An accelerated inexact proximal point algorithm for convex minimization

Bingsheng He* and Xiaoming Yuan †

September 17, 2010

Abstract. The proximal point algorithm (PPA) is classical and popular in the community of Optimization. In practice, inexact PPAs which solves the involved proximal subproblems approximately subject to certain inexact criteria are truly implementable. In this paper, we first propose an inexact PPA with a new inexact criterion for solving convex minimization, and show that the iteration-complexity of this inexact PPA is $O(1/k)$. Then, we show that this inexact PPA is eligible for being accelerated by some influential acceleration schemes proposed by Nesterov. Accordingly, an accelerated inexact PPA with the convergence rate $O(1/k^2)$ is proposed.

Keywords. Convex minimization, proximal point algorithm, inexact, acceleration.

1 Introduction

In this paper, we consider the following convex minimization problem:

$$\min \{f(x) \mid x \in \Omega\}, \tag{1.1}$$

where $f : R^n \to R \cup \{\infty\}$ is a proper and differentiable convex function, and Ω is a convex closed set in R^n. Throughout we assume that the solution set of (1.1) denoted by Ω^* is not empty. We assume additionally that the differential of $f(x)$ (denoted by $\nabla f(x)$) is Lipschitz continuous, *i. e.*, there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall\, x,\, y \in R^n.$$

Let $\lambda > 0$ be a scalar and

$$J_\lambda(x) := \text{Argmin}\,\{f(z) + \frac{1}{2\lambda}\|z - x\|^2 \,\big|\, z \in \Omega\}, \tag{1.2}$$

be the proximal mapping defined in [8]. Then, for solving (1.1), the iterative scheme of the classical proximal point algorithm (PPA) which was originally developed in [6] and concretely popularized in [11] is

$$x^{k+1} = J_{\lambda_k}(x^k) := \text{Argmin}\,\{f(z) + \frac{1}{2\lambda_k}\|z - x^k\|^2 \,\big|\, z \in \Omega\}, \tag{1.3}$$

where the positive numbers $\{\lambda_k\}$ are proximal parameters. Throughout, we assume that $\{\lambda_k\}$ is non-decreasing.

*Department of Mathematics and National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China. This author was supported by the NSFC Grant 10971095 and the NSF of Jiangsu Province Grant BK2008255. Email: hebma@nju.edu.cn

†Department of Mathematics, Hong Kong Baptist University, Hong Kong, China. This author was supported in part by HKRGC 203009. Email: xmyuan@hkbu.edu.hk

The exact version of PPA (1.3) requires to solve exactly the proximal subproblem (1.3) at each iteration, which can be as difficult as solving the original problem (1.1). In [11], Rockafellar showed that the subproblem (1.3) can be practically alleviated to

$$x^{k+1} \approx J_{\lambda_k}(x^k), \tag{1.4}$$

whose convergence is ensured whenever the accuracy of (1.4) is subject to the criteria:

$$\|x^{k+1} - J_{\lambda_k}(x^k)\| \leq \varepsilon_k \quad \text{with} \quad \sum_{k=0}^{\infty} \varepsilon_k < \infty \tag{1.5}$$

or

$$\|x^{k+1} - J_{\lambda_k}(x^k)\| \leq \varepsilon_k \|x^k - x^{k+1}\| \quad \text{with} \quad \sum_{k=0}^{\infty} \varepsilon_k < \infty, \tag{1.6}$$

where $\{\varepsilon_k\}$ should be a sequence of positive numbers. This concrete development on inexact PPA [11] has inspired many other articles in the literature to consider more practical inexact criteria with the purpose of avoiding the computation of $J_{\lambda_k}(x^k)$. We refer to some nice articles, e.g. [2, 3, 4, 7, 11], for convergence analysis of PPA,

The estimation of the convergence rate of PPA has also been addressed in the literature. In [4], the global convergence rate of the exact PPA (1.3) (where $\Omega = R^n$ and f is a proper and lower semicontinuous convex function) was estimated in terms of the objective residual:

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

from which the iteration-complexity $O(1/k)$ is instantly implied provided that $\{\lambda_k\}$ is chosen to be $\lambda_k \geqq \lambda > 0$. Then, in [5], Güler showed that the exact PPA can be accelerated by some acceleration schemes proposed in [9], and thus an accelerated exact PPA with the iteration-complexity $O(1/k^2)$ was proposed. In addition, in [5], Güler proposed some variants of PPAs and discussed the applications of Nesterov's acceleration schemes for these new PPAs. More specifically, let $\Omega := R^n$ in (1.1); $\{y_k\}$ be an auxiliary sequence generated in the spirit of the acceleration technique in [10]; and let

$$J_{\lambda_k}(y_k) := \text{Argmin}\,\{f(z) + \frac{1}{2\lambda_k}\|z - y^k\|^2 \,\big|\, z \in \Omega\}. \tag{1.7}$$

If the new iterate x^{k+1} is generated by

$$x^{k+1} = J_{\lambda_k}(y_k); \tag{1.8}$$

then the global convergence rate of the new exact PPA (1.8) was estimated in terms of the objective residual:

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

from which the iteration-complexity $O(1/k^2)$ is implied if $\{\lambda_k\}$ is chosen to be $\lambda_k \geqq \lambda > 0$. It was also shown in [5] that the proximal subproblem (1.8) can be performed inexactly:

$$x^{k+1} \approx J_{\lambda_k}(y_k) \tag{1.9}$$

subject to

$$\|x^{k+1} - J_{\lambda_k}(y^k)\| \leq \varepsilon_k \quad \text{with} \quad \sum_{k=0}^{\infty} \varepsilon_k < \infty. \tag{1.10}$$

Obviously, the inexact criterion (1.10) is a straightforward extension of the earliest one (1.5). If there exists a constant $M > 0$ such that

$$\lambda_i \leq M\lambda_j \quad \text{whenever } i \leq j;$$

and for some $\sigma > 0$ such that

$$\varepsilon_k = O(1/k^{\sigma}),\ k = 0, 1, 2, \cdots;$$

then the global convergence rate of the inexact PPA (1.9) subject to (1.10) was estimated in [5]:

$$f(x_k) - \min_{x \in R^n} f(x) \le O(\frac{1}{k^2}) + O(\frac{1}{k^{2\sigma-1}}).$$

The iteration-complexity of $O(1/k^2)$ is thus obtained whenever $\sigma \ge 3/2$.

Note that the applicability of the inexact criterion (1.10) in practice is prohibitively limited due to (a): the expensiveness or unavailability of $J_{\lambda_k}(y^k)$; (b): the summable requirement on ε_k essentially requires increasing accuracy for solving the subproblems; (c): the accuracy of solving the subproblems are controlled by absolute errors, rather than relative errors which are more likely to induce attractive numerical performance, as shown widely in the literature. Thus, we are inspired to develop an accelerated inexact PPA whose convergence rate is the same $O(1/k^2)$ as in [5], while its inexact criterion for executing (1.9) is more implementable than (1.10). This is the aim of the paper.

In the following, we first propose a new inexact criterion for performing (1.9) and give some remarks in Section 2. Then, in Section 3, we first derive an inexact PPA for (1.1) where the inexact proximal subproblem (1.9) is implemented subject to the new inexact criterion, and then show that the iteration-complexity of this new inexact PPA is $O(1/k)$. In Section 4, we show that the new inexact PPA can be accelerated by Nesterov's acceleration schemes and thus an accelerated inexact PPA with the convergence rate $O(1/k^2)$ is proposed. Finally, some conclusions are given in Section 5.

2 A new inexact criterion

In this section, we propose a new inexact criterion, which is more implementable than (1.10), for performing the inexact proximal subproblem (1.9). This new criterion is the essential difference of our methods to be presented from the PPAs in [5].

As we have mentioned, to propose accelerated inexact PPAs with the convergence rate $O(1/k^2)$, an auxiliary sequence $\{y^k\}$ should be generated by some Nesterov's acceleration schemes (see [5]). With the auxiliary sequence $\{y^k\}$, the proximal subproblem at each iteration of this type methods is (1.8). In views of the optimality condition, it is easy to verify that solving (1.8) amounts to solving the following projection equation

$$x = P_{\Omega}[y^k - \lambda_k \nabla f(x)], \tag{2.1}$$

where P_{Ω} denotes the projection operator onto Ω under the Euclidean distance. Thus, an inexact PPA is to seek an approximate solution of (1.8), denoted by z^{k+1}, such that

$$z^{k+1} \approx P_{\Omega}[y^k - \lambda_k \nabla f(z^{k+1})]. \tag{2.2}$$

Let $\{z^k\}$ be such a sequence satisfies (2.2) (to be specified later). We denote

$$x^{k+1} := P_{\Omega}[y^k - \lambda_k \nabla f(z^{k+1})]. \tag{2.3}$$

Then, it is clear that x^{k+1} is the exact solution of the subproblem (1.8) if $x^{k+1} = z^{k+1}$ or $\nabla f(x^{k+1}) = \nabla f(z^{k+1})$.

We are now ready to present our new inexact criterion to perform the inexact proximal subproblem (1.9).

A new inexact criterion

For given $y^k \in R^n$ and $\lambda_k > 0$, let z^{k+1} be given by (2.2) and x^{k+1} be given by (2.3). We require that

$$(z^{k+1} - x^{k+1})^T(\nabla f(z^{k+1}) - \nabla f(x^{k+1})) \le \frac{1}{2\lambda_k}\|y^k - x^{k+1}\|^2. \tag{2.4}$$

The condition (2.4) is the acceptance condition for generating an approximate solution of the proximal subproblem (1.8) by the inexact PPAs to be proposed. We call such an iterate x^{k+1} satisfying (2.4) an acceptance vector.

Remark 2.1. As the proximal subproblem (1.3) is a strongly convex problem, such a sequence $\{z^k\}$ satisfying (2.2) is ensured by many existing methods. In particular, if $\lambda_k \le 1/2L$ where L is the Lipschitz constant of $\nabla f(x)$, we can easily take $z^{k+1} = y^k$. By doing so, the inexact criterion (2.4) is met:

$$\begin{aligned}(y^k - x^{k+1})^T \lambda_k(\nabla f(y^k) - \nabla f(x^{k+1})) &\le \|y^k - x^{k+1}\| \cdot \frac{1}{2L}\|\nabla f(y^k) - \nabla f(x^{k+1})\| \\ &\le \frac{1}{2}\|y^k - x^{k+1}\|^2.\end{aligned}$$

Remark 2.2. The inexact PPAs in [5, 11] take $x^{k+1} = z^{k+1}$. In the case of $\Omega = R^n$, (2.2) implies that

$$\xi^{k+1} := \lambda_k f(z^{k+1}) + (z^{k+1} - y^k) \approx 0. \tag{2.5}$$

In [11] (see Page 880) and [5] (see Definition 3.1 in Page 656 and Page 660), the inexact criterion is set as:

$$\|\xi^k\| \le \varepsilon_k \qquad \text{and} \qquad \sum_{k=0}^{\infty} \varepsilon_k < \infty.$$

The difference of the proposed inexact criterion is that we take

$$x^{k+1} = y^k - \lambda_k f(z^{k+1}).$$

Use the notation of ξ^k (2.5), we have

$$x^{k+1} = z^{k+1} - \xi^{k+1},$$

and thus our criterion (2.4) becomes

$$(\xi^{k+1})^T \lambda_k(\nabla f(x^{k+1} + \xi^{k+1}) - \nabla f(x^{k+1})) \le \frac{1}{2}\|y^k - x^{k+1}\|^2.$$

Remark 2.3. Since we assume that ∇f is Lipschitz continuous, the proposed inexact criterion (2.4) is guaranteed if we ensure that

$$\|\xi^{k+1}\| \le \sqrt{\frac{1}{2\lambda_k L}} \cdot \|y^k - x^{k+1}\|.$$

3 An inexact PPA

In this section, we present an inexact PPA whose proximal subproblems are solved subject to the new inexact criterion (2.4). Then, we prove the convergence of the new inexact PPA and show that the iteration-complexity of this inexact PPA is $O(1/k)$.

Before presenting the inexact PPA, we first prove a property which play important roles in the proof of convergence and the estimation of the convergence rate for the methods to be proposed.

Lemma 3.1. *For given y^k and $\lambda_k > 0$, let x^{k+1} be given by (2.3) subjected to the inexact criterion* (2.4)*. Then, we have*

$$2\lambda_k(f(x) - f(x^{k+1})) \ge \|y^k - x^{k+1}\|^2 + 2(x - y^k)^T(y^k - x^{k+1}), \quad \forall\, x \in \Omega. \tag{3.1}$$

Proof. First, using the convexity of f, we have

$$f(x) \quad \ge \quad f(z^{k+1}) + (x - z^{k+1})^T \nabla f(z^{k+1}). \tag{3.2}$$

Using the convexity of f again and by a manipulation, we have

$$\begin{aligned} f(x^{k+1}) &\le f(z^{k+1}) + (x^{k+1} - z^{k+1})^T \nabla f(x^{k+1}) \\ &= f(z^{k+1}) + (x^{k+1} - z^{k+1})^T \nabla f(z^{k+1}) \\ &\quad + (x^{k+1} - z^{k+1})^T \big(\nabla f(x^{k+1}) - \nabla f(z^{k+1})\big) \\ &\le f(z^{k+1}) + (x^{k+1} - z^{k+1})^T \nabla f(z^{k+1}) + \frac{1}{2\lambda_k}\|y^k - x^{k+1}\|^2. \end{aligned} \tag{3.3}$$

The last inequality is due to the acceptance acceptance condition (2.4). It follows from (3.2) and (3.3) that

$$\begin{aligned} f(x) - f(x^{k+1}) &\ge f(z^{k+1}) + (x - z^{k+1})^T \nabla f(z^{k+1}) \\ &\quad - \big(f(z^{k+1}) + (x^{k+1} - z^{k+1})^T \nabla f(z^{k+1}) + \frac{1}{2\lambda_k}\|y^k - x^{k+1}\|^2\big) \\ &= (x - x^{k+1})^T \nabla f(z^{k+1}) - \frac{1}{2\lambda_k}\|y^k - x^{k+1}\|^2. \end{aligned} \tag{3.4}$$

On the other hand, since x^{k+1} is the projection of $[y^k - \lambda_k \nabla f(z^{k+1})]$ on Ω (see (2.3)), it follows that

$$(x - x^{k+1})^T \{[y^k - \lambda_k \nabla f(z^{k+1})] - x^{k+1}\} \le 0, \quad \forall\, x \in \Omega,$$

from which we obtain

$$(x - x^{k+1})^T \lambda_k \nabla f(z^{k+1}) \ge (x - x^{k+1})^T (y^k - x^{k+1}), \quad \forall\, x \in \Omega. \tag{3.5}$$

For the first term of the right-hand side of (3.4), it follows from (3.5) that

$$(x - x^{k+1})^T \nabla f(z^{k+1}) \ge \frac{1}{\lambda_k}(x - x^{k+1})^T (y^k - x^{k+1}), \quad \forall x \in \Omega. \tag{3.6}$$

Substituting (3.6) in (3.4), we obtain

$$\begin{aligned} f(x) - f(x^{k+1}) &\ge \frac{1}{\lambda_k}(x - x^{k+1})^T (y^k - x^{k+1}) - \frac{1}{2\lambda_k}\|y^k - x^{k+1}\|^2 \\ &= \frac{1}{2\lambda_k}\|y^k - x^{k+1}\|^2 + \frac{1}{\lambda_k}(x - y^k)^T (y^k - x^{k+1}), \end{aligned}$$

and the assertion of this lemma is proved. □

In the following, we present the new inexact PPA where the approximate solutions of the proximal subproblems are subject to the proposed inexact criterion (2.4) and the auxiliary sequence of $\{y^k\}$ can be avoided by simply taking $y^k \equiv x^k$.

Algorithm 1: An inexact PPA for (1.1)
Step 0. Take $x^0 \in \Omega$.
Step k. $(k \ge 0)$ $x^{k+1} = P_\Omega[x^k - \lambda_k f(z^{k+1})]$ where z^{k+1} satisfies (2.2) and the inexact criterion (2.4) is satisfied.

The following theorem states the global convergence of the proposed Algorithm 1.

Theorem 3.2. *Let $\{x^k\}$ be generated by the proposed Algorithm 1. Then, we have*

$$f(x^{k+1}) \le f(x^k) - \frac{1}{2\lambda_k}\|x^k - x^{k+1}\|^2, \tag{3.7}$$

and

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 - 2\lambda_k (f(x^{k+1}) - f(x^*)). \tag{3.8}$$

Thus, $\{x^k\}$ is convergent.

Proof. By using (3.1) for $x = x^k$ and $y^k = x^k$, we obtain the first assertion of this theorem. By setting $y^k = x^k$ and $x = x^*$, we have

$$2\lambda_k(f(x^*) - f(x^{k+1})) \geq \|x^k - x^{k+1}\|^2 - 2(x^k - x^*)^T(x^k - x^{k+1}).$$

Applying the relation

$$\|a - b\|^2 - 2(a - c)^T(a - b) = \|b - c\|^2 - \|a - c\|^2,$$

with

$$a = x^k, \quad b = x^{k+1} \quad \text{and} \quad c = x^*,$$

to the last inequality, we get

$$2\lambda_k(f(x^*) - f(x^{k+1})) \geq \|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2.$$

The second assertion follows from the above inequality directly. From (3.8) we have

$$\sum_{l=0}^{k-1} 2\lambda_l\big(f(x^l) - f(x^*)\big) \leq \|x^0 - x^*\|^2 - \|x^k - x^*\|^2 \leq \|x^0 - x^*\|^2,$$

and thus

$$\lim_{k\to\infty} \big(f(x^k) - f(x^*)\big) = 0.$$

Combining the fact that $\{x^k\}$ is bounded, the sequence $\{x^k\}$ converges to a solution point. □

In the following, we show that the iteration-complexity of the proposed Algorithm 1 is $O(1/k)$.

Theorem 3.3. *Let $\{x^k\}$ be generated by the proposed Algorithm 1. Then, we have*

$$f(x^k) - f(x^*) \leq \frac{\|x^0 - x^*\|^2}{2k\lambda_0}, \quad \forall x^* \in \Omega^*, \ \forall k \geq 1. \tag{3.9}$$

Proof. Because $y^k = x^k$, it follows from (3.8) that, for all $l \geq 0$, we have

$$2\lambda_l(f(x^*) - f(x^{l+1})) \geq \|x^{l+1} - x^*\|^2 - \|x^l - x^*\|^2, \quad \forall x^* \in \Omega^*.$$

Using the fact that $f(x^*) - f(x^l) \leq 0$ and summing the above inequality over $l = 0, \ldots, k-1$, we obtain

$$2\lambda_0\Big(kf(x^*) - \sum_{l=0}^{k-1} f(x^{l+1})\Big) \geq \|y^k - x^*\|^2 - \|x^0 - x^*\|^2. \tag{3.10}$$

It follows from (3.7) that

$$2\lambda_0\big(f(x^l) - f(x^{l+1})\big) \geq \frac{\lambda_0}{\lambda_l}\|x^l - x^{l+1}\|^2.$$

Multiplying the last inequality by l and summing over $l = 0, \ldots, k-1$, it follows that

$$2\lambda_0 \sum_{l=0}^{k-1}\Big(lf(x^l) - (l+1)f(x^{l+1}) + f(x^{l+1})\Big) \geq \sum_{l=0}^{k-1} \frac{\lambda_0}{\lambda_l} l\|x^l - x^{l+1}\|^2,$$

which simplifies to

$$2\lambda_0\Big(-kf(x^k) + \sum_{l=0}^{k-1} f(x^{l+1})\Big) \geq \sum_{l=0}^{k-1} \frac{\lambda_0}{\lambda_l} l\|x^l - x^{l+1}\|^2. \tag{3.11}$$

Adding (3.10) and (3.11), we get

$$2k\lambda_0\big(f(x^*) - f(x^k)\big) \geq \|x^k - x^*\|^2 - \|x^0 - x^*\|^2 + \sum_{l=0}^{k-1} \frac{\lambda_0}{\lambda_l} l\|x^l - x^{l+1}\|^2,$$

and hence it follows that

$$f(x^k) - f(x^*) \leq \frac{\|x^0 - x^*\|^2}{2k\lambda_0}.$$

The proof is complete. □

4 An accelerated inexact PPA

In this section, we accelerate the proposed Algorithm 1 with the acceleration scheme in [9]. Thus, an accelerated inexact PPA with the iteration-complexity $O(1/k^2)$ for solving (1.1) is proposed.

Algorithm 2: An accelerated inexact PPA
Step 0. Take $\lambda > 0$, $x^1 \in R^n$. Set $y^1 = x^1$, $t_1 = 1$.
Step k. ($k \geq 1$) For given y^k, let $x^{k+1} = P_\Omega[x^k - \lambda_k f(z^{k+1})]$ where z^{k+1} satisfies (2.2) and the inexact criterion (2.4) is satisfied.
Set

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \tag{4.1a}$$

and

$$y^{k+1} = x^{k+1} + \Big(\frac{t_k - 1}{t_{k+1}}\Big)(x^{k+1} - x^k). \tag{4.1b}$$

To derive the iteration-complexity of the proposed Algorithm 2, we need to prove some properties of the corresponding sequence.

Lemma 4.1. *The sequences $\{x^k\}$ and $\{y^k\}$ generated by the proposed Algorithm 2 satisfy*

$$2t_k^2 v_k - 2t_{k+1}^2 v_{k+1} \geq \frac{1}{\lambda_{k+1}}\|u^{k+1}\|^2 - \frac{1}{\lambda_{k+1}}\|u^k\|^2, \ \forall k \geq 1, \tag{4.2}$$

where $v_k := f(x^{k+1}) - f(x^*)$ and $u^k := t_k x^{k+1} - (t_k - 1)x^k - x^*$.

Proof. By using Lemma 3.1 for $k+1$, $x = x^k$ and $x = x^*$ we get

$$2\lambda_{k+1}\big(f(x^{k+1}) - f(x^{k+2})\big) \geq \|y^{k+1} - x^{k+2}\|^2 + 2(x^{k+1} - y^{k+1})^T(y^{k+1} - x^{k+2}),$$

and

$$2\lambda_{k+1}\big(f(x^*) - f(x^{k+2})\big) \geq \|y^{k+1} - x^{k+2}\|^2 + 2(x^* - y^{k+1})^T(y^{k+1} - x^{k+2}).$$

Using the definition of v_k, we get

$$2\lambda_{k+1}(v_k - v_{k+1}) \geq \|y^{k+1} - x^{k+2}\|^2 + 2(x^{k+1} - y^{k+1})^T(y^{k+1} - x^{k+2}), \tag{4.3}$$

and

$$-2\lambda_{k+1}v_{k+1} \geq \|y^{k+1} - x^{k+2}\|^2 + 2(x^{k+1} - y^{k+1})^T(y^{k+1} - x^{k+2}). \tag{4.4}$$

To get a relation between v_k and v_{k+1}, we multiply (4.3) by $(t_{k+1} - 1)$ and add it to (4.4):

$$\begin{aligned}&2\lambda_{k+1}\big((t_{k+1} - 1)v_k - t_{k+1}v_{k+1}\big)\\ &\quad\geq\ t_{k+1}\|x^{k+2} - y^{k+1}\|^2 + 2(x^{k+2} - y^{k+1})^T\big(t_{k+1}y^{k+1} - (t_{k+1} - 1)x^{k+1} - x^*\big).\end{aligned}$$

Multiplying the last inequality by t_{k+1} and using

$$t_k^2 = t_{k+1}^2 - t_{k+1} \qquad \big(\text{and thus} \quad t_{k+1} = (1 + \sqrt{1 + 4t_k^2})/2 \quad \text{as in} \quad (4.1a)\big),$$

which yields

$$\begin{aligned}&2\lambda_{k+1}\big(t_k^2 v_k - t_{k+1}^2 v_{k+1}\big)\\ &\quad\geq\ \|t_{k+1}(x^{k+2} - y^{k+1})\|^2 + 2t_{k+1}(x^{k+2} - y^{k+1})^T\big(t_{k+1}y^{k+1} - (t_{k+1} - 1)x^{k+1} - x^*\big).\end{aligned}$$

Applying the relation

$$\|b - a\|^2 + 2(b - a)^T(a - c) = \|b - c\|^2 - \|a - c\|^2$$

to the right-hand side of the last inequality with

$$a := t_{k+1}y^{k+1}, \qquad b := t_{k+1}x^{k+2}, \qquad c := (t_{k+1} - 1)x^{k+1} + x^*,$$

we get

$$2\lambda_{k+1}(t_k^2 v_k - t_{k+1}^2 v_{k+1}) \geq \|t_{k+1}x^{k+2} - (t_{k+1}-1)x^{k+1} - x^*\|^2 - \|t_{k+1}y^{k+1} - (t_{k+1}-1)x^{k+1} - x^*\|^2.$$

In order to write the above inequality in the form (4.2) with $u^k = t_k x^{k+1} - (t_k - 1)x^k - x^*$, we need only to set

$$t_{k+1}y^{k+1} - (t_{k+1}-1)x^{k+1} - x^* = t_k x^{k+1} - (t_k - 1)x^k - x^*.$$

From the last equality we obtain

$$y^{k+1} = x^{k+1} + \Big(\frac{t_k - 1}{t_{k+1}}\Big)(x^{k+1} - x^k).$$

This is just the form (4.1b) in the accelerated multi-step version of the APPA . □

Since we have assumed that the sequence $\{\lambda_k\}$ is non-decreasing, it follows from (4.2) that

$$2t_k^2 v_k - 2t_{k+1}^2 v_{k+1} \geq \frac{1}{\lambda_{k+1}}\|u^{k+1}\|^2 - \frac{1}{\lambda_k}\|u^k\|^2, \quad \forall\, k \geq 1. \tag{4.5}$$

To proceed the proof of the main theorem, we need the following Lemma 4.2 and Lemma 4.3, which have also been considered in [1]. We omit their proofs as they are trivial.

Lemma 4.2. *Let $\{a_k\}$ and $\{b_k\}$ be positive sequences of reals satisfying*

$$a_k - a_{k+1} \geq b_{k+1} - b_k \quad \forall\, k \geq 1.$$

Then, $a_k \leq a_1 + b_1$ for every $k \geq 1$.

Lemma 4.3. *The positive sequence $\{t_k\}$ generated by*

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \qquad \text{with} \qquad t_1 = 1$$

satisfies

$$t_k \geq \frac{k+1}{2}, \quad \forall\, k \geq 1.$$

Now, we are ready to show that the proposed Algorithm 2 is convergent with the rate $O(1/k^2)$.

Theorem 4.4. *Let $\{x^k\}$ and $\{y^k\}$ be generated by the proposed Algorithm 2. Then, for any $k \geq 1$, we have*

$$f(x^k) - f(x^*) \leq \frac{2\|x^1 - x^*\|^2}{\lambda_1 k^2}, \;\; \forall\, x^* \in \Omega^*. \tag{4.6}$$

Proof. Let us define the quantities

$$a_k := 2t_k^2 v_k, \qquad b_k := \frac{1}{\lambda_k}\|u^k\|^2.$$

By using Lemma 4.1 and Lemma 4.2, we obtain

$$2t_k^2 v_k \leq a_1 + b_1,$$

which combined with the definition v_k and $t_k \geq (k+1)/2$ (by Lemma 4.3) yields

$$f(x^{k+1}) - f(x^*) = v_k \leq \frac{2(a_1 + b_1)}{(k+1)^2}. \tag{4.7}$$

Since $t_1 = 1$, and using the definition of u_k given in Lemma 4.1, we have

$$\lambda_1 a_1 = 2\lambda_1 t_1^2 v_1 = 2\lambda_1 v_1 = 2\lambda_1\big(f(x^2) - 2f(x^*)\big), \qquad \lambda_1 b_1 = \|u^1\|^2 = \|x^2 - x^*\|.$$

Setting $x = x^*$ and $k = 1$ in (3.1), we have

$$2\lambda_1(f(x^2) - f(x^*)) \leq 2(y^1 - x^*)^T(y^1 - x^2) - \|y^1 - x^2\|^2 = \|y^1 - x^*\|^2 - \|x^2 - x^*\|^2.$$

Therefore, we have

$$\begin{aligned}\lambda_1(a_1 + b_1) &= 2\lambda(f(x^2) - f(x^*)) + \|x^2 - x^*\|^2 \\ &\leq \|y^1 - x^*\|^2 - \|x^2 - x^*\|^2 + \|x^2 - x^*\|^2 \\ &= \|x^1 - x^*\|^2.\end{aligned}$$

Substituting it in (4.7), the assertion is proved. □

Based on Theorem 4.4, for obtaining an ε-optimal solution (denoted by $\tilde{x}$) in the sense that $f(\tilde{x}) - f(x^*) \leq \varepsilon$, the number of iterations required by the proposed Algorithm 2 is at most $\lceil C/\sqrt{\varepsilon} - 1\rceil$ where $C = 2\|x^1 - x^*\|^2/\lambda$.

5 Conclusions

In this paper we mainly show that the influential acceleration schemes of Nesterov's can be applied to accelerate some inexact variants of the classical proximal point algorithm (PPA) with implementable inexact criteria. As a result, an accelerated inexact PPA with the convergence rate $O(1/k^2)$ is yielded. We are thus inspired to consider the possibility of accelerating some other methods which are closely related to PPA, e.g. the augmented Lagrangian method.

References

[1] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Science, 2 (2009), pp. 183-202.

[2] J. V. Burke and M. J. Qian, A variable metric proximal point algorithm for monotone operators, SIAM J. Control and Optimization, 37 (1998), pp. 353-375.

[3] J. Eckstein, Approximate iterations in Bregman-function-based proximal algorithms, Math. Program., 83 (1998), pp. 113-123.

[4] O. Güler, On the convergence of the proximal point algorithm for convex minimization, SIAM J. Control Optim., 29(2)(1991), pp. 403-419.

[5] O. Güler, New proximal point algorithms for convex minimization, SIAM J. Optimi., 2 (1992), pp. 649-664.

[6] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, Rev. Francaise d'Inform. Recherche Oper., 4 (1970), pp. 154-159.

[7] R. D. C. Monteiro and B. F. Svaiter, Convergence rate of inexact proximal point methods with relative error criteria for convex optimization, manuscript, 2010.

[8] J. J. Moreau, Proximaté et dualité dans un espace Hilbertien, Bull. Soc. Math. France, 93 (1965), pp. 273-299.

[9] Y. E. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543-547.

[10] Y. E. Nesterov, On an approach to the construction of optimal methods ofminimization ofsmooth convexfunctions, Ekonom. Mat. Metody, 24 (1988), pp. 509-517.

[11] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim., 14 (1976), pp. 877-898.