Mathematical Programming 50 (1991) 359–366
North-Holland

Finite termination of the proximal point algorithm

Michael C. Ferris*
Computer Sciences Department, University of Wisconsin, Madison, WI 53706, USA

Received 21 November 1988
Revised manuscript received 3 July 1989

This paper concerns the notion of a sharp minimum on a set and its relationship to the proximal point algorithm. We give several equivalent definitions of the property and use the notion to prove finite termination of the proximal point algorithm.

Key words: Sharp minima, proximal point, finite termination.

1. Introduction

In this paper we are concerned principally with the convex programming problem

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

where ϕ is a closed, convex function defined on $\mathbb{R}^n$, having values in $\mathbb{R}$ and S is a closed, convex set in $\mathbb{R}^n$. We write $\bar{S}$ for the optimal solution set of (1), $\bar{S} := \arg\min_{x\in S} \phi(x)$ and assume this set to be non-empty, in order that a projection operation onto this set is well defined. In order to simplify our analysis, let us define

$$\phi_S(x) := \phi(x) + \psi(x \mid S)$$

and note that this is a closed convex function, since the indicator function of the set S, $\psi(\cdot \mid S)$, is respectively closed and convex if and only if S is closed and convex. We can now rewrite problem (1) as

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

The following is our central definition.

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

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

where $P(x \mid \bar{S}) \in \arg\min_{y\in\bar{S}} \|y - x\|$.

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.

Note that the definition holds independently of any convexity assumptions on S and ϕ, provided that the projection exists. Furthermore, it is clear that for any norm $\phi(P(x\,|\,\bar{S}))$ is constant for all $x \in S$, so we can replace $\phi(P(x\,|\,\bar{S}))$ by $\bar{\phi}$ in the above definition. It has been shown (Ferris, 1988) that the definition is independent of the choice of norm on the space $\mathbb{R}^n$. We note that this is a generalization of the concept of a sharp minimum at a point (Polyak, 1987), which is that ϕ has a sharp minimum at $\bar{x} \in S$, if there exists an $\alpha > 0$ with

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

Amongst other things, this implies that $\bar{x}$ is the unique minimum of ϕ on S.

It has been shown (Rockafellar, 1976) that a sharp minimum at $\bar{x}$ is a sufficient condition for finite termination of the proximal point algorithm. We show that the notion of a sharp minimum on $\bar{S}$ is sufficient for finite termination in the next section.

A key result relating finite termination of the proximal point algorithm to a sharp minimum point is that ϕ has a sharp minimum at $\bar{x}$ on S, if and only if $0 \in \operatorname{int} \partial\phi_S(\bar{x})$. We quote a generalization of this result for a sharp minimum on $\bar{S}$. The proof of this result can be found in Burke and Ferris (1991) and Ferris (1988). We use the following notation, $\partial\phi_S(\bar{S}) := \bigcup_{x \in \bar{S}} \partial\phi_S(x)$.

Theorem 2. *ϕ has a sharp minimum on $\bar{S}$ if and only if there exists $\varepsilon > 0$ with*

$$\varepsilon\left(B \cap \bigcup_{y \in \bar{S}} N(y\,|\,\bar{S})\right) \subseteq \partial\phi_S(\bar{S}). \qquad \square$$

The notation we use is, for the most part, standard. The following partial list is provided for the reader's convenience. Superscripts are used to distinguish between vectors, e.g., x^1, x^2, etc., and $\langle \cdot, \cdot \rangle$ is used to denote the inner product. For $C \subseteq \mathbb{R}^m$, bdry C is the boundary of C, $\psi(\cdot\,|\,C)$ is the indicator function for C, $\psi^*(\cdot\,|\,C)$ is the support functional for C and cone C is the cone generated by C. If C is, moreover, convex then $N(x\,|\,C)$ is the normal cone to C at $x \in C$.

For $f : \mathbb{R}^n \to \mathbb{R}$, f convex, $\partial f(x)$ is the subdifferential of f at x,

$$\partial f(x) := \{x^* \,|\, f(z) \geq f(x) + \langle x^*, z - x\rangle\}.$$

The symbol $\|\cdot\|$ denotes a given norm, and B denotes the associated closed unit ball. Throughout this paper, we assume the given norm to be the Euclidean norm. For $C \subseteq \mathbb{R}^n$ define $\operatorname{dist}(x\,|\,C) := \inf\{\|x - y\| \,|\, y \in C\}$ and if C is a closed convex set, then given $x \in \mathbb{R}^n$, we write $P(x\,|\,C)$ for the projection of x on C, $P(x\,|\,C) := \arg\min\{\|z - x\| \,|\, z \in C\}$.

2. Proximal point and sharp minima

The notion of a proximal point was introduced by Moreau (1965), 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 (1) has a sharp minimum on $\bar{S}$. 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 (1973) 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,

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

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

$$x^{i+1}=J_{\lambda_i}x^i, \tag{3}$$

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

$$\underset{x\in S}{\text{minimize}}\ \phi(x)+\frac{1}{2\lambda_i}\|x-x^i\|^2 \tag{4}$$

since the minimizer of (4) satisfies $0\in\partial\phi_S(x^{i+1})+(1/\lambda_i)(x^{i+1}-x^i)$, and hence x^{i+1} is of the form given by (3). Furthermore, x^{i+1} is unique by the strict convexity of the objective of (4), and a little algebra shows that $x^{i+1}+\lambda_i v^{i+1}=x^i$, where $v^{i+1}\in\partial\phi_S(x^{i+1})$. We wish to make two remarks. The first is to note the difference between (4) and Ekeland's Principle (Ekeland, 1974). Ekeland used a distance function as the perturbation, whereas here the square of the distance function is added. In (4) the perturbation function is differentiable, whereas this is not the case for the perturbation in Ekeland (1974). Second, note the difference between (4) and the perturbation results of Mangasarian (1984) for linear programs. In (4) the perturbation is defined in terms of the iterates of the algorithm, whereas in Mangasarian (1984) the perturbation function is fixed, independent of the algorithm and the iteration.

The following properties of the iterates of the algorithm follow easily.

Lemma 3 (Rockafellar, 1976).

(a) $\|v^i\|$ *is non-increasing for* $i=1,2,\ldots$.

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

$$\|x^{i+1}-z\|^2+\lambda_i^2\|v^{i+1}\|^2\leqslant\|x^i-z\|^2. \qquad \square$$

The following technical lemma is important for the proof of the ensuing theorem.

Lemma 4. *For any* $\lambda > 0$, *let* $y = \lambda(z - P(z\,|\,\bar{S}))$, *and suppose that* $y \in \partial\phi_S(w)$ *for some* $w \in \bar{S}$. *Then* $y \in \partial\phi_S(P(z\,|\,\bar{S}))$.

Proof. We note that

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

and therefore

$$w, P(z\,|\,\bar{S}) \in \bar{S} \Rightarrow \phi_S(x) - \phi_S(P(z\,|\,\bar{S})) \geqslant \langle y, x - w\rangle \quad \forall x \in \mathbb{R}^n.$$

It follows from the definition of $P(\cdot\,|\,\bar{S})$ that $\langle z - P(z\,|\,\bar{S}),\ w - P(z\,|\,\bar{S})\rangle \leqslant 0$ so that multiplication by $\lambda > 0$ gives $\langle y, w - P(z\,|\,\bar{S})\rangle \leqslant 0$. Hence, for all $x \in \mathbb{R}^n$,

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

that is, $y \in \partial\phi_S(P(z\,|\,\bar{S}))$. □

Lemma 5. *Suppose* $\exists \varepsilon > 0$ *such that*

$$\varepsilon\left(B \cap \bigcup_{y \in \bar{S}} N(y\,|\,\bar{S})\right) \subseteq \partial\phi_S(\bar{S}).$$

If $\|w\| \leqslant \varepsilon$ *and* $w \in \partial\phi_S(z)$, *then* $z \in \bar{S}$.

Proof. Let $\varepsilon > 0$ be chosen so that $\varepsilon(B \cap \bigcup_{y \in \bar{S}} N(y\,|\,\bar{S})) \subseteq \partial\phi_S(\bar{S})$. We choose w with $\|w\| < \varepsilon$ and $w \in \partial\phi_S(z)$.

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

$$y = \varepsilon(z - P(z\,|\,\bar{S}))/\|z - P(z\,|\,\bar{S})\|$$

so that $y \in \varepsilon(B \cap \bigcup_{y \in \bar{S}} N(y\,|\,\bar{S}))$. 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

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

from whence it follows that

$$\begin{aligned}&\frac{\varepsilon}{\|z - P(z\,|\,\bar{S})\|}\langle z - z^y, z - P(z\,|\,\bar{S})\rangle \\ &\qquad = \langle z - z^y, y\rangle \leqslant \langle z - z^y, w\rangle \leqslant \|z - z^y\|\,\|w\|.\end{aligned}$$

Hence,

$$\|w\| \geqslant \varepsilon \frac{\langle z - z^y, z - P(z\,|\,\bar{S})\rangle}{\|z - P(z\,|\,\bar{S})\|\,\|z - z^y\|}.$$

The result now follows using Lemma 4 which enables us to take $z^y = P(z\,|\,\bar{S})$, and hence derive the contradiction $\|w\| \geqslant \varepsilon$. The proof is now complete, since this gives $z = P(z\,|\,\bar{S})$. □

The following result is the central one of this section.

Theorem 6. *Suppose* (1) *has a sharp minimum on $\bar{S}$. Let $\{\lambda_i\}$ be any positive sequence which is bounded below and let $x^0 \in \mathbb{R}^n$. Then the proximal point algorithm terminates in a finite number of iterations.*

Proof. It follows from Theorem 2 that $\exists \varepsilon > 0$ with

$$\varepsilon\left(B \cap \bigcup_{y \in \bar{S}} N(y \mid \bar{S})\right) \subseteq \partial\phi_S(\bar{S}).$$

Let $\lambda_i \geq \lambda > 0$ for the given sequence. Then, for any $z \in \bar{S}$ we have from Lemma 3 that the sequence $\{\|x^i - z\|\}$ is bounded and hence converges. If we invoke the lemma again for $i = 0, \dots, N$ and sum, the following inequality results:

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

Using the above observations, it is clear that

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

so that

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

by the non-increasing property of $\|v^i\|$ given in Lemma 3. Hence, there exists a sufficiently large but finite N such that

$$\|v^{N+1}\|^2 \leq \frac{M}{\lambda^2(N+1)} < \varepsilon^2.$$

It then follows from Lemma 5 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 7. *Let $\{x^k\}$ be generated by the proximal point algorithm and let*

$$\bar{S}_k := \{x \in S \mid \phi(x) \leq \phi(x^k)\}.$$

Then

$$x^k = P(x^{k-1} \mid \bar{S}_k).$$

In particular, if the proximal point algorithm terminates in a finite number of iterations, k say, then $x^k = P(x^{k-1} \mid \bar{S})$.

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

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

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

$$0 \geq \langle v^k, P(x^{k-1} | \bar{S}_k) - x^k \rangle,$$

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

$$0 \geq \langle x^{k-1} - x^k, P(x^{k-1} | \bar{S}_k) - x^k \rangle.$$

The definition of $P(\cdot | \bar{S}_k)$ gives

$$0 \geq \langle x^{k-1} - P(x^{k-1} | \bar{S}_k), x^k - P(x^{k-1} | \bar{S}_k) \rangle,$$

so by adding the above inequalities we get

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

and hence $x^k = P(x^{k-1} | \bar{S}_k)$. □

The final theorem shows that a sharp minimum is sufficient for one step termination of the proximal point algorithm.

Theorem 8. *Suppose* (1) *has a sharp minimum on* $\bar{S}$. *Then for any given* x^0, *the proximal point algorithm terminates in one iteration for a sufficiently large choice of* λ.

Proof. It follows from Theorem 2 that there exists $\varepsilon > 0$ with $\varepsilon(B \cap \bigcup_{y \in \bar{S}} N(y | \bar{S})) \subseteq \partial \phi_S(\bar{S})$. We assume that $x^1 \neq P(x^0 | \bar{S})$. Define

$$y = \varepsilon(x^1 - P(x^0 | \bar{S})) / \| x^1 - P(x^0 | \bar{S}) \|$$

so that $y \in \varepsilon(B \cap \bigcup_{y \in \bar{S}} N(y | \bar{S}))$, and hence that for some $z^y \in \bar{S}$, $y \in \partial \phi_S(z^y)$. The monotonicity of $\partial \phi_S$ gives

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

from whence it follows that

$$\langle x^1 - z^y, y \rangle \leq \langle x^1 - z^y, v^1 \rangle \leq \| x^1 - z^y \| \, \| v^1 \|.$$

Hence

$$\| v^1 \| \geq \varepsilon \frac{\langle x^1 - z^y, x^1 - P(x^0 | \bar{S}) \rangle}{\| x^1 - P(x^0 | \bar{S}) \| \, \| x^1 - z^y \|} = \varepsilon, \tag{5}$$

the last equality by taking $z^y = P(x^0 | \bar{S})$ which is possible from Lemma 4. Now choose $\lambda \geq \operatorname{dist}(x^0 | \bar{S}) / \varepsilon$. By Lemma 3 we see that

$$\| x^1 - P(x^0 | \bar{S}) \|^2 + \lambda^2 \| v^1 \|^2 \leq \| x^0 - P(x^0 | \bar{S}) \|^2,$$

so that

$$\| x^1 - P(x^0 | \bar{S}) \|^2 \leq [\operatorname{dist}(x^0 | \bar{S})]^2 - \lambda^2 \| v^1 \|^2 \leq \lambda^2 \varepsilon^2 - \lambda^2 \| v^1 \|^2 \leq 0.$$

The last inequality follows from (5). But this is a contradiction, and so $x^1 = P(x^0 | \bar{S})$. □

Theorems 6 and 8 are generalizations of the corresponding results first obtained for linear programs by Polyak and Tretiyakov (1972).

A practical computational method for updating the sequence $\{\lambda_i\}$ of Theorem 6 is the one used by De Leone and Mangasarian (1988), and is given as follows:

$$\lambda_{i+1} = \begin{cases} \lambda_i & \text{if } \|x^{i+1} - x^i\| \leq \mu \|x^i - x^{i-1}\|, \ 0 < \mu < 1, \\ \nu\lambda_i & \text{otherwise, } \nu > 1. \end{cases}$$

This scheme is often used for updating the penalty parameter in an augmented Lagrangian algorithm. Furthermore, computational results with this method and extended parallel versions of the method are given by Polyak and Tretiyakov (1972). More problems having the property of a sharp minimum (which include amongst others non-degenerate monotone linear complementarity problems) are being investigated (Burke and Ferris, 1991; Mangasarian, 1990) to exploit the results of this paper fully.

Acknowledgement

I should like to thank Professor O. Mangasarian and Ms. L. Brady for many useful discussions during the preparation of this paper. I am grateful to Professor J. Burke for his helpful remarks regarding the proof of Theorem 2, and for his suggestions on improvements to the paper. Since the main results of this paper previously appeared (Ferris, 1988), the author has discovered an alternative proof of a similar result due to Bertsekas and Tsitsiklis (1989) which first appeared in 1989.

References

D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

H. Brézis, *Opérateurs Maximaux Monotones* (North-Holland, Amsterdam, 1973).

J.V. Burke and M.C. Ferris, "The sharpness of functions on sets," in preparation (1991).

R. De Leone and O.L. Mangasarian, "Serial and parallel solution of large scale linear programs by augmented Lagrangian successive overrelaxation," in: A. Kurzhanski, K. Neumann and D. Pallaschke, eds., *Optimization, Parallel Processing and Applications. Lecture Notes in Economics and Mathematical Systems No. 304* (Springer-Verlag, Berlin, 1988) pp. 103–124.

I. Ekeland, "On the variational principle," *Journal of Mathematical Analysis and Applications* 47 (1974) 324–353.

M.C. Ferris, "Weak sharp minima and penalty functions in mathematical programming," Technical Report 779, Computer Sciences Department, University of Wisconsin (Madison, WI, 1988).

O.L. Mangasarian, "Normal solutions of linear programs," *Mathematical Programming Study* 22 (1984) 206–216.

O.L. Mangasarian, "Error bounds for nondegenerate monotone linear complementarity problems," *Mathematical Programming* (*Series B*) 48 (1990) 437–445.

J.-J. Moreau, "Proximité et dualité dans un espace Hilbertien," *Bulletin de la Société Mathématique de France* 93 (1965) 273–299.

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

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

R.T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization* 14 (1976) 877–898.