WEAK SHARP MINIMA
AND EXACT PENALTY FUNCTIONS

by

Michael C. Ferris

Computer Sciences Technical Report #779

June 1988

Weak Sharp Minima and Exact Penalty Functions

Michael C. Ferris*

Computer Sciences Department

University of Wisconsin

Madison, Wisconsin 53706

June 1988

Abstract. This paper concerns the notion of a weak sharp minimum, a generalization of the sharp minimum property described by Polyak[11].

We give several equivalent definitions of the property and use the notion to prove finite termination of the proximal point algorithm. Several results concerning the application of penalty functions to problems which are possibly infeasible are proven and we show that the penalty problem having a weak sharp minimum is sufficient to guarantee exactness.

Some applications to convex and concave programs are described briefly.

Key words. Weak sharp minima, penalty functions, proximal point, finite termination

Abbreviated title. Weak Sharp Minima

*This material is based on research supported by National Science Foundation Grants DCR–8521228 and CCR–8723091 and Air Force Office of Scientific Research Grant AFOSR–86–0172

1 Introduction

In this paper we shall be concerned principally with the general programming problem

$$\underset{x \in S}{\text{minimize}} \quad \phi(x) \tag{1}$$

where ϕ is a lower semicontinuous function defined on $\mathbb{R}^n$, having values in $\mathbb{R}$ and S is a closed set in $\mathbb{R}^n$. For the most part we will be concerned with the convex program, where we assume further that ϕ and S are convex. In this case, ϕ_S, defined by

$$\phi_S(x) := \phi(x) + \psi_S(x)$$

is a lower semicontinuous closed convex function, since the indicator function of the set S, ψ_S, is respectively closed and convex if and only if S is closed and convex. Thus (1) can be rewritten as

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \phi_S(x) \tag{2}$$

We write $\bar{S}$ for the optimal solution set of (1), $\bar{S} := \arg\min_{x \in S} \phi(x)$. We shall frequently require this set to be nonempty, in order that a projection operation onto this set is well–defined.

The principal concern of this work is a property that we call **weak sharp minima**. The motivation behind this definition comes from the literature where the idea of a **sharp minimum** or the equivalent notion of **strong uniqueness** has been used. The name sharp minimum, due to Polyak[11], is the subject of one of his papers[10]. References to strong uniqueness abound in the literature of approximation theory, for example [9], [4] and [1]. The definition is as follows.

Definition 1 *A function ϕ defined on a set S is said to have a* **sharp minimum** *at $\bar{x} \in S$ if there exists an $\alpha > 0$ such that*

$$\phi(x) - \phi(\bar{x}) \geq \alpha \|x - \bar{x}\| \quad \textit{for all } x \in S,$$

where $\|\cdot\|$ represents some norm on $\mathbb{R}^n$.

The essential idea is that we can place a norm function underneath ϕ, which touches ϕ at $\bar{x}$ and has a strictly positive slope. Perhaps the principal shortcoming of the concept of a sharp minimum is the implied uniqueness of the solution of the problem. This immediately rules out one of the most fundamental problems of optimization, namely linear programming. In contrast, our new concept of a weak sharp minimum, which we define below, avoids this serious difficulty.

We quickly describe our notation. All vectors are column vectors and given a vector x, we denote its ith component by x_i. We say $x \geq 0$ if one has $x_i \geq 0$ for all i. For a given scalar λ, we define $(\lambda)_+ := \max\{0, \lambda\}$. If $x \in \mathbb{R}^n$, we write x_+ for the vector whose ith component is $(x_i)_+$. Superscripts are used to distinguish between vectors, e.g., x^1, x^2, etc. We use $\langle , \rangle$ for the inner product. If K is any set, we define the indicator function of that set, ψ_K, as

$$\psi_K(x) := \begin{cases} 0 & \text{if } x \in K \\ \infty & \text{if } x \notin K \end{cases}$$

and

$$\operatorname{cone} K := \{\mu x \mid x \in K, \mu > 0\}$$

The distance function, $d(\cdot, K)$, is defined by

$$d(x, K) := \inf\{\|x - y\| \mid y \in K\}$$

where $\|\cdot\|$ represents a given norm. If K is a closed convex set in $\mathbb{R}^n$, then given $x \in \mathbb{R}^n$, we write $P_K(x)$ for the projection of x on K

$$P_K(x) := \arg\min\{\|z - x\| \mid z \in K\}$$

and the normal cone of K at x is given by

$$N_K(x) := \{x^* \mid \langle x^*, y - x\rangle \leq 0 \qquad \forall y \in K\}$$

B represents the closed unit ball in $\mathbb{R}^n$.

2 Subgradients and weak sharp minima

The following is our central definition.

Definition 2 *Let $\bar{S}$ be the non-empty optimal solution set of (1). We say that ϕ has a* **weak sharp minimum** *(on S) if for some norm on $\mathbb{R}^n$ there exists an $\alpha > 0$ such that for all $x \in S$*

$$\phi(x) - \phi(P_{\bar{S}}(x)) \geq \alpha \|x - P_{\bar{S}}(x)\|$$

where $P_{\bar{S}}(\cdot)$ is the projection onto $\bar{S}$ under the given norm.

Note that the definition holds independently of any convexity assumptions on S and ϕ, provided that the projection is well defined. Furthermore, we note that for any norm $\phi(P_{\bar{S}}(p)x)$ is a constant for every $x \in S$, so we can rewrite the above using

$$\bar{\phi} := \phi(P_{\bar{S}}(x))$$

Our first task is to show that this definition is independent of the choice of norm on the space $\mathbb{R}^n$. This essentially relies upon the equivalence of norms in finite dimensional spaces.

Lemma 3 *ϕ has a weak sharp minimum (on S) $\iff$*

$$\phi(x) - \bar{\phi} \geq \beta_2 \|x - P_{\bar{S}}(x)\|_2 \quad \text{for some } \beta_2 > 0$$

Proof It follows from the definition that

$$\phi(x) - \bar{\phi} \geq \beta_p \left\|x - P_{p,\bar{S}}(x)\right\|_p$$

for some norm $\|\|_p$. Hence

$$\phi(x) - \bar{\phi} \geq \beta_p \gamma_{p,2} \left\|x - P_{p,\bar{S}}(x)\right\|_2 \geq \beta_p \gamma_{p,2} \left\|x - P_{2,\bar{S}}(x)\right\|_2$$

by the equivalence of norms on $\mathbb{R}^n$ and the definition of the projection operator. ∎

We recall the definition of a sharp minimum (given by Definition 1) in order to motivate Definition 2, and the term "weak" sharp minimum. The following lemma, which is easy to prove, gives the connection between these properties when the function ϕ is convex.

Lemma 4 *Suppose that ϕ is a convex function defined on a convex set S.*

(a) *ϕ has a sharp minimum at $\bar{x}$ on S*

(b) *$\bar{x}$ is the unique minimum of ϕ on S, that is* $\arg\min_{x \in S} \phi(x) = \{\bar{x}\}$

(c) $0 \in \operatorname{int} \partial\phi_S(\bar{x})$

(d) *ϕ has a weak sharp minimum on S*

(e) *ϕ has a weak sharp minimum on S and* $\arg\min_{x \in S} \phi(x) = \{\bar{x}\}$

Then

$$(a) \iff (c) \implies (b)$$

and

$$(e) \implies (a) \implies (d)$$

We now prove a generalization of the above result for a weak sharp minimum. In the sequel it is assumed that $\|\cdot\| \equiv \|\cdot\|_2$. In this section, ϕ and S are convex, and we assume $\bar{S}$ is nonempty. We use the following notation

$$\partial\phi_S(\bar{S}) := \bigcup_{x \in \bar{S}} \partial\phi_S(x)$$

To generalize the subgradient condition given above, we introduce the following set

$$K_{\bar{S}} := \operatorname{cone}\{x - P_{\bar{S}}(x) \mid x \in \mathbb{R}^n\}$$

The use of this set enables us to ignore the directions of recession of the optimal solution set. The following lemmas are essentially technical in nature. They are useful in the proofs of the main results of this paper and show more clearly how $K_{\bar{S}}$ relates to our problem. The first lemma gives a characterization of the projection operator.

Lemma 5 *Let C be a non-empty, closed convex set in $\mathbb{R}^n$ and $p \in C$. Then*

$$\langle z - p, c - p \rangle \leq 0 \text{ for all } c \in C \iff p = P_C(z)$$

An easy corollary of this result is

Corollary 6 *A necessary and sufficient condition for the linear functional $\langle x^*, \cdot \rangle$ to attain its supremum over $\bar{S}$ is that $x^* \in K_{\bar{S}}$.*

Furthermore, when $\bar{S}$ is compact, the following holds

Lemma 7 *If $\bar{S}$ is compact then $K_{\bar{S}} = \mathbb{R}^n$.*

Proof Clearly, $K_{\bar{S}} \subseteq \mathbb{R}^n$. For the converse, let $y \in \mathbb{R}^n$. The function $\langle y, \cdot \rangle$ is continuous and hence, by compactness of $\bar{S}$, the problem

$$\underset{z \in \bar{S}}{\text{maximize}} \quad \langle y, z \rangle$$

has a solution $z^* \in \bar{S}$. Let $\bar{z} = z^* + y$, so that it remains to prove $z^* = P_{\bar{S}}(\bar{z})$. However, $\langle y, z^* \rangle \geq \langle y, z \rangle$ for all $z \in \bar{S}$, so that

$$\langle \bar{z} - z^*, z - z^* \rangle \leq 0 \quad \text{for all } z \in \bar{S}$$

implying that $z^* = P_{\bar{S}}(\bar{z})$, by Lemma 5. ∎

We can now relate the distance function given in the introduction to the set $K_{\bar{S}}$.

Lemma 8

$$d(y, \bar{S}) = \sup_{x^* \in B \bigcap K_{\bar{S}}} \{\langle x^*, y \rangle - \psi^*_{\bar{S}}(x^*)\}$$

Proof By [13, page 146]

$$\begin{aligned} d(y,\bar{S}) &= \sup_{x^* \in B} \{\langle x^*, y\rangle - \psi_{\bar{S}}^*(x^*)\} \\ &= \sup_{x^* \in B \bigcap K_{\bar{S}}} \{\langle x^*, y\rangle - \psi_{\bar{S}}^*(x^*)\} \end{aligned}$$

We now show the reverse inequality. Let $y \in \mathbb{R}^n$ and define $z := y - P_{\bar{S}}(x)$. It follows from the definition that $d(y,\bar{S}) \leq \|z\|$ and

$$\begin{aligned} \sup_{x^* \in B \bigcap K_{\bar{S}}} \{\langle x^*, y\rangle - \psi_{\bar{S}}^*(x^*)\} &\geq \left\langle \frac{z}{\|z\|}, y \right\rangle - \psi_{\bar{S}}^*(\frac{z}{\|z\|}) \\ &= \left\langle \frac{z}{\|z\|}, y \right\rangle + \left\langle \frac{z}{\|z\|}, P_{\bar{S}}(y) \right\rangle - \psi_{\bar{S}}^*(\frac{z}{\|z\|}) \\ &= \|z\| \end{aligned}$$

the last inequality following from Lemma 5. ∎

The main result of this section follows. We relate the definition of a weak sharp minimum to the subdifferentials of the optimal solution set.

Theorem 9 *ϕ has a weak sharp minimum on S if and only if there exists $\epsilon > 0$ with*

$$\epsilon(B \bigcap K_{\bar{S}}) \subseteq \partial\phi_S(\bar{S})$$

Proof ($\Rightarrow$) Suppose that ϕ has a weak sharp minimum (on S). Let $x^* \in B \bigcap K_{\bar{S}}$. It follows from the definition of a weak sharp minimum, that $\forall y \in S$

$$\phi(y) \geq \bar{\phi} + \alpha d(y,\bar{S})$$

and hence

$$\phi_S(y) \geq \bar{\phi} + \alpha d(y,\bar{S}) \quad \forall y$$

where the distance and indicator functions are defined in the introduction. By Lemma 8

$$\begin{aligned}\phi_S(y) &\geq \bar{\phi} + \alpha d(y, \bar{S}) \\ &= \bar{\phi} + \alpha \sup_{z^* \in B \bigcap K_{\bar{S}}} (\langle z^*, y\rangle - \psi^*_{\bar{S}}(z^*))\end{aligned}$$

and since $x^* \in B \bigcap K_{\bar{S}}$

$$\begin{aligned}\phi_S(y) &\geq \bar{\phi} + \alpha(\langle x^*, y\rangle - \psi^*_{\bar{S}}(x^*)) \\ &= \bar{\phi} + \alpha(\langle x^*, y\rangle - \sup_{z \in \bar{S}} \langle x^*, z\rangle)\end{aligned}$$

However, by Corollary 6, the supremum is attained at a particular point $\bar{x}$ depending on x^*. Thus

$$\phi_S(y) \geq \bar{\phi} + \alpha(\langle x^*, y\rangle - \langle x^*, \bar{x}(x^*)\rangle)$$

and since $\bar{x}(x^*) \in \bar{S}$, it follows that

$$\phi_S(y) \geq \phi_S(\bar{x}(x^*)) + (\langle \alpha x^*, y - \bar{x}(x^*)\rangle)$$

Hence $\alpha x^* \in \partial\phi_S(\bar{x}(x^*)) \subseteq \partial\phi_S(\bar{S})$. Now x^* arbitrary in $B \bigcap K_{\bar{S}}$ implies $\alpha(B \bigcap K_{\bar{S}}) \subseteq \partial\phi_S(\bar{S})$.

($\Leftarrow$) Suppose that $\epsilon(B \bigcap K_{\bar{S}}) \subseteq \operatorname{int} \partial\phi_S(\bar{S})$ and let $x \in S$. This means that

$$\forall x^* \in B \bigcap K_{\bar{S}} \quad \exists \bar{x} \in \bar{S} \quad \text{such that} \quad \epsilon x^* \in \partial\phi_S(\bar{x})$$

Hence, for all y

$$\phi_S(y) \geq \bar{\phi} + \langle \epsilon x^*, y - \bar{x}\rangle \tag{3}$$

which gives

$$0 \quad \geq \quad \langle x^*, y - \bar{x}\rangle \quad \forall y \in \bar{S}$$

and hence $\psi^*_{\bar{S}}(x^*) = \langle x^*, \bar{x}\rangle$. It now follows from (3) and the linearity of $\langle , \rangle$ that for all y

$$\begin{aligned}\phi_S(y) &\geq \bar{\phi} + (\langle \epsilon x^*, y\rangle - \langle \epsilon x^*, \bar{x}\rangle) \\ &= \bar{\phi} + \epsilon(\langle x^*, y\rangle - \psi^*_{\bar{S}}(x^*))\end{aligned}$$

so that, for all y

$$\begin{aligned} \phi_S(y) &\geq \bar{\phi} + \epsilon \sup_{x^* \in B \bigcap K_{\bar{S}}} (\langle x^*, y \rangle - \psi_{\bar{S}}^*(x^*)) \\ &= \bar{\phi} + \epsilon d(y, \bar{S}) \end{aligned}$$

by Lemma 8. Since $x \in S$, we find $\phi(x) \geq \bar{\phi} + \epsilon \|x - P_{\bar{S}}(x)\|$. ∎

Corollary 10 *If $\bar{S}$ is compact then a weak sharp minimum is equivalent to*

$$0 \in \operatorname{int} \partial \phi_S(\bar{S})$$

3 Penalty functions with empty feasible sets

Consider the constrained minimization problem

$$\begin{array}{ll} \text{minimize} & \phi(x) \\ \text{subject to} & x \in S := X_0 \bigcap X_1 \end{array} \tag{4}$$

where X_0 and X_1 are subsets of $\mathbb{R}^n$ and $\phi\colon X_0 \to \mathbb{R}$. We propose to prove a series of results concerning the application of penalty functions to (4), with the minimum of assumptions, much in the spirit of Mangasarian[6]. In particular, the aim will be to remove the standard assumption of feasibility of the problem, that is to say that the feasible set, S, may be empty.

A well–known approach to solve an optimization problem is to use a penalty function and to minimize a combination of the objective function and a penalty parameter multiplied by some function which penalizes the constraint violation, increasing the penalty parameter to infinity in order to obtain a solution to the original problem. It is contended here that this approach may not necessarily be the best manner of treating the problem, especially for an ill–posed problem for which the feasible region may be empty. More precisely, one should try to find the point which minimizes the objective amongst those points which minimize the constraint violation, since this approach will give an answer which may be of some use in the case of an empty or nearly empty feasible set. However, in this section, we show that under relatively few assumptions, this is indeed what the penalty method produces.

To this end, we associate with (4), the classical exterior penalty problem

$$\underset{x \in X_0}{\text{minimize}} \quad P(x,\alpha) := \phi(x) + \alpha Q(x) \tag{5}$$

where α is in $\mathbb{R}_+$, and $Q: X_0 \to \mathbb{R}_+$ is a **penalty function** for S, that is, $Q(x) = 0$ for $x \in S$ and $Q(x) > 0$, otherwise.

Lemma 11 *Suppose that for some $\alpha^* \geq 0$, $P(x,\alpha^*)$ is bounded below on X_0. Then*

$$\inf_{x \in S} \phi(x) > -\infty$$

Proof If $S = \emptyset$, then the result is trivial. Otherwise, assume there exits a sequence $\{x^i\} \subseteq S$, such that $\lim_{i\to\infty} \phi(x^i) = -\infty$. However, on S, $Q(x) = 0$, so that this sequence means that $P(x^i, \alpha^*)$ is unbounded below, which is a contradiction. ∎

The following elementary but very useful monotonicity result is found in [6]. The statement of the proposition given here is simplified by the observation made in Lemma 11.

Proposition 12 *Let $x^i \in X_0$ be a solution of*

$$\underset{x \in X_0}{\text{minimize}} \quad P(x,\alpha_i)$$

for $i = 1,2$ with $\alpha_2 > \alpha_1 \geq 0$. Then

$$\begin{aligned} Q(x^1) &\geq Q(x^2) \\ \phi(x^1) &\leq \phi(x^2) \\ P(x^1,\alpha_1) &\leq P(x^2,\alpha_2) \end{aligned}$$

The following theorem goes some way to proving the result that we outlined in the introduction of this section. However, there is no guarantee here that the sequence $\{x^i\}$ converges, or even has an accumulation point – this will be considered later in this section. Furthermore, the inequality given by (6) may be strict, and we have to wait until we have described the concept of sharp penalty functions in order to make this an equality.

Theorem 13 *Let $\{\alpha_i\}$ be an increasing unbounded sequence of positive numbers and let $\{x^i\}$ be a corresponding sequence of points in $X_0 \setminus S$ such that*

$$-\infty < P(x^i, \alpha_i) = \min_{x \in X_0} P(x, \alpha_i).$$

If we define $\bar{Q} := \lim_{i \to \infty} Q(x^i)$, then

$$\bar{Q} = \inf_{x \in X_0} Q(x)$$

Further, if we let

$$X_Q := \{x \in X_0 \mid Q(x) \leq \bar{Q}\}$$

and assume $X_Q \neq \emptyset$ then

$$-\infty < \lim_{i \to \infty} \phi(x^i) \leq \inf_{x \in X_Q} \phi(x) \tag{6}$$

Proof It follows from Proposition 12 that the sequence $\{Q(x^i)\}$ is nonincreasing and bounded below by 0, and hence converges to some $\bar{Q} \geq 0$ and $Q(x^i) \geq \bar{Q}$, for $i = 1, 2, \ldots$. Clearly $\bar{Q} \geq \inf_{x \in X_0} Q(x)$, so we suppose that $\bar{Q} > \inf_{x \in X_0} Q(x)$ and obtain a contradiction. Let $\epsilon = \bar{Q} - \inf_{x \in X_0} Q(x) > 0$. By the definition of infimum we know there exists $\bar{x} \in X_0$ such that $Q(\bar{x}) \leq \bar{Q} - \epsilon/2$. Then, for all i, by the definition of x^i

$$\phi(x^i) + \alpha_i Q(x^i) \leq \phi(\bar{x}) + \alpha_i Q(\bar{x})$$

which implies that

$$\alpha_i(Q(x^i) - Q(\bar{x})) \leq \phi(\bar{x}) - \phi(x^i)$$

and so

$$\begin{aligned}
\alpha_i \frac{\epsilon}{2} &\leq \alpha_i(\bar{Q} - Q(\bar{x})) \\
&\leq \alpha_i(Q(x^i) - Q(\bar{x})) \\
&\leq \phi(\bar{x}) - \phi(x^i)
\end{aligned}$$

But Lemma 11 shows that $\phi(x^i)$ is bounded below, so that $\phi(\bar{x})$ is unbounded above, a contradiction.

For the final part of the theorem we note from Proposition 12 that the sequence $\{\phi(x^i)\}$ is nondecreasing. For any $\epsilon > 0$, pick $x(\epsilon) \in X_Q$ such that

$$\phi(x(\epsilon)) < \inf_{x \in X_Q} \phi(x) + \epsilon$$

Then

$$\begin{aligned}
\epsilon + \inf_{x \in X_Q} \phi(x) > \phi(x(\epsilon)) &= P(x(\epsilon), \alpha_i) - \alpha_i \bar{Q} \\
&\geq P(x^i, \alpha_i) - \alpha_i \bar{Q} \\
&= \phi(x^i) + \alpha_i Q(x^i) - \alpha_i \bar{Q} \\
&\geq \phi(x^i)
\end{aligned}$$

Thus $\phi(x^i) \leq \inf_{x \in X_Q} \phi(x)$ so that $\{\phi(x^i)\}$ converges to $\bar{\phi}$ and $\phi(x^i) \leq \bar{\phi} \leq \inf_{x \in X_Q} \phi(x)$. ∎

Note that if the feasible set S, is nonempty, then the following lemma shows that the results of Mangasarian[6] can be recovered as special cases of the theory which is presented here. The proof is trivial and is omitted.

Lemma 14 *If $S \neq \emptyset$ then $X_Q = S$.*

The following lemma clears up a loose end in Theorem 13. We specify a condition on Q which guarantees that either $X_Q \neq \emptyset$ or the sequence $\{x^i\}$ does not have an accumulation point. In the latter case, there is really nothing left to do; it would seem that the choice of penalty function was not a good one.

Lemma 15 *Suppose that Q is lower semicontinuous. Then either the sequence $\{x^i\}$ given in Theorem 13 has no accumulation point, or $X_Q \neq \emptyset$.*

Proof Suppose $\{x^i\}$ has an accumulation point $\bar{x}$. Then, by lower semicontinuity

$$\inf_{y \in X_0} Q(y) = \lim_{j \to \infty} Q(x^{i_j}) \geq Q(\bar{x})$$

which implies that $\bar{x} \in X_Q$. ∎

We can also prove a stronger result in the case that the sequence $\{x^i\}$ has an accumulation point, namely that the accumulation point actually solves the original problem. In fact we show that the accumulation point does exactly what we had hoped for in the introduction to this section, it solves

$$\underset{x \in X_Q}{\text{minimize}} \quad \phi(x)$$

Theorem 16 *Let $\{\alpha_i\}$ be an increasing unbounded sequence of positive numbers and let $\{x^i\}$ be a corresponding sequence of points in $X_0 \setminus S$ such that*

$$P(x^i, \alpha_i) = \min_{x \in X_0} P(x, \alpha_i)$$

with an accumulation point $\bar{x} \in X_0$ and hence $X_Q \neq \emptyset$. If ϕ and Q are lower semicontinuous at $\bar{x}$, then $Q(\bar{x}) = \bar{Q}$, and $\bar{x}$ solves

$$\underset{x \in X_Q}{\text{minimize}} \quad \phi(x) \tag{7}$$

Furthermore,

$$\lim_{j \to \infty} \alpha_{i_j}(Q(x^{i_j}) - \bar{Q}) = 0 \tag{8}$$

where $\{x^{i_j}\}$ is a subsequence converging to $\bar{x} \in X_0$.

Proof Let $\{x^{i_j}\}$ be a subsequence converging to $\bar{x} \in X_0$. It follows from Theorem 13 and the lower semicontinuity of Q that

$$\begin{aligned} 0 &= \lim_{j\to\infty} (Q(x^{i_j}) - \bar{Q}) \\ &\geq Q(\bar{x}) - \bar{Q} \\ &\geq 0 \end{aligned}$$

Hence $Q(\bar{x}) = \bar{Q}$ and $\bar{x} \in X_Q$. We invoke Theorem 13 again and the lower semicontinuity of ϕ to give

$$\phi(\bar{x}) \leq \lim_{j\to\infty} \phi(x^{i_j}) \leq \inf_{x\in X_Q} \phi(x)$$

Since $\bar{x} \in X_Q$, it follows that $\bar{x}$ solves (7). To establish (8) we note that

$$\begin{aligned} 0 &\geq P(x^{i_j}, \alpha_{i_j}) - P(\bar{x}, \alpha_{i_j}) \\ &= \phi(x^{i_j}) - \phi(\bar{x}) + \alpha_{i_j}(Q(x^{i_j}) - \bar{Q}) \end{aligned}$$

Hence

$$\begin{aligned} \phi(\bar{x}) - \phi(x^{i_j}) &\geq \alpha_{i_j}(Q(x^{i_j}) - \bar{Q}) \\ &\geq 0 \end{aligned}$$

If we let $j \to \infty$ and recall that ϕ is lower semicontinuous at $\bar{x}$ we conclude that

$$\lim_{j\to\infty} \alpha_{i_j}(Q(x^{i_j}) - \bar{Q}) = 0$$

as required. ∎

4 Proximal point and weak sharp minima

The notion of a proximal point was introduced by Moreau[8] and has been extensively analysed by various researchers. In the sequel we give a brief description of the algorithm. We proceed to extend some of the results found in the literature on the finite termination of the algorithm to the case where the problem being considered has a weak sharp minimum. It is assumed throughout that S is a non–empty closed convex set and that ϕ is a proper closed convex function. It is easy to see from the definition of the subdifferential $\partial\phi_S$ that $\bar{x}$ is an optimal solution of the general convex programming problem (1) if and only if $0 \in \partial\phi_S(\bar{x})$. Therefore, the minimization problem (1) can be solved by finding a solution of a generalized equation $0 \in \partial\phi_S(x)$. It was shown by Brézis [3] that provided ϕ_S is a proper closed convex function, then the subdifferential is a maximal monotone operator and the resolvent, J_λ, defined by

$$J_\lambda = (I + \lambda\partial\phi_S)^{-1}$$

is a contraction and single–valued. Furthermore

$$0 \in \partial\phi_S(x) \iff x = J_\lambda x \text{ for some } \lambda > 0$$

The minimization problem (1) has thus been transformed into the problem of finding a fixed point of the resolvent J_λ of the subdifferential $\partial\phi_S$. For any given starting point x^0, the proximal point method generates the following sequence of iterates $\{x^i\}$, obtained by the relation

$$x^{i+1} = J_{\lambda_i} x^i \tag{9}$$

where $\{\lambda_i\}$ is a sequence of positive numbers with $\lambda_i \geq \lambda > 0$, for all i. In fact, x^{i+1} is the optimal solution of

$$\underset{x\in S}{\text{minimize}} \quad \phi(x) + \tfrac{1}{2\lambda_i}\left\|x - x^i\right\|^2$$

which is equivalent to

$$\underset{x\in \mathbb{R}^n}{\text{minimize}} \quad \phi_S(x) + \tfrac{1}{2\lambda_i}\left\|x - x^i\right\|^2 \tag{10}$$

since the minimizer of (10) must satisfy

$$0 \in \partial\phi_S(x^{i+1}) + \frac{1}{\lambda_i}(x^{i+1} - x^i)$$

and hence x^{i+1} satisfies (9). Furthermore, x^{i+1} is unique by the strong convexity of the objective of (10), and a little algebra shows that

$$x^{i+1} + \lambda_i v^{i+1} = x^i \tag{11}$$

where $v^{i+1} \in \partial \phi_S(x^{i+1})$. The following property follows easily from the above definitions.

Lemma 17 (a) $\|v^i\|$ *is non–increasing for* $i = 1, 2, \ldots$

(b) *If* $\bar{S} \neq \emptyset$ *then for any* $z \in \bar{S}$

$$\left\|x^{i+1} - z\right\|^2 + \lambda_i^2 \left\|v^{i+1}\right\|^2 \leq \left\|x^i - z\right\|^2$$

Proof (a): By monotonicity of the subdifferential, we see

$$0 \leq \left\langle v^{i+1} - v^i, x^{i+1} - x^i \right\rangle$$

so that using (11)

$$0 \leq \left\langle v^{i+1} - v^i, -\lambda_i v^{i+1} \right\rangle$$

Hence

$$\left\|v^{i+1}\right\|^2 \leq \left\|v^{i+1}\right\| \left\|v^i\right\|$$

so that

$$\left\|v^{i+1}\right\| \leq \left\|v^i\right\|$$

(b): Let $z \in \bar{S}$. Then by (11)

$$\left\|x^{i+1} + \lambda_i v^{i+1} - z\right\|^2 = \left\|x^i - z\right\|^2$$

Thus

$$\left\|x^{i+1} - z\right\|^2 + 2\lambda_i \left\langle v^{i+1} - 0, x^{i+1} - z \right\rangle + \lambda_i^2 \left\|v^{i+1}\right\|^2 = \left\|x^i - z\right\|^2$$

but by monotonicity, and $0 \in \partial\phi(z)$

$$\left\langle v^{i+1} - 0, x^{i+1} - z \right\rangle \geq 0$$

so we get

$$\left\| x^{i+1} - z \right\|^2 + \lambda_i^2 \left\| v^{i+1} \right\|^2 \leq \left\| x^i - z \right\|^2$$

as required. ∎

We are now able to prove the following lemma which is important for the proof of the ensuing theorem.

Lemma 18 *For any $\lambda > 0$, let $y = \lambda(z - P_{\bar{S}}(z))$ and suppose that $y \in \partial\phi_S(w)$ for some $w \in \bar{S}$. Then $y \in \partial\phi_S(P_{\bar{S}}(z))$.*

Proof We note that

$$y \in \partial\phi_S(w) \implies \phi_S(x) - \phi_S(w) \geq \langle y, x - w \rangle \quad \forall x \in \mathbb{R}^n$$

and therefore

$$w, P_{\bar{S}}(z) \in \bar{S} \implies \phi_S(x) - \phi_S(P_{\bar{S}}(z)) \geq \langle y, x - w \rangle \quad \forall x \in \mathbb{R}^n$$

Now it follows from Lemma 5 that $\langle z - P_{\bar{S}}(z), w - P_{\bar{S}}(z) \rangle \leq 0$ so that multiplication by $\lambda > 0$ gives $\langle y, w - P_{\bar{S}}(z) \rangle \leq 0$. Hence for all $x \in \mathbb{R}^n$

$$\begin{aligned} \phi_S(x) - \phi_S(P_{\bar{S}}(z)) &\geq \langle y, x - w \rangle + \langle y, w - P_{\bar{S}}(z) \rangle \\ &= \langle y, x - P_{\bar{S}}(z) \rangle \end{aligned}$$

that is, $y \in \partial\phi_S(P_{\bar{S}}(z))$. ∎

Theorem 19 *Suppose $\exists \epsilon > 0$ such that*

$$\epsilon(B \bigcap K_{\bar{S}}) \subseteq \partial\phi_S(\bar{S})$$

If $\|w\| \leq \epsilon$ and $w \in \partial\phi_S(z)$ then $z \in \bar{S}$.

Proof Let $\epsilon > 0$ be chosen so that $\epsilon(B \bigcap K_{\bar{S}}) \subseteq \partial\phi_S(\bar{S})$. We choose w with $\|w\| < \epsilon$ and $w \in \partial\phi_S(z)$.

Let us assume that $z \neq P_{\bar{S}}(z)$. We proceed to obtain a contradiction. Define

$$y = \frac{\epsilon(z - P_{\bar{S}}(z))}{\|z - P_{\bar{S}}(z)\|}$$

so that $y \in \epsilon(B \bigcap K_{\bar{S}})$. It then follows that $\exists z^y \in \bar{S}$ with the property that $y \in \partial\phi_S(z^y)$. The monotonicity of $\partial\phi_S$ gives that

$$0 \leq \langle z - z^y, w - y \rangle$$

from whence it follows that

$$\begin{aligned}\frac{\epsilon}{\|z - P_{\bar{S}}(z)\|}\langle z - z^y, z - P_{\bar{S}}(z)\rangle &= \langle z - z^y, y \rangle \\ &\leq \langle z - z^y, w \rangle \\ &\leq \|z - z^y\| \, \|w\|\end{aligned}$$

Hence

$$\|w\| \geq \epsilon \frac{\langle z - z^y, z - P_{\bar{S}}(z)\rangle}{\|z - P_{\bar{S}}(z)\| \, \|z - z^y\|}$$

The result now follows using Lemma 18 which enables us to take $z^y = P_{\bar{S}}(z)$ and hence derive the contradiction $\|w\| \geq \epsilon$. The proof is now complete since this gives $z = P_{\bar{S}}(z)$. ∎

The following result is the central one of this section.

Corollary 20 *Suppose* $\exists \epsilon > 0$ *such that*

$$\epsilon(B \bigcap K_{\bar{S}}) \subseteq \partial\phi_S(\bar{S})$$

Let $\{\lambda_i\}$ *be any positive sequence which is bounded below and let* $x^0 \in \mathrm{I\!R}^n$. *Then the proximal point algorithm terminates in a finite number of iterations.*

Proof Let $\lambda_i \geq \lambda > 0$ for the given sequence. Then for any $z \in \bar{S}$ we have from Lemma 17 that the sequence $\{\|x^i - z\|\}$ is bounded and hence converges. If we invoke the lemma again for $i = 0, \ldots, N$ and sum, the following inequality results

$$\left\|x^{N+1} - z\right\|^2 + \sum_{i=0}^{N} \lambda_i^2 \left\|v^{i+1}\right\|^2 \leq \left\|x^0 - z\right\|^2$$

Using the above observations, it is clear that

$$\sum_{i=0}^{N} \lambda_i^2 \left\|v^{i+1}\right\|^2 \leq M$$

so that

$$\lambda^2 \left\|v^{N+1}\right\|^2 (N+1) \leq M$$

by non–increasing property of $\|v^i\|$ given in Lemma 17. Hence, there exists a sufficiently large but finite N such that

$$\left\|v^{N+1}\right\|^2 \leq \frac{M}{\lambda^2(N+1)} < \epsilon^2$$

where ϵ is given in Theorem 19. It then follows from Theorem 19 that x^{N+1} is in the solution set. ∎

The following relationship between the solution of the proximal point method and the previous iterate is an aid to understanding the algorithm. It states that, in fact, the proximal point algorithm terminates with the closest point in the solution set to the last non–optimal iterate. This result should add some clarification to the naming of the proximal point algorithm, since it attempts to find the minimizer of ϕ on S which is proximal to x^i.

Theorem 21 *Suppose that the proximal point algorithm terminates in a finite number of iterations, k say. Then $x^k = P_{\bar{S}}(x^{k-1})$.*

Proof Note from the definition of the proximal point algorithm that, for all $x \in \mathbb{R}^n$

$$\phi_S(x) - \phi_S(x^k) \geq \left\langle v^k, x - x^k \right\rangle$$

Substituting $P_{\bar{S}}(x^{k-1})$ for x, we get

$$0 \geq \left\langle v^k, P_{\bar{S}}(x^{k-1}) - x^k \right\rangle$$

and since $\lambda_{k-1} > 0$ we see that

$$0 \geq \left\langle x^{k-1} - x^k, P_{\bar{S}}(x^{k-1}) - x^k \right\rangle$$

Lemma 5 gives

$$0 \geq \left\langle x^{k-1} - P_{\bar{S}}(x^{k-1}), x^k - P_{\bar{S}}(x^{k-1}) \right\rangle$$

so by adding the above inequalities we get

$$\begin{aligned} 0 &\geq \left\langle x^k - x^{k-1}, x^k - P_{\bar{S}}(x^{k-1}) \right\rangle + \left\langle x^{k-1} - P_{\bar{S}}(x^{k-1}), x^k - P_{\bar{S}}(x^{k-1}) \right\rangle \\ &= \left\| x^k - P_{\bar{S}}(x^{k-1}) \right\|^2 \end{aligned}$$

and hence $x^k = P_{\bar{S}}(x^{k-1})$. ∎

The next two theorems provide a strong link between the subgradient condition associated with a weak sharp minimum and the proximal point algorithm.

Theorem 22 *Suppose* $\exists \epsilon > 0$ *such that*

$$\epsilon(B \bigcap K_{\bar{S}}) \subseteq \partial \phi_S(\bar{S})$$

Then for any given x^0*, the proximal point algorithm terminates in one iteration for a sufficiently large choice of* λ.

Proof Let $\epsilon > 0$ be chosen so that $\epsilon(B \bigcap K_{\bar{S}}) \subseteq \partial\phi_S(\bar{S})$. We assume that $x^1 \neq P_{\bar{S}}(x^0)$. Define

$$y = \frac{\epsilon(x^1 - P_{\bar{S}}(x^0))}{\|x^1 - P_{\bar{S}}(x^0)\|}$$

so that $y \in \epsilon(B \bigcap K_{\bar{S}})$, and hence that for some $z^y \in \bar{S}$, $y \in \partial\phi_S(z^y)$. The monotonicity of $\partial\phi_S$ gives

$$0 \leq \left\langle x^1 - z^y, v^1 - y \right\rangle$$

from whence it follows that

$$\begin{aligned} \left\langle x^1 - z^y, y \right\rangle &\leq \left\langle x^1 - z^y, v^1 \right\rangle \\ &\leq \left\| x^1 - z^y \right\| \left\| v^1 \right\| \end{aligned}$$

Hence

$$\left\| v^1 \right\| \geq \epsilon \frac{\langle x^1 - z^y, x^1 - P_{\bar{S}}(x^0) \rangle}{\|x^1 - P_{\bar{S}}(x^0)\| \, \|x^1 - z^y\|} = \epsilon \tag{12}$$

the last equality by taking $z^y = P_{\bar{S}}(z)$ which is possible from Lemma 18. Now choose $\lambda > d(x^0, \bar{S})/\epsilon$. By Lemma 17 we see that

$$\left\| x^1 - P_{\bar{S}}(x^0) \right\|^2 + \lambda^2 \left\| v^1 \right\|^2 \leq \left\| x^0 - P_{\bar{S}}(x^0) \right\|^2$$

so that

$$\begin{aligned} \left\| x^1 - P_{\bar{S}}(x^0) \right\|^2 &\leq [d(x^0, \bar{S})]^2 - \lambda^2 \left\| v^1 \right\|^2 \\ &< \lambda^2 \epsilon^2 - \lambda^2 \left\| v^1 \right\|^2 \\ &\leq 0 \end{aligned}$$

The last inequality follows from (12). But this is a contradiction and so $x^1 = P_{\bar{S}}(x^0)$. ∎

Corollary 20 and Theorem 22 are generalizations of the corresponding results first obtained for linear programs by Polyak and Tretiyakov[12].

Theorem 23 *Suppose that the proximal point algorithm terminates in one step for any $x^0 \in \mathbb{R}^n$ and that $\partial\phi_S(\bar{S})$ is compact. Then $\exists\epsilon > 0$ such that*

$$\epsilon(B \bigcap K_{\bar{S}}) \subseteq \partial\phi_S(\bar{S})$$

Proof Let

$$y = \frac{x^0 - P_{\bar{S}}(x^0)}{\|x^0 - P_{\bar{S}}(x^0)\|}$$

for some x^0. We show $\exists\bar{\lambda}$ with $\frac{y}{\lambda} \in \partial\phi_S(\bar{S})$ for all $\lambda \geq \bar{\lambda}$. The hypothesis of the theorem guarantees the existence of $\lambda(x^0)$ with

$$P_{\bar{S}}(x^0) = x^1 = x^0 - \lambda(x^0)v^1(x^0)$$

However, by properties of the proximal point algorithm,

$$v^1(x^0) \in \partial\phi_S(x^1) = \partial\phi_S(P_{\bar{S}}(x^0)) \subseteq \partial\phi_S(\bar{S})$$

Since $\partial\phi_S(\bar{S})$ is compact, it follows that, for every x^0, generated as above, $\|v^1(x^0)\| \leq M$ which gives

$$\begin{aligned} \left|\lambda(x^0)\right| &= \frac{1}{\|v^1(x^0)\|} \\ &\geq \frac{1}{M} =: \bar{\lambda} \end{aligned}$$

Hence, for all $\lambda \geq \bar{\lambda}$

$$\frac{y}{\lambda} = v^1(x^0) \in \partial\phi_S(\bar{S})$$

The result now follows easily. ∎

Corollary 24 *If proximal point terminates in one iteration for any $x^0 \in \mathbb{R}^n$ and $\bar{S}$ is compact then* $0 \in \operatorname{int}\partial\phi_S(\bar{S})$.

Proof $\partial\phi_S(\bar{S})$ is compact by [13, Theorem 24.7] and $K_{\bar{S}} = \mathbb{R}^n$ by Lemma 7. ∎

5 Perturbations and weak sharp minima

In this section, our interest lies primarily with the perturbed problem $P(\epsilon)$ defined as follows:

$$\underset{x \in S}{\text{minimize}} \quad \phi(x) + \epsilon f(x) \tag{13}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and ϵ is a nonnegative real number. For convenience we define the optimal solution set of (13) as $\bar{S}(\epsilon)$. The following result is elementary and easy to prove.

Lemma 25

$$\left.\begin{array}{c} \epsilon > 0 \\ \bar{x} \in \bar{S}(\epsilon) \subseteq \bar{S} \end{array}\right\} \implies \bar{x} \in \arg\min_{x \in \bar{S}} f(x)$$

Proof Suppose to the contrary that $\bar{x} \notin \arg\min_{x \in \bar{S}} f(x)$. Then there is some $\bar{y} \in \bar{S}$ such that $f(\bar{y}) < f(\bar{x})$. However, for every $x \in S$,

$$\phi(\bar{x}) + \epsilon f(\bar{x}) \leq \phi(x) + \epsilon f(x).$$

In particular, since $\bar{S} \subseteq S$, $\phi(\bar{x}) + \epsilon f(\bar{x}) \leq \phi(\bar{y}) + \epsilon f(\bar{y})$, so that $\bar{x}, \bar{y} \in \bar{S}$ implies $\epsilon f(\bar{x}) \leq \epsilon f(\bar{y})$. This is a contradiction since $\epsilon > 0$. ∎

The converse result needs more work and relies upon the notion of a weak sharp minimum. We shall need another definition, which will prove to be a central property throughout the thesis.

Definition 26 *For the problem*

$$\underset{x \in S}{\text{minimize}} \quad \phi(x)$$

and for a perturbation function f, the associated perturbed problem (13) is said to have the **finite perturbation property** *if there exists a positive $\bar{\epsilon}$ such that for all $\epsilon \in (0, \bar{\epsilon}]$*

$$\bar{S}_f = \bar{S}(\epsilon) \subseteq \bar{S}$$

where $\bar{S}_f := \arg\min_{x \in \bar{S}} f(x)$.

The proof of the following theorem is adapted from Mangasarian and Meyer[7].

Theorem 27 *Suppose that ϕ is a convex function on S, and that ϕ has a weak sharp minimum on S. Let $\phi(x) + \epsilon^* f(x)$ be bounded below on S for some $\epsilon^* > 0$. Then the finite perturbation property holds provided that any of the following is true:*

(a) A local Lipschitz property is satisfied by f, namely there exist positive numbers k and K such that

$$f(P_{\bar{S}}(x)) - f(x) \le K \|x - P_{\bar{S}}(x)\| \text{ for } x \in S \text{ and } \|x - P_{\bar{S}}(x)\| \le k \tag{14}$$

(b) f is convex on some open convex set containing S, and $\bar{S}$ is compact

(c) f is strictly convex on some open convex set containing S

Proof Let $\bar{x} \in \arg\min_{x \in \bar{S}} f(x)$. We will first establish that $\bar{x} \in \bar{S}(\epsilon) \subseteq \bar{S}$ for sufficiently small $\epsilon \ge 0$ under hypothesis (a) by showing that for sufficiently small $\epsilon \ge 0$

$$\phi(\bar{x}) + \epsilon f(\bar{x}) \quad < \quad \phi(x) + \epsilon f(x) \quad \text{for } x \in S \setminus \bar{S} \tag{15}$$

and

$$\phi(\bar{x}) + \epsilon f(\bar{x}) \quad \le \quad \phi(x) + \epsilon f(x) \quad \text{for } x \in \bar{S} \tag{16}$$

Inequality (16) holds because $\bar{x}$ minimizes f on $\bar{S}$. To establish (15), let $x \in S \setminus \bar{S}$; thus $x \neq P_{\bar{S}}(x)$, and we consider the two following cases.

Case 1: $0 < \|P_{\bar{S}}(x) - x\| \le k$. The strict inequality (15) follows for $\epsilon \in [0, \alpha/K)$ since

$$\begin{aligned}
\epsilon(f(\bar{x}) - f(x)) &\le \epsilon(f(P_{\bar{S}}(x)) - f(x)) && \text{(definitions of } \bar{x} \text{ and } P_{\bar{S}}(x)) \\
&\le \epsilon K \|P_{\bar{S}}(x) - x\| && \text{(by (14) and } \|x - P_{\bar{S}}(x)\| \le k) \\
&< \alpha \|P_{\bar{S}}(x) - x\| && (\epsilon \in [0, \alpha/K) \text{ and } x \neq P_{\bar{S}}(x)) \\
&\le \phi(x) - \phi(P_{\bar{S}}(x)) && \text{(by weak sharp minimum)} \\
&= \phi(x) - \phi(\bar{x}) && \text{(definition of } P_{\bar{S}}(x))
\end{aligned}$$

Case 2: $\|P_{\bar{S}}(x) - x\| > k$. Let ν be such that

$$\phi(x) + \epsilon^* f(x) \geq \nu + 1 > \nu$$

for $x \in S$, so that $f(x) > \nu/\epsilon^* - \phi(x)/\epsilon^*$. By defining

$$\rho \;:=\; -\nu/\epsilon^* + f(\bar{x}) + \phi(\bar{x})/\epsilon^* \;>\; 0$$

and noting that $\phi(\bar{x}) = \phi(P_{\bar{S}}(x))$ for all $x \in S$, we have that

$$f(\bar{x}) - f(x) \;<\; [\phi(x) - \phi(P_{\bar{S}}(x))]/\epsilon^* + \rho \quad \text{for } x \in S$$

Since $\|P_{\bar{S}}(x) - x\| > k$, it follows that

$$f(\bar{x}) - f(x) \;<\; [\phi(x) - \phi(P_{\bar{S}}(x))]/\epsilon^* + (\rho/k)\,\|x - P_{\bar{S}}(x)\|$$

and so, for $x \in S$

$$\gamma[f(\bar{x}) - f(x)] \;<\; \frac{\gamma}{\epsilon^*}[\phi(x) - \phi(P_{\bar{S}}(x))] + \frac{\gamma\rho}{k}\,\|x - P_{\bar{S}}(x)\| \tag{17}$$

Consequently for γ small enough, that is, $\gamma \in [0, \alpha k/\rho]$, the last term of (17) is less than or equal to $\alpha\,\|x - P_{\bar{S}}(x)\|$. Therefore, for such γ

$$\begin{aligned}\gamma[f(\bar{x}) - f(x)] \;&<\; \frac{\gamma}{\epsilon^*}[\phi(x) - \phi(P_{\bar{S}}(x))] + \alpha\,\|x - P_{\bar{S}}(x)\| \\ &\leq\; \left[1 + \frac{\gamma}{\epsilon^*}\right][\phi(x) - \phi(\bar{x})]\end{aligned}$$

the last inequality coming from the weak sharp minimum assumption on ϕ. If we let

$$\epsilon \;=\; \frac{\gamma}{(1 + \gamma/\epsilon^*)}$$

then, for sufficiently small ϵ, (15) follows in this second case as well.

Now note that by choice of k, hypothesis (b) implies (a) (see Theorem 10.4 of [13]), so that the proof will be completed by showing that the result holds under hypothesis (c) even when $\bar{S}$ is not compact. Let

$$T = \{x \mid \|x - \bar{x}\| \leq k\}$$

where k is some strictly positive number, let $S' = S \cap T$, and let $\bar{S}' = \bar{S} \cap T$. Note that S' is a compact convex set and that $\bar{S}'$ is the set of optimal solutions of $\min_{x \in S'} \phi(x)$, so that the preceding arguments imply that there exists an $\epsilon' > 0$ such that $\bar{x} \in \bar{S}'(\epsilon) \subseteq \bar{S}'$ for $\epsilon \in [0, \epsilon']$ where $\bar{S}'(\epsilon)$ denotes the solution set of $\min_{x \in S \cap T} \phi(x) + \epsilon f(x)$. By the convexity of $\phi(x) + \epsilon f(x)$, $\bar{x} \in \bar{S}'(\epsilon)$ implies that $\bar{x} \in \bar{S}(\epsilon)$ (since a local solution of $P(\epsilon)$ must also be a global solution), and by strict convexity it follows that $\bar{S}(\epsilon) = \{\bar{x}\}$. Thus $\bar{S}(\epsilon) \subseteq \bar{S}$ for $\epsilon \in [0, \epsilon']$, and since $\bar{x} \in \bar{S}(\epsilon)$, the theorem is established under hypothesis (c). ∎

The following theorem and corollary connects the perturbation results obtained in this section to the results obtained for the proximal point algorithm in the previous section.

Theorem 28 *Let x^0 be given and let*

$$f(x) = \frac{1}{2} \left\| x - x^0 \right\|^2$$

Then the proximal point algorithm terminates in one step for sufficiently large λ provided that the finite perturbation property (Definition 26) holds.

Proof f is strictly convex implies that $\arg\min_{x \in \bar{S}} f(x) = \{P_{\bar{S}}(x^0)\}$. Hence $\exists \bar{\epsilon} > 0$ such that for all $\epsilon \in (0, \bar{\epsilon}]$

$$P_{\bar{S}}(x^0) \in \bar{S}(\epsilon) = \arg\min_{x \in S} \phi(x) + \frac{\epsilon}{2} \left\| x - x^0 \right\|^2$$

Thus for all $\lambda > 1/\bar{\epsilon}$

$$P_{\bar{S}}(x^0) \in \arg\min_{x \in S} \phi(x) + \frac{1}{2\lambda} \left\| x - x^0 \right\|^2 = \{x^1\}$$

Hence, $x^1 = P_{\bar{S}}(x^0)$, as was required. ∎

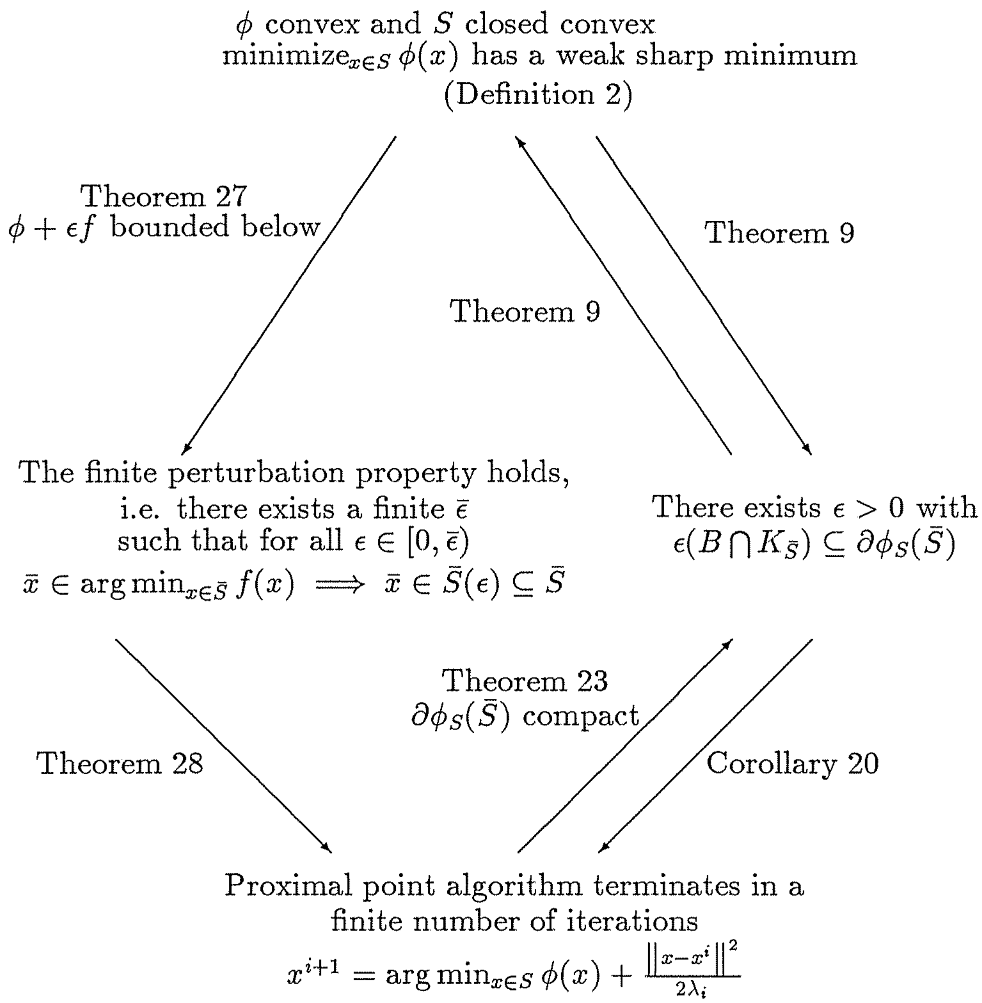

Figure 1: Connection between various results

Corollary 29 *Under the hypotheses of Theorem 27, the proximal point algorithm terminates in one step for sufficiently large λ.*

Proof Use Theorem 27 and Theorem 28 ∎

Figure 1 summarizes briefly the results given in the previous sections.

6 Sharp penalty functions

We now assume that

$$S = \{x \in \mathbb{R}^n \mid \mathrm{g(x)} \leq 0\}$$

is a closed, convex (by convexity of g) but possibly empty set. The property of a weak sharp minima is closely related to the exactness of penalty functions in constrained optimization. We shall briefly discuss the connection between these concepts in this section.

In order to motivate our discussion, we consider some related notions which are found in the literature.

Definition 30 *A set of constraints, $g(x) \leq 0$, is called* **sharp** *if there is a positive number α such that for each x there is a p with $g(p) \leq 0$ and $\|x - p\| \leq \alpha \|g(x)_+\|$.*

Note that the equivalence of norms on $\mathbb{R}^n$ ensures that the property of being sharp does not depend on the norm used. Brady[2] gives conditions under which a set of constraints is sharp. The name sharp constraints is justified by the following lemma.

Lemma 31 *Suppose g is a closed, convex function. The set of constraints $g(x) \leq 0$ is sharp if and only if the problem*

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \|g(x)_+\| \tag{18}$$

has a weak sharp minimum.

Proof The optimal solution set of (18), S, is closed and convex. If problem (18) has a weak sharp minimum, then take $p = P_S(x)$, which exists by the definition, and the result follows easily. Conversely, suppose that the constraints $g(x) \leq 0$ are sharp. Then S is nonempty and it follows that the projection operation onto this set is well–defined. Thus, since p is in the optimal solution set, it is immediate that

$$\begin{aligned} \|x - P_S(x)\| &\leq \|x - p\| \\ &\leq \alpha \|g(x)_+\| \\ &\leq \alpha \left[\|g(x)_+\| - \|g(P_S(x))_+\|\right] \end{aligned}$$

and so (18) has a weak sharp minimum. ∎

Although this observation identifies a large class of problems with weak sharp minima (in fact any set of constraints which has a condition constant leads to a problem with a weak sharp minimum) the definition of sharp constraints is too restrictive for our purposes since it implies the problems being considered are feasible. This leads to the following more general definition.

Definition 32 *Suppose we have a closed, convex set, S, which may be empty. We define a* **penalty function** *for S to be any function, Q, satisfying*

(a) $Q(x) = 0$ *for all* $x \in S$

(b) $Q(x) > 0$ *for* $x \notin S$

We say that Q is a **sharp penalty function** *for S if*

$$\underset{x \in X_0}{\text{minimize}} \quad Q(x), \tag{19}$$

has a weak sharp minimum.

We note that X_0 is some subset of $\mathbb{R}^n$. This is a weaker notion than sharp constraints as the following lemma shows.

Lemma 33 *Suppose g is a closed, convex function and the set of constraints $g(x) \leq 0$ is sharp. Then $Q(x) = \|g(x)_+\|$ is a sharp penalty function.*

Proof Clearly $Q(x)$ as defined above is a penalty function. Since $g(x) \leq 0$ are sharp constraints, the result follows from Lemma 31. ∎

However, the sharp penalty function is precisely the notion which is needed in order to make the inequality (6) in Theorem 13 an equality. The penalty problem under consideration is

$$\underset{x \in X_0}{\text{minimize}} \quad P(x, \alpha) := \phi(x) + \alpha Q(x)$$

Theorem 34 *Let $\{\alpha_i\}$ be an increasing unbounded sequence of positive numbers and let $\{x^i\}$ be a corresponding sequence of points in $X_0 \setminus S$ such that*

$$-\infty < P(x^i, \alpha_i) = \min_{x \in X_0} P(x, \alpha_i)$$

If we define $\bar{Q} := \lim_{i \to \infty} Q(x^i)$, then

$$\bar{Q} = \inf_{x \in X_0} Q(x)$$

Further, if we let

$$\bar{S}_Q := \{x \in X_0 \mid Q(x) \leq \bar{Q}\}$$

and assume $\bar{S}_Q \neq \emptyset$ and ϕ is Lipschitz continuous on X_0, that is, $\exists K > 0$

$$|\phi(y) - \phi(x)| \leq K \|y - x\|$$

for all $x, y \in X_0$ and the penalty function Q is sharp, then

$$\lim_{i \to \infty} \phi(x^i) = \inf_{x \in \bar{S}_Q} \phi(x)$$

Proof The first part of the Theorem is a restatement of Theorem 13. It therefore only remains to prove equality in the last statement of the theorem. By the definition of weak sharp minimum, we have

$$\mu \left\| x^i - P_{\bar{S}_Q}(x^i) \right\| \leq Q(x^i) - Q(P_{\bar{S}_Q}(x^i))$$

Hence, by the Lipschitz continuity of ϕ we see that

$$\begin{aligned} \phi(P_{\bar{S}_Q}(x^i)) - \phi(x^i) &\leq \left| \phi(x^i) - \phi(P_{\bar{S}_Q}(x^i)) \right| \\ &\leq K \left\| x^i - P_{\bar{S}_Q}(x^i) \right\| \\ &\leq \frac{K}{\mu}(Q(x^i) - \bar{Q}) \end{aligned}$$

This inequality, coupled with the fact that $P_{\bar{S}_Q}(x^i) \in \bar{S}_Q$ leads to the inequality

$$\phi(x^i) + \frac{K}{\mu}(Q(x^i) - \bar{Q}) \geq \phi(P_{\bar{S}_Q}(x^i)) \geq \inf_{x \in \bar{S}_Q} \phi(x)$$

By Theorem 13, $\lim_{i \to \infty} Q(x^i) = \bar{Q}$, so taking the limit leads to the conclusion that

$$\lim_{i \to \infty} \phi(x^i) \geq \inf_{x \in \bar{S}_Q} \phi(x)$$

The equality now follows from Theorem 13. ∎

We now perturb the penalty function by the objective function ϕ. For ease of notation we will assume that $X_0 = \mathbb{R}^n$. Consider the problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad Q(x) + \epsilon\phi(x) \tag{20}$$

with assumed optimal solution set, $\bar{S}_Q(\epsilon)$. The following is an immediate corollary to Theorem 27.

Theorem 35 *Let Q be a sharp penalty function for S. Suppose that one of the following holds for the function ϕ:*

(a) the local Lipschitz property, namely there exists positive numbers k and K such that

$$\phi(P_{\bar{S}_Q}(x)) - \phi(x) \leq K \left\| x - P_{\bar{S}_Q}(x) \right\| \quad \textit{whenever} \quad \left\| x - P_{\bar{S}_Q}(x) \right\| \leq k$$

(b) ϕ is convex on $\mathbb{R}^n$ and $\bar{S}_Q$ is compact

(c) ϕ is strictly convex on $\mathbb{R}^n$

and that $Q + \epsilon^*\phi$ is bounded below on $\mathbb{R}^n$ for some $\epsilon^* > 0$. Then there exists an $\bar{\epsilon} > 0$ such that for all $\epsilon \in (0, \bar{\epsilon}]$

$$\bar{S}_Q(\phi) = \bar{S}_Q(\epsilon) \subseteq \bar{S}_Q$$

where $\bar{S}_Q(\phi) := \arg\min_{x \in \bar{S}_Q} \phi(x)$. (i.e. the finite perturbation property holds for this problem.)

The following theorem gives the result that we have been promising, that is, the sharp penalty function property is an exact penalty.

Theorem 36 *Suppose that one of the conditions (a), (b) or (c) of Theorem 35 holds for the function ϕ, and suppose that Q is a sharp penalty function for S. Let $\{\alpha_i\}$ be an increasing unbounded sequence of positive numbers and let $\{x^i\}$ be a corresponding sequence of points in $\mathbb{R}^n \setminus S$ such that*

$$-\infty < P(x^i, \alpha_i) = \min_{x \in \mathbb{R}^n} P(x, \alpha_i)$$

Then there exists $\bar{\alpha}$, such that for all $\alpha \geq \bar{\alpha}$, the corresponding x^i solves

$$\underset{x \in \bar{S}_Q}{\text{minimize}} \quad \phi(x)$$

Proof Since

$$P(x, \alpha_i) = \alpha_i(Q(x) + \frac{1}{\alpha_i}\phi(x))$$

is bounded below on $\mathbb{R}^n$, it follows that $Q + \epsilon^*\phi$ is bounded below on $\mathbb{R}^n$ for $\epsilon^* = 1/\alpha_i$. Thus invoking Theorem 35, we see that there exists an $\bar{\epsilon} > 0$ such that for all $\epsilon \in (0, \bar{\epsilon}]$

$$\bar{S}_Q(\phi) = \bar{S}_Q(\epsilon) \subseteq \bar{S}_Q$$

where $\bar{S}_Q(\phi) := \arg\min_{x \in \bar{S}_Q} \phi(x)$. Thus

$$\begin{aligned} & x^i \in \arg\min_{x \in \mathbb{R}^n} P(x, 1/\epsilon_i) \\ \iff \quad & x^i \in \arg\min_{x \in \mathbb{R}^n} [\phi(x) + \frac{1}{\epsilon_i} Q(x)] \\ \iff \quad & x^i \in \bar{S}_Q(\phi) = \bar{S}_Q(\epsilon_i) = \arg\min_{x \in \mathbb{R}^n} [Q(x) + \epsilon_i \phi(x)] \end{aligned}$$

Hence α_i must only be larger than $1/\bar{\epsilon}$ in order to give $x^i \in \bar{S}_Q(\phi)$ as was required. ∎

The following corollary is an immediate consequence of Lemma 14.

Corollary 37 *If $S \neq \emptyset$, then for all $\alpha \geq \bar{\alpha}$ the corresonding x^i is a solution to the original problem (1).*

Finally, we give an example which shows we have extended the idea of a sharp constraint to the case where the feasible region may be empty. Consider the following example

$$\begin{array}{ll} \text{minimize} & \phi(x) \\ \text{subject to} & g(x) = x_+ + 3 \leq 0 \end{array}$$

Let $Q(x) = \|g(x)_+\|$, so that Q is a penalty function for $g(x) \leq 0$. Problem (19) has a weak sharp minimum since $\bar{S}_Q = \mathbb{R}_-$ and thus for $x > 0$,

$$Q(x) - Q(P_{\mathbb{R}_-}(x)) = |x+3| - 3 = |x| = \|x - 0\| = \left\|x - P_{\mathbb{R}_-}(x)\right\|$$

We see that Q is a sharp penalty function (even though the constraints are not sharp). We consider the objective to be $\phi(x) := \|x\|^2$, so that $\arg\min_{x \in \bar{S}_Q} \phi(x) = \{0\}$. The theory then guarantees that $0 \in \bar{S}_Q(\epsilon)$ for sufficiently small ϵ, which is true in the example since for all $\epsilon > 0$, $\bar{S}_Q(\epsilon) = \{0\}$.

7 Related notions

In this section we show some further examples of problems with weak sharp minima, in addition to those we have previously mentioned, which have been mainly associated with condition constants.

The first example is that of linear programming. In the appendix to their paper[7], Mangasarian and Meyer prove that any linear programming problem has a weak sharp minimum. They also prove a theorem very similar to Theorem 27 for the linear programming case when $\phi(x)$ is taken as $p^T x$ for some $p \in \mathbb{R}^n$ and $S = \{x \mid Ax \geq b\}$ for some $A \in \mathbb{R}^{m \times n}$ with $b \in \mathbb{R}^m$. Using the linearity of ϕ they show that hypothesis (c) of Theorem 27 can be weakened to just convexity of f.

We have seen that the notions of weak sharp minimum, finite perturbation property and finite convergence of the proximal point algorithm are equivalent in the convex setting. If we remove the restriction of convexity, then, in certain cases, we can still recover the finite perturbation property which will prove to be useful in our computations. In particular, Mangasarian and Meyer[7] have proved the following result:

Proposition 38 (Mangasarian and Meyer[7])
Let $S = \{x \in \mathbb{R}^{\mathrm{n}} \mid \mathrm{g(x)} \leq 0, \mathrm{h(x)} = 0\}$ *and*

$$\underset{x \in S}{\text{minimize}} \quad \phi(x) \tag{21}$$

have a nonempty set $\tilde{S}$ *of local optimal solutions satisfying a constraint qualification. Let* $\tilde{x} \in \tilde{S}$ *and define* $\tilde{\phi} = \phi(\tilde{x})$ *and* N *as the open ball around* $\tilde{x}$ *satisfying* $\phi(x) \geq \phi(\tilde{x})$ *for all* $x \in N \bigcap S$*. Let* ϕ*,* g*,* h *and* f *be differentiable on* $\tilde{S}$ *and let the nonlinear program*

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & \begin{array}{c} x \in S \bigcap N \\ \phi(x) \leq \tilde{\phi} \end{array} \end{array} \tag{22}$$

have a Karush–Kuhn–Tucker point $(\bar{x}, \bar{v}, \bar{s}, \bar{\gamma}) \in \mathbb{R}^{\mathrm{n+m+k+1}}$*. Then there exists an* $\bar{\epsilon} > 0$ *such that for each* ϵ *in* $(0, \bar{\epsilon}]$ *there exists a* $(\bar{u}, \bar{r})\colon [0, \bar{\epsilon}] \to \mathbb{R}^{\mathrm{m+k}}$ *such that* $(\bar{x}, \bar{u}(\epsilon), \bar{r}(\epsilon))$ *is a Karush–Kuhn–Tucker point for the perturbed problem*

$$\underset{x \in S}{\text{minimize}} \quad \phi(x) + \epsilon f(x)$$

and $\bar{x}$ *is also a local solution of the nonlinear program (21).*

Proposition 39 (Mangasarian[5])
Suppose h *is linear and* ϕ *and* g *are concave at* $\bar{x}$*. Then (22) has a Karush–Kuhn–Tucker point.*

Using the above results, the following corollary is immediate.

Corollary 40 *Suppose h is linear and ϕ and g are concave. Then (21) has the finite perturbation property around every local solution, i.e. for each local solution of (21), there exists a set N and a positive $\bar{\epsilon}$ such that for all $\epsilon \in (0, \bar{\epsilon}]$*

$$\bar{S}_f = \bar{S}(\epsilon) \bigcap N \subseteq \bar{S} \bigcap N$$

where $\bar{S}_f := \arg\min_{x \in \bar{S} \bigcap N} f(x)$.

8 Conclusions

The notion of a weak sharp minimum has been introduced and used to unify a number of important concepts of mathematical programming. In particular, we have shown it to be the fundamental property needed in the finite termination of the proximal point algorithm, exactness of penalty functions and for the finite perturbation property to hold.

Acknowledgement

I should like to thank Professor O. Mangasarian and Ms. L. Brady for many useful discussions during the preparation of this paper.

References

[1] D.A. Ault, F.R. Deutsch, P.D. Morris, and J.E. Olson. Interpolating subspaces in approximation theory. In A. Talbot, editor, *Approximation Theory*, Academic Press, London, 1970.

[2] L. Brady. *Condition Constants for Solutions of Convex Inequalities.* PhD thesis, University of Wisconsin, Madison, Wisconsin, 1988.

[3] H. Brézis. *Opérateurs Maximaux Monotones.* North–Holland, 1973.

[4] E.W. Cheney. *Introduction to Approximation Theory.* McGraw–Hill, New York, 1966.

[5] O.L. Mangasarian. *Nonlinear Programming.* McGraw–Hill, New York, 1969.

[6] O.L. Mangasarian. Some applications of penalty functions in mathematical programming. In R. Conti, E. De Giorgi, and F. Giannessi, editors, *Optimization and Related Fields*, pages 307–329, Springer–Verlag, Heidelberg, 1986. Lecture Notes in Mathematics 1190.

[7] O.L. Mangasarian and R.R. Meyer. Nonlinear perturbation of linear programs. *SIAM Journal on Control and Optimization*, 17(6):745–752, November 1979.

[8] J.–J. Moreau. Proximité et dualité dans un espace Hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

[9] D.J. Newman and H.S. Shapiro. Some theorems on Čebyšev approximation. *Duke Mathematical Journal*, 30:673–682, 1963.

[10] B.T. Polyak. Sharp minima. December 1979. A Talk given at the IIASA Workshop on Generalized Lagrangians and their Applications, IIASA, Laxenburg, Austria.

[11] B.T. Polyak. *Introduction to Optimization.* Optimization Software, Inc., Publications Division, New York, 1987.

[12] B.T. Polyak and N.V. Tretiyakov. Concerning an iterative method for linear programming and its economic interpretation. *Economics and Mathematical Methods*, 8(5):740–751, 1972. (Russian).

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