ASYMPTOTIC CONVERGENCE ANALYSIS OF SOME INEXACT PROXIMAL POINT ALGORITHMS FOR MINIMIZATION

Ciyou Zhu *

Mathematics and Computer Science Division, Argonne National Laboratory
9700 South Cass Avenue, Argonne, IL 60439

August, 1993; revised July, 1994

Abstract. In this paper, we prove that the inexact proximal point algorithm (PPA) in the form of Bonnans–Gilbert–Lemaréchal–Sagastizábal's "general algorithmic pattern" (GAP–1) converges linearly under mild conditions. Based on this essential result, we prove the linear convergence for the bundle method without requiring the differentiability of the objective function or the uniqueness of the solution. We also prove the linear convergence for Correa–Lemaréchal's "implementable form" of PPA and derive its rate of convergence.

As applications, we propose another variant (GAP–2) of inexact PPA, which shares the same convergence property as GAP–1 but makes more sense numerically. In the framework of GAP–2, we develop linearly convergent inexact PPA schemes with fixed number of inner-loop iterations, for solving hemiquadratic extended linear-quadratic programming problems, and for minimizing convex C^1 functions without requiring strong convexity.

Keywords. Proximal point algorithm, bundle method, linear convergence, extended linear-quadratic programming.

---

* This research was supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38 and the National Science Foundation under Contract ASC-9213149.

1. Introduction.

Consider the convex programming problem

$$(\mathcal{P}) \qquad \text{minimize } f(x) \text{ over all } x \in \mathbb{R}^n,$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a closed proper convex function [12]. As a *blanket assumption*, we assume that the solution set of $(\mathcal{P})$, denoted by $\bar{X}$, is nonempty. Let $\langle \cdot, \cdot \rangle_M$ be the inner product associated with a symmetric positive definite matrix $M \in \mathbb{R}^{n \times n}$

$$\langle x, y \rangle_M = x^T M y, \tag{1.1}$$

where "T" indicates the transpose, and let $|\cdot|_M$ be the norm induced by this inner product. A general iteration scheme for solving $(\mathcal{P})$ is the proximal point algorithm (PPA) [6].

Proximal Point Algorithm (Exact PPA). *Specify starting points* x^0. *For* $k = 0, 1, 2, \ldots,$ *generate the sequence* $\{x^k\}$ *as*

$$x^{k+1} := \operatorname*{argmin}_{x \in \mathbb{R}^n} \{ f(x) + \frac{1}{2} |x - x^k|_M^2 \}. \tag{1.2}$$

The PPA is especially useful as a "regularization" procedure for the problem where the objective function is not strongly convex. Some commonly used minimization algorithms may lose their convergence properties or even become undefined when f is not strongly convex. With the PPA as an "outer loop" of the iteration, however, these algorithms can still be used in solving the "inner loop" subproblems of calculating the argmin on the right-hand side of (1.2). The objective functions in these subproblems are strongly convex because of the augmented term $\frac{1}{2c}|x - x^k|_M^2$.

In a more general sense, the PPA can be formulated as a procedure for solving the multivalue equation $0 \in T_M(x)$ related to a *maximal monotone operator* (associated with the M-inner product) $T_M : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ [13]. The operator

$$J_M = (I + T_M)^{-1} \tag{1.3}$$

is single valued from all of $\mathbb{R}^n$ into $\mathbb{R}^n$ [8]. Hence the iteration

$$x^{k+1} := J_M(x^k) \tag{1.4}$$

is well defined. Moreover, $J_M(x) = x$ if and only if $0 \in T_M(x)$. The iteration (1.4) converges to a solution of $0 \in T_M$ [6, 7]. Let T_M be the subdifferential mapping $\partial_M f$ related to f in the M-inner product. Then problem $(\mathcal{P})$ is equivalent to the multivalue equation $0 \in T_M(x)$, and iteration (1.4) reduces precisely to (1.2). See [13] for detailed discussions.

In notation, we shall always use subscript "M" to indicate any concept associated with the inner product $\langle \cdot, \cdot \rangle_M$, such as $| \cdot |_M$, $\partial_M f$, $\mathbb{B}_M(z, \delta)$ (the closed ball centered at z with radius δ) and $\text{dist}_M(\cdot, \cdot)$ (the distance function) etc. When M is the identity matrix I, this M-inner product reduces to the ordinary Euclidean inner product, and the subscript will be dropped. To simplify the notations, the ordinary Euclidean inner product will underlie most parts of the paper unless otherwise specified.

One problem in the implementation of PPA is the calculation of argmin on the right-hand side of (1.2). Usually, it is also an iterative process. Since this subproblem has to be solved for each outer-loop iteration, it is computationally expensive to take x^{k+1} as the exact argmin in (1.2). In [13], Rockafellar proposed a more realistic version of the PPA, where an approximation, instead of the exact argmin in (1.2), is taken as the next iterate x^{k+1}.

Asymptotically Accurate Proximal Point Algorithm [13]**.** *Choose a positive sequence $\{\delta_k\}$ such that $\sum \delta_k < +\infty$. Specify starting points $x^0 \in X$. For $k = 0, 1, 2, \ldots$, generate the sequence $\{x^k\}$ in such a way that*

$$|x^{k+1} - \bar{x}^{k+1}| \leq \delta_k, \tag{1.5}$$

where

$$\bar{x}^{k+1} := \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \{ f(x) + \frac{1}{2} |x - x^k|^2 \}. \tag{1.6}$$

This iteration scheme converges to an optimal solution of $(\mathcal{P})$ [13]. If in addition, x^{k+1} satisfies

$$|x^{k+1} - \bar{x}^{k+1}| \leq \delta_k |x^{k+1} - x^k|, \tag{1.7}$$

then the convergence will be linear under the assumptions that the optimal solution set of $(\mathcal{P})$ is a singleton and that the inverse of the mapping $T = \partial \hat{f}$ is Lipschitz continuous in some neighborhood of 0. Other convergence results regarding the asymptotically accurate PPA may be found in Luque [5], where conditions similar to (1.7) are also imposed.

Unfortunately, the criterion (1.7) with $\sum \delta_k < +\infty$ will eventually drive

$$\frac{|x^{k+1} - \bar{x}^{k+1}|}{|x^{k+1} - x^k|} \to 0, \text{ as } k \to \infty.$$

Hence, the subproblem of finding the argmin in (1.6) will still have to be solved to high accuracy, thus posing potential difficulty for the inner loop. Intuitively it might not be wise to spend much computational effort in solving a single subproblem to very high accuracy while leaving the "base point" x^k of the inner-loop iteration unchanged.

For any fixed $z \in \mathbb{R}^n$, let $\tilde{f}_M(\cdot; z) : \mathbb{R}^n \to \mathbb{R}$ be the function defined as

$$\tilde{f}_M(x; z) = f(x) + \frac{1}{2}|x - z|_M^2. \tag{1.8}$$

Denote

$$\tilde{f}^*_{M,z} = \min_{x \in \mathbb{R}^n} \tilde{f}_M(x; z). \tag{1.9}$$

In [1], Bonnans–Gilbert–Lemaréchal–Sagastizábal proposed a "general algorithmic pattern" (GAP–1) for inexact PPA, where f is simply required to decrease by a fixed fraction of $f(z) - \tilde{f}^*_{M,z}$ in the inner iteration based at z. With the simplified notations

$$\tilde{f}_k(x) = \tilde{f}_{M_k}(x; x^k) \quad \text{and} \quad \tilde{f}^*_k = \min_{x \in \mathbb{R}^n} \tilde{f}_k(x) \tag{1.10}$$

for any given M_k and x^k, GAP–1 can be written in the following form.

General Algorithmic Pattern – 1 (GAP–1) [1]. *Choose some descent parameter $\eta \in (0, 1)$. Specify a starting point x^0 and a symmetric positive definite matrix M^0. For $k = 0, 1, 2, \ldots$, generate the sequence $\{x^k\}$ as follows: if $x^k \in \bar{X}$, then stop; otherwise*

Step 1. *compute x^{k+1} satisfying*

$$\frac{f(x^{k+1}) - \tilde{f}^*_k}{f(x^k) - \tilde{f}^*_k} \le \eta; \tag{1.11}$$

Step 2. *choose a symmetric positive definite matrix M_{k+1}.*

The convergence of GAP–1 was established in [1, Theorem 2.3]. Let $\lambda_{\max}(M_k)$ be the largest eigenvalue of M_k.

Theorem 1 (Convergence of GAP–1) [1, Theorem 2.3]**.** *Assume the solution set $\bar{X}$ of $(\mathcal{P})$ is nonempty and bounded. Then the sequence $\{x^k\}$ generated by GAP–1 is bounded. Moreover, if*

$$\sum_{k=1}^{\infty}\big(\lambda_{\max}(M_k)\big)^{-1} = \infty, \tag{1.12}$$

then any accumulation point of $\{x^k\}$ belongs to $\bar{X}$. The same properties hold also for the sequence $\{\bar{x}^k\}$ of the minimizer in (1.12).

However, there are two important question remain unanswered:

(i) How fast would the inexact PPA in the pattern of GAP–1 converge?

(ii) How much computational effort does it cost numerically to meet the inner-loop stopping criterion (1.11) of GAP–1?

In this paper, we are going to analyze the asymptotic rate of convergence for several inexact PPAs, where our results on GAP–1 will play a central role. Before proceeding further, we first propose another variant of inexact PPA, from which the answer to the second question will become clearer.

General Algorithmic Pattern – 2 (GAP–2). *Choose an inner-loop algorithm for minimizing the regularized objective function $\tilde{f}(\cdot\,; z)$ for fixed z, as well as a descent parameter $\eta \in (0,1)$. Specify a starting point x^0 and a symmetric positive definite matrix M^0. For $k = 0, 1, 2, \ldots$, generate the sequence $\{x^k\}$ as follows: if $x^k \in \bar{X}$, then stop; otherwise*

Step 1. *minimize $\tilde{f}(y; x^k)$ starting from $y^{k,0} = x^k$ to generate $y^{k,1}, y^{k,2}, \ldots$, until a point y^{k,m_k} satisfying*

$$\frac{\tilde{f}_k(y^{k,m_k}) - \tilde{f}_k^*}{\tilde{f}_k(y^{k,0}) - \tilde{f}_k^*} \le \eta \tag{1.13}$$

is obtained; take

$$x^{k+1} = y^{k,m_k}; \tag{1.14}$$

Step 2. *choose a symmetric positive definite matrix M_{k+1}.*

Now the stopping criterion (1.13) in GAP–2 is related to the values of the regularized functions $\tilde{f}_k$ only. Obviously, this criterion can be satisfied in a finite number of iterations if the inner-loop algorithm itself has at least a linear rate on $\tilde{f}_k$. Therefore people can make the decision as whether to terminate the inner-loop iteration by simply counting the number of iterations. On the other hand, by noting $\tilde{f}_k(y^{k,0}) = f(x^k)$ and $\tilde{f}_k(y^{k,m_k}) = \tilde{f}_k(x^{k+1}) \geq f(x^{k+1})$, it is easy to see that the inner-loop stopping criterion of GAP–2 implies (1.11) in GAP–1. Hence GAP–2 shares the same convergence property of GAP–1, as stated in Theorem 1.

A prominent feature of GAP–1 and GAP–2 is that no asymptotically stringent stopping criteria were imposed on the inner-loop iteration. It is clear from the stopping criterion (1.13) that the regularized subproblems need only to be solved to finite precision constantly. An even more surprising fact is that both GAP–1 and GAP–2 will converge linearly under mild local growth conditions on the inverse of the subgradient mapping ∂f. We will prove this result in Section 2 without any differentiability assumption or strong convexity assumption on the objective function f.

The rest of the paper is arranged as follows. As applications of the essential result in Section 2, we prove the linear convergence for the proximal form of bundle method in Section 3 and for Correa–Lemaréchal's "implementable form" of PPA in Section 4. In Section 5, we apply GAP–2 to solve the hemiquadratic extended linear-quadratic programming problems. In Section 6, we derive linearly convergent iteration schemes in the framework of GAP–2 for minimizing convex C^1 functions without requiring strong convexity. We also derive certain "suboptimal" choices for the proximal parameter c in these applications when the matrices M_k takes the form $M_k = \frac{1}{c}I$. For more sophisticated updating schemes of M_k, the reader is referred to [1], which contains a complete bibliography in this regard. The applications of the new results in this paper on those more sophisticated schemes will be discussed elsewhere.

2. Rate of Convergence for GAP–1 and GAP–2.

In this section, we investigate the rate of convergence for the sequence $\{f(x^k)\}$ generated by either GAP–1 or GAP–2. We show that they converge at least linearly with a rate related to the descent parameter η, as well as certain growth condition on the inverse of the subdifferential mapping ∂f near 0.

Keep the notations developed in Section 1. Let f^* be the optimal value of $(\mathcal{P})$

$$f^* = f(\bar{X}). \tag{2.1}$$

Denote

$$\bar{x}^{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \tilde{f}_k(x). \tag{2.2}$$

The following lemma gives a general result on the relationship between the outer-loop residue $f(x^k) - f^*$ and the inner-loop residue $\tilde{f}_k(x^k) - \tilde{f}_k^*$.

Lemma 2 (Relation between inner- and outer-loop residues)**.** *Let $\{x^k\}$ be a sequence in $\mathbb{R}^n \setminus \bar{X}$ such that $\operatorname{dist}(\bar{X}, x^k) \to 0$ as $k \to \infty$. Let $\{M_k\}$ be a sequence of symmetric positive definite matrices in $\mathbb{R}^{n\times n}$ with*

$$\lambda_{\max}(M_k) \le \Lambda_{\max} \tag{2.3}$$

for some $\Lambda_{\max} > 0$ and sufficiently large k. Suppose there exist constants $\alpha \ge 0$, $\delta > 0$ and $r \ge 1$ such that the inverse of the subdifferential mapping $T = \partial f$ satisfies the growth condition

$$\forall w \in \mathbb{B}(0,\delta), \quad \forall x \in T^{-1}(w), \qquad \operatorname{dist}(\bar{X}, x) \le \alpha |w|^r. \tag{2.4}$$

If the sequence $\{\lambda_{\min}(M_k)\}$ has a lower bound $\Lambda_{\min} > 0$, then

$$\tilde{f}_k(x^k) - \tilde{f}_k^* \le f(x^k) - f^* \le 2\big(1 + \alpha\Lambda_{\max}^r \operatorname{dist}^{r-1}(\bar{X}, x^k)\big)\big(\tilde{f}_k(x^k) - \tilde{f}_k^*\big) \tag{2.5}$$

for sufficiently large k. Hence for any $\varepsilon > 0$, there exists k' such that for all $k \ge k'$,

$$f(x^k) - f^* \le \begin{cases} 2(1+\alpha\Lambda_{\max})\big(\tilde{f}_k(x^k) - \tilde{f}_k^*\big) & \text{if } r = 1, \\ 2(1+\varepsilon)\big(\tilde{f}_k(x^k) - \tilde{f}_k^*\big) & \text{if } r > 1. \end{cases} \tag{2.6}$$

Proof. The first half of (2.5) follows directly from $\tilde{f}_k(x^k) = f(x^k)$ and $\tilde{f}_k^* \ge f^*$. Now we prove the second half of (2.5). Recall that $\bar{x}^{k+1} = J_{M_k}(x^k)$ and $\bar{x}^{k+1} \ne x^k$ unless $x^k \in \bar{X}$, where $J_{M_k} = (I + T_{M_k})^{-1}$ is the proximal mapping related to the operator $T_{M_k} = \partial_{M_k} f$. According to Güler [3, Lemma 2.2]

$$f(\bar{x}^{k+1}) - f(z) \le \frac{1}{2}|z - x^k|_{M_k}^2 - \frac{1}{2}|z - \bar{x}^{k+1}|_{M_k}^2 - \frac{1}{2}|x^k - \bar{x}^{k+1}|_{M_k}^2 \quad \forall z. \tag{2.7}$$

For any $\bar{x} \in \bar{X}$, $z = \bar{x}$ in (2.7) yields

$$\tilde{f}_k(\bar{x}^{k+1}) - f(\bar{x}) \le \frac{1}{2}|x^k - \bar{x}|_{M_k}^2 - \frac{1}{2}|\bar{x}^{k+1} - \bar{x}|_{M_k}^2. \tag{2.8}$$

On the other hand, $z = x^k$ in (2.7) yields

$$\tilde{f}_k(x^k) - \tilde{f}_k(\bar{x}^{k+1}) \geq \frac{1}{2}|x^k - \bar{x}^{k+1}|^2_{M_k}. \tag{2.9}$$

Combining (2.8) and (2.9), we get

$$\frac{\tilde{f}_k(\bar{x}^{k+1}) - f(\bar{x})}{\tilde{f}_k(x^k) - \tilde{f}_k(\bar{x}^{k+1})} \leq \frac{|x^k - \bar{x}|^2_{M_k} - |\bar{x}^{k+1} - \bar{x}|^2_{M_k}}{|x^k - \bar{x}^{k+1}|^2_{M_k}}. \tag{2.10}$$

But

$$\begin{aligned}\frac{|x^k - \bar{x}|^2_{M_k} - |\bar{x}^{k+1} - \bar{x}|^2_{M_k}}{|x^k - \bar{x}^{k+1}|^2_{M_k}} &= \frac{|x^k - \bar{x}^{k+1}|^2_{M_k} + 2\langle x^k - \bar{x}^{k+1}, \bar{x}^{k+1} - \bar{x}\rangle_{M_k}}{|x^k - \bar{x}^{k+1}|^2_{M_k}} \\ &\leq \frac{|x^k - \bar{x}^{k+1}|^2_{M_k} + 2|x^k - \bar{x}^{k+1}|_{M_k}|\bar{x}^{k+1} - \bar{x}|_{M_k}}{|x^k - \bar{x}^{k+1}|^2_{M_k}} \\ &= \frac{|x^k - \bar{x}^{k+1}|_{M_k} + 2|\bar{x}^{k+1} - \bar{x}|_{M_k}}{|x^k - \bar{x}^{k+1}|_{M_k}}. \end{aligned} \tag{2.11}$$

Substituting (2.11) in (2.10), we get

$$\frac{\tilde{f}_k(\bar{x}^{k+1}) - f(\bar{x})}{\tilde{f}_k(x^k) - \tilde{f}_k(\bar{x}^{k+1})} \leq \frac{|x^k - \bar{x}^{k+1}|_{M_k} + 2|\bar{x}^{k+1} - \bar{x}|_{M_k}}{|x^k - \bar{x}^{k+1}|_{M_k}}$$

for any $\bar{x} \in \bar{X}$. Hence

$$\frac{\tilde{f}_k(\bar{x}^{k+1}) - f^*}{\tilde{f}_k(x^k) - \tilde{f}_k(\bar{x}^{k+1})} \leq 1 + \frac{2\mathrm{dist}_{M_k}(\bar{X}, \bar{x}^{k+1})}{|x^k - \bar{x}^{k+1}|_{M_k}}. \tag{2.12}$$

Now, according to [13, Proposition 1],

$$|J_{M_k}(z) - J_{M_k}(z')|^2_{M_k} + |(I - J_{M_k})(z) - (I - J_{M_k})(z')|^2_{M_k} \leq |z - z'|_{M_k}, \tag{2.13}$$

$$(I - J_{M_k})(z) \in T_{M_k}\big(J_{M_k}(z)\big), \tag{2.14}$$

for all z and z'. Let $z = x^k$ and $z' = \bar{x}$ in (2.13) and (2.14). By (2.13),

$$|\bar{x}^{k+1} - \bar{x}|^2_{M_k} + |x^k - \bar{x}^{k+1}|^2_{M_k} \leq |x^k - \bar{x}|^2_{M_k} \quad \forall \bar{x} \in \bar{X}.$$

Hence

$$|x^k - \bar{x}^{k+1}|_{M_k} \leq \mathrm{dist}_{M_k}(\bar{X}, x^k) \leq \Lambda^{\frac{1}{2}}_{\max}\mathrm{dist}(\bar{X}, x^k). \tag{2.15}$$

Therefore

$$|x^k - \bar{x}^{k+1}| \le \Lambda_{\min}^{\frac{1}{2}} |x^k - \bar{x}^{k+1}|_{M_k} \to 0 \text{ as } \operatorname{dist}(\bar{X}, x^k) \to 0.$$

But $x^k - \bar{x}^{k+1} \in T_{M_k}(\bar{x}^{k+1})$ by (2.14). Then

$$f(z) \ge f(\bar{x}^{k+1}) + \langle x^{k+1} - \bar{x}^{k+1}, z - \bar{x}^{k+1} \rangle_{M_k} \quad \forall z,$$

or equivalently,

$$f(z) \ge f(\bar{x}^{k+1}) + \langle M^k(x^{k+1} - \bar{x}^{k+1}), z - \bar{x}^{k+1} \rangle \quad \forall z,$$

which yields $M^k(x^k - \bar{x}^{k+1}) \in T(\bar{x}^{k+1})$, or in other words,

$$\bar{x}^{k+1} \in T^{-1}\big(M^k(x^k - \bar{x}^{k+1})\big).$$

Hence it follows from (2.4) that

$$\Lambda_{\max}^{-\frac{1}{2}} \operatorname{dist}_{M_k}(\bar{X}, \bar{x}^{k+1}) \le \operatorname{dist}(\bar{X}, \bar{x}^{k+1}) \le \alpha |M^k(x^k - \bar{x}^{k+1})|^r \le \alpha \Lambda_{\max}^{\frac{r}{2}} |(x^k - \bar{x}^{k+1})|_{M^k}^r \tag{2.16}$$

for sufficiently large k. Combining (2.16) and (2.12), we obtain

$$\frac{\tilde{f}_k(\bar{x}^{k+1}) - f^*}{\tilde{f}_k(x^k) - \tilde{f}_k(\bar{x}^{k+1})} \le 1 + 2\alpha \Lambda_{\max}^{\frac{r+1}{2}} |x^k - \bar{x}^{k+1}|_{M_k}^{r-1},$$

from which the second half of (2.5) follows by (2.15). □

Lemma 2 may be viewed as a quantization of the relationship

$$J(x) = x \iff 0 \in T(x).$$

The result can also be used in the optimality test of the algorithm. If an estimation on $\tilde{f}_k(x^k) - \tilde{f}_k^*$ for the subproblem is available, then an estimation on $f(x^k) - f^*$ for the original problem can be obtained via (2.6).

The growth condition (2.4) is in fact a local upper Lipschitz condition with modulus α and order r on T^{-1}. It is a common requirement when convergence rates are under considerations. Both Rockafellar [13] and Luque [5] used similar conditions in their analysis of PPAs. Actually, the condition in [13] on the convergence of the asymptotically accurate PPA implies the growth condition with $r = 1$. See Zhu [19] for more discussion on the growth conditions.

Theorem 3 (Rate of convergence for GAP–1 and GAP–2)**.** *Assume that the solution set of $(\mathcal{P})$ is nonempty and bounded, and there exist constants $\alpha \geq 0$, $\delta > 0$, and $r \geq 1$ such that the inverse of the subdifferential mapping $T = \partial f$ satisfies the growth condition*

$$\forall w \in \mathbb{B}(0,\delta), \quad \forall x \in T^{-1}(w), \qquad \operatorname{dist}(\bar{X}, x) \leq \alpha |w|^r. \tag{2.17}$$

Let $\{x^k\} \subset \mathbb{R}^n \setminus \bar{X}$ be a sequence generated by GAP–1 (or GAP–2). If the eigenvalues of the symmetric positive definite matrices $\{M_k\}$ in the algorithm satisfies

$$\Lambda_{\max} \geq \lambda_{\max}(M_k) \geq \lambda_{\min}(M_k) \geq \Lambda_{\min} > 0 \tag{2.18}$$

with some $\Lambda_{\max} > \Lambda_{\min} > 0$ for sufficiently large k, then the sequence $\{x^k\}$ satisfies

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{1-\eta}{2\big(1 + \alpha \Lambda_{\max}^r \operatorname{dist}^{r-1}(\bar{X}, x^k)\big)} \tag{2.19}$$

for sufficiently large k, where $\eta \in (0,1)$ is the descent parameter in the stopping criterion of the algorithm. Hence the sequence $\{f(x^k)\}$ of the objective value converges to the optimal value of $(\mathcal{P})$ linearly in the sense that

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{1-\eta}{2(1 + \alpha/c^r)} \tag{2.20}$$

for all sufficiently large k, and

$$\limsup_{k\to\infty} \frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq \frac{1+\eta}{2} \quad \text{if } r > 1. \tag{2.21}$$

Proof. We need only to prove the theorem for GAP–1, since the stopping criterion (1.13) in GAP–2 implies the stopping criterion (1.11) in GAP–1. Observe that $\operatorname{dist}(\bar{X}, x^k) \to 0$ as $k \to \infty$ by Theorem 1. Now we prove the conclusions on the rate of convergence by using Lemma 2.

It follows from (1.11) that

$$f(x^k) - f(x^{k+1}) \geq (1-\eta)\big(\tilde{f}_k(x^k) - \tilde{f}_k^*\big) \tag{2.22}$$

for sufficiently large k. However, by the second half of (2.5),

$$f(x^k) - f^* \leq \big(2 + (2\alpha\Lambda_{\max}^r)\operatorname{dist}^{r-1}(\bar{X}, x^k)\big)\big(\tilde{f}_k(x^k) - \tilde{f}_k^*\big) \tag{2.23}$$

for sufficiently large k. Observe that $f(x^k) - f^* > 0$ for $x^k \notin \bar{X}$. Hence dividing (2.22) by (2.23), we have

$$\frac{f(x^k) - f(x^{k+1})}{f(x^k) - f^*} \geq \frac{1 - \eta}{2\big(1 + \alpha \Lambda_{\max}^r \operatorname{dist}^{r-1}(\bar{X}, x^k)\big)}$$

for sufficiently large k, which yields (2.19). The inequalities (2.20) and (2.21) follow from (2.19) immediately. □

According to Theorem 3, a linear rate of the inner-loop iteration on the "regularized" subproblem will induce a linear rate on the outer-loop iteration, provided the growth condition (2.17) is satisfied. Hence, algorithms possessing linear or superlinear rate of convergence on "regular" problems can be extended to solve the "singular" problems via the inexact PPA.

Theorem 3 also throws a light on the choice of the *proximal parameter* c when M_k in the algorithm takes the form $M_k = \frac{1}{c} I$. For $r = 1$, which is the common case corresponding to an upper Lipschitz continuous inverse of the subdifferential mapping ∂f in some neighborhood of 0, the rate given by (2.20) depends on the proximal parameter c, as well as on the inner-loop rate represented by η. Once the inner-loop algorithm is chosen, the relationship between η and c can be determined. Then the selection of c can be optimized by minimizing the right-hand side of (2.20). Two typical examples will be given in Sections 5 and 6.

3. Linear Convergence on the Proximal Form of Bundle Methods.

In this section, we assume the objective function f is finite valued. Correa–Lemaréchal's proximal form of bundle methods [2] for finding the minimum of f may be written as follows.

Proximal Form of Bundle Methods [2]. *Choose some tolerance $\mu \in (0, 1)$ and a positive sequence $\{c_k\}$. Specify a starting point x^0 and a convex function $\varphi^0 \leq f$. Set $k = j = 0$.*

Step 1. Solve for y

$$\min \big\{\varphi^j(y) + \frac{1}{2c_k}|y - x^k|^2\big\} \tag{3.1}$$

to obtain the unique optimal solution y^j, as well as $\gamma^j := (x^k - y^j)/c_k \in \partial \varphi^j(y^j)$.

Step 2. Compute $f(y^j)$. If

$$f(x^k) - f(y^j) \geq \mu[f(x^k) - \varphi^j(y^j)], \tag{3.2}$$

then take a *descent-step* by letting $x^{k+1} := y^j$ and $k := k+1$; otherwise take a *null-step* by doing nothing here.

Step 3. If a descent-step was made in Step 2, then choose a convex function $\varphi^{j+1} : \mathbb{R}^n \to \mathbb{R}$ satisfying $\varphi^{j+1} \leq f$; otherwise choose φ^{j+1} satisfying

$$\varphi^{j+1} \leq f, \tag{3.3}$$

$$\varphi^j(y^j) + \langle \gamma^j, \cdot - y^j \rangle \leq \varphi^{j+1}, \tag{3.4}$$

$$f(y^j) + \langle g^j, \cdot - y^j \rangle \leq \varphi^{j+1}, \tag{3.5}$$

where $g^j \in \partial f(y^j)$. Go to Step 1 with $j := j+1$.

Correa and Lemaréchal proved in [2, Theorem 4.4] that if c_k is bounded and $\sum c_k = +\infty$, then this iteration scheme

(i) either reaches some $x^k \in \bar{X}$ in finite iterations;

(ii) or generates $\{x^k\}$ converging to some $\bar{x} \in \bar{X}$.

Now we prove further that, if the proximal parameters c_k are bounded away from zero, then the algorithm converges at least linearly for the problems satisfying the growth conditions (2.17) with a nonempty and bounded solution set.

Proposition 4 (Rate of convergence on Correa–Lemaréchal's bundle methods). *Suppose the optimal solution set $\bar{X}$ is nonempty and bounded, and $x^k \notin \bar{X}$ in Correa–Lemaréchal's proximal form of bundle methods for all k. Assume that there exist constants $\alpha \geq 0$, $\delta > 0$, and $r \geq 1$ such that the inverse of the subdifferential mapping $\mathcal{T} = \partial f$ satisfies the growth condition (2.17) in Theorem 3. If there exist positive constants $c_{\max}$ and $c_{\min}$ such that*

$$c_{\max} \geq c_k \geq c_{\min} > 0 \tag{3.6}$$

for sufficiently large k, then the sequence $\{x^k\}$ generated by Correa–Lemaréchal's proximal form of bundle method satisfies

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{\mu}{2\big(1 + (\alpha/c_{\min}^r)\mathrm{dist}^{r-1}(\bar{X}, x^k)\big)} \tag{3.7}$$

for all sufficiently large k. Hence the sequence $\{f(x^k)\}$ of the objective value converges to the optimal value f^* linearly

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{\mu}{2(1 + \alpha/c_{\min}^r)} \tag{3.8}$$

for all sufficiently large k, and

$$\limsup_{k\to\infty} \frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{\mu}{2} \quad \textit{if} \quad r > 1. \tag{3.9}$$

Proof. According to [1, Proposition 2.2] condition (3.2) implies (1.11). Hence the iteration falls in the category of GAP–1 with $M_k = I/c_k$, and the conclusion follows immediately from Theorem 3. $\square$

4. Linear Convergence on Correa–Lemaréchal's Implementable PPA.

In this section, We prove the linear convergence of Correa–Lemaréchal's "implementable form" of PPA [2, Algorithm 3.3]. Let $\partial_\varepsilon f$ denote the ε-subdifferential of f. Recall that

$$g \in \partial_\varepsilon f(x) \iff f(z) \geq f(x) + \langle z - x, g\rangle - \varepsilon \quad \forall\, z.$$

With the notations in this paper, Correa–Lemaréchal's implementable form of PPA for minimizing $f(x)$ goes as follows.

Correa–Lemaréchal's implementable form of PPA [2]**.** *Choose a positive sequence $\{c_k\}$ and an algorithm to generate a minimizing sequence $\{w^j\}$ of $\tilde{f}(\cdot\,; z)$ for fixed z. Choose $\kappa > 1$ and $\mu \in (0, 1)$. Start from x^0, set $k = 0$.*

Step 1. *Set $j = 0$; start from some $w^{k,j} = w^{k,0}$.*

Step 2. *Set*

$$\varepsilon(x^k, w^{k,j}) = \kappa\big[f(x^k) - f(w^{k,j}) - \frac{\mu}{c_k}|w^{k,j} - x^k|^2\big]. \tag{4.1}$$

If

$$\frac{x^k - w^{k,j}}{c_k} \in \partial_{\varepsilon(x^k, w^{k,j})} f(x^k), \tag{4.2}$$

then go to Step 3; otherwise compute $w^{k,j+1}$, increase j by 1, and execute Step 2 again.

Step 3. *Set $x^{k+1} = w^{k,j}$, increase k by 1, and loop to Step 1.*

Correa and Lemaréchal pointed out in [2] that (4.1) and (4.2) together imply

$$\varepsilon(x^k, w^{k,j}) = \kappa\big[f(x^k) - f(w^{k,j}) - \frac{\mu}{c_k}|w^{k,j} - x^k|^2\big] \geq 0. \tag{4.3}$$

They also proved that if f is strongly coercive, that is,

$$f(z)/|z| \to +\infty \text{ when } |z| \to \infty, \tag{4.4}$$

then

(i) for each k, Step 2 eventually exit to Step 3, unless x^k is already optimal [2, Theorem 3.2];

(ii) $\liminf_{k\to\infty} f(x^k) = f^*$ [2, Proposition 3.2].

Here we prove further that, if the proximal parameters c_k are bounded above and bounded away from zero, then the algorithm converges at least linearly for the problems satisfying the growth conditions (2.17) with a nonempty and bounded solution set.

Proposition 5 (Rate of convergence on Correa–Lemaréchal's form of PPA)**.** *Suppose the optimal solution set $\bar{X}$ is nonempty and bounded, and $x^k \notin \bar{X}$ in Correa–Lemaréchal's implementable PPA for all k. Assume that f is strongly coercive in the sense of (4.4), and there exist constants $\alpha \geq 0$, $\delta > 0$, and $r \geq 1$ such that the inverse of the subdifferential mapping $\mathcal{T} = \partial f$ satisfies the growth condition (2.17) in Theorem 3. If there exist positive constants $c_{\max}$ and $c_{\min}$ such that*

$$c_{\max} \geq c_k \geq c_{\min} > 0 \tag{4.5}$$

for sufficiently large k, then the sequence $\{x^k\}$ generated by Correa–Lemaréchal's implementable PPA satisfies

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{1}{(2\kappa + 1/\mu)\big(1 + (\alpha/c_{\min}^r)\mathrm{dist}^{r-1}(\bar{X}, x^k)\big)} \tag{4.6}$$

for sufficiently large k. Hence the sequence $\{f(x^k)\}$ of the objective value converges to the optimal value f^* linearly

$$\frac{f(x^{k+1})-f^*}{f(x^k)-f^*} \leq 1-\frac{1}{(2\kappa+1/\mu)(1+\alpha/c_{\min}^r)} \tag{4.7}$$

for all sufficiently large k, and

$$\limsup_{k\to\infty}\frac{f(x^{k+1})-f^*}{f(x^k)-f^*} \leq 1-\frac{1}{2\kappa+1/\mu} \quad \textit{if} \quad r>1. \tag{4.8}$$

Proof. We prove the proposition by showing that the iteration scheme falls in the category of GAP–1 with

$$\eta = 1-\frac{1}{\kappa+1/(2\mu)} \qquad \text{and} \qquad M_k = \frac{1}{c_k}I. \tag{4.9}$$

Then the conclusions follow directly from Theorem 3.

Observe that (4.2) in the algorithm with $\varepsilon(x^k, w^j)$ given by (4.1) yields

$$\begin{aligned} f(z) \geq & f(x^k) + \frac{1}{c_k}\langle x^k - x^{k+1}, z - x^k\rangle - \varepsilon(x^k, x^{k+1}) \quad \forall\, z \\ = & f(x^k) + \frac{1}{c_k}\langle x^k - x^{k+1}, z - x^k\rangle - \kappa\big[f(x^k) - f(x^{k+1}) - \frac{\mu}{c_k}|x^{k+1}-x^k|^2\big] \quad \forall\, z, \end{aligned}$$

or equivalently,

$$\begin{aligned} \tilde{f}_k(z) \geq f(x^k) + \frac{1}{2}|z - x^k|_{M_k} + \langle x^k - x^{k+1}, z - x^k\rangle_{M_k} \\ - \kappa\big[f(x^k) - f(x^{k+1}) - \mu|x^{k+1}-x^k|^2_{M_k}\big] \quad \forall\, z, \end{aligned}$$

where $M_k = I/c_k$ and $\tilde{f}_k(z) = \tilde{f}_{M_k}(z; x^k)$. Let $z = \bar{x}^{k+1} = \operatorname*{argmin}_{x\in\mathbb{R}^n} \tilde{f}_k(x)$. We have

$$\begin{aligned} \tilde{f}_k(\bar{x}^{k+1}) \geq f(x^k) + \frac{1}{2}|\bar{x}^{k+1} - x^k|_{M_k} + \langle x^k - x^{k+1}, \bar{x}^{k+1} - x^k\rangle_{M_k} \\ - \kappa\big[f(x^k) - f(x^{k+1}) - \mu|x^{k+1}-x^k|^2_{M_k}\big]. \end{aligned}$$

Therefore, with the notation $\tilde{f}_k^* = \min_{x\in\mathbb{R}^n} \tilde{f}_k(x)$,

$$\begin{aligned} f(x^k) - \tilde{f}_k^* \leq \kappa\big[f(x^k) - f(x^{k+1}) - \mu|x^{k+1}-x^k|^2_{M_k}\big] \\ - \frac{1}{2}|\bar{x}^{k+1} - x^k|_{M_k} - \langle x^k - x^{k+1}, \bar{x}^{k+1} - x^k\rangle_{M_k}. \end{aligned} \tag{4.10}$$

However

$$|\bar{x}^{k+1} - x^k|^2_{M_k} + 2\langle x^k - x^{k+1}, \bar{x}^{k+1} - x^k\rangle_{M_k} = -|x^k - x^{k+1}|^2_{M_k} + |\bar{x}^{k+1} - x^{k+1}|^2_{M_k}. \tag{4.11}$$

Substituting (4.11) in (4.10), we obtain

$$\begin{aligned} f(x^k) - \tilde{f}_k^* \le \kappa\big[f(x^k) - f_k(x^{k+1}) - \mu|x^{k+1} - x^k|^2_{M_k}\big] \\ + \frac{1}{2}|x^{k+1} - x^k|^2_{M_k} - \frac{1}{2}|\bar{x}^{k+1} - x^{k+1}|^2_{M_k}. \end{aligned}$$

Hence, for $\mu \in (0,1)$, there holds

$$f(x^k) - \tilde{f}_k^* \le \kappa\big[f(x^k) - f(x^{k+1})\big] + \frac{1}{2}|x^{k+1} - x^k|^2_{M_k}. \tag{4.12}$$

On the other hand,

$$f(x^k) - f(x^{k+1}) - \mu|x^{k+1} - x^k|^2_{M_k} \ge 0$$

by (4.3). Hence

$$\frac{1}{\mu}\big(f(x^k) - f_k(x^{k+1})\big) \ge |x^{k+1} - x^k|^2_{M_k}. \tag{4.13}$$

Substituting (4.13) in (4.12), we obtain

$$f(x^k) - \tilde{f}_k^* \le \big(\kappa + \frac{1}{2\mu}\big)\big[f(x^k) - f(x^{k+1})\big].$$

Then

$$\frac{f(x^k) - \tilde{f}_k^*}{f(x^k) - f(x^{k+1})} \le \kappa + \frac{1}{2\mu},$$

or equivalently,

$$\frac{f(x^{k+1}) - \tilde{f}_k^*}{f(x^k) - \tilde{f}_k^*} \le 1 - \frac{1}{\kappa + 1/(2\mu)}.$$

Therefore condition (1.11) in GAP–1 holds with η given by (4.9). $\square$

5. Application in Hemiquadratic Extended Linear-Quadratic Problems. Let

$$L(u,v) = p{\cdot}u + \tfrac{1}{2}u{\cdot}Pu + q{\cdot}v - \tfrac{1}{2}v{\cdot}Qv - v{\cdot}Ru, \tag{5.1}$$

where $p \in \mathbb{R}^n$, $q \in \mathbb{R}^m$, and $R \in \mathbb{R}^{m\times n}$. The matrices $P \in \mathbb{R}^{n\times n}$ and $Q \in \mathbb{R}^{m\times m}$ are symmetric and positive semidefinite. (In this section, the Euclidean inner product is indicated simply by "$\cdot$" for convenience.) The associated primal and dual extended linear-quadratic problems (ELQPs) are

$$(\mathcal{P}_L) \qquad \text{minimize } f(u) \text{ over all } u \in U, \text{ where } f(u) := \sup_{v\in V} L(u,v),$$

$$(\mathcal{Q}_L) \qquad \text{maximize } g(v) \text{ over all } v \in V, \text{ where } g(v) := \inf_{u\in U} L(u,v),$$

where U and V are nonempty polyhedral (convex) sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. Problems $(\mathcal{P}_L)$ and $(\mathcal{Q}_L)$ have a common optimal value if either has a finite optimal value; in this case, the problems are equivalent to the saddle point problem of the Lagrangian L on $U \times V$ [16, Theorem 1]. The objective functions f and g in these problems are piecewise linear-quadratic and may have discontinuities in the first- or second-order derivatives in general.

The primal-dual steepest descent algorithm is a newly developed method for large-scale ELQPs with special structure [20, 17]. It has a linear rate on *fully quadratic* problems, that is, problems with positive definite P and Q matrices [17]. Now with the aid of the inexact PPA, we can extend this algorithm to solve problems where one of P and Q is singular (*hemiquadratic* case). The resulting iteration scheme will retain a linear rate of convergence.

Suppose that in (5.1) the matrix P is singular, while the matrix Q is positive definite. Then f is not necessarily strongly convex, and may even have discontinuities in its second-order derivatives. However, we can apply GAP–2 on the primal problem $(\mathcal{P}_L)$. (The variable u corresponds to the variable x in Sections 1 and 2, and the set $\bar{U}$ correspond to the set $\bar{X}$, respectively.) The objective function of the inner-loop subproblem is

$$\tilde{f}(u;z) = f(u) + \frac{1}{2c}|u-z|^2 = \sup_{v\in V} \tilde{L}(u,v;z),$$

where $\tilde{L}$ is the fully quadratic augmented Lagrangian

$$\begin{aligned}\tilde{L}(u,v;z) &= L(u,v) + \frac{1}{2c}|u-z|^2 \\ &= (p - \frac{1}{c}z)\cdot u + \frac{1}{2}u\cdot(P + \frac{1}{c}I)u + q\cdot v - \frac{1}{2}v\cdot Qv - v\cdot Ru + \frac{1}{2c}|z|^2.\end{aligned} \tag{5.2}$$

Let $F(\cdot) : U \to V$ and $G(\cdot;\cdot) : V \times U \to U$ be the mappings

$$F(u) = \operatorname*{argmax}_{v\in V} \tilde{L}(u,v;z) = \operatorname*{argmax}_{v\in V}\{(q - Ru)\cdot v - \tfrac{1}{2}v\cdot Qv\}, \tag{5.3}$$

$$G(v;z) = \operatorname*{argmin}_{u\in U} \tilde{L}(u,v;z) = \operatorname*{argmin}_{u\in U}\{(p - \frac{1}{c}z - R^T v)\cdot u + \tfrac{1}{2}u\cdot(P + \frac{1}{c}I)u\}, \tag{5.4}$$

respectively. For fixed z, define the mapping $\tilde{D}(\cdot;z) : U \to U$ as

$$\tilde{D}(u;z) = G\big(F(u);z\big). \tag{5.5}$$

The primal part of the algorithm [17], applied to the subproblem of minimizing $\tilde{f}(u;z)$ on U, goes as follows. Specify some starting point $u^0 \in U$. At u^k, generate u^{k+1} by

$$u^{k+1} := (1-\lambda_k)u^k + \lambda_k \tilde{D}(u^k;z), \tag{5.6}$$

where there are three possible rules to determine the step length λ_k [17]:

(i) Perfect line search

$$\lambda_k := \operatorname*{argmin}_{\lambda\in[0,1]} \tilde{f}\big((1-\lambda)u^k + \lambda\tilde{D}(u^k;z);z\big);$$

(ii) Fixed step lengths

$$\lambda_k := \min\{1, \frac{1}{2\gamma^2}\}, \text{ where } \gamma = |Q^{-\frac{1}{2}}R(P + \frac{1}{c}I)^{-\frac{1}{2}}|; \tag{5.7}$$

(iii) Adaptive step lengths

$$\begin{aligned}\lambda_k := \max\Big\{\, \theta^j \;\Big|\; &\tilde{f}\big((1-\theta^j)u^k + \theta^j\tilde{D}(u^k;z)\big) - \tilde{f}(u^k;z) \\ &\le \big(\tilde{f}(u^k;z) - \tilde{g}(F(u^k);z)\big)(-\tfrac{1}{2}\theta^j),\ j\in\{0,1,2,\ldots\}\,\Big\},\end{aligned}$$

with $\theta \in (0,1)$ and $\tilde{g}(v;z) = \min_{u\in U}\tilde{L}(u,v;z)$. The corresponding single step rates are [17, Theorems 3.1 and 3.2]

$$\frac{\tilde{f}(u^{k+1};z) - \tilde{f}_z^*}{\tilde{f}(u^k;z) - \tilde{f}_z^*} = \tilde{\eta}_{(i),(ii)} := \begin{cases}\gamma^2 & \text{if } 0 \le \gamma^2 < \frac{1}{2}, \\ 1 - \frac{1}{4\gamma^2} & \text{if } \gamma^2 \ge \frac{1}{2},\end{cases} \tag{5.8}$$

for rules (i) and (ii) and

$$\frac{\tilde{f}(u^{k+1};z)-\tilde{f}_z^*}{\tilde{f}(u^k;z)-\tilde{f}_z^*}=\tilde{\eta}_{(iii)}:=\begin{cases}\frac{\theta}{2} & \text{if } 0\le\gamma^2<\frac{1}{2},\\ 1-\frac{\theta}{4\gamma^2} & \text{if } \gamma^2\ge\frac{1}{2},\end{cases} \tag{5.9}$$

for rule (iii), respectively. An iteration scheme based on GAP–2 with $M_k = I/c$ goes as follows.

Linearly Convergent Inexact PPA for Hemiquadratic ELQP. *Choose an integer $m \geq 1$. Specify the starting point $u^0 \in U$. For $k = 0, 1, 2, \ldots$, generate the sequence $\{u^k\}$ as follows: if $u^k \in \bar{U}$, then stop; otherwise*

(i) minimize $\tilde{f}(w; u^k)$ over U starting from $w^{k,0} = u^k$ to generate $w^{k,1}, w^{k,2}, \ldots, w^{k,m}$ by using the iteration (5.6);

(ii) take $u^{k+1} = w^{k,m}$.

To verify that the growth condition in Theorem 3 is satisfied, observe that the subdifferentials ∂f is a polyhedral multifunction in the sense of Robinson [10] (cf. [14, p. 787]). Hence, the subdifferential $\partial\hat{f} = \partial f + \partial N_U$ is also a polyhedral multifunction, where $N_U(u)$ is the normal cone of U at u. (The equality follows from [12, Theorem 23.8].) Then, by [10, Corollary], there exist $\alpha \geq 0$ and $\delta > 0$ such that the growth condition (2.17) is satisfied with $r = 1$.

The rate of the outer loop can be obtained by first raising the single-step ratio in (5.8) or (5.9) to the m-th power to get the descent parameter η in the inner-loop stopping criterion (1.13), and then substituting this η in Theorem 3. For instance, if step rule (i) or (ii) is used in the inner loop, we have

$$\frac{\tilde{f}_k(w^{k,t+1})-\tilde{f}_k^*}{\tilde{f}_k(w^{k,t})-\tilde{f}_k^*}\le\tilde{\eta}_{(i),(ii)} \quad \text{for} \quad t=0,1,\ldots,m-1, \tag{5.10}$$

where $\tilde{\eta}_{(i),(ii)}$ is given by (5.8). Note that $u^{k+1} = w^{k,m}$*and* $u^k = w^{k,0}$ in the algorithm. Hence, multiplying (5.10) for $t = 0, 1, \ldots, m-1$, we obtain

$$\frac{\tilde{f}(u^{k+1})-\tilde{f}_k^*}{\tilde{f}_k(u^k)-\tilde{f}^*}\le(\tilde{\eta}_{(i),(ii)})^m.$$

According to Theorem 3, the outer iteration converges with a linear rate

$$\frac{f(u^{k+1})-f^*}{f(u^k)-f^*}\le 1-\frac{1-(\tilde{\eta}_{(i),(ii)})^m}{2(1+\alpha/c)}, \tag{5.11}$$

where

$$\tilde{\eta}_{(i),(ii)} = \tilde{\eta}_{(i),(ii)}\big(\gamma(c)\big)$$

depends on c by (5.8) and (5.7). A "suboptimal" choice of c could be obtained by minimizing the right-hand side of (5.11) in c.

The inner-loop algorithm used above has been extended, in [18], to solve the minimax problems represented by a more general Lagrangian than (5.1). It is not difficult to derive a linearly convergent inexact PPA for those minimax problems by following the patterns in this section.

6. Applications in Unconstrained Minimization.

Consider the unconstrained minimization of a convex C^1 function

$$(\mathcal{P}_S) \qquad\qquad \text{minimize } \ f(x) \ \text{ over all } \ x \in \mathbb{R}^n,$$

where f is *not* necessarily strongly convex. Most commonly used minimization algorithms with a linear (or better) rate require that there exists a (convex) neighborhood $\mathcal{N}$ of $\bar{X}$ such that

$$\sigma_1|y-x|^2 \le \langle y-x, \nabla f(y) - \nabla f(x)\rangle \le \sigma_n|y-x|^2 \quad \forall\, y \in \mathcal{N},\ \forall\, x \in \mathcal{N} \tag{6.1}$$

for some $\sigma_n \ge \sigma_1 > 0$. Their convergence properties are usually not guaranteed if $\sigma_1 = 0$ in (6.1), a situation that occurs when f is not strongly convex. Now, by putting these algorithms in the inner loop of the inexact PPA, we can make these algorithms work in the latter case, while still retaining a linear rate of convergence.

Take the steepest descent method as an example. The algorithm searches along the direction $d^k = -\nabla f(x^k)$ to generate the new iterate

$$x^{k+1} := x^k + \lambda_k d^k, \ \text{ where } \ \lambda_k = \operatorname*{argmin}_{\lambda \in [0,+\infty)} f(x^k + \lambda d^k).$$

If (6.1) is satisfied, then it is not difficult to prove the inequality

$$f(y) \le f(z) + \langle \nabla f(z), y - z\rangle + \frac{\sigma_n}{2}|y-z|^2 \quad \forall\, y \in \mathcal{N},\ \forall\, x \in \mathcal{N}. \tag{6.2}$$

(Cf. the techneque used in [9, p. 86–87], with minor adaptations.) By taking $y = x^k + \sigma_n^{-1} d^k$ and $z = x^k$, we have

$$\begin{aligned} f(x^{k+1}) &\le f(x^k + \sigma_n^{-1} d^k) \\ &\le f(x^k) + \sigma_n^{-1}\langle \nabla f(x^k), d^k\rangle + \frac{\sigma_n^{-1}}{2}|d^k|^2 \\ &\le f(x^k) - \frac{1}{2\sigma_n}|\nabla f(x^k)|^2. \end{aligned} \tag{6.3}$$

On the other hand, by taking $y = x^k$ and $z = \bar{x}$ in (6.2), we obtain

$$f(x^k) \le f(\bar{x}) + \frac{\sigma_n}{2}|x^k - \bar{x}|^2. \tag{6.4}$$

Combining (6.3) and (6.4), we have

$$\frac{f(x^k) - f(x^{k+1})}{f(x^k) - f(\bar{x})} \ge \frac{1}{\sigma_n^2}\frac{|\nabla f(x^k)|^2}{|x^k - \bar{x}|^2}.$$

But

$$|\nabla f(x^k)||x^k - \bar{x}| \ge \langle \nabla f(x^k), x^k - \bar{x}\rangle \ge \sigma_1 |x^k - \bar{x}|^2$$

by (6.1). Hence the sequence of objective values will eventually converge with a linear rate

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \le 1 - \big(\frac{\sigma_1}{\sigma_n}\big)^2. \tag{6.5}$$

Now consider problems with $\sigma_1 = 0$ in (6.1). Suppose the optimal solution set $\bar{X}$ of $(\mathcal{P}_S)$ is nonempty and bounded. Let

$$\tilde{f}(x; z) = f(x) + \frac{1}{2c}|x - z|^2.$$

A possible iteration scheme based on GAP–2 with $M_k = I/c$ goes as follows.

Linearly Convergent Inexact PPA for Unconstrainted Minimization. *Choose an integer $m \ge 1$. Specify starting point $x^0 \in X$. For $k = 0, 1, 2, \ldots$, generate the sequence $\{x^k\}$ as follows: if $x^k \in \bar{X}$, then stop; otherwise*

(i) minimize $\tilde{f}(w; x^k)$ starting from $w^{k,0} = x^k$ to generate $w^{k,1}, w^{k,2}, \ldots, w^{k,m}$ by using the steepest descent algorithm;

(ii) take $x^{k+1} = w^{k,m}$.

Similar to (6.1), now the regularized objective function $\tilde{f}(\cdot; z)$ of the inner-loop subproblem satisfies

$$c^{-1}|y-x|^2 \leq \langle y-x, \nabla \tilde{f}(y;z) - \nabla \tilde{f}(x;z)\rangle \leq (c^{-1}+\sigma_n)|y-x|^2 \quad \forall\, y \in \mathcal{N}, \forall\, x \in \mathcal{N}. \tag{6.6}$$

Without loss of generality, suppose $\mathcal{N}$ is convex. By taking limits on both sides of the inequality

$$f^* + \frac{1}{2c}|w^{k,t} - x^k| \leq f(w^{k,t}) + \frac{1}{2c}|w^{k,t} - x^k| = \tilde{f}_k(w^{k,t}) \leq \tilde{f}_k(x^k) = f(x^k)$$

as $k \to \infty$, it is easy to see that, for any $t \in \{0, 1, \ldots, m\}$,

$$\mathrm{dist}(w^{k,t}, \bar{X}) \to 0 \text{ as } k \to \infty.$$

Hence all the points $w^{k,0}, w^{k,2}, \ldots, w^{k,m}$ will be in $\mathcal{N}$ for sufficiently large k. Therefore, similar to (6.5) for f, we have

$$\frac{\tilde{f}_k(w^{k,t+1}) - \tilde{f}_k^*}{\tilde{f}_k(w^{k,t}) - \tilde{f}_k^*} \leq 1 - \frac{1}{(1+c\sigma_n)^2} \tag{6.7}$$

for $\tilde{f}_k$ by (6.6), where $t = 0, 1, \ldots, m-1$. Note that $x^{k+1} = w^{k,m}$ and $x^k = w^{k,0}$ in the algorithm. Multiplying (6.7) for $t = 0, 1, \ldots, m-1$, we obtain

$$\frac{\tilde{f}(x^{k+1}) - \tilde{f}_k^*}{\tilde{f}_k(x^k) - \tilde{f}^*} \leq \big(1 - \frac{1}{(1+c\sigma_n)^2}\big)^m. \tag{6.8}$$

Hence the stopping criterion (1.17) of GAP–2 is satisfied with an η equals to the right-hand side of (6.8). According to Theorem 3, the outer iteration converges with a linear rate

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{1}{2(1+\alpha/c^r)}\Big(1 - \big(1 - \frac{1}{(1+c\sigma_n)^2}\big)^m\Big), \tag{6.9}$$

provided the growth condition (3.17) is satisfied. A "suboptimal" selection of c could be obtained by minimizing the right-hand side of (6.9). In the special case of $m = 1$, this iteration scheme reduces to Iusem-Svaiter's method in [4], and (6.9) gives a linear rate for their method when f satisfies the conditions in this paper.

Note that instead of the steepest descent method, one can use the conjugate gradient method or the BFGS method in the inner-loop iteration as well, to get the "inexact PPA version of the CG method" or the "inexact PPA version of the BFGS method," correspondingly. These iteration schemes can be used on problems without strong convexity, where the ordinary methods may fail. The rate of convergence will depend on the progress of the inner loop resulting from those individual algorithms.

Acknowledgments. The author is indebted to Professor C. Lemaréchal for his insightful comments and encouragements. The preprints sent to the author by him were very helpful in preparing the revised version of this paper.

References.

[1] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal and C. A. Sagastizábal, "A family of variable metric proximal methods," Math. Programming, to appear.

[2] R. Correa and C. Lemaréchal, "Convergence of some algorithms for convex minimization," Math. Programming 62 (1993), pp. 261–275.

[3] O. Güler, "On the convergence of the proximal point algorithm for convex minimization," SIAM J. Control Opt. 29 (1991), pp. 403–419.

[4] A. N. Iusem and B. F. Svaiter, "'A proximal regularization of the steepest descent method," RAIRO, to appear. C. A. Sagastizábal, "A family of variable metric proximal methods," Math. Programming, to appear.

[5] F. J. Luque, "Asymptotic convergence analysis of the proximal point algorithm," SIAM J. Control Opt. 22 (1984), pp. 277–293.

[6] B. Martinet, "Regularisation d'inéquations variationelles par approximations successives," Rev. Francaise Inf. Rech. Oper., 1970, pp. 154–159.

[7] B. Martinet, "Determination approchée d'un point fixe d'une application pseudo-contractante," C.R. Acad. Sci. Paris 274 (1972), pp. 163–165.

[8] G. J. Minty, "Monotone (nonlinear) operators in Hilbert space," Duke Math. J. 29 (1962), pp. 341-346.

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

[10] S. M. Robinson, "Some continuity properties of polyhedral multifunctions," Math. Programming Studies 14 (1981), pp. 206–214.

[11] R. T. Rockafellar, "On the maximality of sums of nonlinear monotone operators," Trans. Amer. Math. Soc. 149 (1970), pp. 75–88.

[12] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, N.J., 1970.

[13] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," SIAM J. Control Opt. 14 (1976), pp. 877–898.

[14] R. T. Rockafellar, "Linear-quadratic programming and optimal control," SIAM J. Control Opt. 25 (1987), pp. 781–814.

[15] R. T. Rockafellar, "Computational schemes for solving large-scale problems in extended linear-quadratic programming," Math. Programming 48 (1990), pp. 447-474.

[16] R. T. Rockafellar and R. J-B Wets, "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming," Math. Programming Studies 28 (1986), pp. 63–93.

[17] C. Zhu, "On the primal-dual steepest descent algorithm for extended linear-quadratic programming," SIAM J. Optimization, accepted.

[18] C. Zhu, "Solving large scale minimax problems with the primal-dual steepest descent algorithm," Mathematical Programming, accepted.

[19] C. Zhu, "Asymptotic convergence analysis of the forward-backward splitting algorithm," Mathematics of Operations Research, accepted.

[20] C. Zhu and R. T. Rockafellar, "Primal-dual projected gradient algorithms for extended linear-quadratic programming," SIAM J. Optimization 3 (1993), pp. 751–783.