JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 97, No. 2, pp. 357-383, MAY 1998

Convergence Analysis of Some Methods for Minimizing a Nonsmooth Convex Function[1,2]

J. R. BIRGE,[3] L. QI,[4] AND Z. WEI[5]

Communicated by P. Tseng

Abstract. In this paper, we analyze a class of methods for minimizing a proper lower semicontinuous extended-valued convex function $f: \Re^n \to \Re \cup \{\infty\}$. Instead of the original objective function f, we employ a convex approximation f_{k+1} at the kth iteration. Some global convergence rate estimates are obtained. We illustrate our approach by proposing (i) a new family of proximal point algorithms which possesses the global convergence rate estimate $f(x_k) - \min_{x\in\Re^n} f(x) = O(1/(\sum_{j=0}^{k-1} \sqrt{\lambda_j})^2)$ even if the iteration points are calculated approximately, where $\{\lambda_k\}_{k=0}^{\infty}$ are the proximal parameters, and (ii) a variant proximal bundle method. Applications to stochastic programs are discussed.

Key Words. Nonsmooth convex optimization, proximal point method, bundle algorithm, stochastic programming.

1. Introduction

Consider the following optimization problem:

$$\min\{f(x): x \in \Re^n\}, \tag{1}$$

where $f: \Re^n \to \Re \cup \{\infty\}$ is a proper lower semicontinuous extended-valued convex function.

[1]This material is based on work supported by the National Science Foundation, Award DDM 92-15921, and the Australian Research Council.

[2]The authors thank Professor Paul Tseng and two referees for helpful comments. Lemma 3.1 and the proof of Lemma 3.3(i) were improved because of referee suggestions.

[3]Professor, Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, Michigan.

[4]Associate Professor, School of Mathematics, University of New South Wales, Sydney, NSW, Australia.

[5]PhD Student, School of Mathematics, University of New South Wales, Sydney, NSW, Australia.

The Moreau–Yosida approximation F_λ of f is defined by

$$F_\lambda(x) = \min\{f(y) + (1/2\lambda)\|y - x\|^2 : y \in \Re^n\},$$

where λ is a real positive number. As proved by Moreau (Ref. 1), F_λ is a differentiable convex function defined in the whole space $\Re^n$ and possesses the same set of minimizers as f in (1). Using these properties, Martinet (Ref. 2) presented a proximal point algorithm for solving (1): start from an initial point $x_0 \in \Re^n$ and generate $\{x_k\}_{k=0}^{\infty}$ by solving

$$x_{k+1} = \operatorname{argmin}\{f(x) + (1/2\lambda_k)\|x - x_k\|^2 : x \in \Re^n\}, \tag{2}$$

where $\{\lambda_k\}_{k=0}^{\infty}$ is a sequence of positive numbers.

Under some additional reasonable assumptions, Rockafellar proved in Ref. 3 the local superlinear convergence of the proximal point algorithm for finding a zero of an arbitrary maximal monotone operator even if the iteration points are calculated approximately. When his results are applied to a lower semicontinuous proper convex function f, two general criteria for generating x_{k+1} are that

$$\operatorname{dist}(0, S_k(x_{k+1})) \le \sigma_{1k}/\lambda_k, \qquad \sum_{k=0}^{\infty} \sigma_{1k} < \infty, \tag{3}$$

and that

$$\operatorname{dist}(0, S_k(x_{k+1})) \le (\sigma_{2k}/\lambda_k)\|x_{k+1} - x_k\|, \qquad \sum_{k=0}^{\infty} \sigma_{2k} < \infty, \tag{4}$$

where

$$S_k(x) = \partial f(x) + (1/\lambda_k)(x - x_k). \tag{5}$$

For recent convergence results of the proximal point algorithm, we refer the reader to Refs. 4–12.

Güler presented in Ref. 10 two different proximal point algorithms which used an idea introduced by Nesterov (Ref. 13) for smooth convex minimization. His methods generate an additional sequence $\{y_k\}_{k=0}^{\infty}$ of points in $\Re^n$, and calculate x_{k+1} from

$$x_{k+1} = \operatorname{argmin}\{f(x) + (1/2\lambda_k)\|x - y_k\|^2 : x \in \Re^n\}. \tag{6}$$

He also showed that the minimization in (6) can be performed inexactly by a modification of (3), i.e.,

$$\operatorname{dist}(0, \partial f(x_{k+1}) + (1/\lambda_k)(x_{k+1} - y_k)) \le \sigma_{3k}/\lambda_k, \tag{7}$$

where $\sigma_{3k} = O(1/k^\sigma)$ for some $\sigma > 1/2$.

Lemaréchal combined the proximal point method with the bundle method in Ref. 14; also see Refs. 15–17. In his algorithm, a sequence

$\{x_k\}_{k=0}^{\infty}$ is generated by a sequence of convex functions $\{f_k\}_{k=0}^{\infty}$. More precisely,

$$x_{k+1} = \operatorname{argmin}\{f_k(x) + (1/2\lambda_k)\|x - x_k\|^2 : x \in \mathfrak{R}^n\}, \tag{8}$$

where f_k is a bundle linearization function of f. For the bundle method, also see Refs. 18–25.

In this paper, we study procedures that use a sequence of approximate objective functions $\{f_k\}_{k=0}^{\infty}$. Such approximation is necessary in some optimization problems [for example, stochastic programs (see Refs. 26–33), where the objective functions are too complex for exact evaluation]. For stochastic programs, the objective function involves the expected value

$$f(x) = E[F(x, \omega)] = \int_{\Omega} F(x, \omega)\mathscr{P}(d\omega), \tag{9}$$

where ω is a random vector defined on a probability space $(\Omega, \mathscr{A}, \mathscr{P})$. Thus, the precise evaluation of f and its subgradients involves multidimensional integration. To avoid the computational burden associated with this evaluation, the objective function f is replaced by a sequence of approximate functions $\{f_k\}_{k=0}^{\infty}$; see Refs. 26–33.

The remainder of the paper is organized as follows. In Section 2, we describe a model algorithm and give some global convergence rate estimates. As an application, in Section 3 we present a family of proximal point algorithms which calculate x_{k+1} with $u_{k+1} \in \partial f(x_{k+1})$ by

$$\begin{aligned} &\|u_{k+1} + (1/\lambda_k)(x_{k+1} - y_k)\| \\ &\leq \sigma_{4k}\|u_{k+1}\| + (\sigma_{5k}/\lambda_k)\|x_{k+1} - y_k\|, \end{aligned} \tag{10}$$

where σ_{4k} and σ_{5k} are some numbers in $(0, 1)$. In particular, under the same conditions, we obtain the following global convergence rate estimate obtained in Ref. 10 [with exact minimization (6)]:

$$f(x_k) - \min_{x \in \mathfrak{R}^n} f(x) = O(1/(\textstyle\sum_{j=0}^{k-1} \sqrt{\lambda_j})^2). \tag{11}$$

Note the following: (i) we obtain the same convergence results using the inexact minimization (10); for example, for all k, $\sigma_{4k} = \sigma_{5k} = 1/5$, or $\sigma_{4k} = 0$ and $\sigma_{5k} \in [0, 1/3]$, or $\sigma_{4k} \in [0, 1/2]$ and $\sigma_{5k} = 0$; of course, an essential difference from the proximal point algorithms in the literature is the fact that σ_{4k} and σ_{5k} can be bounded away from zero; and (ii) the convergence rate (11) for algorithm (10) is higher than that obtained for (7) in Ref. 10. Some applications in stochastic programs are discussed in Section 3. As another application of the results in Section 2, we present in Section 4 a variant proximal bundle method which calculates x_{k+1} by

$$x_{k+1} = \operatorname{argmin}\{f_{k+1}(x) + (1/2\lambda_k)\|x - y_k\|^2 : x \in \Re^n\}. \tag{12}$$

2. Model Method Algorithm

We describe briefly the main idea of the methods in Ref. 10. The idea there is to generate recursively a sequence $\{\varphi_k\}_{k=0}^{\infty}$ of simple convex quadratic functions that approximate f such that, at step $k \geq 0$, for all $x \in \Re^n$,

$$\varphi_{k+1}(x) - f(x) \leq (1 - \alpha_k)[\varphi_k - f(x)], \tag{13}$$

where $\alpha_k \in [0, 1)$. If (13) is satisfied for each $k \geq 0$, then

$$\varphi_k(x) - f(x) \leq \left(\prod_{i=0}^{k-1} (1 - \alpha_i)\right)[\varphi_0(x) - f(x)]. \tag{14}$$

If, at step k,

$$f(x_k) \leq \varphi_k^* := \min\{\varphi_k(z) : z \in \Re^n\}, \tag{15}$$

then from (14),

$$f(x_k) - f(x) \leq \left(\prod_{i=0}^{k-1} (1 - \alpha_i)\right)[\varphi_0(x) - f(x)], \tag{16}$$

which implies that, if $\prod_{i=0}^{k-1}(1 - \alpha_i) \to 0$, then $\{x_k\}_{k=0}^{\infty}$ is a minimizing sequence for f.

The aim of this section is to extend the above idea to a sequence of proper lower semicontinuous extended-valued convex functions $\{f_k\}_{k=0}^{\infty}$, which approximate f at step k. We assume that:

(A1) for any $x \in \operatorname{dom} f$ and all $k \geq 0$, $f_k(x) < +\infty$.

Let $x_0 \in \Re^n$, $\{x_k\}_{k=0}^{\infty} \subset \Re^n$, $u_{k+1} \in \partial f_{k+1}(x_{k+1})$, and let a constant $a > 0$ be given. For given $\alpha_k \in [0, 1)$, we define

$$\varphi_0(x) = f_0(x_0) + (a/2)\|x - x_0\|^2, \tag{17}$$

$$\begin{aligned} \varphi_{k+1}(x) = {} & (1 - \alpha_k)\varphi_k(x) + \alpha_k[f_{k+1}(x_{k+1}) + u_{k+1}^T(x - x_{k+1})] \\ & - (1 - \alpha_k)[f_k(x_k) - f_{k+1}(x_k)]. \end{aligned} \tag{18}$$

The following two lemmas slightly extend related results in Ref. 10. The proofs follow closely those of Ref. 10.

Lemma 2.1. For all k, the quadratic functions $\varphi_k(x)$ satisfy the following inequality:

$$\varphi_{k+1}(x)-f_{k+1}(x) \\ \leq(1-\alpha_k)[\varphi_k(x)-f_k(x)]+(1-\alpha_k)\delta_{k+1}(x), \tag{19}$$

where

$$\delta_{k+1}(x)=f_k(x)-f_{k+1}(x)+f_{k+1}(x_k)-f_k(x_k), \\ k=0, 1, 2, \ldots. \tag{20}$$

Proof. From the definition of φ_{k+1}, we have

$$\begin{aligned}
&\varphi_{k+1}(x)-f_{k+1}(x)\\
&=(1-\alpha_k)[\varphi_k(x)-f_k(x)]+(1-\alpha_k)f_k(x)-f_{k+1}(x)\\
&+\alpha_k[f_{k+1}(x_{k+1})-u_{k+1}^T(x-x_{k+1})]-(1-\alpha_k)[f_k(x_k)-f_{k+1}(x_k)]\\
&=(1-\alpha_k)[\varphi_k(x)-f_k(x)]\\
&+(1-\alpha_k)[f_k(x)-f_k(x_k)-f_{k+1}(x)+f_{k+1}(x_k)]\\
&-\alpha_k[f_{k+1}(x)-f_{k+1}(x_{k+1})-u_{k+1}^T(x-x_{k+1})].
\end{aligned}$$

Using the convexity of f_{k+1}, the conclusion (19) follows. □

Denote $\delta_0=0$, $\epsilon_0=0$, $\alpha_{-1}=0$, and

$$\epsilon_{k+1}(x)=(1-\alpha_{k-1})\epsilon_k(x)-\delta_{k+1}(x), \qquad k=0, 1, 2, \ldots. \tag{21}$$

For all j such that $1\leq j\leq k$, let

$$\Delta_{kj}=\prod_{i=1}^{j}(1-\alpha_{k-i}).$$

It is easy to show that

$$\epsilon_{k+1}(x)=\delta_{k+1}(x)+\sum_{j=1}^{k}\Delta_{kj}\delta_{k+1-j}(x). \tag{22}$$

From Lemma 2.1, we have

$$\varphi_{k+1}(x)-f_{k+1}(x) \\ \leq\left(\prod_{i=0}^{k}(1-\alpha_i)\right)[\varphi_0(x)-f_0(x)]+(1-\alpha_k)\epsilon_{k+1}(x). \tag{23}$$

Since for all k, the quadratic function φ_k can be written in canonical form (Ref. 10), we may let

$$\varphi_k(x) = \varphi_k^* + (a_k/2)\|x - v_k\|^2,$$

which combined with (18) yields

$$a_{k+1} = (1 - \alpha_k)a_k, \tag{24}$$

$$v_{k+1} = v_k - (\alpha_k/a_{k+1})u_{k+1}. \tag{25}$$

From (17), we have $a_0 = a$ and $v_0 = x_0$.

Lemma 2.2. If $\varphi_k^* \geq f_k(x_k)$, then

$$\varphi_{k+1}^* \geq f_{k+1}(x_{k+1}) + u_{k+1}^T[(1 - \alpha_k)x_k + \alpha_k v_k - x_{k+1} - (\alpha_k^2/2a_{k+1})u_{k+1}]. \tag{26}$$

Proof. From the definition of φ_{k+1}^*, (24), and (25), we have

$$\begin{aligned}
\varphi_{k+1}^* &= \varphi_{k+1}(v_{k+1}) \\
&= (1 - \alpha_k)\varphi_k(v_{k+1}) + \alpha_k[f_{k+1}(x_{k+1}) + u_{k+1}^T(v_{k+1} - x_{k+1})] \\
&\quad - (1 - \alpha_k)[f_k(x_k) - f_{k+1}(x_k)] \\
&= (1 - \alpha_k)\varphi_k^* + [(1 - \alpha_k)a_k/2]\|v_{k+1} - v_k\|^2 \\
&\quad - (1 - \alpha_k)[f_k(x_k) - f_{k+1}(x_k)] \\
&\quad + \alpha_k f_{k+1}(x_{k+1}) + \alpha_k u_{k+1}^T(v_{k+1} - x_{k+1}) \\
&\geq (1 - \alpha_k)[f_{k+1}(x_k) - f_{k+1}(x_{k+1})] \\
&\quad + (a_{k+1}/2)\|(\alpha_k/a_{k+1})u_{k+1}\|^2 \\
&\quad + \alpha_k u_{k+1}^T(v_k - (\alpha_k/a_{k+1})u_{k+1} - x_{k+1}) + f_{k+1}(x_{k+1}),
\end{aligned}$$

by the assumption in the lemma. Using the convexity of f_{k+1}, we have

$$\begin{aligned}
\varphi_{k+1}^* &\geq f_{k+1}(x_{k+1}) + (1 - \alpha_k)u_{k+1}^T(x_k - x_{k+1}) \\
&\quad + \alpha_k u_{k+1}^T(v_k - x_{k+1}) - (\alpha_k^2/2a_{k+1})\|u_{k+1}\|^2 \\
&= f_{k+1}(x_{k+1}) \\
&\quad + u_{k+1}^T[(1 - \alpha_k)x_k + \alpha_k v_k - x_{k+1} - (\alpha_k^2/2a_{k+1})u_{k+1}].
\end{aligned}$$

So, (26) follows. □

Letting

$$y_k = (1 - \alpha_k)x_k + \alpha_k v_k, \tag{27}$$

and choosing x_{k+1} with $u_{k+1} \in \partial f_{k+1}(x_{k+1})$ such that

$$q_{k+1} = u_{k+1}^T [y_k - x_{k+1} - (\alpha_k^2/2a_{k+1})u_{k+1}] \geq 0, \tag{28}$$

we have the following lemma.

Lemma 2.3. Suppose that (27) and (28) hold. Then, for $k = 0, 1, 2, \ldots,$

$$\varphi_k^* \geq f_k(x_k). \tag{29}$$

Furthermore, for $k = 0, 1, 2, \ldots,$

$$\varphi_k^* \geq f_k(x_k) + q_k. \tag{30}$$

Proof. We prove (29) by induction. From the definition of φ_0, (29) holds for $k = 0$. Suppose that (29) holds for k. From Lemma 2.2 and (28), we have that (29) holds for $k+1$. Then, Lemma 2.2, (26), and (29) imply that (30) holds. □

Algorithm 2.1. Model Method Algorithm (MMA).

Step 0. Initialization. Select an initial point $x_0 \in \text{dom } f$. Let $v_0 = x_0$, $a_0 = a > 0$, $\alpha_0 \in (0, 1)$, and $k = 0$.

Step 1. Set
$y_k = (1 - \alpha_k)x_k + \alpha_k v_k$.

Step 2. Generate f_{k+1} satisfying assumption (A1). Then, compute x_{k+1} with $u_{k+1} \in \partial f_{k+1}(x_{k+1})$ such that (28) holds, that is,

$$q_{k+1}^{MMA} = u_{k+1}^T \{y_k - x_{k+1} - [\alpha_k^2/2(1 - \alpha_k)a_k]u_{k+1}\} \geq 0.$$

Set

$$a_{k+1} = (1 - \alpha_k)a_k,$$

$$v_{k+1} = v_k - (\alpha_k/a_{k+1})u_{k+1}.$$

Choose

$$\alpha_{k+1} \in (0, 1).$$

Step 3. Increase k by 1, and go to Step 1.

It is worth noting that, for any $y_k \in \Re^n$, any $\alpha_k \in (0, 1)$, and any $a_k > 0$, we can always find x_{k+1} and $u_{k+1} \in \partial f_{k+1}(x_{k+1})$ such that $q_{k+1}^{MMA} \geq 0$. In fact, since

$$\partial f_{k+1}(x) + \{1/[\alpha_k^2/(a_k(1 - \alpha_k))]\}(x - y_k)$$

is a strongly monotone mapping with modulus $1/[\alpha_k^2/(a_k(1-\alpha_k))]$, there is a unique solution x_{k+1} such that

$$0 \in \partial f_{k+1}(x_{k+1}) + \{1/[\alpha_k^2/(a_k(1-\alpha_k))]\}(x_{k+1} - y_k).$$

Let $u_{k+1} \in \partial f_{k+1}(x_{k+1})$ such that

$$0 = u_{k+1} + \{1/[\alpha_k^2/(a_k(1-\alpha_k))]\}(x_{k+1} - y_k).$$

Then, (x_{k+1}, u_{k+1}) is a desired solution.

From (23) and Lemma 2.3, we obtain the following convergence rate estimate.

Theorem 2.1. Suppose that $\{x_k\}_{k=0}^{\infty}$ is generated by (MMA). Then, for all $x \in \text{dom } f$, $k \geq 1$,

$$f_k(x_k) - f_k(x) + q_k^{MMA} \leq \beta_k[\varphi_0(x) - f_0(x)] + (1 - \alpha_{k-1})\epsilon_j(x),$$

where

$$\beta_k = \prod_{i=0}^{k-1} (1 - \alpha_i).$$

The model method algorithm gives a variety of choices on different (a) approximation sequences $\{f_k\}_{k=0}^{\infty}$, (b) solution methods for q_k^{MMA}, and (c) parameters α_k. In the remainder of this section, we discuss (a). We will discuss (b) and (c) in Section 3.

Denote $\delta_0^0 = 0$, $\epsilon_0^0 = 0$, $\alpha_{-1}^0 = 0$,

$$\delta_{k+1}^0 = f_{k+1}(x_k) - f_k(x_k), \qquad k = 0, 1, 2, \ldots, \tag{31}$$

$$\epsilon_{k+1}^0 = (1 - \alpha_{k-1})\epsilon_k^0 + \delta_{k+1}^0, \qquad k = 0, 1, 2, \ldots. \tag{32}$$

Then,

$$\epsilon_{k+1}^0 = \delta_{k+1}^0 + \sum_{j=1}^{k} \Delta_{kj}\delta_{k+1-j}^0. \tag{33}$$

Corollary 2.1. Suppose that $\{x_k\}_{k=0}^{\infty}$ is generated by (MMA) with $f_k(x) \leq f_{k+1}(x)$ for $x \in \text{dom } f$ and $k \geq 0$. Suppose that f_k satisfies the following condition:

(A2) There is an index set K such that, for all $x \in \text{dom } f$,

$$\limsup_{k \in K, k \to \infty} f_{k+1}(x) \leq f(x). \tag{34}$$

Suppose that

$$\lim_{k\in K, k\to\infty} \beta_{k+1}=0, \tag{35}$$

$$\lim_{k\in K, k\to\infty} (1-\alpha_k)\epsilon_{k+1}^0=0, \tag{36}$$

$$\limsup_{k\in K, k\to\infty} f_{k+1}(x_{k+1})\geq \limsup_{k\in K, k\to\infty} f(x_k)$$

$$[\limsup_{k\in K, k\to\infty} f_{k+1}(x_{k+1})\geq \limsup_{k\in K, k\to\infty} f(y_k)]. \tag{37}$$

Then, $\{x_k\}_{k\in K}[\{yu_k\}_{k\in K}]$ is a minimizing sequence for f.

Proof. From the assumption that $f_k(x)\leq f_{k+1}(x)$, we have, for $x\in\operatorname{dom} f$ and $k\geq 0$, that $\delta_k(x)\leq\delta_k^0$ and $\epsilon_k(x)\leq\epsilon_k^0$. From Theorem 2.1 and the assumptions in this corollary, we have, for all $x\in\operatorname{dom} f$,

$$\limsup_{k\in K, k\to\infty} f(x_k)\leq f(x),$$

which implies that $\{x_k\}_{k\in K}$ is a minimizing sequence for f. The same property can be obtained for $\{y_k\}_{k\in K}$. □

The following result indicates that, for any bounded sequence $\{\delta_k^0\}_{k=0}^{\infty}$, we can choose $\alpha_k\in(0, 1)$ such that (35) and (36) hold.

Lemma 2.4. Suppose that $\{|\delta_k^0|\}_{k=0}^{\infty}$ is bounded.

(I) If there is $\bar{\alpha}>0$ such that, for all $k\geq 0$, $\alpha_k\geq\bar{\alpha}>0$, then (35) holds and $\{|\epsilon_k^0|\}_{k=0}^{\infty}$ is bounded.

(II) If $\alpha_k\to 1$, then (36) holds.

Proof. From the boundedness of δ_k^0, we have $M>0$ such that, for all $k\geq 0$, $|\delta_k^0|\leq M$. From the definition of Δ_{kj} we have that, for all j: $1\leq j\leq k$,

$$\Delta_{kj}=\prod_{i=1}^{j} (1-\alpha_{k-i})\leq(1-\bar{\alpha})^j.$$

This inequality combined with (32) yields

$$|\epsilon_{k+1}^0|\leq M\left(1+\sum_{j=1}^{k} (1-\bar{\alpha})^j\right),$$

which implies that $\{|\epsilon_k^0|\}_{k=0}^{\infty}$ is bounded. Since

$$0<\beta_k=\prod_{j=0}^{k-1}(1-\alpha_j)\leq(1-\tilde{\alpha})^k,$$

(35) holds. Since the condition in (II) implies the condition in (I), the conclusion (II) follows from $\alpha_k\to 1$ and the fact that $\{|\epsilon_k^0|\}_{k=0}^{\infty}$ is bounded. □

Denote

$$f^*=\inf\{f(x): x\in\Re^n\},$$

$$X^*=\{x; x\in\Re^n, f(x)=f^*\},$$

$$f_0^*=\inf\{f_0(x): x\in X^*\}.$$

Corollary 2.2. Suppose that $\{x_k\}_{k=0}^{\infty}$ is generated by (MMA) with $f_k(x)\leq f_{k+1}(x)$ for $x\in\text{dom}\, f$ and $k\geq 0$; let $f^*>-\infty$ and $f_0^*>-\infty$. Suppose that f_k satisfies the following condition:

(A2)* For all $k\geq 0$, there is a constant $b_k\geq 0$, dependent on k and f, such that, for any $x\in\text{dom}\, f$,

$$|f(x)-f_k(x)|\leq b_k. \tag{38}$$

Then, the following results hold.

(I) For any $x\in\text{dom}\, f$,

$$f(x_k)-f(x)+q_k^{MMA} \leq\beta_k[\varphi_0(x)-f_0(x)]+(1-\alpha_{k-1})\epsilon_k^0+2b_k. \tag{39}$$

In particular, we have the convergence rate estimate

$$f(x_k)-f^*+q_k^{MMA} \leq\beta_k[f_0(x_0)-f_0^*+(a/2)\rho(x_0,X^*)^2]+(1-\alpha_{k-1})\epsilon_k^0+2b_k. \tag{40}$$

(II) If

$$\lim_{k\to r} b_k=0, \tag{41}$$

and if (35) and (36) hold, then $\{x_k\}_{k=0}^{\infty}$ is a minimizing sequence for f. In particular, if there is $r\in(0,1)$ such that $\alpha_k=1-r$ and $b_k=r^k$, for all $k\geq 0$, then $\{x_k\}_{k=0}^{\infty}$ is a minimizing sequence for f. Moreover,

$$f(x_k)-f^*=O(kr^k). \tag{42}$$

(III) Suppose that (35), (36), and (41) hold. If X^* is a nonempty compact subset in $\Re^n$, then $\{x_k\}_{k=0}^{\infty}$ is bounded and every accumulation point of $\{x_k\}_{k=0}^{\infty}$ is a minimizer for f.

Proof. It is easy to prove conclusions (I) by Theorem 2.1. From (39), we have that $\{x_k\}_{k=0}^{\infty}$ is a minimizing sequence for f if (35), (36), and (41) hold. We prove (42) now. Suppose that $k \geq 1$. From the definitions of δ_k^0, Δ_{kj}, and (38), we have

$$\delta_k^0 \leq r^{k-1} + r^k,$$

$$\Delta_{kj} = r^j.$$

The above relations and (33) yield

$$\epsilon_k^0 \leq k(r^{k-1} + r^k).$$

This inequality combined with $\beta_k = r^k$ and (39) yields

$$f(x_k) - f^* + q_k \leq [f_0(x_0) - f_0^* + (a/2)\,\rho(x_0, X^*)^2 + (1+r)k + 2]r^k,$$

which implies (42). Conclusion (III) follows from the fact that $\{x_k\}_{k=0}^{\infty}$ is a minimizing sequence for f and that X^* is compact. □

If $f_k = f$, for all $k \geq 0$, then $\delta_k^0 = 0$. In this case, we can choose $b_k = 0$. By noting Theorem 2.1, we have the following corollary.

Corollary 2.3. Suppose that $\{x_k\}_{k=0}^{\infty}$ is generated by (MMA). If $f_k = f$, for any $k \geq 0$, then

$$f(x_k) - f(x) + q_k^{MMA} \leq \beta_k [f(x_0) - f(x) + (a/2)\|x - x_0\|^2]. \quad (43)$$

Consequently,

$$f(x_k) - f^* \leq \beta_k [f(x_0) - f^* + (a/2)\rho(x_0, X^*)^2], \quad (44)$$

and (MMA) converges $[f(x_k) \to f^*]$ if $\beta_k \to 0$. Furthermore, (MMA) has the following global convergence rate estimate:

$$f(x_k) - f^* \leq O(\beta_k),$$

where

$$\rho(z, W) = \min\{\|z - w\| : w \in W\}.$$

Remark 2.1. Noting the definitions of ϵ_k^0 and $\epsilon_k^0(x)$, we can prove that results similar to Corollary 2.2 hold even when $f_k(x) \leq f_{k+1}(x)$ is not true.

Remark 2.2. The results in Corollary 2.2 or Remark 2.1 are useful for solving some convex complex problems. The following examples can be viewed as general models arising from stochastic programming (see Refs. 27–34).

Example 2.1. Suppose that f has the following structure:

$$f(x) = h_0(x) + \sum_{k=1}^{\infty} p_k h_k(x) + \Theta_\chi(x),$$

where $\chi \subset \Re^n$ is a nonempty compact convex subset of $\Re^n$ and Θ_χ is the indicator function of χ, i.e.,

$$\Theta_\chi(x) = \begin{cases} 0, & \text{if } x \in \chi, \\ +\infty, & \text{otherwise.} \end{cases}$$

Assume that:

(B1) for $j \geq 0$, $h_j : \Re^n \to \Re$ is a convex function;
(B2) for $j \geq 1$, $h_j(x) \geq 0$ and there is a constant $M_0 > 0$ such that

$$\sup_{k \geq 1} \{h_j(x) : x \in \chi\} \leq M_0;$$

(B3) for $j \geq 1$, $p_j \geq 0$ and

$$\sum_{j=1}^{\infty} p_j < +\infty.$$

Let

$$f_k(x) = h_0(x) + \sum_{i=1}^{k} p_i h_i(x) + \Theta_\chi(x),$$

and solve (1) by (MMA). If $h_j(x) \geq 0$ or $p_j \geq 0$ is not true, we can also solve this example by noting Remark 2.1.

Example 2.2. Another problem is

$$\min\left\{ f(x) = \int_\Omega F(x, y)\, dy + \Theta_\chi(x) : x \in \Re^n \right\}, \tag{45}$$

where Ω is a compact subset of $\Re^m$ and $F(\cdot, \cdot) : \Re^n \times \Re^m \to \Re$ is of bounded variation on Ω in the sense of Hardy and Krause; $F(\cdot, y)$ is convex for any given $y \in \Omega$. In this case, according to some integration rules (see Refs. 35–38 for details), we can choose, $\Omega_j^k \subset \Omega$ and $y_j^k \in \Omega_j^k$, for j such that $0 \leq j \leq k-1$, so that

$$f_k(x) = \sum_{j=0}^{k-1} F(x, y_j^k)\mu(\Omega_j^k) + \Theta_\chi(x)$$

satisfies the following conditions:

(C1) $\bigcup_{j=0}^{k-1} \Omega_j^k = \Omega$, for all $k \geq 1$, and $\operatorname{int} \Omega_i^k \cap \operatorname{int} \Omega_j^k = \varnothing$, for all $0 \leq i < j \leq k-1$, where int denote the interior of a set;

(C2) $\lim_{k \to \infty} \sup\{|f(x) - f_k(x)| : x \in \chi\} = 0$.

From Remark 2.1, we can solve (45) by (MMA).

3. New Proximal Point Algorithms

A new family of proximal point algorithms with four parameters (λ_k, α_k, σ_{4k}, σ_{5k}) is proposed in this section by introducing a new solution method for (28). In fact, Güler (Ref. 10) gave method (6) for solving (28). We use (10) here to solve (28). The method described below is based on Lemma 3.1, which claims that we can choose λ_k, σ_{4k}, σ_{5k} such that the solution set of (10) is contained in the solution set of (28).

Lemma 3.1. Let u and v be vectors of $\Re^n$; let $\tau \in [0, 1]$ and $t \in [0, 1)$. Assume that

$$\|u + v\| \leq \tau \|u\| + t\|v\|. \tag{46}$$

Then,

$$-u^T v \geq [1 - (\tau + t)/(1 - t)]\|u\|^2. \tag{47}$$

Proof. The inequality

$$(u + v)^T u \leq \|u + v\| \, \|u\|$$

and (46) imply that

$$-u^T v \geq (1 - \tau)\|u\|^2 - t\|u\| \, \|v\|. \tag{48}$$

On the other hand, from (46), we have

$$\|v\| \leq [(1 + \tau)/(1 - t)]\|u\|,$$

which combined with (48) yields (47). □

From Lemma 3.1, we have the following corollary.

Corollary 3.1. Let u, v, w be vectors of $\Re^n$; let $\lambda>0$, $\tau\in[0, 1]$, and $t\in[0, 1)$. Suppose that

$$\|u+(1/\lambda)(v-w)\|\leq\tau\|u\|+(t/\lambda)\|v-w\|.$$

Then,

$$u^T(w-v)\geq[1-(\tau+t)/(1-t)]\lambda\|u\|^2.$$

Let

$$\Psi(\tau; t)=1-(\tau+t)/(1-t).$$

From Corollary 3.1, we now state our general proximal point algorithm.

Algorithm 3.1. General Proximal Point Algorithm (GPPA).

Step 0. Initialization. Select an initial point $x_0\in\text{dom } f$. Let $v_0=x_0$, $a_0=a>0$, and $k=0$.

Step 1. Choose $\lambda_k>0$, $\alpha_k\in(0, 1)$, $\sigma_{4k}\in[0, 1]$, and $\sigma_{5k}\in[0, 1)$ such that

$$\alpha_k^2/2a_k(1-\alpha_k)\leq\lambda_k\Psi(\sigma_{4k}, \sigma_{5k}); \tag{49}$$

set

$$y_k=(1-\alpha_k)x_k+\alpha_k v_k.$$

Step 2. Generate f_{k+1} satisfying (A1). Then, compute x_{k+1} with $u_{k+1}\in\partial f_{k+1}(x_{k+1})$ such that

$$\|u_{k+1}+(1/\lambda_k)(x_{k+1}-y_k)\|$$
$$\leq\sigma_{4k}\|u_{k+1}\|+(\sigma_{5k}/\lambda_k)\|x_{k+1}-y_k\|. \tag{50}$$

Set

$$a_{k+1}=(1-\alpha_k)a_k,$$

$$v_{k+1}=v_k-(\alpha_k/a_{k+1})u_{k+1}.$$

Step 3. Increase k by 1, and go to Step 1.

Theorem 3.1. Suppose that $\sigma_{4k}\in[0, 1]$ and $\sigma_{5k}\in[0, 1)$. If (x_{k+1}, u_{k+1}) with $u_{k+1}\in\partial f_{k+1}(x_{k+1})$ is a solution of (50), then

$$u_{k+1}^T(y_k-x_{k+1})\geq\Psi(\sigma_{4k}, \sigma_{5k})\lambda_k\|u_{k+1}\|^2. \tag{51}$$

Therefore, if (49) holds, then (x_{k+1}, u_{k+1}) is a solution for (28). Moreover,

$$q_{k+1}^{MMA} \geq q_{k+1}^{GPPA} = (\lambda_k/2)\{2\Psi(\sigma_{4k}, \sigma_{5k}) - [\alpha_k^2/(1-\alpha_k)a_k\lambda_k]\}\|u_{k+1}\|^2 \geq 0. \quad (52)$$

Proof. For any given k, let

$$u = u_{k+1}, \quad v = x_{k+1}, \quad w = y_k, \quad \lambda = \lambda_k,$$

$$\tau = \sigma_{4k}, \quad t = \sigma_{5k}.$$

From Lemma 3.1, we have (51). From the definition of q_k^{MMA} and (51), if (49) holds, then $q_k^{MMA} \geq 0$, so that (x_{k+1}, u_{k+1}) is a solution for (28); (52) follows the definitions of q_k^{MMA} and q_k^{GPPA} □

We are interested in finding σ_{4k}, σ_{5k}, and α_k such that β_k tends to 0 as fast as possible for any given sequence $\{\lambda_k\}_{k=0}^{\infty}$ (see Ref. 10). From the definition of β_k, this is equivalent to having α_k as large as possible. To find such α_k, for any $c>0$, set

$$c - \alpha_k^2/(1-\alpha_k)a_k\lambda_k = 0, \quad (53)$$

or

$$\alpha_k^2 + ca_k\lambda_k\alpha_k - ca_k\lambda_k = 0.$$

Therefore,

$$\alpha_k(c) = [\sqrt{(ca_k\lambda_k)^2 + 4ca_k\lambda_k} - ca_k\lambda_k]/2. \quad (54)$$

Similarly to the proof of Lemma 2.2 in Ref. 10, we can prove the following lemma.

Lemma 3.2. For all k,

$$1\bigg/\left[1 + \sqrt{ca}\cdot\sum_{j=0}^{k-1}\sqrt{\lambda_j}\right]^2 \leq \beta_k(c) \leq 1\bigg/\left[1 + (\sqrt{ca}/2)\sum_{j=0}^{k-1}\sqrt{\lambda_j}\right]^2. \quad (55)$$

Let

$$\Sigma(c) = \{(\tau, t): \tau \in [0, 1], t \in [0, 1), \Psi(\tau, t) \geq c/2\}.$$

Since for any $c \in (0, 2]$,

$$\{(\tau, t): t = \tau \leq (2-c)/(6-c)\} \subset \Sigma(c),$$

$\Sigma(c) \neq \varnothing$. From Corollary 2.3, we have the following convergence rate result.

Theorem 3.2. Suppose that, for all k, $f_k = f$, α_k satisfies (53). If $(\sigma_{4k}, \sigma_{5k}) \in \Sigma(c)$, then for any $x \in \operatorname{dom} f$, (GPPA) for any c has the global convergence rate estimate

$$\begin{aligned} & f(x_k) - f(x) + (\lambda_{k-1}/2)[2\Psi(\sigma_{4(k-1)}, \sigma_{5(k-1)}) - c]\|u_k\|^2 \\ & \leq [f(x_0) - f(x) + (a/2)\|x - x_0\|^2] \Big/ \left[1 + (\sqrt{ca}/2) \sum_{j=0}^{k-1} \sqrt{\lambda_j}\right]^2 \\ & \leq O\left(1 \Big/ \left(\sum_{j=0}^{k-1} \sqrt{\lambda_j}\right)^2\right). \end{aligned} \tag{56}$$

Therefore,

$$\begin{aligned} & f(x_k) - f^* \\ & \leq [4(f(x_0) - f^* + (a/2)\rho(x_0, X^*)^2)] \Big/ \left[ca\left(\sum_{j=0}^{k-1} \sqrt{\lambda_j}\right)^2\right]. \end{aligned} \tag{57}$$

Algorithm (GPPA) for any c converges $[f(x_k) \to f^*]$ if

$$\sum_{k=0}^{\infty} \sqrt{\lambda_k} = \infty. \tag{58}$$

In particular, if $\lambda_k \geq \lambda > 0$, for all $k \geq 0$, then

$$\begin{aligned} f(x_k) - f^* & \leq (4/a\lambda c k^2)[f(x_0) - f^* + (a/2)\rho(x_0, X^*)^2] \\ & = O(1/k^2). \end{aligned} \tag{59}$$

Since

$$\begin{aligned} \alpha_k'(c) = 2a_k\lambda_k / [(2 + ca_k\lambda_k + \sqrt{(ca_k\lambda_k)^2 + 4ca_k\lambda_k}) \\ \times \sqrt{(ca_k\lambda_k)^2 + 4ca_k\lambda_k}] > 0, \end{aligned}$$

$\alpha_k(c)$ is an increasing function of c. On the other hand, since $\Psi(0, 0) = 1$ and $t \in (0, 1)$, for all $\tau \in (0, 1]$,

$$\Psi(\tau, t) = 1 - (t + \tau)/(1 - t) < 1.$$

Therefore, $c = 2$, i.e.,

$$\alpha_k(2) = \sqrt{(a_k\lambda_k)^2 + 2a_k\lambda_k} - a_k\lambda_k$$

is the best choice for $\beta_k \to 0$ as fast as possible for a given sequence $\{\lambda_k\}_{k=0}^{\infty}$. From Lemma 3.2, we have

$$1\Big/\left[1+\sqrt{2a}\sum_{j=0}^{k-1}\sqrt{\lambda_j}\right]^2 \le \beta_k(2)$$

$$\le 1\Big/\left[1+(\sqrt{2a}/2)\sum_{j=0}^{k-1}\sqrt{\lambda_j}\right]^2 \tag{60}$$

and the following result.

Corollary 3.2. Suppose that, for all $k, f_k = f$, $c=2$, and α_k satisfies (53). If $\sigma_{4k}=\sigma_{5k}=0$, then for any $x \in \operatorname{dom} f$, (GPPA) for $c=2$ has the global convergence rate estimate

$$f(x_k)-f(x)$$

$$\le [f(x_0)-f(x)+(a/2)\|x-x_0\|^2]\Big/\left[1+(\sqrt{2a}/2)\sum_{j=0}^{k-1}\sqrt{\lambda_j}\right]^2. \tag{61}$$

Therefore,

$$f(x_k)-f^*$$

$$\le 2[f(x_0)-f^*+(a/2)\rho(x_0,X^*)^2]\Big/\left[a\left(\sum_{j=0}^{k-1}\sqrt{\lambda_j}\right)^2\right]. \tag{62}$$

In Ref. 10, Güler selected $c=1$ and gave the convergence rate results (56)–(59) with the calculation of x_{k+1} performed exactly by (6). Let

$$\Sigma_{E1}=\{(\tau,t)\colon \tau=0 \text{ and } t\in[0,1/3];\ \tau\in[0,1/2] \text{ and } t=0\}.$$

Since $\Psi(0;1/3)=\Psi(1/3,0)=1/2$,

$$\Psi(0,t)=1-t/(1-t)=2-1/(1-t) \qquad [\text{also } \Psi(\tau;0)=1-\tau]$$

is decreasing for $t\in[0,1)$, $\tau\in[0,1/3)$, $\Sigma_{E1}\subset\Sigma(1)$. Therefore, we have the following corollary.

Corollary 3.3. Suppose that, for all k, $f_k=f$, $c=1$, and α_k satisfies (53). If $(\sigma_{4k},\sigma_{5k})\in\Sigma_{E1}$, then for any $x\in\operatorname{dom} f$, (GPPA) has the global convergence rate estimate (56), (57), and (59) for the case $c=1$.

From (62) and (57) with $c=1$, we can deduce that the convergence rate of (GPPA) obtained for $c=2$ is twice faster than that obtained for $c=1$.

Remark 3.1. From Theorem 3.2 in this paper and Theorem 2.3 in Ref. 10, we obtained, for (GPPA) with $c=1$ the same convergence rates as those obtained in Ref. 10 for the proximal point algorithm (PPA), but (GPPA) with $c=1$ can be executed inexactly and the convergence rate obtained in this paper is higher than the convergence rate obtained for the algorithm with inexact minimization in Ref. 10; see Theorem 3.2 in this paper and Theorem 3.3 in Ref. 10; furthermore, we do not need that σ_{4k} and σ_{5k} tend to 0. Of course, from a practical point of view, it is also essential to replace $\sigma_{4k}=0$ and $\Sigma_{k=0}^{\infty}\sigma_{5k}<\infty$ [see (4)] by the looser relation which allows σ_{4k} and σ_{5k} bounded away from zero. The looser relation in this paper yields that (GPPA) is not always a standard proximal point algorithm, since x_{k+1} cannot be an approximate minimizer of $f(x)+(1/2\lambda_k)\|x-y_k\|^2$.

In the following, we will give another choice for α_k and prove that $\{x_k\}_{k=0}^{\infty}$ is an asymptotically regular sequence [$\|x_{k+1}-x_k\|\to 0$] under the condition that X^* is a compact set. This result has not been discussed in Ref. 10 and does not appear clear for this type of algorithm. We only prove it in a special case of the choice for α_k.

Algorithm 3.2. Special case of (GPPA).

Step 0. Initialization. Select an initial point $x_0\in\text{dom}\, f$. Let $v_0=x_0$, $a_0=a>0$, $\lambda_0>0$, and $k=0$.

Step 1. For x_k, v_k, a_k, λ_k, set

$$\alpha_k=a_k\lambda_k/(1+a_k\lambda_k),$$

$$y_k=(1-\alpha_k)x_k+\alpha_k v_k.$$

Step 2. Compute

$$x_{k+1}=\text{argmin}\{f(z)+(1/2\lambda_k)\|z-y_k\|^2: z\in\Re^n\},$$

$$v_{k+1}=v_k+(x_{k+1}-y_k),$$

$$a_{k+1}=(1-\alpha_k)a_k,$$

and choose $\lambda_{k+1}>0$.

Step 3. Increase k by 1, and go to Step 1.

It is not hard to show that Algorithm 3.2 is a special case of Algorithm 3.1. In fact, since

$$\alpha_k=a_k\lambda_k/(1+a_k\lambda_k),$$

we have

$$\lambda_k = \alpha_k / a_k(1-\alpha_k) \geq \alpha_k^2/2a_k(1-\alpha_k),$$

which implies that (49) holds; on the other hand, in this algorithm, we can let

$$u_{k+1} = -(1/\lambda_k)(x_{k+1} - y_k);$$

therefore,

$$-(\alpha_k/a_{k+1})u_{k+1} = (\alpha_k/a_{k+1})(1/\lambda_k)(x_{k+1} - y_k).$$

This relation combined with the definition of λ_k and the construction of a_k yields

$$-(\alpha_k/a_{k+1})u_{k+1} = x_{k+1} - y_k,$$

which implies that

$$v_{k+1} = v_k - \alpha_k(u_{k+1}/a_{k+1}).$$

From the above discussions, we can deduce that Algorithm 3.2 is a special case of Algorithm 3.1.

Lemma 3.3. Suppose that $\{\lambda_k\}_{k=0}^{\infty}$, $\{\alpha_k\}_{k=0}^{\infty}$, and $\{\beta_k\}_{k=0}^{\infty}$ are generated by Algorithm 3.2. Then, the following results hold:

(I) For $k = 1, 2, \ldots,$

$$\beta_k = 1\bigg/\left(1 + a\sum_{i=0}^{k-1} \lambda_i\right).$$

(II) If $\left\{k\lambda_k \bigg/ \sum_{j=0}^{k-1} \lambda_j\right\}_{k=1}^{\infty}$ is bounded,

$$\lim_{k\to\infty} k\alpha_k^2\left(\sum_{i=1}^{k} \beta_i \lambda_{i-1}\right) = 0, \tag{63}$$

$$\lim_{k\to\infty} \beta_{k+1}\lambda_k = 0. \tag{64}$$

Proof.

(I) For any given $i \geq 0$, from $1 - \alpha_i = 1/(1 + a_i\lambda_i)$ and $a_{i+1} = (1-\alpha_i)a_i$, we have

$$1/(1 + a_i\lambda_i) = a_{i+1}/a_i,$$

i.e.,

$$\lambda_i = 1/a_{i+1} - 1/a_i.$$

Hence,

$$\sum_{i=0}^{k-1} \lambda_i = 1/a_k - 1/a_0,$$

which combined with $a_0 = a$ and $\beta_k = a_k/a$ yields conclusion (I).

(II) Set

$$c_k = k\alpha_k^2 \left(\sum_{i=1}^{k} \beta_i \lambda_{i-1} \right).$$

Then,

$$c_k = ka^2 \left[\sum_{i=1}^{k} \lambda_{i-1} \Big/ \left(1 + a \sum_{j=0}^{i-1} \lambda_j\right)\right]\left[\lambda_k \Big/ \left(1 + a \sum_{i=0}^{k} \lambda_i\right)\right]^2.$$

Since $\{k\lambda_k / \sum_{j=0}^{k-1} \lambda_j\}_{k=1}^{\infty}$ is bounded, there is $M_1 > 0$, such that, for all $i \geq 1$,

$$i\lambda_i \Big/ \sum_{j=0}^{i-1} \lambda_j \leq M_1.$$

Noting that $\lambda_i > 0$, for all $i \geq 0$, we have

$$c_k \leq (M_1^3/a)(1/k) \sum_{i=1}^{k} (1/i).$$

Conclusion (63) follows using

$$(1/k) \sum_{i=1}^{k} (1/i) \to 0.$$

Since

$$\begin{aligned} \beta_{k+1}\lambda_k &= (1/k)\{k\lambda_k/[1 + a(\lambda_0 + \cdots + \lambda_k)]\} \\ &\leq (1/ak)k\lambda_k/(\lambda_0 + \cdots + \lambda_{k-1}), \end{aligned}$$

(64) follows from the assumption. □

Lemma 3.4. Suppose that $\{x_k\}_{k=0}^{\infty}$ and $\{y_k\}_{k=0}^{\infty}$ are generated by Algorithm 3.2. If the conditions of Lemma 3.3 hold, X^* is a nonempty compact set, and

$$\sum_{k=0}^{\infty} \lambda_k = +\infty, \tag{65}$$

then

$$\lim_{k\to\infty} \|x_{k+1} - y_k\| = 0.$$

Proof. From Corollary 2.3, we have

$$\begin{aligned} &f(x_k) - f(x) + (1/2\lambda_{k-1})\|x_k - y_{k-1}\|^2 \\ &\le \beta_k[f(x_0) - f(x) + (a/2)\|x - x_0\|^2]. \end{aligned} \tag{66}$$

From (65) and Corollary 2.2, we have that $\{x_k\}_{k=0}^{\infty}$ is a bounded sequence; since X^* is nonempty, $f^* > -\infty$, which implies that $\{f(x_k)\}_{k=0}^{\infty}$ is bounded from below. Set $x = x_k$ in (66); we have

$$\begin{aligned} &(1/2\lambda_{k-1})\|x_k - y_{k-1}\|^2 \\ &\le \beta_k[f(x_0) - f(x_k) + (a/2)\|x_k - x_0\|^2], \end{aligned} \tag{67}$$

which implies that $\{f(x_k)\}_{k=0}^{\infty}$ is bounded from above. The boundedness of $\{x_k\}_{k=0}^{\infty}$ and $\{|f(x_k)|\}_{k=0}^{\infty}$ with (67) yield the desired conclusion. □

Theorem 3.3. Suppose that the conditions of Lemma 3.4 hold. Then, $\{x_k\}_{k=0}^{\infty}$ is an asymptotically regular sequence.

Proof. Since

$$\|v_k - v_0\|^2 \le \left(\sum_{i=1}^{k} \|v_i - v_{i-1}\|\right)^2 \le k \sum_{i=1}^{k} \|v_i - v_{i-1}\|^2,$$

we have

$$\begin{aligned} &\|\alpha_k v_k - \alpha_k v_0\|^2 \\ &\le k\alpha_k^2 \sum_{i=1}^{k} \|v_i - v_{i-1}\|^2 \\ &= k\alpha_k^2 \sum_{i=1}^{k} \|x_i - y_{i-1}\|^2 \\ &\le 2 \sup\{f(x_0) - f(x_i) + (a/2)\|x_i - x_0\|^2, i = 0, \dots, k\} k\alpha_k^2 \sum_{i=1}^{k} \beta_i \lambda_{i-1}. \end{aligned}$$

Using (64), we have

$$\alpha_k = a_k\lambda_k/(1 + a_k\lambda_k) = a\beta_{k+1}\lambda_k \to 0.$$

This result and Lemma 3.3 yield that $\alpha_k\|v_k\| \to 0$. Hence,

$$\|y_k - x_k\| = \|-\alpha_k x_k + \alpha_k v_k\| \to 0.$$

This conclusion, the fact that

$$\|x_{k+1}-x_k\| \leq \|x_{k+1}-y_k\| + \|y_k - x_k\|,$$

and Lemma 3.4 yield $x_{k+1}-x_k \to 0$. □

From Theorem 3.3, Remark 14.1.1 in Ref. 39, and the boundedness of $\{x_k\}_{k=0}^{\infty}$, we have the following corollary.

Corollary 3.4. Suppose that the conditions of Lemma 3.4 hold. Then, either the accumulation set of $\{x_k\}_{k=0}^{\infty}$ is a singleton or it is a connected set.

Remark 3.2. From (66), we can deduce that the convergence rate obtained for Algorithm 3.2 is lower than the convergence rate obtained for Algorithm 3.1 with $c\in(0, 2]$, so we may hope that, for any $c\in(0, 2]$. Algorithm 3.1 with $f_k=f$ has also the properties that $\|x_{k+1}-y_k\|\to 0$ and $\|x_{k+1}-x_k\|\to 0$ if X^* is a nonempty compact set.

Remark 3.3. From (66) and (I) of Lemma 3.3, we obtain for Algorithm 3.2 the same global convergence rate estimate

$$f(x_k) - \min_{x\in\Re^n} f(x) = O\left(1\Big/\sum_{j=0}^{k-1} \lambda_j\right)$$

as obtained for (2) in Ref. 9. In Ref. 9, it was shown that the condition (65) is necessary and sufficient for the convergence of the classical proximal point algorithm (2), but we do not know whether (65) is still a necessary condition for convergence of Algorithm 3.2.

4. Proximal Bundle Method Algorithm

In this section, we give a variant proximal bundle method algorithm by combining (GPPA) with the bundle method. In iteration $k+1$, x_{k+1} is calculated by the formula

$$x_{k+1} = \text{argmin}\{f_{k+1} + (1/2\lambda_k)\|x-y_k\|^2 : x\in\Re^n\},$$

where f_{k+1} is a bundle linearization function of f. More precisely, for $k\geq 0$,

$$f_{k+1} = \max\{f_k(x), f(x_k) + g_k^T(x-x_k)\}, \tag{68}$$

where

$$f_0(x) = f(x_0) + g_0^T(x - x_0) \tag{69}$$

and $g_k \in \partial f(x_k)$.

Algorithm 4.1. Proximal Bundle Method Algorithm (PBMA).

Step 0. Initialization. Select an initial point $x_0 \in \text{dom } f$. Let $v_0 = x_0$, $a_0 = a > 0$, $f_0(x) = f(x_0) + g_0^T(x - x_0)$, and $k = 0$.

Step 1. For x_k, v_k, a_k, choose $\lambda_k > 0$ and $\alpha_k \in (0, 1)$ such that

$$\lambda_k = \alpha_k / a_k(1 - \alpha_k). \tag{70}$$

Set

$$y_k = (1 - \alpha_k)x_k + \alpha_k v_k.$$

Step 2. Compute $g_k \in \partial f(x_k)$. Generate f_{k+1} by the formula (68). Compute

$$x_{k+1} = \text{argmin}\{f_{k+1}(z) + (1/2\lambda_k)\|z - y_k\|^2 : z \in \Re^n\},$$

$$v_{k+1} = v_k + (x_{k+1} - y_k),$$

$$a_{k+1} = (1 - \alpha_k)a_k.$$

Step 3. Increase k by 1, and go to Step 1.

It is not hard to show that (PBMA) is a special case of (GPPA). From Corollary 2.1 and Lemma 2.4, we have the following property for (PBMA), but the global convergence of (PBMA) under some more reasonable conditions is not clear for us at this moment.

Proposition 4.1. Suppose that $\{x_k\}_{k=0}^{\infty}$ is generated by (PBMA). Suppose that, for all $k \geq 0$, we choose $\alpha_k = 1 - r^k$, where $r \in (0, 1)$ is a constant. Assume that:

(i) $\{f(x_k) - f_k(x_k)\}_{k=0}^{\infty}$ is bounded from above;
(ii) there is an index K, such that

$$\lim_{k \in K, k \to \infty} g_k^T(x_{k+1} - x_k) = 0.$$

Then, $\{x_k\}_{k \in K}$ is a minimizing sequence for f.

Proof. From the convexity of f and the construction of f_k, we have, for all $k \geq 0$, for all $x \in \Re^n$,

$$f_k(x) \leq f(x),$$

which implies that

$$f_{k+1}(x_k) = \max\{f_k(x_k), f(x_k)\} = f(x_k).$$

Hence,

$$|\delta_{k+1}^0| = |f_{k+1}(x_k) - f_k(x_k)| = f(x_k) - f_k(x_k).$$

From assumption (i), $\{|\delta_k^0|\}_{k=0}^{\infty}$ is bounded. From the definitions of α_k and λ_k, we can deduce that (35) and (36) hold by using Lemma 2.4. Using the construction of f_k once again, we have

$$f_{k+1}(x_{k+1}) \geq f(x_k) + g_k^T(x_{k+1} - x_k),$$

which combined with assumption (ii) yields that

$$\limsup_{k \in K, k \to \infty} f_{k+1}(x_{k+1}) \geq \limsup_{k \in K, k \to \infty} f(x_k).$$

Hence, the conclusion follows from Corollary 2.1. □

Remark 4.1. It is worth noting that, in Proposition 4.1, we must assume that α_k is close enough to one and λ_k is big enough. Furthermore, we assume conditions (i) and (ii). These are disadvantages of this method, but the method has one difference from the original bundle methods (see Refs. 15, 19, and 21 for details): the calculations of $\{x_k\}_{k=0}^{\infty}$ are based on (12), not based on (8). Since the convergence rates obtained up to now for the original proximal algorithm [$\{x_k\}_{k=0}^{\infty}$ generated by (2)] are lower than the convergence rate obtained for (GPPA) (see Refs. 9, 10, and Section 2 of this paper), we hope that (PBMA) has a higher convergence rate than the original bundle methods. Modifications of the convergence assumptions by using null-step techniques (Refs. 18–25) and numerical texts will be our further research topic.

Remark 4.2. It is possible to give another chocie for f_k. In fact, we can choose

$$f_0(x) = f(y_0) + \bar{g}_0^T(x - y_0),$$

$$f_{k+1} = \max\{f_k(x), f(y_k) + \bar{g}_k^T(x - y_k)\},$$

where $\bar{g}_k \in \partial f(y_k)$. From Corollary 2.1 and Lemma 2.4, we can give the same property as Proposition 4.1 for this method.

References

1. MOREAU, J. J., *Proximité et Dualité dans un Espace Hilbertien*, Bulletin de la Societé Mathematique de France, Vol. 93, pp. 273–299, 1965.

2. MARTINET, B., *Regularisation d'Inéquations Variationelles par Approximations Successives*, Revenue Francaise d'Informatique et de Recherche Operationelle, Vol. 4, pp. 154–159, 1970.
3. ROCKAFELLAR, R. T., *Monotone Operators and Proximal Point Algorithm*, SIAM Journal on Control and Optimization, Vol. 14, pp. 877–898, 1976.
4. BERTSEKAS, D. P., and TSENG, P., *Partial Proximal Minimization Algorithms for Convex Programming*, SIAM Journal on Optimization, Vol. 4, pp. 551–572, 1994.
5. CHEN, G., and TEBOULLE, M., *A Proximal-Based Decomposition Method for Convex Minimization Problems*, Mathematical Programming, Vol. 64, pp. 81–101, 1994.
6. ECKSTEIN, J., and BERTSEKAS, D. P., *On the Douglas–Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators*, Mathematical Programming, Vol. 55, pp. 293–318, 1992.
7. FUKUSHIMA, M., *A Descent Algorithm for Nonsmooth Convex Programming*, Mathematical Programming, Vol. 30, pp. 163–175, 1984.
8. FUKUSHIMA, M., and QI, L., *A Globally and Superlinearly Convergent Algorithm for Nonsmooth Convex Minimization*, SIAM Journal on Optimization, Vol. 6, pp. 463–472, 1996.
9. GÜLER, O., *On the Convergence of the Proximal Point Algorithm for Convex Minimization*, SIAM Journal on Control and Optimization, Vol. 29, pp. 403–419, 1991.
10. GÜLER, O., *New Proximal Point Algorithms for Convex Minimization*, SIAM Journal on Optimization, Vol. 4, pp. 649–664, 1994.
11. ROCKAFELLAR, R. T., *Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming*, Mathematics of Operations Research, Vol. 1, pp. 97–116. 1976.
12. ZHU, C., *Modified Proximal Point Algorithm for Extended Linear-Quadratic Programming*, Computational Optimization and Applications, Vol. 1, pp. 185–206, 1992.
13. NESTEROV, Y. E., *On an Approach to the Construction of the Optimal Methods of Minimization of Smooth Convex Functions*, Ekonomicheskie i Matematicheskie Metody, Vol. 24, pp. 509–517, 1988.
14. LEMARÉCHAL, C., *Nonsmooth Optimization and Descent Methods*, Research Report RR-78-4, International Institute of Applied Systems Analysis, Laxenburg, Austria, 1977.
15. KIWIEL, K. C., *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics, Springer, Berlin, Germany, Vol. 1133, 1985.
16. LEMARÉCHAL, C., *Nondifferentiable Optimization*, Handbooks in Operations Research and Management Science, Edited by G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd, North Holland, Amsterdam, Holland, Vol. 1, pp. 529–572, 1989.
17. HIRIART-URRUTY, J., and LEMARÉCHAL, C., *Convex Analysis and Minimization Algorithms, Vols. 1–2*, Springer Verlag, Berlin, Germany, 1993.
18. AUSLENDER, A., *Numerical Methods for Nondifferentiable Convex Optimization*, Mathematical Programming Study, Vol. 30, pp. 102–126, 1986.

19. CORREA, R., and LEMARÉCHAL, C., *Convergence of Some Algorithms for Convex Minimization*, Mathematical Programming, Vol. 62, pp. 261–275, 1993.
20. KIWIEL, K. C., *An Aggregate Subgradient Method for Nonsmooth Convex Minimization*, Mathematical Programming, Vol. 27, pp. 320–341, 1983.
21. KIWIEL, K. C., *Proximity Control in Bundle Methods for Convex Nondifferentiable Minimization*, Mathematical Programming, Vol. 46, pp. 105–122, 1990.
22. LEMARÉCHAL, C., *Constructing Bundle Methods for Convex Optimization*, Fermat Days 85: Mathematics for Optimization, Edited by J. B. Hiriart-Urruty, North Holland, Amsterdam, Holland, pp. 201–240, 1986.
23. LEMARÉCHAL, C., and MIFFLIN, R., *A Globally and Superlinearly Convergent Algorithm for One-Dimensional Minimization of Convex Functions*, Mathematical Programming, Vol. 24, pp. 241–256, 1982.
24. LEMARÉCHAL, C., and SAGASTIZÁBAL, C., *An Approach to Variable-Metric Methods*, Proceeding of the 16th IFIP Conference on System Modelling and Optimization, Springer Verlag, Berlin, Germany, pp. 144–162, 1994.
25. MIFFLIN, R., *A Quasi-Second-Order Proximal Bundle Algorithm*, Mathematical Programming, Vol. 73, pp. 51–72, 1996.
26. AU, K. T., HIGLE, J. L., and SEN, S., *Inexact Subgradient Methods with Applications in Stochastic Programming*, Mathematical Programming, Vol. 63, pp. 65–82, 1994.
27. BIRGE, J. R., and QI, L., *Subdifferential Convergence in Stochastic Programs*, SIAM Journal on Optimization, Vol. 5, pp. 436–453, 1995.
28. BIRGE, J. R., and WETS, R. J. B., *Designing Approximation Schemes for Stochastic Optimization Problems in Particular Stochastic Programs with Recourse*, Mathematical Programming Study, Vol. 27, pp. 54–102, 1986.
29. BIRGE, J. R., CHEN, X., QI, L., and WEI, Z., *A Stochastic Newton Method for Stochastic Quadratic Programs with Recourse*, Applied Mathematics Report AMR 94/9, University of New South Wales, 1994.
30. CHEN, X., QI, L., and WOMERSLEY, R. S., *Newton's Method for Quadratic Stochastic Programs with Recourse*, Journal of Computational and Applied Mathematics, Vol. 60, pp. 29–46, 1995.
31. HIGLE, J. L., and SEN, S., *Stochastic Decomposition: An Algorithm for Two-Stage Linear Programs with Recourse*, Mathematics of Operations Research, Vol. 3, pp. 650–669, 1991.
32. KALL, P., and WALLACE, S. T., *Stochastic Programming*, Wiley, New York, New York, 1994.
33. KING, A. J., and WETS, R. J. B., *Epi-Consistency of Convex Stochastic Programs*, Stochastic and Stochastic Reports, Vol. 34, pp. 83–92, 1990.
34. QI, L., and WOMERSLEY, R. S., *An SQP Algorithm for Extended Linear-Quadratic Problems in Stochastic Programming*, Annals of Operation Research, Vol. 56, pp. 251–285, 1995.
35. JOE, S., and SOAN, I. H., *Imbedded Lattice Rules for Multidimensional Integration*, SIAM Journal on Numerical Analysis, Vol. 29, pp. 1191–1135, 1992.
36. NIEDERREITER, H., *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, Pennsylvania, 1992.

37. SLOAN, I. H., and JOE, S., *Lattice Methods for Multiple Integration*, Clarendon Press, Oxford, 1994.
38. SPANIER, J., and MAIZE, E. H., *Quasi-Random Methods for Estimating Integrals Using Relatively Small Samples*, SIAM Review, Vol. 36, pp. 18–44, 1994.
39. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, New York, 1970.