FASTER LAGRANGIAN-BASED METHODS: A UNIFIED PREDICTION-CORRECTION FRAMEWORK *

TAO ZHANG†, YONG XIA†, AND SHIRU LI† §

Abstract. Motivated by the prediction-correction framework constructed by He and Yuan [SIAM J. Numer. Anal. 50: 700-709, 2012], we propose a unified prediction-correction framework to accelerate Lagrangian-based methods. More precisely, for strongly convex optimization, general linearized Lagrangian method with indefinite proximal term, alternating direction method of multipliers (ADMM) with the step size of Lagrangian multiplier not larger than $(1+\sqrt{5})/2$ (or 2 when the objective of the composite convex optimization is the sum of a strongly convex and a linear function), linearized ADMM with indefinite proximal term, symmetric ADMM, and multi-block ADMM type method (assuming the gradient of one block is Lipschitz continuous) can achieve $O(1/k^2)$ convergence rate in the ergodic sense. The non-ergodic convergence rate is also established.

Key words. Convex programming, Augmented Lagrangian method, Alternating direction method of multipliers, Convergence rate.

AMS subject classifications. 47H09, 47H10, 90C25, 90C30

1. Introduction.

Augmented Lagrangian method (ALM) [20, 26] is a classic and efficient tool for solving the convex programming problem with linear equality constraints:

$$\text{(P1)} \quad \min\{f(x): \ Ax = b\},$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is closed, proper, convex, but not necessarily smooth, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Even recently, (P1) has a wide range of applications in compressed sensing, image processing, machine learning and so on. There are some well-known variants of ALM, for example, the proximal ALM introduced by Rockafeller [27, 28] and the linearized ALM with different special proximity terms.

As another popular Lagrangian-based method, the alternating direction method of multipliers (ADMM) [6, 7] plays a great role in efficiently solving large-scale structural case of (P1):

$$\text{(P2)} \quad \min\{f(x) = f_1(x_1) + f_2(x_2): \ Ax := A_1x_1 + A_2x_2 = b\},$$

where $A_1 \in \mathbb{R}^{m\times n_1}$, $A_2 \in \mathbb{R}^{m\times n_2}$, $b \in \mathbb{R}^m$, $f_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ and $f_2 : \mathbb{R}^{n_2} \to \mathbb{R}$ are proper, closed and convex. Following the idea of proximal ALM, proximal ADMM [5] is proposed to make the subproblems easier to solve by adding proximity terms. With a carefully selected special proximity term, it leads to the linearized ADMM [34]. For the ADMM type methods, there are numerous references on the $O(1/k)$ convergence rate in the ergodic sense [2, 3, 16, 23] and in the non-ergodic sense [2, 18, 21].

The first accelerated gradient method for unconstrained convex optimization with $O(1/k^2)$ convergence rate is due to Nesterov [24]. It is further generalized to composite convex optimization with simple proximal operator [1, 32]. For (P2) with an additional assumption that both f_1 and f_2 are strongly convex, ADMM is accelerated by introducing the extrapolation technique [8]. Under one function being strongly

*This research was supported by National Natural Science Foundation of China under grants 11822103, 11771056 and Beijing Natural Science Foundation Z180005.

†School of Mathematical Sciences, Beihang University, Beijing, 100191, P. R. China ((T. Zhang) shuxuekuangwu@buaa.edu.cn, (Y. Xia) yxia@buaa.edu.cn ,(S.R. Li) lishiru@buaa.edu.cn).

§Corresponding author.

convex, ADMM with adaptive parameters also enjoys an $O(1/k^2)$ convergence rate [33]. As shown in [31], the strongly convex assumption can be replaced with gradient Lipschitz continuity. For $O(1/k^2)$ convergence rate of linearized ADMM type methods under strongly convex assumption, we refer to [3, 25]. Recently, Sabach and Teboulle [29] proposed a class of Lagrangian-based methods with faster convergence by the so-called *nice primal algorithmic map*.

Convergence analysis of Lagrangian-based methods used to be a heavy task. Until 2012, He and Yuan [17] established a unified prediction-correction framework to simplify the analysis of the Lagrangian-based methods, from which $O(1/k)$ convergence rate can be obtain in the ergodic and non-ergodic sense. Motivated by this framework, we propose in this paper a prediction-correction framework to analyze the accelerate versions of the Lagrangian-based methods with $O(1/k^2)$ convergence rate. For (P1) and (P2) with f, f_1 or f_2 being strongly convex, we succeed in establishing $O(1/k^2)$ convergence rate in ergodic sense of Lagrangian-based methods including general linearized ALM with indefinite proximal term, ADMM with step size of Lagrangian multiplier not larger than $(1+\sqrt{5})/2$ (or 2 when f_1 is linear), linearized ADMM with indefinite proximal term, symmetric ADMM. It is a bit surprise that the $O(1/k^2)$ convergence rate in ergodic sense of ADMM type methods can be established for solving multi-block optimization problem

$$\text{(P3)}\quad \min\{f(x)=\sum_{i=1}^m f_i(x_i):\ Ax:=\sum_{i=1}^m A_i x_i = b\},$$

where $m\geq 2$, $f_i:\mathbb{R}^{n_i}\to\mathbb{R}$ is closed proper convex for $i\in[1,2,\cdots,m]$, f_m is addition gradient Lipschitz continuous, $A_i\in\mathbb{R}^{m\times n_i}$ and $b\in\mathbb{R}^m$. Non-ergodic $O(1/k^2)$ convergence rate is also considered in this paper.

The remainder of this paper is organized as follows. Section 2 presents our unified prediction-correction framework for accelerated. Then we establish faster ergodic convergence rate under strongly convex and gradient Lipschitz continuous assumptions in Sections 3 and 4, respectively. Section 5 gives faster non-ergodic convergence rate. Conclusions are made in Section 6.

Notation. Let x_i be the i-th part of x, i.e., $x=(x_1,\cdots,x_m)\in\mathbb{R}^n$. Denote by $D\succ(\succeq)0$ the symmetric and positive (semi)definite matrix in $\mathbb{R}^{n\times n}$. For $D\succ 0$, we define the matrix norm as $\|x\|_D=\sqrt{x^TDx}$. Denote by $\sigma_{\max}(D)$ and $\sigma_{\min}(D)$ the maximal and minimal eigenvalue of D, respectively. Let $\partial f(x)$ be the subdifferential of the convex function $f(x)$. Denote by $\nabla f(x)$ the gradient of the smooth function $f(x)$. The following two definitions are standard.

Definition 1.1. *$f:\mathbb{R}^n\to\mathbb{R}$ is σ-strongly convex if there is a $\sigma>0$ such that*

$$f(y)\geq f(x)+\langle f'(x),y-x\rangle+\frac{\sigma}{2}\|y-x\|^2,\ f'(x)\in\partial f(x),\ \forall x,y\in\mathbb{R}^n.$$

Definition 1.2. *$f:\mathbb{R}^n\to\mathbb{R}$ is L-gradient Lipschitz continuous if f is differentiable and there is a constant $L>0$ such that*

$$f(y)\leq f(x)+\langle\nabla f(x),y-x\rangle+\frac{L}{2}\|y-x\|^2,\ \forall x,y\in\mathbb{R}^n.$$

If f is L-gradient Lipschitz continuous, we have

$$f(y)\geq f(x)+\langle\nabla f(x),y-x\rangle+\frac{1}{2L}\|\nabla f(x)-\nabla f(y)\|^2,\ \forall x,y\in\mathbb{R}^n. \tag{1.1}$$

2. The unified prediction-correction framework for acceleration. We consider (P1). The Lagrangian function of (P1) is:

$$L(x,\lambda) = f(x) - \lambda^T(Ax - b),$$

where λ is the Lagrange multiplier. We call (x^*, λ^*) a saddle point of $L(x, \lambda)$ if

$$L(x^*,\lambda) \le L(x^*,\lambda^*) \le L(x,\lambda^*).$$

As shown in [9, 10, 11, 17], the saddle point (x^*, λ^*) can be alternatively characterized as a solution point of the following variational inequality (VI):

$$f(x) - f(x^*) + (u - u^*)^T F(u^*) \ge 0, \quad \forall x, \tag{2.1}$$

where

$$u = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(u) = \begin{pmatrix} -A^T\lambda \\ Ax - b \end{pmatrix}. \tag{2.2}$$

The following framework due to He and Yuan [17] has been widely used in [10, 11, 19] for solving (2.1).

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that

$$f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) \ge (v - \widetilde{v}^k)^T Q(v^k - \widetilde{v}^k), \ \forall u, \tag{2.3}$$

where $Q^T + Q \succ 0$ (while Q is not necessarily symmetric).
[Correction Step.] Update v^{k+1} by

$$v^{k+1} = v^k - M(v^k - \widetilde{v}^k). \tag{2.4}$$

The selections of Q and M are crucial. The following conditions are sufficient to guarantee the convergence.

[Convergence Conditions.] For the matrices Q and M used in (2.3) and (2.4), respectively, there exists a matrix $H \succ 0$ such that

$$HM = Q, \tag{2.5}$$

$$G := Q^T + Q - M^T HM \succeq 0. \tag{2.6}$$

LEMMA 2.1 ([10, 11, 17, 19]). *Under convergence condition* (2.5) *and* (2.6)*, for the prediction-correction framework* (2.3) *and* (2.4)*, we have*

$$f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) \ge \frac{1}{2}(\|v^{k+1} - v\|_H^2 - \|v^k - v\|_H^2) + \frac{1}{2}\|v^k - \widetilde{v}^k\|_G^2, \ \forall u.$$

Then, the ergodic $O(1/K)$ convergence rate can be easily obtained by adding up the inequalities presented in the following lemma from $k = 0$ to $k = K$. In order to pursue faster convergence rate, we propose the following framework with dynamically updated Q^k and M^k. Moreover, we need an extra item in (2.3), see (2.7), where the new variable z can be set as x, x_i, or $\nabla f(x)$, see details in Sections 3 and 4.

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that

$$f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) \geq (v - \widetilde{v}^k)^T Q^k (v^k - \widetilde{v}^k) + \frac{\sigma}{2}\|z^k - z\|^2, \ \forall u, \tag{2.7}$$

where $\sigma > 0$, and $(Q^k)^T + Q^k \succ 0$ (while Q^k is not necessarily symmetric).
[Correction Step.] Update v^{k+1} by

$$v^{k+1} = v^k - M^k(v^k - \widetilde{v}^k). \tag{2.8}$$

Similarly, we give the following dynamic convergence conditions.

[Convergence Conditions.] For the matrices Q^k and M^k used in (2.7) and (2.8), respectively, there exists a matrix $H^k \succ 0$ such that

$$H^k M^k = Q^k, \tag{2.9}$$

$$G^k := (Q^k)^T + Q^k - (M^k)^T H^k M^k \succeq 0. \tag{2.10}$$

With the dynamic convergence conditions (2.9) and (2.10), we can obtain Lemma 2.2 similar to Lemma 2.1.

Lemma 2.2. *Under convergence condition* (2.9) *and* (2.10)*, for the prediction-correction framework* (2.7) *and* (2.8)*, we have*

$$\begin{aligned} & f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) \\ \geq & \frac{1}{2}(\|v^{k+1} - v\|_{H^k}^2 + \sigma\|z^k - z\|^2 - \|v^k - v\|_{H^k}^2) + \frac{1}{2}\|v^k - \widetilde{v}^k\|_{G^k}^2, \ \forall u. \end{aligned} \tag{2.11}$$

Note that if H^k is not a constant matrix, it may fail to establish the convergence by directly adding up (2.11) from $k = 0$ to $k = K$. Therefore, we need the following additional convergence condition.

[Convergence Condition.]

$$\begin{aligned} & r^k(\|v^{k+1} - v^*\|_{H^k}^2 + \sigma\|z^k - z^*\|^2 - \|v^k - v^*\|_{H^k}^2 + \|v^k - \widetilde{v}^k\|_{G^k}^2) \\ \geq & \|v^{k+1} - v^*\|_{H_0^{k+1}}^2 - \|v^k - v^*\|_{H_0^k}^2 + \Theta^{k+1} - \Theta^k, \end{aligned} \tag{2.12}$$

where $r^k > 0$, $H_0^k \succ 0$ and $\Theta^k \geq 0$.

Notice that throughout this paper, we have

$$f(x^*) - f(\widetilde{x}^k) + (u^* - \widetilde{u}^k)^T F(\widetilde{u}^k) \equiv f(x^*) - f(\widetilde{x}^k) + (\lambda^*)^T (A\widetilde{x}^k - b).$$

We give the following two technique lemmas for convergence rate analysis.

Lemma 2.3 ([33]). *Let (x^*, λ^*) be the saddle point, given a function ϕ and a fixed point $\widetilde{x}$, it holds that*

$$\begin{aligned} & f(\widetilde{x}) - f(x^*) - \lambda^T (A\widetilde{x} - b) \leq \phi(\lambda), \ \forall \lambda \\ \Longrightarrow & f(\widetilde{x}) - f(x^*) + \rho\|A\widetilde{x} - b\| \leq \sup_{\|\lambda\| \leq \rho} \phi(\lambda), \ \forall \rho > 0. \end{aligned}$$

LEMMA 2.4 ([33]). *Let (x^*,λ^*) be the saddle point, and $\|\lambda^*\|\le\rho$. For any $\epsilon\ge 0$,*

$$f(\widetilde{x})-f(x^*)+\rho\|A\widetilde{x}-b\|\le\epsilon,$$
$$\Longrightarrow \|A\widetilde{x}-b\|\le\frac{\epsilon}{\rho-\|\lambda^*\|}\ \text{and}\ -\frac{\|\lambda^*\|\epsilon}{\rho-\|\lambda^*\|}\le f(\widetilde{x})-f(x^*)\le\epsilon.$$

THEOREM 2.5. *Under convergence condition* (2.9) *and* (2.10)*, for prediction-correction framework* (2.7) *and* (2.8) *satisfying* (2.12)*, we have*

$$f(\widetilde{X}^K)-f(x^*)\le O\left(1\Big/\sum_{k=0}^{K}r^k\right),\quad \|A\widetilde{X}^K-b\|\le O\left(1\Big/\sum_{k=0}^{K}r^k\right),$$

where $\widetilde{X}^K=(\sum_{k=0}^K r^k\widetilde{x}^k)/(\sum_{k=0}^K r^k)$. In particular, $O(1/k^2)$ convergence rate is established with the setting $r^k=O(k)$.

Proof. Multiplying both sides of (2.11) by r^k and then using convergence condition (2.12) yields that

$$r^k\{f(x^*)-f(\widetilde{x}^k)+(u^*-\widetilde{u}^k)^TF(\widetilde{u}^k)\}\ge\frac{1}{2}\Big(\|v^{k+1}-v^*\|^2_{H_0^{k+1}}-\|v^k-v^*\|^2_{H_0^k}+\Theta^{k+1}-\Theta^k\Big).$$

Because $(u^*-\widetilde{u}^k)^TF(\widetilde{u}^k)\equiv(\lambda^*)^T(A\widetilde{x}^k-b)$, by adding both side of the above inequalities from $k=0$ to $k=K$, we obtain

$$\sum_{k=0}^{K}r^k\{f(\widetilde{x}^k)-f(x^*)-(\lambda^*)^T(A\widetilde{x}^k-b)\}\le C,$$

where C is a constant. Dividing both sides of the above inequality by $\sum_{k=0}^K r^k$ and then using the convexity of f and Lemmas 2.3 and 2.4 completes the proof. ∎

3. Faster ergodic convergence rate for strongly convex optimization.

Based on Theorem 2.5, we establish $O(1/k^2)$ ergodic convergence rate of several Lagrangian-based methods for solving strongly convex case of (P1) and (P2).

3.1. General proximal ALM.

The general proximal ALM (GPALM) for solving (P1) reads as:

$$\text{(GPALM)}\quad\begin{cases}x^{k+1}=\arg\min_x\{L_{\beta^k}(x,\lambda^k)+\frac{1}{2}\|x-x^k\|^2_{D^k}\},\\ \lambda^{k+1}=\lambda^k-\gamma\beta^k(Ax^{k+1}-b),\quad\gamma\in(0,2],\end{cases}\tag{3.1}$$

where $\beta^k>0$, D^k is symmetric, and $L_{\beta^k}(x,\lambda)$ is the augmented Lagrangian function of (P1) defined as

$$L_{\beta^k}(x,\lambda)=f(x)-\lambda^T(Ax-b)+\frac{\beta^k}{2}\|Ax-b\|^2.$$

Let u and $F(u)$ be defined in (2.2), $v=u$. Define the artificial vector $\widetilde{v}^k=(\widetilde{x}^k,\widetilde{\lambda}^k)$ as

$$\widetilde{x}^k=x^{k+1}\ \text{and}\ \widetilde{\lambda}^k=\lambda^k-\beta^k(A\widetilde{x}^k-b).\tag{3.2}$$

Similar as the analysis in [12], we can write GPALM (3.1) as:

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that

$$f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) \geq (v - \widetilde{v}^k)^T Q^k (v^k - \widetilde{v}^k), \ \forall u,$$
$$\text{where} \ \ Q^k = \begin{pmatrix} D^k & 0 \\ 0 & \frac{1}{\beta^k} I \end{pmatrix}. \tag{3.3}$$

According to the dual step in GPALM (3.1), we obtain

$$\lambda^{k+1} = \lambda^k - \gamma\beta^k (Ax^{k+1} - b) = \lambda^k - \gamma(\lambda^k - \widetilde{\lambda}^k). \tag{3.4}$$

Therefore, u^{k+1} generated by GPALM can be viewed as the output after correcting $\widetilde{u}^k$ by the following scheme:

[Correction Step.] Update v^{k+1} by

$$v^{k+1} = v^k - M^k (v^k - \widetilde{v}^k),$$
$$\text{where} \ \ M^k = \begin{pmatrix} I & 0 \\ 0 & \gamma I \end{pmatrix}. \tag{3.5}$$

We define

$$\begin{aligned} H^k &= \begin{pmatrix} D^k & 0 \\ 0 & \frac{1}{\gamma\beta^k} I \end{pmatrix}, \\ G^k &= (Q^k)^T + Q^k - (M^k)^T H^k M^k = \begin{pmatrix} D^k & 0 \\ 0 & \frac{2-\gamma}{\beta^k} I \end{pmatrix}. \end{aligned} \tag{3.6}$$

If D^k is symmetric positive semidefinite, then H^k and G^k satisfy (2.9) and (2.10). By Theorem 2.5, it is sufficient to prove the following result.

THEOREM 3.1. *Let $D^k = D_0/\beta^k$ with $D_0 \succeq 0$. Then $\{v^{k+1}\}$ generated by GPALM* (3.1) *satisfies convergence condition* (2.12) *with $r^k = \beta^k$, $\sigma = 0$, $\Theta^k = 0$ and*

$$H_0^k \equiv \begin{pmatrix} D_0 & 0 \\ 0 & \frac{1}{\gamma} I \end{pmatrix}.$$

Proof. According to Lemma 2.2, (2.11) holds. Multiplying both sides of (2.11) (with $\sigma = 0$) by β^k and then ignoring nonnegative term yields that

$$\begin{aligned} &\beta^k \Big(f(x^*) - f(\widetilde{x}^k) + (u^* - \widetilde{u}^k)^T F(\widetilde{u}^k) \Big) \\ \geq &\frac{\beta^k}{2} \Big(\|v^{k+1} - v^*\|_{H^k}^2 - \|v^k - v^*\|_{H^k}^2 + \|v^k - \widetilde{v}^k\|_{G^k}^2 \Big) \\ \geq &\frac{1}{2} \Big(\|v^{k+1} - v^*\|_{H_0^{k+1}}^2 - \|v^k - v^*\|_{H_0^k}^2 \Big). \end{aligned}$$

The proof is complete. □

Next we consider the possible indefinite case $D^k \not\succeq 0$. For $\tau \in [\frac{2+\gamma}{4}, 1]$, we define

$$D^k = \tau r \beta^k I - \beta^k A^T A := D_1^k - (1-\tau)\beta^k A^T A, \tag{3.7}$$
$$D_1^k := \tau r \beta^k I - \tau\beta^k A^T A. \tag{3.8}$$

Using the optimality condition of x-subproblem in GPALM (3.1) and noting that f is σ-strongly convex, we obtain, $\forall x$,

$$f(x)-f(x^{k+1})+(x-x^{k+1})^T\{-A^T\lambda+\beta^k A^T(Ax^{k+1}-b)+D^k(x^{k+1}-x^k)\} \geq \frac{\sigma}{2}\|x^{k+1}-x\|^2.$$

As a result, our prediction-correction scheme on GPALM can be simplified as:

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that

$$f(x)-f(\widetilde{x}^k)+(u-\widetilde{u}^k)^T F(\widetilde{u}^k) \geq (u-\widetilde{u}^k)^T Q^k(u^k-\widetilde{u}^k)+\frac{\sigma}{2}\|\widetilde{x}^k-x\|^2, \ \forall u, \quad (3.9)$$

where Q^k is defined in (3.3).
[Correction Step.] Update v^{k+1} by

$$v^{k+1} = v^k - M^k(v^k-\widetilde{v}^k), \quad (3.10)$$

where M^k is defined in (3.5).

Setting $\widetilde{z}^k = \widetilde{x}^k = x^{k+1}$ and $z = x^*$ in Lemma 2.2 yields that

$$\begin{aligned} f(x^*)-f(\widetilde{x}^k)+(u^*-\widetilde{u}^k)^T F(\widetilde{u}^k) \geq & \frac{1}{2}\Big(\|v^{k+1}-v^*\|^2_{H^k}+\sigma\|x^{k+1}-x^*\|^2 \\ & -\|v^k-v^*\|^2_{H^k}\Big)+\frac{1}{2}\|v^k-\widetilde{v}^k\|^2_{G^k}. \end{aligned}$$

According to (3.2), (3.4), (3.7) and (3.8), we have

$$\begin{aligned} & \|v^k-\widetilde{v}^k\|^2_{G^k} = \|x^k-x^{k+1}\|^2_{D^k}+\frac{2-\gamma}{\beta^k}\|\lambda^k-\widetilde{\lambda}^k\|^2 \\ = & \|x^k-x^{k+1}\|^2_{D_1^k}+(2-\gamma)\beta^k\|Ax^{k+1}-b\|^2-(1-\tau)\beta^k\|x^k-x^{k+1}\|^2_{A^TA} \\ = & \|x^k-x^{k+1}\|^2_{D_1^k}+(2-\gamma)\beta^k\|x^{k+1}-x^*\|^2_{A^TA}-(1-\tau)\beta^k\|x^k-x^{k+1}\|^2_{A^TA}. \end{aligned} \quad (3.11)$$

According to Theorem 2.5, it is sufficient to prove the following result.

THEOREM 3.2. *Let D^k be given in* (3.7), $r > \|A^TA\|^2$ *and*

$$\beta^k(\tau r\beta^k+\sigma) \geq \tau r(\beta^{k+1})^2, \ \beta^{k+1} \geq \beta^k, \ \forall k. \quad (3.12)$$

Then $\{v^{k+1}\}$ *generated by GPALM* (3.1) *satisfies convergence condition* (2.12) *with*

$$H_0^k = \begin{pmatrix} \beta^k D_1^k+(1-\tau)(\beta^k)^2A^TA & 0 \\ 0 & \frac{1}{\gamma}I \end{pmatrix},$$

$r^k = \beta^k$, $z = x$, $\widetilde{z}^k = x^{k+1}$ *and* $\Theta^k = 0$.

Proof. Let

$$e = (\|v^{k+1}-v^*\|^2_{H^k}+\sigma\|x^{k+1}-x^*\|^2-\|u^k-u^*\|^2_{H^k})+\|v^k-\widetilde{v}^k\|^2_{G^k}.$$

According to the structure of H^k and G^k and (3.11), we obtain

$$\begin{aligned} e = & \|x^{k+1}-x^*\|^2_{D_1^k}-(1-\tau)\beta^k\|x^{k+1}-x^*\|^2_{A^TA}+\frac{1}{\gamma\beta^k}\|\lambda^{k+1}-\lambda^*\|^2+\sigma\|x^{k+1}-x^*\|^2 \\ & -\Big(\|x^k-x^*\|^2_{D_1^k}-(1-\tau)\beta^k\|x^k-x^*\|^2_{A^TA}+\frac{1}{\gamma\beta^k}\|\lambda^k-\lambda^*\|^2\Big) \\ & +\Big(\|x^k-x^{k+1}\|^2_{D_1^k}+(2-\gamma)\beta^k\|x^{k+1}-x^*\|^2_{A^TA}\Big)-(1-\tau)\beta^k\|x^k-x^{k+1}\|^2_{A^TA}. \end{aligned} \quad (3.13)$$

Baised on the fact that

$$\|x^k - x^{k+1}\|^2_{A^TA} \le 2\|x^k - x^*\|^2_{A^TA} + 2\|x^{k+1} - x^*\|^2_{A^TA},$$

we obtain

$$\begin{aligned} e \ge\, & \|x^{k+1} - x^*\|^2_{D_1^k} + (1-\tau)\beta^k\|x^{k+1} - x^*\|^2_{A^TA} + \frac{1}{\gamma\beta^k}\|\lambda^{k+1} - \lambda^*\|^2 \\ & + \sigma\|x^{k+1} - x^*\|^2 - \left(\|x^k - x^*\|^2_{D_1^k} + (1-\tau)\beta^k\|x^k - x^*\|^2_{A^TA} + \frac{1}{\gamma\beta^k}\|\lambda^k - \lambda^*\|^2\right) \\ & + \|x^k - x^{k+1}\|^2_{D_1^k} + (4\tau - \gamma - 2)\beta^k\|x^{k+1} - x^*\|^2_{A^TA}. \end{aligned} \tag{3.14}$$

Note that $D_1^k \succ 0$ with $r > \|A^TA\|^2$. For any $\tau \in [\frac{2+\gamma}{4}, 1]$ with $\gamma \in (0, 2]$, under the assumption (3.12), we have

$$\begin{aligned} & \beta^k\{D_1^k + (1-\tau)\beta^kA^TA + \sigma I\} = \beta^k\{\tau r\beta^kI + (1-2\tau)\beta^kA^TA + \sigma I\} \\ \succeq & \beta^k\{\tau r\beta^kI + \sigma I\} + (1-2\tau)(\beta^{k+1})^2A^TA \succeq \tau r(\beta^{k+1})^2 + (1-2\tau)(\beta^{k+1})^2A^TA \\ = & \beta^{k+1}\{D_1^{k+1} + (1-\tau)\beta^{k+1}A^TA\}. \end{aligned} \tag{3.15}$$

Ignoring nonnegative terms and multiplying both sides of (3.14) by β^k and combining (3.15) yields that

$$\begin{aligned} \beta^ke \ge\, & \|x^{k+1} - x^*\|^2_{\beta^kD_1^k} + (1-\tau)(\beta^k)^2\|x^{k+1} - x^*\|^2_{A^TA} + \frac{1}{\gamma}\|\lambda^{k+1} - \lambda^*\|^2 \\ +\sigma\beta^k\|x^{k+1} - x^*\|^2 & - \left(\|x^k - x^*\|^2_{\beta^kD_1^k} + (1-\tau)(\beta^k)^2\|x^k - x^*\|^2_{A^TA} + \frac{1}{\gamma}\|\lambda^k - \lambda^*\|^2\right) \\ \ge\, & \|x^{k+1} - x^*\|^2_{\beta^{k+1}D_1^{k+1}} + (1-\tau)(\beta^{k+1})^2\|x^{k+1} - x^*\|^2_{A^TA} + \frac{1}{\gamma}\|\lambda^{k+1} - \lambda^*\|^2 \\ & - \left(\|x^k - x^*\|^2_{\beta^kD_1^k} + (1-\tau)(\beta^k)^2\|x^k - x^*\|^2_{A^TA} + \frac{1}{\gamma}\|\lambda^k - \lambda^*\|^2\right) \\ =\, & \|v^{k+1} - v^*\|^2_{H_0^{k+1}} - \|v^k - v^*\|^2_{H_0^k}. \end{aligned}$$

The proof is complete. □

REMARK 1. *There exists some $\delta > 0$ such that $\beta^k = \delta k$ satisfies* (3.12).

3.2. ADMM.

We consider solving (P2). Let

$$u = \begin{pmatrix} x_1 \\ x_2 \\ \lambda \end{pmatrix}, \quad v = \begin{pmatrix} x_2 \\ \lambda \end{pmatrix} \quad \text{and} \quad F(u) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ Ax - b \end{pmatrix}. \tag{3.16}$$

We define the artificial vector $\widetilde{v}^k = (\widetilde{x}_2^k, \widetilde{\lambda}^k)$ with

$$\widetilde{x}_1^k = x_1^{k+1},\ \widetilde{x}_2^k = x_2^{k+1} \text{ and } \widetilde{\lambda}^k = \lambda^k - \beta^k(A_1\widetilde{x}_1^k + A_2x_2^k - b). \tag{3.17}$$

3.2.1. ADMM.

The classic algorithm ADMM solves (P2) by the following iterative scheme:

$$\text{(ADMM)} \quad \begin{cases} x_1^{k+1} = \arg\min_{x_1}\{L_{\beta^k}(x_1, x_2^k, \lambda^k)\}, \\ x_2^{k+1} = \arg\min_{x_2}\{L_{\beta^k}(x_1^{k+1}, x_2, \lambda^k)\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta^k(A_1x_1^{k+1} + A_2x_2^{k+1} - b). \end{cases} \tag{3.18}$$

From the optimality condition of x_2-subproblem in ADMM (3.18) and σ-strongly convex of f_2, we obtain, $\forall x_2$,

$$f_2(x_2)-f_2(x_2^{k+1})+(x_2-x_2^{k+1})^T\{-A_2^T\lambda^k+\beta^k A_2^T(A_1x_1^{k+1}+A_2x_2^{k+1}-b)\}\geq\frac{\sigma}{2}\|x_2^{k+1}-x_2\|^2. \tag{3.19}$$

Hence, similar as the analysis in [11], we can write the framework about ADMM (3.18) as

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that

$$f(x)-f(\widetilde{x}^k)+(u-\widetilde{u}^k)^T F(\widetilde{u}^k)\geq(v-\widetilde{v}^k)^T Q^k(v^k-\widetilde{v}^k)+\frac{\sigma}{2}\|\widetilde{x}_2^k-x_2\|^2,\ \forall u,$$

where $Q^k=\begin{pmatrix}\beta^k A_2^TA_2 & 0\\ -A_2 & \frac{1}{\beta^k}I\end{pmatrix}$.

According to the dual step in ADMM and (3.17) we obtain

$$\lambda^{k+1}=\lambda^k-\gamma\beta^k(A_1x_1^{k+1}+A_2x_2^{k+1}-b)=\lambda^k-\gamma(\lambda^k-\widetilde{\lambda}^k)+\gamma\beta^kA_2(x_2^k-\widetilde{x}_2^k). \tag{3.20}$$

Thence, v^{k+1} generated by ADMM can be viewed as being corrected from $\widetilde{v}^k$:

[Correction Step.] Update v^{k+1} by

$$v^{k+1}=v^k-M^k(v^k-\widetilde{v}^k),$$
$$\text{where } M^k=\begin{pmatrix}I & 0\\ -\gamma\beta^kA_2 & \gamma I\end{pmatrix}. \tag{3.21}$$

We define

$$H^k=\begin{pmatrix}\beta^kA_2^TA_2 & 0\\ 0 & \frac{1}{\gamma\beta^k}I\end{pmatrix},$$
$$G^k=((Q^k)^T+Q^k)-(M^k)^TH^kM^k=\begin{pmatrix}(1-\gamma)\beta^kA_2^TA_2 & -(1-\gamma)A_2^T\\ -(1-\gamma)A_2 & \frac{2-\gamma}{\beta^k}I\end{pmatrix}.$$

Based on Lemma 2.2, if H^k and G^k satisfy (2.9) and (2.10), we have

$$\begin{aligned}&f(x^*)-f(\widetilde{x}^k)+(u^*-\widetilde{u}^k)^TF(\widetilde{u}^k)\\ &\geq\frac{1}{2}\Big(\|v^{k+1}-v^*\|_{H^k}^2+\sigma\|x_2^{k+1}-x_2^*\|^2-\|v^k-v^*\|_{H^k}^2\Big)+\frac{1}{2}\|v^k-\widetilde{v}^k\|_{G^k}^2.\end{aligned} \tag{3.22}$$

According to the structure of G^k, we have

$$\begin{aligned}\|v^k-\widetilde{v}^k\|_{G^k}^2=&(1-\gamma)\beta^k\|x_2^k-x_2^{k+1}\|_{A_2^TA_2}^2+\frac{2-\gamma}{\beta^k}\|\lambda^k-\widetilde{\lambda}^k\|^2\\ &-2(1-\gamma)(x_2^k-x_2^{k+1})^T(A_2)^T(\lambda^k-\widetilde{\lambda}^k).\end{aligned} \tag{3.23}$$

Based on (3.17) and (3.20), we obtain

$$\begin{aligned}\|v^k-\widetilde{v}^k\|^2_{G^k}=&\beta^k\|A_2(x_2^k-x_2^{k+1})\|^2+\frac{2-\gamma}{\gamma^2\beta^k}\|\lambda^k-\lambda^{k+1}\|^2\\&+\frac{2}{\gamma}(x_2^k-x_2^{k+1})^TA_2^T(\lambda^k-\lambda^{k+1})\\=&\beta^k\|A_2(x_2^k-x_2^{k+1})\|^2+(2-\gamma)\beta^k\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2\\&+2\beta^k(x_2^k-x_2^{k+1})^TA_2^T(A_1x_1^{k+1}+A_2x_2^{k+1}-b).\end{aligned}\tag{3.24}$$

According to Theorem 2.5, it is sufficient to prove the following result.

THEOREM 3.3. *Suppose that f_2 is σ-strongly convex. For $\gamma\in(0,(1+\sqrt{5})/2]$ and $\{\beta^k\}$ satisfying*

$$\beta^k\left(\beta^k+\frac{\sigma}{\sigma_{\max}(A_2^TA_2)}\right)\ge(\beta^{k+1})^2,\ \frac{(\beta^k)^3\sigma_{\max}(A_2^TA_2)}{\beta^k\sigma_{\max}(A_2^TA_2)+\sigma}\le(\beta^{k-1})^2,\ \beta^{k+1}\ge\beta^k,\tag{3.25}$$

$\{v^{k+1}\}$ generated by ADMM (3.18) satisfies convergence condition (2.12) with

$$H_0^k=\begin{pmatrix}(\beta^k)^2A_2^TA_2 & 0\\ 0 & \frac{1}{\gamma}I\end{pmatrix},$$

$r^k=\beta^k$, $z=x_2$, $\widetilde{z}^k=x_2^{k+1}$, and $\Theta^k=(1-\gamma)^2(\beta^{k-1})^2\|A_1x_1^k+A_2x_2^k-b\|^2$.

Proof. According to (3.19), we obtain, $\forall x_2$,

$$f_2(x_2)-f(x_2^k)+(x_2-x_2^k)^T\{-A_2^T\lambda^{k-1}+\beta^{k-1}A_2^T(A_1x_1^k+A_2x_2^k-b)\}\ge\frac{\sigma}{2}\|x_2^k-x_2\|^2.\tag{3.26}$$

Putting (3.19) with $x_2=x_2^k$ and (3.26) with $x_2=x_2^{k+1}$ together gives

$$\begin{aligned}(x_2^k-x_2^{k+1})^TA_2^T\{\lambda^{k-1}-\lambda^k-\beta^{k-1}(A_1x_1^k+A_2x_2^k-b)+\beta^k(A_1x_1^{k+1}+A_2x_2^{k+1}-b)\}\\\ge\sigma\|x_2^k-x_2^{k+1}\|^2.\end{aligned}$$

According to dual sequence updating, we obtain

$$\begin{aligned}(x_2^k-x_2^{k+1})^TA_2^T\{\beta^k(A_1x_1^{k+1}+A_2x_2^{k+1}-b)-(1-\gamma)\beta^{k-1}(A_1x_1^k+A_2x_2^k-b)\}\\\ge\sigma\|x_2^k-x_2^{k+1}\|^2.\end{aligned}$$

By rearranging the above inequality, we obtain

$$\begin{aligned}&\beta^k(x_2^k-x_2^{k+1})^TA_2^T(A_1x_1^{k+1}+A_2x_2^{k+1}-b)\\&\ge(1-\gamma)\beta^{k-1}(x_2^k-x_2^{k+1})^TA_2^T(A_1x_1^k+A_2x_2^k-b)+\sigma\|x_2^k-x_2^{k+1}\|^2\\&\ge(1-\gamma)\beta^{k-1}(x_2^k-x_2^{k+1})^TA_2^T(A_1x_1^k+A_2x_2^k-b)+\sigma'\|A_2(x_2^k-x_2^{k+1})\|^2,\end{aligned}\tag{3.27}$$

where $\sigma'=\sigma/\sigma_{\max}(A_2^TA_2)$. As $\{\beta^k\}$ is non-decreasing, we have

$$\begin{aligned}&(1-\gamma)\beta^{k-1}(x_2^k-x_2^{k+1})^TA_2^T(A_1x_1^k+A_2x_2^k-b)\\&\ge-\frac{(\beta^k+\sigma')}{2}\|A_2(x_2^k-x_2^{k+1})\|^2-\frac{(1-\gamma)^2(\beta^{k-1})^2}{2(\beta^k+\sigma')}\|A_1x_1^k+A_2x_2^k-b\|^2\\&\ge-\frac{(\beta^k+\sigma')}{2}\|A_2(x_2^k-x_2^{k+1})\|^2-\frac{(1-\gamma)^2(\beta^k)^2}{2(\beta^k+\sigma')}\|A_1x_1^k+A_2x_2^k-b\|^2.\end{aligned}\tag{3.28}$$

Note that

$$
\begin{aligned}
& -\frac{(\beta^k)^2}{\beta^k+\sigma'}\|A_1x_1^k+A_2x_2^k-b\|^2=-\frac{(\beta^k)(\beta^k+\sigma'-\sigma')}{\beta^k+\sigma'}\|A_1x_1^k+A_2x_2^k-b\|^2\\
=&\beta^k\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2-(\beta^k-\frac{\sigma'\beta^k}{\beta^k+\sigma'})\|A_1x_1^k+A_2x_2^k-b\|^2\\
&-\beta^k\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2.
\end{aligned} \tag{3.29}
$$

According to (3.24), (3.27), (3.28) and (3.29), we obtain

$$
\begin{aligned}
&\|v^k-\widetilde{v}^k\|_{G^k}^2=\beta^k\|A_2(x_2^k-x_2^{k+1})\|^2+(2-\gamma)\beta^k\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2\\
&+2\beta^k(x_2^k-x_2^{k+1})^TA_2^T(A_1x_1^{k+1}+A_2x_2^{k+1}-b)\\
&\geq(1+\gamma-\gamma^2)\beta^k\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2\\
&+(1-\gamma)^2\{\beta^k\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2-(\beta^k-\frac{\sigma'\beta^k}{\beta^k+\sigma'})\|A_1x_1^k+A_2x_2^k-b\|^2\}.
\end{aligned} \tag{3.30}
$$

Then, according to $\gamma\in(0,\frac{1+\sqrt{5}}{2}]$ and condition (3.25), we have

$$
\beta^k\|v^k-\widetilde{v}^k\|_{G^k}^2\geq(1-\gamma)^2\{(\beta^k)^2\|A_1x_1^{k+1}+A_2x_2^{k+1}-b\|^2-(\beta^{k-1})^2\|A_1x_1^k+A_2x_2^k-b\|^2\}.
$$

Again using condition (3.25), we have

$$
\beta^k\{\|v^{k+1}-v^*\|_{H^k}^2+\sigma\|x_2^{k+1}-x_2^*\|^2-\|v^k-v^*\|_{H^k}^2\}\geq\|v^{k+1}-v^*\|_{H_0^{k+1}}^2-\|v^k-v^*\|_{H_0^k}^2.
$$

The proof is complete. ∎

REMARK 2. *There are scalars $\delta>0$ such that $\beta^k=\delta k$ satisfies* (3.25)*, we can obtain $O(1/k^2)$ convergence. Note that this result has been pointed out in [33] for the spectial case $\gamma=1$, we extend it to $\gamma\in(0,(1+\sqrt{5})/2]$ by the prediction-correction framework* (2.7) *and* (2.8) *satisfying convergence condition* (2.12)*.*

Next we consider an interesting case of (P2) with $f_1(x_1)=g^Tx_1$ where $g\in\mathbb{R}^{n_1}$. It is reported in [6] that ADMM (3.18) is convergent with fixed β^k and $\gamma\in(0,2)$. Based on Theorem 2.5 and the following result, we can establish $O(1/K^2)$ ergodic convergence rate when f_2 is additional σ-strongly convex.

THEOREM 3.4. *If $f_1(x_1)=g^Tx_1$ with $g\in\mathbb{R}^{n_1}$ and f_2 is σ-strongly convex, for $\gamma\in(0,2]$ and*

$$
\beta^k\Big(\beta^k+\frac{\sigma}{\sigma_{\max}(A_2^TA_2)}\Big)\geq(\beta^{k+1})^2, \tag{3.31}
$$

then $\{v^{k+1}\}$ generated by ADMM (3.18) *satisfies that*

$$
\begin{aligned}
&g^TX_1^K+f_2(X_2^K)-g^Tx_1^*-f_2(x_2^*)\leq O(1/\Pi^K),\\
&\|A_1X_1^K+A_2X_2^K-b\|\leq O(1/\Pi^K),
\end{aligned}
$$

where $X_1^K=\sum_{k=0}^{k=K}\beta^kx_1^k/\Pi^K$, $X_2^K=\sum_{k=0}^{k=K}\beta^kx_2^k/\Pi^K$ and $\Pi^K=\sum_{k=0}^{k=K}\beta^k$.

Proof. Let (x_1^*,x_2^*,λ^*) be a saddle point, we can obtain

$$
\begin{cases}
g=A_1^T\lambda^*, & \text{(3.32a)}\\
f_2(x_2)-f_2(x_2^*)-(x_2-x_2^*)^TA_2^T\lambda^*\geq0, & \text{(3.32b)}\\
A_1x_1^*+A_2x_2^*=b. & \text{(3.32c)}
\end{cases}
$$

From the optimality condition we obtain

$$\begin{cases} g - A_1^T(\lambda^k - \beta^k(A_1x_1^{k+1} + A_2x_2^k - b)) = 0, & \text{(3.33a)} \\ f_2(x_2) - f_2(x_2^{k+1}) - (x_2 - x_2^{k+1})^T A_2^T(\lambda^k - \beta^k(A_1x_1^{k+1} + A_2x_2^{k+1} - b)) & \\ \qquad \geq \frac{\sigma}{2}\|x_2^{k+1} - x_2\|^2, & \text{(3.33b)} \\ \lambda^{k+1} = \lambda^k - \gamma\beta^k(A_1x_1^{k+1} + A_2x_2^{k+1} - b). & \text{(3.33c)} \end{cases}$$

According to (3.32a), (3.32c) and (3.33a) we obtain

$$A_1^T(\lambda^k - \lambda^* - \beta^k A_2(x_2^k - x_2^*) - \beta^k A_1(x_1^{k+1} - x_1^*)) = 0. \tag{3.34}$$

Let $P \in \mathbb{R}^{m\times m}$ be the projection matrix on the range space of A_1. We have $PA_1 = A_1$ and $P^2 = P$. Then it follows from (3.34) that

$$A_1(x_1^{k+1} - x_1^*) = \frac{1}{\beta^k}P(\lambda^k - \lambda^*) - P(A_2(x_2^k - x_2^*)). \tag{3.35}$$

According to (3.33b) and (3.35) we obtain

$$\begin{aligned} & f_2(x_2^*) - f_2(x_2^{k+1}) - (x_2^* - x_2^{k+1})^T A_2^T \lambda^* \\ & \geq (x_2^* - x_2^{k+1})^T A_2^T(\lambda^k - \lambda^* - \beta^k(A_1x_1^{k+1} + A_2x_2^{k+1} - b)) + \frac{\sigma}{2}\|x_2^{k+1} - x_2\|^2 \\ & = (x_2^* - x_2^{k+1})^T A_2^T((I-P)(\lambda^k - \lambda^*) + \beta^k P(A_2(x_2^k - x_2^*))) + \beta^k\|x_2^{k+1} - x_2^*\|^2_{A_2^TA_2} \\ & + \frac{\sigma}{2}\|x_2^{k+1} - x_2\|^2. \end{aligned} \tag{3.36}$$

Since $P^2 = P$, we have

$$(x_2^* - x_2^{k+1})^T A_2^T P(A_2(x_2^k - x_2^*)) \geq -\frac{1}{2}(\|P(A_2(x_2^* - x_2^{k+1}))\|^2 + \|P(A_2(x_2^k - x_2^*))\|^2). \tag{3.37}$$

According to (3.32c), (3.33c) and (3.35) we can obtain

$$\lambda^{k+1} - \lambda^* = \lambda^k - \lambda^* + \gamma(\beta^k P(A_2(x_2^k - x_2^*)) - P(\lambda^k - \lambda^*) - \beta^k A_2(x_2^{k+1} - x_2^*)).$$

Then, based on the fact $P^2 = P$, we obtain

$$(I-P)(\lambda^{k+1} - \lambda^*) = (I-P)(\lambda^k - \lambda^*) - \gamma\beta^k(I-P)(A_2(x_2^{k+1} - x_2^*)). \tag{3.38}$$

This means

$$\begin{aligned} \|(I-P)(\lambda^{k+1} - \lambda^*)\|^2 = & \|(I-P)(\lambda^k - \lambda^*)\|^2 + \gamma^2(\beta^k)^2\|(I-P)(A_2(x_2^{k+1} - x_2^*))\|^2 \\ & - 2\gamma\beta^k\langle\lambda^k - \lambda^*, (I-P)(A_2(x_2^{k+1} - x_2^*))\rangle. \end{aligned} \tag{3.39}$$

Combining (3.36), (3.37) and (3.39), we obtain

$$\begin{aligned} & f_2(x_2^*) - f_2(x_2^{k+1}) - (x_2^* - x_2^{k+1})^T A_2^T \lambda^* \\ & \geq \frac{1}{2\gamma\beta^k}\|(I-P)(\lambda^{k+1} - \lambda^*)\|^2 - \frac{1}{2\gamma\beta^k}\|(I-P)(\lambda^k - \lambda^*)\|^2 + \frac{\sigma}{2}\|x_2^{k+1} - x_2^*\|^2 \\ & + \frac{\beta^k}{2}\|P(A_2(x_2^{k+1} - x_2^*))\|^2 - \frac{\beta^k}{2}\|P(A_2(x_2^k - x_2^*))\|^2 \\ & + \beta^k(1 - \frac{\gamma}{2})\|(I-P)(A_2(x_2^{k+1} - x_2^*))\|^2. \end{aligned} \tag{3.40}$$

According to (3.32a) and (3.32c), we have

$$\begin{aligned} & g^T x_1^* + f_2(x_2^*) - g^T x_1^{k+1} - f_2(x_2^{k+1}) + (\lambda^*)^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \\ =& f_2(x_2^*) - f_2(x_2^{k+1}) - (x_2^* - x_2^{k+1})^T A_2^T \lambda^*. \end{aligned} \tag{3.41}$$

Note that

$$\begin{aligned} \frac{\sigma}{2}\|x_2^{k+1} - x_2^*\|^2 &\geq \frac{\sigma}{2\sigma_{\max}(A_2^T A_2)}\|A_2(x_2^{k+1} - x_2^*)\|^2 \\ &\geq \frac{\sigma}{2\sigma_{\max}(A_2^T A_2)}\|P(A_2(x_2^{k+1} - x_2^*))\|^2. \end{aligned} \tag{3.42}$$

Combining (3.31), (3.40), (3.41) and (3.42), we can deduce

$$\begin{aligned} & \beta^k \Big(g^T x_1^* + f_2(x_2^*) - g^T x_1^{k+1} - f_2(x_2^{k+1}) + (\lambda^*)^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \Big) \\ \geq & \frac{1}{2\gamma}\|(I-P)(\lambda^{k+1} - \lambda^*)\|^2 - \frac{1}{2\gamma}\|(I-P)(\lambda^k - \lambda^*)\|^2 + \frac{(\beta^{k+1})^2}{2}\|P(A_2(x_2^{k+1} - x_2^*))\|^2 \\ & - \frac{(\beta^k)^2}{2}\|P(A_2(x_2^k - x_2^*))\|^2 + (\beta^k)^2(1 - \frac{\gamma}{2})\|(I-P)(A_2(x_2^{k+1} - x_2^*))\|^2. \end{aligned}$$

Adding both sides of the above inequality from $k=0$ to $k=K$ implies that

$$g^T X_1^K + f_2(X_2^K) - g^T x_1^* - f_2(x_2^*) - (\lambda^*)^T (A_1 X_1^K + A_2 X_2^K - b) \leq O\Big(1/\sum_{k=0}^{k=K} \beta^k\Big).$$

Combining Lemmas 2.3 and 2.4 completes the proof. □

Remark 3. *There are scalars $\delta > 0$ such that $\beta^k = \delta k$ satisfies* (3.31) *hence $O(1/k^2)$ convergence rate can be obtained.*

3.2.2. Linearized ADMM. We study the linearized ADMM (LADMM):

$$(\text{LADMM}) \quad \begin{cases} x_1^{k+1} = \arg\min\limits_{x_1} L_{\beta^k}(x_1, x_2^k, \lambda^k), \\ x_2^{k+1} = \arg\min\limits_{x_2} \{L_{\beta^k}(x_1^{k+1}, x_2, \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|_{D^k}^2\}, \\ \lambda^{k+1} = \lambda^k - \beta^k(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases} \tag{3.43}$$

where

$$D^k = \tau r \beta^k I - \beta^k A_2^T A_2. \tag{3.44}$$

Then $D^k = \tau D_1^k + (\tau - 1)\beta^k A_2^T A_2$, where

$$D_1^k = r\beta^k I - \beta^k A_2^T A_2.$$

The $O(1/k)$ convergence rate is established in [13] when $\tau \in [\frac{3}{4}, 1]$, $\beta^k = \beta$ and $r > \|A_2\|^2$. It means that D^k may be indefinite so that more efficient numerical performance could be expected in general. The main result of this section is to establish $O(1/k^2)$ convergence rate under strongly convex assumption.

Using the optimality condition of x_2-subproblem in LADMM (3.43) and the fact that f_2 is σ-strongly convex, we obtain, $\forall x_2$,

$$\begin{aligned} f_2(x_2) - f(x_2^{k+1}) + (x_2 - x_2^{k+1})^T \{ & -A_2^T \lambda^k + \beta^k A_2^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \\ & + D^k (x_2^{k+1} - x_2^k)\} \geq \frac{\sigma}{2}\|x_2^{k+1} - x_2\|^2. \end{aligned} \tag{3.45}$$

Hence, similar as the analysis in [13], we can write LADMM as the following scheme:

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that,

$$f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) \geq (v - \widetilde{v}^k)^T Q^k (v^k - \widetilde{v}^k) + \frac{\sigma}{2}\|\widetilde{x}_2^k - x_2\|^2, \ \forall u,$$

$$\text{where} \ \ Q^k = \begin{pmatrix} \tau r \beta^k I & 0 \\ -A_2 & \frac{1}{\beta^k} I \end{pmatrix}. \tag{3.46}$$

[Correction Step.] Update v^{k+1} by

$$v^{k+1} = v^k - M^k(v^k - \widetilde{v}^k), \tag{3.47}$$

$$\text{where} \ \ M^k = \begin{pmatrix} I & 0 \\ -\beta^k A_2 & I \end{pmatrix}. \tag{3.48}$$

We define

$$H^k = \begin{pmatrix} \tau r \beta^k & 0 \\ 0 & \frac{1}{\beta^k} I \end{pmatrix},$$

$$G^k = (Q^k)^T + Q^k - (M^k)^T H^k M^k = \begin{pmatrix} D^k & 0 \\ 0 & \frac{1}{\beta^k} I \end{pmatrix}.$$

Since $\lambda^{k+1} = \lambda^k - \beta^k(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) = \lambda^k - (\lambda^k - \widetilde{\lambda}^k) + \beta^k A_2(x_2^k - \widetilde{x}_2^k)$, we have

$$\begin{aligned} &\|v^k - \widetilde{v}^k\|_{G^k}^2 = \|x^k - x^{k+1}\|_{D^k}^2 + \frac{1}{\beta^k}\|\lambda^k - \widetilde{\lambda}^k\|^2 \\ =&\tau r \beta^k \|x_2^k - x_2^{k+1}\|^2 + \frac{1}{\beta^k}\|\lambda^k - \lambda^{k+1}\|^2 + 2(x_2^k - x_2^{k+1})^T A_2^T (\lambda^k - \lambda^{k+1}). \end{aligned} \tag{3.49}$$

According to Theorem 2.5, it is sufficient to show the following result.

THEOREM 3.5. *If f_2 is σ-strongly convex, for D^k given in* (3.44) *with $\tau \in [\frac{3}{4}, 1]$, $r > \|A_2\|^2$ and*

$$\beta^k(\tau r \beta^k + \sigma) \geq \tau r (\beta^{k+1})^2, \ \beta^{k+1} \geq \beta^k, \ \forall k, \tag{3.50}$$

then $\{v^{k+1}\}$ generated by LADMM (3.43) *satisfies convergence condition* (2.12) *with*

$$H_0^k = \begin{pmatrix} \tau r (\beta^k)^2 & 0 \\ 0 & I \end{pmatrix},$$

$r^k = \beta^k$, $z = x_2$, $\widetilde{z}^k = x_2^{k+1}$, *and* $\Theta^k = \frac{1}{2}\|x_2^{k-1} - x_2^k\|^2_{\tau\beta^k D_1^k + (1-\tau)(\beta^k)^2 A_2^T A_2}$.

Proof. According to (3.45), we obtain, $\forall x_2$,

$$f_2(x_2) - f(x_2^k) + (x_2 - x_2^k)^T\{-A_2^T \lambda^k + D^{k-1}(x_2^k - x_2^{k-1})\} \geq \frac{\sigma}{2}\|x_2^k - x_2\|^2. \tag{3.51}$$

Putting (3.45) with $x_2 = x_2^k$ and (3.51) with $x_2 = x_2^{k+1}$ together yields that

$$(x_2^k - x_2^{k+1})^T\{A_2^T(\lambda^k - \lambda^{k+1}) + D^k(x_2^{k+1} - x_2^k) - D^{k-1}(x_2^k - x_2^{k-1})\} \geq \sigma\|x_2^k - x_2^{k+1}\|^2. \tag{3.52}$$

Since $r > \|A_2\|^2$, $D_1^k \succ 0$. By using Cauchy-Schwarz inequality, for $\tau \in [\frac{3}{4}, 1]$ and non-decrease of $\{\beta^k\}$, we have

$$\begin{aligned}
&(x_2^k - x_2^{k+1})^T D^{k-1}(x_2^k - x_2^{k-1}) = (x_2^k - x_2^{k+1})^T \{\tau D_1^{k-1} - (1-\tau)\beta^{k-1} A_2^T A_2\}(x_2^k - x_2^{k-1}) \\
&\geq -\frac{1}{2}\|x_2^k - x_2^{k+1}\|^2_{\tau D_1^{k-1} + (1-\tau)\beta^{k-1} A_2^T A_2} - \frac{1}{2}\|x_2^{k-1} - x_2^k\|^2_{\tau D_1^{k-1} + (1-\tau)\beta^{k-1} A_2^T A_2} \\
&\geq -\frac{1}{2}\|x_2^k - x_2^{k+1}\|^2_{\tau D_1^{k} + (1-\tau)\beta^{k} A_2^T A_2} - \frac{1}{2}\|x_2^{k-1} - x_2^k\|^2_{\tau D_1^{k} + (1-\tau)\beta^{k} A_2^T A_2}.
\end{aligned} \tag{3.53}$$

Combining (3.52) and (3.53), we obtain

$$\begin{aligned}
&(x_2^k - x_2^{k+1})^T A_2^T (\lambda^k - \lambda^{k+1}) \\
&\geq (x_2^k - x_2^{k+1})^T \{D^k (x_2^k - x_2^{k+1}) + D^{k-1}(x_2^k - x_2^{k-1})\} + \sigma\|x_2^k - x_2^{k+1}\|^2 \\
&\geq \|x_2^k - x_2^{k+1}\|^2_{D^k} - \frac{1}{2}\|x_2^k - x_2^{k+1}\|^2_{\tau D_1^k + (1-\tau)\beta^k A_2^T A_2} \\
&- \frac{1}{2}\|x_2^{k-1} - x_2^k\|^2_{\tau D_1^k + (1-\tau)\beta^k A_2^T A_2} + \sigma\|x_2^k - x_2^{k+1}\|^2 \\
&= \frac{1}{2}\|x_2^k - x_2^{k+1}\|^2_{\tau D_1^k + (1-\tau)\beta^k A_2^T A_2} + \sigma\|x_2^k - x_2^{k+1}\|^2 \\
&- \frac{1}{2}\|x_2^{k-1} - x_2^k\|^2_{\tau D_1^k + (1-\tau)\beta^k A_2^T A_2} - 2(1-\tau)\beta^k\|x_2^k - x_2^{k+1}\|^2_{A_2^T A_2}.
\end{aligned} \tag{3.54}$$

Again using Cauchy-Schwarz inequality and $\tau \in [\frac{3}{4}, 1]$, we obtain

$$\begin{aligned}
(x_2^k - x_2^{k+1})^T A_2^T (\lambda^k - \lambda^{k+1}) &\geq -\frac{\beta^k}{2(5-4\tau)}\|x_2^k - x_2^{k+1}\|^2_{A_2^T A_2} - \frac{5-4\tau}{2\beta^k}\|\lambda^k - \lambda^{k+1}\|^2 \\
&\geq -(\tau - \frac{1}{2})\beta^k\|x_2^k - x_2^{k+1}\|^2_{A_2^T A_2} - \frac{5-4\tau}{2\beta^k}\|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned} \tag{3.55}$$

Then based on (3.49), (3.54) and (3.55), we obtain

$$\begin{aligned}
&\|v^k - \widetilde{v}^k\|^2_{G^k} \\
&= \tau r \beta^k \|x_2^k - x_2^{k+1}\|^2 + \frac{1}{\beta^k}\|\lambda^k - \lambda^{k+1}\|^2 + 2(x_2^k - x_2^{k+1})^T A_2^T (\lambda^k - \lambda^{k+1}) \\
&\geq \frac{1}{2}\|x_2^k - x_2^{k+1}\|^2_{\tau D_1^k + (1-\tau)\beta^k A_2^T A_2} + \sigma\|x_2^k - x_2^{k+1}\|^2 - \frac{1}{2}\|x_2^{k-1} - x_2^k\|^2_{\tau D_1^k + (1-\tau)\beta^k A_2^T A_2} \\
&+ \tau\|x_2^k - x_2^{k+1}\|^2_{D_1^k} + 2(\tau - \frac{3}{4})\{\beta^k\|x_2^k - x_2^{k+1}\|^2_{A_2^T A_2} + \frac{1}{\beta^k}\|\lambda^k - \lambda^{k+1}\|^2\}.
\end{aligned} \tag{3.56}$$

Again using $\tau \in [\frac{3}{4}, 1]$, we have

$$\tau D_1^k + (1-\tau)\beta^k A_2^T A_2 = \tau r \beta^k I + (1-2\tau)\beta^k A_2^T A_2 \succeq \tau r \beta^k I + (1-2\tau)\beta^{k+1} A_2^T A_2.$$

Hence, according to condition (3.50), we have

$$\begin{aligned}
&\beta^k\{\tau D_1^k + (1-\tau)\beta^k A_2^T A_2 + \sigma I\} \succeq \tau r (\beta^k)^2 I + 2\sigma I \beta^k + (1-2\tau)(\beta^{k+1})^2 A_2^T A_2 \\
&\succeq \tau r (\beta^{k+1})^2 I + (1-2\tau)(\beta^{k+1})^2 A_2^T A_2 = \tau \beta^{k+1} D_1^{k+1} + (1-\tau)(\beta^{k+1})^2 A_2^T A_2.
\end{aligned} \tag{3.57}$$

Ignoring nonnegative terms and multiplying both sides of (3.56) by β^k and then combining it with (3.57) yields that

$$\begin{aligned}\beta^k\|v^k-\widetilde{v}^k\|^2_{G^k} \geq & \frac{1}{2}\|x_2^k-x_2^{k+1}\|^2_{\tau\beta^{k+1}D_1^{k+1}+(1-\tau)(\beta^{k+1})^2A_2^TA_2} \\ & -\frac{1}{2}\|x_2^{k-1}-x_2^k\|^2_{\tau\beta^kD_1^k+(1-\tau)(\beta^k)^2A_2^TA_2}.\end{aligned} \tag{3.58}$$

According to the structure of H^k and condition (3.50), we have

$$\beta^k\{\|v^{k+1}-v^*\|^2_{H^k}+\sigma\|x_2^{k+1}-x_2^*\|^2-\|v^k-v^*\|^2_{H^k}\} \geq \|v^{k+1}-v^*\|^2_{H_0^{k+1}}-\|v^k-v^*\|^2_{H_0^k}. \tag{3.59}$$

Combining (3.58) and (3.59) completes the proof. □

REMARK 4. *There is a $\delta>0$ such that $\beta^k=\delta k$ satisfies* (3.50).

3.2.3. Symmetric ADMM. We consider the symmetric ADMM (SADMM) for solving (P2):

$$\text{(SADMM)} \quad \begin{cases} x_1^{k+1} = \arg\min_{x_1}\{L_{\beta^k}(x_1,x_2^k,\lambda^k), \\ \bar{\lambda}^k = \lambda^k-\gamma_1^k\beta^k(A_1x_1^{k+1}+A_2x_2^k-b), \\ x_2^{k+1} = \arg\min_{x_2}\{L_{\beta^k}(x_1^{k+1},x_2,\bar{\lambda}^k), \\ \lambda^{k+1} = \bar{\lambda}^k-\gamma_2^k\beta^k(A_1x_1^{k+1}+A_2x_2^{k+1}-b). \end{cases} \tag{3.60}$$

With fixed γ_1, γ_2 and β^k, SADMM has an $O(1/k)$ convergence rate [11, 10]. We extend this result to $O(1/k^2)$ with dynamic parameters under proper conditions.

Using the optimality condition of x_2-subproblem in SADMM (3.60) and the fact that f_2 is σ-strongly convex, we obtain, $\forall x_2$,

$$f_2(x_2)-f(x_2^{k+1})+(x_2-x_2^{k+1})^T\{-A_2^T\bar{\lambda}^k+\beta^kA_2^T(A_1x_1^{k+1}+A_2x_2^{k+1}-b)\} \geq \frac{\sigma}{2}\|x_2^{k+1}-x_2\|^2. \tag{3.61}$$

Similar to the analysis in [11], we can write SADMM as the scheme:

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that

$$f(x)-f(\widetilde{x}^k)+(u-\widetilde{u}^k)^TF(\widetilde{u}^k) \geq (v-\widetilde{v}^k)^TQ^k(v^k-\widetilde{v}^k)+\frac{\sigma}{2}\|\widetilde{x}_2^k-x_2\|^2,\ \forall u,$$
$$\text{where}\ \ Q^k=\begin{pmatrix}\beta^kA_2^TA_2 & -\gamma_1^kA_2^T \\ -A_2 & \frac{1}{\beta^k}I\end{pmatrix}. \tag{3.62}$$

[Correction Step.] Update v^{k+1} by

$$v^{k+1}=v^k-M^k(v^k-\widetilde{v}^k),$$
$$\text{where}\ \ M^k=\begin{pmatrix}I & 0 \\ -\gamma_2^k\beta^kA_2 & (\gamma_1^k+\gamma_2^k)I\end{pmatrix}. \tag{3.63}$$

We define

$$H^k = \frac{1}{\gamma_1^k+\gamma_2^k}\begin{pmatrix} (\gamma_1^k+\gamma_2^k-\gamma_1^k\gamma_2^k)\beta^k A_2^TA_2 & -\gamma_1^k A_2^T \\ -\gamma_1^k A_2 & \frac{1}{\beta^k} I \end{pmatrix},$$

$$G^k = (Q^k)^T + Q^k - (M^k)^T H^k M^k = \begin{pmatrix} (1-\gamma_2^k)\beta^k A_2^TA_2 & -(1-\gamma_2^k)A_2^T \\ -(1-\gamma_2^k)A_2 & \frac{2-\gamma_1^k-\gamma_2^k}{\beta^k} I \end{pmatrix}.$$

Clearly, $G^k \succeq 0$ if $(1-\gamma_1^k)(1-\gamma_2^k) \geq 0$. According to Theorem 2.5, it is sufficient to prove the following result.

THEOREM 3.6. *Suppose f_2 is σ-strongly convex.*

(i) *For $\gamma_1^k+\gamma_2^k=1$, $\gamma_1 = 1/\beta^k$, $\beta^k \geq 1$ and*

$$\beta^k\Big(\beta^k + \frac{\sigma}{\sigma_{\max}(A_2^TA_2)}\Big) \geq (\beta^{k+1})^2,\ \ \beta^{k+1} \geq \beta^k,\ \ \forall k, \tag{3.64}$$

$\{v^{k+1}\}$ generated by SADMM (3.60) satisfies convergence condition (2.12) with

$$H_0^k = \begin{pmatrix} ((\beta^k)^2-\beta^k+1)A_2^TA_2 & -A_2^T \\ -A_2 & I \end{pmatrix},$$

$r^k = \beta^k$, $z = x_2$, $\widetilde{z}^k = x_2^{k+1}$ and $\Theta^k = 0$.

(ii) *For $\gamma_2^k = 1$, $\gamma_1 = 1/\beta^k$ and*

$$(\beta^k)^2 + \frac{\sigma}{\sigma_{\max}(A_2^TA_2)}(1+\beta^k) \geq (\beta^{k+1})^2, \tag{3.65}$$

$\{v^{k+1}\}$ generated by SADMM (3.60) satisfies convergence condition (2.12) with

$$H_0^k = \begin{pmatrix} (\beta^k)^2 A_2^TA_2 & -A_2^T \\ -A_2 & I \end{pmatrix},$$

$r^k = \beta^k + 1$, $z = x_2$, $\widetilde{z}^k = x_2^{k+1}$ and $\Theta^k = 0$.

Proof. The conditions given in either (i) or (ii) imply that $H \succeq 0$ and $G^k \succeq 0$.

(i) Since $\gamma_1^k+\gamma_2^k = 1$, $\gamma_1 = 1/\beta^k$, we have

$$H^k = \frac{1}{\beta^k}\begin{pmatrix} ((\beta^k)^2-\beta^k+1)A_2^TA_2 & -A_2^T \\ -A_2 & I \end{pmatrix}.$$

According to (3.64), we have

$$\begin{aligned}((\beta^k)^2-\beta^k+1)A_2^TA_2 + \sigma\beta^k I &\succeq ((\beta^k)^2-\beta^{k+1}+1)A_2^TA_2 + \frac{\sigma}{\sigma_{\max}(A_2^TA_2)}\beta^k A_2^TA_2 \\ &\succeq ((\beta^{k+1})^2-\beta^{k+1}+1)A_2^TA_2.\end{aligned}$$

Hence,

$$\beta^k\{\|v^{k+1}-v^*\|_{H^k}^2 + \sigma\|x_2^{k+1}-x_2^*\|^2 - \|v^k-v^*\|_{H^k}^2\} \geq \|v^{k+1}-v^*\|_{H_0^{k+1}}^2 - \|v^k-v^*\|_{H_0^k}^2.$$

(ii) Since $\gamma_2^k = 1$ and $\gamma_1 = 1/\beta^k$, we obtain

$$H^k = \frac{1}{1+\beta^k}\begin{pmatrix} (\beta^k)^2 A_2^TA_2 & -A_2^T \\ -A_2 & I \end{pmatrix}.$$

Hence, under condition (3.65), we obtain

$$
\begin{aligned}
(\beta^k)^2 A_2^T A_2 + (1+\beta^k)\sigma I &\succeq \Big((\beta^k)^2 + (1+\beta^k)\frac{\sigma}{\sigma_{\max}(A_2^T A_2)}\Big) A_2^T A_2 \\
&\succeq (\beta^{k+1})^2 A_2^T A_2.
\end{aligned}
$$

Then it holds that

$$
\begin{aligned}
&(1+\beta^k)\Big(\|v^{k+1}-v^*\|_{H^k}^2 + \sigma\|x_2^{k+1}-x_2^*\|^2 - \|v^k - v^*\|_{H^k}^2\Big) \\
&\geq \|v^{k+1}-v^*\|_{H_0^{k+1}}^2 - \|v^k - v^*\|_{H_0^k}^2.
\end{aligned}
$$

The proof is complete as $G^k \succeq 0$. □

REMARK 5. *There exists $\delta > 0$ such that $\beta^k = \delta k$ satisfies either condition* (3.64) *or* (3.65). *We can multiply the objective function by a positive scalar to adjust the strongly convex parameter to satisfying $\beta^k \geq 1$,*

4. Acceleration under gradient Lipschitz continuous assumption. We study (P3). Let

$$
x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad u = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ \lambda \end{pmatrix}, \quad v = \begin{pmatrix} A_2 x_2 \\ \vdots \\ A_m x_m \\ \lambda \end{pmatrix} \quad \text{and} \quad F(u) = \begin{pmatrix} -A_1^T\lambda \\ -A_2^T\lambda \\ \vdots \\ -A_m^T\lambda \\ Ax - b \end{pmatrix}. \tag{4.1}
$$

It has been shown in [4] that classic ADMM for three block may not be convergent. We consider the following two prediction-correction frameworks [14, 15] with dynamic β^k for solving (P3):

[Prediction Step.]

$$
\begin{cases}
\widetilde{x}_1^k = \arg\min_{x_1}\{L_{\beta^k}(x_1, x_2^k, .., x_m^k, \lambda^k), \\
\widetilde{x}_2^k = \arg\min_{x_2}\{L_{\beta^k}(\widetilde{x}_1^k, x_2, x_3^k, .., x_m^k, \lambda^k), \\
\cdots \\
\widetilde{x}_j^k = \arg\min_{x_j}\{L_{\beta^k}(\widetilde{x}_1^k, \cdots, x_j, x_{j+1}^k, .., x_m^k, \lambda^k), \\
\cdots \\
\widetilde{x}_m^k = \arg\min_{x_m}\{L_{\beta^k}(\widetilde{x}_1^k, \cdots, \widetilde{x}_{m-1}^k, x_m, \lambda^k), \\
\widetilde{\lambda}^k = \lambda^k - \beta^k(A_1\widetilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b).
\end{cases} \tag{4.2}
$$

There are two different correction steps.

[Correction Step.] (Forward substitution procedure)

$$
v^{k+1} = v^k - M^k(v^k - \widetilde{v}^k), \tag{4.3}
$$

where

$$M^k = \alpha \begin{pmatrix} I & 0 & \cdots & \cdots & 0 \\ I & I & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ I & \cdots & I & I & 0 \\ -\beta^k I & \cdots & -\beta^k I & -\beta^k I & I \end{pmatrix} \; and \; \alpha \in (0,1]. \tag{4.4}$$

[Correction Step.] (Backward substitution procedure)

$$(P^k)^T(v^{k+1} - v^k) = -N^k(v^k - \widetilde{v}^k), \tag{4.5}$$

where

$$P^k = \begin{pmatrix} \sqrt{\beta^k} I & 0 & \cdots & \cdots & 0 \\ \sqrt{\beta^k} I & \sqrt{\beta^k} I & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \sqrt{\beta^k} I & \cdots & \sqrt{\beta^k} I & \sqrt{\beta^k} I & 0 \\ 0 & \cdots & 0 & 0 & \frac{1}{\sqrt{\beta^k}} I \end{pmatrix}, \quad N^k = \alpha \begin{pmatrix} \sqrt{\beta^k} I & 0 & \cdots & \cdots & 0 \\ 0 & \sqrt{\beta^k} I & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \sqrt{\beta^k} I & 0 \\ -\sqrt{\beta^k} I & \cdots & -\sqrt{\beta^k} I & -\sqrt{\beta^k} I & \frac{1}{\sqrt{\beta^k}} I \end{pmatrix}$$

and $\alpha \in (0,1]$.

Since P^k is invertible, we can rewrite correct step (4.5) (Backward substitution procedure) as:

$$v^{k+1} = v^k - M^k(v^k - \widetilde{v}^k), \quad M^k = (P^k)^{-T} N^k. \tag{4.6}$$

The λ-subproblem in (4.2) can be simply rearranged as

$$\lambda^k - \widetilde{\lambda}^k = \beta^k(A_1\widetilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b).$$

Then according to forward substitution procedure (4.3), we have

$$\begin{aligned} \lambda^{k+1} &= \lambda^k - \alpha \left(\sum_{j=2}^m \beta^k A_j(\widetilde{x}_j^k - x_j^k) + (\lambda^k - \widetilde{\lambda}^k) \right) \\ &= \lambda^k - \alpha\beta^k (\sum_{i=1}^m A_i \widetilde{x}_i^k - b). \end{aligned}$$

Similarly, on the basis of backward substitution procedure (4.5), we can reduce

$$\begin{aligned} \frac{1}{\sqrt{\beta^k}}(\lambda^{k+1} - \lambda^k) &= -\frac{\alpha}{\sqrt{\beta^k}} \left(\sum_{j=2}^m \beta^k A_j(\widetilde{x}_j^k - x_j^k) + (\lambda^k - \widetilde{\lambda}^k) \right) \\ &= -\frac{\alpha\beta^k}{\sqrt{\beta^k}} (\sum_{i=1}^m A_i \widetilde{x}_i^k - b). \end{aligned}$$

Hence, both two correction steps satisfy that:

$$\lambda^{k+1} = \lambda^k - \alpha\beta^k(\sum_{i=1}^m A_i\widetilde{x}_i^k - b).$$

Then using the optimality condition of x_m-subproblem in prediction step (4.2) and the fact that f_m is L-gradient Lipschitz continuous, we have, $\forall x_m$,

$$\begin{aligned}&f_m(x_m) - f_m(\widetilde{x}_m^k) + (x_m - \widetilde{x}_m^k)^T\{-A_m^T\lambda^k + \beta^k A_m^T(\sum_{i=1}^m A_i\widetilde{x}_i^k - b)\}\\ &\geq (x_m - \widetilde{x}_m^k)^T\{\nabla f_m(\widetilde{x}_m^k) - A_m^T\lambda^k + \beta^k A_m^T(\sum_{i=1}^m A_i\widetilde{x}_i^k - b)\}\\ &+ \frac{1}{2L}\|\nabla f_m(\widetilde{x}_m^k) - \nabla f_m(x_m)\|^2,\end{aligned} \tag{4.7}$$

and

$$0 = \nabla f_m(\widetilde{x}_m^k) - A_m^T\lambda^k + \beta^k A_m^T(\sum_{i=1}^m A_i\widetilde{x}_i^k - b) = \nabla f_m(\widetilde{x}_m^k) + (\frac{1}{\alpha} - 1)A_m^T\lambda^k - \frac{1}{\alpha}A_m^T\lambda^{k+1}. \tag{4.8}$$

Since $0 = \nabla f_m(x_m^*) - A_m^T\lambda^*$, it follows from (4.8) that

$$\begin{aligned}\frac{1}{2L}\|\nabla f_m(\widetilde{x}_m^k) - \nabla f_m(x_m^*)\|^2 &= \frac{1}{2L}\|A_m^T\Big((1 - \frac{1}{\alpha})\lambda^k + \frac{1}{\alpha}\lambda^{k+1} - \lambda^*\Big)\|^2\\ &\geq \frac{\sigma_{\min}(A_m A_m^T)}{2L}\|(1 - \frac{1}{\alpha})\lambda^k + \frac{1}{\alpha}\lambda^{k+1} - \lambda^*\|^2.\end{aligned}$$

Hence, according to (4.7) and (4.8), and then based on a proof similar to that in [15], we can rewrite prediction step (4.2) as

[Prediction Step.] With given v^k, find $\widetilde{u}^k$ such that

$$\begin{aligned}&f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k)\\ &\geq (v - \widetilde{v}^k)^T Q^k(v^k - \widetilde{v}^k) + \frac{1}{2L}\|\nabla f_m(\widetilde{x}_m^k) - \nabla f_m(x_m)\|^2, \ \forall u,\end{aligned} \tag{4.9}$$

where

$$Q^k = \begin{pmatrix} \beta^k I & 0 & \cdots & \cdots & 0\\ \beta^k I & \beta^k I & \ddots & & \vdots\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ \beta^k I & \cdots & \beta^k I & \beta^k I & 0\\ -I & \cdots & -I & -I & \frac{1}{\beta^k}I \end{pmatrix}. \tag{4.10}$$

For prediction-correction framework (4.2) and (4.3), let H^k and G^k satisfy (2.9) and (2.10), then we have

$$H^k = \frac{1}{\alpha}\text{diag}(\beta^k I, \cdots, \beta^k I, \frac{1}{\beta^k}I). \tag{4.11}$$

Then it holds that

$$f(x^*) - f(\widetilde{x}^k) + (u^* - \widetilde{u}^k)^T F(\widetilde{u}^k) \geq \frac{1}{2}\Big(\|v^{k+1} - v^*\|_{H^k}^2 - \|v^k - v^*\|_{H^k}^2\Big) + \frac{1}{2}\|v^k - \widetilde{v}^k\|_{G^k}^2 + \frac{\sigma_{\min}(A_m A_m^T)}{2L}\|(1-\frac{1}{\alpha})\lambda^k + \frac{1}{\alpha}\lambda^{k+1} - \lambda^*\|^2.$$

Based on Theorem 2.5, it is sufficient to prove the following theorem.

THEOREM 4.1. *If f_m is L-gradient Lipschitz continuous, and $\{\beta^k\}$ satisfies that*

$$\frac{1}{(\beta^k)^2} + \frac{\sigma_{\min}(A_m A_m^T)}{L\beta^k} \geq \frac{1}{(\beta^{k+1})^2} + (1-\alpha)\frac{\sigma_{\min}(A_m A_m^T)}{L\beta^{k+1}}, \tag{4.12}$$

then $\{v^{k+1}\}$ generated by prediction-correction framework (4.2) *and* (4.3) *satisfies convergence condition* (2.12) *with*

$$H_0^k = \frac{1}{\alpha}\mathrm{diag}\left(I, \cdots, I, \left(\frac{1}{(\beta^k)^2} + (\frac{1}{\alpha} - 1)\frac{\sigma_{\min}(A_m A_m^T)}{L\beta^k}\right) I\right), \tag{4.13}$$

$r^k = \frac{1}{\beta^k}$, $z^* = \nabla f_m(x_m^*)$, $\widetilde{z}^k = \nabla f_m(\widetilde{x}_m^k)$, $\sigma = \frac{\sigma_{\min}(A_m A_m^T)}{L}$ *and* $\Theta^k = 0$.

Proof. We can verify that

$$\|(1-\frac{1}{\alpha})\lambda^k + \frac{1}{\alpha}\lambda^{k+1} - \lambda^*\|^2 = (1-\frac{1}{\alpha})\|\lambda^k - \lambda^*\|^2 + \frac{1}{\alpha}\|\lambda^{k+1} - \lambda^*\|^2 - (1-\frac{1}{\alpha})\frac{1}{\alpha}\|\lambda^k - \lambda^{k+1}\|^2.$$

Therefor, according to $\alpha \in (0,1]$, (4.12) and $G^k \succeq 0$, we obtain

$$\frac{1}{\beta^k}\{\|v^{k+1} - v^*\|_{H^k}^2 - \|v^k - v^*\|_{H^k}^2 + \|v^k - \widetilde{v}^k\|_{G^k}^2 + \frac{\sigma_{\min}(A_m A_m^T)}{L}\|(1-\frac{1}{\alpha})\lambda^k + \frac{1}{\alpha}\lambda^{k+1} - \lambda^*\|^2\} \geq \|v^{k+1} - v^*\|_{H_0^{k+1}}^2 - \|v^k - v^*\|_{H_0^k}^2.$$

Then we complete the proof. □

For prediction-correction framework (4.2) and (4.5), let H^k and G^k satisfy (2.9) and (2.10), then we have

$$H^k = \frac{1}{\alpha}\begin{pmatrix} \beta^k I & \cdots & \beta^k I & 0 \\ \beta^k I & \cdots & \beta^k I & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \beta^k I & \cdots & \beta^k I & 0 \\ 0 & \cdots & 0 & \frac{1}{\beta^k} I \end{pmatrix}. \tag{4.14}$$

Based on a proof similar to that of Theorem 4.1, we have

THEOREM 4.2. *If f_m is L-gradient Lipschitz continuous, and $\{\beta^k\}$ satisfies* (4.12)*, then $\{v^{k+1}\}$ generated by the prediction-correction framework* (4.2) *and* (4.5) *satisfies convergence condition* (2.12) *with*

$$H_0^k = \frac{1}{\alpha}\begin{pmatrix} I & \cdots & I & 0 \\ I & \cdots & I & 0 \\ \vdots & \ddots & \vdots & \vdots \\ I & \cdots & I & 0 \\ 0 & \cdots & 0 & \left(\frac{1}{(\beta^k)^2} + (1-\alpha)\frac{\sigma_{\min}(A_m A_m^T)}{L\beta^k}\right) I \end{pmatrix}, \tag{4.15}$$

$r^k = \frac{1}{\beta^k}$, $z^* = \nabla f_m(x_m^*)$, $\widetilde{z}^k = \nabla f_m(\widetilde{x}_m^k)$, $\sigma = \frac{\sigma_{\min}(A_m A_m^T)}{L}$ *and* $\Theta^k = 0$.

REMARK 6. *There exists a* $\delta > 0$ *such that* $\beta^k = \delta k$ *satisfies condition* (4.12), *hence* $O(1/k^2)$ *ergodic convergence rate can be obtained. If* $m = 2$ *and* $\alpha = 1$, *prediction-correction framework* (4.2), (4.3) *and* (4.5) *reduce to ADMM* (3.18) *with* $\gamma = 1$. *In this special case, the* $O(1/k^2)$ *ergodic convergence rate under gradient Lipschitz continuous assumption is given in [31]. Based on our prediction-correction framework, we have extended this result to multi-block type ADMM with only one block being gradient Lipschitz continuous.*

Let us consider the following proximal ADMM for solving problem (P2) with $f_1(x_1) = g^T x_1$,

$$\begin{cases} x_1^{k+1} = \arg\min\limits_{x_1}\{L_{\beta^k}(x_1, x_2^k, \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_{D^k}^2\}, \\ x_2^{k+1} = \arg\min\limits_{x_2} L_{\beta^k}(x_1^{k+1}, x_2, \lambda^k), \\ \lambda^{k+1} = \lambda^k - \beta^k(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b), \end{cases}$$

where $D^k \succeq 0$. Let (3.16) and (3.17) hold with $v = (x_1, x_2, \lambda)$. Then v^k satisfies prediction-correction framework (2.7) and (2.8) with

$$Q^k = \begin{pmatrix} D^k & 0 & 0 \\ 0 & \beta^k A_2^T A_2 & 0 \\ 0 & -A_2 & \frac{1}{\beta^k} I \end{pmatrix}, \quad M^k = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\beta^k A_2 & I \end{pmatrix}.$$

Convergence condition (2.9) and (2.10) can be verified with

$$H^k = \begin{pmatrix} D^k & 0 & 0 \\ 0 & \beta^k A_2^T A_2 & 0 \\ 0 & 0 & \frac{1}{\beta^k} I \end{pmatrix}.$$

If f_2 is gradient Lipschitz continuous, in order to satisfy convergence condition (2.12), we select $D^k = \beta^k I$. Then x_1-subproblem becomes

$$g - A_1^T \lambda + \beta^k A_1^T (A_1 x_1^{k+1} + A_2 x_2^k - b) + \beta^k (x_1^{k+1} - x_1^k).$$

Hence, we calculate the inverse of $A_1^T A_1 + I$ only once in all iteration. We can refer to [22] for more result about $O(1/k^2)$ in this situation.

If f_2 is strongly convex, in order to satisfy convergence condition (2.12), we select $D^k = I/\beta^k$. Then x_1-subproblem becomes

$$g - A_1^T \lambda + \beta^k A_1^T (A_1 x_1^{k+1} + A_2 x_2^k - b) + 1/\beta^k (x_1^{k+1} - x_1^k).$$

Hence, we need calculate the inverse of $A_1^T A_1 + 1/(\beta^k)^2 I$ in every iteration. Actually, convergence condition (2.12) holds if we add extra proximal term ($\frac{1}{2}\|x_2 - x_2^k\|_{r\beta^k I - \beta^k A_2^T A_2}^2$ with $r > \|A_2\|^2$) in x_2-subproblem while this may not hold when f_2 is gradient Lipschitz continuous.

5. From ergodic to non-ergodic. We establish the convergence rate from ergodic to non-ergodic. According to our framework given in Section 2, by combining accelerated gradient technology given in [32] and algorithm framework given in [29], we give $O(1/k^2)$ ergodic convergence rate under strongly convex. We also extend the algorithm framework given in [29] under one function being gradient Lipschitz continuous and accelerate the convergence rate of multi-block ADMM type algorithms.

5.1. Faster non-ergodic convergence rate under strongly convex. We first consider the following prediction-correction framework.

[Prediction Step.] With given v^k, $\hat{y}^k$ and $\bar{x}^k$, find $\widetilde{u}^k$ such that

$$
\begin{aligned}
&f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) + (y - \widetilde{y}^k)^T \nabla g(\hat{y}^k) + c^k (A(x - \widetilde{x}^k))^T (A\bar{x}^k - b) \\
&\geq (v - \widetilde{v}^k)^T Q^k (v^k - \widetilde{v}^k) + (y - \widetilde{y}^k)^T P^k (y^k - \widetilde{y}^k) + \frac{\sigma}{2}\|\widetilde{z}^k - z\|^2, \ \forall u,
\end{aligned}
\tag{5.1}
$$

where $y, z \in \{x_i, x\}$.
[Correction Step.] Update v^{k+1} by

$$
v^{k+1} = v^k - M^k (v^k - \widetilde{v}^k). \tag{5.2}
$$

If $Ax^* = b$, H^k and G^k satisfy (2.9) and (2.10), then we can obtain that

$$
\begin{aligned}
&f(x^*) - f(\widetilde{x}^k) + (u^* - \widetilde{u}^k)^T F(\widetilde{u}^k) + (y^* - \widetilde{y}^k)^T \nabla g(\hat{y}^k) + c^k (b - A\widetilde{x}^k)^T (A\bar{x}^k - b) \\
&\geq (v^* - \widetilde{v}^k)^T Q^k (v^k - \widetilde{v}^k) + (y^* - \widetilde{y}^k)^T P^k (y^k - \widetilde{y}^k) + \frac{\sigma}{2}\|\widetilde{z}^k - z^*\|^2 \\
&\geq \frac{1}{2}(\|v^{k+1} - v^*\|_{H^k}^2 + \sigma\|\widetilde{z}^k - z^*\|^2 - \|v^k - v^*\|_{H^k}^2 + \|v^k - \widetilde{v}^k\|_{G^k}^2) \\
&+ \frac{1}{2}(\|\widetilde{y}^k - y^*\|_{P^k}^2 - \|y^k - y^*\|_{P^k}^2 + \|y^k - \widetilde{y}^k\|_{P^k}^2).
\end{aligned}
\tag{5.3}
$$

We give the following lemma to show the convergence rate of prediction-correction framework (5.1) and (5.2).

Lemma 5.1. *Suppose g is L-gradient Lipschitz continuous, $Ax^* = b$, $\widetilde{y}^k = y^{k+1}$, $P^k = (L/\alpha^k)I$, $c^k = (\alpha^{k-1})^2$, $\alpha^k \geq 1$, $(\alpha^{k+1})^2 - (\alpha^k)^2 = \alpha^{k+1}$,*

$$
\bar{x}^{k+1} = (1 - 1/\alpha^k)\bar{x}^k + \widetilde{x}^k/\alpha^k, \tag{5.4}
$$
$$
\bar{y}^{k+1} = (1 - 1/\alpha^k)\bar{y}^k + \widetilde{y}^k/\alpha^k, \tag{5.5}
$$
$$
\hat{y}^k = (1 - 1/\alpha^k)\bar{y}^k + y^k/\alpha^k, \tag{5.6}
$$

and $\{v^{k+1}\}$ generated by the prediction-correction framework (5.1) and (5.2) satisfies the convergence condition (2.9), (2.10) and (2.12) with $r^k = \alpha^k$ and $\Theta^{k+1} - \Theta^k = (\alpha^k)^2\|A\widetilde{x}^k - b\|^2$, then it holds that

$$
\begin{aligned}
&f(\bar{x}^{k+1}) - f(x^*) + g(\bar{y}^{k+1}) - g(y^*) \leq O(1/(\alpha^k)^2), \\
&\|A\bar{x}^{k+1} - b\| \leq O(1/(\alpha^k)^2).
\end{aligned}
$$

Proof. Since g is assumed to be L-gradient Lipschitz continuous, we have

$$
g(\bar{y}^{k+1}) \leq g(\hat{y}^k) + \nabla g(\hat{y}^k)^T (\bar{y}^{k+1} - \hat{y}^k) + \frac{L}{2}\|\bar{y}^{k+1} - \hat{y}^k\|^2.
$$

By (5.5)-(5.6), we obtain

$$\begin{aligned} g(\bar{y}^{k+1}) \leq& g(\hat{y}^k) + (1-\frac{1}{\alpha^k})\nabla g(\hat{y}^k)^T(\bar{y}^k - \hat{y}^k) + \frac{1}{\alpha^k}\nabla g(\hat{y}^k)^T(\widetilde{y}^k - \hat{y}^k) \\ &+ \frac{L}{2(\alpha^k)^2}\|\widetilde{y}^k - y^k\|^2 \\ =&(1-\frac{1}{\alpha^k})[g(\hat{y}^k) + \nabla g(\hat{y}^k)^T(\bar{y}^k - \hat{y}^k)] + \frac{1}{\alpha^k}[g(\hat{y}^k) + \nabla g(\hat{y}^k)^T(y - \hat{y}^k)] \\ &+ \frac{1}{\alpha^k}\nabla g(\hat{y}^k)^T(\widetilde{y}^k - y) + \frac{L}{2(\alpha^k)^2}\|\widetilde{y}^k - y^k\|^2 \\ \leq&(1-\frac{1}{\alpha^k})g(\bar{y}^k) + \frac{1}{\alpha^k}g(y) + \frac{1}{\alpha^k}\nabla g(\hat{y}^k)^T(\widetilde{y}^k - y) + \frac{L}{2(\alpha^k)^2}\|\widetilde{y}^k - y^k\|^2. \end{aligned} \tag{5.7}$$

As $P^k = (L/\alpha^k)I$, $\{v^{k+1}\}$ satisfies convergence condition (2.12), we multiply both sides of (5.3) by β^k and then obtain

$$\begin{aligned} &\alpha^k[f(x^*) - f(\widetilde{x}^k) + (u^* - \widetilde{u}^k)^T F(\widetilde{u}^k) + (y^* - \widetilde{y}^k)^T\nabla g(\hat{y}^k) + c^k(b - A\widetilde{x}^k)^T(A\bar{x}^k - b)] \\ &\geq \frac{1}{2}[\|v^{k+1} - v^*\|^2_{H_0^{k+1}} - \|v^k - v^*\|^2_{H_0^k} + (\alpha^k)^2\|A\widetilde{x}^k - b\|^2] \\ &+ \frac{L}{2}[\|\widetilde{y}^k - y^*\|^2 - \|y^k - y^*\|^2 + \|y^k - \widetilde{y}^k\|^2]. \end{aligned} \tag{5.8}$$

It follows from (5.4) that

$$\begin{aligned} \frac{(\alpha^k)^4}{2}\|A\bar{x}^{k+1} - b\|^2 = &\frac{(\alpha^k)^4}{2}(1-\frac{1}{\alpha^k})^2\|A\bar{x}^k - b\|^2 \\ &+ \frac{(\alpha^k)^2}{2}\|A\widetilde{x}^k - b\|^2 + (\alpha^k)^3(1-\frac{1}{\alpha^k})(A\bar{x}^k - b)^T(A\widetilde{x}^k - b), \end{aligned}$$

which implies that

$$\begin{aligned} \frac{(\alpha^k)^4}{2}\|A\bar{x}^{k+1} - b\|^2 =&\frac{(\alpha^{k-1})^4}{2}\|A\bar{x}^k - b\|^2 + \frac{(\alpha^k)^2}{2}\|A\widetilde{x}^k - b\|^2 \\ &+ \alpha^k(\alpha^{k-1})^2(A\bar{x}^k - b)^T(A\widetilde{x}^k - b). \end{aligned} \tag{5.9}$$

Define

$$S^{k+1} =: f(\bar{x}^{k+1}) - f(x^*) + g(\bar{y}^{k+1}) - g(y^*) - (\lambda^*)^T(A\bar{x}^{k+1} - b) + \frac{(\alpha^k)^2}{2}\|A\bar{x}^{k+1} - b\|^2.$$

Then according to (5.7), (5.8) and (5.9), we have

$$\begin{aligned} &(\alpha^k)^2 S^{k+1} \leq (\alpha^{k-1})^2 S^k + \alpha^k[f(\widetilde{x}^k) - f(x^*) + \nabla g(\hat{y}^k)^T(\widetilde{y}^k - y^*) - (\lambda^*)^T(A\widetilde{x}^k - b)] \\ &+ \frac{(\alpha^k)^2}{2}\|A\widetilde{x}^k - b\|^2 + \alpha^k(\alpha^{k-1})^2(A\bar{x}^k - b)^T(A\widetilde{x}^k - b) + \frac{L}{2}\|\widetilde{y}^k - y^k\|^2 \\ \leq&(\alpha^{k-1})^2 S^k + \frac{1}{2}[\|v^k - v^*\|^2_{H_0^k} - \|v^{k+1} - v^*\|^2_{H_0^{k+1}}] + \frac{L}{2}[\|y^k - y^*\|^2 - \|\widetilde{y}^k - y^*\|^2]. \end{aligned}$$

Then according to $\widetilde{y}^k = y^{k+1}$, we obtain

$$(\alpha^k)^2 S^{k+1} + \frac{1}{2}\|v^{k+1} - v^*\|^2_{H_0^{k+1}} + \frac{L}{2}\|y^{k+1} - y^*\|^2 < \infty.$$

Then combining Lemmas 2.3 and 2.4 completes the proof. □

REMARK 7. *The condition* $(\alpha^{k+1})^2 - (\alpha^k)^2 = \alpha^{k+1}$ *is equivalent to*

$$\alpha^{k+1} = \frac{\sqrt{1+4(\alpha^k)^2}+1}{2},$$

which dates back to the setting in accelerated gradient method [1, 24]. By induction, we obtain $\alpha^k \geq (k+2)/2$. *Hence,* $O(1/k^2)$ *convergence rate is established.*

We give the following examples to clarify that the following algorithms satisfy the framework (5.1) and (5.2).

EXAMPLE 1. *Consider the following problem*

$$\min\{f(x)+g(x):\ Ax=b\},$$

where $f, g : \mathbb{R}^n \to \mathbb{R}$ *is closed, proper, convex,* g *is* L*-gradient Lipschitz continuous,* $A \in \mathbb{R}^{m\times n}$*, and* $b \in \mathbb{R}^m$*. It can be solved by the following algorithm:*

$$\begin{cases} \hat{x}^k = (1-\frac{1}{\alpha^k})\bar{x}^k + \frac{1}{\alpha^k}x^k, \\ \hat{\lambda}^k = \lambda^k - c^k(A\bar{x}^k - b), \\ x^{k+1} = \arg\min\limits_{x}\{f(x) + \nabla g(\hat{x}^k)^T x - (\hat{\lambda}^k)^T(Ax) + \frac{\beta^k}{2}\|Ax-b\|^2 \\ \qquad\qquad + \frac{1}{2}\|x-x^k\|^2_{D^k+\frac{L}{\alpha^k}I}\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta^k(Ax^{k+1}-b), \quad \gamma\in(0,2), \\ \bar{x}^{k+1} = (1-\frac{1}{\alpha^k})\bar{x}^k + \frac{1}{\alpha^k}x^{k+1}, \end{cases} \tag{5.10}$$

where $\beta^k > 0$ *and* $D^k \succeq 0$*. If* f *is* σ*-strongly convex, then algorithm* (5.10) *satisfies the framework* (5.1) *and* (5.2) *with* $v = u$*,* $F(u)$ *given in* (2.2)*,* $\widetilde{x}^k$ *and* $\widetilde{v}^k$ *given in* (3.2)*,* Q^k *given in* (3.3)*,* M^k *given in* (3.5)*,* $y = x$ *and* $z = x$*. Faster convergence can be verified under the condition given in Lemma 5.1.*

We give the proof of Example 1 in Appendix A.

REMARK 8. *Recently, Sabach and Teboulle [29] proposed an algorithm similar to* (5.10) *with* $\hat{x}^k = x^k$ *and* x*-subproblem being replaced by*

$$x^{k+1} = \arg\min_x\{f(x) + \nabla g(\hat{x}^k)^T x - (\hat{\lambda}^k)^T(Ax) + \frac{\beta^k}{2}\|Ax-b\|^2 + \frac{1}{2}\|x-x^k\|^2_{D^k+L\alpha^k I}\}.$$

If $f=0$*,* $A = D^k = 0$ *and* $b=0$*, Sabach and Teboulle's algorithm reduces to:*

$$x^{k+1} = x^k - \frac{1}{L\alpha^k}\nabla g(x^k),$$

which is similar to the subgradient algorithm [30], while our algorithm (5.10) *reduces to the accelerated gradient algorithm given in [32].*

EXAMPLE 2. *Consider the following problem*

$$\min\{f_1(x_1)+f_2(x_2)+g(x_2):\ Ax := A_1x_1 + A_2x_2 = b\},$$

where $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ *is closed, proper, convex for* $i \in \{1,2\}$*,* g *is* L*-gradient Lipschitz continuous, closed, proper, convex. Consider the following algorithm for solving the* *above optimization problem:*

$$\begin{cases} \hat{x}^k = (1-\frac{1}{\alpha^k})\bar{x}^k + \frac{1}{\alpha^k}x^k, \\ \hat{\lambda}^k = \lambda^k - c^k(A\bar{x}^k - b), \\ x_1^{k+1} = \arg\min\limits_{x_1}\{f_1(x_1) - (\hat{\lambda}^k)^T(A_1x_1) + \frac{\beta^k}{2}\|A_1x_1 + A_2x_2^k - b\|^2\}, \\ x_2^{k+1} = \arg\min\limits_{x_2}\{f_2(x_2) + \nabla g(\hat{x}_2^k)^T x_2 - (\hat{\lambda}^k)^T(A_2x) \\ \qquad\qquad + \frac{\beta^k}{2}\|A_1x_1^{k+1} + A_2x_2 - b\|^2 + \frac{1}{2}\|x_2 - x_2^k\|^2_{D^k + \frac{L}{\alpha^k}I}\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta^k(Ax^{k+1} - b), \quad \gamma \in (0,1), \\ \bar{x}^{k+1} = (1-\frac{1}{\alpha^k})\bar{x}^k + \frac{1}{\alpha^k}x^{k+1}, \end{cases} \tag{5.11}$$

where $\beta^k > 0$ and $D^k \succeq 0$. If f_2 is σ-strongly convex, then the above algorithm satisfies the framework (5.1) *and* (5.2) *with u, v, $F(u)$ given in* (3.16)*, $\widetilde{x}^k$ and $\widetilde{v}^k$ given in* (3.17)*,*

$$Q^k = \begin{pmatrix} \beta^k A_2^T A_2 + D^k & 0 \\ -A_2 & \frac{1}{\beta^k}I \end{pmatrix},$$

M^k given in (3.21)*, $y = x_2$ and $z = x_2$. Then faster convergence can be verified by the condition given in Lemma 5.1.*

We give the proof of Example 2 in Appendix B.

5.2. Faster non-ergodic convergence rate under gradient Lipschitz continuous. We consider the following prediction-correction framework.

[Prediction Step.] With given v^k and $\bar{x}^k$, find $\widetilde{u}^k$ such that

$$\begin{aligned} &f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) + c^k(A(x - \widetilde{x}^k))^T(A\bar{x}^k - b) \\ &\geq (v - \widetilde{v}^k)^T Q^k(v^k - \widetilde{v}^k) + \frac{1}{2L}\|\widetilde{z}^k - z\|^2, \ \forall u. \end{aligned} \tag{5.12}$$

[Correction Step.] Update v^{k+1} by

$$v^{k+1} = v^k - M^k(v^k - \widetilde{v}^k). \tag{5.13}$$

If $Ax^* = b$, H^k and G^k satisfy (2.9) and (2.10), we can verify that

$$\begin{aligned} &f(x^*) - f(\widetilde{x}^k) + (u^* - \widetilde{u}^k)^T F(\widetilde{u}^k) + c^k(b - A\widetilde{x}^k)^T(A\bar{x}^k - b) \\ &\geq (v^* - \widetilde{v}^k)^T Q^k(v^k - \widetilde{v}^k) + \frac{1}{2L}\|\widetilde{z}^k - z^*\|^2 \\ &\geq \frac{1}{2}(\|v^{k+1} - v^*\|^2_{H^k} + \frac{1}{L}\|\widetilde{z}^k - z^*\|^2 - \|v^k - v^*\|^2_{H^k} + \|v^k - \widetilde{v}^k\|^2_{G^k}). \end{aligned} \tag{5.14}$$

We give the following lemma to show the convergence rate of prediction-correction framework (5.12) and (5.13).

LEMMA 5.2. *Suppose $Ax^* = b$, $c^k = (1-\alpha^k)$, $0 < \alpha^k \leq 1$ with $\eta > 0$, $1/(\alpha^{k+1})^2 - 1/(\alpha^k)^2 = 1/\alpha^{k+1}$,*

$$\bar{x}^{k+1} = (1-\alpha^k)\bar{x}^k + \alpha^k\widetilde{x}^k$$

and v^{k+1} generated by the prediction-correction framework (5.12) *and* (5.13) *satisfies the convergence condition* (2.12) *with $r^k = 1/\alpha^k$ and $\Theta^{k+1} - \Theta^k = -\frac{(c^k)^2}{\alpha^k}\|A\bar{x}^k - b\|^2 + \|A\widetilde{x}^k - b\|^2$, then*

$$f(\bar{x}^{k+1}) - f(x^*) \le O((\alpha^k)^2), \quad \|A\bar{x}^{k+1} - b\| \le O((\alpha^k)^2).$$

Proof. By the fact $\bar{x}^{k+1} = (1-\alpha^k)\bar{x}^k + \alpha^k\widetilde{x}^k$, we have

$$\begin{aligned}\frac{1}{2(\alpha^k)^2}\|A\bar{x}^{k+1} - b\|^2 &= \frac{1}{2(\alpha^k)^2}(1-\alpha^k)^2\|A\bar{x}^k - b\|^2 + \frac{1}{2}\|A\widetilde{x}^k - b\|^2 \\ &\quad + \frac{1}{\alpha^k}(1-\alpha^k)(A\bar{x}^k - b)^T(A\widetilde{x}^k - b) \\ &= \frac{1}{2(\alpha^{k-1})^2}\|A\bar{x}^k - b\|^2 - \frac{\alpha^k}{2(\alpha^{k-1})^2}\|A\bar{x}^k - b\|^2 + \frac{1}{2}\|A\widetilde{x}^k - b\|^2 \\ &\quad + \frac{1}{\alpha^k}(1-\alpha^k)(A\bar{x}^k - b)^T(A\widetilde{x}^k - b).\end{aligned} \tag{5.15}$$

Because v^{k+1} satisfy the convergence condition (2.12) with $r^k = \frac{1}{\alpha^k}$ and $\Theta^{k+1} - \Theta^k = -\frac{(c^k)^2}{\alpha^k}\|A\bar{x}^k - b\|^2 + \|A\widetilde{x}^k - b\|^2$, we have

$$\begin{aligned}&\frac{1}{\alpha^k}[f(x^*) - f(\widetilde{x}^k) + (\lambda^*)^T(A\widetilde{x}^k - b) + c^k(b - A\widetilde{x}^k)^T(A\bar{x}^k - b)] \\ =&\frac{1}{\alpha^k}[f(x^*) - f(\widetilde{x}^k) + (u^* - \widetilde{u}^k)^T F(\widetilde{u}^k) + c^k(b - A\widetilde{x}^k)^T(A\bar{x}^k - b)] \\ \ge&\frac{1}{2\alpha^k}[\|v^{k+1} - v^*\|^2_{H^k} + \frac{1}{L}\|\widetilde{z}^k - z^*\|^2 - \|v^k - v^*\|^2_{H^k} + \|v^k - \widetilde{v}^k\|^2_{G^k}] \\ \ge&\frac{1}{2}[\|v^{k+1} - v^*\|^2_{H_0^{k+1}} - \|v^k - v^*\|^2_{H_0^k} - \frac{(c^k)^2}{\alpha^k}\|A\bar{x}^k - b\|^2 + \|A\widetilde{x}^k - b\|^2].\end{aligned} \tag{5.16}$$

Define

$$S^{k+1} := f(\bar{x}^{k+1}) - f(x^*) - (\lambda^*)^T(A\bar{x}^{k+1} - b) + \frac{1}{2}\|A\bar{x}^{k+1} - b\|^2.$$

Because $\frac{(c^k)^2}{\alpha^k} = \frac{(\alpha^k)^3}{(\alpha^{k-1})^4} \le \frac{\alpha^k}{(\alpha^{k-1})^2}$, then we obtain

$$\begin{aligned}\frac{1}{(\alpha^k)^2}S^{k+1} \le&\frac{1}{(\alpha^{k-1})^2}S^k + \frac{1}{\alpha^k}[f(\widetilde{x}^k) - f(x^*) - (\lambda^*)^T(A\widetilde{x}^k - b)] + \frac{1}{2}\|A\widetilde{x}^k - b\|^2 \\ &+ \frac{1-\alpha^k}{\alpha^k}(A\bar{x}^k - b)^T(A\widetilde{x}^k - b) - \frac{\alpha^k}{2(\alpha^{k-1})^2}\|A\bar{x}^k - b\|^2 \\ \le&\frac{1}{(\alpha^{k-1})^2}S^k + \frac{1}{2}[\|v^k - v^*\|^2_{H_0^k} - \|v^{k+1} - v^*\|^2_{H_0^{k+1}}].\end{aligned}$$

Hence, we obtain

$$1/(\alpha^k)^2 S^{k+1} + 1/2\|v^{k+1} - v^*\|^2_{H_0^{k+1}} < \infty.$$

Then combining Lemma 2.3 and Lemma 2.4, we get the conclusion. ∎

We give the following example to clarify that the following algorithms satisfy the framework (5.12) and (5.13).

Example 3. *Let u, v and $F(u)$ satisfy* (4.1)*, we consider the following algorithm to solve (P3):*

$$\begin{cases} \hat{\lambda}^k = \lambda^k - c^k(A\bar{x}^k - b); \\ \widetilde{x}_1^k = \arg\min_{x_1}\{L_{\beta^k}(x_1, x_2^k, .., x_m^k, \hat{\lambda}^k); \\ \widetilde{x}_2^k = \arg\min_{x_2}\{L_{\beta^k}(\widetilde{x}_1^k, x_2, x_3^k, .., x_m^k, \hat{\lambda}^k); \\ ... \\ \widetilde{x}_j^k = \arg\min_{x_j}\{L_{\beta^k}(\widetilde{x}_1^k, ..., x_j, x_{j+1}^k, .., x_m^k, \hat{\lambda}^k); \\ ... \\ \widetilde{x}_m^k = \arg\min_{x_m}\{L_{\beta^k}(\widetilde{x}_1^k, ..., \widetilde{x}_{m-1}^k, x_m, \hat{\lambda}^k); \\ \widetilde{\lambda}^k = \lambda^k - \beta^k(A_1\widetilde{x}_1^k + \sum_{j=2}^m A_j x_j^k - b); \\ v^{k+1} = v^k - M^k(v^k - \widetilde{v}^k); \\ \bar{x}^{k+1} = (1-\alpha^k)\bar{x}^k + \alpha^k\widetilde{x}^k. \end{cases} \tag{5.17}$$

Where M^k is given in (4.4) *or* (4.6) *and v is given in* (4.1)*. If f_m is L-gradient Lipschitz continuous, then the above algorithm satisfies the framework* (5.12) *and* (5.13) *with Q^k given in* (4.10)*, $z = \nabla f_m(x_m)$. Then convergence can be verified under the condition given in Lemma 5.2.*

We give the proof of Example 3 in Appendix C.

6. Conclusions. We substantially extend He and Yuan's unified prediction-correction framework to analyze $O(1/k^2)$ convergence rate for Lagrangian-based methods under strongly convex assumption, such as general linearized ALM with indefinite proximal term, ADMM with step size of Lagrangian multiplier not larger than $(1+\sqrt{5})/2$ (or 2 when one of the two summation function is linear), linearized ADMM with indefinite proximal term, and symmetric ADMM. For multi-block optimization problem, if one block of the objective function is gradient Lipschitz continuous, we establish $O(1/k^2)$ convergence rate in ergodic sense of ADMM type methods. We also extend our framework to establish non-ergodic $O(1/k^2)$ convergence rate. The future work is to explore combining extrapolation technique and prediction-correction framework for Lagrangian-based methods. Besides, further applications of our unified prediction-correction framework are expected.

Appendix A. Proof of Example 1. According to optimality of x-subproblem in (5.10) and σ-strongly convexity of f, we have

$$f(x) - f(x^{k+1}) + (x - x^{k+1})^T\{\nabla g(\hat{x}^k) - A^T(\lambda^k - c^k(A\bar{x}^k - b)) + \beta^k A^T(Ax^{k+1} - b) + (D^k + \frac{L}{\alpha^k}I)(x^{k+1} - x^k)\} \geq \frac{\sigma}{2}\|x^{k+1} - x\|^2, \ \forall x. \tag{A.1}$$

According to $\widetilde{x}^k = x^{k+1}$, then (A.1) can be rewritten as

$$f(x) - f(\widetilde{x}^k) + (x - \widetilde{x}^k)^T\{-A^T\widetilde{\lambda}^k\} + (x - \widetilde{x}^k)^T\nabla g(\hat{x}^k) + c^k(A(x - \widetilde{x}^k))^T(A\bar{x}^k - b) \geq (x - \widetilde{x}^k)^T(D^k + \frac{L}{\alpha^k}I)(x^k - \widetilde{x}^k) + \frac{\sigma}{2}\|x^{k+1} - x\|^2, \ \forall x. \tag{A.2}$$

Note that $\widetilde{\lambda}^k$ defined in (3.2) can be rewritten as

$$(\lambda - \widetilde{\lambda}^k)^T\{(A\widetilde{x}^k - b) + \frac{1}{\beta^k}(\widetilde{\lambda}^k - \lambda^k)\} \geq 0, \ \forall \lambda. \tag{A.3}$$

According to $v = u$ and $F(u)$ defined in (2.2), we can rewrite (A.2) and (A.3) as

$$f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) + (x - \widetilde{x}^k)^T \nabla g(\hat{x}^k) + c^k (A(x - \widetilde{x}^k))^T (A\bar{x}^k - b) \geq (v - \widetilde{v}^k)^T \begin{pmatrix} D^k & 0 \\ 0 & \frac{1}{\beta^k} I \end{pmatrix} (v^k - \widetilde{v}^k) + \frac{L}{\alpha^k}(x - \widetilde{x}^k)^T (x^k - \widetilde{x}^k) + \frac{\sigma}{2}\|\widetilde{x}^k - x\|^2, \ \forall u. \tag{A.4}$$

Then according to (3.4),

$$v^{k+1} = v^k - \begin{pmatrix} I & 0 \\ 0 & \gamma I \end{pmatrix} (v^k - \widetilde{v}^k). \tag{A.5}$$

Then (A.4) and (A.5) satisfies prediction-correction framework (5.1) and (5.2). According to (3.11), G^k defined in (2.10) satisfy that

$$\|v^k - \widetilde{v}^k\|_{G^k}^2 = \|x^k - x^{k+1}\|_{D^k}^2 + \frac{2-\gamma}{\beta^k}\|\lambda^k - \widetilde{\lambda}^k\|^2 \geq (2-\gamma)\beta^k \|A\widetilde{x}^k - b\|^2.$$

Hence, let $\beta^k = \alpha^k/(2-\gamma)$. Similar as the analysis in Section 3, we can verify that

$$\begin{aligned} &\alpha^k [\|v^{k+1} - v^*\|_{H^k}^2 + \sigma\|x^{k+1} - x^*\|^2 - \|v^k - v^*\|_{H^k}^2 + \|v^k - \widetilde{v}^k\|_{G^k}^2] \\ \geq &\|v^{k+1} - v^*\|_{H_0^{k+1}}^2 - \|v^k - v^*\|_{H_0^k}^2 + (\alpha^k)^2 \|A\widetilde{x}^k - b\|^2. \end{aligned}$$

Convergence condition (2.12) can be verified with $r^k = \alpha^k$ and $\Theta^{k+1} - \Theta^k = (\alpha^k)^2 \|A\widetilde{x}^k - b\|^2$.

Appendix B. Proof of Example 2. According to x_1-subproblem in (5.11) and $\widetilde{x}^k$, $\widetilde{\lambda}^k$ defined in (3.17), we have

$$f_1(x_1) - f_1(\widetilde{x}_1^k) + (x_1 - \widetilde{x}_1^k)^T \{-A_1^T \widetilde{\lambda}^k + c^k A_1^T (A\bar{x}^k - b)\} \geq 0, \ \forall x_1. \tag{B.1}$$

According to x_2-subproblem in (5.11), σ-strongly convexity of f_2 and $\widetilde{x}^k$, $\widetilde{\lambda}^k$ defined in (3.17), we have

$$f_2(x_2) - f_2(\widetilde{x}_2^k) + (x_2 - \widetilde{x}_2^k)^T \{\nabla g(\hat{x}_2^k) - A_2^T \widetilde{\lambda}^k + c^k A_2^T (A\bar{x}^k - b) - \beta^k A_2^T A_2 (x_2^k - \widetilde{x}_2^k) - (D^k + \frac{L}{\alpha^k} I)(x_2^k - \widetilde{x}_2^k)\} \geq \frac{\sigma}{2}\|\widetilde{x}_2^k - x_2\|^2, \ \forall x_2. \tag{B.2}$$

Note that $\widetilde{\lambda}^k$ defined in (3.17) can be rewritten as

$$(\lambda - \widetilde{\lambda}^k)^T \{(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b) + A_2(x_2^k - \widetilde{x}_2^k) - \frac{1}{\beta^k}(\lambda^k - \widetilde{\lambda}^k)\} \geq 0, \ \forall \lambda. \tag{B.3}$$

According to u, v and $F(u)$ defined in (3.16), we can rewrite (B.1), (B.2) and (B.3) as

$$f(x) - f(\widetilde{x}^k) + (u - \widetilde{u}^k)^T F(\widetilde{u}^k) + (x_2 - \widetilde{x}_2^k)^T \nabla g(\hat{x}_2^k) + c^k (A(x - \widetilde{x}^k))^T (A\bar{x}^k - b) \geq (v - \widetilde{v}^k)^T \begin{pmatrix} \beta^k A_2^T A_2 + D^k & 0 \\ A_2 & \frac{1}{\beta^k} I \end{pmatrix} (v^k - \widetilde{v}^k) + \frac{L}{\alpha^k}(x_2 - \widetilde{x}_2^k)^T (x_2^k - \widetilde{x}_2^k) + \frac{\sigma}{2}\|\widetilde{x}_2^k - x_2\|^2, \tag{B.4}$$

where $f(x) = f_1(x_1) + f_2(x_2)$. Based on (3.20),

$$v^{k+1} = v^k - \begin{pmatrix} I & 0 \\ -\gamma\beta^k A_2 & \gamma I \end{pmatrix} (v^k - \widetilde{v}^k). \tag{B.5}$$

Then (B.4) and (B.5) satisfy prediction-correction framework (5.1) and (5.2). According to the structure of G^k satisfying (2.10), we can verify that

$$\|v^k - \widetilde{v}^k\|_{G^k}^2 \geq \eta\beta^k\|A\widetilde{x}^k - b\|^2, \ \eta > 0.$$

Hence, let $\beta^k = \alpha^k/\eta$. Similar as the analysis in Appendix B, convergence condition (2.12) can be verified with $r^k = \alpha^k$ and $\Theta^{k+1} - \Theta^k = (\alpha^k)^2\|A\widetilde{x}^k - b\|^2$.

Appendix C. Proof of Example 3. According to x_i-subproblems in (5.17) for $i \in [1, m-1]$, we have, $\forall x_i$

$$f_i(x_i)-f_i(\widetilde{x}_i^k)+(x_i-\widetilde{x}_i^k)^T\{-A_i^T\lambda^k+c^kA_i^T(A\bar{x}^k-b)+\beta^kA_i^T(\sum_{j=1}^{i}A_j\widetilde{x}_j^k+\sum_{j=i+1}^{m}A_jx_j^k-b)\} \geq 0. \tag{C.1}$$

Based on $\widetilde{\lambda}^k = \lambda^k - \beta^k(A\widetilde{x}_1^k + \sum_{j=2}^m Ax_j^k - b)$, we have, for $i \in [1, m-1]$, $\forall x_i$,

$$f_i(x_i) - f_i(\widetilde{x}_i^k) + (x_i - \widetilde{x}_i^k)^T\{-A_i^T\widetilde{\lambda}^k + c^kA_i^T(A\bar{x}^k - b) + \beta^kA_i^T(\sum_{j=2}^{i}A_j(\widetilde{x}_j^k - x_j^k))\} \geq 0. \tag{C.2}$$

According to x_m-subproblem in (5.17) and f_m being L-gradient Lipschitz continuous, we have

$$\begin{aligned}
&f_m(x_m) - f_m(\widetilde{x}_m^k) + (x_m - \widetilde{x}_m^k)^T\{-A_m^T\widetilde{\lambda}^k + c^kA_m^T(A\bar{x}^k - b) + \beta^kA_m^T(\sum_{j=2}^{m}A_j(\widetilde{x}_j^k - x_j^k))\} \\
&\geq (x_m - \widetilde{x}_m^k)^T\{\nabla f_m(\widetilde{x}_m^k) - A_m^T\lambda^k + c^kA_m^T(A\bar{x}^k - b) + \beta^kA_m^T(\sum_{i=1}^{m}A_i\widetilde{x}_i^k - b)\} \\
&+ \frac{1}{2L}\|\nabla f_m(\widetilde{x}_m^k) - \nabla f_m(x_m)\|^2, \ \forall x_m.
\end{aligned} \tag{C.3}$$

Because of the optimality condition of x_m-subproblem and $\lambda^{k+1} = \lambda^k - \alpha\beta^k(\sum_{i=1}^m A_i\widetilde{x}_i^k - b)$, we obtain

$$\begin{aligned}
0 &= \nabla f_m(\widetilde{x}_m^k) - A_m^T\hat{\lambda}^k + \beta^kA_m^T(\sum_{i=1}^{m}A_i\widetilde{x}_i^k - b) \\
&= \nabla f_m(\widetilde{x}_m^k) + (\frac{1}{\alpha} - 1)A_m^T\lambda^k - \frac{1}{\alpha}A_m^T\lambda^{k+1} + c^kA_m^T(A\bar{x}^k - b).
\end{aligned} \tag{C.4}$$

Note that $\widetilde{\lambda}^k$ defined in (4.2) can be rewritten as

$$(\lambda - \widetilde{\lambda}^k)^T\{(\sum_{i=1}^{m}A_i\widetilde{x}_i^k - b) - \sum_{j=2}^{m}A_j(\widetilde{x}_j^k - x_j^k) - \frac{1}{\beta^k}(\lambda^k - \widetilde{\lambda}^k)\} \geq 0, \ \forall\lambda. \tag{C.5}$$

On the basis of (C.2), (C.3), (C.4) and (C.5), we have

$$
\begin{aligned}
&f(x)-f(\widetilde{x}^k)+\begin{pmatrix} x_1-\widetilde{x}_1^k\\ x_2-\widetilde{x}_2^k\\ \vdots\\ x_i-\widetilde{x}_i^k\\ \vdots\\ x_m-\widetilde{x}_m^k\\ \lambda-\widetilde{\lambda}^k\end{pmatrix}^T \left\{\begin{pmatrix} -A_1^T\widetilde{\lambda}^k\\ -A_2^T\widetilde{\lambda}^k\\ \vdots\\ -A_i^T\widetilde{\lambda}^k\\ \vdots\\ -A_m^T\widetilde{\lambda}^k\\ \sum_{i=1}^m A_i\widetilde{x}_m^k-b\end{pmatrix}+\begin{pmatrix} c^kA_1(A\bar{x}^k-b)\\ c^kA_2(A\bar{x}^k-b)\\ \vdots\\ c^kA_i(A\bar{x}^k-b)\\ \vdots\\ c^kA_m(A\bar{x}^k-b)\\ 0\end{pmatrix}\right.\\
&\quad\left.+\begin{pmatrix} 0\\ \beta^kA_2\sum_{j=2}^2 A_j(\widetilde{x}_j^k-x_j^k)\\ \vdots\\ \beta^kA_i\sum_{j=2}^i A_j(\widetilde{x}_j^k-x_j^k)\\ \vdots\\ \beta^kA_m\sum_{j=2}^m A_j(\widetilde{x}_j^k-x_j^k)\\ -\sum_{j=2}^m A_j(\widetilde{x}_j^k-x_j^k)+\frac{1}{\beta^k}(\widetilde{\lambda}^k-\lambda^k)\end{pmatrix}\right\}\geq\frac{1}{2L}\|\nabla f_m(\widetilde{x}_m^k)-\nabla f_m(x_m)\|^2.
\end{aligned}
\tag{C.6}
$$

According to x, u, v and $F(u)$ defined in (4.1), we obtain that (C.6) satisfies (5.12). Then algorithm (5.17) fits prediction-correction framework (5.12) and (5.13). Similar as the analysis given in [15], algorithm (5.17) under framework (5.12) and (5.13) with G^k defined in (2.10) satisfies that, for M^k defined in (4.4), there exist $\delta_1>0$, such that

$$
\|v^k-\widetilde{v}^k\|_{G^k}^2\geq\delta_1\|\mathrm{diag}(\sqrt{\beta^k}I,...,\sqrt{\beta^k}I,\frac{1}{\sqrt{\beta^k}}I)M^k(v^k-\widetilde{v}^k)\|^2\geq\alpha^2\delta_1\beta^k\|\sum_{i=1}^m A_i\widetilde{x}_i^k-b\|^2.
$$

For M^k defined in (4.6), there exist $\delta_2>0$ such that

$$
\|v^k-\widetilde{v}^k\|_{G^k}^2\geq\delta_2\|N^k(v^k-\widetilde{v}^k)\|\alpha^2\geq\alpha^2\delta_2\beta^k\|\sum_{i=1}^m A_i\widetilde{x}_i^k-b\|^2.
$$

According to $0=\nabla f_m(x_m^*)-A_m^T\lambda^*$ and combining (C.4), we have

$$
\begin{aligned}
&\frac{1}{L}\|\nabla f_m(\widetilde{x}_m^k)-\nabla f_m(x_m^*)\|^2=\frac{1}{L}\|A_m^T\Big((1-\frac{1}{\alpha})\lambda^k+\frac{1}{\alpha}\lambda^{k+1}-\lambda^*-c^k(A\bar{x}^k-b)\Big)\|^2\\
&\geq\sigma'\|(1-\frac{1}{\alpha})\lambda^k+\frac{1}{\alpha}\lambda^{k+1}-\lambda^*-c^k(A\bar{x}^k-b)\|^2\\
&\geq\Big(\sigma'-\frac{\sigma'^2}{\sigma'+1}\Big)\|(1-\frac{1}{\alpha})\lambda^k+\frac{1}{\alpha}\lambda^{k+1}-\lambda^*\|^2-(c^k)^2\|A\bar{x}^k-b\|^2,
\end{aligned}
$$

where $\sigma'=\frac{\sigma_{min}(A_mA_m^T)}{L}$ and the last inequation holds by the fact

$$
\begin{aligned}
&\sigma'\|s-t\|^2=\sigma'\|s\|^2+\sigma'\|t\|^2-2\sigma's^Tt\\
\geq&\sigma'\|s\|^2+\sigma'\|t\|^2-(\sigma'+1)\|s\|^2-\frac{\sigma'^2}{\sigma'+1}\|t\|^2,\quad\forall s,t.
\end{aligned}
$$

Let $\beta^k = \alpha^k/(\alpha^2\delta_i),\ i=1,2.$ Then

$$\begin{aligned}
&\frac{1}{\alpha^k}[\|v^{k+1}-v^*\|^2_{H^k} + \frac{1}{L}\|\nabla f_m(\widetilde{x}^k_m) - \nabla f_m(x^*_m)\|^2 - \|v^k - v^*\|^2_{H^k} + \|v^k - \widetilde{v}^k\|^2_{G^k}] \\
\geq & \frac{1}{\alpha^k}[\|v^{k+1}-v^*\|^2_{H^k} + \frac{\sigma''}{\alpha}\|\lambda^{k+1}-\lambda^*\|^2 - \|v^k - v^*\|^2_{H^k} - \sigma''(\frac{1}{\alpha}-1)\|\lambda^k - \lambda^*\|^2] \\
&- \frac{(c^k)^2}{\alpha^k}\|A\bar{x}^k - b\|^2 + \|A\widetilde{x}^k - b\|^2 \\
\geq & \frac{1}{\alpha^2\delta_i}[\|v^{k+1}-v^*\|^2_{H_0^{k+1}} - \|v^k - v^*\|^2_{H_0^k}] - \frac{(c^k)^2}{\alpha^k}\|A\bar{x}^k - b\|^2 + \|A\widetilde{x}^k - b\|^2,
\end{aligned} \tag{C.7}$$

where $\sigma'' = \sigma' - \frac{\sigma'^2}{\sigma'+1}$, H^k and H_0^k defined in (4.11) and (4.13) or (4.14) and (4.15), and the last inequation holds by the following assumption:

$$\frac{1}{(\beta^k)^2} + \frac{\sigma''}{\beta^k} \geq \frac{1}{(\beta^{k+1})^2} + (1-\alpha)\frac{\sigma''}{\beta^{k+1}}.$$

Then v^{k+1} generated by the prediction-correction framework (5.12) and (5.13) satisfies the convergence condition (2.12) with $r^k = \frac{1}{\alpha^k}$ and $\Theta^{k+1} - \Theta^k = -\frac{(c^k)^2}{\alpha^k}\|A\bar{x}^k - b\|^2 + \|A\widetilde{x}^k - b\|^2$.

REFERENCES

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

[2] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40 (2011), pp. 120–145.

[3] ———, *On the ergodic convergence rates of a first-order primal-dual algorithm*, Math. Program., 159 (2016), pp. 253–287.

[4] C. H. Chen, B. S. He, Y. Y. Ye, and X. M. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, Math. Program., 155 (2016), pp. 57–79.

[5] J. Eckstein, *Some saddle-function splitting methods for convex programming*, Optim. Method Softw., 4 (1994), pp. 75–83.

[6] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Comput. Math. with Appl., 2 (1976), pp. 17–40.

[7] R. Glowinski and A. Marroco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires*, ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 9 (1975), pp. 41–76.

[8] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, *Fast alternating direction optimization methods*, SIAM J. Imaging Sci., 7 (2014), pp. 1588–1623.

[9] B. S. He, *My 20 years research on alternating directions method of multipliers*, Oper. Res. Trans, 22 (2018), pp. 1–31.

[10] B. S. He, H. Liu, Z. R. Wang, and X. M. Yuan, *A strictly contractive Peaceman-Rachford splitting method for convex programming*, SIAM J. Optim., 24 (2014), pp. 1011–1040.

[11] B. S. He, F. Ma, and X. M. Yuan, *Convergence study on the symmetric version of ADMM with larger step sizes*, SIAM J. Imaging Sci., 9 (2016), pp. 1467–1501.

[12] ———, *Optimal proximal augmented Lagrangian method and its application to full Jacobian splitting for multi-block separable convex minimization problems*, IMA J. Numer. Anal., 40 (2020), pp. 1188–1216.

[13] ———, *Optimally linearizing the alternating direction method of multipliers for convex programming*, Comput. Optim. Appl., 75 (2020), pp. 361–388.

[14] B. S. He, M. Tao, and X. M. Yuan, *Alternating direction method with Gaussian back substitution for separable convex programming*, SIAM J. Optim., 22 (2012), pp. 313–340.

[15] ———, *Convergence rate analysis for the alternating direction method of multipliers with a*

substitution procedure for separable convex programming, Math. Oper. Res., 42 (2017), pp. 662–691.

[16] B. S. He and X. M. Yuan, *On the o(1/n) convergence rate of the Douglas-Rachford alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[17] ———, *On the o(1/t) convergence rate of the alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[18] ———, *On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers*, Numer. Math., 130 (2015), pp. 567–577.

[19] ———, *A class of ADMM-based algorithms for three-block separable convex programming*, Comput. Optim. Appl., 70 (2018), pp. 791–826.

[20] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.

[21] H. Li and Z. C. Lin, *Accelerated alternating direction method of multipliers: an optimal o(1/k) nonergodic analysis*, J. Sci. Comput., 79 (2019), pp. 671–699.

[22] S. R. Li, Y. Xia, and T. Zhang, *Alternating direction method of multipliers for convex programming: a lift-and-permute scheme*, arXiv:2203.16271, (2022).

[23] R. D. C. Monteiro and B. F. Svaiter, *Iteration-complexity of block-decomposition algorithms and the alternating direction method of multipliers*, SIAM J. Optim., 23 (2013), pp. 475–507.

[24] Y. E. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, in Dokl. akad. nauk Sssr, vol. 269, 1983, pp. 543–547.

[25] Y. Ouyang, Y. M. Chen, G. H. Lan, and E. Pasiliao Jr, *An accelerated linearized alternating direction method of multipliers*, SIAM J. Imaging Sci., 8 (2015), pp. 644–681.

[26] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, Optimization, (1969), pp. 283–298.

[27] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.

[28] ———, *Monotone operators and the proximal point algorithm*, SIAM journal on control and optimization, 14 (1976), pp. 877–898.

[29] S. Sabach and M. Teboulle, *Faster Lagrangian-based methods in convex optimization*, SIAM J. Optim., 32 (2022), pp. 204–227.

[30] N. Z. Shor, *Minimization methods for non-differentiable functions*, vol. 3, Springer Science & Business Media, 2012.

[31] W. Tian and X. Yuan, *An alternating direction method of multipliers with a worst-case* $O(1/n^2)$ *convergence rate*, Math. Comput., 88 (2019), pp. 21–56.

[32] P. Tseng, *Approximation accuracy, gradient methods, and error bound for structured convex optimization*, Math. Program., 125 (2010), pp. 263–295.

[33] Y. Y. Xu, *Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming*, SIAM J. Optim., 27 (2017), pp. 1459–1484.

[34] J. Yang and X. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comput., 82 (2013), pp. 301–329.