IMPROVED LAGRANGIAN-PPA BASED PREDICTION CORRECTION METHOD FOR LINEARLY CONSTRAINED CONVEX OPTIMIZATION

YANFEI YOU
Aliyun School of Big Data, Changzhou University
Jiangsu, China

SUHONG JIANG*
School of Accounting, Nanjing University of Finance and Economics
Jiangsu, China

(Communicated by Hoang Xuan Phu)

Abstract. This paper presents an improved Lagrangian-PPA based prediction correction method to solve linearly constrained convex optimization problem. At each iteration, the predictor is achieved by minimizing the proximal Lagrangian function with respect to the primal and dual variables. These optimization subproblems involved either admit analytical solutions or can be solved by a fast algorithm. The new update is generated by using the information of the current iterate and the predictor, as well as an appropriately chosen stepsize. Compared with the existing PPA based method, the parameters are relaxed. We also establish the convergence and convergence rate of the proposed method. Finally, numerical experiments are conducted to show the efficiency of our Lagrangian-PPA based prediction correction method.

1. Introduction. In this paper, we concentrate on the following linearly constrained optimization problem:

\[ \min \{ \theta(x) \mid Ax = b \ (\text{or} \ Ax \geq b), \ x \in \mathcal{X} \} \]

where \( \theta(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) is a convex function, \( A \in \mathbb{R}^{m \times n} \), \( b \in \mathbb{R}^m \), and \( \mathcal{X} \subset \mathbb{R}^n \) is a closed convex set. Problem (1) plays a vital role in numerical optimization and has found many applications in matrix completion, statistics, engineering and finance. An illustrative example is the nearest correlation matrix problem, that is to find an approximated correlation matrix to any given matrix \( C \in \mathbb{R}^{n \times n} \), i.e.,

\[ \min \{ \frac{1}{2} \| X - C \|_F^2 \mid \text{diag}(X) = e, \ X \in \mathcal{S}^n_+ \}, \]

where \( e \in \mathbb{R}^n \) is the vector whose entries are all 1s, \( \mathcal{S}^n_+ \) denotes the cone of positive definite symmetric matrices, \( \text{diag}(X) \) is the vector of diagonal elements of \( X \), and

2020 Mathematics Subject Classification. Primary: 90C25, 90C30; Secondary: 65K05, 65K10.

Key words and phrases. Linearly constrained convex optimization, Lagrangian-PPA, proximal point algorithm, contraction methods, convergence rate.

The first author is supported by the NSFC grant 11701048. The second author is supported by the NSFC grant 12101297.

* Corresponding author: Suhong Jiang.
∥ · ∥_F denotes the matrix Frobenius norm \( \|X\|_F = (\text{trace}(X^\top X))^{\frac{1}{2}} \). Clearly, model (2) takes a special form of (1).

It is well known that proximal point algorithm (short as PPA) is an efficient approach for solving the problem (1), which originally proposed by Martinet [19, 20] and Rockafellar [21]. Although PPA has a number of advantages, solving the PPA subproblem is almost as difficult as the entire original problem. To overcome the drawback of classical PPA, a customized PPA was proposed in [13] and further analyzed in [11]. Along another line, You et al. proposed a simple method named L-PPA in [22]. In [9], He gave a review on the developed customized PPA in H-norm (H is a positive definite matrix), and obtained more general and simple proof of contractive convergence by using the variational inequality approach. In [17], Ma et al. proposed a class of customized proximal point algorithms for linearly constrained convex optimization problems, which contained several existing customized proximal point algorithms. A new customized proximal point algorithm for linearly constrained convex optimization problem has been proposed in [15], which do not involve relaxation step.

In this paper, we aim to develop an efficient proximal point operator based first-order algorithm to solve problem (1). Throughout this paper, we assume that the set collected by the KKT points, defined by \( \Omega^* \), is nonempty; and further assume the proximal point operator of \( \theta \)

\[
p^\theta_p(a) := \arg\min_x \{ \theta(x) + \frac{r}{2}\|x-a\|^2 \mid x \in \mathcal{X} \}
\]

either admits an analytical form or can be efficiently solved up to a high precision. To deal with the linear constraints and use the proximal point operator information, a typical method is to reformulate the linearly constrained optimization problem (1) as an equivalent variational inequality [6]. Specifically, the Lagrangian function of (1) is

\[
L(x, \lambda) = \theta(x) - \lambda^\top(Ax - b),
\]

where \( \lambda \) is the Lagrangian multiplier defined on \( \Lambda \) with

\[
\Lambda = \begin{cases} \mathbb{R}^m, & \text{for the equality constraints} \quad Ax = b, \\ \mathbb{R}^m_+, & \text{for the inequality constraints} \quad Ax \geq b. \end{cases}
\]

Let \((x^*, \lambda^*)\) be a saddle point of the Lagrangian function, then we have

\[
L_{x \in \mathcal{X}}(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L_{x \in \mathcal{X}^*}(x, \lambda^*).
\]

Given the nonempty attribute of the KKT points set \( \Omega^* \), solving (1) is equivalent to finding a saddle point \((x^*, \lambda^*)\) of the Lagrangian function (4). This amounts to

\[
\begin{cases} x^* \in \mathcal{X}, & \theta(x) - \theta(x^*) + (x - x^*)^\top(-A^\top \lambda^*) \geq 0, \quad \forall x \in \mathcal{X}, \\ \lambda^* \in \Lambda, & (\lambda - \lambda^*)^\top(Ax^* - b) \geq 0, \quad \forall \lambda \in \Lambda,
\end{cases}
\]

or equivalently, a compact mixed variational inequality:

\[
\text{MVI}(\Omega, F, \theta) \quad u^* \in \Omega, \quad \theta(x) - \theta(x^*) + (u - u^*)^\top F(u^*) \geq 0, \quad \forall u \in \Omega.
\]

where

\[
u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} -A^\top \lambda \\ Ax - b \end{pmatrix}, \quad \text{and} \quad \Omega = \mathcal{X} \times \Lambda.
\]

It is easily verified that \( F \) is a skew-symmetric linear operator and then problem (6) is a monotone variational inequality (abbreviated as MVI). In this paper, we treat the linearly constrained convex optimization (1) in the frame of MVI(\( \Omega, F, \theta \)).
where and

\[ \theta(x) - \theta(x^{k+1}) + (u - u^{k+1})^T \{ F(u^{k+1}) + r_k(u^{k+1} - u^k) \} \geq 0, \forall u \in \Omega. \] (8)

The generated sequence \( \{u^k\} \) satisfies

\[ \|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - r_k(2 - r_k)\|u^k - u^{k+1}\|^2, \forall u^* \in \Omega^*, \]

and is Fejér monotone with respect to the solution set \( \Omega^* \). However, solving the PPA subproblem (8) is almost as difficult as the MVI (6), therefore the classical PPA only has theoretical meaning and is not often used in practical computation. To alleviate the difficulty, a customized PPA was proposed in [13] and further analyzed in [11]. The \( k \)-th iteration of the customized PPA begins with a given \( u^k = (x^k, \lambda^k) \), and firstly generates a \( \tilde{u}^k \) via the following scheme:

\[ \tilde{u}^k \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^T \{ F(\tilde{u}^k) + G(\tilde{u}^k - u^k) \} \geq 0, \forall u \in \Omega, \] (9)

where

\[ G = \begin{pmatrix} rI_n & -\nu A^T \\ -A^T & sI_m \end{pmatrix}, \quad \nu \in [-1, 1]. \] (10)

Surprisingly, the variational inequality (9) can be decomposed as two easily solvable problems:

\[ \tilde{x}^k \in \mathcal{X}, \quad \theta(x) - \theta(\tilde{x}^k) + (x - \tilde{x}^k)^T \{ -A^T(\tilde{\lambda}^k + \nu(\tilde{\lambda}^k - \lambda^k)) + r(\tilde{x}^k - x^k) \} \geq 0, \forall x \in \mathcal{X}, \] (11a)

and

\[ \tilde{\lambda}^k \in \Lambda, \quad (\lambda - \tilde{\lambda}^k)^T \{ (Ax^k - b) + s(\tilde{\lambda}^k - \lambda^k) \} \geq 0, \forall \lambda \in \Lambda, \] (11b)

where the solution of (11b) can be obtained directly by

\[ \tilde{\lambda}^k = P_\Lambda[\lambda^k - \frac{1}{s}(Ax^k - b)], \]

with \( P_\Lambda \) being the Euclidean projector onto \( \Lambda \), and the solution of (11a) is \( p_\mathcal{X}^G(x^k + \frac{1}{\nu} A^T(2\tilde{\lambda}^k - \lambda^k)) \). Moreover, by updating

\[ u^{k+1} = u^k - \gamma(u^k - \tilde{u}^k), \gamma \in (0, 2) \]

and restricting

\[ \frac{4rs}{(1 + \theta)^2} > \|A^T\|, \theta \in (-1, 1), \] (12)

the authors in [13] prove that the generated sequence \( \{u^k\} \) satisfies

\[ \|u^{k+1} - u^*\|_G^2 \leq \|u^k - u^*\|_G^2 - \gamma(2 - \gamma)\|u^k - \tilde{u}^k\|_G^2, \forall u^* \in \Omega^*. \]

A typical choice of \( \nu \) is 1 and to ensure the convergence the parameters \( r \) and \( s \) need to satisfy the following condition \( rs > \|A^T\| \).

Along another line, a simple method named L-PPA in [22] for solving (6) has been proposed. Given \( u^k = (x^k, \lambda^k) \), L-PPA updates a predictor \( \tilde{u}^k = (\tilde{x}^k, \tilde{\lambda}^k) \) by minimizing (resp. maximizing) the Lagrangian function with proximal terms as following:

\[ \tilde{x}^k = \text{Argmin}\{L(x, \lambda^k) + \frac{r}{2}\|x - x^k\|^2 | x \in \mathcal{X}\}, \] (13a)
and
\[ \hat{\lambda}^k = \text{Arg max}\{L(\hat{x}^k, \lambda) - \frac{s}{2}\|\lambda - \hat{\lambda}^k\|^2 | \lambda \in \Lambda \}. \] (13b)

If one directly takes \( \tilde{u}^k \) as the new iterate, it corresponds the classical Arrow-Hurwicz method [1]. The Arrow-Hurwicz method has been successfully applied in Rudin Osher and Fatemi (ROF) image denoising problem [23], but its convergence is only established in the situation that \( rs \) is very large [5]. To overcome this difficulty, the authors in [22] only consider \( \tilde{u}^k \) as a predictor and update the new iterate as:
\[ u^{k+1} = u^k - \beta_k W(u^k - \tilde{u}^k), \] (14)
where \( \beta_k \in (0, 2) \) and \( W = \begin{pmatrix} I_n & \frac{1}{r}A^T \\ 0 & I_m \end{pmatrix} \). The restriction on the parameters \( r \) and \( s \) is relaxed to \( rs \geq \frac{1}{2}\|A^TA\| \), which is also weaker than that in [13].

In this paper, similar to [22], we use the proximal minimization of Lagrangian function to generate a predictor like (13). However, we develop a new search direction in the correction step and further relax the restriction on the parameters \( r \) and \( s \), which often can accelerate the proximal point operator based algorithms. To verify the efficiency of our algorithm, we conduct numerical experiments on the nearest correlation matrix problems and the results show the proposed methods take fewer iterations and also spend much less computing time compared with PPA (11) and L-PPA (14).

The remainder of this paper is organized as follows. Section 2 introduces the predictor based on the Lagrangian function. In Section 3, we present our method based on contraction and prove its convergence. An \( o(1/\varepsilon) \) convergence rate of the proposed method is also established in ergodic sense. In Section 4, we implement our proposed method to solve the correction matrix problem and compare its performance with PPA and L-PPA. Section 5 concludes this paper.

2. Predictor using Lagrangian function. In this section, we discuss the implementation and the properties of the predictor generated by (13). By direct computation, the predictor \((\hat{x}^k, \hat{\lambda}^k)\) takes the following form

For given \((x^k, \lambda^k)\) and \( r, s > 0 \),
\[ \hat{x}^k = p^r_{\theta}(x^k + \frac{1}{r}A^T \lambda^k) \] (15a)
and
\[ \hat{\lambda}^k = P_{\Lambda}[\lambda^k - \frac{1}{s}(Ax^k - b)], \] (15b)
which can be easily obtained in the implementation.

The following lemma established in [22] is useful for our subsequent discussions.

**Lemma 2.1.** For given \( u^k = (x^k, \lambda^k) \), let \( \tilde{u}^k \) be generated by (15). Then we have \( \tilde{u}^k \in \Omega \), \( \theta(x) - \theta(\tilde{x}^k) + (u - \tilde{u}^k)^\top F(\tilde{u}^k) \geq (u - \tilde{u}^k)^\top Q(u^k - \tilde{u}^k), \) \( \forall u \in \Omega \), (16) where
\[ Q = \begin{pmatrix} rI_n & A^T \\ 0 & sI_m \end{pmatrix}. \] (17)

**Lemma 2.2.** For given \( u^k = (x^k, \lambda^k) \), let \( \tilde{u}^k \) be generated by (15). Then we have
\( (u^k - u^*)^\top Q(u^k - \tilde{u}^k) \geq (u^k - \tilde{u}^k)^\top Q(u^k - \tilde{u}^k), \) \( \forall u^* \in \Omega^* \). (18)
Next, we introduce two additional parameters $\tau \in (\frac{1}{2}, \frac{7}{2})$ and $\eta \in [0, 1]$ satisfying
\[
rs > \frac{\tau^2 - \eta + \eta^2}{(2\tau - 1)^2} A^T A,
\]
where $r$ and $s$ are the parameters in (15). With this preparation, we are able to prove the following lemma which is key to our proposed method.

**Lemma 2.3.** For given $u^k = (x^k, \lambda^k)$, let $\tilde{u}^k$ be generated by (15). Then we have
\[
(u^k - \tilde{u}^k)^T Q (u^k - \tilde{u}^k) \geq \frac{7 - 2\tau}{4(2\tau + 1)} u^k - \tilde{u}^k \|_{\tilde{R}},
\]
where
\[
R = \begin{pmatrix} rI_n & 0 \\ 0 & sI_m \end{pmatrix}.
\]

**Proof.** We observe the right hand side of (18). Because
\[
(u^k - \tilde{u}^k)^T Q (u^k - \tilde{u}^k)
\]
\[
= r \|x^k - \tilde{x}^k\|^2 + s \|\lambda^k - \tilde{\lambda}^k\|^2 + (x^k - \tilde{x}^k)^T A^T (\lambda^k - \tilde{\lambda}^k)
\]
\[
\geq r \|x^k - \tilde{x}^k\|^2 - \frac{1}{2} \|x^k - \tilde{x}^k\| \cdot A^T (\lambda^k - \tilde{\lambda}^k)
\]
\[
+ s \|\lambda^k - \tilde{\lambda}^k\|^2 - \frac{1}{2} \|\lambda^k - \tilde{\lambda}^k\| \cdot A(x^k - \tilde{x}^k)
\]
\[
\geq r \|x^k - \tilde{x}^k\|^2 - \frac{1}{4} \left\{ r \|x^k - \tilde{x}^k\|^2 + \frac{1}{r} \|A^T (\lambda^k - \tilde{\lambda}^k)\|^2 \right\}
\]
\[
+ s \|\lambda^k - \tilde{\lambda}^k\|^2 - \frac{1}{4} \left\{ s \|\lambda^k - \tilde{\lambda}^k\|^2 + \frac{1}{s} \|A(x^k - \tilde{x}^k)\|^2 \right\}
\]
\[
= \frac{3}{4} \left\{ r \|x^k - \tilde{x}^k\|^2 + s \|\lambda^k - \tilde{\lambda}^k\|^2 \right\} - \frac{1}{4} \left\{ \frac{1}{s} \|A(x^k - \tilde{x}^k)\|^2 + \frac{1}{r} \|A^T (\lambda^k - \tilde{\lambda}^k)\|^2 \right\},
\]
using the condition (19) and $\eta \in [0, 1]$, we obtain
\[
rs > \frac{\tau^2 - \eta + \eta^2}{(2\tau - 1)^2} A^T A \geq \frac{2\tau + 1}{4(2\tau - 1)} A^T A
\]
and
\[
(u^k - \tilde{u}^k)^T Q (u^k - \tilde{u}^k) \geq \frac{7 - 2\tau}{4(2\tau + 1)} \left\{ r \|x^k - \tilde{x}^k\|^2 + s \|\lambda^k - \tilde{\lambda}^k\|^2 \right\}
\]
\[
= \frac{7 - 2\tau}{4(2\tau + 1)} \|u^k - \tilde{u}^k\|^2_{\tilde{R}}.
\]
The lemma is proved. \hfill \square

3. **Lagrangian-PPA based contraction method.** In this section, we construct a contraction method based on the predictor (15). The sequence $\{u^k\}$ generated by the contraction methods in [7, 8, 10, 12] is Fejér monotone with respect to the solution set $\Omega^*$. We will prove a similar conclusion in this section.

**Definition 3.1.** Define two matrices $H$ and $M$ by
\[
H = \tau \begin{pmatrix} rI + \frac{\eta^2}{\tau} A^T \Lambda^{-1} A & \eta A^T \Lambda^{-1} \\ \eta A^T \Lambda^{-1} & \eta \Lambda^{-1} \end{pmatrix}, \quad M = \frac{1}{\tau} \begin{pmatrix} I_n & \frac{1-s}{r \tau} A^T \\ -\frac{s}{r} A & I_m \end{pmatrix},
\]
where $\Lambda = I + \frac{(1-\eta)\eta}{rs} AA^T$. 
It is not difficult to show
\[ Q = H M. \] (22)
and \( H \) is a symmetric positive definite matrix. For given \( u^k \) and the predictor \( \hat{u}^k \) generated by (15), we let \( u^{k+1}(\alpha) \) be an \( \alpha \)-dependent new iterate defined by
\[ u^{k+1}(\alpha) = u^k - \alpha M(u^k - \hat{u}^k). \] (23)
For the new iterate, a contraction-like property can be established in the following lemma.

**Lemma 3.2.** For given \( u^k = (x^k, \lambda^k) \), let \( \hat{u}^k \) be generated by (15) and \( u^{k+1}(\alpha) \) be defined by (23). Then for any \( \alpha \geq 0 \), we have
\[ \|u^k - u^*\|_H^2 - \|u^{k+1}(\alpha) - u^*\|_H^2 \geq q(\alpha), \quad \forall u^* \in \Omega^*, \] where
\[ q(\alpha) = 2\alpha(u^k - \hat{u}^k)^\top Q(u^k - \hat{u}^k) - \alpha^2\|M(u^k - \hat{u}^k)\|_H^2. \] (25)

**Proof.** For any but fixed \( u^* \in \Omega^* \), we define
\[ \vartheta(\alpha) = \|u^k - u^*\|_H^2 - \|u^{k+1}(\alpha) - u^*\|_H^2. \] (26)
By using (23), we get
\[ \vartheta(\alpha) = \|u^k - u^*\|_H^2 - \|(u^k - u^*) - \alpha M(u^k - \hat{u}^k)\|_H^2 \\
= 2\alpha(u^k - u^*)^\top H M(u^k - \hat{u}^k) - \alpha^2\|M(u^k - \hat{u}^k)\|_H^2. \]
To the term \((u^k - u^*)^\top H M(u^k - \hat{u}^k)\) in the right hand side of the last equation, using (18) and \( Q = H M \) (see (22)) gives
\[ \vartheta(\alpha) \geq q(\alpha), \]
immediately and the lemma is proved. \[\square\]

Note that \( q(\alpha) \) in (25) is a quadratic function and it reaches its maximum at
\[ \alpha_k^* = \frac{(u^k - \hat{u}^k)^\top Q(u^k - \hat{u}^k)}{\|M(u^k - \hat{u}^k)\|_H^2}. \] (27)
Motivated by this, we are ready to present our Lagrangian-PPA prediction-correction method.

**Algorithm :** Lagrangian-PPA based prediction contraction method.
Let \( \tau \in (\frac{1}{2}, \frac{3}{2}) \), \( \gamma \in (0, 2) \), and \( \eta \in [0, 1] \). Let \( M \) and \( H \) be defined in (21). Take \( u^0 \in \mathbb{R}^{n+m} \). Choose \( r \) and \( s \) according to (19).

**Prediction step:** Generated the predictor \( \hat{u}^k \) via solving the primal dual procedure (15).

**Correction step:** Correct the predictor and generate the new iterate \( u^{k+1} \) via
\[ u^{k+1} = u^k - \alpha_k M(u^k - \hat{u}^k), \] (28a)
where
\[ \alpha_k^* = \frac{(u^k - \hat{u}^k)^\top Q(u^k - \hat{u}^k)}{\|M(u^k - \hat{u}^k)\|_H^2}, \quad \alpha_k = \gamma \alpha_k^*. \] (28b)
To prove the convergence of our method, we need to estimate the bound of the stepsizes.\[\square\]
Lemma 3.3. Under the condition (19), we have
\[ \frac{1}{2} < \alpha^*_k \leq c_0, \quad \forall k \geq 0, \]  
where \( c_0 = \frac{1}{2} \|Q + Q^\top\|_H^2 \) and \( \lambda_{\min}(M^\top HM) \) is the minimum eigenvalue of \( M^\top HM \).

Proof. First, we prove the left hand of (29). By using (27), it suffices to prove the following equivalent assertion:
\[ 2(u^k - \hat{u}^k)^\top Q(u^k - \hat{u}^k) - \|M(u^k - \hat{u}^k)\|_H^2 > 0. \]  
Note that
\[ 2(u^k - \hat{u}^k)^\top Q(u^k - \hat{u}^k) - \|M(u^k - \hat{u}^k)\|_H^2 = (u^k - \hat{u}^k)^\top (Q + Q^\top - M^\top HM)(u^k - \hat{u}^k). \]
Using \( Q = HM \) (see (22)), we have
\[ Q^\top + Q - M^\top HM = Q^\top + Q - M^\top Q. \]
Next, we prove that the above matrix is positive definite. By a manipulation, we get
\[
Q^\top + Q - M^\top Q &= \left( \begin{array}{cc}
2rI_n & A^\top \\
A & 2sI_m
\end{array} \right) - \frac{1}{\tau} \left( \begin{array}{cc}
I_n & -\frac{2}{\tau}A^\top \\
\frac{1-\eta}{\tau}A & I_m
\end{array} \right) \left( \begin{array}{cc}
rI_n & A^\top \\
0 & sI_m
\end{array} \right) \\
&= \frac{1}{\tau} \left( \begin{array}{cc}
2rI_n & \tau A^\top \\
\tau A & 2sI_m
\end{array} \right) - \left( \begin{array}{cc}
rI_n & (1-\eta)A^\top \\
(1-\eta)A & sI_m + \frac{1-\eta}{\tau}AA^\top
\end{array} \right)
\]
Setting \( P = \left( \begin{array}{cc}
\frac{I_n}{(2\tau - 1)r} & 0 \\
-\frac{1-\eta}{2(\tau - 1)r}A & I_m
\end{array} \right) \), then we have
\[
Q^\top + Q - M^\top Q &= \frac{1}{\tau} P^{-1} \left( \begin{array}{cc}
(2\tau - 1)rI_n & (\tau - 1 + \eta)A^\top \\
(\tau - 1 + \eta)A & (2\tau - 1)sI_m - \frac{1-\eta}{\tau}AA^\top
\end{array} \right) \left( P \right)^{-1} \left( \begin{array}{cc}
rI_n & A^\top \\
0 & sI_m - \frac{1-\eta}{\tau}AA^\top
\end{array} \right) \\
&= \frac{1}{\tau} P^{-1} \left( \begin{array}{cc}
(2\tau - 1)rI_n & 0 \\
0 & (2\tau - 1)sI_m - \frac{1-\eta}{\tau^2}AA^\top
\end{array} \right) \left( P \right)^{-1}.
\]
According to the condition (19), we know \( Q^\top + Q - M^\top Q \) is symmetric positive definite.

Second, we prove the right hand of (29). By using (27), we get
\[
\alpha^*_k &= \frac{(u^k - \hat{u}^k)^\top Q(u^k - \hat{u}^k)}{\|M(u^k - \hat{u}^k)\|_H^2} \\
&= \frac{1}{2} \frac{(u^k - \hat{u}^k)^\top (Q + Q^\top)(u^k - \hat{u}^k)}{\|M(u^k - \hat{u}^k)\|_H^2} \\
&\leq \frac{1}{2} \frac{\|Q + Q^\top\|_2^2 \|u^k - \hat{u}^k\|_2^2}{\lambda_{\min}(M^\top HM)\|u^k - \hat{u}^k\|_2^2} \\
&= \frac{1}{2} \frac{\|Q + Q^\top\|_2^2}{\lambda_{\min}(M^\top HM)} = c_0.
\]
Thus, we complete the proof of this assertion. \( \Box \)

The next theorem proves the Fejér monotone of the sequence generated by our proposed method, which is key to our convergence analysis.
\textbf{Theorem 3.4.} For given $u^k = (x^k, \lambda^k)$, let $\tilde{u}^k$ be generated by (15) and $u^{k+1}$ be updated by (28). Then we have

$$\|u^{k+1} - u^*\|^2_H \leq \|u^k - u^*\|^2_H - \frac{(7 - 2\tau)(2 - \gamma)}{8(2\tau + 1)} \|u^k - \tilde{u}^k\|^2_H. \quad (31a)$$

and

$$\|u^{k+1} - u^*\|^2_H \leq \|u^k - u^*\|^2_H - \frac{(7 - 2\tau)(2 - \gamma)\lambda_{\min}((M^{-1})^\top R M^{-1})}{8(2\tau + 1)c_0^2} \|[u^k - u^{k+1}]\|^2. \quad (31b)$$

\textbf{Proof.} From (24) and (28), it follows that

$$\|u^k - u^*\|^2_H - \|u^{k+1} - u^*\|^2_H \geq q(\gamma u^*_k).$$

By using (25), (27) and $\alpha^* \geq \frac{1}{2}$, we obtain

$$q(\gamma u^*_k) \geq \frac{1}{2}(2 - \gamma)(u^k - \tilde{u}^k)^\top Q(u^k - \tilde{u}^k). \quad (32)$$

In addition, it follows from (20) that

$$(u^k - \tilde{u}^k)^\top Q(u^k - \tilde{u}^k) \geq \frac{7 - 2\tau}{4(2\tau + 1)} \|u^k - \tilde{u}^k\|^2_H.$$ 

Substituting it in (32), we obtain

$$q(\gamma u^*_k) \geq \frac{(7 - 2\tau)(2 - \gamma)}{8(2\tau + 1)} \|u^k - \tilde{u}^k\|^2_H.$$ 

Using $u^k - \tilde{u}^k = \frac{1}{\alpha_k} M^{-1}(u^k - u^{k+1})$, we have

$$\|u^k - u^*\|^2_H - \|u^{k+1} - u^*\|^2_H \geq \frac{(7 - 2\tau)(2 - \gamma)}{8(2\tau + 1)} \|u^k - u^k\|^2_H \geq \frac{(7 - 2\tau)(2 - \gamma)}{8(2\tau + 1)c_0^2} \|M^{-1}(u^k - u^{k+1})\|^2_H \geq \frac{(7 - 2\tau)(2 - \gamma)\lambda_{\min}((M^{-1})^\top R M^{-1})}{8(2\tau + 1)c_0^2} \|[u^k - u^{k+1}]\|^2.$$ 

Thus, the theorem is proved.

The inequality (31) indicates that the proposed algorithm is a contraction method. With the contraction of the sequence, we are able to establish the convergence and convergence rate of our Lagrangian-PPA based prediction correction method.

\textbf{Theorem 3.5.} The sequence $\{u^k\}$ generated by the Lagrangian-PPA based prediction correction method converges to some $u^\infty$ which belongs to $\Omega^*$, and the proposed algorithm can achieve an approximated solution $u^k$ such that

$$\|u^k - u^{k+1}\| \leq \varepsilon$$ 

at most $o(\frac{1}{\tau})$ iterations.

\textbf{Proof.} According to (31), it holds that $\{u^k\}$ is bounded and

$$\lim_{k \to \infty} \|u^k - \tilde{u}^k\| = 0. \quad (34)$$
So, \( \{ \tilde{u}^k \} \) is also bounded. Let \( u^\infty \) be a cluster point of \( \{ \tilde{u}^k \} \) and \( \{ \tilde{u}^k_j \} \) be a subsequence which converges to \( u^\infty \). It follows from (16) that
\[
\tilde{u}^k_j \in \Omega, \quad \theta(x) - \theta(\tilde{x}^k_j) + (u - \tilde{u}^k_j)^\top F(\tilde{x}^k_j) \geq (u - u^k_j)^\top Q(u^k_j - \tilde{u}^k_j), \quad \forall u \in \Omega.
\]
Since the matrix \( Q \) is nonsingular, it follows from the continuity of \( \theta(x) \) and \( F(u) \) that
\[
u^\infty \in \Omega, \quad \theta(x) - \theta(x^\infty) + (u - u^\infty)^\top F(u^\infty) \geq 0, \quad \forall u \in \Omega.
\]
This indicates that \( u^\infty \) is a solution of MVI(\( \Omega, F \)) and then (31) is valid with \( u^* \) replaced by \( u^\infty \). Then, we have
\[
\|u^{k+1} - u^\infty\|_H \leq \|u^k - u^\infty\|_H,
\]
and thus \( \{ u^k \} \) converges to \( u^\infty \).

Furthermore, according to (31b), we can easily obtain that
\[
\frac{(7 - 2\tau)\gamma(2 - \gamma)\lambda_{\text{min}}((M^{-1})^\top R M^{-1})}{8(2\tau + 1)r^2_0} \sum_{k=0}^{\infty} \| (u^k - u^{k+1})^2 \| \leq \|u^0 - u^*\|_H^2.
\]
and therefore
\[
\sum_{k=0}^{\infty} \|u^k - u^{k+1}\|^2 < \infty.
\]
By invoking Lemma 1.2 in [4], we know
\[
\min_{1 \leq j \leq k} \| u^j - u^{j+1} \|^2 = o(1/k).
\]
The proof is complete.

**Remark 1.** It is noteworthy that \( \|u^k - u^{k+1}\|^2 \) serves a measure of optimality. On one hand, we can easily show that \( u^k \) is optimal if and only if \( u^k = u^{k+1} \). On the other hand, the square residual of KKT system can be bounded by \( \|u^k - u^{k+1}\|^2 \).

**Remark 2.** A major contribution of this paper is that we relax the restriction of the parameters \( r \) and \( s \). Specifically, by using (19), we get
\[
\begin{align*}
\frac{\tau^2 - \eta + \eta^2}{(2\tau - 1)^2} & \in \left[ \frac{2}{7}, 1 \right], \quad \text{if } \tau = 1, \eta \in [0, 1] \\
\frac{\tau^2 - \eta + \eta^2}{(2\tau - 1)^2} & \in \left[ \frac{5}{12}, \frac{4}{5} \right], \quad \text{if } \tau = 2, \eta \in [0, 1] \\
\frac{\tau^2 - \eta + \eta^2}{(2\tau - 1)^2} & \in \left[ \frac{7}{20}, \frac{9}{25} \right], \quad \text{if } \tau = 3, \eta \in [0, 1].
\end{align*}
\]
Obviously, \( \frac{\tau^2 - \eta + \eta^2}{(2\tau - 1)^2} \) is less than \( \frac{1}{2} \) when \( \tau \in [\frac{3}{12}, \frac{7}{12}] \), and hence we loosen the restriction on \( rs \), compared with L-PPA proposed in [22] and the customized PPA in [13].

4. **Numerical results.** In this section, we apply the proposed method to the nearest correlation matrix problem and matrix completion problem, and report the results of experiments to demonstrate the efficiency of our method. In these experiments, we evaluate these methods in terms of their numbers of iterations and computation time. All the codes were written by Matlab R2018a version and all the numerical experiments were performed on a Lenovo desktop computer with Intel (R) Core(TM) i7-8550U CPU with 1.99 GHz and 16.00 GB RAM.
4.1. Nearest correlation matrix problem. Recall that the optimization model of the nearest correlation matrix problem is of the following form:
\[
\min \left\{ \frac{1}{2} \| X - C \|_F^2 \mid \text{diag}(X) = e, X \in S_n^+ \right\}.
\] (39)

Here, we apply our proposed method to solve (2), and its subproblem is specified into:
\[
\tilde{X}^k = \arg \min \{ \| X - \frac{1}{1+r} (rX^k + \text{diag}(\lambda^k) + C) \|_F^2 \mid X \in S_n^+ \},
\] (40a)
and
\[
\tilde{\lambda}^k = \lambda^k - \frac{1}{s} (\text{diag}(\tilde{X}^k) - e).
\] (40b)

It is easy to see that (40a) admits a closed-form solution
\[
\tilde{X}^k = P_{S_n^+} \left[ \frac{1}{1+r} (rX^k + \text{diag}(\lambda^k) + C) \right],
\] (41)
where \( P_{S_n^+} \) denotes the projection operator onto \( S_n^+ \) which can be completed by an eigenvalue decomposition. In our experiments, we apply singular value decomposition (SVD) to conduct the eigenvalue decomposition. We constructed test data sets like those of [14]. In order to illustrate the efficiency of our proposed method, we compare its performance with PPA [13] and L-PPA [22] using the experiments with \( n \) ranging from 500 to 4000. To perform those algorithms, we need to set the parameters and the stopping criterion.

• The Parameters. Note that \( A = I \), we used initial stepsize \( s = 0.5, r = 1.01/s \) for PPA, \( s = 0.5 \times 0.8, r = 0.65/s \) for L-PPA. And the parameters of our algorithm are set as follows: \( \tau \in \{2, 3\}, s = 0.5, \eta = 0.9, \gamma = 1.5, \) and \( r = \left( \frac{\tau^2 - \eta^2}{(2\tau - 1)^2} + 0.01 \right)/s. \)

• Stopping criterion. The stopping criterion is
\[
\max \{ \max_{ij} |X^k_{ij} - \tilde{X}^k_{ij}|, \max_j |\lambda^k_j - \tilde{\lambda}^k_j| \} \leq 10^{-5}.
\]

The comparisons between these algorithms for solving (2) are presented in Table 1.

| Table 1. Performance of PPA, L-PPA and Our Method |
|-----------------------------------------------|
| Size | PPA | L-PPA | Our Method \( \tau = 2 \) | Our Method \( \tau = 3 \) |
|------|-----|------|-----------------|-----------------|
| \( n \) | It. | CPU | It. | CPU | It. | CPU | It. | CPU |
| 500 | 27 | 2.85 | 22 | 2.36 | 19 | 2.01 | 20 | 2.06 |
| 1000 | 31 | 17.01 | 25 | 14.42 | 18 | 10.23 | 18 | 10.28 |
| 1500 | 36 | 60.95 | 29 | 52.34 | 19 | 37.72 | 17 | 32.19 |
| 2000 | 41 | 178.37 | 33 | 152.42 | 20 | 88.30 | 20 | 89.05 |
| 3000 | 49 | 698.51 | 37 | 541.60 | 25 | 369.14 | 25 | 357.59 |
| 4000 | 59 | 1917.01 | 44 | 2552.53 | 31 | 934.66 | 31 | 934.66 |

As can be seen clearly from the table 1, our proposed method outperforms PPA and L-PPA both in terms of iterations and CPU time. Compared our method with PPA and L-PPA, our method is much faster when \( \tau = 2 \) or \( \tau = 3 \). Therefore, wider taking value area of \( rs \) of our method is meaningful.

All methods take more iterations with the matrix dimension \( n \) increasing. However, the growth rate of iteration number of our method is slower than PPA and L-PPA. No matter what \( \tau \) is, the difference of iteration number between \( n = 500 \) and \( n = 4000 \) is small. On the whole, our method is a simple and efficient method which is able to linearly constrained convex optimization problem.
4.2. Matrix completion problem. In this subsection we report experiments which demonstrate the effective performance of our proposed method on matrix completion problem [2, 3]. Considering the recovery of a low-rank matrix $X$ from under-sampled data, the matrix completion problem can be cast as the following convex programming:

$$\min \{ \|X\|_* \mid X_{\Pi} = M_{\Pi} \},$$

(42)

where $M \in \mathbb{R}^{l \times n}$ is a given matrix, $\|X\|_*$ denotes the nuclear norm which is defined as the sum of all singular values of $X$, and $(\cdot)_{\Pi} : \mathbb{R}^{l \times n} \to \mathbb{R}^{l \times n}$ is the sampling operator defined by $X_{ij} = M_{ij}$ if $(i, j) \in \Pi$ and 0 otherwise. Applying our method to this problem, the predictor takes the form

$$\tilde{X}^k = \text{Argmin} \left\{ \frac{1}{r} \|X\|_* + \frac{1}{2} \|X - (X^k + \frac{1}{r} \lambda^k_{\Pi})\|_F^2 \right\};$$

(43)

and

$$\tilde{\lambda}^k_{\Pi} = \lambda^k_{\Pi} - \frac{1}{s} (\tilde{X}^k_{\Pi} - M_{\Pi})$$

In fact, the closed-form solution of (43) is given by:

$$\tilde{\sigma}_i = \max(\sigma_i - \frac{1}{r}, 0),$$

and

$$\tilde{X}^k = U^k \tilde{\Lambda}^k (V^k)^T,$$

where $\tilde{\Lambda} = \text{diag}(\tilde{\sigma}_1, \tilde{\sigma}_2, \ldots, \tilde{\sigma}_n)$.

Its computational cost is dominated by singular value decomposition, hence, we combine a Lanczos method PROPACK [16] to accomplish the singular value decomposition. We conducted the test examples of (42) in the way suggested by [3] and compared our method with PPA [13] and L-PPA [22].

- **The Parameters.** Since $\|A^T A\| \leq 1$, we used initial stepsize $s = 160$, $r = 1.01/s$ for PPA, $s = 160 \times 0.8$, $r = 0.65/s$ for L-PPA. And the parameters of our algorithm are set as follows: $\tau = 3$, $s = 140$, $\eta = 0.9$, $\gamma = 1.5$, and $r = (\frac{\tau^2 - n + n^2}{(2\tau - 1)} + 0.01)/s$. All initial points were set to 0.

- **Stopping criterion.** As [3], the stopping criterion is

$$\frac{\|X^k_{\Pi} - M_{\Pi}\|_F}{\|M_{\Pi}\|_F} \leq 10^{-4}.$$

In Table 2, we report the numerical performance of these methods in various scenarios. In the following, $sr$ denotes the sampling ratio as [18] and $fr = \text{rank} * (l + n - \text{rank})/(sr * l * n)$ is the freedom of set, all test matrices were square.

| Problems | PPA | L-PPA | Our Method |
|----------|-----|-------|------------|
| n rank sr fr | It. CPU | It. CPU | It. CPU |
| 500 10 0.16 0.25 | 57 1.82 | 41 1.26 | 40 0.94 |
| 1000 10 0.12 0.17 | 70 11.7 | 52 8.59 | 41 5.17 |
| 1000 20 0.16 0.25 | 63 19.17 | 46 10.25 | 36 7.51 |
| 2000 10 0.12 0.08 | 91 95.32 | 67 50.54 | 50 46.84 |
| 2000 20 0.16 0.13 | 82 117.73 | 61 99.71 | 46 62.14 |
From Table 2, we observe that our method outperformed PPA and L-PPA in both iteration numbers and CPU times. Our numerical results show that our method was competitive and it provided more robust performance than PPA. The above experimental results confirm the efficiency of our method, in particular showing its nice convergent behavior.

5. Conclusion. This paper proposes a Lagrangian-PPA based prediction correction method for solving linearly constrained convex optimization problem. We establish the convergence and the convergence rate of this method. Compared with the existing PPA based methods such as the customized PPA [13] and L-PPA [22], we enlarge the range of the parameters $r$ and $s$, which often can accelerate the speed of the algorithms. To illustrate the efficiency of our method, we conduct numerical experiments on nearest correlation matrix problem and matrix completion problem, and compare their performance with L-PPA and the customized PPA.

REFERENCES

[1] K. J. Arrow, L. Hurwicz and H. Uzawa, Studies in linear and non-linear programming, In Stanford Mathematical Studies in the Social Sciences, II, Stanford University Press, Stanford, 1958.

[2] J. F. Cai, E. J. Candès and Z. W. Shen, A singular value thresholding algorithm for matrix completion, *SIAM J. Optim.*, 20 (2010), 1956–1982.

[3] E. J. Candès and B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.*, 9 (2008), 717–772.

[4] W. Deng, M. J. Lai and W. Yin, On the $o(1/k)$ convergence and parallelization of the alternating direction method of multipliers, preprint, *arXiv:1312.3040*.

[5] E. Esser, X. Q. Zhang and T. F. Chan, A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science, *SIAM J. Imaging Sci.*, 3 (2010), 1015–1046.

[6] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I*, Springer Series in Operations Research, Springer Verlag, New York, 2003.

[7] B. S. He, A class of projection and contraction methods for monotone variational inequalities, *Appl. Math. Optim.*, 35 (1997), 69–76.

[8] B. S. He, Inexact implicit methods for monotone general variational inequalities, *Math. Program. Ser. A*, 86 (1999), 199–217.

[9] B. S. He, PPA-like contraction methods for convex optimization: A framework using variational inequality approach, *J. Oper. Res. Soc. China*, 3 (2015), 391–420.

[10] B. S. He, X. L. Fu and Z. K. Jiang, Proximal point algorithm using a linear proximal term, *J. Optim. Theory Appl.*, 141 (2009), 299–239.

[11] B. S. He and Y. Shen, On the convergence rate of customized proximal point algorithm for convex optimization and saddle-point problem (in Chinese), *Scientia Sinica Mathematica*, 42 (2012), 515–525.

[12] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, *Pacific J. Optim.*, 4 (2008), 195–212.

[13] B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imaging Sci.*, 5 (2012), 119–149.

[14] B. S. He and X. M. Yuan, A contraction method with implementable proximal regularization for linearly constrained convex programming, *Optimization Online*, 2010.

[15] B. Jiang, Z. Peng and K. Deng, Two new customized proximal point algorithms without relaxation for linearly constrained convex optimization, *Bull. Iranian Math. Soc.*, 46 (2020), 865–892.

[16] R. M. Larsen, Propack software for large and sparse svd calculation, 2005.

[17] F. Ma and M. Ni, A class of customized proximal point algorithms for linearly constrained convex optimization, *Comput. Appl. Math.*, 37 (2018), 896–911.

[18] S. Q. Ma, D. Goldfarb and L. F. Chen, Fixed point and Bregman iterative methods for matrix rank minimization, *Math. Program.*, 128 (2011), 321–353.
[19] B. Martinet, Regularisation d’inéquations variationnelles par approximations succesives, *Revue Française Informat. Recherche Opérationnelle*, 4 (1970), 154–158.

[20] B. Martinet, Détermination approchée d’un point fixe d’une application pseudo-contractante. Cas de l’application prox, *C. R. Acad. Sci. Paris*, 274 (1972), 163–165.

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

[22] Y. F. You, X. L. Fu and B. S. He, Lagrangian-PPA based contraction methods for linearly constrained convex optimization, *Pac. J. Optim.*, 10 (2014), 199–213.

[23] M. Zhu and T. Chan, An efficient primal-dual hybrid gradient algorithm for total variation image restoration, *UCLA CAM Report*, 34 (2008), 8–34.

Received November 2020; 1st revision May 2021; 2nd revision July 2021; early access October 2021.

E-mail address: yyf@cczu.edu.cn

E-mail address: jiangsuhong1985@163.com