RECEIVED
FFR 28 1996
OSTI

LINEARLY CONVERGENT
INEXACT PROXIMAL POINT ALGORITHM
FOR MINIMIZATION

Ciyou Zhu *

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

August, 1993; revised April, 1994

Abstract. In this paper, we propose a linearly convergent inexact PPA for minimization, where the inner loop stops when the relative reduction on the residue (defined as the objective value minus the optimal value) of the inner loop subproblem meets some *preassigned* constant. This inner loop stopping criterion can be achieved in a *fixed* number of iterations if the inner loop algorithm has a linear rate on the regularized subproblems. Therefore the algorithm is able to avoid the computationally expensive process of solving the inner loop subproblems exactly or asymptotically accurately – a process required by most of the other linearly convergent PPAs.

As applications of this inexact PPA, we develop *linearly convergent* iteration schemes for minimizing functions with singular Hessian matrices, and for solving hemiquadratic extended linear-quadratic programming problems. We also prove that Correa-Lemaréchal's "implementable form" of PPA [2, Algorithm 3.3] converges linearly under mild conditions.

Keywords. Convex programming, proximal point algorithm, extended linear-quadratic programming, singular Hessian matrix.

* 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.

MASTER
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

The submitted manuscript has been authored by a contractor of the U. S. Government under contract No. W-31-109-ENG-38. Accordingly, the U. S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U. S. Government purposes.

1. Introduction.

Consider the convex programming problem

$$(\mathcal{P}) \quad \text{minimize } f(x) \text{ over all } x \in X$$

in the n -dimensional Euclidean space, where $X \subset \mathbb{R}^n$ is a closed convex set and $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a convex function. Denote the Euclidean norm in $\mathbb{R}^n$ by $|\cdot|$. A general iteration scheme for solving $(\mathcal{P})$ is the proximal point algorithm (PPA) [5].

Proximal Point Algorithm (Exact PPA). Choose a positive constant c (proximal parameter). Specify starting points x^0 . For $k = 0, 1, 2, \dots$, generate the sequence $\{x^k\}$ as

$$x^{k+1} := \operatorname{argmin}_{x \in X} \left\{ f(x) + \frac{1}{2c} |x - x^k|^2 \right\}. \quad (1.1)$$

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 (1.1). The objective functions in these subproblems are strongly convex because of the augmented term $\frac{1}{2c} |x - x^k|^2$.

In a more general sense, the PPA can be formulated as a procedure for solving the multivalued equation $0 \in T(x)$ related to a *maximal monotone operator* $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ [12]. The operator

$$J_c = (I + cT)^{-1} \quad (1.2)$$

is single valued from all of $\mathbb{R}^n$ into $\mathbb{R}^n$ [7]. Hence the iteration

$$x^{k+1} := J_c(x^k) \quad (1.3)$$

is well defined. Moreover, $J_c(x) = x$ if and only if $0 \in T(x)$. The iteration (1.3) converges to a solution of $0 \in T$ [5, 6]. For the minimization problem $(\mathcal{P})$, define the extended-valued convex function $\hat{f} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ as

$$\hat{f}(x) = f(x) + \delta_X(x), \quad (1.4)$$

where δ_X is the indicator function of the set X . Then the subdifferential mapping $\partial \hat{f}$ is maximal monotone [10]. The problem of minimizing $\hat{f}$ is equivalent to the multivalued equation $0 \in T(x)$ with $T = \partial \hat{f}$, and iteration (1.3) reduces precisely to (1.1). See [12] for detailed discussions.

One problem in the implementation of PPA is the calculation of argmin on the right-hand side of (1.1). 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 (1.1). In [12], Rockafellar proposed a more realistic version of the PPA, where an approximation, instead of the exact argmin in (1.1), is taken as the next iterate x^{k+1} .

Asymptotically Accurate Proximal Point Algorithm [12]. Choose a positive constant (proximal parameter) c , and a positive sequence $\{\delta_k\}$ such that $\sum \delta_k < +\infty$. Specify starting points $x^0 \in X$. For $k = 0, 1, 2, \dots$, generate the sequence $\{x^k\}$ in such a way that

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

where

$$\bar{x}^{k+1} := \operatorname{argmin}_{x \in X} \{f(x) + \frac{1}{2c}|x - x^k|^2\}. \quad (1.6)$$

This iteration scheme converges to an optimal solution of $(\mathcal{P})$ [12]. If in addition, x^{k+1} satisfies

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

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 [4], where conditions similar to (1.7) are also imposed.

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

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

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.

In [2], Correa and Lemaréchal proposed an “implementable form” of PPA, where they use an ε -subdifferential relation to stop the inner loop iteration. (See Section 6.) However, this stopping criterion will still be tightened asymptotically as $\varepsilon \rightarrow 0$ in the algorithm. Besides that, the rate of convergence of the algorithm was not discussed there.

Another question in the implementation of PPA is the choice of the proximal parameter c . According to the analysis of Rockafellar [12] and Luque [4], larger c will in general lead to faster convergence for the outer-loop iteration. However, a larger c often results in numerically more difficult subproblems and, therefore, a slower progress in the inner loop. Hence, a reasonable choice of c should be based on the analysis of the combined effects of c on both the inner and the outer loops.

In this paper, we develop some linearly convergent variants of PPA, where the inner-loop subproblems need not to be solved exactly or asymptotically accurately, but with limited relative precision during the whole process. We call these new variants *inexact* proximal point algorithms, in contrast to the PPA in its exact form or in its asymptotically accurate forms. In Section 2, we formulate two versions of inexact PPAs and prove their convergence. In Section 3, we prove the linear convergence of the sequence $\{f(x^k)\}$ of objective values generated by the inexact PPA under mild assumptions. In other words, we *get rid of any exactly or asymptotically accurately solved subproblems, while still achieving linear convergence for the objective values*. In Section 4, we derive linearly convergent iteration schemes in the framework of the inexact PPA for minimizing objective functions with singular Hessian matrices. In Section 5, we apply the inexact PPA to solve the hemiquadratic extended linear-quadratic programming problems. We also derive certain “suboptimal” choices for the proximal parameter c in these applications. We conclude the paper by showing, in Section 6, that Correa–Lemaréchal’s inner loop stopping criterion in their “implementable form” of PPA [2, Algorithm 3.3] implies the stopping criterion in our algorithm. Thus the linear convergence of Correa–Lemaréchal’s algorithm follows immediately from the results of this paper.

2. Linearly Convergent Inexact Proximal Point Algorithms.

In this section, we formulate two linearly convergent versions of inexact PPAs and prove their convergence. The first of these two versions applies to the general convex objective functions, while the second version takes advantage of the differentiability of the objective function when it is continuously differentiable.

A prominent feature of these inexact PPAs is the *fixed* relative accuracy in the inner loop stopping criterion. For fixed $c > 0$ and $z \in \mathbb{R}^n$, let the function $\tilde{f}(\cdot; z) : \mathbb{R}^n \rightarrow \mathbb{R}$ be defined as

$$\tilde{f}(x; z) = f(x) + \frac{1}{2c}|x - z|^2. \quad (2.1)$$

For convenience, denote

$$\tilde{f}_k(\cdot) = \tilde{f}(\cdot; x^k) \quad \text{and} \quad \tilde{f}_k^* = \min_{x \in X} \tilde{f}_k(x). \quad (2.2)$$

At the base point x^k , the minimization of the regularized objective function $\tilde{f}_k$ will be stopped as soon as a point $\tilde{y}^{k+1}$ satisfying

$$\frac{\tilde{f}_k(\tilde{y}^{k+1}) - \tilde{f}_k^*}{\tilde{f}_k(x^k) - \tilde{f}_k^*} \leq \eta \quad (2.3)$$

is generated in the inner loop, where η is a constant in $(0, 1)$. Obviously starting from x^k , this stopping criterion will be satisfied within a fixed number of iterations if the inner loop algorithm itself has a linear rate on the regularized function $\tilde{f}_k$.

Of course, stopping criterion (2.3) in the inner loop does not suffice to guarantee the convergence of the outer loop in general. A novel idea in our algorithm is to use a *bench mark point* in the updating to secure the convergence. With the combined effects of the bench mark point and the inner loop stopping criterion (2.3), we are able to obtain *both the convergence and a linear rate* for the whole process without requiring the proximal calculation being carried out exactly or asymptotically accurately.

We denote the Euclidean inner product by $\langle \cdot, \cdot \rangle$, and denote the line segment between two points w^1 and w^2 by $[w^1, w^2]$. The *optimal solution set* $\bar{X}$ of $(\mathcal{P})$ is a closed convex set in $\mathbb{R}^n$. Let f^* be the corresponding *optimal value*. For given z and $[w^1, w^2] \subset X$, define

$$\tilde{f}_{z, [w^1, w^2]}^* = \min_{x \in [w^1, w^2]} \tilde{f}(x; z).$$

Inexact Proximal Point Algorithm – 1 (for general convex f).

(i) Choose an inner loop algorithm for minimizing the regularized objective function $\tilde{f}(\cdot; z)$ on X for fixed z , as well as a constant $\eta \in (0, 1)$ for the stopping criterion.

(ii) Choose a positive constant $\beta \in (0, 1]$, and a mapping $D : X \rightarrow X$ such that for all $z \in X \setminus \bar{X}$, $D(x) \neq x$ and $d(z) = D(z) - z$ being a descent direction of $\tilde{f}(\cdot; z)$ at z . For any $x \in X$ and $y \in X$, let

$$A(x) = \{ y \in [x, D(x)] \mid \tilde{f}(y; x) \leq (1 - \beta)\tilde{f}(x; x) + \beta\tilde{f}_{x, [x, D(x)]}^* \}. \quad (2.4)$$

Specify starting point $x^0 \in X$. For $k = 0, 1, 2, \dots$, generate the sequence $\{x^k\}$ as follows: if $x^k \in \bar{X}$, then stop; otherwise

Step 1. generate a bench mark point $y^{k+1} \in A(x^k)$;

Step 2. start the minimization of $\tilde{f}_k(x)$ at x^k to get some $\tilde{y}^{k+1}$ satisfying the inner loop stopping criterion (2.3)

$$\frac{\tilde{f}_k(\tilde{y}^{k+1}) - \tilde{f}_k^*}{\tilde{f}_k(x^k) - \tilde{f}_k^*} \leq \eta$$

Step 3. take

$$x^{k+1} = \begin{cases} \tilde{y}^{k+1} & \text{if } \tilde{f}_k(\tilde{y}^{k+1}) \leq \tilde{f}_k(y^{k+1}), \\ y^{k+1} & \text{otherwise;} \end{cases}$$

and repeat.

Observe that no exact line search is required in the generation of the bench mark points. By (2.4), y^{k+1} needs only to reduce the regularized objective function $\tilde{f}_k$ to a certain extent (determined by β) on the line segment $[x^k, D(x^k)]$. Moreover, we will show later that some of the inner loop algorithms provide such a bench mark point automatically in the first iteration. In those cases, there would be no extra computations involved in Step 1.

Obviously, stopping criterion (2.3) could be replaced by any other conditions, as long as these conditions imply (2.3). One possibility is to impose an upper bound for the number of inner loop iterations if the inner loop algorithm itself has a linear rate on the regularized function $\tilde{f}_k$. In Section 6, we will prove that Correa-Lemaréchal's stopping criterion in their "implementable form" of PPA [2, Algorithm 3.3] also implies (2.3) with certain $\eta \in (0, 1)$.

If the objective function f in $(\mathcal{P})$ is differentiable in an open neighborhood containing X , then the formula in Step 1 of the above algorithm for bench mark points can be replaced by the Armijo step length strategy.

Inexact Proximal Point Algorithm – 2 (for C^1 objective function).

(i) Choose an inner loop algorithm for minimizing the regularized objective function $\tilde{f}(\cdot; z)$ on X for fixed z , as well as a constant $\eta \in (0, 1)$ for the stopping criterion.

(ii) Choose $\theta \in (0, 1)$, $\tau \in (0, 1)$, and a mapping $D : X \rightarrow X$ such that for all $z \in X \setminus \bar{X}$, $D(z) \neq z$ and $\langle \nabla \tilde{f}(z; z), D(z) - z \rangle < 0$, where $\nabla \tilde{f}(z; z)$ is the gradient of $\tilde{f}(\cdot; z)$ at z . Let

$$A'(x) = x + \lambda(x)d(x), \quad (2.5)$$

where $d(x) = D(x) - x$ and

$$\lambda(x) = \max \{ \theta^j \mid \tilde{f}(x + \theta^j d(x); x) - \tilde{f}(x; x) \leq \tau \theta^j \langle \nabla \tilde{f}(x; x), d(x) \rangle, j \in \{0, 1, 2, \dots\} \}$$

for $x \notin \bar{X}$.

Specify starting point $x^0 \in X$. For $k = 0, 1, 2, \dots$, generate the sequence $\{x^k\}$ as follows: if $x^k \in \bar{X}$, then stop; otherwise

Step 1. generate the bench mark point $y^{k+1} = A'(x^k)$;

Step 2. start the minimization of $\tilde{f}_k(x)$ at x^k to get some $\tilde{y}^{k+1}$ satisfying the inner loop stopping criterion (2.3)

$$\frac{\tilde{f}_k(\tilde{y}^{k+1}) - \tilde{f}_k^*}{\tilde{f}_k(x^k) - \tilde{f}_k^*} \leq \eta$$

Step 3. take

$$x^{k+1} = \begin{cases} \tilde{y}^{k+1} & \text{if } f(\tilde{y}^{k+1}) \leq f(y^{k+1}), \\ y^{k+1} & \text{otherwise;} \end{cases}$$

and repeat.

Proposition 1 (convergence of inexact PPA-1). Suppose the optimal solution set $\bar{X}$ is nonempty and bounded. If the mapping $D : X \rightarrow X$ in the inexact PPA-1 is continuous on $X \setminus \bar{X}$, and if for all $z \in X \setminus \bar{X}$, $d(z) = D(z) - z$ is a descent direction of $\tilde{f}(\cdot; z)$ at z , then the sequence $\{x^k\} \subset X$ generated by the inexact PPA-1 either terminates at a point in $\bar{X}$ or satisfies

$$\text{dist}(x^k, \bar{X}) \rightarrow 0 \text{ and } f(x^k) \rightarrow f^* \text{ as } k \rightarrow \infty.$$

Proof. Let

$$B(y) = \{x \in X \mid f(x) \leq f(y)\}. \quad (2.6)$$

Then

$$x^{k+1} \in B(y^{k+1}) \in (B \circ A)(x^k)$$

by Step 3 of the algorithm. In the following, we prove the proposition by verifying that for the algorithmic mapping A and B defined in (2.4) and (2.6) respectively, all the conditions of [1, Theorem 7.3.4] on the convergence of the iteration $x^{k+1} \in (B \circ A)(x^k)$ are satisfied.

First, we claim that A is closed on $X \setminus \bar{X}$ in the sense that $w^i \in X, w^i \rightarrow w \in X \setminus \bar{X}$, and $y^i \in A(w^i), y^i \rightarrow y$, implies $y \in A(w)$ [1]. Now $y^i \in A(w^i)$ means

$$\tilde{f}(y^i; w^i) \leq (1 - \beta)\tilde{f}(w^i; w^i) + \beta\tilde{f}_{w^i, [w^i, D(w^i)]}^*. \quad (2.7)$$

Let $\lambda \in [0, 1]$ be a real number such that

$$(1 - \lambda)w + \lambda D(w) \in \operatorname{argmin}_{z \in [w, D(w)]} \tilde{f}(z; w).$$

Then, by the continuity of convex function in its effective domain [11, Theorem 10.1],

$$\tilde{f}((1 - \lambda)w^i + \lambda D(w^i); w^i) \rightarrow \min_{z \in [w, D(w)]} \tilde{f}(z; w), \text{ as } w^i \rightarrow w.$$

But $\tilde{f}_{w^i, [w^i, D(w^i)]}^* \leq \tilde{f}((1 - \lambda)w^i + \lambda D(w^i); w^i)$. Therefore (2.7) yields

$$\tilde{f}(y^i; w^i) \leq (1 - \beta)\tilde{f}(w^i; w^i) + \beta\tilde{f}((1 - \lambda)w^i + \lambda D(w^i); w^i). \quad (2.8)$$

Taking the limit on both sides of (2.8), we obtain

$$\tilde{f}(y; w) \leq (1 - \beta)\tilde{f}(w; w) + \beta\tilde{f}_{w, [w, D(w)]}^*,$$

which means that $y \in A(w)$. Hence the algorithmic mapping A is closed on $X \setminus \bar{X}$.

Second, we claim that

$$f(y) < f(z) \quad \forall y \in A(z), z \in X \setminus \bar{X}. \quad (2.9)$$

Recall that $D(z) - z$ is a descent direction of $\tilde{f}(\cdot; z)$ at z for all $z \in X \setminus \bar{X}$. Hence

$$\tilde{f}_{z, [z, D(z)]}^* < \tilde{f}(z; z) \quad \forall z \in X \setminus \bar{X},$$

which implies

$$\tilde{f}(y; z) < \tilde{f}(z; z) \quad \forall y \in A(z), z \in X \setminus \bar{X},$$

by (2.4). But $\tilde{f}(z; z) = f(z)$ and $\tilde{f}(y; z) \geq f(y)$. Hence (2.9) follows.

Third, we claim that for any $x^0 \in X$, the sequence $\{x^k\}$ generated by the inexact PPA-1 is contained in a compact subset of X . Note that $\bar{X}$ is a level set of $\hat{f}$ defined in (1.4). But all the level sets of a proper, lower semicontinuous convex function have the same recession cone [11, Theorem 8.7]. Hence, the boundedness of $\bar{X}$ implies the boundedness of

$$\{x \mid f(x) \leq f(x^0)\} \cap X = \{x \mid \hat{f}(x) \leq \hat{f}(x^0)\}. \quad (2.10)$$

Obviously, the set on the left-hand side of (2.10) is a closed subset of X containing the sequence $\{x^k\}$. This completes the proof. $\square$

Proposition 2 (convergence of inexact PPA-2). *Suppose that the optimal solution set $\bar{X}$ is nonempty and bounded and that the objective function f is continuously differentiable in some open neighborhood of X . If the mapping $D : X \rightarrow X$ in the inexact PPA-2 is continuous on $X \setminus \bar{X}$, and $\langle \nabla \tilde{f}(z; z), D(z) - z \rangle < 0$ for all $z \in X \setminus \bar{X}$, where $\nabla \tilde{f}(z; z)$ is the gradient of $\tilde{f}(\cdot; z)$ at z , then the sequence $\{x^k\} \subset X$ generated by the inexact PPA-2 either terminates at a point in $\bar{X}$ or satisfies*

$$\text{dist}(x^k, \bar{X}) \rightarrow 0 \text{ and } f(x^k) \rightarrow f^* \text{ as } k \rightarrow \infty.$$

Proof. Let

$$B(y) = \{x \in X \mid f(x) \leq f(y)\}.$$

Then

$$x^{k+1} \in B(y^{k+1}) = (B \circ A')(x^k)$$

by Step 3 of the algorithm. Note that in (2.5), A' is a descent algorithmic mapping

$$f(A'(x)) \leq \tilde{f}(A'(x); x) < \tilde{f}(x; x) = f(x) \quad \forall x \in X \setminus \bar{X}, \quad (2.11)$$

and B does not increasing the objective value. Let $\{x^k\}$ be an infinite sequence generated by the algorithm. Then $\{x^k\}$ is contained in the set

$$\{x \mid f(x) \leq f(x^0)\} \cap X = \{x \mid \hat{f}(x) \leq \hat{f}(x^0)\},$$

which is compact as has been shown in the proof of Proposition 1. Let $\{x^k\}_{\mathcal{K}}$ be a convergent subsequence of $\{x^k\}$,

$$x^k \rightarrow w \text{ as } k \rightarrow \infty, k \in \mathcal{K},$$

where $\mathcal{K}$ is a subset of $\{0, 1, 2, \dots\}$. Then the nonincreasing sequence of objective values $\{f(x^k)\}$ converges to $f(w)$ as $k \rightarrow \infty$. To prove the proposition, we need only to prove that $w \in \bar{X}$.

By contradiction, suppose $w \notin \bar{X}$. Then $x^k \notin \bar{X}$ for sufficiently large $k \in \mathcal{K}$ since $\bar{X}$ is a closed set. Let

$$\lambda^* := \max \{ \theta^j \mid \tilde{f}(w + \theta^j d(w); w) - \tilde{f}(w; w) \leq \tau \theta^j \langle \nabla \tilde{f}(w; w), d(w) \rangle, j \in \{0, 1, 2, \dots\} \},$$

$$\lambda_k := \max \{ \theta^j \mid \tilde{f}_k(x^k + \theta^j d(x^k)) - \tilde{f}_k(x^k) \leq \tau \theta^j \langle \nabla \tilde{f}_k(x^k), d(x^k) \rangle, j \in \{0, 1, 2, \dots\} \},$$

where the simplified notation $\tilde{f}_k(\cdot)$ is used for the function $\tilde{f}(\cdot; x^k)$. Suppose $\lambda^* = \theta^{j'}$. Then

$$\tilde{f}_k(x^k + \theta^j d(x^k)) - \tilde{f}_k(x^k) < \tau \theta^j \langle \nabla \tilde{f}_k(x^k), d(x^k) \rangle \quad \forall j > j',$$

$$\tilde{f}_k(x^k + \theta^j d(x^k)) - \tilde{f}_k(x^k) > \tau \theta^j \langle \nabla \tilde{f}_k(x^k), d(x^k) \rangle \quad \forall j < j',$$

for sufficiently large $k \in \mathcal{K}$. Hence $\lambda_k \in \{\lambda^*, \theta \lambda^*\}$ and

$$\tilde{f}_k(A'(x^k)) - \tilde{f}_k(x^k) \leq \tau \lambda_k \langle \nabla \tilde{f}_k(x^k), d(x^k) \rangle \leq \tau \theta \lambda^* \langle \nabla \tilde{f}_k(x^k), d(x^k) \rangle \quad (2.12)$$

for sufficiently large $k \in \mathcal{K}$. By (2.11) and the equality $\nabla \tilde{f}_k(x^k) = \nabla f(x^k)$, it follows from (2.12) that

$$f(A'(x^k)) - f(x^k) \leq \tau \lambda_k \langle \nabla f(x^k), d(x^k) \rangle \leq \tau \theta \lambda^* \langle \nabla f(x^k), d(x^k) \rangle.$$

Taking $\limsup$ as $k \rightarrow \infty$ with $k \in \mathcal{K}$, we get

$$\limsup_{k \rightarrow \infty, k \in \mathcal{K}} f(A'(x^k)) - f(w) \leq \tau \theta \lambda^* \langle \nabla f(w), d(w) \rangle.$$

Let $\varepsilon = -\tau \theta \lambda^* \langle \nabla f(w), d(w) \rangle$. Then $\varepsilon > 0$ by the assumption in the inexact PPA-2. But $x^{k+1} \in (B \circ A')(x^k)$ implies $f(x^{k+1}) \leq f(A'(x^k))$. Hence

$$f(x^{k+1}) - f(w) \leq -\varepsilon/2 < 0 \text{ for all sufficiently large } k \in \mathcal{K}.$$

This contradicts the fact that the nonincreasing sequence $\{f(x^k)\}$ converges to $f(w)$ as $k \rightarrow \infty$. $\square$

3. Rate of Convergence for the Inexact PPA.

In this section, we investigate the rate of convergence for the sequence $\{f(x^k)\}$ generated by the inexact PPAs. We show that the rate is related to the progress in the inner loop minimization, as well as the growth condition on the inverse of the subdifferential mapping $\partial \hat{f}$ near 0.

The results in this section will also be useful in deriving convergence rates for some other variants of PPAs. In order to facilitate such applications in Section 6, we here consider a slightly more general problem

$$(\mathcal{P}_\varphi) \quad \text{minimize } \varphi(x) \text{ over all } x \in \mathbb{R}^n,$$

where $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ is a proper low semicontinuous convex function. Suppose the optimal solution set $\bar{\mathcal{X}} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \varphi(x)$ is nonempty. For fixed $z \in \mathbb{R}^n$, let

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

Obviously problem $(\mathcal{P}_\varphi)$ includes $(\mathcal{P})$ as a special case, and the regularized function $\tilde{\varphi}(x; z)$ coincides with $\tilde{f}(x; z)$ for $x \in X$ when $\varphi = \hat{f}$ as defined in (1.4). We denote

$$\varphi^* = \varphi(\bar{\mathcal{X}}), \quad \tilde{\varphi}_k(\cdot) = \tilde{\varphi}(\cdot; x^k) \quad \text{and} \quad \tilde{\varphi}_k^* = \min_{x \in \mathbb{R}^n} \tilde{\varphi}_k(x).$$

Let $\mathcal{X}$ be the effective domain [11] of φ ,

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid \varphi(x) < +\infty\}.$$

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

Lemma 3 (relation between inner and outer loop residues). *Suppose there exists $\alpha > 0$, $\delta > 0$ and $r \geq 1$ such that the inverse of the subdifferential mapping $\mathcal{T} = \partial \varphi$ satisfies the growth condition*

$$\forall w \in \mathbb{B}(0, \delta), \quad \forall x \in \mathcal{T}^{-1}(w), \quad \operatorname{dist}(\bar{\mathcal{X}}, x) \leq \alpha|w|^r, \quad (3.1)$$

where $\mathbb{B}(0, \delta)$ is the closed ball centered at 0 with radius δ in $\mathbb{R}^n$. Then for any sequence $\{x^k\} \subset \mathcal{X} \setminus \bar{\mathcal{X}}$ such that $\operatorname{dist}(\bar{\mathcal{X}}, x^k) \rightarrow 0$ as $k \rightarrow \infty$,

$$\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^* \leq \varphi(x^k) - \varphi^* \leq 2(1 + (\alpha/c^r)\operatorname{dist}^{r-1}(\bar{\mathcal{X}}, x^k))(\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^*) \quad (3.2)$$

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

$$\varphi(x^k) - \varphi^* \leq \begin{cases} 2(1 + \alpha/c)(\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^*) & \text{if } r = 1, \\ 2(1 + \varepsilon)(\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^*) & \text{if } r > 1. \end{cases} \quad (3.3)$$

Proof. The first half of (3.2) follows directly from $\tilde{\varphi}_k(x^k) = \varphi(x^k)$ and $\tilde{\varphi}_k^* \geq \varphi^*$. Now we prove the second half of (3.2). Let $\bar{x}^{k+1} = \underset{x \in \mathcal{X}}{\operatorname{argmin}} \tilde{\varphi}_k(x)$. Then $\bar{x}^{k+1} = \mathcal{J}_c(x^k)$ and $\bar{x}^{k+1} \neq x^k$ unless $x^k \in \bar{\mathcal{X}}$, where $\mathcal{J}_c = (I + c\mathcal{T})^{-1}$ is the proximal mapping related to the operator $\mathcal{T}$. According to Güler [3, Lemma 2.2]

$$\varphi(\bar{x}^{k+1}) - \varphi(z) \leq \frac{1}{2c}|z - x^k|^2 - \frac{1}{2c}|z - \bar{x}^{k+1}|^2 - \frac{1}{2c}|x^k - \bar{x}^{k+1}|^2 \quad \forall z \in \mathcal{X}. \quad (3.4)$$

For any $\bar{x} \in \bar{\mathcal{X}}$, $z = \bar{x}$ in (3.4) yields

$$\tilde{\varphi}_k(\bar{x}^{k+1}) - \varphi(\bar{x}) \leq \frac{1}{2c}|x^k - \bar{x}|^2 - \frac{1}{2c}|\bar{x}^{k+1} - \bar{x}|^2. \quad (3.5)$$

On the other hand, $z = x^k$ in (3.4) yields

$$\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(\bar{x}^{k+1}) \geq \frac{1}{2c}|x^k - \bar{x}^{k+1}|^2. \quad (3.6)$$

Combining (3.5) and (3.6), we get

$$\frac{\tilde{\varphi}_k(\bar{x}^{k+1}) - \varphi(\bar{x})}{\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(\bar{x}^{k+1})} \leq \frac{|x^k - \bar{x}|^2 - |\bar{x}^{k+1} - \bar{x}|^2}{|x^k - \bar{x}^{k+1}|^2}. \quad (3.7)$$

But

$$\begin{aligned} \frac{|x^k - \bar{x}|^2 - |\bar{x}^{k+1} - \bar{x}|^2}{|x^k - \bar{x}^{k+1}|^2} &= \frac{|x^k - \bar{x}^{k+1}|^2 + 2\langle x^k - \bar{x}^{k+1}, \bar{x}^{k+1} - \bar{x} \rangle}{|x^k - \bar{x}^{k+1}|^2} \\ &\leq \frac{|x^k - \bar{x}^{k+1}|^2 + 2|x^k - \bar{x}^{k+1}||\bar{x}^{k+1} - \bar{x}|}{|x^k - \bar{x}^{k+1}|^2} \\ &= \frac{|x^k - \bar{x}^{k+1}| + 2|\bar{x}^{k+1} - \bar{x}|}{|x^k - \bar{x}^{k+1}|}. \end{aligned} \quad (3.8)$$

Substituting (3.8) in (3.7), we get

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

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

$$\frac{\tilde{\varphi}_k(\bar{x}^{k+1}) - \varphi^*}{\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(\bar{x}^{k+1})} \leq 1 + \frac{2\text{dist}(\bar{\mathcal{X}}, \bar{x}^{k+1})}{|x^k - \bar{x}^{k+1}|}, \quad (3.9)$$

where φ^* is the optimal value of $(\mathcal{P}_\varphi)$.

Now, according to [12, Proposition 1],

$$|\mathcal{J}_c(z) - \mathcal{J}_c(z')|^2 + |(I - \mathcal{J}_c)(z) - (I - \mathcal{J}_c)(z')|^2 \leq |z - z'|, \quad (3.10)$$

$$(I - \mathcal{J}_c)(z) \in c\mathcal{T}(\mathcal{J}_c(z)), \quad (3.11)$$

for all $z \in \mathcal{X}$ and $z' \in \mathcal{X}$. Let $z = x^k$ and $z' = \bar{x}$ in (3.10) and (3.11). By (3.10),

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

Hence $|x^k - \bar{x}^{k+1}| \leq \text{dist}(\bar{\mathcal{X}}, x^k)$. Therefore

$$|x^k - \bar{x}^{k+1}| \rightarrow 0 \text{ as } \text{dist}(\bar{\mathcal{X}}, x^k) \rightarrow 0.$$

But $x^k - \bar{x}^{k+1} \in c\mathcal{T}(\bar{x}^{k+1})$ by (3.11), or, in other words,

$$\bar{x}^{k+1} \in \mathcal{T}^{-1}\left(\frac{x^k - \bar{x}^{k+1}}{c}\right).$$

Hence it follows from (3.1) that

$$\text{dist}(\bar{\mathcal{X}}, \bar{x}^{k+1}) \leq \frac{\alpha}{c^r} |x^k - \bar{x}^{k+1}|^r \quad (3.12)$$

for sufficiently large k . Substituting (3.12) in (3.9), we obtain

$$\frac{\tilde{\varphi}_k(\bar{x}^{k+1}) - \varphi^*}{\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(\bar{x}^{k+1})} \leq 1 + \frac{2\alpha}{c^r} |x^k - \bar{x}^{k+1}|^{r-1},$$

from which (3.2) follows, since $\tilde{\varphi}_k(x^k) = \varphi(x^k)$ and $|x^k - \bar{x}^{k+1}| \leq \text{dist}(\bar{\mathcal{X}}, x^k)$. $\square$

Lemma 3 may be viewed as a quantization of the relationship

$$\mathcal{J}_c(x) = x \iff 0 \in \mathcal{T}(x).$$

The result can also be used in the optimality test of the algorithm. If an estimation on $\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^*$ for the subproblem is available, then an estimation on $\varphi(x^k) - \varphi^*$ for the original problem can be obtained via (3.3).

The growth condition (3.1) is in fact a local upper Lipschitz condition with modulus α and order r on $\mathcal{T}^{-1}$. It is a common requirement when convergence rates are under considerations. Both Rockafellar [12] and Luque [4] used similar conditions in their analysis of PPAs. Actually, the condition in [12] on the convergence of the asymptotically accurate PPA implies the growth condition with $r = 1$. See Zhu [18] for more discussions on the growth conditions.

Theorem 4 (relation between inner and outer loop rates). Suppose $\{x^k\} \subset \mathcal{X} \setminus \bar{\mathcal{X}}$ is a sequence converging to $\bar{\mathcal{X}}$ in the sense that $\text{dist}(\bar{\mathcal{X}}, x^k) \rightarrow 0$. Assume the growth condition (3.1) on the inverse of the subdifferential mapping $T = \partial\varphi$ is satisfied with constants $\alpha > 0$, $\delta > 0$ and $r \geq 1$. Let $\{\tilde{y}^k\} \subset \mathcal{X}$ be a sequence satisfying

$$\frac{\tilde{\varphi}_k(\tilde{y}^{k+1}) - \tilde{\varphi}_k^*}{\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^*} \leq \eta \quad (3.13)$$

with some $\eta \in [0, 1)$ for all sufficiently large k . Then

$$\frac{\varphi(\tilde{y}^{k+1}) - \varphi^*}{\varphi(x^k) - \varphi^*} \leq 1 - \frac{1 - \eta}{2(1 + (\alpha/c^r)\text{dist}^{r-1}(\bar{\mathcal{X}}, x^k))} \quad (3.14)$$

for all sufficiently large k .

Proof. It follows from (3.13) that

$$\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(\tilde{y}^{k+1}) \geq (1 - \eta)(\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^*) \quad (3.15)$$

for sufficiently large k . However, by the second half of (3.2),

$$\varphi(x^k) - \varphi^* \leq (2 + (2\alpha/c^r)\text{dist}^{r-1}(\bar{\mathcal{X}}, x^k))(\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^*) \quad (3.16)$$

for sufficiently large k . Observe that $\varphi(x^k) - \varphi^* > 0$ for $x^k \notin \bar{\mathcal{X}}$. Hence dividing (3.15) by (3.16), we have

$$\frac{\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(\tilde{y}^{k+1})}{\varphi(x^k) - \varphi^*} \geq \frac{1 - \eta}{2(1 + (\alpha/c^r)\text{dist}^{r-1}(\bar{\mathcal{X}}, x^k))}$$

for sufficiently large k . But $\tilde{\varphi}_k(x^k) = \varphi(x^k)$ and $\tilde{\varphi}_k(\tilde{y}^{k+1}) \geq \varphi(\tilde{y}^{k+1})$. Therefore

$$\frac{\varphi(x^k) - \varphi(\tilde{y}^{k+1})}{\varphi(x^k) - \varphi^*} \geq \frac{1 - \eta}{2(1 + (\alpha/c^r)\text{dist}^{r-1}(\bar{\mathcal{X}}, x^k))}$$

for sufficiently large k , from which (3.14) follows. $\square$

Corollary 5 (rate of convergence of inexact PPA-1 and PPA-2). Suppose there exist constants $\alpha > 0$, $\delta > 0$ and $r \geq 1$ such that the inverse of the subdifferential mapping $T = \partial\hat{f}$ related to $\hat{f}$ in (1.4) satisfies the growth condition

$$\forall w \in \mathbb{B}(0, \delta), \quad \forall x \in T^{-1}(w), \quad \text{dist}(\bar{\mathcal{X}}, x) \leq \alpha|w|^r, \quad (3.17)$$

where $\mathbb{B}(0, \delta)$ is the closed ball centered at 0 with radius δ in $\mathbb{R}^n$. Let $\{x^k\} \subset X \setminus \bar{X}$ be a sequence generated by PPA-1 (or PPA-2). If the conditions in Proposition 1 (or Proposition 2) for the convergence of PPA-1 (or PPA-2 respectively) are satisfied, then the sequence $\{f(x^k)\}$ of the objective value converges to the optimal value of $(\mathcal{P})$ linearly in the sense that

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{1 - \eta}{2(1 + \alpha/c^r)} \quad (3.18)$$

for all sufficiently large k , and

$$\limsup_{k \rightarrow \infty} \frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq \frac{1 + \eta}{2} \text{ if } r > 1, \quad (3.19)$$

where $\eta \in (0, 1)$ is the constant in the stopping criterion of the algorithm.

Proof. By Proposition 1 (or Proposition 2 respectively), we have $\text{dist}(\bar{X}, x^k) \rightarrow 0$ as $k \rightarrow \infty$. Now invoke Theorem 4 for $\varphi = \hat{f}(x) = f(x) + \delta_X(x)$. (3.18) and (3.19) follow directly from (3.14), since $f(x^{k+1}) \leq f(\tilde{y}^{k+1})$ in the algorithm. $\square$

According to Corollary 5, 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 (3.17) 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.

Corollary 5 also throws a light on the choice of the proximal parameter c . For $r = 1$, which is the common case corresponding to an upper Lipschitz continuous inverse of the subdifferential mapping $\partial \hat{f}$ in some neighborhood of 0, the rate given by (3.18) 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 choice of c can be optimized by minimizing the right-hand side of (3.18).

4. Applications In Unconstrained Minimization.

Consider the unconstrained minimization of a convex C^2 function with possibly singular Hessian matrix

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

Most commonly used minimization algorithms with a linear (or superlinear) rate require that the Hessian matrix be uniformly positive definite and bounded in some neighborhood $\mathcal{N}$ of $\bar{X}$, that is, that there exist $\sigma_n \geq \sigma_1 > 0$ such that

$$\sigma_1 |w|^2 \leq \langle w, \nabla^2 f(x) w \rangle \leq \sigma_n |w|^2 \quad \forall w \in \mathcal{N}, \quad \forall x \in \mathbb{R}^n. \quad (4.1)$$

Their convergence properties are usually not guaranteed if the Hessian matrix is singular at some $\bar{x} \in \bar{X}$, a situation that occurs, for instance, when the optimal solution is not unique. Now, by putting these algorithms in the inner loop of the inexact PPA, we can extend them to solve problems with singular Hessian matrices, 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

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

If (4.1) is satisfied (without loss of generality, suppose $\mathcal{N}$ is convex), then

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k + \sigma_n^{-1} d^k) \\ &= f(x^k) + \sigma_n^{-1} \langle \nabla f(x^k), d^k \rangle + \frac{\sigma_n^{-2}}{2} \langle d^k, \nabla^2 f(w) d^k \rangle \\ &\leq f(x^k) - \frac{1}{2\sigma_n} |\nabla f(x^k)|^2 \end{aligned} \quad (4.2)$$

for x^k sufficiently close to $\bar{X}$, where $w \in [x^k, x^k + \sigma_n^{-1} d^k]$. On the other hand, the Taylor expansion at an optimal solution $\bar{x} \in \bar{X}$ gives

$$f(x^k) \leq f(\bar{x}) + \frac{\sigma_n}{2} |x^k - \bar{x}|^2. \quad (4.3)$$

Combining (4.2) and (4.3), we have

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

But

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

(cf. [8, p. 86]). Hence the sequence of objective values will eventually converge with a linear rate

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \left(\frac{\sigma_1}{\sigma_n}\right)^2. \quad (4.4)$$

Now consider problems with singular Hessian matrices. Suppose the optimal solution set $\bar{X}$ of $(\mathcal{P}_S)$ is nonempty and bounded. Let

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

One possible iteration scheme based on Inexact PPA-1 goes as follows.

A Linearly Convergent Inexact PPA for Unconstrained Minimization.

Choose an integer $m \geq 1$. Specify starting point $x^0 \in X$. For $k = 0, 1, 2, \dots$, 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 $w^{k,1}, w^{k,2}, \dots, w^{k,m}$ by using the steepest descent algorithm;
- (ii) define $\tilde{y}^{k+1} := w^{k,m}$, $y^{k+1} := w^{k,1}$; take

$$x^{k+1} = \begin{cases} \tilde{y}^{k+1} & \text{if } f(\tilde{y}^{k+1}) \leq f(y^{k+1}), \\ y^{k+1} & \text{otherwise.} \end{cases}$$

Obviously the point $w^{k,1}$ is taken as the bench mark in Step 1 of PPA-1, and the inner loop minimization in Step 2 is stopped when the number of iteration exceeds m . Now $\nabla \tilde{f}(x; x) = \nabla f(x)$, and $\nabla \tilde{f}(x; x) = 0$ if and only if $x \in \bar{X}$. Define

$$D(x) = x - c \nabla \tilde{f}(x; x) = x - c \nabla f(x).$$

The first step of the steepest descent method in the inner loop of minimizing $\tilde{f}_k(u)$ yields

$$y^{k+1} = \operatorname{argmin}\{\tilde{f}_k(x) \mid x \in [x^k, D(x^k)]\}. \quad (4.5)$$

Therefore the conditions in Proposition 1 are satisfied, and the algorithm converges in the sense that $\operatorname{dist}(x^k, \bar{X}) \rightarrow 0$ and $f(x^k) \rightarrow f^*$ as $k \rightarrow \infty$, unless the iteration terminates with an $x^k \in \bar{X}$.

Now for any bounded neighborhood $\mathcal{N}$ of $\bar{X}$, there exists $\sigma_n > 0$ such that the objective function $\tilde{f}(\cdot; z)$ of the inner-loop subproblem satisfies

$$c^{-1}|w|^2 \leq \langle w, \nabla^2 \tilde{f}(x; z)w \rangle \leq (c^{-1} + \sigma_n)|w|^2 \quad \forall w \in \mathbb{R}^n, \forall x \in \mathcal{N}, \forall z \in \mathcal{N}. \quad (4.6)$$

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

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

Hence all the points $w^{k,0}, w^{k,2}, \dots, w^{k,m}$ will be in $\mathcal{N}$ for sufficiently large k . Therefore, similar to (4.4) for f , we have

$$\frac{\tilde{f}_k(w^{k,t+1}) - \tilde{f}_k^*}{\tilde{f}_k(w^{k,t}) - \tilde{f}_k^*} \leq 1 - \frac{1}{(1 + c\sigma_n)^2} \quad (4.7)$$

for $\tilde{f}_k$ by (4.6), where $t = 0, 1, \dots, m-1$. Note that $\tilde{y}^{k+1} = w^{k,m}$ and $x^k = w^{k,0}$ in the algorithm. Multiplying (4.7) for $t = 0, 1, \dots, m-1$, we obtain

$$\frac{\tilde{f}(\tilde{y}^{k+1}) - \tilde{f}^*}{\tilde{f}(x^k) - \tilde{f}^*} \leq \left(1 - \frac{1}{(1 + c\sigma_n)^2}\right)^m. \quad (4.8)$$

Hence the stopping criterion in PPA-1 is satisfied with an η equals to the right-hand side of (4.8). According to Corollary 5, the outer iteration converges with a linear rate

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{1}{2(1 + \alpha/c^r)} \left(1 - \left(1 - \frac{1}{(1 + c\sigma_n)^2}\right)^m\right), \quad (4.9)$$

provided the growth condition (3.17) is satisfied. A “suboptimal” choice of c could be obtained by minimizing the right-hand side of (4.9).

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-1 version of the CG method” or the “inexact PPA-1 version of the BFGS method” correspondingly. These iteration schemes can be used on

the singular problems where the corresponding ordinary methods may fail. The convergence of these schemes follow immediately from the analysis in this section, since the first step in either the CG or the BFGS methods is identical to the first step in the steepest descent method. The rate of convergence will depend on the inner loop progress resulting from those individual algorithms.

5. Application In Hemiquadratic Extended Linear-Quadratic Problems.

Let

$$L(u, v) = p \cdot u + \frac{1}{2} u \cdot P u + q \cdot v - \frac{1}{2} v \cdot Q v - v \cdot R u, \quad (5.1)$$

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

$$\begin{aligned} (\mathcal{P}_L) \quad & \text{minimize } f(u) \text{ over all } u \in U, \text{ where } f(u) := \sup_{v \in V} L(u, v), \\ (\mathcal{Q}_L) \quad & \text{maximize } g(v) \text{ over all } v \in V, \text{ where } g(v) := \inf_{u \in U} L(u, v), \end{aligned}$$

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 $U \times V$ [15, 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 [19, 16]. It has a linear rate on *fully quadratic* problems, that is, the problems with positive definite P and Q matrices [16]. 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 $\text{dom } f = \mathbb{R}^n$, and we can apply the inexact PPA-1 on the primal problem $(\mathcal{P}_L)$. (The variable u corresponds to the variable x in Sections 2 and 3,

and the sets U and $\bar{U}$ correspond to the sets X and $\bar{X}$, 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 augmented Lagrangian

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

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

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

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

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

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

The primal part of the algorithm [16], 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

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

where there are three possible rules to determine the step length λ_k [16]:

(i) Perfect line search

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

(ii) Fixed step lengths

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

(iii) Adaptive step lengths

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

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 [16, Theorems 3.1, 3.2]

$$\frac{\tilde{f}(u^{k+1}; z) - \tilde{f}_z^*}{\tilde{f}(u^k; z) - \tilde{f}_z^*} = \tilde{\eta}_{(i),(ii)} := \begin{cases} \gamma^2 & \text{if } 0 \leq \gamma^2 < \tfrac{1}{2}, \\ 1 - \frac{1}{4\gamma^2} & \text{if } \gamma^2 \geq \tfrac{1}{2}, \end{cases} \quad (5.7)$$

for rules (i) and (ii), and

$$\frac{\tilde{f}(u^{k+1}; z) - \tilde{f}_z^*}{\tilde{f}(u^k; z) - \tilde{f}_z^*} = \tilde{\eta}_{(iii)} := \begin{cases} \frac{\theta}{2} & \text{if } 0 \leq \gamma^2 < \tfrac{1}{2}, \\ 1 - \frac{\theta}{4\gamma^2} & \text{if } \gamma^2 \geq \tfrac{1}{2}, \end{cases} \quad (5.8)$$

for rule (iii) respectively. A possible iteration scheme based on the Inexact PPA-1 goes as follows.

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

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

(ii) define $\tilde{y}^{k+1} := w^{k,m}$, $y^{k+1} := w^{k,1}$; take

$$u^{k+1} = \begin{cases} \tilde{y}^{k+1} & \text{if } f(\tilde{y}^{k+1}) \leq f(y^{k+1}), \\ y^{k+1} & \text{otherwise.} \end{cases}$$

In this scheme, the point $w^{k,1}$ is taken as the bench mark in Step 1 of PPA-1, and the inner loop minimization in Step 2 is stopped when the number of iteration exceeds m . Define

$$D(u) = \tilde{D}(u; u).$$

It follows from Rockafellar [14, p. 469] that D is continuous. Moreover, by [19, Proposition 3.1], $d(u) = D(u) - u$ is a descent direction of $\tilde{f}(\cdot; u)$ at u for $u \notin \bar{U}$. Hence the conditions in Proposition 1 are satisfied by either the step length rule (i),

or (ii), or (iii). Therefore the algorithm converges in the sense that $\text{dist}(u^k, \bar{U}) \rightarrow 0$ and $f(u^k) \rightarrow f^*$ as $k \rightarrow \infty$, unless the iteration terminates with an $u^k \in \bar{U}$.

To verify that the growth condition in Corollary 5 is satisfied, observe that the subdifferentials ∂f is a polyhedral multifunction in the sense of Robinson [9] (cf. [13, 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 [11, Theorem 23.8].) Then, by [9, Corollary], there exist $\alpha > 0$ and $\delta > 0$ such that the growth condition (3.17) is satisfied with $r = 1$.

The rate of the outer loop can be obtained by first raising the single step ratio in (5.7) or (5.8) to the m -th power to get the constant η in the inner loop stopping criterion, and then substituting this η in Corollary 5. For instance, if step rule (i) or (ii) is used in the inner loop, we have

$$\frac{\tilde{f}_k(w^{k,t+1}) - \tilde{f}_k^*}{\tilde{f}_k(w^{k,t}) - \tilde{f}_k^*} \leq \tilde{\eta}_{(i),(ii)} \quad \text{for } t = 0, 1, \dots, m-1, \quad (5.9)$$

where $\tilde{\eta}_{(i),(ii)}$ is given by (5.7). Note that $\tilde{y}^{k+1} = w^{k,m}$ and $u^k = w^{k,0}$ in the algorithm. Hence, multiplying (5.9) for $t = 0, 1, \dots, m-1$, we obtain

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

According to Corollary 5, the outer iteration converges with a linear rate

$$\frac{f(x^{k+1}) - f^*}{f(x^k) - f^*} \leq 1 - \frac{1 - (\tilde{\eta}_{(i),(ii)})^m}{2(1 + \alpha/c)}. \quad (5.10)$$

A "suboptimal" choice of c could be obtained by minimizing the right-hand side of (5.10).

The inner loop algorithm used above has been extended, in [17], to solve the minimax problems represented by a more general Lagrangian than (5.1). Therefore, it is not difficult to derive a linearly convergent inexact PPA for those minimax problems by following the patterns in this section.

6. 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 $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be a proper low semicontinuous convex function. Let $\partial_\varepsilon \varphi$ denote the ε -subdifferential of φ . Recall that

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

For simplicity of the notation, we will still use a fixed proximal constant c . The extension of the results to the case of a sequence of indexed $\{c_k\}$ would be straight forward. With the notations developed in Section 3, Correa-Lemaréchal's implementable form of PPA for minimizing $\varphi(x)$ goes as follows:

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

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

Step 2. Set

$$\varepsilon(x^k, w^{k,j}) = \kappa [\varphi(x^k) - \varphi(w^{k,j}) - \frac{\mu}{c} |w^{k,j} - x^k|^2]. \quad (6.1)$$

If

$$\frac{x^k - w^{k,j}}{c} \in \partial_{\varepsilon(x^k, w^{k,j})} \varphi(x^k) \quad (6.2)$$

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 (6.1) and (6.2) imply

$$\varepsilon(x^k, w^{k,j}) = \kappa [\varphi(x^k) - \varphi(w^{k,j}) - \frac{\mu}{c} |w^{k,j} - x^k|^2] \geq 0. \quad (6.3)$$

They also proved that if φ is strongly coercive, i.e.,

$$\varphi(z)/|z| \rightarrow +\infty \text{ when } |z| \rightarrow \infty, \quad (6.4)$$

then

- (i) for each k , Step 2 eventually exit to Step 3, unless x^k is already optimal [2, Theorem 3.2];
- (ii) $\liminf_{k \rightarrow \infty} \varphi(x^k) = \varphi^*$ [2, Proposition 3.2].

Now we prove further that, when the parameters κ and μ fall in certain ranges, the algorithm converges at least linearly for the problems satisfying the growth conditions (3.1).

Proposition 6 (rate of convergence on Correa-Lemaréchal's form of PPA). Suppose the optimal solution set $\bar{\mathcal{X}}$ is nonempty and bounded, and $x^k \notin \bar{\mathcal{X}}$ in Correa-Lemaréchal's implementable PPA for all k . If φ is strongly coercive in the sense of (6.4), and there exist constants $\alpha > 0$, $\delta > 0$ and $r \geq 1$ such that the inverse of the subdifferential mapping $\mathcal{T} = \partial\varphi$ satisfies the growth condition (3.1) in Lemma 3, then the sequence $\{x^k\} \subset \mathcal{X}$ generated by Correa-Lemaréchal's implementable PPA satisfies

$$\frac{\varphi(x^{k+1}) - \varphi^*}{\varphi(x^k) - \varphi^*} \leq 1 - \frac{1}{(2\kappa + 2/(2\mu - 1))(1 + (\alpha/c^r)\text{dist}^{r-1}(\bar{\mathcal{X}}, x^k))} \quad (6.5)$$

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

$$\frac{\varphi(x^{k+1}) - \varphi^*}{\varphi(x^k) - \varphi^*} \leq 1 - \frac{1}{(2\kappa + 2/(2\mu - 1))(1 + \alpha/c^r)} \quad (6.6)$$

for all sufficiently large k , and

$$\limsup_{k \rightarrow \infty} \frac{\varphi(x^{k+1}) - \varphi^*}{\varphi(x^k) - \varphi^*} \leq 1 - \frac{1}{2\kappa + 2/(2\mu - 1)} \quad \text{if } r > 1. \quad (6.7)$$

Proof. We prove the proposition by showing that $\text{dist}(x^k, \bar{\mathcal{X}}) \rightarrow 0$ as $k \rightarrow \infty$, and (3.13) holds with

$$\eta = 1 - \frac{1}{\kappa + 1/(2\mu - 1)}. \quad (6.8)$$

Then the conclusions of the proposition follow directly from Theorem 4.

First observe that $\liminf_{k \rightarrow \infty} \varphi(x^k) = \varphi^*$ implies $\lim_{k \rightarrow \infty} \varphi(x^k) = \varphi^*$, since the sequence $\{\varphi(x^k)\}$ is monotonically decreasing by (6.3). By contradiction, suppose

$\text{dist}(x^k, \bar{\mathcal{X}}) \not\rightarrow 0$ as $k \rightarrow \infty$. Then there exists a subsequence $\{x^k\}_{\mathcal{K}}$, where $\mathcal{K}$ is an infinite subset of the set of all positive integers, such that

$$x^k \rightarrow \hat{x} \notin \bar{\mathcal{X}} \text{ as } k \rightarrow \infty, k \in \mathcal{K},$$

for some $\hat{x} \in \mathcal{X}$. Hence

$$\varphi(x^k) \rightarrow \varphi(\hat{x}) > \varphi^* \text{ as } k \rightarrow \infty, k \in \mathcal{K},$$

which contradicts the fact that $\lim_{k \rightarrow \infty} \varphi(x^k) = \varphi^*$. Therefore under the assumptions of the proposition, we have

$$\text{dist}(x^k, \bar{\mathcal{X}}) \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Next observe that (6.2) in the algorithm with $\varepsilon(x^k, w^j)$ given by (6.1) yields

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

or equivalently,

$$\begin{aligned} \tilde{\varphi}_k(z) &\geq \tilde{\varphi}_k(x^k) + \frac{1}{2c} |z - x^k| + \frac{1}{c} \langle x^k - x^{k+1}, z - x^k \rangle \\ &\quad - \kappa [\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(x^{k+1}) + (\frac{1}{2c} - \frac{\mu}{c}) |x^{k+1} - x^k|^2] \quad \forall z. \end{aligned}$$

Let $z = \bar{x}^{k+1} = \underset{x \in X}{\text{argmin}} \tilde{\varphi}_k(x)$. We have

$$\begin{aligned} \tilde{\varphi}_k(\bar{x}^{k+1}) &\geq \tilde{\varphi}_k(x^k) + \frac{1}{2c} |\bar{x}^{k+1} - x^k| + \frac{1}{c} \langle x^k - x^{k+1}, \bar{x}^{k+1} - x^k \rangle \\ &\quad - \kappa [\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(x^{k+1}) + (\frac{1}{2c} - \frac{\mu}{c}) |x^{k+1} - x^k|^2]. \end{aligned}$$

Therefore, with the notation $\tilde{\varphi}_k^* = \min_{x \in X} \tilde{\varphi}_k(x)$,

$$\begin{aligned} \tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^* &\leq \kappa [\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(x^{k+1}) + (\frac{1}{2c} - \frac{\mu}{c}) |x^{k+1} - x^k|^2] \\ &\quad - \frac{1}{2c} |\bar{x}^{k+1} - x^k| - \frac{1}{c} \langle x^k - x^{k+1}, \bar{x}^{k+1} - x^k \rangle. \end{aligned} \quad (6.9)$$

However

$$|\bar{x}^{k+1} - x^k|^2 + 2\langle x^k - x^{k+1}, \bar{x}^{k+1} - x^k \rangle = -|x^k - x^{k+1}|^2 + |\bar{x}^{k+1} - x^{k+1}|^2. \quad (6.10)$$

Substituting (6.10) in (6.9), we obtain

$$\begin{aligned} \tilde{\varphi}_k(x^k) - \tilde{\varphi}_k^* &\leq \kappa [\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(x^{k+1}) + (\frac{1-2\mu}{2c})|x^{k+1} - x^k|^2] \\ &\quad + \frac{1}{2c}|x^{k+1} - x^k| - \frac{1}{2c}|\bar{x}^{k+1} - x^{k+1}|. \end{aligned}$$

Hence, for $\mu \in (\frac{1}{2}, 1)$, there holds

$$\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(\bar{x}^{k+1}) \leq \kappa [\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(x^{k+1})] + \frac{1}{2c}|x^{k+1} - x^k|^2. \quad (6.11)$$

On the other hand,

$$\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(x^{k+1}) + (\frac{1-2\mu}{2c})|x^{k+1} - x^k|^2 = \varphi(x^k) - \varphi(x^{k+1}) - \frac{\mu}{c}|x^{k+1} - x^k|^2 \geq 0$$

by (6.3). Hence

$$\frac{1}{2\mu-1}(\tilde{\varphi}_k(x^k) - \tilde{\varphi}_k(x^{k+1})) \geq \frac{1}{2c}|x^{k+1} - x^k|^2. \quad (6.12)$$

Substituting (6.12) in (6.11), we obtain

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

Then

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

or equivalently,

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

Therefore the condition (3.13) in Theorem 4 holds with η given by (6.8). $\square$

Acknowledgments. After this paper was written, the author became aware of A. N. Iusem and B. F. Svaiter's paper in press by RAIRO with the title "A proximal regularization of the steepest descent method," where a regularized objective function was used in the line search of the steepest descent method. Iusem and Svaiter were able to prove the convergence of their method for pseudo-convex objective functions without requiring the boundedness of the level sets. However, their paper presents no rate of convergence results. In contrast, the example in Section 4 of the present paper with $m = 1$ shows that their method has a linear rate when the objective function satisfies the conditions in this paper.

References.

- [1] M. S. Bazaraa and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, Wiley, 1979.
- [2] R. Correa and C. Lemarchal, "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] F. J. Luque, "Asymptotic convergence analysis of the proximal point algorithm," *SIAM J. Control Opt.* 22 (1984), pp. 277-293.
- [5] B. Martinet, "Regularisation d'inéquations variationnelles par approximations successives," *Rev. Francaise Inf. Rech. Oper.*, 1970, pp. 154-159.
- [6] B. Martinet, "Détermination approchée d'un point fixe d'une application pseudo-contractante," *C.R. Acad. Sci. Paris* 274 (1972), pp. 163-165.
- [7] G. J. Minty, "Monotone (nonlinear) operators in Hilbert space," *Duke Math. J.* 29 (1962), pp. 341-346.
- [8] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York 1970.
- [9] S. M. Robinson, "Some continuity properties of polyhedral multifunctions," *Math. Programming Studies* 14 (1981), pp. 203-214.
- [10] R. T. Rockafellar, "On the maximality of sums of nonlinear monotone operators," *Trans. Amer. Math. Soc.* 149 (1970), pp. 75-88.
- [11] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, N.J., 1970.
- [12] R. T. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM J. Control Opt.* 14 (1976), pp. 877-898.
- [13] R. T. Rockafellar, "Linear-quadratic programming and optimal control," *SIAM J. Control Opt.* 25 (1987), pp. 781-814.
- [14] R. T. Rockafellar, "Computational schemes for solving large-scale problems in extended linear-quadratic programming," *Math. Programming* 48 (1990), pp.

447-474.

- [15] 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.
- [16] C. Zhu, "On the primal-dual steepest descent algorithm for extended linear-quadratic programming," *SIAM J. Optimization*, accepted.
- [17] C. Zhu, "Solving large scale minimax problems with the primal-dual steepest descent algorithm," *Mathematical Programming*, accepted.
- [18] C. Zhu, "Asymptotic convergence analysis of the forward-backward splitting algorithm," *Mathematics of Operations Research*, accepted.
- [19] C. Zhu and R. T. Rockafellar, "Primal-dual projected gradient algorithms for extended linear-quadratic programming," *SIAM J. Optimization* 3 (1993), pp. 751-783.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.