J Optim Theory Appl (2012) 154:536–548
DOI 10.1007/s10957-011-9948-6

An Accelerated Inexact Proximal Point Algorithm for Convex Minimization

Bingsheng He · Xiaoming Yuan

Received: 23 September 2010 / Accepted: 19 October 2011 / Published online: 10 November 2011

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

Keywords Convex minimization · Proximal point algorithm · Inexact · Acceleration

1 Introduction

The proximal point algorithm (PPA) dates back to the work [1], and it was introduced to the optimization literature in [2], and then was promoted substantially in [3]. In this paper, we consider the application of PPA to convex minimization problems. More specifically, we propose to solve the PPA subproblems approximately subject to a new inexact criterion and show the $O(1/k)$ iteration-complexity of this new

Communicated by Jen-Chih Yao.

B. He
Department of Mathematics and National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210093, China
e-mail: hebma@nju.edu.cn

X. Yuan (✉)
Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong, China
e-mail: xmyuan@hkbu.edu.hk

inexact PPA. In addition, we demonstrate that the new inexact PPA can be accelerated by Nesterov's acceleration schemes and the iteration-complexity of the accelerated inexact PPA is $O(1/k^2)$.

The paper is organized as follows. In Sect. 2, we provide some background and illustrate our motivation. In Sect. 3, we propose a new inexact criterion for solving proximal subproblems approximately. Then a new inexact PPA with this inexact criterion is proposed in Sect. 4. The global convergence and the $O(1/k)$ iteration-complexity of this new inexact PPA are also proved in this section. In Sect. 5, we show that the new inexact PPA can be accelerated by Nesterov's acceleration schemes, and thus an accelerated inexact PPA with the $O(1/k^2)$ iteration-complexity is proposed. Finally, some conclusions are given in Sect. 6.

2 Motivation

We consider the following convex minimization problem:

$$\min \{ f(x) \mid x \in \Omega \}, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a proper closed convex function and Ω is a closed convex subset in $\mathbb{R}^n$. Throughout, we assume the solution set of (1) denoted by Ω^* to be nonempty.

Let $\lambda > 0$ be a scalar and

$$J_\lambda(x) := \operatorname{Argmin} \left\{ f(z) + \frac{1}{2\lambda} \|z - x\|^2 \mid z \in \Omega \right\}, \tag{2}$$

be the proximal operator defined in [1]. Then, for solving (1), the iterative scheme of PPA is

$$x^{k+1} = J_{\lambda_k}(x^k) := \operatorname{Argmin} \left\{ f(z) + \frac{1}{2\lambda_k} \|z - x^k\|^2 \mid z \in \Omega \right\}, \tag{3}$$

where the positive numbers $\{\lambda_k\}$ are proximal parameters. Throughout, the sequence $\{\lambda_k\}$ is assumed to be nondecreasing.

The exact version of PPA (3) requires to solve exactly the proximal subproblem (3) at each iteration, which can be as difficult as solving the original problem (1). In [3], Rockafellar showed that the subproblem (3) can be practically alleviated to

$$x^{k+1} \approx J_{\lambda_k}(x^k), \tag{4}$$

whose convergence is ensured whenever the accuracy of (4) is subject to the criterion

$$\|x^{k+1} - J_{\lambda_k}(x^k)\| \le \varepsilon_k \quad \text{with} \sum_{k=0}^{\infty} \varepsilon_k < \infty \tag{5}$$

or

$$\|x^{k+1} - J_{\lambda_k}(x^k)\| \le \varepsilon_k \|x^k - x^{k+1}\| \quad \text{with} \sum_{k=0}^{\infty} \varepsilon_k < \infty, \tag{6}$$

where $\{\varepsilon_k\}$ should be a sequence of positive numbers. This concrete development on inexact PPA [3] 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., [3–7] for convergence analysis of PPA.

The estimate of the convergence rate of PPA has also been addressed in the literature. In [6], the global convergence rate of the exact PPA (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 $O(1/k)$ iteration-complexity is instantly implied provided that $\{\lambda_k\}$ is chosen to be $\lambda_k \geq \lambda > 0$. Then, in [8], Güler showed that the exact PPA can be accelerated by some acceleration schemes proposed in [9], and thus an accelerated exact PPA with the $O(1/k^2)$ iteration-complexity was proposed. In addition, in [8], 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); $\{y_k\}$ be an auxiliary sequence generated in the spirit of the acceleration technique in [10]; and let

$$J_{\lambda_k}(y_k) := \operatorname{Argmin}\left\{ f(z) + \frac{1}{2\lambda_k}\left\|z - y^k\right\|^2 \mid z \in \Omega \right\}. \tag{7}$$

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

$$x^{k+1} = J_{\lambda_k}(y_k); \tag{8}$$

then the global convergence rate of the new exact PPA (8) was estimated in term 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 $O(1/k^2)$ iteration-complexity is implied if $\{\lambda_k\}$ is chosen to be $\lambda_k \geq \lambda > 0$. It was also shown in [8] that the proximal subproblem (8) can be performed inexactly:

$$x^{k+1} \approx J_{\lambda_k}(y_k) \tag{9}$$

subject to

$$\left\|x^{k+1} - J_{\lambda_k}\left(y^k\right)\right\| \leq \varepsilon_k \quad \text{with} \sum_{k=0}^{\infty} \varepsilon_k < \infty. \tag{10}$$

Obviously, the inexact criterion (10) is a straightforward extension of the earliest one (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

$$\varepsilon_k = O\big(1/k^\sigma\big), \quad k = 0, 1, 2, \ldots;$$

then the global convergence rate of the inexact PPA (9) subject to (10) was estimated in [8]:

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

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

The applicability of the inexact criterion (10) in practice may be 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 [8], while its inexact criterion for executing (9) is more implementable than (10). This is the aim of the paper.

3 A New Inexact Criterion

In this section, we propose a new inexact criterion for performing the inexact proximal subproblem (9). To yield an inexact criterion with easy operability, our discussion is under the additional assumption that $f(x)$ is differentiable and its differential (denoted by $\nabla f(x)$) is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

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

But the framework of the coming analysis can be extended to the nonsmooth case of $f(x)$.

As we have mentioned, to propose accelerated inexact PPAs with the $O(1/k^2)$ iteration-complexity, an auxiliary sequence $\{y^k\}$ should be generated by some Nesterov's acceleration schemes (see [8]). With the auxiliary sequence $\{y^k\}$, the proximal subproblem at each iteration of this type of methods is (8). In view of the optimality condition, it is easy to verify that solving (8) amounts to solving the following projection equation:

$$x = P_\Omega\big[y^k - \lambda_k \nabla f(x)\big], \tag{11}$$

where P_Ω denotes the projection operator onto Ω under the Euclidean distance. Thus, an inexact PPA is to seek an approximate solution of (8), denoted by z^{k+1}, such that

$$z^{k+1} \approx P_\Omega\big[y^k - \lambda_k \nabla f\big(z^{k+1}\big)\big]. \tag{12}$$

Let $\{z^k\}$ be a sequence satisfying (12) (the inexactness will be specified later). We denote

$$x^{k+1} := P_\Omega\big[y^k - \lambda_k \nabla f\big(z^{k+1}\big)\big]. \tag{13}$$

Then it is clear that x^{k+1} is the exact solution of the subproblem (8) if

$$x^{k+1} = z^{k+1}$$

or

$$\nabla f\left(x^{k+1}\right) = \nabla f\left(z^{k+1}\right).$$

We are now ready to present our new inexact criterion to perform the inexact proximal subproblem (9).

A new inexact criterion

For given $y^k \in \mathbb{R}^n$ and $\lambda_k > 0$, let z^{k+1} be given by (12) and x^{k+1} be given by (13). We require that

$$\left(z^{k+1} - x^{k+1}\right)^T \left(\nabla f\left(z^{k+1}\right) - \nabla f\left(x^{k+1}\right)\right) \le \frac{1}{2\lambda_k} \left\| y^k - x^{k+1} \right\|^2. \tag{14}$$

The condition (14) is the acceptance condition for generating an approximate solution of the proximal subproblem (8) by the inexact PPAs to be proposed. We call such an iterate x^{k+1} satisfying (14) an acceptance vector.

Remark 3.1 As the proximal subproblem (3) is a strongly convex problem, such a sequence $\{z^k\}$ satisfying (12) 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 (14)

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

is met.

Remark 3.2 The inexact PPAs in [3, 8] take $x^{k+1} = z^{k+1}$. In the case of $\Omega = R^n$, (12) implies that

$$\xi^{k+1} := \lambda_k f\left(z^{k+1}\right) + \left(z^{k+1} - y^k\right) \approx 0. \tag{15}$$

In [3] (see p. 880) and [8] (see Definition 3.1 in pp. 656 and 660), the inexact criterion is set as

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

The difference of the proposed inexact criterion is that we take

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

Using the notation of ξ^k (15), we have

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

and thus our criterion (14) becomes

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

Remark 3.3 Since ∇f is assumed to be Lipschitz continuous with the constant L, the proposed inexact criterion (14) is guaranteed if we ensure that

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

4 An Inexact PPA

In this section, we present an inexact PPA whose proximal subproblems are solved subject to the new inexact criterion (14). Then we prove the convergence of the new inexact PPA and show its $O(1/k)$ iteration-complexity. Before presenting the inexact PPA, we first prove a proposition which plays important roles in the coming analysis.

Lemma 4.1 *For given y^k and $\lambda_k > 0$, let x^{k+1} be given by (13) subjected to the inexact criterion (14). Then we have*

$$2\lambda_k\left(f(x) - f\left(x^{k+1}\right)\right) \ge \left\| y^k - x^{k+1} \right\|^2 + 2\left(x - y^k\right)^T \left(y^k - x^{k+1}\right), \quad \forall x \in \Omega. \tag{16}$$

Proof First, using the convexity of f, we have

$$f(x) \ge f\left(z^{k+1}\right) + \left(x - z^{k+1}\right)^T \nabla f\left(z^{k+1}\right). \tag{17}$$

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

$$\begin{aligned} f\left(x^{k+1}\right) &\le f\left(z^{k+1}\right) + \left(x^{k+1} - z^{k+1}\right)^T \nabla f\left(x^{k+1}\right) \\ &= f\left(z^{k+1}\right) + \left(x^{k+1} - z^{k+1}\right)^T \nabla f\left(z^{k+1}\right) \\ &\quad + \left(x^{k+1} - z^{k+1}\right)^T \left(\nabla f(x^{k+1}) - \nabla f\left(z^{k+1}\right)\right) \\ &\le f\left(z^{k+1}\right) + \left(x^{k+1} - z^{k+1}\right)^T \nabla f\left(z^{k+1}\right) + \frac{1}{2\lambda_k} \left\| y^k - x^{k+1} \right\|^2. \end{aligned} \tag{18}$$

The last inequality is due to the acceptance condition (14). It follows from (17) and (18) that

$$f(x) - f\left(x^{k+1}\right) \geq f\left(z^{k+1}\right) + \left(x - z^{k+1}\right)^T \nabla f\left(z^{k+1}\right) - \left(f\left(z^{k+1}\right) + \left(x^{k+1} - z^{k+1}\right)^T \nabla f\left(z^{k+1}\right) + \frac{1}{2\lambda_k}\left\|y^k - x^{k+1}\right\|^2\right) = \left(x - x^{k+1}\right)^T \nabla f\left(z^{k+1}\right) - \frac{1}{2\lambda_k}\left\|y^k - x^{k+1}\right\|^2. \tag{19}$$

On the other hand, since x^{k+1} is the projection of $[y^k - \lambda_k \nabla f(z^{k+1})]$ on Ω (see (13)), it follows that

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

from which we obtain

$$\left(x - x^{k+1}\right)^T \lambda_k \nabla f\left(z^{k+1}\right) \geq \left(x - x^{k+1}\right)^T \left(y^k - x^{k+1}\right), \quad \forall x \in \Omega. \tag{20}$$

For the first term of the right-hand side of (19), it follows from (20) that

$$\left(x - x^{k+1}\right)^T \nabla f\left(z^{k+1}\right) \geq \frac{1}{\lambda_k}\left(x - x^{k+1}\right)^T \left(y^k - x^{k+1}\right), \quad \forall x \in \Omega. \tag{21}$$

Substituting (21) in (19), we obtain

$$f(x) - f\left(x^{k+1}\right) \geq \frac{1}{\lambda_k}\left(x - x^{k+1}\right)^T \left(y^k - x^{k+1}\right) - \frac{1}{2\lambda_k}\left\|y^k - x^{k+1}\right\|^2 = \frac{1}{2\lambda_k}\left\|y^k - x^{k+1}\right\|^2 + \frac{1}{\lambda_k}\left(x - y^k\right)^T \left(y^k - x^{k+1}\right),$$

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 (14) 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)

Step 0. Take $x^0 \in \Omega$.
Step k. ($k \geq 0$) $x^{k+1} = P_\Omega[x^k - \lambda_k f(z^{k+1})]$ where z^{k+1} satisfies (12) and the inexact criterion (14) is satisfied.

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

Theorem 4.1 *Let $\{x^k\}$ be generated by the proposed Algorithm 1. Then we have*

$$f\left(x^{k+1}\right) \leq f\left(x^k\right) - \frac{1}{2\lambda_k}\left\|x^k - x^{k+1}\right\|^2, \tag{22}$$

and

$$\|x^{k+1}-x^*\|^2 \le \|x^k-x^*\|^2 - 2\lambda_k\big(f(x^{k+1})-f(x^*)\big), \quad \forall x^* \in \Omega^*. \tag{23}$$

Thus, $\{x^k\}$ *is convergent to a solution point of* (1).

Proof By using (16) for $x=x^k$ and $y^k=x^k$, we obtain the first assertion of this theorem. Let $x^* \in \Omega^*$ and set $y^k=x^k$ and $x=x^*$ in (16), we have

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

Applying the relation

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

with

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

to the last inequality, we get the assertion (22) immediately.

The second assertion follows from the above inequality directly. From (23), we have

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

and thus

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

Since $\{x^k\}$ is bounded, the sequence $\{x^k\}$ converges to $x^* \in \Omega^*$. □

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

Theorem 4.2 *Let* $\{x^k\}$ *be generated by the proposed Algorithm* 1. *Then we have*

$$f(x^k)-f(x^*) \le \frac{\|x^0-x^*\|^2}{2k\lambda_0}, \quad \forall x^* \in \Omega^*,\ \forall k \ge 1. \tag{24}$$

Proof Because $y^k=x^k$, it follows from (23) that, for all $l \ge 0$, we have

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

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

$$2\lambda_0\left(kf(x^*) - \sum_{l=0}^{k-1} f(x^{l+1})\right) \ge \|x^k-x^*\|^2 - \|x^0-x^*\|^2. \tag{25}$$

It follows from (22) that

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

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

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

which simplifies to

$$2\lambda_0\left(-kf\big(x^k\big)+\sum_{l=0}^{k-1}f\big(x^{l+1}\big)\right)\geq\sum_{l=0}^{k-1}\frac{\lambda_0}{\lambda_l}l\big\|x^l-x^{l+1}\big\|^2. \tag{26}$$

Adding (25) and (26), we get

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

and hence it follows that

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

The proof is complete. □

5 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 $O(1/k^2)$ iteration-complexity is proposed for solving (1).

Algorithm 2: An accelerated inexact PPA

Step 0. Let $\{\lambda_k\}$ be a nondecreasing and positive sequence and $x^1\in\mathbb{R}^n$. Set $y^1=x^1$, $t_1=1$.

Step k. ($k\geq 1$) For given y^k, $x^{k+1}=P_\Omega[y^k-\lambda_k f(z^{k+1})]$ where z^{k+1} satisfies (12) and the inexact criterion (14) is satisfied.

Set

$$t_{k+1}=\frac{1+\sqrt{1+4t_k^2}}{2}, \tag{27a}$$

and

$$y^{k+1}=x^{k+1}+\left(\frac{t_k-1}{t_{k+1}}\right)\big(x^{k+1}-x^k\big). \tag{27b}$$

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

Lemma 5.1 *The sequences* $\{x^k\}$ *and* $\{y^k\}$ *generated by the proposed Algorithm* 2 *satisfy*

$$2t_k^2 v_k - 2t_{k+1}^2 v_{k+1} \geq \frac{1}{\lambda_{k+1}}\|u^{k+1}\|^2 - \frac{1}{\lambda_{k+1}}\|u^k\|^2, \quad \forall k \geq 1, \tag{28}$$

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 4.1 for $k+1$, $x = x^k$ and $x = x^*$, we get

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

and

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

Using the definition of v_k, we get

$$2\lambda_{k+1}(v_k - v_{k+1}) \geq \|y^{k+1} - x^{k+2}\|^2 + 2\big(x^{k+1} - y^{k+1}\big)^T\big(y^{k+1} - x^{k+2}\big) \tag{29}$$

and

$$-2\lambda_{k+1}v_{k+1} \geq \|y^{k+1} - x^{k+2}\|^2 + 2\big(x^* - y^{k+1}\big)^T\big(y^{k+1} - x^{k+2}\big). \tag{30}$$

To get a relation between v_k and v_{k+1}, we multiply (29) by $(t_{k+1} - 1)$ and add it to (30):

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

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

$$t_k^2 = t_{k+1}^2 - t_{k+1} \quad \Big(\text{and thus } t_{k+1} = \big(1 + \sqrt{1 + 4t_k^2}\big)/2 \text{ as in (27a)}\Big),$$

which yields

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

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

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

In order to write the above inequality in the form (28) 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

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

This is just the accelerated step (27b) of the proposed accelerated inexact PPA. □

Since we have assumed that the sequence $\{\lambda_k\}$ is nondecreasing, it follows from (28) that

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

To proceed the proof of the main theorem, we need the following Lemmas 5.2 and 5.3, which have also been considered in [11]. We omit their proofs as they are trivial.

Lemma 5.2 *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 5.3 *The positive sequence $\{t_k\}$ generated by*

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

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 5.1 *Let $\{x^k\}$ and $\{y^k\}$ be generated by the proposed Algorithm 2. Then we have*

$$f\left(x^k\right) - f\left(x^*\right) \leq \frac{2\|x^1 - x^*\|^2}{\lambda_1 k^2}, \quad \forall x^* \in \Omega^*, \ \forall k \geq 1. \tag{31}$$

Proof Let us define the quantities

$$a_k := 2t_k^2 v_k \quad \text{and} \quad b_k := \frac{1}{\lambda_k}\|u^k\|^2.$$

By using Lemmas 5.1 and 5.2, we obtain

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

which, combining with the definition v_k and $t_k \ge (k+1)/2$ (by Lemma 5.3), yields

$$f(x^{k+1}) - f(x^*) = v_k \le \frac{2(a_1 + b_1)}{(k+1)^2}. \tag{32}$$

Since $t_1 = 1$, and using the definition of u_k given in Lemma 5.1, we have

$$\begin{aligned}
&\lambda_1 a_1 = 2\lambda_1 t_1^2 v_1 = 2\lambda_1 v_1 = 2\lambda_1\big(f(x^2) - 2f(x^*)\big) \quad \text{and}\\
&\lambda_1 b_1 = \|u^1\|^2 = \|x^2 - x^*\|^2.
\end{aligned}$$

Setting $x = x^*$ and $k = 1$ in (16), we have

$$\begin{aligned}
2\lambda_1\big(f(x^2) - f(x^*)\big) &\le 2(y^1 - x^*)^T(y^1 - x^2) - \|y^1 - x^2\|^2\\
&= \|y^1 - x^*\|^2 - \|x^2 - x^*\|^2.
\end{aligned}$$

Therefore, we have

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

Substituting it in (32), the assertion is proved. □

Based on Theorem 5.1, for obtaining an ε-optimal solution (denoted by $\tilde{x}$) in the sense that $f(\tilde{x}) - f(x^*) \le \varepsilon$, the number of iterations required by the proposed Algorithm 2 is at most $\lceil\sqrt{C/\varepsilon}\rceil$ where $C = 2\|x^1 - x^*\|^2/\lambda_1$. That is, the $O(1/k^2)$ iteration-complexity of Algorithm 2 is proved.

6 Conclusions

We show that Nesterov's acceleration schemes 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 $O(1/k^2)$ iteration-complexity is yielded. We are thus inspired to consider the possibility of accelerating some other methods which are related to PPA, e.g., the augmented Lagrangian method and the alternating direction method.

Acknowledgements The research was supported by the National Natural Science Foundation of China (NSFC) Grant 10971095, the NSF of Jiangsu Province Grant BK2008255, and an internal HKBU grant FRG2/10-11/069.

References

1. Moreau, J.J.: Proximaté et dualité dans un espace Hilbertien. Bull. Soc. Math. Fr. **93**, 273–299 (1965)
2. Martinet, B.: Regularisation, d'inéquations variationelles par approximations succesives. Rev. Fra. Inform. Rech. Opér. **4**, 154–159 (1970)
3. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
4. Burke, J.V., Qian, M.J.: A variable metric proximal point algorithm for monotone operators. SIAM J. Control Optim. **37**, 353–375 (1998)
5. Eckstein, J.: Approximate iterations in Bregman-function-based proximal algorithms. Math. Program. **83**, 113–123 (1998)
6. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)
7. Monteiro, R.D.C., Svaiter, B.F.: Convergence rate of inexact proximal point methods with relative error criteria for convex optimization. Manuscript (2010)
8. Güler, O.: New proximal point algorithms for convex minimization. SIAM J. Optim. **2**, 649–664 (1992)
9. Nesterov, Y.E.: A method for solving the convex programming problem with convergence rate $O(1/k^2)$. Dokl. Akad. Nauk SSSR **269**, 543–547 (1983)
10. Nesterov, Y.E.: On an approach to the construction of optimal methods of minimization of smooth convex functions. Èkon. Mat. Metody **24**, 509–517 (1988)
11. Beck, A., Teboulle, M.: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**, 183–202 (2009)