Some Methods for Solving Nonsmooth Convex Minimization Problems

J.R. Birge
Department of Industrial and Operations Engineering
University of Michigan, Ann Arbor MI 48109-2117 USA

L. Qi and Z. Wei
University of New South Wales
Sydney NSW 2052 Australia

January 30, 1996

Revised
Technical Report 95-3

Some Methods for Solving Nonsmooth Convex Minimization Problems [1]

J. R. BIRGE
Department of Industrial and Operations Engineering
University of Michigan, Ann Arbor, MI 48109, USA

L. QI and Z. WEI
School of Mathematics,
University of New South Wales, Sydney NSW 2052, Australia

January 30, 1996

Abstract

In this paper, we propose and analyze a class of methods for minimizing a proper lower semicontinuous extended-valued convex function $f : R^n \to R\bigcup\{\infty\}$. Instead of the original objective function f, we employ a convex approximation f_{k+1} at the k-th iteration. Some basic global convergence rate estimates are obtained according to the properties of the approximation sequence $\{f_k\}_{k=0}^{\infty}$. We illustrate the applicability of our approach by proposing (i) a new family of proximal point algorithms which possesses the global convergence rate estimate $f(x_k) - \min_{x\in R^n} f(x) = O(1/(\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2)$ even if the iteration points are calculated approximately, where $\{\lambda_k\}_{k=0}^{\infty}$ are the proximal parameters, and (ii) a new bundle method which is globally convergent under some assumptions.

Key Words. Nonsmooth convex optimization, proximal point method, bundle algorithm, global convergence.

1. Introduction

Consider the following optimization problem

$$\min\{f(x) : x \in R^n\}, \tag{1.1}$$

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

The Moreau-Yosida approximation F_λ of f is defined by

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

where λ is a real positive number.

As proved by Moreau [?], F_λ is a differentiable convex function defined in the whole space of R^n that possesses the same set of minimizers as the problem in (1.1). Using these properties, Martinet [?] presented a proximal point algorithm for solving (1.1): start from an initial point $x_0 \in R^n$ and generate $\{x_k\}_{k=0}^{\infty}$ by solving

$$x_{k+1} = \text{argmin}\{f(x) + \frac{1}{2\lambda_k}\|x - x_k\|^2 : x \in R^n\}, \tag{1.2}$$

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

[1]This material is based on work supported by the National Science Foundation under Award Number DDM-9215921 and the Australian Research Council.

Rockafellar made a major contribution to proximal point methods in [?] by proving, under some additional reasonable assumptions, 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, which is an important consideration in practice. As an application, Rockafellar applied the results to a lower semicontinuous proper convex function f. In this case, the two general criteria for generating x_{k+1} are defined by

$$\mathrm{dist}(0, S_k(x_{k+1})) \leq \frac{\sigma_{1,k}}{\lambda_k}, \qquad \sum_{k=0}^{\infty} \sigma_{1,k} < \infty \tag{1.3}$$

and

$$\mathrm{dist}(0, S_k(x_{k+1})) \leq \frac{\sigma_{2,k}}{\lambda_k} \|x_{k+1} - x_k\|, \qquad \sum_{k=0}^{\infty} \sigma_{2,k} < \infty, \tag{1.4}$$

where

$$S_k(x) = \partial f(x) + \frac{1}{\lambda_k}(x - x_k). \tag{1.5}$$

For a survey of recent convergence results of the proximal point algorithm we refer the reader to [?, ?, ?, ?, ?, ?, ?, ?, ?].

Güler presented in [?] two different proximal point algorithms which used an idea introduced by Nesterov [?] for smooth convex minimization. The main difference from the classical proximal point algorithm in (1.3, 1.4, 1.5) is that the methods given in [?] generate an additional sequence $\{y_k\}_{k=0}^{\infty}$ of points in R^n, and calculate x_{k+1} from

$$x_{k+1} = \mathrm{argmin}\{f(x) + \frac{1}{2\lambda_k}\|x - y_k\|^2 : x \in R^n\}. \tag{1.6}$$

Güler also showed that the minimization in (1.6) can be performed inexactly by a modification of (1.3), i.e.,

$$\mathrm{dist}(0, \partial f(x_{k+1}) + \frac{1}{\lambda_k}(x_{k+1} - y_k)) \leq \frac{\sigma_{3,k}}{\lambda_k}, \tag{1.7}$$

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

Lemaréchal combined the proximal point method with the bundle method in his pioneering work [?], (also see [?], [?] and [?]). In Lemaréchal's algorithm, a sequence $\{x_k\}_{k=0}^{\infty}$ is generated by a sequence of convex functions, $\{f_k\}_{k=0}^{\infty}$. More precisely,

$$x_{k+1} = \mathrm{argmin}\{f_k(x) + \frac{1}{2\lambda_k}\|x - x_k\|^2 : x \in R^n\} \tag{1.8}$$

where f_k is a bundle linearization function of f. The power of the bundle methods has yielded many results (see [?, ?, ?, ?, ?, ?, ?, ?, ?]).

In this paper, we study procedures that permit the solution of (1.1) via the construction of a sequence of objective function approximations, $\{f_k\}_{k=0}^{\infty}$. Such approximations are necessary in many optimization problems (for example, stochastic programming, see [?, ?, ?, ?, ?, ?, ?, ?]) where the objective functions are too complex for exact evaluation. Within the context of stochastic programming, the objective function involves the expected value

$$f(x) = E[F(x, \omega)] = \int_{\Omega} F(x, \omega) \mathcal{P}(d\omega), \tag{1.9}$$

where ω is a random vector defined on a probability space $(\Omega, \mathcal{A}, \mathcal{P})$. Thus, the precise evaluation of f and its subgradients involves multidimensional integration. To avoid the computational burden associated with this evaluation, it is customary to replace the objective function, f, with a sequence of approximations, $\{f_k\}_{k=0}^{\infty}$, see [?, ?, ?, ?, ?, ?, ?, ?].

The remainder of the paper is organized as follows. In the next section, we describe the model algorithm and give some basic global convergence rate estimates. As the first important 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 the following form

$$\|u_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - y_k))\| \leq \sigma_{4,k}\|u_{k+1}\| + \frac{\sigma_{5,k}}{\lambda_k}\|x_{k+1} - y_k\|, \tag{1.10}$$

where $\sigma_{4,k} \in [0,1]$ and $\sigma_{5,k} \in [0,1)$. In particular, we obtain, under the same conditions, the following global convergence rate estimate obtained by Güler [?](with exact minimization (1.6)),

$$f(x_k) - \min_{x \in R^n} f(x) = O(\frac{1}{(\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2}). \tag{1.11}$$

Noting that (i) we obtain the same convergence results using inexact minimization (??) (for examples, for all k, $\sigma_{4,k} = 0$ and $\sigma_{5,k} \in [0, \sqrt{5} - 2]$ or $\sigma_{4,k} \in [0, \frac{1}{6}]$ and $\sigma_{5,k} = 0$) and (ii) the convergence rate (??) for algorithm (??) is higher than that obtained for (1.7) in [?]. Moreover, some applications of the results in stochastic programming are discussed in the same section. As another application of the results in Section 2, in Section 4, we present a new bundle method which calculate x_{k+1} from

$$x_{k+1} = \operatorname{argmin}\{f_{k+1}(x) + \frac{1}{2\lambda_k}\|x - y_k\|^2 : x \in R^n\}. \tag{1.12}$$

The proposed new bundle method may have a better convergence rate than the original bundle methods since the convergence rate (??) obtained for algorithm (1.6) is higher than

$$f(x_k) - \min_{x \in R^n} f(x) = O(\frac{1}{\sum_{j=0}^{k-1}\lambda_j}), \tag{1.13}$$

which is obtained for (1.2) in [?]. Under some additional assumptions, we establish the global convergence for the proposed bundle method.

2. The model of the methods

We begin our research with a brief description of the main idea given in the methods in [?] which is in turn suggested by Nesterov [?] for smooth convex minimization problems. The idea of the algorithms in [?] is to generate recursively a sequence $\{\varphi_k\}_{k=0}^{\infty}$ of simple convex quadratic functions that approximate f in such a way that at step $k \geq 0$, for all $x \in R^n$

$$\varphi_{k+1}(x) - f(x) \leq (1 - \alpha_k)(\varphi_k - f(x)), \tag{2.1}$$

where α_k is a number in the interval $[0, 1)$.

If (??) is satisfied for each $k \geq 0$, then

$$\varphi_k(x) - f(x) \leq (\prod_{i=0}^{k-1}(1 - \alpha_i))(\varphi_0(x) - f(x)). \tag{2.2}$$

If, at step k, we have at hand a point x_k such that

$$f(x_k) \leq \varphi_k^* := \min\{\varphi_k(z) : z \in R^n\}, \tag{2.3}$$

then we obtain from (??) that

$$f(x_k) - f(x) \le (\prod_{i=0}^{k-1}(1-\alpha_i))(\varphi_0(x) - f(x)), \tag{2.4}$$

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, $\{f_k\}$. Since f is too complex for exact evaluation in some applications, so in each iteration, we assume that there is a simple convex function f_k at hand, which is viewed as an approximation for f at step k. Denote $X = \{x : x \in R^n, f(x) < +\infty\}$. Our asymptotic analysis is based on the following assumption:

(A1): for any $x \in X$ and all $k \ge 0$, $f_k(x) < +\infty$.

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

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

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

The following two lemmas can be viewed as slight extensions of the related results in [?]. The proofs follow closely those of [?].

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

$$\varphi_{k+1}(x) - f_{k+1}(x) \le (1-\alpha_k)[\varphi_k(x) - f_k(x)] + (1-\alpha_k)\delta_{k+1}(x), \tag{2.5}$$

where

$$\delta_{k+1}(x) = f_k(x) - f_{k+1}(x) + f_{k+1}(x_k) - f_k(x_k).$$

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

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

Using the convexity of f_{k+1}, the conclusion (??) follows. 2

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

$$\delta_{k+1}(x) = f_k(x) - f_{k+1}(x) + f_{k+1}(x_k) - f_k(x_k), \quad k = 0,1,2,..., \tag{2.6}$$

and

$$\epsilon_{k+1}(x) = (1-\alpha_{k-1})\epsilon_k(x) + \delta_{k+1}(x), \quad k = 0,1,2,\dots\,. \tag{2.7}$$

For all $j : 1 \le j \le k$, let

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

It is easy to show that

$$\epsilon_{k+1}(x) = \delta_{k+1}(x) + \sum_{j=1}^{k} \Delta_{k,j}\delta_{k+1-j}(x). \tag{2.8}$$

¿From Lemma 2.1, we have

$$\varphi_{k+1}(x) - f_{k+1}(x) \le (\prod_{i=0}^{k}(1-\alpha_i))[\varphi_0(x) - f_0(x)] + (1-\alpha_k)\epsilon_{k+1}(x). \tag{2.9}$$

Letting

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

we have

$$a_{k+1} = (1-\alpha_k)a_k, \tag{2.10}$$

$$v_{k+1} = v_k - \frac{\alpha_k}{a_{k+1}}u_{k+1}, \tag{2.11}$$

where $a_0 = a$ and $v_0 = x_0$.

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

$$\varphi_{k+1}^* \ge f_{k+1}(x_{k+1}) + u_{k+1}^T[(1-\alpha_k)x_k + \alpha_k v_k - x_{k+1} - \frac{\alpha_k^2}{2a_{k+1}}u_{k+1}]. \tag{2.12}$$

Proof: From the definition of φ_{k+1}^*, (??) and (??), we have

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

by the assumption in the lemma.

Using the convexity of f_{k+1}, we have

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

So (??) follows. 2

Letting

$$y_k = (1-\alpha_k)x_k + \alpha_k v_k, \tag{2.13}$$

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

$$q_{k+1} = u_{k+1}^T[y_k - x_{k+1} - \frac{\alpha_k^2}{2a_{k+1}}u_{k+1}] \geq 0, \tag{2.14}$$

we have the following lemma:

Lemma 2.3 *Suppose that (??) and (??) hold. Then for $k = 0, 1, 2, ...,$*

$$\varphi_k^* \geq f_k(x_k). \tag{2.15}$$

Furthermore, for $k = 0, 1, 2, ...,$

$$\varphi_k^* \geq f_k(x_k) + q_k. \tag{2.16}$$

Proof: We prove (??) by induction. From the definitions of φ_0 and ϵ_0, (??) holds for $k = 0$. Suppose that (??) holds for k. From Lemma 2.2 and (??), we have that (??) holds for $k+1$.

Lemma 2.2 and (??) imply that (??) holds. 2

The Model Method (MM).

Step 0 (Initialization). Select an initial point $x_0 \in X$. 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 (A1). Then compute x_{k+1} with $u_{k+1} \in \partial f_{k+1}(x_{k+1})$ such that (??) holds, that is

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

Set

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

and

$$v_{k+1} = v_k - \frac{\alpha_k}{a_{k+1}}u_{k+1}.$$

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 $y_k \in R^n$, 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 $q_{k+1}^{MM} \geq 0$. In fact, since

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

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

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

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

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

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

¿From (??) and Lemma 2.3, we obtain the following basic convergence rate estimate.

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

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

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

Theorem 2.1 suggests many possibilities for obtaining convergence methods according to different (a) approximation sequences $\{f_k\}_{k=0}^{\infty}$, (b) solution methods for q_k^{MM}, and (c) choices of α_k. In the remainder of this section, we give some results based on the properties of $\{f_k\}_{k=0}^{\infty}$. We discuss (b) and (c) in the next section.

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

$$\delta_{k+1}^0 = f_{k+1}(x_k) - f_k(x_k), \quad k = 0, 1, 2, ..., \tag{2.17}$$

and

$$\epsilon_{k+1}^0 = (1 - \alpha_{k-1})\epsilon_k^0 + \delta_{k+1}^0, \quad k = 0, 1, 2, ... \tag{2.18}$$

Then

$$\epsilon_{k+1}^0 = \delta_{k+1}^0 + \sum_{j=1}^{k} \Delta_{k,j}\delta_{k+1-j}^0. \tag{2.19}$$

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

(A2) There is an index set K such that for all $x \in X$,

$$\limsup_{k \in K, k \to \infty} f_{k+1}(x) \leq f(x). \tag{2.20}$$

If

$$\lim_{k \in K, k \to \infty} \beta_{k+1} = 0, \tag{2.21}$$

$$\lim_{k \in K, k \to \infty} (1 - \alpha_k)\epsilon_{k+1}^0 = 0 \tag{2.22}$$

and

$$\limsup_{k \in K, k \to \infty} f_{k+1}(x_{k+1}) \geq \limsup_{k \in K, k \to \infty} f(x_k) \quad (\limsup_{k \in K, k \to \infty} f_{k+1}(x_{k+1}) \geq \limsup_{k \in K, k \to \infty} f(y_k)), \tag{2.23}$$

then $\{x_k\}_{k \in K}$ ($\{y_k\}_{k \in K}$) is a minimizing sequence for f.

Proof: From the assumption that $f_k(x) \leq f_{k+1}(x)$ we have, for $x \in X$ 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 $x \in X$,

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

which implies that $\{x_k\}_{k \in K}$ is a minimizing sequence for f. The same property can be obtained for $\{y_k\}_{k \in K}$. 2

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 (??) and (??) 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 (??) holds and* $\{|\epsilon_k^0|\}_{k=0}^{\infty}$ *is bounded.*

(II). If $\alpha_k \to 1$, *then (??) 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 $\Delta_{k,j}$ we have, for all $j : 1 \leq j \leq k$, that

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

This inequality combined with (??) yields

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

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

Since $0 < \beta_k = \Pi_{j=0}^{k-1}(1 - \alpha_j) \leq (1 - \bar{\alpha})^k$, (??) holds.

Since the condition in (II) implies the condition in (I), the conclusion of (II) follows from $\alpha_k \to 1$ and that $\{|\epsilon_k^0|\}_{k=0}^{\infty}$ is bounded. 2

Denote

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

$$X^* = \{x : x \in R^n, f(x) = f^*\},$$

and

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

Corollary 2.2 *Suppose that* $\{x_k\}_{k=0}^{\infty}$ *is generated by (MM) with* $f_k(x) \leq f_{k+1}(x)$ *for* $x \in X$ *and* $k \geq 0$; $f^* > -\infty$ *and* $f_0^* > -\infty$. *If* f_k *satisfies*

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

$$|f(x) - f_k(x)| \leq b_k, \tag{2.24}$$

then the following results (I), (II), and (III) hold.

(I). For any $x \in X$,

$$f(x_k) - f(x) + q_k^{MM} \leq \beta_k[\varphi_0(x) - f_0(x)] - (1 - \alpha_{k-1})\epsilon_k^0 + 2b_k. \tag{2.25}$$

In particular, we have the convergence rate estimate

$$f(x_k) - f^* + q_k^{MM} \leq \beta_k[f_0(x_0) - f_0^* + \frac{a}{2}\rho(x_0, X^*)^2] - (1 - \alpha_{k-1})\epsilon_k^0 + 2b_k. \tag{2.26}$$

(II). If

$$\lim_{k\to\infty} b_k = 0, \tag{2.27}$$

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

$$f(x_k) - f^* = O(kr^k). \tag{2.28}$$

(III). Suppose that (??), (??) and (??) hold. If X^* *is a nonempty compact subset in* R^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 the conclusions of (I) by Theorem 2.1. From (??), we have that $\{x_k\}_{k=0}^{\infty}$ is a minimizing sequence for f if (??), (??) and (??) hold. We prove (??) now. Suppose that $k \geq 1$. ¿From the definitions of $\delta_k^0, \Delta_{k,j}$ and (??), we have

$$\delta_k^0 \geq -(r^{k-1} + r^k)$$

and

$$\Delta_{k,j} = r^j.$$

The above two relations and (??) yield

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

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

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

which implies (??).

The conclusion (III) follows that $\{x_k\}_{k=0}^{\infty}$ is a minimizing sequence for f and that X^* is compact. 2

If for all $k \geq 0, f_k = f$, 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 (MM). If for any $k \geq 0$, $f_k = f$, then*

$$f(x_k) - f(x) + q_k^{MM} \leq \beta_k[f(x_0) - f(x) + \frac{a}{2}\|x - x_0\|^2]. \tag{2.29}$$

Consequently,

$$f(x_k) - f^* \leq \beta_k[f(x_0) - f^* + \frac{a}{2}\rho(x_0, X^*)^2]. \tag{2.30}$$

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

Remark 2.1 Noting that the definitions of ϵ_k^0 and $\epsilon_k^0(x)$, using the same way as Corollary 2.2, we can prove the similar results for Corollary 2.2 hold even if $f_k(x) \leq f_{k+1}(x)$ 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 two examples can be viewed as two general models arising from stochastic programming (see [?, ?, ?, ?, ?, ?]).

Example 1. Suppose f has the following structure

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

where $\mathcal{X} \subset R^n$ is a nonempty compact convex subset of R^n and $\Theta_{\mathcal{X}}$ is the indicator function of $\mathcal{X}$, i.e.,

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

We assume the following:

b1) For $j \geq 0$, $h_j : R^n \to R$ is a convex function.

b2) For $j \geq 1, h_j(x) \geq 0$ and there is a constant $M_0 > 0$, such that

$$\sup_{k\geq 1}\{|h_j(x)| : x \in \mathcal{X}\} \leq M_0.$$

b3) For $j \geq 1, p_j \geq 0$ and

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

Then, we can let

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

and solve the problem (1.1) by (MM). 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. Another type of these problems can be formulated as

$$\min\{f(x) = \int_{\Omega} F(x,y)dy + \Theta_{\mathcal{X}}(x) : x \in R^n\}, \tag{2.31}$$

where Ω is a compact subset of R^m and $F(.,.) : R^n \times R^m \to R$ is of bounded variation on Ω in the sense of Hardy and Krause; $F(.,y)$ is convex for any given $y \in \Omega$.

In this case, according to some integration rules (see [?, ?, ?, ?] for details) we can choose, for $j : 0 \leq j \leq k-1$, $\Omega_j^k \subset \Omega$ and $y_j^k \in \Omega_j^k$, such that

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

satisfies

c1) For all $k \geq 1$, $\Omega_i^k \bigcap_{i\neq j} \Omega_j^k = \emptyset$ and $\bigcup_{j=0}^{k-1} \Omega_j^k = \Omega$.

c2) $\lim_{k\to\infty} \sup\{|f(x) - f_k(x)| : x \in \mathcal{X}\} = 0$.

¿From Remark 2.1, we can solve (??) by (MM).

3. New proximal point algorithms

A new family of proximal point algorithms with four parameters ($\lambda_k, \alpha_k, \sigma_{4,k}$ and $\sigma_{5,k}$) is proposed in this section by devoting a new solution method for (??). In fact, Güler [?] gave one method (1.6) for solving (??). We use (??) here to solve (??). The method described below is based on one important property (Lemma 3.1) which claims that we can choose λ_k, $\sigma_{4,k}$ and $\sigma_{5,k}$ such that the solution set of (??) is contained in the solution set of (??).

Lemma 3.1 *Let u, v and w be vectors of R^n, $\lambda > 0, \tau \in [0,1]$ and $t \in [0,1)$. If*

$$\|u + \frac{1}{\lambda}(v-w)\| \leq \tau\|u\| + \frac{t}{\lambda}\|v-w\|, \tag{3.1}$$

then

$$u^T(w-v) \geq \frac{(1-\tau)^2(1-t)^2 - \tau(1+\tau)(1+t)^2}{(1-t)(1+t)^2}\lambda\|u\|^2. \tag{3.2}$$

Proof: The inequality

$$(u + \frac{1}{\lambda}(v-w))^T(v-w) \le \|u + \frac{1}{\lambda}(v-w)\|\|v-w\|$$

and (??) imply that

$$u^T(w-v) \ge \frac{1-t}{\lambda}\|v-w\|^2 - \tau\|u\|\|v-w\|. \tag{3.3}$$

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

$$\frac{1+\tau}{1-t}\lambda\|u\| \ge \|v-w\| \ge \frac{1-\tau}{1+t}\lambda\|u\|$$

which combining with (??) yields (??). 2

Let

$$\Psi(\tau;t) = \frac{(1-\tau)^2(1-t)^2 - \tau(1+\tau)(1+t)^2}{(1-t)(1+t)^2}.$$

¿From Lemma 3.1, we may now state our proximal point algorithm.

The General Proximal Point Algorithm (GPPA).
Initialization. Select an initial point $x_0 \in X$. Let $v_0 = x_0, a_0 = a > 0$, and $k = 0$.

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

$$\frac{\alpha_k^2}{2a_k(1-\alpha_k)} \le \lambda_k\Psi(\sigma_{4,k};\sigma_{5,k}), \tag{3.4}$$

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

$$\|u_{k+1} + \frac{1}{\lambda_k}(x_{k+1} - y_k)\| \le \sigma_{4,k}\|u_{k+1}\| + \frac{\sigma_{5,k}}{\lambda_k}\|x_{k+1} - y_k\|. \tag{3.5}$$

Set

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

and

$$v_{k+1} = v_k - \frac{\alpha_k}{a_{k+1}}u_{k+1}.$$

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

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

$$u_{k+1}^T(y_k - x_{k+1}) \ge \Psi(\sigma_{4,k};\sigma_{5,k})\lambda_k\|u_{k+1}\|^2. \tag{3.6}$$

Therefore, (x_{k+1}, u_{k+1}) is a solution for (??). Moreover,

$$q_{k+1}^{MM} \ge q_{k+1}^{GPPA} = \frac{\lambda_k}{2}[2\Psi(\sigma_{4,k};\sigma_{5,k}) - \frac{\alpha_k^2}{(1-\alpha_k)a_k\lambda_k}]\|u_{k+1}\|^2 \ge 0. \tag{3.7}$$

Proof: For any given k, let $u = u_{k+1}, v = x_{k+1}, w = y_k, \lambda = \lambda_k, \tau = \sigma_{4,k}$, and $t = \sigma_{5,k}$. ¿From Lemma 3.1, we have (??).

¿From the definition of q_k^{MM} and (??), if (??) holds, then $q_k^{MM} \geq 0$ so that (x_{k+1}, u_{k+1}) is a solution for (??). (??) follows the definitions of q_k^{MM} and q_k^{GPPA}. 2

By the above conclusions, we may state global convergence results similar to Corollary 2.1 – Corollary 2.3 but do not repeat these results here. We are more interested in finding $\sigma_{4,k}, \sigma_{5,k}$ and α_k such that β_k tends to 0 as fast as possible for any given sequence $\{\lambda_k\}_{k=0}^{\infty}$(see [?]). 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

$$c - \frac{\alpha_k^2}{(1-\alpha_k)a_k\lambda_k} = 0, \tag{3.8}$$

or

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

Therefore,

$$\alpha_k(c) = \frac{\sqrt{(ca_k\lambda_k)^2 + 4ca_k\lambda_k} - ca_k\lambda_k}{2}. \tag{3.9}$$

Similarly to the proof of Lemma 2.2 in [?], we can prove the following lemma.

Lemma 3.2

$$\frac{1}{(1+\sqrt{ca}\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2} \leq \beta_k(c) \leq \frac{1}{(1+(\sqrt{ca}/2)\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2}. \tag{3.10}$$

Let

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

Since for any $c \in (0,2], \{(\tau, 0) : \tau \in [0, \frac{2-c}{6}]\} \subset \Sigma(c), \Sigma(c) \neq \emptyset$. ¿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 (??). If $(\sigma_{4,k}, \sigma_{5,k}) \in \Sigma(c)$, then for any $x \in X$, (GPPA-(c)) has the global convergence rate estimate*

$$\begin{aligned} f(x_k) - f(x) + \frac{\lambda_{k-1}}{2}[2\Psi(\sigma_{4,k-1};\sigma_{5,k-1}) - c]\|u_k\|^2 &\leq \frac{f(x_0) - f(x) + (a/2)\|x - x_0\|^2}{(1+(\sqrt{ca}/2)\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2} \\ &\leq O(\frac{1}{(\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2}). \end{aligned} \tag{3.11}$$

Therefore

$$f(x_k) - f^* \leq \frac{4(f(x_0) - f^* + (a/2)\rho(x_0, X^*)^2)}{ca(\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2}. \tag{3.12}$$

The algorithm (GPPA-(c)) converges ($f(x_k) \to f^*$) if

$$\sum_{k=0}^{\infty}\sqrt{\lambda_k} = \infty. \tag{3.13}$$

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

$$\begin{aligned} f(x_k) - f^* &\leq \frac{4/(a\lambda)}{ck^2}(f(x_0) - f^* + (a/2)\rho(x_0, X^*)^2) \\ &= O(\frac{1}{k^2}). \end{aligned} \tag{3.14}$$

Since

$$\alpha'_k(c) = \frac{2a_k\lambda_k}{(2 + ca_k\lambda_k + \sqrt{(ca_k\lambda_k)^2 + 4ca_k\lambda_k})\sqrt{(ca_k\lambda_k)^2 + 4ca_k\lambda_k}} > 0,$$

$\alpha_k(c)$ is an increasing function about c.

On the other hand, since $\Psi(0,0) = 1$ and for all $\tau \in (0,1]$, $t \in (0,1)$,

$$\Psi(\tau;t) = \frac{(1-\tau)^2(1-t)}{(1+t)^2} - \frac{\tau(1+\tau)}{(1-t)} < 1.$$

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

$$\frac{1}{(1 + \sqrt{2a}\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2} \leq \beta_k(2) \leq \frac{1}{(1 + (\sqrt{2a}/2)\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2} \tag{3.15}$$

and the following result.

Corollary 3.1 *Suppose that for all k, $f_k = f$, $c = 2$ and α_k satisfies (??). If $\sigma_{4,k} = \sigma_{5,k} = 0$, then for any $x \in X$, (GPPA-(2)) has the global convergence rate estimate*

$$f(x_k) - f(x) \leq \frac{f(x_0) - f(x) + (a/2)\|x - x_0\|^2}{(1 + (\sqrt{2a}/2)\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2}. \tag{3.16}$$

Therefore,

$$f(x_k) - f^* \leq \frac{2(f(x_0) - f^* + (a/2)\rho(x_0, X^*)^2)}{a(\sum_{j=0}^{k-1}\sqrt{\lambda_j})^2}. \tag{3.17}$$

In [?], Güler selected $c = 1$ and gave the convergence rate results (??)-(??) with the calculation of x_{k+1} performed exactly by (1.6). Let

$$\Sigma_{E1} = \{(\tau, t) : \tau = 0 \text{ and } t \in [0, \sqrt{5} - 2]; \tau \in [0, \frac{1}{6}] \text{ and } t = 0\}.$$

Since $\Psi(0; \sqrt{5} - 2) = \Psi(\frac{1}{6}; 0) = \frac{1}{2}$, $\Psi(0,t) = \frac{1-t}{(1+t)^2}$ (also $\Psi(\tau;0) = 1 - 3\tau$) is decreasing for $t \in [0,1)$ ($\tau \in [0, \frac{1}{3})$), $\Sigma_{E1} \subset \Sigma(1)$. Therefore, we have

Corollary 3.2 *Suppose for all k, $f_k = f$, $c = 1$ and α_k satisfies (??). If $(\sigma_{4,k}, \sigma_{5,k}) \in \Sigma_{E1}$, then for any $x \in X$, (GPPA-E1) has the global convergence rate estimate (??), (??) and (??) for the case $c = 1$.*

¿From (??) and (??) with $c = 1$, we can deduce that the convergence rate obtained for (GPPA-(2)) is twice faster than that obtained for (GPPA-E1).

Remark 3.1 From Theorem 3.2 in this paper and Theorem 2.3 in [?], we obtained, for (GPPA-E1), the same convergence rates as those obtained in [?] for **The Proximal Point Algorithm** (PPA), but our algorithm (GPPA-E1) is more practical than (PPA) because (1) we calculate x_{k+1} by (??), and (2) the calculation x_{k+1} by (1.6) can be almost as difficult as the original minimization problem (1.1). On the other hand, the convergence rate obtained for **the algorithm with inexact minimization** in [?] is lower than the convergence rate obtained for (GPPA-E1) in this paper (see Theorem 3.2 in this paper and Theorem 3.3 in [?]); furthermore, we do not need that $\sigma_{4,k}$ and $\sigma_{5,k}$ tend to 0 (One cannot deduce that $\sigma_{4,k}\|u_{k+1}\| + \frac{\sigma_{5,k}}{\lambda_k}\|x_{k+1} - y_k\| = O(\frac{1}{k^\sigma})$ for some $\sigma > \frac{1}{2}$).

Remark 3.2 There are many possibilities for obtaining convergent proximal point algorithms. In fact, the two conditions to guarantee these possibilities are $q_k \geq 0$ and $\beta_k \to 0$. By viewing Corollary 2.1 of [?], we can deduce that (GPPA) has the same possibilities for guaranteed convergence as (PPA).

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 [?] and does not appear clear for this type algorithm. We only prove it in a special case of the choice for α_k.

Algorithm 1.
Step 0 (Initialization). Select an initial point $x_0 \in X$. 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

$$\alpha_k = \frac{a_k \lambda_k}{1 + a_k \lambda_k},$$

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

Step 2. Compute

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

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

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

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 1 is a special case of (GPPA). In fact, since $\alpha_k = \frac{a_k \lambda_k}{1 + a_k \lambda_k}$, we have

$$\lambda_k = \frac{\alpha_k}{a_k(1 - \alpha_k)} \geq \frac{\alpha_k^2}{2a_k(1 - \alpha_k)},$$

which implies (??) holds; on the other hand, in this algorithm, we can let $u_{k+1} = -\frac{1}{\lambda_k}(x_{k+1} - y_k)$, therefore,

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

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

$$-\frac{\alpha_k}{a_{k+1}} u_{k+1} = x_{k+1} - y_k,$$

which implies that

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

¿From above discussions, we can deduce that Algorithm 1 is a special case of (GPPA).

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 1.*

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

$$\beta_k = \frac{1}{1 + a\sum_{i=0}^{k-1} \lambda_i}.$$

(II) If $\{\frac{k\lambda_k}{\sum_{j=0}^{k-1}\lambda_j}\}_{k=1}^{\infty}$ *is bounded, then*

$$\lim_{k\to\infty} k\alpha_k^2(\sum_{i=1}^{k}\beta_i\lambda_{i-1}) = 0 \tag{3.18}$$

and

$$\lim_{k\to\infty}\beta_{k+1}\lambda_k = 0. \tag{3.19}$$

Proof : (I) For $i \geq 0$, since $1-\alpha_i = \frac{1}{1+a_i\lambda_i}$ and $a_{i+1} = (1-\alpha_i)a_i$, we have

$$1-\alpha_i = \frac{1+a\sum_{j=0}^{i-1}\lambda_j}{1+a\sum_{j=0}^{i}\lambda_j}.$$

These equations yield the desired conclusion (I).

(II) Set

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

Then

$$c_k = ka^2[\sum_{i=1}^{k}\frac{\lambda_{i-1}}{1+a\sum_{j=0}^{i-1}\lambda_j}][\frac{\lambda_k}{1+a\sum_{i=0}^{k}\lambda_i}]^2.$$

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

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

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

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

Conclusion (??) follows using $\frac{\sum_{i=1}^{k}\frac{1}{i}}{k} \to 0$.

Since

$$\beta_{k+1}\lambda_k = \frac{1}{k}\frac{k\lambda_k}{1+a(\lambda_0+\ldots+\lambda_k)} \leq \frac{1}{ak}\frac{k\lambda_k}{\lambda_0+\ldots+\lambda_{k-1}},$$

(??) follows the assumption. 2

Lemma 3.4 *Suppose that* $\{x_k\}_{k=0}^{\infty}$ *and* $\{y_k\}_{k=0}^{\infty}$ *are generated by Algorithm 1. If the conditions of Lemma 3.3 hold,* X^* *is a nonempty compact set and*

$$\sum_{k=0}^{\infty}\lambda_k = +\infty, \tag{3.20}$$

then

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

Proof : ¿From Corollary 2.3, we have

$$f(x_k) - f(x) + \frac{1}{2\lambda_{k-1}}\|x_k - y_{k-1}\|^2 \leq \beta_k[f(x_0) - f(x) + \frac{a}{2}\|x - x_0\|^2]. \tag{3.21}$$

¿From (??) 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 (??), we have

$$\frac{1}{2\lambda_{k-1}}\|x_k - y_{k-1}\|^2 \le \beta_k[f(x_0) - f(x_k) + \frac{a}{2}\|x_k - x_0\|^2], \tag{3.22}$$

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 (??) yield the desired conclusion. 2

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

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

we have

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

Using (??), we have $\alpha_k = \frac{a_k\lambda_k}{1+a_k\lambda_k} = a\beta_{k+1}\lambda_k \to 0$. 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, as well as the fact that $\|x_{k+1} - x_k\| \le \|x_{k+1} - y_k\| + \|y_k - x_k\|$ and Lemma 3.4, yields $x_{k+1} - x_k \to 0$. 2

¿From Theorem 3.3, Remark 14.1.1 in [?] and the boundedness of $\{x_k\}$, we have the following corollary.

Corollary 3.3 *Suppose that the conditions of Lemma 3.4 hold. Then either the accumulation set of $\{x_k\}_{k=0}^{\infty}$ is singleton or it is a connected set.*

Remark 3.3 From (??), we can deduce that the convergence rate obtained for Algorithm 1 is lower than the convergence rate obtained for (GPPA-(c)) ($c \in (0, 2]$), so we may hope that for any $c \in (0, 2]$, (GPPA-(c))(with $f_k = f$) also has 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 3.4 From (??) and (I) of Lemma 3.3, we obtained, for Algorithm 1, the same global convergence rate estimate ((??)) as obtained for (1.2) by [?]. In [?], it was shown that the condition (??) is necessary and sufficient for the convergence of the classical proximal point algorithm (1.2), but we do not know whether (??) is still a necessary condition for convergence of Algorithm 1.

4. A new bundle method

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

$$x_{k+1} = \text{argmin}\{f_{k+1} + \frac{1}{2\lambda_k}\|x - y_k\|^2 : x \in R^n\}$$

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

$$f_{k+1} = \max\{f_k(x), f(x_k) + g_k^T(x - x_k)\}, \tag{4.1}$$

where

$$f_0(x) = f(x_0) + g_0^T(x - x_0) \tag{4.2}$$

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

The Bundle Method(BM)
Step 0 (Initialization). Select an initial point $x_0 \in X$. 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

$$\lambda_k = \frac{\alpha_k}{a_k(1-\alpha_k)}. \tag{4.3}$$

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 (??). Compute

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

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

and

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

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

It is not hard to show that (BM) is a special case of (GPPA). ¿From Corollary 2.1 and Lemma 2.4, we have the following convergence result.

Theorem 4.1 *Suppose that $\{x_k\}_{k=0}^{\infty}$ is generated by (BM). Suppose that for all $k \geq 0$, we choose $\alpha_k = 1 - r^k$; where $r \in (0,1)$ is a constant. If*

(1) $\{f(x_k) - f_k(x_k)\}_{k=0}^{\infty}$ is bounded from above;

(2) there is an index K, such that $\{\|g_k\|\}_{k\in K}$ is a bounded subsequence and

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

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 $x \in R^n$,

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

which implies that

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

Hence

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

¿From the assumption (1), $\{|\delta_k^0|\}_{k=0}^{\infty}$ is bounded.

¿From the definitions of α_k and λ_k, we can deduce that (??) and (??) 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 the assumptions in (2) 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 Corollary 2.1. 2

Remark 4.1 It is worth noting that in the convergence result of (BM), we must assume that α_k is close enough to one and λ_k is big enough. Furthermore, we assume some weak conditions (1) and (2). These are disadvantages in this method, but the method has two differences from the original bundle methods (see [?, ?, ?] for details): (i) It has no null steps. This means that it is not always necessary to have inner iterations from the current point x_k to the next point x_{k+1}. (ii) The calculations of $\{x_k\}_{k=0}^{\infty}$ are based on (??), not based on (1.8). Since the convergence rates obtained up to now for the original proximal algorithm ($\{x_k\}_{k=0}^{\infty}$ generated by (1.2)) are lower than the convergence rate obtained for (GPPA) (see [?, ?] and Section 2 of this paper), we hope that (BM) has a higher convergence rate than the original bundle methods.

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

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

and

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

where $\bar{g}_k \in \partial f(y_k)$. From Corollary 2.1 and Lemma 2.4, we can give the convergence result of this method, which is very close to Theorem 4.1.

Acknowledgements. The authors are thankful to Claude Lemaréchal and two referees for their helpful comments. Indeed, Condition (II) of Lemma 3.3 is suggested by one referee, while Corollary 3.3 is suggested by another referee.

References

[1] A. Auslender, Numerical methods for nondifferentiable convex optimizations, *Mathematical Programming Study* 30 (1986) 102-126.

[2] K. T. Au, J. L. Higle and S. Sen, Inexact subgradient methods with applications in stochastic programming, *Mathematical Programming* 63 (1994) 65-82.

[3] D. P. Bertsekas and P. Tseng, Partial proximal minimization algorithms for convex programming, *SIAM Journal on Optimization* 4 (1994) 551-572.

[4] J. R. Birge and L. Qi, Subdifferential convergence in stochastic programs, *SIAM Journal on Optimization* 5 (1995) 436-453.

[5] J. R. Birge and R. J-B. Wets, Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse, *Mathematical Programming Study* 27 (1986) 54-102.

[6] J. R. Birge, X. Chen, L. Qi and Z. Wei, A stochastic Newton method for stochastic quadratic program with recourse, AMR 94/9, Applied Mathematics Report, University of New South Wales, 1994.

[7] J. R. Birge, S. M. Pollock and L. Qi, A quadratic recourse function for the two-stage stochastic program, Technical Report 95-13, Department of Industrial and Operations Engineering, The University of Michigan, 1995.

[8] G. Chen and M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Mathematical Programming* 64 (1994) 81-101.

[9] X. Chen, L. Qi and R. S. Womersley, Newton's method for quadratic stochastic programs with recourse, *Journal of Computational and Applied Mathematics* 60 (1995) 29-46.

[10] F. H. Clark, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[11] R. Correa and C. Lemaréchal, Convergence of some algorithms for convex minimization, *Mathematical Programming* 62 (1993) 261-275.

[12] J. Eckstein and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Mathematical Programming* 55 (1992) 293-318

[13] M. Fukushima, A descent algorithm for nonsmooth convex programming, *Mathematical Programming* 30 (1984) 163-175.

[14] M. Fukushima and L. Qi, A globally and superlinearly convergent algorithm for nonsmooth convex minimization, to appear in: *SIAM Journal on Optimization.*

[15] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM Journal on Control and Optimization* 29 (1991) 403-419.

[16] O. Güler, New proximal point algorithms for convex minimization, *SIAM Journal on Optimization* 4 (1992) 649-664.

[17] C. D. Ha, A generalization of the proximal point algorithm, *SIAM Journal on Control and Optimization* 28 (1990) 503-512.

[18] J. L. Higle and S. Sen, Stochastic decomposition: An algorithm for two-stage linear programs with recourse, *Mathematics of Operations Research* 3 (1991) 650-669.

[19] J. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms: I and II,* (Springer-Verlag, Berlin, 1993).

[20] S. Joe and I. H. Sloan, Imbedded lattice rules for multidimensional integration, *SIAM Journal on Numerical Analysis* 29 (1992) 1119-1135.

[21] P. Kall and S. T. Wallace, Stochastic Programming, Wiley, New York, 1994.

[22] A. J. King and R. J-B Wets, Epi-consistency of convex stochastic programs, *Stochastic and Stochastic Reports* 34 (1990) 83-92.

[23] K. C. Kiwiel, An aggregate subgradient method for nonsmooth convex minimization, *Mathematical Programming* 27 (1983) 320-341.

[24] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics 1133 (Springer, Berlin, 1985).

[25] K. C. Kiwiel, Proximity control in bundle methods for convex nondifferentiable minimization, *Mathematical Programming* 46 (1990) 105-122.

[26] C. Lemaréchal, Nonsmooth optimization and descent methods, Research Report Rr-78-4, International Institute of Applied Systems Analysis, Laxenburg, Austria, 1977.

[27] C. Lemaréchal, Constructing bundle methods for convex optimization, in: J. B. Hiriart-Urruty, ed., *Fermat Days 85: Mathematics for Optimization*, (North-Holland, Amsterdam, 1986) 201-240.

[28] C. Lemaréchal, Nondifferentiable optimization, in Handbooks in Operations Research and Management Science, Vol. 1, *Optimization*, G. L. Nemhauser, A. H. G. Rinnooy Kan and M. J. Todd, eds., (North-Holland, Amsterdam, 1989) 529-572.

[29] C. Lemaréchal and R. Mifflin, A globally and superlinearly convergent algorithm for one-dimensional minimization of convex functions, *Mathematical Programming* 24 (1982) 241-256.

[30] C. Lemaréchal and C. Sagastizábal, An approach to variable metric methods, in: *Proceeding of the 16th IFIP Conference on System Modelling and Optimization*, Springer-Verlag, Berlin, to appear.

[31] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, *Revue Francaise d'Informatique et de Recherche Operationelle* (1970) 154-159.

[32] R. Mifflin, A quasi-second-order proximal bundle algorithm, to appear in: *Mathematical Programming.*

[33] J. J. Moreau, Proximite et dualite dans un espace hilbertien, *Bulletin de la Societe Mathematique de France* 93 (1965) 273-299.

[34] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992.

[35] Y. E. Nesterov, On an approach to the construction of the optimal methods of minimization of smooth convex functions, *Ekonom. i Mat. Metody,* 24 (1988) 509-517.

[36] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970.

[37] L. Qi and R. S. Womersley, An SQP algorithm for extended linear-quadratic problems in stochastic programming, *Annals Operation Research* 56 (1995) 251-285.

[38] R. T. Rockafellar, Monotone operators and proximal point algorithm, *SIAM Journal on Control and Optimization* 14 (1976) 877-898.

[39] R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Mathematics of Operations Research* 1 (1976) 97-116.

[40] I. H. Sloan and S. Joe, Lattice Methods for Multiple Integration, to be published by Oxford University Press.

[41] J. Spanier and E. H. Maize, Quasi-random methods for estimating integrals using relatively small samples, *SIAM Review* 36 (1994), 18-44.

[42] C. Zhu, Modified proximal point algorithm for extended linear-quadratic programming, *Computational Optimization and Applications* 1 (1992) 185-206.