# SOLVING SEPARABLE CONVEX OPTIMIZATION PROBLEMS: FASTER PREDICTION-CORRECTION FRAMEWORK*

TAO ZHANG†, YONG XIA†, AND SHIRU LI†

**Abstract.** He and Yuan's prediction-correction framework [SIAM J. Numer. Anal. 50: 700-709, 2012] is able to provide convergent algorithms for solving separable convex optimization problems at a rate of $O(1/t)$ ($t$ represents iteration times) in both ergodic (the average of iteration) and pointwise senses. This paper presents a faster prediction-correction framework at a rate of $O(1/t)$ in the non-ergodic sense (the last iteration) and $O(1/t^2)$ in the pointwise sense. Based the faster prediction-correction framework, we give three faster algorithms which enjoy $O(1/t)$ in the non-ergodic sense of primal-dual gap and $O(1/t^2)$ in the pointwise sense. The first algorithm updates dual variable twice when solving two-block separable convex optimization with equality linear constraints. The second algorithm solves multi-block separable convex optimization problems with linear equality constraints in Gauss-Seidel way. The third algorithm solves minmax problems with larger step sizes.

**Key words.** Prediction-correction, Separable convex optimization, Non-ergodic sense, Pointwise sense.

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

## 1. Introduction.

He and Yuan presented a prediction-correction framework [19] (or see [21]) to analyze the convergence rate of the alternating direction method of multipliers (ADMM) [7, 8]. Representing ADMM in their prediction-correction framework immediately leads to an $O(1/t)$ ($t$ represents the number of iterations) convergence rate of the primal-dual gap in the ergodic sense* In a series of follow-up works, He et al. [11, 12, 13, 15, 20] further proved that the prediction-correction framework enjoys an $O(1/t)$ convergence rate in the pointwise sense.

The prediction-correction framework actually provides a unified approach for developing and analyzing algorithms that solve the following two separable convex optimization problems.

Example 1. *The multi-block separable convex optimization problem with equality constraints:*

$$\text{(P1)} \qquad \min_{x_i} \left\{ f(x) = \sum_{i=1}^{m} f_i(x_i) : \ (Ax :=) \sum_{i=1}^{m} A_i x_i = b \right\},$$

*where* $m \geq 1$, $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ *is a closed proper convex function,* $A_i \in \mathbb{R}^{l \times n_i}$, $\sum_{i=1}^{m} n_i = n$ *and* $b \in \mathbb{R}^l$*. Define the Lagrangian function of* (P1)*:*

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

*This work was supported by the National Natural Science Foundation of China under Grant 12171021 and the Fundamental Research Funds for the Central Universities.

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

*In this paper, for given iteration sequence $\{v^t\}$, the ergodic sense represents the average of $\{v^t\}$; the non-ergodic sense represents the last iteration sequence of $\{v^t\}$; pointwise sense represents $\|v^t - v^{t-1}\|^2$.

*where* $\lambda \in \mathbb{R}^l$ *is the corresponding Lagrange multiplier. For convenience, we define*

$$\theta(u) := f(x) = \sum_{i=1}^{m} f_i(x_i),\ u := x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix},\ w := \begin{pmatrix} x_1 \\ \vdots \\ x_m \\ \lambda \end{pmatrix},\ F(w) := \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_m^T\lambda \\ Ax-b \end{pmatrix}.$$

*Then according to the optimality condition,* (P1) *is equivalent to find* $w^* = (x_1^*, \ldots, x_m^*, \lambda^*)$ *such that*

$$\text{(P1')} \qquad 0 \in T(w^*),\ \text{where } T(w) := \begin{pmatrix} \partial f_1(x_1) \\ \vdots \\ \partial f_m(x_m) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\lambda \\ \vdots \\ -A_m^T\lambda \\ Ax-b \end{pmatrix}.$$

*Noting that, for a given point* $\hat{w} := \begin{pmatrix} \hat{x} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \hat{u} \\ \hat{\lambda} \end{pmatrix}$, *it holds that*

$$\theta(\hat{u}) - \theta(u) - (w-\hat{w})^T F(w) = L(\hat{x}, \lambda) - L(x, \hat{\lambda}),$$

*which is exactly the primal dual gap.*

EXAMPLE 2. *The min-max problem:*

$$\text{(P2)} \qquad \min_x \max_y \left\{ \Phi(x,y) := f(x) - y^T Ax - g(y) \right\},$$

*where* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R}^m \to \mathbb{R}$ *are closed proper convex functions and* $A \in \mathbb{R}^{m\times n}$. *For convenience, we define*

$$\theta(u) := f(x) + g(y),\ u = w := \begin{pmatrix} x \\ y \end{pmatrix},\ F(w) := \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}.$$

*Then according to the optimality condition,* (P2) *is equivalent to find* $w^* = (x^*, y^*)$ *such that*

$$\text{(P2')} \qquad 0 \in T(w^*),\ \text{where } T(w) := \begin{pmatrix} \partial f(x) \\ \partial g(y) \end{pmatrix} + \begin{pmatrix} -A^T y \\ Ax \end{pmatrix}.$$

*Noting that, for a given point* $\hat{w} := \hat{u} = \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix}$, *it holds that,*

$$\theta(\hat{u}) - \theta(u) - (w-\hat{w})^T F(w) = \Phi(\hat{x}, y) - \Phi(x, \hat{y}),$$

*which is exactly the primal dual gap.*

In this paper, we suppose that the optimal solutions set of (P1) or (P2) are nonempty and bounded.

Quite a few well-known algorithms that follow this framework include the augmented Lagrangian method (ALM) [23, 35], the proximal ALM [36, 37], ADMM, the linear ADMM [43], and the strictly contractive Peaceman-Rachford splitting method [11, 13]. These algorithms are designed to solve two-block separable convex optimization problems with equality constraints. For solving multi-block separable convex optimization problems with equality constraints, the Jacobian ALM [10, 42], ADMM

TABLE 1.1
*Algorithms for solving* (P1)

| | Problem | Rate of primal dual gap function values or feasibility measure | Rate in pointwise sense |
|---|---|---|---|
| ADMM | $(\text{P1})_2$ | $O(1/t)$(ergodic)<br>$O(1\sqrt{t})$(non-ergodic) | $O(1/t)$ |
| Li and Lin [26] | $(\text{P1})_2$ | $O(1/t)$(non-ergodic) | - |
| Tran-Dinh and Zhu [39] | $(\text{P1})_2$ | $O(1/t)$(non-ergodic) | - |
| Valkonen [41] | $(\text{P1})_2$ | $O(1/t)$(non-ergodic) | - |
| Sabach and Teboulle [38] | $(\text{P1})_2$ | $O(1/t)$(non-ergodic) | - |
| Luo [30] | $(\text{P1})_2$ | $O(1/t)$(non-ergodic) | - |
| Attouch et al. [1, 2] | $(\text{P1})_2$ | - | $O(1/t^2)$ |
| **This paper** | (P1) | $O(1/t)$(non-ergodic) | $O(1/t^2)$ |

$(\text{P1})_2$ represents (P1) with $m = 2$.

with substitution [16, 17], and the ADMM-type algorithm [18], are all based on the prediction-correction framework. It is interesting to note that the analysis of the divergence of multi-block ADMM [6] is also based on the prediction-correction framework. For solving min-max problems (P2), Chambolle-Pock (CP) algorithm and their variants [4, 5, 12, 14] could fall into the prediction-correction framework.

On the other hand, there are faster algorithms for solving unconstrained or simple-constrained convex optimization problems. Nesterov [34] is the first to present an accelerated gradient method for unconstrained convex optimization with an $O(1/t^2)$ convergence rate. This method has been extended to composite convex optimization problems that involve the simple proximal operator [3, 40] (see [32] for an explanation of second-order differential equations). Attouch [1, 2] introduced a faster algorithm enjoying $O(1/t^2)$ in the pointwise sense for solving monotone inclusions problems by the continuous dynamical system approaches.

Note that, by the prediction-correction framework, He and Yuan [19] give $O(1/t)$ ergodic convergence rate of the primal-dual gap for the classical ADMM. As for the non-ergodic convergence rate, only $O(1/\sqrt{t})$ non-ergodic convergence rate of classical ADMM is given in [28, Section 3.4.5.1]. This is slower than the ergodic case. Actually, as shown in [9], Golowich et al. give the theoretical guarantee that non-ergodic convergence is slower than the ergodic sense when solving saddle point problems. Actually, the non-ergodic convergence is important in theory and also in practice.

For improving the convergence rates in the non-ergodic case when solving linear constraints optimization problems only under convex assumption, Li and Lin [26] and Tran-Dinh and Zhu [39] and Valkonen [41] present accelerated ADMM with $O(1/t)$ non-ergodic convergence rate when solving (P1) with $m = 2$. The same non-ergodic rate result can also be obtained by the continuous dynamical system approaches [30]. By introducing the so called *nice primal algorithmic map*, Sabach and Teboulle [38] present a class of Lagrangian-based methods with $O(1/t)$ non-ergodic convergence rates of both the function values and the feasibility measure when solving (P1) with $m = 1$ or 2. Note that all the papers [26, 30, 38, 39, 41] do not give the $O(1/t^2)$ convergence rate in the pointwise sense. We can refer to Table 1.1 for more details.

Specially, for the special case of (P1) with $m = 1$ and equality constraints, Luo

Table 1.2
*Algorithms for solving* (P2)

| | Condition | Rate of primal dual gap or function values, feasibility measure | Rate in pointwise sense |
|---|---|---|---|
| Chambolle and Pock [4] | $rs > \rho(A^TA)$ | $O(1/t)$(ergodic) | $O(1/t)$ |
| Jiang et al. [25] Li and Yan [27] | $rs > 0.75\rho(A^TA)$ | $O(1/t)$(ergodic) | - |
| He et al. [12] | $rs > 0.75\rho(A^TA)$ | $O(1/t)$(ergodic) | $O(1/t)$ |
| Attouch et al. [1, 2] | $rs > \rho(A^TA)$ | - | $O(1/t^2)$ |
| [26, 38, 39, 41] | $rs > \rho(A^TA)$ | $O(1/t)$(non-ergodic) | - |
| **This paper** | $rs > 0.75\rho(A^TA)$ | $O(1/t)$(non-ergodic) | $O(1/t^2)$ |

[29, 31], He et al. [22] and Boţ et al. [24] give accelerated ALM with $O(1/t^2)$ non-ergodic convergence of the function values and the feasibility measure. Only Boţ et al. [24] give the convergence of the iterates under smooth objective functions. When solving the special case of (P1) with $m = 2$ and equality constraints, even if ADMM-type algorithms have easily solvable subproblems, the subproblems of accelerated ALM given in [22, 29, 31] may not be easily solvable. Without taking into account of the multi-block structure is why the convergence rate is faster than $O(1/t)$. The special case of (P1) with $m = 1$ is not the main work of this paper.

Consider the problem (P2). Chambolle and Pock [4] give easily solvable subproblems algorithm (CP) with $O(1/t)$ ergodic convergence of primal-dual gap under $rs > \rho(A^TA)^\dagger$. Actually, CP can be seen as proximal point algorithm (PPA) for solving (P2). Hence $O(1/t)$ in pointwise sense of iteration can be obtained. Jiang et al. [25] and Li and Yan [27] extend CP with $O(1/t)$ ergodic convergence under $rs > 0.75\rho(A^TA)$ without pointwise convergence rate. Numerically, it indicates that small $rs$ will accelerate convergence when $\rho(A^TA)$ is large. Recently, He et al. [12] give a CP type algorithm for solving (P2) with both $O(1/t)$ in ergodic and pointwise sense under $rs > 0.75\rho(A^TA)$. Howere, we point out that the accelerated algorithms given in [1, 2, 26, 38, 39, 41] for solving (P2) is under the assumption $rs > \rho(A^TA)$. For more details, we can refer to Table 1.2.

Considering that He and Yuan's prediction-correction framework only enjoys $O(1/t)$ convergence rate in ergodic sense and pointwise sense, and there is no unify framework with non-ergodic convergence rate when solving (P1) and (P2), it is necessary to establish a framework with faster convergence rate (non-ergodic sense and pointwise sense).

**Contributions** We list in the following the contributions of this paper:

1. We present a faster prediction-correction framework for solving (P1) and (P2). Different from the algorithms which rely on *nice primal algorithmic map* in [38] or the continuous dynamical system approaches in [1, 2, 30], our ingredient is the prediction-correction framework in [19] which enjoys $O(1/t)$ in ergodic sense of the primal-dual gap and $O(1/t)$ in the pointwise sense. It is worth noting that our framework provides algorithms at a rate of $O(1/t)$ in non-ergodic sense of primal-dual gap and $O(1/t^2)$ in the pointwise sense.

$^\dagger\rho(A^TA)$ represents the spectrum of $A^TA$. $1/r$ and $1/s$ actually serve as the stepsizes in iteration algorithm.

2. He et al. [11, 13] propose a ADMM-type algorithm with dual variable updating twice for better numerical performance when solving (P1) with $m = 2$ and equality constraints (or see Remark 2). Based on the proposed faster prediction-correction framework, we give a faster ADMM algorithm with dual variable updating twice for solving this problems in Section 4.1. This is the first paper that give algorithm updates dual variable twice with non-ergodic convergence.
3. Based on the proposed faster prediction-correction framework, we give a faster ADMM-type algorithm in Gauss-Seidel way for solving (P1) in Section 4.2. The existing accelerated algorithms with non-ergodic convergence rates solve separable convex optimization problems with two-block equality constraints, for example, [26, 30, 38, 39, 41]. Considering multi-block ADMM is divergent as shown in [6]. It seems that expanding the accelerated algorithms in [26, 30, 38, 39, 41] to multi-block cases in Gauss-Seidel way is not easy. This is the first paper to give a faster algorithm in Gauss-Seidel way with non-ergodic convergence rate for solving multi-block structure convex optimization problems with equality constraints. We can refer to Table 1.1 for detailed comparison for solving (P1).
4. Based on the proposed faster prediction-correction framework, we give a faster CP type algorithm for solving (P2) in Section 4.3. The faster CP type algorithm enjoys faster convergence such as $O(1/t)$ in non-ergodic sense, and $O(1/t^2)$ the pointwise sense compared to the non-accelerated algorithms given in [12, 25, 27]. The convergence rate is established under the condition $rs > 0.75\rho(A^TA)$ compared to the accelerated algorithms [1, 2, 26, 38, 39, 41] with the condition $rs > \rho(A^TA)$. We can refer to Table 1.2 for detailed comparison for solving (P2).

**2. He and Yuan's prediction-correction framework.** With $\theta(u)$, $u$, $w$, $F(w)$ and $T(w)$ defined in Example 1 or 2, the fundmental algorithm for solving (P1') and (P2') (or (P1) and (P2)) is the proximal point algorithm (PPA), which was originally introduced by Martinet [33], reads as:

**PPA.** With a given $w^k$, find $w^{k+1}$ such that

$$0 \in T(w^{k+1}) + Q(w^{k+1} - w^k), \tag{2.1}$$

where $Q$ is a symmetric defined matrix.

For general matrix $Q$, the subproblem of PPA may not be easily solved. Hence, an additional algorithm to solve the subproblem of PPA is necessary. Consider the separable structure of (P1') and (P2'), we can solve each block of (P1') or (P2') in Jacobi or Gauss-Seidel way in order to reduce computation cost in every iteration (such as ADMM).

Thus, we consider the following prediction-correction type PPA with special selection of the scaled matrixs. Noting that, by special selection of the scaled matrixs, many famous algorithms fall into the following prediction-correction framework, including the augmented Lagrangian method (ALM) [23, 35], the proximal ALM [36, 37], ADMM, the linear ADMM [43], and the strictly contractive Peaceman-Rachford splitting method [11, 13], the Jacobian ALM [10, 42], ADMM with substitution [16, 17] for solving multi-block of (P1), and Chambolle-Pock (CP) algorithm and their variants

[4, 5, 12, 14].

**[Prediction step.]** With a given $w^k$, find $\widetilde{w}^k$ such that

$$0 \in T(\widetilde{w}^k) + L^T QL(\widetilde{w}^k - w^k), \tag{PS}$$

where $Q^T + Q \succ 0$ (noting that $Q$ is not necessarily symmetric), $L$ is a matrix.
**[Correction step.]** Update $v^{k+1} = Lw^{k+1}$ by

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

where $\widetilde{v}^k = L\widetilde{w}^k$.

REMARK 1. *Three algorithms satisfying* (PS)-(CS) *are given in Section 2.1-2.3. The definitions of L, Q and M for different algorithms are also given.*

Then there exists $\widetilde{g}^k \in (T - F)(\widetilde{w}^k)$ such that

$$\begin{aligned} &(w - \widetilde{w}^k)^T(\widetilde{g}^k + F(\widetilde{w}^k) + L^T QL(\widetilde{w}^k - w^k) = 0, \ \forall w \\ \Longrightarrow &\theta(u) - \theta(\widetilde{u}^k) + (w - \widetilde{w}^k)^T F(\widetilde{w}^k) \geq (w - \widetilde{w}^k)^T L^T QL(w^k - \widetilde{w}^k), \ \forall w, \end{aligned} \tag{2.2}$$

where the inequality using convexity of $\theta$. Hence prediction-correction framework (PS)-(CS) infers the following framework (in variational form) due to He and Yuan [19]. It is fundamental in providing convergent algorithms for solving (P1) and (P2). We can also refer to [21] for a detailed understanding.

**[Prediction step.]** With a given $w^k$, find $\widetilde{w}^k$ such that

$$\theta(u) - \theta(\widetilde{u}^k) + (w - \widetilde{w}^k)^T F(\widetilde{w}^k) \geq (w - \widetilde{w}^k)^T L^T QL(w^k - \widetilde{w}^k), \ \forall w, \tag{2.3}$$

where $Q^T + Q \succ 0$ (noting that $Q$ is not necessarily symmetric) and $L$ is a matrix.
**[Correction step.]** Update $v^{k+1} = Lw^{k+1}$ by

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

where $\widetilde{v}^k = L\widetilde{w}^k$.

To ensure the convergence of algorithms satisfying the above framework, He and Yuan [21, 19] add some assumptions on the selection of the matrices $Q$ and $M$. The following is such a commonly used condition.

**[Convergence Condition.]** For the given matrice $Q$ and nonsingular matrice $M$, setting

$$H := QM^{-1} \succ 0, \tag{CC1}$$

$$G := Q^T + Q - M^T HM \succ 0. \tag{CC2}$$

The convergence rates of prediction-correction framework (2.3)-(2.4) are established as follows.

THEOREM 2.1 ($O(1/t)$ ergodic convergence rate). ([19, 21]) *Let* $\{\widetilde{w}^k\}$ *be generated by prediction-correction framework* (2.3)-(2.4) *under* (CC1)-(CC2). *Then we have*

$t = 1, 2, \dots,$

$$\theta(\bar{u}^t) - \theta(u) + (\bar{w}^t - w)^T F(w) \leq \frac{1}{2(t+1)} \|v^0 - Lw\|_H^2, \ \forall w,$$

*where* $\bar{w}^t = \frac{1}{t+1} \sum_{k=0}^{t} \widetilde{w}^k$.

THEOREM 2.2 ($O(1/t)$ in the pointwise sense). ([11, 12, 13, 15, 20, 21]) *Let* $\{\widetilde{w}^k\}$ *be generated by prediction-correction framework* (2.3)-(2.4) *under* (CC1)-(CC2)*. Then we have*

$$\|M(v^t - \widetilde{v}^t)\|_H^2 \leq O(1/t), \ t = 1, 2, \dots.$$

In the following, we give some algorithms that satisfying (PS)-(CS) for solving (P1) and (P2) for better understanding of prediction-correction framework (PS)-(CS).

**2.1. Algorithm satisfying** (PS)**-**(CS) **for solving** (P1) **with** $m = 2$**.** Consider (P1) with $m = 2$, i.e.,

$$\min_{x_1, x_2} \{f_1(x_1) + f_2(x_2): \ A_1x_1 + A_2x_2 = b\}. \tag{2.5}$$

The strictly contractive Peaceman-Rachford splitting method [11, 13] for solving (2.5) is given by:

(2.6)

$$\begin{cases} x_1^{k+1} \in \arg\min\limits_{x_1} \left\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2}\|A_1x_1 + A_2x_2^k - b\|^2 + \frac{1}{2}\|x_1 - x_1^k\|_P^2 \right\}, \\ \lambda^{k+\frac{1}{2}} = \lambda^k - r\beta(A_1x_1^{k+1} + A_2x_2^k - b), \\ x_2^{k+1} \in \arg\min\limits_{x_2} \left\{ f_2(x_2) - x_2^T A_2^T \lambda^{k+\frac{1}{2}} + \frac{\beta}{2}\|A_1x_1^{k+1} + A_2x_2 - b\|^2 \right\}, \\ \lambda^{k+1} = \lambda^{k+\frac{1}{2}} - s\beta(A_1x_1^{k+1} + A_2x_2^{k+1} - b), \ P \succ 0, \ \beta > 0. \end{cases}$$

REMARK 2. *Algorithm* (2.6) *reduces to ADMM when* $r = 0$ *and* $P = 0$*. Different from ADMM, Algorithm* (2.6) *updates* $\lambda$ *twice and convergence is established under weak conditions in* (2.8)*. As shown in [11, 13], Algorithm* (2.6) *enjoys better numerical performance compared to ADMM. Algorithm* (2.6) *with* $r = s = 1$ *and* $\alpha = 0$ *reduces to Peaceman-Rachford splitting method (PRSM) for solving the dual of* (2.5)*. Howere, the convergence of PRSM is established under the strongly convex assumption. In order to remove the strongly convex assumption, He et al. [11, 13] propose some conditions on* $r$ *and* $s$ *to ensure convergence (or see* (2.8)*).*

*We can set* $P = \alpha I_{n_1} - \beta A_1^T A_1$ *with* $\alpha > \|A_1\|^2$ *in order to enjoy easily solved subproblems.*

For convenience, we define

(2.7)

$$L = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & 0 & I_l \end{pmatrix}, \ Q = \begin{pmatrix} P & 0 & 0 \\ 0 & \beta A_2^T A_2 & -rA_2^T \\ 0 & -A_2 & \frac{1}{\beta} I_l \end{pmatrix}, \ M = \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -s\beta A_2 & (r+s)I_l \end{pmatrix}.$$

THEOREM 2.3. *For* $L$, $Q$, $M$ *defined in* (2.7) *and* $\theta(u)$, $u$, $w$, $F(w)$, $T(w)$ *defined in Example 1 with* $m = 2$*, it holds that:*

(1) *Algorithm* (2.6) *satisfies* (PS)-(CS)*.*
(2) $H$ *and* $G$ *satisfying* (CC1)-(CC2) *are positive definite if* $A_2$ *is full column rank,*

$$r \in (-1, 1), \ s \in (0, 1) \text{ and } r + s > 0. \tag{2.8}$$

*Proof.* Setting

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

in Algorithm (2.6).
Proof of (1). The optimality condition of the $x_1$-subproblem reads as:

$$\begin{aligned} 0 &\in \partial f_1(\widetilde{x}_1^k) - A_1^T[\lambda^k - \beta(A_1\widetilde{x}_1^k + A_2x_2^k - b)] + \alpha(\widetilde{x}_1^k - x_1^k) \\ &= \partial f_1(\widetilde{x}_1^k) - A_1^T\widetilde{\lambda}^k + P(\widetilde{x}_1^k - x_1^k). \end{aligned}$$

The optimality condition of the $x_2$-subproblem reads as:

$$\begin{aligned} 0 &\in \partial f_2(\widetilde{x}_2^k) - A_2^T[\lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2x_2^k - b)] + \beta A_2^T[A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b] \\ &= \partial f_2(\widetilde{x}_2^k) - A_2^T\widetilde{\lambda}^k + \beta A_2^T A_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k). \end{aligned}$$

The definition of $\widetilde{\lambda}^k$ can be rewritten as:

$$0 = A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b - A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k).$$

Combining the above three relations, we obtain

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P(\widetilde{x}_1^k - x_1^k) \\ \beta A_2^T A_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k) \\ -A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},$$

which is equivalent to

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P & 0 & 0 \\ 0 & \beta A_2^T A_2 & -rA_2^T \\ 0 & -A_2 & \frac{1}{\beta}I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}.$$

Then prediction step holds. It holds that

$$\begin{aligned} \lambda^{k+1} =& \lambda^{k+\frac{1}{2}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b) \\ =& \lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2x_2^k - b) - s\beta(A_1\widetilde{x}_1^k + A_2x_2^k - b) + s\beta A_2(x_2^k - \widetilde{x}_2^k) \\ =& \lambda^k - r(\lambda^k - \widetilde{\lambda}^k) - s(\lambda^k - \widetilde{\lambda}^k) + s\beta A_2(x_2^k - \widetilde{x}_2^k). \end{aligned} \tag{2.9}$$

Thus, together with $\widetilde{x}_1^k = x_1^{k+1}$ and $\widetilde{x}_2^k = x_2^{k+1}$, we have

$$\begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^k \\ x_2^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -s\beta A_2 & (r+s)I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}.$$

The correction step holds.
Proof of (2). For the positive definiteness of $H$ and $G$ satisfying (CC1)-(CC2), we can refer to [13, Lemma 4.1] and [13, Page 1480] for a similar proof. □

**2.2. Algorithm satisfying** (PS)-(CS) **for solving** (P1)**.** The following algorithm for solving (P1) was first presented by He et al. [18].

**[Prediction step.]** With given $\beta > 0$ and $(x_1^k, x_2^k, \dots, x_m^k, \lambda^k)$, find $(\widetilde{x}_1^k, \dots, \widetilde{x}_m^k, \widetilde{\lambda}^k)$ by

(2.10)
$$\begin{cases} \widetilde{x}_1^k = \arg\min_{x_1} \left\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta}{2} \|A_1(x_1 - x_1^k)\|^2 \right\}, \\ \widetilde{x}_2^k = \arg\min_{x_2} \left\{ f_2(x_2) - x_2^T A_2^T \lambda^k + \frac{\beta}{2} \|A_1(\widetilde{x}_1^k - x_1^k) + A_2(x_2 - x_2^k)\|^2 \right\}, \\ \quad \vdots \\ \widetilde{x}_i^k = \arg\min_{x_i} \left\{ f_i(x_i) - x_i^T A_2^T \lambda^k + \frac{\beta}{2} \| \sum_{j=1}^{i-1} A_j(\widetilde{x}_j^k - x_j^k) + A_i(x_i - x_i^k)\|^2 \right\}, \\ \quad \vdots \\ \widetilde{x}_m^k = \arg\min_{x_m} \left\{ f_m(x_m) - x_m^T A_2^T \lambda^k + \frac{\beta}{2} \| \sum_{j=1}^{m-1} A_j(\widetilde{x}_j^k - x_j^k) + A_m(x_m - x_m^k)\|^2 \right\} \\ \widetilde{\lambda}^k = \arg\max_{\lambda} \left\{ -\lambda^T (\sum_{j=1}^m A_j \widetilde{x}_j^k - b) - \frac{1}{2\beta} \|\lambda - \lambda^k\|^2 \right\}. \end{cases}$$

**[Correction step.]**Update $(A_1 x_1^{k+1}, A_2 x_2^{k+1}, \dots, A_m x_m^{k+1}, \lambda^{k+1})$ by

(2.11)
$$\begin{pmatrix} \sqrt{\beta} A_1 x_1^{k+1} \\ \sqrt{\beta} A_2 x_2^{k+1} \\ \vdots \\ \sqrt{\beta} A_m x_m^{k+1} \\ \frac{1}{\sqrt{\beta}} \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} A_1 x_1^{k} \\ \sqrt{\beta} A_2 x_2^{k} \\ \vdots \\ \sqrt{\beta} A_m x_m^{k} \\ \frac{1}{\sqrt{\beta}} \lambda^{k} \end{pmatrix} - \begin{pmatrix} \alpha I_l & -\alpha I_l & 0 & \dots & 0 \\ 0 & \alpha I_l & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha I_l & 0 \\ 0 & \dots & 0 & \alpha I_l & 0 \\ -\alpha I_l & 0 & \dots & 0 & I_l \end{pmatrix} \begin{pmatrix} \sqrt{\beta}(A_1 x_1^k - A_1 \widetilde{x}_1^k) \\ \sqrt{\beta}(A_2 x_2^k - A_2 \widetilde{x}_2^k) \\ \vdots \\ \sqrt{\beta}(A_m x_m^k - A_m \widetilde{x}_m^k) \\ \frac{1}{\sqrt{\beta}}(\lambda^k - \widetilde{\lambda}^k) \end{pmatrix}.$$

For convenience, we define

(2.12)
$$L = \mathrm{Diag}\left( \sqrt{\beta} A_1, \quad \sqrt{\beta} A_2, \quad \dots, \sqrt{\beta} A_m, \quad \frac{1}{\sqrt{\beta}} I_l \right),$$
$$Q = \begin{pmatrix} I_l & 0 & \dots & 0 & I_l \\ I_l & I_l & \ddots & \vdots & I_l \\ \vdots & \vdots & \ddots & 0 & \vdots \\ I_l & I_l & \dots & I_l & I_l \\ 0 & 0 & \dots & 0 & I_l \end{pmatrix}, \; M = \begin{pmatrix} \alpha I_l & -\alpha I_l & 0 & \dots & 0 \\ 0 & \alpha I_l & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha I_l & 0 \\ 0 & \dots & 0 & \alpha I_l & 0 \\ -\alpha I_l & 0 & \dots & 0 & I_l \end{pmatrix}.$$

THEOREM 2.4. *For $L$, $Q$, $M$ defined in* (2.12) *and $\theta(u)$, $u$, $w$, $F(w)$, $T(w)$ defined in Example 1, it holds that*

(1) *Algorithm* (2.10)-(2.11) *satisfies* (PS)-(CS).

(2) *$H$ and $G$ satisfying* (CC1)-(CC2) *are positive definite if $\alpha \in (0, 1)$.*

*Proof.* Proof of (1). For $i = 1, 2, \dots, m$, the optimality condition of the $x_i$-subproblem is given by

$$\begin{aligned} 0 &\in \partial f_i(\widetilde{x}_i^k) - A_i^T \lambda^k + \beta A_i^T \sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) \\ &= \partial f_i(\widetilde{x}_i^k) - A_i^T \widetilde{\lambda}^k + \beta A_i^T \sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) + A_i^T(\widetilde{\lambda}^k - \lambda^k). \end{aligned}$$

The optimality condition of the $\lambda$-subproblem is given by

$$0 = (\sum_{j=1}^m A_j \widetilde{x}_j^k - b) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k).$$

Combining the above two relations, we obtain

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ \vdots \\ \partial f_m(\widetilde{x}_m^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + \begin{pmatrix} \beta A_1^T A_1(\widetilde{x}_1^k - x_1^k) + A_1^T(\widetilde{\lambda}^k - \lambda^k) \\ \beta A_2^T \sum_{j=1}^2 A_j(\widetilde{x}_j^k - x_j^k) + A_2^T(\widetilde{\lambda}^k - \lambda^k) \\ \vdots \\ \beta A_m^T \sum_{j=1}^m A_j(\widetilde{x}_j^k - x_j^k) + A_m^T(\widetilde{\lambda}^k - \lambda^k) \\ \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},$$

which is equivalent to

$$0 \in \begin{pmatrix} \partial f_1(\widetilde{x}_1^k) \\ \partial f_2(\widetilde{x}_2^k) \\ \vdots \\ \partial f_m(\widetilde{x}_m^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + L^T \begin{pmatrix} I_l & 0 & \dots & 0 & I_l \\ I_l & I_l & \ddots & \vdots & I_l \\ \vdots & \vdots & \ddots & 0 & \vdots \\ I_l & I_l & \dots & I_l & I_l \\ 0 & 0 & \dots & 0 & I_l \end{pmatrix} L \begin{pmatrix} \widetilde{x}_1^k - x_1^k \\ \widetilde{x}_2^k - x_2^k \\ \vdots \\ \widetilde{x}_m^k - x_m^k \\ \widetilde{\lambda}^k - \lambda^k \end{pmatrix}.$$

Then the prediction step holds. The correction step is easy to verified.
Proof of (2). We can refer to [18, Lemma 7.1, 7.2]. □

**2.3. Algorithm satisfying** (PS)-(CS) **for solving** (P2)**.** The following algorithm is presented by He et al. [12] for solving (P2):

**[Prediction step.]** With given $\beta > 0$ and $(x^k, y^k)$, find $(\widetilde{x}^k, \widetilde{y}^k)$ by

$$\begin{cases} \widetilde{x}^k = \arg\min_x \left\{\Phi(x, y^k) + \frac{r}{2}\|x - x^k\|^2\right\}, \\ \widetilde{y}^k = \arg\max_y \left\{\Phi([\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)], y) - \frac{s}{2}\|y - y^k\|^2\right\}. \end{cases} \tag{2.13}$$

**[Correction step.]** Update $(x^{k+1}, y^{k+1})$ by

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \begin{pmatrix} I_n & 0 \\ -(1-\alpha)\frac{1}{s}A & I_m \end{pmatrix} \begin{pmatrix} x^k - \widetilde{x}^k \\ y^k - \widetilde{y}^k \end{pmatrix}. \tag{2.14}$$

For convenience, we define

$$L = \begin{pmatrix} I_n & 0 \\ 0 & I_m \end{pmatrix}, \quad Q = \begin{pmatrix} rI_n & A^T \\ \alpha A & sI_m \end{pmatrix} \text{ and } M = \begin{pmatrix} I_n & 0 \\ -(1-\alpha)\frac{1}{s}A & I_m \end{pmatrix}. \tag{2.15}$$

THEOREM 2.5. *For $L$, $Q$, $M$ defined in* (2.15) *and $\theta(u)$, $u$, $w$, $F(w)$, $T(w)$ as defined in Example 2, it holds that*

(1) *Algorithm* (2.13)-(2.14) *satisfies* (PS)-(CS).
(2) *$H$ and $G$ satisfying* (CC1)-(CC2) *are positive definite if*

$$rs > (1 - \alpha + \alpha^2)\rho(A^T A), \ \alpha \in [0, 1].$$

*Proof.* Proof of (1). The optimality condition of the $x$-subproblem reads as:

$$\begin{aligned} 0 &\in \partial f(\widetilde{x}^k) - A^T y^k + r(\widetilde{x}^k - x^k) \\ &= \partial f(\widetilde{x}^k) - A^T\widetilde{y}^k + r(\widetilde{x}^k - x^k) + A^T(\widetilde{y}^k - y^k). \end{aligned}$$

The optimality condition of the $y$-subproblem reads as:

$$0 \in \partial g(\widetilde{y}^k) + A[\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)] + s(\widetilde{y}^k - y^k).$$

Combining the above two relations together yields that

$$0 \in \begin{pmatrix} \partial f(\widetilde{x}^k) \\ \partial g(\widetilde{y}^k) \end{pmatrix} + \begin{pmatrix} -A^T\widetilde{y}^k \\ A\widetilde{x}^k \end{pmatrix} + \begin{pmatrix} r(\widetilde{x}^k - x^k) + A^T(\widetilde{y}^k - y^k) \\ \alpha A(\widetilde{x}^k - x^k) + s(\widetilde{y}^k - y^k) \end{pmatrix}.$$

This is equivalent to:

$$0 \in \begin{pmatrix} \partial f(\widetilde{x}^k) \\ \partial g(\widetilde{y}^k) \end{pmatrix} + \begin{pmatrix} -A^T\widetilde{y}^k \\ A\widetilde{x}^k \end{pmatrix} + \begin{pmatrix} rI_n & A^T \\ \alpha A & sI_m \end{pmatrix} \begin{pmatrix} \widetilde{x}^k - x^k \\ \widetilde{y}^k - y^k \end{pmatrix}.$$

Then we obtain the prediction step (4.11) satisfy (FPS) with $L$ and $Q$ defined in (2.15). The correction step is easy to verified.

Proof of (2). We can refer to [12, Proposition 4.1]. □

REMARK 3. *When $\alpha = 1$, the algorithm* (2.13)-(2.14) *simplifies to the CP algorithm introduced in [4]. Convergence is established under the condition $rs > \rho(A^TA)$. By setting $\alpha = \frac{1}{2}$ in algorithm* (2.13)-(2.14)*, the convergence condition is relaxed to $rs > 0.75\rho(A^TA)$. This extension broadens the permissible step size compared to the original CP algorithm.*

## 3. Faster prediction-correction framework.

We first present the following new prediction-correction framework with $\theta(u)$, $u$, $w$, $F(w)$ and $T(w)$ defined in Example 1 or 2, and then establish the convergence rates.

**[Prediction step.]** With given $w^k$ and $\breve{w}^{k-1}$, find $\breve{w}^k$ such that

$$0 \in (T - F)(\breve{w}^k) + F(\widetilde{w}^k) + L^TQL(\widetilde{w}^k - w^k), \tag{FPS}$$

where $Q^T + Q \succ 0$ (noting that $Q$ is not necessarily symmetric), $L$ is a matrix and

$$\widetilde{w}^k = \frac{1}{\tau^k}\breve{w}^k - \frac{1-\tau^k}{\tau^k}\breve{w}^{k-1}, \tag{3.1}$$

and the sequence $\{\tau^k\}$ satisfies the following equality:

$$1/\tau^{k-1} = (1-\tau^k)/\tau^k, \ \tau^{-1} \in (0, 1). \tag{Y}$$

**[Correction step.]** Update $v^{k+1} = Lw^{k+1}$ by

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

where $\widetilde{v}^k = L\widetilde{w}^k$.

According to (FPS), there exists $\breve{g}^k \in (T - F)(\breve{w}^k)$ such that

$$\begin{aligned} &(w - \breve{w}^k)^T(\breve{g}^k + F(\widetilde{w}^k) + L^TQL(\widetilde{w}^k - w^k) = 0, \ \forall w \\ \Longrightarrow &\theta(u) - \theta(\breve{u}^k) + (w - \breve{w}^k)^TF(\widetilde{w}^k) \geq (w - \breve{w}^k)^TL^TQL(w^k - \widetilde{w}^k), \ \forall w, \end{aligned} \tag{3.2}$$

where the inequality using convexity of $\theta$. Hence prediction-correction framework (FPS)-(FCS) infers the following framework (in variational form).

**[Prediction step.]** With given $w^k$ and $\breve{w}^{k-1}$, find $\breve{w}^k$ such that

$$\theta(u) - \theta(\breve{u}^k) + (w - \breve{w}^k)^T F(\widetilde{w}^k) \geq (w - \breve{w}^k)^T L^T Q L (w^k - \widetilde{w}^k), \ \forall w, \tag{3.3}$$

where $Q^T + Q \succ 0$ (noting that $Q$ is not necessarily symmetric), $L$ is a matrix and $\widetilde{w}^k$ defined in (3.1) and $\tau^k$ satisfying (Y).
**[Correction step.]** Update $v^{k+1} = Lw^{k+1}$ by

$$v^{k+1} = v^k - M(v^k - \widetilde{v}^k), \tag{3.4}$$

where $\widetilde{v}^k = L\widetilde{w}^k$.

The following property of the sequence $\{\tau^k\}$ satisfying (Y) is trivial to verify and hence omitted.

LEMMA 3.1. *Let* $\{\tau^k\}$ *satisfy* (Y). *Then* $\tau^k = \frac{1}{\tau^{-1}+k+1}$.

**3.1. $O(1/t)$ non-ergodic convergence rate.** We establish $O(1/t)$ non-ergodic convergence rate of the primal dual gap for the faster prediction-correction framework (FPS)-(FCS) under the conditions (CC1)-(CC2).

LEMMA 3.2. *For the faster prediction-correction framework* (FPS)-(FCS) *under* (CC1)-(CC2), *we have*

$$\begin{aligned} &\frac{1}{\tau^k}[\theta(u) - \theta(\breve{u}^k)] - \frac{1}{\tau^{k-1}}[\theta(u) - \theta(\breve{u}^{k-1})] + (w - \widetilde{w}^k)^T F(\widetilde{w}^k) \\ \geq &\frac{1}{2}\left(\|v^{k+1} - v\|_H^2 - \|v^k - v\|_H^2 + \|v^k - \widetilde{v}^k\|_G^2\right), \ \forall w, \end{aligned} \tag{3.5}$$

*where* $v = Lw$.

*Proof.* Prediction step (FPS) means prediction step (3.3) holds. Multiplying both sides of (3.3) by $(1 - \tau^k)/\tau^k$ with $u = \breve{u}^{k-1}$ and $w = \breve{w}^{k-1}$, and then adding it to (3.3) yields that

$$\begin{aligned} &\frac{1}{\tau^k}[\theta(u) - \theta(\breve{u}^k)] - \frac{1 - \tau^k}{\tau^k}[\theta(u) - \theta(\breve{u}^{k-1})] + (w - \widetilde{w}^k)^T F(\widetilde{w}^k) \\ \overset{(Y)}{=} &\frac{1}{\tau^k}[\theta(u) - \theta(\breve{u}^k)] - \frac{1}{\tau^{k-1}}[\theta(u) - \theta(\breve{u}^{k-1})] + (w - \widetilde{w}^k)^T F(\widetilde{w}^k) \\ \geq &(w - \widetilde{w}^k)^T L^T Q L (w^k - \widetilde{w}^k) = (v - \widetilde{v}^k)^T Q (v^k - \widetilde{v}^k). \end{aligned} \tag{3.6}$$

On the other hand, we can verify that

$$\begin{aligned} &(v - \widetilde{v}^k)^T Q (v^k - \widetilde{v}^k) \overset{(FCS)}{=} (v - \widetilde{v}^k)^T H (v^k - v^{k+1}) \\ = &\frac{1}{2}\left(\|v^{k+1} - v\|_H^2 - \|v^k - v\|_H^2 + \|v^k - \widetilde{v}^k\|_H^2 - \|v^{k+1} - \widetilde{v}^k\|_H^2\right) \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} & \|v^k - \widetilde{v}^k\|_H^2 - \|v^{k+1} - \widetilde{v}^k\|_H^2 \\ = \ & \|v^k - \widetilde{v}^k\|_H^2 - \|(v^k - \widetilde{v}^k) - (v^k - v^{k+1})\|_H^2 \\ \overset{(FCS)}{=} & \|v^k - \widetilde{v}^k\|_H^2 - \|(v^k - \widetilde{v}^k) - M(v^k - \widetilde{v}^k)\|_H^2 \\ = \ & (v^k - \widetilde{v}^k)^T (2HM - M^T H M)(v^k - \widetilde{v}^k) \\ = \ & (v^k - \widetilde{v}^k)^T (Q^T + Q - M^T H M)(v^k - \widetilde{v}^k) \\ = \ & \|v^k - \widetilde{v}^k\|_G^2. \end{aligned} \tag{3.8}$$

Combining (3.6), (3.7) and (3.8) completes the proof. ∎

LEMMA 3.3. *For the faster prediction-correction framework* (FPS)-(FCS) *under* (CC1)-(CC2)*, we have*

$$\begin{aligned}&\frac{1}{\tau^k}[\theta(u)-\theta(\breve{u}^k)+(w-\breve{w}^k)^TF(w)]\\&-\frac{1}{\tau^{k-1}}[\theta(u)-\theta(\breve{u}^{k-1})+(w-\breve{w}^{k-1})^TF(w)]\\\geq&\frac{1}{2}\left(\|v^{k+1}-v\|_H^2-\|v^k-v\|_H^2+\|v^k-\widetilde{v}^k\|_G^2\right),\ \forall w,\end{aligned}\tag{3.9}$$

*where* $v = Lw$.

*Proof.* Based on the definition of $F(w)$ in Example 1 and 2, we have

$$(w-w')^T(F(w)-F(w'))=0,\ \forall w,w'.$$

Then we obtain

$$\begin{aligned}&(w-\widetilde{w}^k)^TF(\widetilde{w}^k)=(w-\widetilde{w}^k)^TF(w)\\=&\ \frac{1}{\tau^k}(w-\breve{w}^k)^TF(w)-\frac{1-\tau^k}{\tau^k}(w-\breve{w}^{k-1})^TF(w)\\\overset{(Y)}{=}&\ \frac{1}{\tau^k}(w-\breve{w}^k)^TF(w)-\frac{1}{\tau^{k-1}}(w-\breve{w}^{k-1})^TF(w).\end{aligned}\tag{3.10}$$

Combining Lemma 3.2 we obtain the conclusion. ∎

Then we obtain the following conclusion immediately.

THEOREM 3.4 ($O(1/t)$ non-ergodic convergence rate). *Let* $\{\breve{w}^t\}$ *be generated by* (FPS)-(FCS) *under* (CC1)-(CC2)*. Then we have*

$$\begin{aligned}&\theta(\breve{u}^t)-\theta(u)-(w-\breve{w}^t)^TF(w)\\\leq&\frac{1}{\tau^{-1}+t+1}\left\{\frac{1}{\tau^0}[\theta(\breve{u}^0)-\theta(u)-(w-\breve{w}^0)^TF(w)]+\frac{1}{2}\|v^1-v\|_H^2\right\},\ t=1,2,\ldots,\ \forall w,\end{aligned}$$

*where* $v = Lw$.

*Proof.* Adding Lemma 3.3 from $k = 1$ to $k = t$, then

$$\begin{aligned}&\frac{1}{\tau^t}[\theta(\breve{u}^t)-\theta(u)-(w-\breve{w}^t)^TF(w)]+\frac{1}{2}\|v^{t+1}-v\|_H^2\\\leq&\frac{1}{\tau^0}[\theta(\breve{u}^0)-\theta(u)-(w-\breve{w}^0)^TF(w)]+\frac{1}{2}\|v^1-v\|_H^2,\ t=1,2,\ldots,\ \forall w.\end{aligned}$$

Based on Lemma 3.1, we obtain the conclusion. ∎

**3.2. $O(1/t^2)$ convergence rate in the pointwise sense.** We establish $O(1/t^2)$ convergence rate in pointwise sense for the faster prediction-correction framework (FPS)-(FCS) under the conditions (CC1)-(CC2).

THEOREM 3.5 ($O(1/t^2)$ in the pointwise sense). *For the faster prediction-correction framework* (FPS)-(FCS) *under* (CC1)-(CC2)*, we have*

$$\|M(\breve{v}^t-\breve{v}^{t-1})\|_H^2\leq O(1/t^2),\ t=1,2,\ldots.$$

*Proof.* Suppose $w^*$ is a saddle point of (P1) or (P2) and $v^* = Lw^*$. Based on the correction step (FCS), we have

$$\begin{aligned} \|v^{k+1} - v^*\|_H^2 =& \|(I-M)(v^k - v^*) + M(\widetilde{v}^k - v^*)\|_H^2 \\ =& \underbrace{\|(I-M)(v^k - v^*)\|_H^2}_{:=A^k} + \underbrace{\|M(\widetilde{v}^k - v^*)\|_H^2}_{:=B^k} \\ &+ 2\underbrace{(v^k - v^*)^T(I-M)^T HM(\widetilde{v}^k - v^*)}_{:=C^k}. \end{aligned} \tag{3.11}$$

According to the definition of $\widetilde{v}^k$, we obtain

$$\begin{aligned} &\|M(\widetilde{v}^k - v^*)\|_H^2 + \frac{1-\tau^k}{\tau^k}\underbrace{\|M(\breve{v}^{k-1} - v^*)\|_H^2}_{:=D^{k-1}} \\ =& \frac{1-\tau^k}{(\tau^k)^2}\underbrace{\|M(\breve{v}^k - \breve{v}^{k-1})\|_H^2}_{:=E^k} + \frac{1}{\tau^k}\underbrace{\|M(\breve{v}^k - v^*)\|_H^2}_{D^k}. \end{aligned} \tag{3.12}$$

Setting $u = u^*$, $w = w^*$ and $v = v^*$ in Lemma 3.3, it follows from (3.11) and (3.12) that

$$\begin{aligned} &\|v^{k+1} - v^*\|_H^2 - \|v^k - v^*\|_H^2 = A^k - A^{k-1} + B^k - B^{k-1} + 2(C^k - C^{k-1}) \\ =& A^k - A^{k-1} + 2(C^k - C^{k-1}) + \frac{1-\tau^k}{(\tau^k)^2}E^k - \frac{1-\tau^{k-1}}{(\tau^{k-1})^2}E^{k-1} \\ &+ \frac{1}{\tau^k}(D^k - D^{k-1}) - \frac{1}{\tau^{k-1}}(D^{k-1} - D^{k-2}) + D^{k-1} - D^{k-2} \\ \le& -\frac{2}{\tau^t}S^t + \frac{2}{\tau^{k-1}}S^{k-1}, \end{aligned} \tag{3.13}$$

where $S^t = \theta(\breve{u}^t) - \theta(u^*) + (\breve{w}^t - w^*)^T F(w^*) \ge 0$. For each $t = 1, 2, \ldots$, summing up both sides of (3.13) from $k = 1$ to $t$ yields that

$$\begin{aligned} &\sum_{k=1}^{t}\left(\|v^{k+1} - v^*\|_H^2 - \|v^k - v^*\|_H^2\right) \\ =& A^t - A^0 + 2(C^t - C^0) + \frac{1-\tau^t}{(\tau^t)^2}E^t - \frac{1-\tau^0}{(\tau^0)^2}E^0 \\ &+ \frac{1}{\tau^t}(D^t - D^{t-1}) - \frac{1}{\tau^0}(D^0 - D^{-1}) + D^{t-1} - D^{-1} \\ \le& -\frac{2}{\tau^t}S^t + \frac{2}{\tau^0}S^0, \end{aligned} \tag{3.14}$$

According to (3.13), $v^k$ is bounded. Therefore, the correction step (FCS) implies that $\widetilde{v}^k$ is also bounded. Then $C^t$ is bounded by its definition. It follows from (3.14) that there is a positive bound $N_1 < \infty$ such that

$$\frac{2}{\tau^t}S^t + A^t + \frac{1-\tau^t}{(\tau^t)^2}E^t + \frac{1}{\tau^t}(D^t - D^{t-1}) + D^{t-1} < N_1. \tag{3.15}$$

Since $S^t$, $A^t$, $N^t$ and $E^t$ are all nonnegative, it implies from (3.15) that

$$\frac{1}{\tau^t}(D^t - D^{t-1}) + D^{t-1} \overset{\text{(Y)}}{=} \frac{1}{\tau^t}D^t - \frac{1}{\tau^{t-1}}D^{t-1} < N_1.$$

Then it holds that

$$\frac{1}{\tau^t} D^t < tN_1 + \frac{1}{\tau^0} D^0.$$

Therefore, we have

$$D^t \le \tau^t t N_1 + \frac{\tau^t}{\tau^0} D^0 < \infty, \ t \to \infty,$$

that is, $D^t$ is bounded, i.e., $D^t \le N_2$ for some $0 < N_2 < \infty$. According to Cauchy-Schwartz inequality, we have

$$\begin{aligned} D^{t-1} - D^t =& -(M(\breve{v}^t - v^*) + M(\breve{v}^{t-1} - v^*))^T HM(\breve{v}^t - \breve{v}^{t-1}) \\ \le& \|M(\breve{v}^t - v^*) + M(\breve{v}^{t-1} - v^*)\|_H \|M(\breve{v}^t - \breve{v}^{t-1})\|_H \\ \le& 2\sqrt{N_2}\sqrt{E^t}. \end{aligned} \tag{3.16}$$

Combining (3.15) and (3.16) implies that

$$\frac{1-\tau^0}{(\tau^t)^2} E^t \le \frac{1-\tau^t}{(\tau^t)^2} E^t \le N_1 + \frac{2}{\tau^t}\sqrt{N_2}\sqrt{E^t}. \tag{3.17}$$

Let $h^t := \sup_{t \ge 0} \sqrt{E^t}/\tau^t$. It follows from (3.17) that

$$(1-\tau^0) h^t \le \frac{N_1}{h^t} + 2\sqrt{N_2} < \infty,$$

i.e., $E^t \le O((\tau^t)^2)$ for $t \to \infty$. By Lemma 3.1, we complete the proof. □

## 4. Applications.

A simple way to design algorithms that satisfy our faster prediction-correction framework (FPS)-(FCS) is to convert algorithms satisfying (PS)-(CS) to faster versions that satisfy (FPS)-(FCS). In particular, we can construct faster algorithms to solve special cases of (P1) and (P2) based on the algorithms that satisfy (PS)-(CS), such as ADMM, the linear ADMM [43], the strictly contractive Peaceman-Rachford splitting method [11, 13], the Jacobian ALM [10, 42], the ADMM with a substitution [16, 17], ADMM-type algorithm [18], CP-type algorithm [12] with a larger step size and so on. In this section, we exemplify three such faster algorithms.

### 4.1. Faster algorithm satisfying (FPS)-(FCS) for solving (P1) with $m = 2$ and equality constraints.

We design a faster algorithm satisfying (FPS)-(FCS) for solving (P1) $m = 2$.

We consider the following algorithm for solving (2.5):

**[Prediction step.]** For given $P \succ 0$, $\beta > 0$, $(\breve{x}_1^{k-1}, \breve{x}_2^{k-1}, \breve{\lambda}^{k-1})$ and $(x_1^k, x_2^k, \lambda^k)$, find $(\breve{x}_1^k, \breve{x}_2^k, \breve{\lambda}^k)$ by

(4.1)

$$\begin{cases} \breve{x}_1^k \in \arg\min\limits_{x_1} \Big\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta\tau^k}{2} \|A_1(\frac{1}{\tau^k} x_1 - \frac{1-\tau^k}{\tau^k} \breve{x}_1^{k-1}) + A_2 x_2^k - b\|^2 \\ \qquad + \frac{\tau^k}{2} \|\frac{1}{\tau^k} x_1 - \frac{1-\tau^k}{\tau^k} \breve{x}_1^{k-1} - x_1^k\|_P^2 \Big\}, \\ \breve{x}_2^k \in \arg\min\limits_{x_2} \big\{ f_2(x_2) - x_2^T A_2^T [\lambda^k - r\beta(A_1 \widetilde{x}_1^k + A_2 x_2^k - b)] \\ \qquad + \frac{\beta\tau^k}{2} \|A_1 \widetilde{x}_1^k + A_2(\frac{1}{\tau^k} x_2 - \frac{1-\tau^k}{\tau^k} \breve{x}_2^{k-1}) - b\|^2 \Big\}, \\ \breve{\lambda}^k = \arg\max\limits_{\lambda} \Big\{ -\frac{\tau^k}{2} \|(\frac{1}{\tau^k}\lambda - \frac{1-\tau^k}{\tau^k} \breve{\lambda}^{k-1}) - [\lambda^k - \beta(A_1 \widetilde{x}_1^k + A_2 x_2^k - b)]\|^2 \Big\}, \end{cases}$$

where $\tau^k$ satisfy (Y).
**[Correction step.]** Update $(x_1^{k+1}, x_2^{k+1}, \lambda^{k+1})$ by

$$\begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^{k} \\ x_2^{k} \\ \lambda^{k} \end{pmatrix} - \begin{pmatrix} I_{n_1} & 0 & 0 \\ 0 & I_{n_2} & 0 \\ 0 & -s\beta A_2 & (r+s)I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}, \tag{4.2}$$

where

$$\begin{pmatrix} \widetilde{x}_1^k \\ \widetilde{x}_2^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} \frac{1}{\tau^k}\breve{x}_1^k - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1} \\ \frac{1}{\tau^k}\breve{x}_2^k - \frac{1-\tau^k}{\tau^k}\breve{x}_2^{k-1} \\ \frac{1}{\tau^k}\breve{\lambda}^k - \frac{1-\tau^k}{\tau^k}\breve{\lambda}^{k-1} \end{pmatrix}.$$

The following theorem clarifies that Algorithm (4.1)-(4.2) satisfies (FPS)-(FCS) with convergence conditions (CC1)-(CC2). Hence the rates of $O(1/t)$ in the non-ergodic sense of the primal-dual gap and $O(1/t^2)$ in the pointwise sense can be obtained by Theorem 3.4 and 3.5.

THEOREM 4.1. *For $L$, $Q$, $M$ defined in* (2.7) *and $\theta(u)$, $u$, $w$, $F(w)$, $T(w)$ defined in Example 1 with $m = 2$, it holds that*

(1) *Algorithm* (4.1)-(4.2) *satisfies* (FPS)-(FCS).
(2) *$H$ and $G$ satisfying* (CC1)-(CC2) *are positive definite if $A_2$ is full column rank,*

$$r \in (-1, 1), \;\; s \in (0, 1) \text{ and } r + s > 0.$$

*Proof.* Proof of (1). The optimality condition of the $\lambda$-subproblem reads as:

$$0 = A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b - A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k). \tag{4.3}$$

or equivalently,

$$\widetilde{\lambda}^k = \lambda^k - \beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b). \tag{4.4}$$

The optimality condition of the $x_1$-subproblem is given by

$$\begin{aligned} 0 &\in \partial f_1(\breve{x}_1^k) - A_1^T[\lambda^k - \beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b)] + P(\widetilde{x}_1^k - x_1^k) \\ &= \partial f_1(\breve{x}_1^k) - A_1^T\widetilde{\lambda}^k + P(\widetilde{x}_1^k - x_1^k). \end{aligned} \tag{4.5}$$

The optimality condition of the $x_2$-subproblem reads as:

$$\begin{aligned} 0 &\in \partial f_2(\breve{x}_2^k) - A_2^T[\lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b)] + \beta A_2^T[A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b] \\ &= \partial f_2(\breve{x}_2^k) - A_2^T\widetilde{\lambda}^k + \beta A_2^T A_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k). \end{aligned} \tag{4.6}$$

Combining (4.5), (4.6) and (4.3), we have

$$0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P(\widetilde{x}_1^k - x_1^k) \\ \beta A_2^T A_2(\widetilde{x}_2^k - x_2^k) - rA_2^T(\widetilde{\lambda}^k - \lambda^k) \\ -A_2(\widetilde{x}_2^k - x_2^k) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},$$

which is equivalent to

$$0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T \widetilde{\lambda}^k \\ -A_2^T \widetilde{\lambda}^k \\ A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b \end{pmatrix} + \begin{pmatrix} P & 0 & 0 \\ 0 & \beta A_2^T A_2 & -rA_2^T \\ 0 & -A_2 & \frac{1}{\beta} I_l \end{pmatrix} \begin{pmatrix} x_1^k - \widetilde{x}_1^k \\ x_2^k - \widetilde{x}_2^k \\ \lambda^k - \widetilde{\lambda}^k \end{pmatrix}.$$

Then we obtain the prediction step (4.1) satisfying (FPS) with $L$ and $Q$ defined in (2.7). The correction step is easy to verified.

Proof of (2). We can refer to Theorem 2.3 (2). ∎

REMARK 4. *According to the correction step* (4.2)*, it holds that*

$$x_1^{k+1} = \widetilde{x}_1^k, \quad x_2^{k+1} = \widetilde{x}_2^k.$$

*and*

$$\begin{aligned} \lambda^{k+1} &= \lambda^k + s\beta A_2(x_2^k - \widetilde{x}_2^k) - (r+s)(\lambda^k - \widetilde{\lambda}^k) \\ &\overset{(4.4)}{=} \underbrace{\lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b)}_{:\lambda^{k+\frac{1}{2}}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b) \\ &= \lambda^{k+\frac{1}{2}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b) \\ &= \lambda^{k+\frac{1}{2}} - s\beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \end{aligned}$$

*Then Algorithm* (4.1)-(4.2) *can be rewritten as:*

$$\begin{cases} \breve{x}_1^k & \in \arg\min\limits_{x_1} \big\{ f_1(x_1) - x_1^T A_1^T \lambda^k \\ & \quad + \frac{\beta\tau^k}{2} \| A_1(\frac{1}{\tau^k} x_1 - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1}) + A_2 x_2^k - b\|^2 + \frac{\tau^k}{2} \| \frac{1}{\tau^k} x_1 - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1} - x_1^k\|_P^2 \Big\}, \\ x_1^{k+1} & = \frac{1}{\tau^k}\breve{x}_1^k - \frac{1-\tau^k}{\tau^k}\breve{x}_1^{k-1}, \\ \lambda^{k+\frac{1}{2}} & = \lambda^k - r\beta(A_1\widetilde{x}_1^k + A_2 x_2^k - b), \\ \breve{x}_2^k & \in \arg\min\limits_{x_2} \Big\{ f_2(x_2) - x_2^T A_2^T \lambda^{k+\frac{1}{2}} \\ & \quad + \frac{\beta\tau^k}{2} \| A_1\widetilde{x}_1^k + A_2(\frac{1}{\tau^k} x_2 - \frac{1-\tau^k}{\tau^k}\breve{x}_2^{k-1}) - b\|^2 \Big\}, \\ x_2^{k+1} & = \frac{1}{\tau^k}\breve{x}_2^k - \frac{1-\tau^k}{\tau^k}\breve{x}_2^{k-1}, \\ \lambda^{k+1} & = \lambda^{k+\frac{1}{2}} - s\beta(A_1\widetilde{x}_1^k + A_2\widetilde{x}_2^k - b). \end{cases}$$

*We can find that this algorithm update $\lambda$ twice in one iteration. Numerically, updating $\lambda$ twice performance better than updating $\lambda$ once. This is the first paper that updates $\lambda$ twice with non-ergodic convergence.*

**4.2. Faster algorithm satisfying** (FPS)**-**(FCS) **for solving** (P1)**.** We design a faster algorithm satisfying (FPS)-(FCS) for solving (P1).

We consider the following algorithm for solving (P1):

**[Prediction step.]** With given $\beta > 0$, $(\breve{x}_1^{k-1}, \breve{x}_2^{k-1}, \ldots, \breve{x}_m^{k-1}, \breve{\lambda}^{k-1})$ and

$(x_1^k, x_2^k, \dots, x_m^k, \lambda^k)$, find $(\breve{x}_1^k, \breve{x}_2^k, \dots, \breve{x}_m^k, \breve{\lambda}^k)$ by

$$\begin{cases} \breve{x}_1^k = \arg\min_{x_1} \left\{ f_1(x_1) - x_1^T A_1^T \lambda^k + \frac{\beta\tau^k}{2} \|A_1([\frac{1}{\tau^k} x_1 - \frac{1-\tau^k}{\tau^k} \breve{x}_1^{k-1}] - x_1^k)\|^2 \right\}, \\ \breve{x}_2^k = \arg\min_{x_2} \{ f_2(x_2) - x_2^T A_2^T \lambda^k \\ \qquad + \frac{\beta\tau^k}{2} \|A_1(\widetilde{x}_1^k - x_1^k) + A_2([\frac{1}{\tau^k} x_2 - \frac{1-\tau^k}{\tau^k} \breve{x}_2^{k-1}] - x_2^k)\|^2 \right\}, \\ \qquad \vdots \\ \breve{x}_i^k = \arg\min_{x_i} \{ f_i(x_i) - x_i^T A_2^T \lambda^k \\ \qquad + \frac{\beta\tau^k}{2} \| \sum_{j=1}^{i-1} A_j(\widetilde{x}_j^k - x_j^k) + A_i([\frac{1}{\tau^k} x_i - \frac{1-\tau^k}{\tau^k} \breve{x}_i^{k-1}] - x_i^k)\|^2 \right\}, \\ \qquad \vdots \\ \breve{x}_m^k = \arg\min_{x_m} \{ f_m(x_m) - x_m^T A_2^T \lambda^k \\ \qquad + \frac{\beta\tau^k}{2} \| \sum_{j=1}^{m-1} A_j(\widetilde{x}_j^k - x_j^k) + A_m([\frac{1}{\tau^k} x_m - \frac{1-\tau^k}{\tau^k} \breve{x}_m^{k-1}] - x_m^k)\|^2 \right\}, \\ \breve{\lambda}^k = \arg\max_{\lambda} \left\{ -\lambda^T(\sum_{j=1}^m A_j \widetilde{x}_j^k - b) - \frac{\tau^k}{2\beta} \|[\frac{1}{\tau^k}\lambda - \frac{1-\tau^k}{\tau^k} \breve{\lambda}^{k-1}] - \lambda^k\|^2 \right\}, \end{cases} \tag{4.7}$$

where $\tau^k$ satisfy (Y).

**[Correction step.]** Update $(A_1 x_1^{k+1}, A_2 x_2^{k+1}, \dots, A_m x_m^{k+1}, \lambda^{k+1})$ by

(4.8)

$$\begin{pmatrix} \sqrt{\beta} A_1 x_1^{k+1} \\ \sqrt{\beta} A_2 x_2^{k+1} \\ \vdots \\ \sqrt{\beta} A_m x_m^{k+1} \\ \frac{1}{\sqrt{\beta}} \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta} A_1 x_1^{k} \\ \sqrt{\beta} A_2 x_2^{k} \\ \vdots \\ \sqrt{\beta} A_m x_m^{k} \\ \frac{1}{\sqrt{\beta}} \lambda^{k} \end{pmatrix} - \begin{pmatrix} \alpha I_l & -\alpha I_l & 0 & \dots & 0 \\ 0 & \alpha I_l & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha I_l & 0 \\ 0 & \dots & 0 & \alpha I_l & 0 \\ -\alpha I_l & 0 & \dots & 0 & I_l \end{pmatrix} \begin{pmatrix} \sqrt{\beta}(A_1 x_1^k - A_1 \widetilde{x}_1^k) \\ \sqrt{\beta}(A_2 x_2^k - A_2 \widetilde{x}_2^k) \\ \vdots \\ \sqrt{\beta}(A_m x_m^k - A_m \widetilde{x}_m^k) \\ \frac{1}{\sqrt{\beta}}(\lambda^k - \widetilde{\lambda}^k) \end{pmatrix},$$

where

$$\begin{pmatrix} \widetilde{x}_1^k \\ \vdots \\ \widetilde{x}_m^k \\ \widetilde{\lambda}^k \end{pmatrix} = \begin{pmatrix} \frac{1}{\tau^k} \breve{x}_1^k - \frac{1-\tau^k}{\tau^k} \breve{x}_1^{k-1} \\ \vdots \\ \frac{1}{\tau^k} \breve{x}_m^k - \frac{1-\tau^k}{\tau^k} \breve{x}_m^{k-1} \\ \frac{1}{\tau^k} \breve{\lambda}^k - \frac{1-\tau^k}{\tau^k} \breve{\lambda}^{k-1} \end{pmatrix}.$$

The following theorem clarifies that Algorithm (4.7)-(4.8) satisfies (FPS)-(FCS) with convergence conditions (CC1)-(CC2). Hence the rates of $O(1/t)$ in the non-ergodic sense of the primal-dual gap and $O(1/t^2)$ in the pointwise sense can be obtained by Theorem 3.4 and 3.5.

THEOREM 4.2. *For $L$, $Q$, $M$ defined in* (2.12) *and $\theta(u)$, $u$, $w$, $F(w)$, $T(w)$ defined in Example 1, it holds that*

(1) *Algorithm* (4.7)-(4.8) *satisfies* (FPS)-(FCS).
(2) *$H$ and $G$ satisfying* (CC1)-(CC2) *are positive definite if $\alpha \in (0, 1)$.*

*Proof.* Proof of (1). For $i = 1, 2, \dots, m$, the optimality condition of the $x_i$-

subproblem is given by

$$
\begin{aligned}
0 &\in \partial f_i(\breve{x}_i^k) - A_i^T\lambda^k + \beta A_i^T\sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) \\
&= \partial f_i(\breve{x}_i^k) - A_i^T\widetilde{\lambda}^k + \beta A_i^T\sum_{j=1}^{i} A_j(\widetilde{x}_j^k - x_j^k) + A_i^T(\widetilde{\lambda}^k - \lambda^k)
\end{aligned} \tag{4.9}
$$

The optimality condition of the $\lambda$-subproblem reads as:

$$
0 = (\sum_{j=1}^{m} A_j\widetilde{x}_j^k - b) + \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k). \tag{4.10}
$$

Combining (4.9) and (4.10), we obtain

$$
0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ \vdots \\ \partial f_m(\breve{x}_m) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + \begin{pmatrix} \beta A_1^T A_1(\widetilde{x}_1^k - x_1^k) + A_1^T(\widetilde{\lambda}^k - \lambda^k) \\ \beta A_2^T\sum_{j=1}^2 A_j(\widetilde{x}_j^k - x_j^k) + A_2^T(\widetilde{\lambda}^k - \lambda^k) \\ \vdots \\ \beta A_m^T\sum_{j=1}^m A_j(\widetilde{x}_j^k - x_j^k) + A_m^T(\widetilde{\lambda}^k - \lambda^k) \\ \frac{1}{\beta}(\widetilde{\lambda}^k - \lambda^k) \end{pmatrix},
$$

which is equivalent to:

$$
0 \in \begin{pmatrix} \partial f_1(\breve{x}_1^k) \\ \partial f_2(\breve{x}_2^k) \\ \vdots \\ \partial f_m(\breve{x}_m) \\ 0 \end{pmatrix} + \begin{pmatrix} -A_1^T\widetilde{\lambda}^k \\ -A_2^T\widetilde{\lambda}^k \\ \vdots \\ -A_m^T\widetilde{\lambda}^k \\ \sum_{j=1}^m A_j\widetilde{x}_j^k - b \end{pmatrix} + L^T\begin{pmatrix} I_l & 0 & \dots & 0 & I_l \\ I_l & I_l & \ddots & \vdots & I_l \\ \vdots & \vdots & \ddots & 0 & \vdots \\ I_l & I_l & \dots & I_l & I_l \\ 0 & 0 & \dots & 0 & I_l \end{pmatrix} L \begin{pmatrix} \widetilde{x}_1^k - x_1^k \\ \widetilde{x}_2^k - x_2^k \\ \vdots \\ \widetilde{x}_m^k - x_m^k \\ \widetilde{\lambda}^k - \lambda^k \end{pmatrix}.
$$

Then the prediction step (4.8) satisfies (FPS) with $L$ and $Q$ defined in (2.12). The correction step is easy to verified.
Proof of (2). We can refer to Theorem 2.4 (2). □

**4.3. Faster algorithm satisfying** (FPS)**-**(FCS) **for solving** (P2)**.** This subsection presents an algorithm satisfying (FPS)-(FCS) for solving (P2).

We consider the following algorithm for solving (P2):

**[Prediction step.]** With given $r, s > 0$, $(\breve{x}^{k-1}, \breve{y}^{k-1})$ and $(x^k, y^k)$, find $(\breve{x}^k, \breve{y}^k)$ by

$$
\begin{cases} \breve{x}^k = \arg\min_x \left\{\Phi(x, y^k) + \frac{r\tau^k}{2}\|(\frac{1}{\tau^k}x - \frac{1-\tau^k}{\tau^k}\breve{x}^{k-1}) - x^k\|^2\right\}, \\ \breve{y}^k = \arg\max_y \left\{\Phi([\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)], y) - \frac{s\tau^k}{2}\|(\frac{1}{\tau^k}y - \frac{1-\tau^k}{\tau^k}\breve{y}^{k-1}) - y^k\|^2\right\}, \end{cases} \tag{4.11}
$$

where $\tau^k$ satisfy (Y).
**[Correction step.]** Update $(x^{k+1}, y^{k+1})$ by

$$
\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \end{pmatrix} - \begin{pmatrix} I_n & 0 \\ -(1-\alpha)\frac{1}{s}A & I_m \end{pmatrix}\begin{pmatrix} x^k - \widetilde{x}^k \\ y^k - \widetilde{y}^k \end{pmatrix}, \tag{4.12}
$$

where

$$\begin{pmatrix}\widetilde{x}^k\\ \widetilde{y}^k\end{pmatrix} = \begin{pmatrix}\frac{1}{\tau^k}\breve{x}^k - \frac{1-\tau^k}{\tau^k}\breve{x}^{k-1}\\ \frac{1}{\tau^k}\breve{y}^k - \frac{1-\tau^k}{\tau^k}\breve{y}^{k-1}\end{pmatrix}.$$

The following theorem clarifies that Algorithm (4.11)-(4.12) satisfies (FPS)-(FCS) with convergence conditions (CC1)-(CC2). Hence the rates of $O(1/t)$ in the non-ergodic sense of the primal-dual gap and $O(1/t^2)$ in the pointwise sense can be obtained by Theorem 3.4 and 3.5.

THEOREM 4.3. *For* $L$, $Q$, $M$ *defined in* (2.12) *and* $\theta(u)$, $u$, $w$, $F(w)$, $T(w)$ *as in Example 2, it holds that*

(1) *Algorithm* (4.11)-(4.12) *satisfies* (FPS)-(FCS).
(2) $H$ *and* $G$ *satisfying* (CC1)-(CC2) *are positive definite if*

$$rs > (1-\alpha+\alpha^2)\rho(A^TA),\ \ \alpha\in[0,1].$$

*Proof.* Proof of (1). The optimality condition of the $x$-subproblem reads as:

$$\begin{aligned}0 &\in \partial f(\breve{x}^k) - A^T y^k + r(\widetilde{x}^k - x^k)\\ &= \partial f(\breve{x}^k) - A^T\widetilde{y}^k + r(\widetilde{x}^k - x^k) + A^T(\widetilde{y}^k - y^k).\end{aligned}$$

The optimality condition of the $y$-subproblem reads as:

$$0 \in \partial g(\breve{y}^k) + A[\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)] + s(\widetilde{y}^k - y^k).$$

Combining the above two relations together yields that

$$0 \in \begin{pmatrix}\partial f(\breve{x}^k)\\ \partial g(\breve{y}^k)\end{pmatrix} + \begin{pmatrix}-A^T\widetilde{y}^k\\ A\widetilde{x}^k\end{pmatrix} + \begin{pmatrix}r(\widetilde{x}^k - x^k) + A^T(\widetilde{y}^k - y^k)\\ \alpha A(\widetilde{x}^k - x^k) + s(\widetilde{y}^k - y^k)\end{pmatrix}.$$

This is equivalent to:

$$0 \in \begin{pmatrix}\partial f(\breve{x}^k)\\ \partial g(\breve{y}^k)\end{pmatrix} + \begin{pmatrix}-A^T\widetilde{y}^k\\ A\widetilde{x}^k\end{pmatrix} + \begin{pmatrix}rI_n & A^T\\ \alpha A & sI_m\end{pmatrix}\begin{pmatrix}\widetilde{x}^k - x^k\\ \widetilde{y}^k - y^k\end{pmatrix}.$$

Then we obtain the prediction step (4.11) satisfies (FPS) with $L$ and $Q$ defined in (2.15). The correction step is easy to verified.
Proof of (2). We can refer to Theorem 2.5 (2). ∎

**5. Conclusions.** We present a faster prediction-correction framework to build $O(1/t)$ convergence rate in the non-ergodic sense and $O(1/t^2)$ convergence rate in the pointwise sense without any additional assumption. In comparison, He and Yuan's framework achieves an $O(1/t)$ convergence rate in both the ergodic and the pointwise senses. Our framework can provide faster algorithms for solving general convex optimization problems. In particular, we present three faster algorithms: ADMM-type algorithm with dual variable updating twice for solving two-block separable convex optimization with equality linear constraints, multi-block ADMM-type algorithm for solving multi-block separable convex optimization problems with linear equality constraints and CP-type algorithm for solving min-max problems with larger step sizes ($rs > 0.75\rho(A^TA)$). Future works include in-depth understanding our framework, for example, from the view point of second-order differential equations, establishing the weak convergence of the iterative sequence and the KKT measure.

## REFERENCES


[1] H. Attouch and S. László, *Newton-like inertial dynamics and proximal algorithms governed by maximally monotone operators*, SIAM J. Optim., 30 (2020), pp. 3252–3283.

[2] H. Attouch and J. Peypouquet, *Convergence of inertial dynamics and proximal algorithms governed by maximal monotone operators*, Math. Program., (2017), pp. 1–42.

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

[4] 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.

[5] A. Chambolle and T. Pock, *On the ergodic convergence rates of a first-order primal-dual algorithm*, Math. Program., 159 (2016), pp. 253–287.

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

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

[8] 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-Math. Model. Num., 9 (1975), pp. 41–76.

[9] N. Golowich, S. Pattathil, C. Daskalakis, and A. Ozdaglar, *Last iterate is slower than averaged iterate in smooth convex-concave saddle point problems*, Conference on Learning Theory (PMLR), (2020), pp. 1758–1784.

[10] B. He, L. Hou, and X. Yuan, *On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming*, SIAM J. Optim., 25 (2015), pp. 2274–2312.

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

[12] B. He, F. Ma, S. Xu, and X. Yuan, *A generalized primal-dual algorithm with improved convergence condition for saddle point problems*, SIAM J. Imaging Sci., 15 (2022), pp. 1157–1183.

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

[14] B. He, F. Ma, and X. Yuan, *An algorithmic framework of generalized primal–dual hybrid gradient methods for saddle point problems*, J. Math. Imaging Vis., 58 (2017), pp. 279–293.

[15] B. He, F. Ma, and X. Yuan, *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.

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

[17] B. He, M. Tao, and X. Yuan, *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.

[18] B. He, S. Xu, and X. Yuan, *Extensions of ADMM for separable convex optimization problems with linear equality or inequality constraints*, arXiv preprint arXiv:2107.01897, (2021).

[19] B. He and X. Yuan, *On the $O(1/t)$ convergence rate of the alternating direction method*, SIAM J. Numer. Anal., 50 (2012), pp. 700–709.

[20] B. He and X. Yuan, *On non-ergodic convergence rate of Douglas¨cRachford alternating direction method of multipliers*, Numer. Math., 130 (2015), pp. 567–577.

[21] B. He and X. Yuan, *On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization*, arXiv:2204.11522, (2022).

[22] X. He, R. Hu, and Y. Fang, *Inertial primal-dual methods for linear equality constrained convex optimization problems*, Numer. Algorithms, 90 (2022), pp. 1669–1690.

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

[24] R. Ioan Boţ, E. Csetnek, and D. Nguyen, *Fast augmented Lagrangian method in the convex regime with convergence guarantees for the iterates*, Math. Program., 200 (2023), pp. 147–197.

[25] F. Jiang, Z. Zhang, and H. He, *Solving saddle point problems: a landscape of primal-dual algorithm with larger stepsizes*, J. Global Optim., 85 (2023), pp. 821–846.

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

[27] H. Li and M. Yan, *On the improved conditions for some primal-dual algorithms*, arXiv:2201.00139v1, (2022).
[28] Z. Lin, H. Li, and C. Fang, *Accelerated Optimization for Machine Learning: First-Order Algorithms*, Springer, Nature Singapore, 2020.
[29] H. Luo, *Accelerated primal-dual methods for linearly constrained convex optimization problems*, arXiv:2109.12604, (2021).
[30] H. Luo, *A unified differential equation solver approach for separable convex optimization: splitting, acceleration and nonergodic rate*, arXiv:2109.13467, (2021).
[31] H. Luo, *A primal-dual flow for affine constrained convex optimization*, ESAIM Control Optim. Calc. Var., 28 (2022).
[32] H. Luo and L. Chen, *From differential equation solvers to accelerated first-order methods for convex optimization*, Math. Program., 195 (2022), pp. 735–781.
[33] B. Martinet, *Regularisation d'inequations variationelles par approximations successives*, Revue Francaise d'informatique et de Recherche operationelle, 4 (1970), pp. 154–159.
[34] Y. Nesterov, *A method for solving the convex programming problem with convergence rate* $O(1/k^2)$, Dokl. Akad. Nauk SSSR, 269 (1983), pp. 543–547.
[35] M. J. D. Powell, *A method for nonlinear constraints in minimization problems*, Optimization, (1969), pp. 283–298.
[36] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Math. Oper. Res., 1 (1976), pp. 97–116.
[37] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Cont. Opt., 14 (1976), pp. 877–898.
[38] S. Sabach and M. Teboulle, *Faster Lagrangian-based methods in convex optimization*, SIAM J. Optim., 32 (2022), pp. 204–227.
[39] Q. Tran-Dinh and Y. Zhu, *Non-stationary first-order primal-dual algorithms with faster convergence rates*, SIAM J. Optim., 30 (2020), pp. 2866–2896.
[40] P. Tseng, *Approximation accuracy, gradient methods, and error bound for structured convex optimization*, Math. Program., 125 (2010), pp. 263–295.
[41] T. Valkonen, *Inertial, corrected, primal-dual proximal splitting*, SIAM J. Optim., 30 (2020), pp. 1391–1420.
[42] S. Xu and B. He, *A parallel splitting ALM-based algorithm for separable convex programming*, Comput. Optim. Appl., 80 (2021), pp. 831–851.
[43] J. Yang and X. Yuan, *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization*, Math. Comput., 82 (2013), pp. 301–329.