Research Article

Convergence Analysis on an Accelerated Proximal Point Algorithm for Linearly Constrained Optimization Problems

Sha Lu^{1,2} and Zengxin Wei³

¹School of Science, East China University of Science and Technology, Shanghai 200237, China

²School of Mathematics and Statistics, Nanning Normal University, Nanning 530001, China

³School of Mathematics and Information Science, Guangxi University, Nanning 530004, China

Correspondence should be addressed to Sha Lu; lusha_nn@126.com

Received 1 September 2020; Revised 17 October 2020; Accepted 20 October 2020; Published 10 November 2020

Academic Editor: Weijun Zhou

Copyright © 2020 Sha Lu and Zengxin Wei. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Proximal point algorithm is a type of method widely used in solving optimization problems and some practical problems such as machine learning in recent years. In this paper, a framework of accelerated proximal point algorithm is presented for convex minimization with linear constraints. The algorithm can be seen as an extension to Güler's methods for unconstrained optimization and linear programming problems. We prove that the sequence generated by the algorithm converges to a KKT solution of the original problem under appropriate conditions with the convergence rate of $O(1/k^2)$.

1. Introduction

Proximal point algorithm (abbr. PPA) is a type of method widely used in solving optimization problems, fixed point problems, maximal monotone operator problems, and so on. The framework of the proximal point method is closely related to many algorithms. It even can do more interpretation and generalization of some other methods and algorithms. In recent years, combining the idea of the proximal point method or the proximal terms with some existing algorithms shows that it can improve the performance of the original algorithms in a certain extent. The main step of the PPA is to compute a subproblem consisting of proximal point operator. In some practical problems with suitable conditions, the proximal point subproblems may be expressed as convex optimization problems with smaller scales or better properties which make them even have closed form solutions. In recent years, the proximal point method, together with its relative models, and several types of PPAs have been used in machine learning, image recognition, signal processing, and so on [1–3].

The original PPA can be seen in Martinet [4, 5] about the research of the fixed point problem and the variational

inequality. For any given x_0 in a Hilbert space, Martinet [4] gave an iterative sequence $\{x_k\}_{k \in \mathbb{N}}$ for variational inequality problem by a PPA based on the following inclusion relation:

$$x_{k+1} \in (I + c_k T)^{-1}(x_k), \quad (1)$$

where c_k is a sequence of positive integers and T is a maximal monotone operator. Later, in the unconstrained minimization of $f(x)$, where $f(\cdot)$ is a proper lower semicontinuous convex function in a Hilbert space, Rockafellar [6, 7] presented a more practical PPA which had the iteration as

$$x_{k+1} \approx \arg \min \left\{ f(x) + \frac{1}{2\beta_k} \|x - x_k\|^2 \right\}. \quad (2)$$

If $\beta_k \geq \beta > 0$, for all k , it was proved that the PPA weakly converges to a zero solution of the maximal monotone operator T . And when $\beta_k \rightarrow \infty$, the convergence rate is superlinear. In 1991, Güler [8] further discussed the PPA and its convergent property. He proved that, under weaker conditions, that is, when $\{\beta_k\}_{k \in \mathbb{N}}$ satisfies $\sum_{k=0}^{\infty} \beta_k \rightarrow \infty$, the PPA is convergent. Furthermore, its global convergence rate can be given as

$$f(x_k) - f(x^*) = O\left(\frac{1}{\sum_{j=0}^{k-1} \beta_j}\right). \quad (3)$$

For solving the convex programming problem with Lipschitz continuous gradients, Nesterov [9] gave an accelerated gradient method, which added a step to the gradient descent direction in order to reduce the times of iterations and obtained the iterative estimation of general first-order methods. The idea of acceleration was then introduced into other optimization algorithms, such as Güler [10], who gave two new PPAs by adding auxiliary point series and adopting acceleration strategy. Compared with (3), his accelerated algorithms obtained better convergence rate:

$$f(x_k) - f(x^*) = O\left(\frac{1}{\left(\sum_{j=0}^{k-1} \sqrt{\beta_j}\right)^2}\right). \quad (4)$$

Güler [11] also applied this method to linear programming and proposed new augmented Lagrangian multiplier algorithms. While the iteration subproblems were solved inexactly, the algorithms not only maintained the global convergent property but also provided faster global convergence rates and can terminate in finite iterations to obtain the primal and dual optimal solutions. In the literature of Birge et al. [12], the authors further generalized Güler's accelerated proximal point method to unconstrained nonsmooth convex optimization problems. They gave a model algorithm and then presented a family of PPAs which generated x_{k+1} under a different rule from the classical PPA. Under weaker conditions, the estimation of the global convergence rate is obtained. They also discussed the application of PPA on stochastic programming and the variant proximal bundle method. More research studies on the proximal point method can be found in the review of Parikh and Boyd [13]. In recent years, much effort has been made to accelerate various first-order methods and the convergence analysis for linearly constrained convex optimization. For example, Ke and Ma [14] proposed an accelerated augmented Lagrangian method for solving the linearly constrained convex programming and showed its convergence rate is $O(1/k^2)$. Xu [15] proposed two accelerated methods for solving structured linearly constrained convex programming and discussed their convergence rate under different conditions. Zhang et al. [16] applied the proximal method of multipliers for equality constrained optimization problems and proved that, under linear independence constraint qualification and the second-order sufficiency optimality condition, it is linearly convergent. And when the penalty parameter increases to $+\infty$, it is superlinear.

Now, we consider an accelerated PPA for solving convex optimization problem with linear equality constraints:

$$(P) \min_{x \in \mathcal{R}^n} f(x), \quad (5)$$

$$\text{s.t. } Ax = b,$$

where $f(x): \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$ is a proper lower semi-continuous (not necessary smooth) convex function and $A \in \mathcal{R}^{m \times n}$, $b \in \mathcal{R}^m$. Our main work is as follows. (1) By using the Lagrange function, the KKT system, and the original-dual relations, we construct appropriate auxiliary sequences by quadratic convex functions $\varphi_k(x)$ and auxiliary points $\{y_k\}$. Then, we update x_k and the Lagrange multipliers, respectively, to extend the accelerated PPA to general convex optimization problems. (2) In the extended algorithm, the parameter α_k which is related to the convergence rate is updated with an introduced constant c . The update of α_k in Güler's algorithm can be seen as a special case with $c = 1$. (3) When the iteration subproblems are solved exactly, the algorithm has the convergence rate of $O(1/k^2)$ in terms of the objective residual of the associated Lagrange function.

The remaining parts of this paper are organized as follows. In Section 2, a framework of the accelerated PPA is presented for constrained optimization problem (5). In Section 3, the global convergence is established under mild assumptions. In Section 4, the convergence rate based on the function values is given. In Section 5, we conclude this paper with some remarks.

2. An Accelerated PPA for Constrained Convex Optimization

For an unconstrained optimization problem,

$$\min_{x \in \mathcal{R}^n} f(x), \quad (6)$$

the classical PPA generates the next iteration point by

$$x_{k+1} := \mathcal{P}_{\beta_k} f(x_k) = \arg \min_{z \in \mathcal{R}^n} \left\{ f(z) + \frac{1}{2\beta_k} \|z - x_k\|^2 \right\}. \quad (7)$$

And the main accelerating idea of Nesterov [9] and Güler [10] is to construct a sequence of auxiliary quadratic convex functions $\varphi_k(x)$ which can be seen as estimations of $f(x)$ but with better functional states. While k increases, the difference between the auxiliary functions and the original objective function is compressed such that, for any $x \in \mathcal{R}^n$,

$$\varphi_k(x) - f(x) \leq (1 - \alpha_{k-1})(\varphi_{k-1}(x) - f(x)), \quad \alpha_{k-1} \in (0, 1). \quad (8)$$

Then, in each iteration of the algorithm, it produces x_k satisfying

$$f(x_k) \leq \varphi_k^* := \varphi_k(v_k^*) = \min_{x \in \mathcal{R}^n} \varphi_k(x), \quad (9)$$

where v_k^* is the minimal point of $\varphi_k(x)$. Thus, while the minimal point x^* exists, since $f(x)$ is proper, it obtains

$$f(x_k) - f(x^*) \leq \varphi_k(x^*) - f(x^*) \leq \prod_{j=0}^{k-1} (1 - \alpha_j) (\varphi_0(x^*) - f(x^*)). \quad (10)$$

Since $\prod_{j=0}^{k-1} (1 - \alpha_j) \rightarrow 0$ ($k \rightarrow +\infty$), the sequence $\{x_k\}_{k \in \mathbb{N}}$ minimizes $f(x)$ and the speed that $f(x)$ converges to f^* is related to $\prod_{j=0}^{k-1} (1 - \alpha_j)$.

The key steps in the accelerated algorithm are the construction of the auxiliary quadratic convex functions with suitable estimation on $f(x)$, the producing of x_k satisfying (15), and the selection of the compressing factor α_k . Inspired by Nesterov's estimate sequences and the PPA given by Güler[10], we consider the constrained optimization problem (5). It is known that the dual problem of (5) is

$$(D) \min_{\lambda \in \mathcal{R}^m} \{f^*(A^\top \lambda) - \lambda^\top b\}, \quad (11)$$

where f^* is the Fenchel conjugate function of f . And the augmented Lagrangian function associated with (5) is defined as

$$\mathcal{L}_\sigma(x, \lambda) := f(x) - \lambda^\top (Ax - b) + \frac{\sigma}{2} \|Ax - b\|^2, \quad (12)$$

where $\lambda \in \mathcal{R}^m$ is the multiplier and $\sigma > 0$ is a penalty parameter. To simplify the discussion, we denote the proximal point about function f at a given y as

$$y^p = \mathcal{P}_\beta f(y) := \operatorname{argmin}_{z \in \mathcal{R}^n} \left\{ f(z) + \frac{1}{2\beta} \|z - y\|^2 \right\}. \quad (13)$$

Firstly, we construct a series of quadratic regular functions for (12) before the accelerated PPA is given. For given x_0, λ_0 , and $a_0 > 0$, let

$$\begin{aligned} \varphi_0(x) &:= \mathcal{L}_\sigma(x_0, \lambda_0) + \frac{a_0}{2} \|x - x_0\|^2 \\ &= f(x_0) - \lambda_0^\top (Ax_0 - b) + \frac{\sigma}{2} \|Ax_0 - b\|^2 + \frac{a_0}{2} \|x - x_0\|^2, \end{aligned} \quad (14)$$

$$\begin{aligned} \varphi_{k+1}(x) &:= (1 - \alpha_k) \varphi_k(x) + \alpha_k \left[f(y_k^p) \right. \\ &\quad \left. + \frac{1}{\beta_k} (y_k - y_k^p)^\top (x - y_k^p) \right] \\ &\quad + (1 - \alpha_k) \left[\lambda_k^\top (Ax_k - b) - \frac{\sigma}{2} \|Ax_k - b\|^2 \right] \\ &\quad - \lambda_k^\top (Ay_k^p - b) + \left(\frac{1}{2} + \alpha_k \right) \sigma \|Ay_k^p - b\|^2, \quad k \geq 0. \end{aligned} \quad (15)$$

Since $f(x)$ is convex and $\varphi_k(x)$ is a quadratic regular function, for any $k \in \mathbb{N}$, $\varphi_k(x)$ can be written in the canonical form:

$$\varphi_k(x) = \varphi_k(v_k^*) + \frac{a_k}{2} \|x - v_k^*\|^2, \quad (16)$$

where v_k^* means the minimizer of $\varphi_k(x)$.

On the contrary, from (15), we have

$$\begin{aligned} \varphi_{k+1}(x) &= (1 - \alpha_k) \left[\varphi_k(v_k^*) + \frac{a_k}{2} \|x - v_k^*\|^2 \right] \\ &\quad + \alpha_k \left[f(y_k^p) + \frac{1}{\beta_k} (y_k - y_k^p)^\top (x - y_k^p) \right] \\ &\quad + (1 - \alpha_k) \left[\lambda_k^\top (Ax_k - b) - \frac{\sigma}{2} \|Ax_k - b\|^2 \right] \\ &\quad - \lambda_k^\top (Ay_k^p - b) + \left(\frac{1}{2} + \alpha_k \right) \sigma \|Ay_k^p - b\|^2. \end{aligned} \quad (17)$$

Comparing the second terms of (16) and (17), it implies

$$a_{k+1} = (1 - \alpha_k) a_k. \quad (18)$$

It is not hard to see that the minimum values and the minimizers between two adjacent quadratic functions have the following relationships:

$$\begin{aligned} \varphi_{k+1}(v_{k+1}^*) &= (1 - \alpha_k) \left[\varphi_k(v_k^*) + \frac{a_k}{2} \|v_{k+1}^* - v_k^*\|^2 \right] \\ &\quad + \alpha_k \left[f(y_k^p) + \frac{1}{\beta_k} (y_k - y_k^p)^\top (v_{k+1}^* - y_k^p) \right] \\ &\quad + (1 - \alpha_k) \left[\lambda_k^\top (Ax_k - b) - \frac{\sigma}{2} \|Ax_k - b\|^2 \right] \\ &\quad - \lambda_k^\top (Ay_k^p - b) + \left(\frac{1}{2} + \alpha_k \right) \sigma \|Ay_k^p - b\|^2, \end{aligned} \quad (19)$$

and

$$v_{k+1}^* = v_k^* - \frac{\alpha_k}{a_{k+1}\beta_k} (y_k - y_k^p). \quad (20)$$

If we generate the iteration points by $x_{k+1} := y_k^p$ for $k \geq 0$, the following lemma shows the relationship between the Lagrange function, the proximal point y_k^p , the constructed quadratic estimation function $\varphi_k(x)$, and its minimizer v_{k+1}^* .

Lemma 1. Suppose for $k \geq 0$, $\varphi_k(x)$, $\mathcal{L}_\sigma(x, \lambda)$, x_k , v_k^* , and y_k^p are defined as previously mentioned. Let $c \in (0, 2)$,

$$\begin{aligned} y_k &= (1 - \alpha_k) x_k + \alpha_k v_k^*, \\ \frac{\alpha_k^2}{(1 - \alpha_k) a_k \beta_k} &= c, \end{aligned} \quad (21)$$

$$\tilde{\lambda}_{k+1} = \tilde{\lambda}_k - \alpha_k \sigma (Ax_{k+1} - b).$$

Then, we have

$$\begin{aligned}\varphi_{k+1}(v_{k+1}^*) &\geq \mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k) + \frac{1}{2\beta_k} (2-c) \|y_k - x_{k+1}\|^2 \\ &\geq \mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k).\end{aligned}\quad (22)$$

Proof. We prove it by induction. While $k = 0$, (22) is obvious. Now, assume

$$\varphi_k(v_k^*) \geq \mathcal{L}_\sigma(x_k, \tilde{\lambda}_{k-1}). \quad (23)$$

Since $f(x)$ is convex, $f(x) + (1/2\beta_k)\|x - y_k\|^2$ is strong convex, from the definition of y_k^p ; then, we have

$$0 \in \partial f(y_k^p) + \frac{1}{\beta_k}(y_k^p - y_k). \quad (24)$$

By convexity, for any $x \in R^n$, it has

$$f(x) \geq f(y_k^p) + \frac{1}{\beta_k}(y_k - y_k^p)^\top (x - y_k^p). \quad (25)$$

Then, by (23),

$$\begin{aligned}\varphi_k(v_k^*) &\geq f(x_k) - \tilde{\lambda}_{k-1}^\top (Ax_k - b) + \frac{\sigma}{2} \|Ax_k - b\|^2 \\ &\geq f(y_k^p) + \frac{1}{\beta_k}(y_k - y_k^p)^\top (x_k - y_k^p) - \tilde{\lambda}_{k-1}^\top (Ax_k - b) \\ &\quad + \frac{\sigma}{2} \|Ax_k - b\|^2.\end{aligned}\quad (26)$$

Substitute (18) and (26) into (20), it implies

$$\begin{aligned}\varphi_{k+1}(v_{k+1}^*) &\geq (1 - \alpha_k) \left[f(y_k^p) + \frac{1}{\beta_k}(y_k - y_k^p)^\top (x_k - y_k^p) - \tilde{\lambda}_{k-1}^\top (Ax_k - b) + \frac{\sigma}{2} \|Ax_k - b\|^2 \right] \\ &\quad + \frac{\alpha_{k+1}}{2} \|v_{k+1}^* - v_k^*\| + \alpha_k \left[f(y_k^p) + \frac{1}{\beta_k}(y_k - y_k^p)^\top (v_{k+1}^* - y_k^p) \right] \\ &\quad + (1 - \alpha_k) \left[\tilde{\lambda}_{k-1}^\top (Ax_k - b) - \frac{\sigma}{2} \|Ax_k - b\|^2 \right] \\ &\quad - \tilde{\lambda}_{k-1}^\top (Ay_k^p - b) + \left(\frac{1}{2} + \alpha_k \right) \sigma \|Ay_k^p - b\|^2 \\ &= f(y_k^p) + \frac{1 - \alpha_k}{\beta_k} (y_k - y_k^p)^\top (x_k - y_k^p) + \frac{\alpha_k^2}{2\alpha_{k+1}\beta_k^2} \|y_k - y_k^p\|^2 \\ &\quad + \frac{\alpha_k}{\beta_k} (y_k - y_k^p)^\top (v_{k+1}^* - y_k^p) - \lambda_k^\top (Ay_k^p - b) + \frac{\sigma}{2} \|Ay_k^p - b\|^2 \\ &= \mathcal{L}_\sigma(y_k^p, \tilde{\lambda}_k) + \frac{\alpha_k^2}{2\alpha_{k+1}\beta_k^2} \|y_k - y_k^p\|^2 \\ &\quad + \frac{1}{\beta_k} (y_k - y_k^p)^\top [(1 - \alpha_k)x_k + \alpha_k v_{k+1}^* - y_k^p].\end{aligned}\quad (27)$$

Notice that, in the last term of (27), by using $y_k = (1 - \alpha_k)x_k + \alpha_k v_k^*$, it turns out

$$(1 - \alpha_k)x_k + \alpha_k v_{k+1}^* - y_k^p = \alpha_k (v_{k+1}^* - v_k^*) + (y_k - y_k^p). \quad (28)$$

Thus, by (20), we gain

$$\begin{aligned}(y_k - y_k^p)^\top [(1 - \alpha_k)x_k + \alpha_k v_{k+1}^* - y_k^p] &= \left(1 - \frac{\alpha_k^2}{\alpha_{k+1}\beta_k} \right) \\ &\quad \cdot \|y_k - y_k^p\|^2.\end{aligned}\quad (29)$$

Substituting this formula into (27), it deduces

$$\begin{aligned}
\varphi_{k+1}(v_{k+1}^*) &\geq \mathcal{L}_\sigma(y_k^p, \tilde{\lambda}_k) + \frac{\alpha_k^2}{2a_{k+1}\beta_k^2} \|y_k - y_k^p\|^2 + \frac{1}{\beta_k} \left(1 - \frac{\alpha_k^2}{a_{k+1}\beta_k}\right) \|y_k - y_k^p\|^2 \\
&= \mathcal{L}_\sigma(y_k^p, \tilde{\lambda}_k) + \frac{1}{\beta_k} \left(1 - \frac{\alpha_k^2}{2a_{k+1}\beta_k}\right) \|y_k - y_k^p\|^2.
\end{aligned} \tag{30}$$

By using of $a_{k+1} = (1 - \alpha_k)a_k$, $\alpha_k^2/(1 - \alpha_k)a_k\beta_k = c < 2$, and $x_{k+1} = y_k^p$, we have

$$\begin{aligned}
\varphi_{k+1}(v_{k+1}^*) &\geq \mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k) + \frac{1}{2\beta_k} (2 - c) \|y_k - x_{k+1}\|^2 \\
&\geq \mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k).
\end{aligned} \tag{31}$$

From Lemma 1, $\varphi_{k+1}(v_{k+1}^*)$ can be seen as an upper bound estimation of the Lagrange function on $(x_{k+1}, \tilde{\lambda}_k)$. v_k^* , y_k , α_k , and $\tilde{\lambda}_k$ all can be updated by explicit formulas. Thus, we give the accelerated PPA for the constrained optimization problem (5) as follows.

Accelerated PPA for constrained convex optimization (C - APPA):

(i) Step 0: let $x_0 \in R^n, \tilde{\lambda}_{-1} = \tilde{\lambda}_0 \in R^m$. Choose $\beta_0 > 0, \sigma > 0, a > 0, c \in (0, 2)$.

Let $v_0^* := x_0, a_0 := a$.

(ii) Step 1: let $\beta_k > 0$,

$$\alpha_k := \frac{\sqrt{(ca_k\beta_k)^2 + 4ca_k\beta_k} - ca_k\beta_k}{2}, \tag{32}$$

$$y_k := (1 - \alpha_k)x_k + \alpha_k v_k^*. \tag{33}$$

(iii) Step 2: compute the proximal point and the updating

$$\begin{aligned}
x_{k+1} &:= \operatorname{argmin}_{x \in R^n} \left\{ f(x) - \tilde{\lambda}_k^\top (Ax - b) + \frac{\sigma}{2} \|Ax - b\|^2 \right. \\
&\quad \left. + \frac{1}{2\beta_k} \|x - y_k\|^2 \right\},
\end{aligned} \tag{34}$$

$$= \operatorname{argmin}_{x \in R^n} \left\{ f(x) + \frac{\sigma}{2} \left\| Ax - b - \frac{\tilde{\lambda}_k}{\sigma} \right\|^2 + \frac{1}{2\beta_k} \|x - y_k\|^2 \right\}, \tag{35}$$

$$\lambda_{k+1} := \tilde{\lambda}_k - \sigma (Ax_{k+1} - b), \tag{36}$$

$$a_{k+1} := (1 - \alpha_k)a_k, \tag{37}$$

$$v_{k+1}^* := v_k^* + \frac{\alpha_k}{a_{k+1}\beta_k} (x_{k+1} - y_k), \tag{38}$$

$$\tilde{\lambda}_{k+1} := \tilde{\lambda}_k + \alpha_k (\lambda_{k+1} - \tilde{\lambda}_k). \tag{39}$$

(iv) Step 3: let $k := k + 1$ and go to Step 1. $\square$

Remark 1. It is obvious that $0 < \alpha_k < 1$ by the definition. Actually, from (10), the convergence of the algorithm relates to $\prod (1 - \alpha_k)$. In the later Theorem 1, it needs to assume the nonzero lower bound such that $\alpha_k \geq \underline{\alpha} > 0$. In particular, when $c\alpha_k\beta_k = 1$, it obtains $\alpha_k = \sqrt{5} - 1/2$.

Remark 2. In the iterations, x_{k+1} can be computed from (34) or (35), which has the same result though they have different quadratic terms. We will use these two different subproblem forms in later discussion on convergence for convenience. One is used to prove the sequence converges to the KKT solution and the other (35), which is more suitable for measuring changes in function values and the constraints, is used in the analysis on the convergence rate.

Remark 3. In the problem (35), scaled proximal terms also can be used such as $\|x - y_k\|_Q^2$, where Q is a positive semidefinite matrix. For some structured function $f(x)$ and appropriately chosen Q , the augmented term can be linearized, and it makes (35) have a closed form solution. Or in practice, we can solve the problem inexactly, but it may not retain $O(1/k^2)$ convergence rate, which we will show in the following sections under the same assumptions. Further discussion will be out of the scope of this paper, and we leave the discussion on inexactly solving cases and their numerical experiments to another paper.

3. Global Convergence

Assumption 1. Assume that f is a proper lower semicontinuous convex function and $x_0 \in R^n$ is a feasible point of (5).

$$\langle \lambda_{k+1}^e + \sigma(1 - \alpha_k)Ax_{k+1}^e, Ax_{k+1}^e \rangle = -\frac{\sigma\alpha_k}{2}\|Ax_{k+1}^e\|^2 - \frac{1}{2\sigma\alpha_k} \left[\|\lambda_{k+1}^e + \sigma(1 - \alpha_k)Ax_{k+1}^e\|^2 - \|\lambda_k^e + \sigma(1 - \alpha_{k-1})Ax_k^e\|^2 \right]. \quad (41)$$

Proof. By the first-order optimization condition of (34) and together with (36), we have

$$A^\top \lambda_{k+1} - \frac{1}{\beta_k} (x_{k+1} - y_k) \in \partial f(x_{k+1}), \quad (42)$$

$$A^\top \lambda_k - \frac{1}{\beta_{k-1}} (x_k - y_{k-1}) \in \partial f(x_k). \quad (43)$$

Since $f(x)$ is proper lower semicontinuous convex function and by Theorem 12.17 of Rockafellar and Wets [18], ∂f is a maximal monotone operator and there exists a self-adjoint positive semidefinite linear operator Σ_f such that, for any $x, x' \in \text{dom}(f)$ with $u \in \partial f(x)$ and $u' \in \partial f(x')$,

$$f(x) \geq f(x') + \langle u', x - x' \rangle + \frac{1}{2}\|x - x'\|_{\Sigma_f}^2, \quad (44)$$

$$\langle u - u', x - x' \rangle \geq \|x - x'\|_{\Sigma_f}^2. \quad (45)$$

Notice that $\bar{x}$ and $\bar{\lambda}$ satisfy the KKT system, and it implies

$$\langle \lambda_{k+1}^e, Ax_{k+1}^e \rangle - \frac{1}{\beta_k} \langle x_{k+1} - y_k, x_{k+1}^e \rangle \geq \|x_{k+1}^e\|_{\Sigma_f}^2. \quad (46)$$

By using of (36) and (39), it is easy to know that

$$Ax_{k+1} - b = A(x_{k+1} - \bar{x}) = \frac{1}{\sigma}(\bar{\lambda}_k - \lambda_{k+1}) = -\frac{1}{\sigma\alpha_k}(\bar{\lambda}_{k+1} - \bar{\lambda}_k). \quad (47)$$

Thus,

$$\langle \lambda_{k+1}^e, Ax_{k+1}^e \rangle + \langle \bar{\lambda}_k - \lambda_{k+1}^e, Ax_{k+1}^e \rangle - \langle \sigma(Ax_{k+1} - b), Ax_{k+1}^e \rangle - \langle \frac{1}{\beta_k} (x_{k+1} - y_k), x_{k+1}^e \rangle \geq \|x_{k+1}^e\|_{\Sigma_f}^2. \quad (48)$$

By (36) and (39) in the algorithm, we have

$$\bar{\lambda}_k = \lambda_k + (\alpha_{k-1} - 1)(\lambda_k - \bar{\lambda}_{k-1}) = \lambda_k - \sigma(\alpha_{k-1} - 1)(Ax_k - b), \quad (49)$$

and

$$\lambda_{k+1} = \lambda_k - \sigma\alpha_{k-1}(Ax_k - b) + \sigma(Ax_k - Ax_{k+1}). \quad (50)$$

Then,

$$(\lambda_{k+1}^e + \sigma(1 - \alpha_k)Ax_{k+1}^e) + \sigma\alpha_k(Ax_{k+1} - b) = \lambda_k^e + \sigma(1 - \alpha_{k-1})Ax_k^e. \quad (51)$$

From the basic relations,

$$2\langle u, v \rangle = \|u\|^2 + \|v\|^2 - \|u - v\|^2 = \|u + v\|^2 - \|u\|^2 - \|v\|^2, \quad (52)$$

it deduces

$$\langle \lambda_{k+1}^e + \sigma(1 - \alpha_k)Ax_{k+1}^e, Ax_{k+1} - b \rangle = \frac{1}{2\sigma\alpha_k} \|\lambda_k^e + \sigma(1 - \alpha_{k-1})Ax_k^e\|^2 - \frac{1}{2\sigma\alpha_k} \|\lambda_{k+1}^e + \sigma(1 - \alpha_k)Ax_{k+1}^e\|^2 - \frac{\sigma\alpha_k}{2} \|Ax_{k+1} - b\|^2. \quad (53)$$

Since $A\bar{x} = b$, $Ax_{k+1} - b = Ax_{k+1}^e$, we gain (41) by (53). $\square$

Lemma 3. Suppose that $\{(x_k, \lambda_k)\}$ is generated by C-APPA and $(\bar{x}, \bar{\lambda})$ is a KKT solution of (5). Then, for $k = 1, 2, \dots$, it has

$$\begin{aligned} & \langle \lambda_k^e - \lambda_{k+1}^e, Ax_{k+1}^e \rangle - \langle \sigma(Ax_{k+1} - b), Ax_{k+1}^e \rangle \\ & \leq -2\|x_{k+1} - x_k\|_{\Sigma_f}^2 - \left(\frac{1}{\beta_k} - \frac{1}{\beta_{k-1}}\right) \|x_{k+1} - x_k\|^2 \\ & \quad - \frac{1}{\beta_k} (\|x_{k+1} - y_k\|^2 - \|x_k - y_k\|^2) + \frac{1}{\beta_{k-1}} \|x_k - y_{k-1}\|^2 \\ & \quad + \sigma\left(\alpha_{k-1} - \frac{1}{2}\right) \|Ax_k^e\|^2 - \sigma\left(\alpha_k - \frac{1}{2}\right) \|Ax_{k+1}^e\|^2 - \frac{\sigma}{2} \|Ax_{k+1}^e - Ax_k^e\|^2. \end{aligned} \quad (54)$$

Proof. From

$$\begin{aligned} & \langle \lambda_{k+1} - \lambda_k, A(x_{k+1} - x_k) \rangle - \left\langle \frac{1}{\beta_k} (x_{k+1} - y_k) - \frac{1}{\beta_{k-1}} (x_k - y_{k-1}), x_{k+1} - x_k \right\rangle \\ & \geq \|x_{k+1} - x_k\|_{\Sigma_f}^2. \end{aligned} \quad (57)$$

By (55), it deduces

$$\begin{aligned} \langle \lambda_{k+1} - \lambda_k, A(x_{k+1} - x_k) \rangle & \geq \frac{1}{2} \left(\frac{1}{\beta_k} - \frac{1}{\beta_{k-1}} \right) \|x_{k+1} - x_k\|^2 \\ & \quad + \frac{1}{2\beta_k} (\|x_{k+1} - y_k\|^2 - \|x_k - y_k\|^2) \\ & \quad - \frac{1}{2\beta_{k-1}} \|x_k - y_{k-1}\|^2 + \|x_{k+1} - x_k\|_{\Sigma_f}^2. \end{aligned} \quad (58)$$

From (50) and (55), we can obtain

$$\begin{aligned} & \left\langle \frac{1}{\beta_k} (x_{k+1} - y_k) - \frac{1}{\beta_{k-1}} (x_k - y_{k-1}), x_{k+1} - x_k \right\rangle \\ & \geq \frac{1}{2} \left(\frac{1}{\beta_k} - \frac{1}{\beta_{k-1}} \right) \|x_{k+1} - x_k\|^2 + \frac{1}{2\beta_k} (\|x_{k+1} - y_k\|^2 - \|x_k - y_k\|^2) \\ & \quad - \frac{1}{2\beta_{k-1}} \|x_k - y_{k-1}\|^2, \end{aligned} \quad (55)$$

and together with (42), (43), and (45), we have

$$\begin{aligned} & \langle A^\top (\lambda_{k+1} - \lambda_k) - \frac{1}{\beta_k} (x_{k+1} - y_k) + \frac{1}{\beta_{k-1}} (x_k - y_{k-1}), x_{k+1} - x_k \rangle \\ & \geq \|x_{k+1} - x_k\|_{\Sigma_f}^2. \end{aligned} \quad (56)$$

Then,

$$\begin{aligned} & \sigma \langle Ax_k^e - Ax_{k+1}^e, Ax_{k+1}^e \rangle \\ & = \langle Ax_{k+1}^e - Ax_k^e, \lambda_{k+1} - \lambda_k \rangle + \sigma(\alpha_{k-1} - 1) \langle Ax_{k+1}^e - Ax_k^e, Ax_k^e \rangle \\ & \geq \|x_{k+1} - x_k\|_{\Sigma_f}^2 + \frac{1}{2} \left(\frac{1}{\beta_k} - \frac{1}{\beta_{k-1}} \right) \|x_{k+1} - x_k\|^2 \\ & \quad + \frac{1}{2\beta_k} (\|x_{k+1} - y_k\|^2 - \|x_k - y_k\|^2) - \frac{1}{2\beta_{k-1}} \|x_k - y_{k-1}\|^2 \\ & \quad + \frac{\sigma(1 - \alpha_{k-1})}{2} \|Ax_k^e\|^2 - \frac{\sigma(1 - \alpha_{k-1})}{2} \|Ax_{k+1}^e\|^2 \\ & \quad + \frac{\sigma(1 - \alpha_{k-1})}{2} \|A(x_{k+1} - x_k)\|^2. \end{aligned} \quad (59)$$

Also by using of (50), it has

$$\sigma(1 - \alpha_k)Ax_{k+1}^e + (\lambda_k^e - \lambda_{k+1}^e) - \sigma(Ax_k - b) = \sigma[(2 - \alpha_k)(Ax_{k+1}^e - (2 - \alpha_{k-1})Ax_k^e)]. \quad (60)$$

Since $0 < \alpha_k < 1$ for all k and by (59), (60), and (50), it deduces

$$\begin{aligned}
& \sigma(1 - \alpha_k) \langle Ax_{k+1}^e, Ax_{k+1}^e \rangle + \langle \lambda_k^e - \lambda_{k+1}^e, Ax_{k+1}^e \rangle - \sigma \langle Ax_k^e, Ax_{k+1}^e \rangle \\
& = 2\sigma \langle (Ax_{k+1}^e - Ax_k^e), Ax_{k+1}^e \rangle - \sigma \langle \alpha_k Ax_{k+1}^e, Ax_{k+1}^e \rangle + \sigma \langle \alpha_{k-1} Ax_k^e, Ax_{k+1}^e \rangle \\
& \leq -2 \|x_{k+1} - x_k\|_{\Sigma_f}^2 - \left(\frac{1}{\beta_k} - \frac{1}{\beta_{k-1}} \right) \|x_{k+1} - x_k\|^2 \\
& \quad - \frac{1}{\beta_k} \left(\|x_{k+1} - y_k\|^2 - \|x_k - y_k\|^2 \right) + \frac{1}{\beta_{k-1}} \|x_k - y_{k-1}\|^2 \\
& \quad + \sigma(1 - \alpha_{k-1}) \langle Ax_{k+1}^e - Ax_k^e, Ax_k^e \rangle - \sigma \alpha_k \|Ax_{k+1}^e\|^2 + \sigma \alpha_{k-1} \langle Ax_{k+1}^e, Ax_k^e \rangle \\
& = -2 \|x_{k+1} - x_k\|_{\Sigma_f}^2 - \left(\frac{1}{\beta_k} - \frac{1}{\beta_{k-1}} \right) \|x_{k+1} - x_k\|^2 \\
& \quad - \frac{1}{\beta_k} \left(\|x_{k+1} - y_k\|^2 - \|x_k - y_k\|^2 \right) + \frac{1}{\beta_{k-1}} \|x_k - y_{k-1}\|^2 \\
& \quad + \sigma \left(\alpha_{k-1} - \frac{1}{2} \right) \|Ax_k^e\|^2 - \sigma \left(\alpha_k - \frac{1}{2} \right) \|Ax_{k+1}^e\|^2 - \frac{\sigma}{2} \|Ax_{k+1}^e - Ax_k^e\|^2.
\end{aligned} \tag{61}$$

Thus, (54) is true.

Now, for $k > 0$, we construct the following ϕ_k and turn to the analysis of the convergence of the algorithm:

$$\begin{aligned}
\phi_k & := \frac{\sigma \alpha_{k-1}}{2} \|Ax_k^e\|^2 + \frac{1}{2\sigma \alpha_k} \|\lambda_k^e + \sigma(1 - \alpha_{k-1})Ax_k^e\|^2 \\
& \quad + \frac{1}{\beta_{k-1}} \|x_k - y_{k-1}\|^2.
\end{aligned} \tag{62}$$

The theorem shows that the sequence $\{\phi_k\}_{k>0}$ decreases monotonically under appropriate conditions, and then C-APPA is convergent since $\phi_k \geq 0$ for all $k \geq 0$. $\square$

Theorem 1. Suppose the solution set of (5) is nonempty. Denote $(\bar{x}, \bar{\lambda})$ as one solution satisfying the KKT system. Let ϕ_k ($k > 0$) be defined as (62) and $\{(x_k, \lambda_k)\}$ is generated by C-APPA. Then,

$$\begin{aligned}
\phi_k - \phi_{k+1} & \geq 2 \|x_{k+1} - x_k\|_{\Sigma_f}^2 + \|x_{k+1}^e\|_{\Sigma_f}^2 \\
& \quad + \frac{1}{2\beta_k} \left[\|x_k - y_k\|^2 + \|x_{k+1} - y_k\|^2 + \|x_{k+1}^e\|^2 \right] \\
& \quad + \frac{\sigma \alpha_{k-1}}{2} \|Ax_{k+1}^e\|^2 + \frac{\sigma}{2} (2 - \alpha_{k-1}) \|A(x_{k+1} - x_k)\|^2.
\end{aligned} \tag{63}$$

Furthermore, for all $k = 1, 2, \dots$, if there exists $\underline{\alpha} > 0$ and $\underline{\beta} > 0$ such that $\alpha_k \geq \underline{\alpha}$ and $\beta_k \geq \underline{\beta}$, then

$$\begin{aligned}
\lim_{k \rightarrow \infty} Ax_k & = b, \\
\lim_{k \rightarrow \infty} x_k & = \bar{x}, \\
\lim_{k \rightarrow \infty} \tilde{\lambda}_k & = \lim_{k \rightarrow \infty} \lambda_k = \bar{\lambda}.
\end{aligned} \tag{64}$$

Proof. From Lemmas 1 and 2, we can substitute (49), (41), and (54) into the left side of inequality (48). Since

$$\begin{aligned}
& \langle \lambda_{k+1}^e, Ax_{k+1}^e \rangle + \left[\sigma(1 - \alpha_k) \langle Ax_{k+1}^e, Ax_{k+1}^e \rangle + \langle \tilde{\lambda}_k^e - \lambda_{k+1}^e, Ax_{k+1}^e \rangle - \langle \sigma(Ax_{k+1} - b), Ax_{k+1}^e \rangle \right] - \langle \frac{1}{\beta_k} (x_{k+1} - y_k), x_{k+1}^e \rangle \\
& \leq \frac{\sigma \alpha_{k-1}}{2} \|Ax_k^e\|^2 - \frac{\sigma \alpha_k}{2} \|Ax_{k+1}^e\|^2 - \frac{1}{2\sigma \alpha_k} \left[\|\lambda_{k+1}^e + \sigma(1 - \alpha_k)Ax_{k+1}^e\|^2 - \|\lambda_k^e + \sigma(1 - \alpha_{k-1})Ax_k^e\|^2 \right] \\
& \quad - 2 \|x_{k+1} - x_k\|_{\Sigma_f}^2 + \frac{1}{\beta_{k-1}} \|x_k - y_{k-1}\|^2 - \frac{1}{\beta_k} \|x_{k+1} - y_k\|^2 \\
& \quad + \frac{1}{\beta_k} \|x_k - y_k\|^2 + \sigma \left(1 - \alpha_k - \frac{\alpha_{k-1}}{2} \right) \|Ax_{k+1}^e\|^2 - \frac{\sigma}{2} (2 - \alpha_{k-1}) \|Ax_{k+1}^e - Ax_k^e\|^2 - \frac{1}{2\beta_k} \|x_{k+1} - y_k\|^2 - \frac{1}{2\beta_k} \|x_{k+1}^e\|^2 + \frac{1}{2\beta_k} \|y_k - \bar{x}\|^2,
\end{aligned} \tag{65}$$

then by using of (62) and (48), we obtain

$$\phi_k - \phi_{k+1} \geq 2\|x_{k+1} - x_k\|_{\Sigma_f}^2 + \|x_{k+1}^e\|_{\Sigma_f}^2 + \frac{1}{2\beta_k} \left[\|x_k - y_k\|^2 + \|x_{k+1} - y_k\|^2 + \|x_{k+1}^e\|^2 \right] + \frac{\sigma\alpha_{k-1}}{2} \|Ax_{k+1}^e\|^2 + \frac{\sigma}{2} (2 - \alpha_{k-1}) \|A(x_{k+1} - x_k)\|^2, \quad (66)$$

which deduces the sequence $\{\phi_k\}_{k>0}$ is nonnegative and monotone nonincreasing. When the sequence is infinite, $\{\sigma\alpha_{k-1}/2 \|Ax_k^e\|^2\}$, $\{1/2\sigma\alpha_k \|\lambda_k^e + \sigma(1 - \alpha_{k-1})Ax_k^e\|^2\}$, and $\{1/\beta_{k-1} \|x_k - y_{k-1}\|^2\}$ are all bounded. Let $k \rightarrow \infty$, and by (66), we have

$$\begin{aligned} \|x_{k+1}^e\|_{\Sigma_f}^2 &\rightarrow 0, \frac{1}{2\beta_k} \|x_k - y_k\|^2 \rightarrow 0, \\ \frac{1}{2\beta_k} \|x_{k+1} - y_k\|^2 &\rightarrow 0, \frac{1}{2\beta_k} \|x_{k+1}^e\|^2 \rightarrow 0, \\ \frac{\sigma\alpha_{k-1}}{2} \|Ax_{k+1}^e\|^2 &\rightarrow 0, \frac{\sigma}{2} (2 - \alpha_{k-1}) \|A(x_{k+1} - x_k)\|^2 \rightarrow 0. \end{aligned} \quad (67)$$

Since $\underline{\alpha}$ and $\underline{\beta}$ are the lower bounds of α_k and β_k , respectively, it can be known that $\{\|Ax_k^e\|\}$, $\{1/2\sigma\alpha_k \|\lambda_k^e + \sigma(1 - \alpha_{k-1})Ax_k^e\|^2\}$, $\{\|x_k - y_k\|\}$, $\{\|x_{k+1} - y_k\|\}$, $\{\|x_{k+1}^e\|\}$, $\{\|Ax_{k+1}^e\|\}$, and $\{\|A(x_{k+1} - x_k)\|\}$ are all bounded. By $x_{k+1}^e = x_{k+1} - \bar{x}$ and $\|x_{k+1}\| \leq \|x_{k+1}^e\| + \|\bar{x}\|$, it has $\{x_{k+1}\}$ is bounded. Also, $\{\|\lambda_k\|\}$ is bounded. We denote the convergent subsequence of $\{(x_k, \lambda_k)\}$ as

$$\lim_{k_i \rightarrow \infty} \{(x_{k_i}, \lambda_{k_i})\} = (x_\infty, \lambda_\infty), \quad \{k_i\} \subseteq \{0, 1, \dots\}. \quad (68)$$

By (42) and (43), it has

$$A^\top \lambda_{k_i} - \frac{1}{\beta_{k_i-1}} (x_{k_i} - y_{k_i-1}) \in \partial f(x_{k_i}). \quad (69)$$

Then, by taking limits of k_i on both sides of (69) simultaneously, it deduces

$$\begin{aligned} A^\top \lambda_\infty &\in \partial f(x_\infty), \\ Ax_\infty - b &= 0. \end{aligned} \quad (70)$$

Assume for general that $(\bar{x}, \bar{\lambda}) = (x_\infty, \lambda_\infty)$. Thus, by the definition of ϕ_k , there must exist subsequence $\{(x_{k_i}, \lambda_{k_i})\}$ such that

$$\lim_{k_i \rightarrow \infty} \phi_{k_i} = 0. \quad (71)$$

Because $\{\phi_k\}$ is nonincreasing and bounded, $\{\phi_k\}_{k \in \mathbb{N}}$ converges to 0. By (62), it has

$$\|\lambda_k^e + \sigma(1 - \alpha_k)Ax_k^e\| \rightarrow 0, \|Ax_k^e\| \rightarrow 0, \quad (72)$$

for $k \rightarrow \infty$. Thus, by $0 \leq \|\lambda_k^e\| \leq \|\lambda_k^e + \sigma(1 - \alpha_k)Ax_k^e\| + \sigma\|Ax_k^e\|$, it obtains

$$\lim_{k \rightarrow \infty} \lambda_k = \bar{\lambda}. \quad (73)$$

Also, by $x_k^e = x_k - \bar{x}$, $A\bar{x} = b$, $\|Ax_k^e\| \rightarrow 0$, $\|x_{k+1}^e\| \rightarrow 0$, we have

$$\lim_{k \rightarrow \infty} x_k = \bar{x}, \lim_{k \rightarrow \infty} Ax_k = b. \quad (74)$$

From

$$\|\bar{\lambda}_k - \bar{\lambda}\| \leq \|\lambda_{k+1} - \bar{\lambda}\| + \sigma\|Ax_{k+1} - b\|, \quad (75)$$

we get $\lim_{k \rightarrow \infty} \bar{\lambda}_k = \bar{\lambda}$. $\square$

4. The Convergence Rate of C-APPA

Suppose that (34) is solved exactly in the iteration and we denote $x_{k+1} = y_k^p$. Let

$$F(x, \tilde{\lambda}_k) = f(x) + \frac{\sigma}{2} \left\| Ax - b - \frac{\tilde{\lambda}_k}{\sigma} \right\|^2, \quad (76)$$

$$\varphi_0(x) = F(x, \tilde{\lambda}_{-1}) + \frac{a_0}{2} \|x - x_0\|^2, \quad (77)$$

$$\begin{aligned} \varphi_{k+1}(x) &= (1 - \alpha_k)\varphi_k(x) + \alpha_k \left[F(x_{k+1}, \tilde{\lambda}_k) \right. \\ &\quad \left. - \frac{1}{\beta_k} (x_{k+1} - y_k)^\top (x - x_{k+1}) \right] \\ &\quad - (1 - \alpha_k) [F(x_k, \tilde{\lambda}_{k-1}) - F(x_k, \tilde{\lambda}_k)]. \end{aligned} \quad (78)$$

Lemma 4. Suppose that x_{k+1} , $\tilde{\lambda}_k$, and $\varphi_k(x)$ are generated by algorithm C-APPA. Then, for all $k = 0, 1, \dots$, it has

$$\begin{aligned} \varphi_{k+1}(x) - F(x, \tilde{\lambda}_k) &\leq (1 - \alpha_k) [\varphi_k(x) - F(x, \tilde{\lambda}_{k-1})] \\ &\quad + (1 - \alpha_k) \theta_{k+1}(x), \end{aligned} \quad (79)$$

where

$$\theta_{k+1}(x) = F(x, \tilde{\lambda}_{k-1}) - F(x, \tilde{\lambda}_k) + F(x_k, \tilde{\lambda}_k) - F(x_k, \tilde{\lambda}_{k-1}). \quad (80)$$

Proof. By (78), it has

$$\begin{aligned} \varphi_{k+1}(x) - F(x, \tilde{\lambda}_k) &= (1 - \alpha_k) [\varphi_k(x) - F(x, \tilde{\lambda}_{k-1})] \\ &\quad + (1 - \alpha_k) [F(x, \tilde{\lambda}_{k-1}) - F(x_k, \tilde{\lambda}_{k-1}) \\ &\quad - F(x, \tilde{\lambda}_k) + F(x_k, \tilde{\lambda}_k)] - \alpha_k [F(x, \tilde{\lambda}_k) \\ &\quad - F(x_{k+1}, \tilde{\lambda}_k) + \frac{1}{\beta_k} (x_{k+1} - y_k)^\top (x - x_{k+1})]. \end{aligned} \quad (81)$$

Because x_{k+1} is the minimizer of (34), we have

$$A^\top \tilde{\lambda}_k - \sigma A^\top (Ax_{k+1} - b) - \frac{1}{\beta_k} (x_{k+1} - y_k) \in \partial f(x_{k+1}). \quad (82)$$

Then, by the definition of $F(x, \tilde{\lambda}_k)$, it implies

$$A^\top \tilde{\lambda}_k - \sigma A^\top (Ax_{k+1} - b) - \frac{1}{\beta_k} (x_{k+1} - y_k) + \sigma A^\top \left(Ax_{k+1} - b - \frac{\tilde{\lambda}_k}{\sigma} \right) = -\frac{1}{\beta_k} (x_{k+1} - y_k) \in \partial F(x_{k+1}, \tilde{\lambda}_k). \quad (83)$$

From the convexity of $F(x, \tilde{\lambda}_k)$, for any $x \in \text{dom}(f)$, we have

$$F(x, \tilde{\lambda}_k) - F(x_{k+1}, \tilde{\lambda}_k) + \frac{1}{\beta_k} (x_{k+1} - y_k)^\top (x - x_{k+1}) \geq 0. \quad (84)$$

Substituting it into (81), then (79) can be proved.

The following lemma shows that, for any $k \geq 0$, if v_k^* is the minimizer of $\varphi_k(x)$ and we write $\varphi_k(x)$ as the canonical form

$$\varphi_k(x) = \varphi_k(v_k^*) + \left(\frac{a_k}{2} \right) \|x - v_k^*\|^2, \quad (85)$$

then $\varphi_k(v_k^*)$ is the upper bound of $F(x, \lambda)$ on x_k . $\square$

Lemma 5. Suppose that $\varphi_k(x)$ and $F(x, \tilde{\lambda}_k)$ are defined as (76)–(78), respectively, and the other sequences are generated by algorithm C-APPA. Then, $\varphi_0(v_0^*) \geq F(x_0, \tilde{\lambda}_{-1})$, and for all $k \geq 1$, it has

$$\varphi_k(v_k^*) \geq F(x_k, \tilde{\lambda}_{k-1}) + \frac{2-c}{2\beta_k} \|x_k - y_{k-1}\|^2 \geq F(x_k, \tilde{\lambda}_{k-1}). \quad (86)$$

Proof. We prove it by induction. By (77) and $v_0^* = x_0$, (86) is true for $k = 0$. Now, suppose for k , $\varphi_k(v_k^*) \geq F(x_k, \tilde{\lambda}_{k-1})$. Then, by (78), (85), (37), and (38), it implies

$$\begin{aligned} \varphi_{k+1}(v_{k+1}^*) &= (1 - \alpha_k) \varphi_k(v_k^*) + \frac{(1 - \alpha_k) a_k}{2} \|v_{k+1}^* - v_k^*\|^2 - (1 - \alpha_k) [F(x_k, \tilde{\lambda}_{k-1}) - F(x_k, \tilde{\lambda}_k)] + \alpha_k F(x_{k+1}, \tilde{\lambda}_k) - \frac{\alpha_k}{\beta_k} (x_{k+1} - y_k)^\top (v_{k+1}^* - x_{k+1}) \\ &\geq (1 - \alpha_k) F(x_k, \tilde{\lambda}_{k-1}) + \frac{\alpha_k^2}{2a_{k+1}\beta_k} \|x_{k+1} - y_k\|^2 - (1 - \alpha_k) [F(x_k, \tilde{\lambda}_{k-1}) - F(x_k, \tilde{\lambda}_k)] + \alpha_k F(x_{k+1}, \tilde{\lambda}_k) - \frac{\alpha_k}{\beta_k} (x_{k+1} - y_k)^\top (v_{k+1}^* - x_{k+1}) \\ &= (1 - \alpha_k) [F(x_k, \tilde{\lambda}_k) - F(x_{k+1}, \tilde{\lambda}_k)] + F(x_{k+1}, \tilde{\lambda}_k) - \frac{\alpha_k^2}{2a_{k+1}\beta_k} \|x_{k+1} - y_k\|^2 - \frac{\alpha_k}{\beta_k} (x_{k+1} - y_k)^\top (v_k^* - x_{k+1}). \end{aligned} \quad (87)$$

Since $F(x, \tilde{\lambda}_k)$ is convex and $-1/\beta_k (x_{k+1} - y_k)^\top \in \partial F(x_{k+1}, \tilde{\lambda}_k)$, then

$$F(x_k, \tilde{\lambda}_k) - F(x_{k+1}, \tilde{\lambda}_k) \geq -\frac{1}{\beta_k} (x_{k+1} - y_k)^\top (x_k - x_{k+1}). \quad (88)$$

Substituting it into (87) and by using of (33), we have

$$\begin{aligned} \varphi_{k+1}(v_{k+1}^*) &\geq F(x_{k+1}, \tilde{\lambda}_k) - (1 - \alpha_k) \frac{1}{\beta_k} (x_{k+1} - y_k)^\top (x_k - x_{k+1}) \\ &\quad - \frac{\alpha_k^2}{2a_{k+1}\beta_k} \|x_{k+1} - y_k\|^2 - \frac{\alpha_k}{\beta_k} (x_{k+1} - y_k)^\top (v_k^* - x_{k+1}) \\ &= F(x_{k+1}, \tilde{\lambda}_k) + \frac{1}{\beta_k} \left(1 - \frac{\alpha_k^2}{2a_{k+1}\beta_k} \right) \|x_{k+1} - y_k\|^2. \end{aligned} \quad (89)$$

Notice that a_{k+1} and α_k are defined as (37) and (33), respectively, and it deduces

$$\frac{\alpha_k^2}{a_{k+1}\beta_k} = \frac{\alpha_k^2}{(1 - \alpha_k)a_k\beta_k} = c < 2, \quad (90)$$

which means

$$\varphi_{k+1}(v_{k+1}^*) \geq F(x_{k+1}, \tilde{\lambda}_k) + \frac{2-c}{2\beta_k} \|x_{k+1} - y_k\|^2 \geq F(x_{k+1}, \tilde{\lambda}_k). \quad (91)$$

The following bounded conclusion of $A_k(c)$ is similar to the one of β_k given in Lemma 2.2 by Güler [10], but it introduces a parameter c in order to have more flexibility. So, here we omit the proving detail for concision. $\square$

Lemma 6. Let $A_k(c) = \prod_{i=0}^{k-1} (1 - \alpha_i)$. Then, for all $k > 0$,

$$\frac{1}{\left(1 + \sqrt{ca_0} \sum_{j=0}^{k-1} \sqrt{\beta_j}\right)^2} \leq A_k(c) \leq \frac{1}{\left(1 + (\sqrt{ca_0}/2) \sum_{j=0}^{k-1} \sqrt{\beta_j}\right)^2}. \quad (92)$$

Now, we give the main convergence conclusion of C-APPA.

Theorem 2. Suppose that the solution set of (5) is nonempty and bounded. Denote $(\bar{x}, \bar{\lambda})$ as a solution satisfying the KKT system and $\inf f(\bar{x}) > -\infty$. Let x_0 be an initial feasible point and $\{(x_k, \lambda_k)\}$ is generated by algorithm C-APPA.

(1) Denote $\Delta_{kj} = \prod_{i=1}^j (1 - \alpha_{k-i})$. Then, it has

$$\begin{aligned} \mathcal{L}_\sigma(x_{k+1}, \bar{\lambda}_k) - \mathcal{L}_\sigma(\bar{x}, \bar{\lambda}) + \sigma(1 - \alpha_k) \sum_{j=0}^k \Delta_{kj} \alpha_{k-j-1} \|Ax_{k-j} - b\|^2 \\ \leq \frac{a_0 A_{k+1}(c)}{2} \|\bar{x} - x_0\|^2. \end{aligned} \quad (93)$$

(2) If, for any $k = 1, 2, \dots$, there exists $\underline{\alpha} > 0$ and $\underline{\beta} > 0$ such that $\alpha_k \geq \underline{\alpha}$ and $\beta_k \geq \underline{\beta}$, then

$$\mathcal{L}_\sigma(x_{k+1}, \bar{\lambda}_k) - f(\bar{x}) \leq \frac{2\|\bar{x} - x_0\|^2}{ck^2 \underline{\beta}} = O\left(\frac{1}{k^2}\right). \quad (94)$$

Proof. Let $\theta_0(x) = 0, p_0(x) = 0, \alpha_{-1} = 0$, and

$$\begin{aligned} p_{k+1}(x) &= (1 - \alpha_{k-1})p_k(x) + \theta_{k+1}(x), \quad k = 0, 1, 2, \dots, \\ \Delta_{k0} &= 1 - \alpha_k, \Delta_{kj} = \prod_{i=1}^j (1 - \alpha_{k-i}), \quad k > 0. \end{aligned} \quad (95)$$

It can be seen that

$$p_{k+1}(x) = \theta_{k+1}(x) + \sum_{j=1}^k \Delta_{kj} \theta_{k+1-j}(x). \quad (96)$$

By (79), it deduces

$$\begin{aligned} \varphi_{k+1}(x) - F(x, \bar{\lambda}_k) &\leq \left(\prod_{i=0}^k (1 - \alpha_i) \right) [\varphi_0(x) - F(x, \bar{\lambda}_{-1})] \\ &\quad + (1 - \alpha_k) p_{k+1}(x). \end{aligned} \quad (97)$$

Also by Lemma 5, we have

$$\begin{aligned} F(x_{k+1}, \bar{\lambda}_k) - F(x, \bar{\lambda}_k) &\leq \varphi_{k+1}(x) - F(x, \bar{\lambda}_k) \\ &\leq \left(\prod_{i=0}^k (1 - \alpha_i) \right) [\varphi_0(x) - F(x, \bar{\lambda}_{-1})] \\ &\quad + (1 - \alpha_k) p_{k+1}(x). \end{aligned} \quad (98)$$

Then, by using of (76)–(78), it obtains

$$\begin{aligned} \mathcal{L}_\sigma(x_{k+1}, \bar{\lambda}_k) - \mathcal{L}_\sigma(\bar{x}, \bar{\lambda}) &= f(x_{k+1}) - \bar{\lambda}_k^\top (Ax_{k+1} - b) + \frac{\sigma}{2} \|Ax_{k+1} - b\|^2 - f(\bar{x}) \leq \left[f(x_{k+1}) + \frac{\sigma}{2} \|Ax_{k+1} - b - \frac{\bar{\lambda}_k}{\sigma}\|^2 \right] \\ &\quad - \left[f(\bar{x}) + \frac{\sigma}{2} \|A\bar{x} - b - \frac{\bar{\lambda}_k}{\sigma}\|^2 \right] \leq \left(\prod_{i=0}^k (1 - \alpha_i) \right) [\varphi_0(\bar{x}) - F(\bar{x}, \bar{\lambda}_{-1})] + (1 - \alpha_k) p_{k+1}(\bar{x}) \\ &= \left(\prod_{i=0}^k (1 - \alpha_i) \right) \left[f(\bar{x}) + \frac{\sigma}{2} \|A\bar{x} - b - \frac{\bar{\lambda}_0}{\sigma}\|^2 + \frac{a_0}{2} \|\bar{x} - x_0\|^2 - f(\bar{x}) - \frac{\sigma}{2} \|A\bar{x} - b - \frac{\bar{\lambda}_0}{\sigma}\|^2 \right] \\ &\quad + (1 - \alpha_k) p_{k+1}(\bar{x}) = \frac{a_0 A_{k+1}(c)}{2} \|\bar{x} - x_0\|^2 + (1 - \alpha_k) p_{k+1}(\bar{x}). \end{aligned} \quad (99)$$

From (81), we have

$$\begin{aligned} \theta_{k+1}(x) &= \frac{\sigma}{2} \left[\left\| Ax - b - \frac{\bar{\lambda}_{k-1}}{\sigma} \right\|^2 - \left\| Ax - b - \frac{\bar{\lambda}_k}{\sigma} \right\|^2 \right] \\ &\quad + \frac{\sigma}{2} \left[\left\| Ax_k - b - \frac{\bar{\lambda}_k}{\sigma} \right\|^2 - \left\| Ax_k - b - \frac{\bar{\lambda}_{k-1}}{\sigma} \right\|^2 \right] \\ &= (\bar{\lambda}_k - \bar{\lambda}_{k-1})^\top (Ax - b) + (\bar{\lambda}_{k-1} - \bar{\lambda}_k)^\top (Ax_k - b), \end{aligned} \quad (100)$$

which obtains

$$\theta_{k+1}(\bar{x}) = (\bar{\lambda}_{k-1} - \bar{\lambda}_k)^\top (Ax_k - b). \quad (101)$$

By (36) and (39), we deduce

$$\bar{\lambda}_k - \bar{\lambda}_{k-1} = \alpha_{k-1} \sigma (Ax_k - b). \quad (102)$$

Thus,

$$\begin{aligned} \theta_{k+1}(\bar{x}) &= -\alpha_{k-1} \sigma \|Ax_k - b\|^2, \\ p_{k+1}(\bar{x}) &= \theta_{k+1}(\bar{x}) + \sum_{j=1}^k \Delta_{kj} \theta_{k+1-j}(\bar{x}) \\ &= -\sigma \left[\alpha_{k-1} \|Ax_k - b\|^2 + \sum_{j=1}^k \Delta_{kj} \alpha_{k-j-1} \|Ax_{k-j} - b\|^2 \right] \leq 0. \end{aligned} \quad (103)$$

From (99), it has

$$\mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k) - \mathcal{L}_\sigma(\bar{x}, \bar{\lambda}) + \sigma(1 - \alpha_k) \sum_{j=0}^k \Delta_{kj} \alpha_{k-j-1} \|Ax_{k-j} - b\|^2 \leq \frac{a_0 A_{k+1}(c)}{2} \|\bar{x} - x_0\|^2. \quad (104)$$

When the assumptions are true, $(x_k, \tilde{\lambda}_k)$ converges to a certain KKT solution $(\bar{x}, \bar{\lambda})$ from Theorem 1. Furthermore, from Lemma 6, it obtains

$$\begin{aligned} \mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k) - f(\bar{x}) &= \mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k) - \mathcal{L}_\sigma(\bar{x}, \bar{\lambda}) \leq \frac{a_0 A_{k+1}(c)}{2} \|\bar{x} - x_0\|^2 \\ &\leq \frac{a_0 \|\bar{x} - x_0\|^2}{\left(1 + (\sqrt{ca}/2) \sum_{j=0}^{k-1} \sqrt{\beta_j}\right)^2} \leq \frac{2a_0 \|\bar{x} - x_0\|^2}{ca \left(\sum_{j=0}^{k-1} \sqrt{\beta_j}\right)^2}. \end{aligned} \quad (105)$$

While $\beta_k \geq \underline{\beta}$, it holds that

$$\mathcal{L}_\sigma(x_{k+1}, \tilde{\lambda}_k) - f(\bar{x}) \leq \frac{2\|\bar{x} - x_0\|^2}{ck^2 \underline{\beta}} = O\left(\frac{1}{k^2}\right). \quad (106)$$

□

5. Conclusions

In this paper, a framework of accelerated PPA is presented for constrained convex optimization. The algorithm can be seen as an extension to Güler's methods for unconstrained optimization and linear programming problems. Different from [11], where the accelerated proximity point method is used on the dual problem of the linear programming, our algorithm directly uses the accelerating technique both on the proximal point operator and the Lagrange multiplier for the original problem in the iterations by constructing appropriate auxiliary sequences with the quadratic convex functions and auxiliary points. We prove that the sequence generated by the algorithm converges to a KKT solution under suitable conditions. Also, it retains a similar convergence rate on objective function values.

This paper has paid more attention on analyzing the convergence of the accelerated PPA for solving convex minimization with linear constraints. In the iterations of C-APPA, computing the proximal mapping of $F(x, \tilde{\lambda}_k)$ is required. Usually, the proximal point of y_k is hard to know exactly. Then, series of x_{k+1} , v_{k+1}^* , y_k , and other auxiliary quantities being updated and calculated by a nonprecision rule should be considered. Several inexactly accelerated PPAs, the convergent properties under some proper conditions, and the numerical results are going to be given in another paper of ours. And more research on different inexact conditions, their further applications, and testings in practice would be further carried out in the future.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors express the gratitude for the cooperation of Congzuo Municipal Bureau of Statistics during this research. Also, this work was supported in part by the Key Project of Guangxi Provincial Natural Science Foundation under Grants 2017GXNSFDA198014, 2018GXNSFDA050014, and 2019AC04004.

References

- [1] P. Ravikumar, A. Agarwal, and M. Wainwright, "Message-passing for graphstructured linear programs: proximal methods and rounding schemes," *Journal of Machine Learning Research*, vol. 11, pp. 1043–1080, 2010.
- [2] N. Pustelnik, J.-C. Pesquet, and C. Chaux, "Relaxing tight frame condition in parallel proximal methods for signal restoration," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 968–973, 2012.
- [3] J. Bai, J. Li, P. Dai et al., "General parameterized proximal point algorithm with applications in statistical learning," *International Journal of Computer Mathematics*, vol. 96, no. 1–4, pp. 199–215, 2019.
- [4] B. Martinet, "Régularisation d'Inéquations variationnelles par approximations successives," *French Journal of Computing and Operational Research*, vol. 4, 1970.
- [5] B. Martinet, "Détermination approchée d'un point fixe d'une application pseudo-contraction," *Comptes rendus de l'Académie des Sciences*, vol. 274A, pp. 163–165, 1972.
- [6] R. Rockafellar, "Monotone operators and the proximal point algorithm," *SIAM Journal on Control and Optimization*, vol. 14, no. 5, pp. 877–898, 1976.
- [7] R. Rockafellar, "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 97–116, 1976.
- [8] O. Güler, "On the convergence of the proximal point algorithm for convex minimization," *SIAM Journal on Control and Optimization*, vol. 29, no. 2, pp. 403–419, 1991.
- [9] Y. Nesterov, "A method for solving the convex programming problem with convergence rate $O(1/k^2)$," *Doklady Akademii Nauk: SSSR*, vol. 269, no. 3, pp. 543–547, 1983.
- [10] O. Güler, "New proximal point algorithms for convex minimization," *SIAM Journal on Optimization*, vol. 2, no. 4, pp. 649–664, 1992.
- [11] O. Güler, "Augmented Lagrangian algorithms for linear programming," *Journal of Optimization Theory and Applications*, vol. 75, no. 3, pp. 445–470, 1992.
- [12] J. Birge, L. Qi, and Z. Wei, "A general approach to convergence properties of some methods for nonsmooth convex optimization," *Applied Mathematics and Optimization*, vol. 38, no. 2, pp. 141–158, 1998.

- [13] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [14] Y.-f. Ke and C.-f. Ma, "An accelerated augmented Lagrangian method for linearly constrained convex programming with the rate of convergence $O(1/k^2)$," *Applied Mathematics-A Journal of Chinese Universities*, vol. 32, no. 1, pp. 117–126, 2017.
- [15] Y. Xu, "Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming," *SIAM Journal on Optimization*, vol. 27, no. 3, pp. 1459–1484, 2017.
- [16] Y. Zhang, J. Wu, and L. Zhang, "The rate of convergence of proximal method of multipliers for equality constrained optimization problems," *Optimization Letters*, vol. 14, no. 6, p. 1599, 2019.
- [17] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NY, USA, 1970.
- [18] R. Rockafellar and R. Wets, *Variational Analysis*, Springer, Berlin, Germany, 1998.