delta] \geq f(\bar{x}) + \alpha \operatorname{dist}(x \mid \bar{S}).$$

Since x and δ were chosen arbitrarily from $\bar{x} + \epsilon\mathbb{B}$ and $(0, \epsilon)$, respectively, this establishes the result. ∎

A Duality correspondences

THEOREM A.1 *[Duality Results] Let C be a nonempty closed convex subset of X and let E and F be nonempty convex subsets of X^*.*

1. *(Aubin and Ekeland, 1984, Proposition 4, page 168) For all $x \in C$,*
$$N_C(x) = \{z \in X^* : \langle z, x\rangle = \psi_C^*(z)\}.$$
2. *(Luenberger, 1968, Theorem 1, page 136) For all $y \in X$,*
$$\operatorname{dist}(y \mid C) = \max_{\|z\|_\circ \leq 1} [\langle z, y\rangle - \psi_C^*(z)].$$
3. *If C is a closed convex cone, then, for all $y \in X$,*
$$\operatorname{dist}(y \mid C) = \psi^*_{\mathbb{B}^\circ \cap C^\circ}(y).$$
4. *For all $y \in X$,*
$$\operatorname{dist}(y \mid C) = \sup_{x \in C} \operatorname{dist}(y \mid x + T_C(x)).$$
5. *(Burke, 1991, Proposition 3.1) The function $\operatorname{dist}(\cdot \mid C)$ is lower semicontinuous and convex with*
$$\operatorname{dom}(\operatorname{dist}(\cdot \mid C)) = X$$
and
$$\partial \operatorname{dist}(y \mid C) = \arg\max_{\|z\|_\circ \leq 1} [\langle z, y\rangle - \psi_C^*(z)]$$
$$= \begin{cases} \mathbb{B}^\circ \bigcap N_C(y), & \text{if } y \in C \\ \operatorname{bdry}(\mathbb{B}^\circ) \bigcap N \ldots (x) & \text{otherwise} \end{cases}$$

6. *Define* $\rho(x) = \operatorname{dist}(x \mid C)$. *Then, for all* $x \in C$ *and* $d \in X$,
$$\rho'(x; d) = \operatorname{dist}(d \mid T_C(x)) = \psi^*_{\mathbb{B}^\circ \cap N_C(x)}(d).$$
7. *(Aubin and Ekeland, 1984, Proposition 7, page 204)*
$$\psi^*_E(d) \le \psi^*_F(d)\ \forall\, d \in X \iff E \subset \operatorname{cl}^*(F).$$
8. *Let* K *be a nonempty closed convex cone in* X. *Then*
$$\psi^*_E(d) \le \psi^*_F(d)\ \forall\, d \in K \iff \psi^*_E(d) \le \psi^*_{F+K^\circ}(d)\ \forall\, d \in X$$
$$\iff E \subset \operatorname{cl}^*(F + K^\circ).$$
9. *(Moreau, 1965, The Moreau Decomposition) Suppose that* X *is a Hilbert space and that* $K \subset X$ *is a nonempty closed convex cone. Then each* $x \in X$ *has a unique representation of the form* $x = x_1 + x_2$ *where* $x_1 \in K$ *and* $x_2 \in K^\circ$ *with* $\langle x_1, x_2 \rangle = 0$. *Indeed, one has* $x_1 = P(x \mid K)$ *and* $x_2 = P(x \mid K^\circ)$.
10. *(Burke and Ferris, 1993, Lemma 2.1) For any nonempty subset* C *of the set* $\bar{S} = \arg\min f_0$ *and any set* $D \subset X^*$ *containing the origin in its interior, we have*
$$D \cap N_{\bar{S}}(x) \subset \partial f_0(x) \quad \forall x \in C, \tag{A.1}$$
if and only if
$$D \cap \Big(\bigcup_{x \in C} N_{\bar{S}}(x) \Big) \subset \bigcup_{x \in C} \partial f_0(x). \tag{A.2}$$

Proof. The proofs of Parts 1, 2, 5, and 7 can be found in the given citations. Part 3 is an immediate consequence of Part 2, and Part 6 is an immediate consequence of Parts 3 and 5 and the fact that $\rho'(x : d) = \psi^*_{\partial \operatorname{dist}(x|C)}(d)$ (Rockafellar, 1974, Part (b), Theorem 11). The following computation shows that Part 4 follows from Parts 1, 2, and 3:

$$\begin{aligned}
&\sup_{x \in C} \operatorname{dist}(y \mid x + T_C(x)) = \sup_{x \in C} \psi^*_{\mathbb{B}^\circ \cap N_C(x)}(y - x) \quad \text{[by Part 3]} \\
&= \sup_{\substack{x \in C,\ z \in N_C(x) \\ \|z\|_\circ \le 1}} [\langle z, y \rangle - \langle z, x \rangle] \\
&= \sup_{\|z\|_\circ \le 1} [\langle z, y \rangle - \psi^*_C(z)] \quad \text{[by Part 1]} \\
&= \operatorname{dist}(y \mid C) \quad \text{[by Part 2]}.
\end{aligned}$$

We now show Part 8 beginning with the first equivalence. First note that if $d \notin K$, then there is a $z \in K^\circ$ such that $\langle z, d \rangle > 0$. Let $w \in F$ and consider $w + \lambda z$ as $\lambda \to \infty$. Since $\langle w + \lambda z, d \rangle \nearrow \infty$ as $\lambda \to \infty$, we have that $\psi^*_{F+K^\circ}(d) = +\infty$. On the other hand, if $d \in K$, then, by the definition of the polar cone,

$$\begin{aligned}
&\psi^*_F(d) \le \psi^*_{F+K^\circ}(d) \\
&= \sup\{\langle w, d \rangle + \langle z, d \rangle : w \in F,\ z \in K^\circ\} \\
&\le \psi^*_F(d).
\end{aligned}$$

Hence, the first equivalence has been established. The second equivalence in

We now establish Part 10. This result is essentially proved in Burke and Ferris (1993), Lemma 2.1, but the statement given here is slightly different. We show that the proof given in Burke and Ferris (1993), Lemma 2.1, works in this setting. Clearly (A.1) implies (A.2) so we need only show the reverse implication. For this, suppose $x \in C$ and $w \in D \bigcap N_{\bar{S}}(x) \neq \emptyset$. By hypothesis, there exists $y \in C$ such that $w \in \partial f_0(y)$. Hence, for any $z \in \bar{S}$, we have

$$f_0(z) \geq f_0(y) + \langle w, z - y \rangle ,$$

or, equivalently,

$$0 \geq \langle w, z - y \rangle \ \forall z \in \bar{S}.$$

Therefore, $w \in N_{\bar{S}}(y)$ so that $w \in N_{\bar{S}}(x) \cap N_{\bar{S}}(y)$ which implies that

$$\langle w, y \rangle = \langle w, x \rangle . \tag{A.3}$$

However, $w \in \partial f_0(y)$ so $f_0(z) - f_0(y) \geq \langle w, z - y \rangle$, for all $z \in X$. Since y, $x \in \bar{S}$, $f_0(y) = f_0(x)$ so that (A.3) gives $f_0(z) - f_0(x) \geq \langle w, z - x \rangle$, for all z, or equivalently, $w \in \partial f_0(x)$. ∎

B Properties of recession functions

LEMMA B.1 *Let f_0 be the essential objective function defined in (2). If $\bar{S} =$ $\arg\min f_0$ is nonempty, then*

$$f_0^\infty = f^\infty + \psi_{S^\infty}, \tag{B.1}$$

$$\mathrm{cl}^*(\mathrm{dom}(f_0^*)) = \mathrm{cl}^*(\mathrm{dom}(f^*) + (S^\infty)^\circ), \tag{B.2}$$

$$\partial f_0^\infty(0) = \mathrm{cl}^*(\partial f^\infty(0) + N_{S^\infty}(0)), \tag{B.3}$$

$$(\bar{S}^\infty)^\circ = \bigcup_{y \in \bar{S}^\infty} N_{\bar{S}^\infty}(y), \tag{B.4}$$

$$\partial f_0^\infty(0) = \bigcup_{y \in \bar{S}^\infty} \partial f_0^\infty(y), \tag{B.5}$$

and

$$\bar{S}^\infty = \arg\min f_0^\infty = \arg\min \{f^\infty(y) \mid y \in S^\infty\} = [\mathrm{cone}(\mathrm{dom}(f_0^*))]^\circ. \tag{B.6}$$

Proof. We first show (B.1). By Rockafellar (1966), Theorem 3B, Part (b), we have $(y, \mu) \in \mathrm{epi}\, f_0^\infty$ if and only if there exists $x \in \mathrm{dom}(f_0) = \mathrm{dom}(f) \cap S$ such that

$$f(x + \lambda y) + \psi_S(x + \lambda y) \leq f(x) + \psi_S(x) + \lambda\mu < \infty \quad \forall \lambda \geq 0,$$

or equivalently,

$y \in S^\infty$ and there exists $x \in \mathrm{dom}(f) \cap S$ such that

But then, again by Rockafellar (1966), Theorem 3B, Part (b), this is equivalent to the statement that $y \in S^\infty$ and $f^\infty(y) \leq \mu$, which, in turn, is equivalent to $(y, \mu) \in \mathrm{epi}(f^\infty + \psi_{S^\infty})$.

Relation (B.2) follows from the fact that

$$\begin{aligned} &\psi^*_{\mathrm{cl}^*(\mathrm{dom}(f_0^*))} = f_0^\infty \\ &= f^\infty + \psi_{S^\infty} \\ &= \psi^*_{\mathrm{cl}^*(\mathrm{dom}(f^*))} + \psi^*_{(S^\infty)^\circ} \\ &= \psi^*_{\mathrm{cl}^*(\mathrm{dom}(f^*)+(S^\infty)^\circ)}, \end{aligned}$$

where the final equality follows from one of the many elementary properties of support functions listed in Aubin and Ekeland (1984), page 31. Relation (B.3) follows by combining (B.2) with (21) and the fact that $(S^\infty)^\circ = N_{S^\infty}(0)$ since S^∞ is a closed convex cone. Relation (B.4) follows from the fact that $\bar{S}^\infty$ is a closed convex cone and so

$$N_{\bar{S}^\infty}(y) \subset (\bar{S}^\infty)^\circ = N_{\bar{S}^\infty}(0) \text{ for every } y \in \bar{S}^\infty.$$

Relation (B.5) follows from (20), which implies that

$$\partial f_0^\infty(y) \subset \mathrm{cl}^*(\mathrm{dom}(f_0^*)) = \partial f_0^\infty(0) \text{ for every } y \in \mathrm{dom}(\partial f_0^\infty).$$

Finally, by Brøndsted and Rockafellar (1965),

$$z \in \partial f(x) \iff x \in \partial f^*(z).$$

Therefore, $\bar{S} = \partial f_0^*(0)$, since $\bar{S} = \arg\min_{x\in S} f$. Consequently, $0 \in \mathrm{dom}(f_0^*)$ and so, by (20), $f_0^\infty(y) = \psi^*_{\mathrm{cl}^*(\mathrm{dom}(f_0^*))}(y) \geq 0$ for all $x \in X$. Hence, by Rockafellar (1966), Theorem 3B, Parts (a) and (d), and (B.1), we have

$$\begin{aligned} &\bar{S}^\infty = \{y \mid f_0^\infty(y) \leq 0\} = \arg\min f_0^\infty \\ &= \arg\min \{f^\infty(y) \mid y \in S^\infty\} = [\mathrm{cone}(\mathrm{dom}(f_0^*))]^\circ, \end{aligned}$$

where the final equivalence comes from the fact that

$$\left\{y \mid f_0^\infty(y) = \psi^*_{\mathrm{cl}^*(\mathrm{dom}(f_0^*))}(y) \leq 0\right\} = [\mathrm{cone}(\mathrm{dom}(f_0^*))]^\circ. \qquad \blacksquare$$

RREMARK. Note that the formula (B.3) always holds. This is an important instance in which the addition formula (6) must always hold.

THEOREM B.2 *Assume that X is a Banach space, and let $f: X \mapsto \overline{\mathbb{R}}$ be a lower semi-continuous convex function that is not everywhere $+\infty$. Then the weak*

Proof. The Brøndsted-Rockafellar theorem, Brøndsted and Rockafellar (1965), implies that $\mathrm{dom}\,(\partial f)$ is dense in $\mathrm{dom}\,(f)$ in the norm topology. Hence the strong closure of $\mathrm{dom}\,(\partial f)$ equals the strong closure of $\mathrm{dom}\,(f)$. Since $\mathrm{dom}\,(f)$ is convex, we have that the weak and strong closures of $\mathrm{dom}\,(f)$ coincide. Therefore,

$$\begin{aligned}&\mathrm{cl}(\mathrm{dom}\,(\partial f)) \subset \text{w-cl}\,(\mathrm{dom}\,(\partial f)) \subset \text{w-cl}\,(\mathrm{dom}\,(f))\\ &= \mathrm{cl}(\mathrm{dom}\,(f)) = \mathrm{cl}(\mathrm{dom}\,(\partial f)).\end{aligned}$$ ■

The following technical lemma is used to relate the results of Theorems 2.3 and 4.1.

LEMMA B.3 *Let C be a nonempty closed convex subset of normed linear space X. Then*

$$\bigcup_{x\in C} N_C\,(x) = \mathrm{dom}\,(\partial\psi_C^*) \subset \mathrm{dom}\,(\psi_C^*) = \mathrm{bar}\,(C)\,, \tag{B.7}$$

and

$$\begin{aligned}&w\text{-}cl\Big(\bigcup_{x\in C} N_C\,(x)\Big) = \mathrm{cl}\,\Big(\bigcup_{x\in C} N_C\,(x)\Big) = \mathrm{cl}\,(\mathrm{bar}\,(C)) \subset \mathrm{cl}^*\,\Big(\bigcup_{x\in C} N_C\,(x)\Big)\\ &\subset \mathrm{cl}^*\,(\mathrm{bar}\,(C)) = (C^\infty)^\circ.\end{aligned} \tag{B.8}$$

In particular, we obtain

$$\Big(\bigcup_{x\in C} N_C\,(x)\Big)^\circ \supset C^\infty. \tag{B.9}$$

If X is assumed to be reflexive, we obtain equality throughout (B.8) as well as in (B.9). If it is further assumed that X is finite-dimensional, then

$$\mathrm{ri}\,(\mathrm{bar}\,(C)) \subset \bigcup_{x\in C} N_C\,(x)\,. \tag{B.10}$$

Proof. Since $\partial\psi_C = N_C$, we obtain from Brøndsted and Rockafellar (1965) that

$$z \in N_C\,(x) \iff x \in \partial\psi_C^*(z).$$

The relation (B.7) immediately follows. Since ψ_C^* is proper lower semi-continuous and convex Aubin and Ekeland (1984), p. 27, and X^* is a Banach space, the first equivalence in (B.8) follows from Theorem B.2 and (B.7). To obtain the second equivalence in (B.8), one combines (B.7) with Brøndsted and Rockafellar (1965), Theorem 2. The inclusions in (B.8) are obvious. The third equivalence in (B.8) follows from (18). The relation (B.9) follows by taking the polar in (B.8). The final inclusion (B.10) is an immediate consequence of Rockafellar (1970), Theorem 23.4. ■

Acknowledgment. The authors wish to thank the anonymous referees for a very thorough reading of the manuscript, for inspiring comments, and for

References

AL-KHAYYAL, F.A. and KYPARISIS, J. (1991) Finite convergence of algorithms for nonlinear programs and variational inequalities. *Journal of Optimization Theory and Applications*, **70** (2), 319–332.

AUBIN, J.-P. and EKELAND, I. (1984) *Applied Nonlinear Analysis*. Wiley-Interscience.

AUSLENDER, A. COMINETTI, R. and CROUZEIX, J.-P. (1993) Convex functions with unbounded level sets and applications to duality theory. *SIAM J. Optimization*, **3**, 669–695.

AUSLENDER, A. and CROUZEIX, J.-P. (1988) Global regularity theorem. *Mathematics of Operations Research*, **13**, 243–253.

BAUSCHKE, H. (1996) Projection algorithms and monotone operators. Ph.D. Thesis, Simon Fraser University, Dept. of Math., Burnaby, British Columbia, V5A 1S6, Canada.

BAUSCHKE, H. BORWEIN, J. and LI, W. (1999) Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization. *Mathematical Programming*, **86**, 135–160.

BONNANS, J.-F. and IOFFE, A.-D. (1995) Quadratic growth and stability in convex programming problems with multiple solutions. *J. Convex Anal.*, **2**, 41–57.

BRØNDSTED, R. and ROCKAFELLAR, R.T. (1965) On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.*, **16**, 605–611.

BURKE, J.V. (1991) An exact penalization viewpoint of constrained optimization. *SIAM J. Control and Optim.*, **29**, 968–998.

BURKE, J.V. and FERRIS, M.C. (1993) Weak sharp minima in mathematical programming. *SIAM J. Control and Optim.*, **31**, 1340–1359.

BURKE, J.V. and FERRIS, M.C. (1995) A Gauss–Newton method for convex composite optimization. *Mathematical Programming*, **71**, 179–194.

BURKE, J.V. and MORÉ, J.J. (1988) On the identification of active constraints. *SIAM J. Numer. Anal.*, **25**, 1197–1211.

BURKE, J.V. and TSENG, P. (1996) A unified analysis of Hoffman's bound via Fenchel duality. *SIAM J. Optimization*, **6**, 265–282.

BURKE, J.V., LEWIS, A. and OVERTON, M. (2000) Optimizing matrix stability. *Proc. of the American Math. Soc.*, **129**, 1635–1642.

BURKE, J.V., LEWIS, A. and OVERTON, M. (2001) Optimal stability and eigenvalue multiplicity. *Foundations of Computational Mathematics*, **1**, 205–225.

CROMME, L. (1978) Strong Uniqueness. *Numerische Mathematik*, **29**, 179–193.

DENG, S. (1997) Computable error bounds for convex inequality systems in reflexive Banach spaces. *SIAM J. Optimization*, **7**, 274–279.

DENG, S. (1998) Global error bounds for convex inequality systems in Banach

EKELAND, I. and TEMAM, R. (1976) *Convex Analysis and Variational Problems.* North Holland.

FERRIS, M.C. (1988) Weak Sharp Minima and Penalty Functions in Mathematical Programming. PhD thesis, University of Cambridge, Cambridge.

FERRIS, M.C. (1990) Iterative linear programming solution of convex programs. *Journal of Optimization Theory and Applications*, **65**, 53–65.

FERRIS, M.C. (1991) Finite termination of the proximal point algorithm. *Mathematical Programming*, **50**, 359–366.

GOWDA, M.S. (1996) An analysis of zero set and global error bound properties of a piecewise affine function via its recession function. *SIAM Journal on Matrix Analysis*, **17**, 594–609.

HENRION, R. and OUTRATA, J. (2001) A subdifferential condition for calmness of multifunctions. *Journal of Mathematical Analysis and Applications*, **258**, 110–130.

HIRIART-URRUTY, J.-B. and LEMARÉCHAL, C. (1993) *Convex Analysis and Minimization Algorithms I, volume 306 of Grundlehren der mathematischen Wissenschaften.* Springer–Verlag.

HOFFMAN, A.J. (1952) On approximate solutions to systems of linear inequalities. *J. Res. Nat. Bur. Standards*, **49**, 263–265.

HU, H. and WANG, Q. (1989) On approximate solutions of infinite systems of linear inequalities. *Linear Algebra and its applications*, **114/115**, 429–438.

JOURANI, A. (2000) Hoffman's error bounds, local controllability, and sensitivity analysis. SIAM J. Control and Optim., **38**, 947–970.

KLATTE, D. (1998) Hoffman's error bound for systems of convex inequalities. in A. V. Fiacco, ed., Mathematical Programming with Data Perturbations. Marcel Dekker Publ.

KLATTE, D. and LI, W. (1999) Asymptotic constraint qualifications and global error bounds for convex inequalities. *Mathematical Programming*, **84**, 137–160.

LEWIS, A.S. and PANG, J.S. (1996) Error bounds for convex inequality systems. in J. P. Crouzeix, ed., Proceedings of the fifth Symposium on Generalized Convexity, Luminy-Marseille.

LI, W. (1997) Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities. *SIAM J. on Optimization*, **7**, 966–978.

LI, C. and WANG, X. (2002) On convergence of the Gauss-Newton method for convex composite optimization. *Mathematical Programming*, **91**, 349–356.

LUO, X.D. and LUO, Z.Q. (1994) Extension of Hoffman's error bound to polynomial systems. *SIAM J. on Optimization*, **4**, 383–392.

LUO, Z.-Q. and PANG, J.-S. (2000) Preface: Error bounds in mathematical programming. *Mathematical Programming*, **88**, 221–222.

LUENBERGER, D.G. (1968) *Optimization by Vector Space Methods.* John Wiley & Sons, Inc..

MANGASARIAN, O.-L. (1985) A condition number for differentiable convex inequalities. *Mathematics of Operations Research*, **10**, 175–179.

MOREAU, J.J. (1965) Proximité et dualité dans un espace Hilbertien. *Bull. Soc. Math. Fr.*, **93**, 273–299.

PANG, J.S. (1997) Error bounds in mathematical programming. *Mathematical Programming*, **79**, 299–332.

POLYAK, B.T. (1979) Sharp Minima. *Institute of Control Sciences Lecture Notes*, Moscow, USSR, 1979; Presented at the IIASA Workshop on Generalized Lagrangians and Their Applications, IIASA, Laxenburg, Austria.

POLYAK, B.T. (1987) *Introduction to Optimization*. Optimization Software, Inc., Publications Division, New York.

ROCKAFELLAR, R.T. (1966) Level sets and continuity of conjugate convex functions. *Trans. Amer. Math. Soc.*, **123**, 46–63.

ROCKAFELLAR, R.T. (1970) *Convex Analysis*. Princeton University Press.

ROCKAFELLAR, R.T. (1974) *Conjugate Duality and Optimization*. SIAM.

ROCKAFELLAR, R.T. and WETS, R.J.-B. (1998) *Variational Analysis*. Springer.

STUDNIARSKI, M. and WARD, D.E. (1999) Weak sharp minima: characterizations and sufficient conditions. *SIAM J. Control and Optim.*, **38**, 219–236.

YE, J.J. (1998) New uniform parametric error bounds. *Journal of Optimization Theory and Applications*, **98**, 197–219.

YE, J.J. and ZHU, D.L. (1995) Optimality conditions for bilevel programming problems. *Optimization*, **33**, 9–27.

YE, J.J. and ZHU, D.L. (1997) A note on optimality conditions for bilevel programming problems. *Optimization*, **39**, 361–366.

YE, J.J. ZHU, D.L. and ZHU, Q.J. (1997) Exact penalization and necessary optimality conditions for generalized bilevel programming problems. *SIAM Journal on Optimization*, **7**, 481–507.