A new insight on the prediction-correction framework with applications to several first-order methods

Jianchao Bai[a,b,*], Shuang Rao[a,c,*]

[a] School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an, 710129, China
[b] Center for Applied Mathematics of Fujian Province, Fuzhou University, Fuzhou, 350116, China
[c] Shenzhen Middle School, Shenzhen, 518200, China

ARTICLE INFO

Keywords:
Convex optimization
Prediction-correction framework
First-order method
Convergence

ABSTRACT

We propose a generalized prediction-correction framework featuring a parameter-free relaxation iteration for solving linearly constrained convex programs. By leveraging variational characterization of the first-order optimality conditions for each resulting subproblem, we establish its global convergence and sublinear convergence rates in both ergodic and nonergodic senses. Furthermore, this new framework is applied to reformulate an indefinite linearized augmented Lagrangian method, the Chambolle-Pock method, and two ADMM-type methods, enabling concise analyses of their convergence conditions.

1. Introduction

Consider the following canonical convex optimization model

$$\min \{\theta(x) \mid Ax = b, x \in \mathcal{X}\}, \tag{1}$$

where $\theta : \mathbb{R}^n \to \mathbb{R}$ is a closed proper convex function (possibly nonsmooth), $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed convex set, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are given. Hereafter, the symbols $\mathbb{R}^{n\times m}, \mathbb{R}^n$ and $\mathbb{R}$ denote the set of $n \times m$ dimensional real matrices, the set of n dimensional real column vectors and the set of real numbers, respectively. A direct extension of (1) is the following separable convex minimization problem

$$\min \left\{\theta_1(y) + \theta_2(z) \mid By + Cz = b, y \in \mathcal{Y}, z \in \mathcal{Z}\right\}, \tag{2}$$

in which $\theta_1 : \mathbb{R}^{n_1} \to \mathbb{R}$ and $\theta_2 : \mathbb{R}^{n_2} \to \mathbb{R}$ are closed proper convex functions, $B \in \mathbb{R}^{m\times n_1}, C \in \mathbb{R}^{m\times n_2}$ are given, $\mathcal{Y} \subseteq \mathbb{R}^{n_1}$ and $\mathcal{Z} \subseteq \mathbb{R}^{n_2}$ are closed convex sets. Constraints for cases with $Ax \geq b$ ($By + Cz \geq b$) can be addressed in a manner analogous to that in [3, 26] by restricting the associated dual variables to the nonnegative space. Throughout this article, the solution sets of the aforementioned problems are assumed to be nonempty.

1.1. Related work

The augmented Lagrangian method [18] is a fundamental method for solving linearly constrained optimization problems. Applying this method to (1) yields

$$\begin{cases} x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \mathcal{L}_\beta(x, \lambda^k), \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} - b), \end{cases} \tag{3}$$

where $\mathcal{L}_\beta(x, \lambda) = \theta(x) - \langle \lambda, Ax - b\rangle + \frac{\beta}{2}\|Ax - b\|^2$ is the augmented Lagrangian function of (1) with $\beta > 0$, and $\lambda \in \mathbb{R}^m$ denotes the Lagrange multiplier. By denoting $\theta = \theta_1 + \theta_2, A = [B, C], x = [y; z], \mathcal{X} = \mathcal{Y}\times\mathcal{Z}$, the problem (2) becomes the form of (1), and hence we have

$$\begin{cases} (y^{k+1}, z^{k+1}) = \arg\min\limits_{(y,z)\in\mathcal{Y}\times\mathcal{Z}} \mathcal{L}_\beta(y, z, \lambda^k), \\ \lambda^{k+1} = \lambda^k - \beta(By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{4}$$

where $\mathcal{L}_\beta(y, z, \lambda) = \theta_1(y) + \theta_2(z) - \langle\lambda, By + Cz - b\rangle + \frac{\beta}{2}\|By + Cz - b\|^2$. Note that solving the subproblem of (4) is challenging since ALM cannot exploit the separable structures of the problem (2) and thus fails to leverage the special properties (such as sparsity, low rank, and local smoothness) of each component objective function. An effective approach to overcome this difficulty is the splitting variant of ALM, namely ADMM [8]. At each iteration, ADMM first sequentially optimizes with respect to one variable while fixing the other, and then updates the Lagrange multiplier.

The key to the performance of ALM lies in how to efficiently tackle the core subproblem. However, neither the subproblem in (3) nor that in (4) can leverage the corresponding proximal operator of each objective function, due to the presence of the quadratic term. To overcome this obstacle, a widely adopted approach is to incorporate a proximal term of the form $\frac{1}{2}\|x - x^k\|^2_{D_0}$ into the objective function of the subproblems. For instance, He et al. [14] proposed the following indefinite linearized ALM (IDL-ALM):

$$\begin{cases} x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \left\{\mathcal{L}_\beta(x, \lambda^k) + \frac{1}{2}\|x - x^k\|^2_{D_0}\right\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} - b), \end{cases} \tag{5}$$

with a stepsize $\gamma\in(0, 2)$. Here, D_0 takes the form of

$$D_0 = D - (1 - \tau)\beta A^\top A \quad \text{with} \quad r > \beta\|A^\top A\|,$$

$D \in \mathbb{R}^{n\times n}$ is an arbitrary positive definite matrix, and the proximal parameter τ satisfies $\frac{2+\gamma}{4} < \tau < 1$. When applying

---

*Corresponding author
jianchaobai@nwpu.edu.cn (J. Bai); raoshuang169@163.com (S. Rao)
ORCID(s): 0000-0002-2394-8852 (J. Bai)

the widely used Chambolle-Pock method [7] to the saddle-point problem of (1), the resulting algorithm reads:

$$\begin{cases} x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \left\{ \theta(x) - \langle \lambda^k, Ax\rangle + \frac{1}{2r}\|x - x^k\|^2 \right\}, \\ \lambda^{k+1} = \lambda^k - s\big[A(2x^{k+1} - x^k) - b\big]. \end{cases} \tag{6}$$

Here, the parameters r and s are called the primal stepsize and dual stepsize respectively, and they satisfy $1/(rs) > \|A^\top A\|$. Notably, the Chambolle-Pock method (6) differs from ALM-type methods: its dual update employs an inertial step with a unit stepsize, while its primal update features a proximity operator.

To utilize a larger dual stepsize and incorporate a proximal term for the nonsmooth subproblem, Ma [22] developed the following proximal ADMM for solving (2):

$$\begin{cases} y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \mathcal{L}_\beta(y, z^k, \lambda^k), \\ z^{k+1} = \arg\min\limits_{z\in\mathcal{Z}} \left\{ \mathcal{L}_\beta(y^{k+1}, z, \lambda^k) + \frac{1}{2}\|z - z^k\|_D^2 \right\}, \\ \lambda^{k+1} = \lambda^k - \gamma\beta(By^{k+1} + Cz^{k+1} - b), \end{cases} \tag{7}$$

where $D \succeq \tau\beta C^\top C$ is a positive definite matrix and the parameters τ and γ are restricted to the following region:

$$\left\{(\tau, \gamma) \mid \tau \geq 0,\ 0 < \gamma < \frac{1 - \tau + \sqrt{\tau^2 + 6\tau + 5}}{2}\right\}. \tag{8}$$

Although a proximal term is employed in (7), the lower bound of the proximal parameter τ is not optimal. To introduce an indefinite proximal term whose proximal parameter can attain the smallest lower bound, an optimal linearized ADMM:

$$\begin{cases} y^{k+1} = \arg\min\limits_{y\in\mathcal{Y}} \mathcal{L}_\beta(y, z^k, \lambda^k) \\ z^{k+1} = \arg\min\limits_{z\in\mathcal{Z}} \left\{ \mathcal{L}_\beta(y^{k+1}, z, \lambda^k) + \frac{1}{2}\|z - z^k\|_{D_0}^2 \right\} \\ \lambda^{k+1} = \lambda^k - \beta(By^{k+1} + Cz^{k+1} - b) \end{cases} \tag{9}$$

was proposed in [13], where $D_0 = \tau r\mathbf{I} - \beta C^\top C$ with $r > \beta\|C^\top C\|$ and $\tau \in (0.75, 1)$.

The dual variable in the aforementioned ADMM-type methods is updated only once. Another variant of ADMM is the symmetric ADMM [12], which features dual variables updated twice with different stepsizes r and s satisfying

$$s \in \left(0, \frac{1 + \sqrt{5}}{2}\right),\ r \in (-1, 1),\ r + s > 0,\ |r| < 1 + s - s^2.$$

As an extension of [12], the generalized symmetric ADMM [1] features a partial proximal term and extends the aforementioned region to a semi-elliptical region.

Recently, a double proximal augmented Lagrangian method [4] was proposed for solving the convex programming problem (1):

$$\begin{cases} x^{k+1} = \arg\min\limits_{x\in\mathcal{X}} \left\{ \begin{array}{l} \theta(x) - \langle \lambda^k, Ax\rangle \\ + \frac{\tau\beta}{2}\|A(x - x^k)\|^2 + \frac{\tau}{2}\|x - x^k\|_D^2 \end{array} \right\}, \\ \lambda^{k+1} = \lambda^k - \beta\big[\gamma(Ax^{k+1} - b) + A(x^{k+1} - x^k)\big], \end{cases} \tag{10}$$

where $D = r\mathbf{I} - \beta A^\top A$ with $r > \beta\|A^\top A\|$ and

$$\tau > \frac{2 + \gamma}{4} + \frac{(\alpha - \frac{\gamma}{2})^2}{2 - \gamma}, \quad \forall \gamma \in (0, 2), \alpha \in \mathbb{R}.$$

The authors further proposed a prediction-correction framework to reformulate (10). Specifically, it corresponds to the framework in Section 3 with $\eta = 1$ and a given nonsingular lower triangular matrix M. Note that unlike existing ALM-type methods, the subproblem in (10) does not depend on the data b. Actually, since ALM converges, we can derive $Ax^k \to b$ and $A(x^{k+1} - x^k) \to 0$. Therefore, the penalty $\frac{\beta}{2}\|A(x - x^k)\|$ is adopted in place of the traditional penalty $\frac{\beta}{2}\|Ax - b\|$, while the term $A(x^{k+1} - x^k)$ is employed in the dual update. Another observation is that the prediction-correction framework proposed in [4] greatly simplifies the convergence analysis of the scheme (10).

1.2. Motivations and contributions

Motivated by the framework proposed in [4], a natural and interesting question arises: *Can this framework be extended to a more general version and applied to simplify the convergence analysis of some existing ALMs and their related methods?* To address this question, we develop a general prediction-correction framework as outlined at the beginning of Section 3. The main contributions and features of our framework are summarized as follows:

- We have developed a new prediction-correction framework and established its global convergence and sublinear convergence rate. *The main innovation lies in that the extra term $\alpha A(\tilde{x}^k - x^k)$ is exploited in the dual prediction.* Our framework can reduce to that in [4] by setting

  $$\eta = 1 \ \text{ and } \ M = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -(1 - \alpha\gamma)\beta A & \gamma\mathbf{I} \end{bmatrix}, \ \forall \alpha \in \mathbb{R}.$$

  By setting $\alpha = t, \eta = 1, M = \mathbf{I}$, our framework except for (13a) reduces to [21, Algorithm 1]. In the second part of Remark 3.1, we provide two feasible methods for selecting a general nonsingular matrix M so as to obtain a customized iterative algorithm.

- Our framework employs a parameter-free iteration for the dual variable based on the residual between the latest primal iteration and its predicted iteration. As explained in the first item of Remark 3.1, the prediction step of the dual variable allows for a forward or backward step. The proposed prediction-correction framework is also applied to reformulate a linearized augmented Lagrangian method, the Chambolle-Pock method, and two ADMM-type methods, and their convergence conditions are concisely discussed.

1.3. Notations

Without loss of generality, we adopt $\langle\cdot,\cdot\rangle$ and $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$ to denote the standard inner product and Euclidean norm, respectively. The bold $\mathbf{I}$ and $\mathbf{0}$ represent the identity matrix and zero matrix/vector, respectively. For any $\mathbf{0} \neq w \in \mathbb{R}^n$, H is positive definite if $w^\top H w > 0$, and H is positive semidefinite if $w^\top H w \geq 0$. We simply denote a positive definite matrix and a positive semidefinite matrix by $H \succ \mathbf{0}$ and $H \succeq \mathbf{0}$, respectively. For a $G \succ \mathbf{0}$, define $\|\mathbf{x}\|_G := \sqrt{\mathbf{x}^\top G \mathbf{x}}$ as its G-weighted norm. For the case that $G = \mathbf{I}$, it reduces to the standard Euclidean norm. For a proper convex function $\theta(x)$, its proximal operator with parameter $\rho > 0$ is defined as

$$\mathbf{prox}_{\rho,\theta}(\cdot) := \arg\min_{x\in\mathcal{X}} \left\{\theta(x) + \frac{\rho}{2}\|x - \cdot\|^2\right\}.$$

2. Variational inequality reformulation

We begin with the following lemma (see also [6, 12]) which will be used to characterize both the saddle-point of the problem and the first-order optimality conditions of the subproblems discussed in the subsequent sections.

Lemma 2.1. *Let $\Phi \subset \mathbb{R}^m$ be a closed convex set and $f, g : \mathbb{R}^m \to \mathbb{R}$ be two convex functions, where g is differentiable. Suppose that the solution set $\Phi^* = \arg\min\{f(x)+g(x) | x \in \Phi\}$ is nonempty. Then, $x^* \in \Phi^*$ if and only if*

$$x^* \in \Phi,\ f(x) - f(x^*) + \left\langle x - x^*, \nabla g(x^*)\right\rangle \geq 0,\ \forall x \in \Phi.$$

Based on Lemma 2.1, we next characterize the saddle-point of the problem (1). We call $(x^*, \lambda^*) \in \Omega := \mathcal{X} \times \mathbb{R}^m$ the saddle point of (1) if

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*), \quad \forall x \in \mathcal{X}, \lambda \in \mathbb{R}^m.$$

Here, $L(x, \lambda) = \theta(x) - \langle\lambda, Ax - b\rangle$ stands for the Lagrangian function of (1).The last two inequalities are equivalent to

$$\begin{cases} x^* = \arg\min\left\{\theta(x) - \langle\lambda^*, Ax - b\rangle \mid x \in \mathcal{X}\right\}, \\ \lambda^* = \arg\min\left\{\theta(x^*) - \langle\lambda, Ax^* - b\rangle \mid \lambda \in \mathbb{R}^m\right\}. \end{cases}$$

Then, applying Lemma 2.1 to the above two problems gives

$$\begin{cases} x^* \in \mathcal{X}, \theta(x) - \theta(x^*) + \left\langle x - x^*, -A^\top\lambda^*\right\rangle \geq 0, \forall x \in \mathcal{X}, \\ \lambda^* \in \mathbb{R}^m, \left\langle \lambda - \lambda^*, Ax^* - b\right\rangle \geq 0, \forall \lambda \in \mathbb{R}^m. \end{cases}$$

Combining these two inequalities yields the following mixed variational inequality

$$\theta(x) - \theta(x^*) + \left\langle w - w^*, \mathcal{J}(w^*)\right\rangle \geq 0, \quad \forall w \in \Omega, \tag{11}$$

where

$$w = \begin{pmatrix} x \\ \lambda \end{pmatrix} \quad \text{and} \quad \mathcal{J}(w) = \begin{pmatrix} -A^\top\lambda \\ Ax - b \end{pmatrix}. \tag{12}$$

An equivalent expression of (11) is $\theta(x) - \theta(x^*) + \left\langle w - w^*, \mathcal{J}(w)\right\rangle \geq 0$ since by the monotonicity of $\mathcal{J}(w)$, it follows that $\left\langle w - \bar{w}, \mathcal{J}(w) - \mathcal{J}(\bar{w})\right\rangle \equiv 0$ for any $w, \bar{w} \in \Omega$. Because the solution set of (1) is nonempty, the solution set of (11), denoted by Ω^*, is also nonempty and can be characterized (see [11, Theorem 2.1]) as

$$\Omega^* = \bigcap_{w\in\Omega} \left\{\bar{w} \mid \theta(x) - \theta(\bar{x}) + \left\langle w - \bar{w}, \mathcal{J}(\bar{w})\right\rangle \geq 0\right\}.$$

Analogous to the above discussions, the saddle-point of (2), denoted by $(y^*, z^*, \lambda^*) \in \mathcal{Y} \times \mathcal{Z} \times \mathbb{R}^m$, satisfies the following variational inequality:

$$\theta(u) - \theta(u^*) + \langle w - w^*, \mathcal{J}(w^*)\rangle \geq 0, \quad \forall w \in \Omega,$$

where $\theta(u) = \theta_1(y) + \theta_2(z)$ and

$$w = \begin{pmatrix} u \\ \lambda \end{pmatrix}, \quad u = \begin{pmatrix} y \\ z \end{pmatrix}, \quad \mathcal{J}(w) = \begin{pmatrix} -B^\top\lambda \\ -C^\top\lambda \\ By + Cz - b \end{pmatrix}.$$

3. A novel prediction-correction framework

In this section, we first present a Generalized Prediction-Correction (GPC) framework, which inherits the prediction step in [4] but features a general correction step. Then, we briefly discuss its global convergence and sublinear convergence rate in the ergodic and nonergodic senses.

GPC: generalized prediction-correction framework.

Prediction Step: Given $\rho > 0, \alpha \in \mathbb{R}$, update $\{\tilde{w}^k := (\tilde{x}^k; \tilde{\lambda}^k)\}$ by

$$\tilde{x}^k = \mathbf{prox}_{\rho,\theta}\big(x^k + A^\top\lambda^k/\rho\big); \tag{13a}$$

$$\tilde{\lambda}^k = \lambda^k - \beta\big[A\tilde{x}^k - b + \alpha A(\tilde{x}^k - x^k)\big]. \tag{13b}$$

Correction Step: Determine a nonsingular matrix M and a scalar $\eta > 0$, update $\{w^{k+1} := (x^{k+1}; \lambda^{k+1})\}$ by

$$w^{k+1} = w^k - \eta M(w^k - \tilde{w}^k). \tag{14}$$

Lemma 3.1. *Let $\{\tilde{w}^k\}$ and $\{w^{k+1}\}$ be the prediction sequence and correction sequence generated by GPC. Let Q be defined in (20). Then, under the conditions that matrices*

$$H = QM^{-1} \ \text{and}\ G = Q^\top + Q - \eta M^\top H M \tag{15}$$

are symmetric positive definite, we have $\tilde{w}^k \in \Omega$ and

$$\theta(x) - \theta(\tilde{x}^k) + \left\langle w - \tilde{w}^k, \mathcal{J}(w)\right\rangle \geq \frac{\eta}{2}\|w^k - \tilde{w}^k\|_G^2 + \frac{1}{2}\Big(\|w - w^{k+1}\|_H^2 - \|w - w^k\|_H^2\Big) \tag{16}$$

for any $w \in \Omega$. Moreover, for any $w^* \in \Omega^*$, it holds that

$$\|w^* - w^k\|_H^2 \geq \|w^* - w^{k+1}\|_H^2 + \eta\|w^k - \tilde{w}^k\|_G^2. \tag{17}$$

Proof. According to Lemma 2.1, the first-order optimality condition of the subproblem in (13a) is $\tilde{x}^k \in \mathcal{X}$ and

$$\theta(x)-\theta(\tilde{x}^k)+\left\langle x-\tilde{x}^k, -A^\top \tilde{\lambda}^k+\rho(\tilde{x}^k-x^k)+A^\top(\tilde{\lambda}^k-\lambda^k)\right\rangle \geq 0$$

for any $x \in \mathcal{X}$. Besides, a straightforward rearrangement of (13b) yields

$$\left\langle \lambda-\tilde{\lambda}^k, A\tilde{x}^k-b+\alpha A(\tilde{x}^k-x^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)\right\rangle = 0 \quad (18)$$

for any $\lambda \in \mathbb{R}^m$. Combine these two relationships with the notations in (12) to obtain

$$\begin{aligned} &\theta(x)-\theta(\tilde{x}^k)+\left\langle w-\tilde{w}^k, \mathcal{J}(\tilde{w}^k)\right\rangle \geq \\ &(w-\tilde{w}^k)^\top Q(w^k-\tilde{w}^k) = \frac{1}{\eta}(w-\tilde{w}^k)^\top H(w^k-w^{k+1}), \end{aligned} \quad (19)$$

where the equality follows from (14) and $H = QM^{-1}$ with

$$Q = \begin{bmatrix} \rho\mathbf{I} & A^\top \\ \alpha A & \frac{1}{\beta}\mathbf{I} \end{bmatrix}. \quad (20)$$

By substituting $a = w, e = \tilde{w}^k, c = w^k$ and $d = w^{k+1}$ into the identity $(a-e)^\top H(c-d) = \frac{1}{2}\left(\|a-d\|_H^2 - \|a-c\|_H^2\right) + \frac{1}{2}\left(\|c-e\|_H^2 - \|d-e\|_H^2\right)$ and applying the result to the right-hand side of (19), we have

$$\begin{aligned} &(w-\tilde{w}^k)^\top H(w^k-w^{k+1}) - \frac{1}{2}\left(\|w-w^{k+1}\|_H^2 - \|w-w^k\|_H^2\right) \\ &= \frac{1}{2}\Big(\|w^k-\tilde{w}^k\|_H^2 - \|w^{k+1}-\tilde{w}^k\|_H^2\Big) \\ &= \frac{1}{2}\Big(\|w^k-\tilde{w}^k\|_H^2 - \|w^{k+1}-w^k+w^k-\tilde{w}^k\|_H^2\Big) \\ &\overset{(14)}{=} \frac{1}{2}\Big(\|w^k-\tilde{w}^k\|_H^2 - \|(w^k-\tilde{w}^k)-\eta M(w^k-\tilde{w}^k)\|_H^2\Big) \\ &= \frac{\eta}{2}\big(w^k-\tilde{w}^k\big)^\top\big(Q^\top+Q-\eta M^\top HM\big)\big(w^k-\tilde{w}^k\big) \\ &\overset{(15)}{=} \frac{\eta}{2}\|w^k-\tilde{w}^k\|_G^2. \end{aligned}$$

So, plugging the last relation into (19), the assertion (16) is confirmed.

Now, letting $w = w^*$ in (16) and using (11), we have

$$\|w^*-w^k\|_H^2 - \|w^*-w^{k+1}\|_H^2 - \eta\|w^k-\tilde{w}^k\|_G^2 \geq 0.$$

Consequently, the relation (17) follows immediately. ■

Similar conditions to (15) can be found in e.g. [24, Lemma 2.7], but our algorithm framework is different from that in [24]. Based on Lemma 3.1, the global convergence of GPC and its ergodic and nonergodic convergence rates can be stated as follows. Since the conclusions are the same as that in [4], we omit the detailed proof and summarize the results to avoid redundancy.

Theorem 3.1. *Let* $\{\tilde{w}^k\}$ *and* $\{w^{k+1}\}$ *be the prediction sequence and correction sequence generated by GPC. Under the conditions in Lemma 3.1, we have*

- $\lim_{k\to\infty} \|w^k - w^{k+1}\| = 0$ *and there exists a* $w^\infty \in \Omega^*$ *such that* $\lim_{k\to\infty} w^k = w^\infty$*;*
- *For any* $(x^*; \lambda^*) \in \Omega^*$*, and* $T, \varsigma > 0$*, we have*

$$\begin{cases} |\theta(x_T)-\theta(x^*)| \leq \frac{\gamma_\varsigma}{2(1+T)}, \\ \|Ax_T-b\| \leq \frac{\gamma_\varsigma}{2(1+T)\left(1+\|\lambda^*\|\right)}, \end{cases}$$

  where $x_T = \frac{1}{T+1}\sum_{k=0}^T \tilde{x}^k$ *and* $\gamma_\varsigma = \inf_{x^*\in\mathcal{X}} \sup_{\|\lambda\|\leq\varsigma} \left\|\begin{pmatrix} x^* \\ \lambda \end{pmatrix} - \begin{pmatrix} x^0 \\ \lambda^0 \end{pmatrix}\right\|_H^2$*;*
- *For any integer* $T > 0$*, there exists a constant* $c > 0$ *such that*

$$\|M(w^k-\tilde{w}^k)\|_H^2 \leq \frac{1}{(T+1)c}\|w^0-w^*\|_H^2, \ \forall w^* \in \Omega^*.$$

Remark 3.1. *We give some remarks on the proposed algorithm and its convergence results:*

(i) *The step in* (13b) *features a flexible iteration* $\alpha A(\tilde{x}^k - x^k)$ *for any* $\alpha \in \mathbb{R}$*. This allows for either a forward step or a backward step. For the case where* $\alpha \in (0, 1)$*,* (13b) *exhibits an inertial step; for the case where* $\alpha \in (0, 2)$*, it exhibits a relaxation step. This new prediction step can be applied to other methods, see Section 4.*

(ii) *A feasible approach to construct* M *is to follow the property of* H*. Since* $H = QM^{-1} \succ \mathbf{0}$*, it should be in the form of*

$$H = QD^{-1}Q^\top, \text{ that is, } M = Q^{-\top}D. \quad (21)$$

So, we deduce $G = Q^\top + Q - \eta D$*, where* $D \succ \mathbf{0}$ *is undetermined. Similar to the discussions by He-Yuan [16], we can first select* D *and* η *such that* $Q^\top + Q \succ \eta D \succ \mathbf{0}$*, then construct* H *and* M *via the equalities in* (21)*. Another feasible approach to construct* M *is to follow the property of* G*. By the condition* $G = Q^\top + Q - \eta M^\top HM \succ \mathbf{0}$*, we have* $\Delta = Q^\top + Q - G \succ \mathbf{0}$*, that is,* $\eta M^\top HM = \Delta$*. Combine it with the relationship* $HM = Q$ *to have*

$$M = Q^{-\top}\Delta/\eta \text{ and } H = \eta Q\Delta^{-1}Q^\top.$$

These discussions suggest that one can also construct a matrix Q *such that* $Q + Q^\top \succ \mathbf{0}$*, conversely; to recover the updates of subproblems and present some new first-order methods enjoying a novel prediction-correction framework.*

(iii) *Although we have presented a general prediction-correction framework for solving* (1)*, this new framework can also be applied to the problem* (2) *with convergence results similar to that in Theorem 3.1.*

4. Applications to some first-order methods

In this section, we apply the proposed framework to reformulate several first-order methods, including the previously mentioned schemes in (5), (6), (7) and (9). We find that the convergence analysis under such a reformulation is much simpler than the original analysis, and some of their convergence conditions have been improved.

4.1. Application to indefinite linearized ALM

We first replace the output of (5) by $(\tilde{x}^k, \tilde{\lambda}^k)$. Then, applying our novel prediction-correction framework to (5) results in the following reformulations:

Prediction-correction reformulation of IDL-ALM (5).

Prediction Step:

$$\tilde{x}^k = \arg\min_{x\in\mathcal{X}} \left\{\mathcal{L}_\beta(x,\lambda^k) + \frac{1}{2}\|x - x^k\|^2_{D_0}\right\}; \tag{22a}$$

$$\tilde{\lambda}^k = \lambda^k - \beta\big[A\tilde{x}^k - b + \alpha A(\tilde{x}^k - x^k)\big]. \tag{22b}$$

Correction Step:

$$w^{k+1} = w^k - \eta M(w^k - \tilde{w}^k), \tag{23}$$

where

$$w^k = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix},\ \tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{\lambda}^k \end{pmatrix},\ M = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \alpha\beta\gamma A & \gamma\mathbf{I} \end{bmatrix}. \tag{24}$$

Remark 4.1. *We provide a reasonable explanation for why* M *takes the form in* (24)*. Combine the update of* λ^{k+1} *in* (5) *and the update of* $\tilde{\lambda}^k$ *in* (22b) *to have*

$$\lambda^{k+1} = \lambda^k - \alpha\beta\gamma A(x^k - \tilde{x}^k) - \gamma(\lambda^k - \tilde{\lambda}^k),$$

which together with the replacement $\tilde{x}^k = x^{k+1}$ *gives* $w^{k+1} = w^k - M(w^k - \tilde{w}^k)$ *with* M *being given in* (24)*. So, the new step* (23) *is obtained by attaching a scalar* $\eta > 0$.

Obviously, the above matrix M is nonsingular for any $\gamma \in (0,2), \beta > 0$ and $\alpha \in \mathbb{R}$. According to Lemma 2.1, the first-order optimality condition of the x-subproblem in (22a) is $\tilde{x}^k \in \mathcal{X}$ and

$$\theta(x) - \theta(\tilde{x}^k) + \left\langle x - \tilde{x}^k, -A^\top\lambda^k + \beta A^\top(A\tilde{x}^k - b) + D_0(\tilde{x}^k - x^k)\right\rangle \geq 0,$$

which, by (22b) and $D_0 = D - (1-\tau)\beta A^\top A$, implies

$$\theta(x) - \theta(\tilde{x}^k) + \left\langle x - \tilde{x}^k, -A^\top\tilde{\lambda}^k + \big[D + (\tau - 1 - \alpha)\beta A^\top A\big](\tilde{x}^k - x^k)\right\rangle \geq 0.$$

Combine it with the previous equality (18) to have

$$\theta(x) - \theta(\tilde{x}^k) + \left\langle w - \tilde{w}^k, \mathcal{J}(\tilde{w}^k)\right\rangle \geq (w - \tilde{w}^k)^\top Q(w^k - \tilde{w}^k) \tag{25}$$

with

$$Q = \begin{bmatrix} D + (\tau - 1 - \alpha)\beta A^\top A & \mathbf{0} \\ \alpha A & \frac{1}{\beta}\mathbf{I} \end{bmatrix}.$$

The inequality in (25) is similar to that in (19), and hence the convergence of IDL-ALM (5) can be guaranteed by Lemma 3.1 if the matrices H and G involved in the subsequent proposition are positive definite.

Proposition 4.1. *For any* $\beta > 0, \eta \in \big(0, \min\{2, 2/\gamma\}\big)$ *and* D *satisfying*

$$D \succeq \left[1 - \tau - \frac{(2-\eta\gamma)(2-\eta)}{4} + \frac{\left(\alpha + \frac{(2-\eta\gamma)(2-\eta)}{2}\right)^2}{(2-\eta\gamma)(2-\eta)}\right]\beta A^\top A, \tag{26}$$

with $\alpha \in \mathbb{R}$ *and* $\gamma \in (0,2)$, *the matrices* H *and* G *defined by* (15) *are symmetric positive definite.*

Proof. First of all, we observe that the matrix M in (24) is nonsingular. So, simple algebra makes the two matrices defined in (15) become

$$H = \begin{bmatrix} D + (\tau - 1 - \alpha)\beta A^\top A & \mathbf{0} \\ \mathbf{0} & \frac{1}{\gamma\beta}\mathbf{I} \end{bmatrix}$$

and

$$G = \begin{bmatrix} (2-\eta)D + \big[(2-\eta)(\tau - 1 - \alpha) - \eta\gamma\alpha^2\big]\beta A^\top A & (1-\eta\gamma)\alpha A^\top \\ (1-\eta\gamma)\alpha A & \frac{2-\eta\gamma}{\beta}\mathbf{I} \end{bmatrix}.$$

For any $\beta, \gamma > 0$ and $\alpha \in \mathbb{R}$, the matrix $H \succ \mathbf{0}$ if

$$D \succeq (\alpha + 1 - \tau)\beta A^\top A. \tag{27}$$

By congruence transformations in linear algebra, we know $G \succ \mathbf{0}$ if and only if

$$\begin{bmatrix} (2-\eta)D + \left\{\big[(2-\eta)(\tau - 1 - \alpha) - \eta\gamma\alpha^2\big] - \frac{\alpha^2(1-\eta\gamma)^2}{2-\eta\gamma}\right\}\beta A^\top A & \mathbf{0} \\ \mathbf{0} & \frac{2-\eta\gamma}{\beta}\mathbf{I} \end{bmatrix}$$

is positive definite, which has been ensured by (26). Moreover, it follows from (26) that the condition in (27) holds automatically. This completes the proof. ■

At the end of this subsection, we note that based on the convergence condition in (26), the parameter τ admits the optimal lower bound as in [14]. More precisely, by taking $\alpha = -\frac{(2-\eta\gamma)(2-\eta)}{2}$ and requiring $D \succeq \left(1 - \tau - \frac{(2-\eta\gamma)(2-\eta)}{4}\right)\beta A^\top A$ to be positive definite, we have from (26) that

$$\tau > 1 - \frac{(2-\eta\gamma)(2-\eta)}{4}.$$

By taking $\eta = 1$, it follows from the last inequality that the optimal lower bound of τ is $\frac{2+\gamma}{4}$ which is the smallest lower bound as pointed out in [14]. By taking $\eta\gamma = 1$, the optimal lower bound of τ is $\frac{2+\eta}{4}$ which is still the smallest lower bound. Similarly, by taking $\gamma = 1$, we have $\tau > 1 - \frac{(2-\eta)^2}{4}$, whose lower bound can be smaller than $\frac{2+\eta}{4}$ if $\eta < 1$.

4.2. Application to Chambolle-Pock method

Similar to the discussions in Section 4.1, we first replace the output of (6) by $(\tilde{x}^k, \tilde{\lambda}^k)$. Then, applying the proposed prediction-correction framework to reformulate the popular Chambolle-Pock method (6) results in the following iterations:

Prediction-correction reformulation of Chambolle-Pock method (6).

Prediction Step:

$$\begin{aligned}\tilde{x}^k &= \arg\min_{x\in\mathcal{X}} \left\{\theta(x) - \langle \lambda^k, Ax\rangle + \frac{1}{2r}\|x - x^k\|^2\right\};\\ \tilde{\lambda}^k &= \lambda^k - s\big[A\tilde{x}^k - b + \alpha A(\tilde{x}^k - x^k)\big].\end{aligned}$$

Correction Step: $w^{k+1} = w^k - \eta M(w^k - \tilde{w}^k)$ where

$$w^k = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix},\ \tilde{w}^k = \begin{pmatrix} \tilde{x}^k \\ \tilde{\lambda}^k \end{pmatrix},\ M = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ (\alpha-1)sA & \mathbf{I} \end{bmatrix}. \tag{29}$$

Analogous to the analysis in Section 4.1, we can still obtain the variational inequality (25) but with the matrix Q being replaced by

$$Q = \begin{bmatrix} \frac{1}{r}\mathbf{I} & A^\top \\ \alpha A & \frac{1}{s}\mathbf{I} \end{bmatrix}.$$

So, the convergence of the Chambolle-Pock method (6) can be ensured by Lemma 3.1 if the matrices H and G involved in the following proposition are positive definite.

Proposition 4.2. *For any $\eta \in (0,2), \alpha \in \mathbb{R}$ and positive parameters r and s satisfying*

$$\frac{1}{rs} > \left[1 - \frac{(2-\eta)^2}{4} + \left(\alpha + 1 - 2\eta + \frac{\eta^2}{2}\right)^2\right]\|A^\top A\|, \tag{30}$$

the matrices H and G defined in (15) are symmetric positive definite.

Proof. By the structure of the matrix M in (29), we know it is inverse. And simple algebra shows

$$H = QM^{-1} = \begin{bmatrix} \frac{1}{r}\mathbf{I} + (1-\alpha)sA^\top A & A^\top \\ A & \frac{1}{s}\mathbf{I} \end{bmatrix}$$

and

$$G = \begin{bmatrix} \frac{2-\eta}{r}\mathbf{I} + \alpha(1-\alpha)s\eta A^\top A & [1+(1-\eta)\alpha]A^\top \\ [1+(1-\eta)\alpha]A & \frac{2-\eta}{s}\mathbf{I} \end{bmatrix}.$$

It is not difficult to check $H \succ \mathbf{0}$ for any $\frac{1}{rs} > \alpha\|A^\top A\|$ with $\alpha \in \mathbb{R}$, and $G \succ \mathbf{0}$ for any

$$\frac{(2-\eta)^2}{rs} > \left\{\alpha(\alpha-1)\eta(2-\eta) + \big[1+(1-\eta)\alpha\big]^2\right\}\|A^\top A\|.$$

The last inequality amounts to the inequality in (30). ■

At the end of this subsection, it is clear that by setting $\alpha = 2\eta - 1 - \eta^2/2$, the inequality in (30) reduces to $\frac{1}{rs} > \frac{\eta(4-\eta)}{4}\|A^\top A\|$, which indicates both parameters r and s can be any positive values as η goes to zero. Besides, by taking $\eta = 1$, we will have $\frac{1}{rs} > \frac{3}{4}\|A^\top A\|$ which is the improved lower bound as stated in [15].

4.3. Application to proximal ADMM

Analogous to the discussions in the above subsections, we apply our novel framework to the proximal ADMM (7), yielding the following iterations:

Prediction-correction reformulation of proximal ADMM (7).

Prediction Step:

$$\begin{aligned}\tilde{y}^k &= \arg\min_{y\in\mathcal{Y}} \mathcal{L}_\beta(y, z^k, \lambda^k);\\ \tilde{z}^k &= \arg\min_{z\in\mathcal{Z}} \left\{\mathcal{L}_\beta(\tilde{y}^k, z, \lambda^k) + \frac{1}{2}\|z - z^k\|_D^2\right\};\\ \tilde{\lambda}^k &= \lambda^k - \beta\big[B\tilde{y}^k + Cz^k - b + \alpha B(\tilde{y}^k - y^k) + \alpha C(\tilde{z}^k - z^k)\big].\end{aligned}$$

Correction Step: $w^{k+1} = w^k - \eta M(w^k - \tilde{w}^k)$ where

$$w^k = \begin{pmatrix} y^k \\ z^k \\ \lambda^k \end{pmatrix},\ \tilde{w}^k = \begin{pmatrix} \tilde{y}^k \\ \tilde{z}^k \\ \tilde{\lambda}^k \end{pmatrix},\ M = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \alpha\gamma\beta B & (\alpha-1)\gamma\beta C & \gamma\mathbf{I} \end{bmatrix}. \tag{32}$$

Similar to the analysis in Section 4.1, we can obtain

$$\theta(u) - \theta(\tilde{u}^k) + \left\langle w - \tilde{w}^k, \mathcal{J}(\tilde{w}^k)\right\rangle \geq (w - \tilde{w}^k)^\top Q(w^k - \tilde{w}^k) \tag{33}$$

with

$$\tilde{u}^k = \begin{pmatrix} \tilde{y}^k \\ \tilde{z}^k \end{pmatrix},\ Q = \begin{bmatrix} -\alpha\beta B^\top B & -\alpha\beta B^\top C & \mathbf{0} \\ -\alpha\beta C^\top B & D - (\alpha-1)\beta C^\top C & \mathbf{0} \\ \alpha B & (\alpha-1)C & \frac{1}{\beta}\mathbf{I} \end{bmatrix}.$$

Then, the convergence of (7) will be ensured by Lemma 3.1, if the matrices H and G involved in the following proposition are positive definite.

Proposition 4.3. *For any $\beta > 0, \gamma \in (0,2), \eta \in \left(0, \min\{2, \frac{2}{\gamma}\}\right)$, $\tau > \frac{1-(2-\eta)(2-\eta\gamma)}{(2-\eta)(2-\eta\gamma)}$ and*

$$\frac{1}{1+\tau} - (2-\eta\gamma)(2-\eta) \leq \alpha < 0, \tag{34}$$

the matrices H and G defined in (15) are symmetric positive definite.

Proof. Simple algebra shows the inverse of matrix M given by (32) is

$$M^{-1} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ -\alpha\beta B & -(\alpha-1)\beta C & \frac{1}{\gamma}\mathbf{I} \end{bmatrix}.$$

Consequently, we derive by $D \succeq \tau\beta C^\top C$ that

$$\begin{aligned} H &= \begin{bmatrix} -\alpha\beta B^\top B & -\alpha\beta B^\top C & \mathbf{0} \\ -\alpha\beta C^\top B & D - (\alpha-1)\beta C^\top C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\beta\gamma}\mathbf{I} \end{bmatrix} \\ &\succeq \begin{bmatrix} -\alpha\beta B^\top B & -\alpha\beta B^\top C & \mathbf{0} \\ -\alpha\beta C^\top B & (\tau+1-\alpha)\beta C^\top C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\beta\gamma}\mathbf{I} \end{bmatrix}. \end{aligned}$$

Given $\beta, \gamma > 0$, it is easy to check $H \succ \mathbf{0}$ for any $\alpha < 0$ and $\tau > -1$.

Next, we show the positive definiteness of G under the conditions in Proposition 4.3. We can first obtain

$$G = \begin{bmatrix} -(2-\eta+\eta\gamma\alpha)\alpha\beta B^\top B & -[2-\eta+\eta\gamma(\alpha-1)]\alpha\beta B^\top C & (1-\eta\gamma)\alpha B^\top \\ -[2-\eta+\eta\gamma(\alpha-1)]\alpha\beta C^\top B & (2-\eta)D-[2-\eta+\eta\gamma(\alpha-1)](\alpha-1)\beta C^\top C & (1-\eta\gamma)(\alpha-1)C^\top \\ (1-\eta\gamma)\alpha B & (1-\eta\gamma)(\alpha-1)C & \frac{2-\eta\gamma}{\beta}\mathbf{I} \end{bmatrix}.$$

By congruence transformations, G can be transformed into $\mathrm{diag}(\tilde{G}, \frac{2-\eta\gamma}{\beta}\mathbf{I})$ with

$$\tilde{G} = \begin{bmatrix} -\left(2-\eta+\frac{\alpha}{2-\eta\gamma}\right)\alpha\beta B^\top B & -\left(2-\eta+\frac{\alpha-1}{2-\eta\gamma}\right)\alpha\beta B^\top C \\ -\left(2-\eta+\frac{\alpha-1}{2-\eta\gamma}\right)\alpha\beta C^\top B & (2-\eta)D-\left(2-\eta+\frac{\alpha-1}{2-\eta\gamma}\right)(\alpha-1)\beta C^\top C \end{bmatrix}$$

and moreover

$$\tilde{G} \succeq \beta \begin{bmatrix} B & \\ & C \end{bmatrix}^\top \begin{bmatrix} -\left(2-\eta+\frac{\alpha}{2-\eta\gamma}\right)\alpha & -\left(2-\eta+\frac{\alpha-1}{2-\eta\gamma}\right)\alpha \\ -\left(2-\eta+\frac{\alpha-1}{2-\eta\gamma}\right)\alpha & \left\{(2-\eta)\tau-\left(2-\eta+\frac{\alpha-1}{2-\eta\gamma}\right)(\alpha-1)\right\} \end{bmatrix} \begin{bmatrix} B & \\ & C \end{bmatrix}.$$

For any $\beta > 0, \gamma \in (0,2), \tau > \frac{1-(2-\eta)(2-\eta\gamma)}{(2-\eta)(2-\eta\gamma)}$ and α satisfying (34), the following inequalities hold:

$$\begin{cases} -\left(2-\eta+\frac{\alpha}{2-\eta\gamma}\right)\alpha \geq 0, \\ [(2-\eta)(2-\eta\gamma)+\alpha]\tau-[1-(2-\eta)(2-\eta\gamma)-\alpha] \geq 0. \end{cases}$$

These discussions imply $\tilde{G} \succeq \mathbf{0}$ and hence $G \succ \mathbf{0}$. ■

Remark 4.2. *The admissible range of γ within the proximal ADMM [22] corresponds to*

$$0 < \gamma < \frac{1-\tau+\sqrt{\tau^2+6\tau+5}}{2}. \tag{35}$$

Even though the region in (35) differs from the interval $(0,2)$ in Proposition 4.3, $\lim\limits_{\tau\to+\infty} \frac{1-\tau+\sqrt{\tau^2+6\tau+5}}{2} = 2$. It is worth noting that the parameter τ in our algorithm satisfies the condition $\tau > \frac{1-(2-\eta)(2-\eta\gamma)}{(2-\eta)(2-\eta\gamma)}$. This means that τ can take negative values (when $\gamma \in (1,2), \eta = 1$), while it is nonnegative in [22]. Based on the condition $\tau > \frac{1-(2-\eta)(2-\eta\gamma)}{(2-\eta)(2-\eta\gamma)}$, we obtain

$$\frac{1}{1+\tau}-(2-\eta\gamma)(2-\eta) \leq \alpha < 0,$$

which indicates that the parameter α is non-positive. Additionally, considering the range of τ in Proposition 4.3 and the relation $D \succeq \tau\beta B^\top B$ that D can be positive indefinite because $\tau > \frac{1-(2-\eta)(2-\eta\gamma)}{(2-\eta)(2-\eta\gamma)} > -1$.

4.4. Application to optimal linearized ADMM

At the beginning of this subsection, we emphasize the differences between the proximal ADMM (7) and the optimal linearized ADMM (9):

- The difference in proximal matrix: the proximal matrix in (7) satisfies $D \succeq \tau\beta C^\top C$ for any $\tau \geq 0$, while the proximal matrix in (9) takes the form $D_0 = \tau r\mathbf{I} - \beta C^\top C$ with $r > \beta\|C^\top C\|$ and $\tau \in (0.75, 1)$.
- The difference in dual stepsize: the dual stepsize in (7) belongs to the region as in (8), while it equals 1 in (9).

Based on these observations, the reformulation of (9) is analogous to that in Section 4.3, namely, the reformulation in Section 4.3 with $\gamma = 1$ and D replaced by D_0. For the sake of conciseness, we omit this reformulation.

By similar discussions to Section 4.3, we can still obtain (33) but with

$$Q = \begin{bmatrix} -\alpha\beta B^\top B & -\alpha\beta B^\top C & \mathbf{0} \\ -\alpha\beta C^\top B & D_0-(\alpha-1)\beta C^\top C & \mathbf{0} \\ \alpha B & (\alpha-1)C & \frac{1}{\beta}\mathbf{I} \end{bmatrix}.$$

And the matrices H and G defined by (15) will become

$$H = \begin{bmatrix} -\alpha\beta B^\top B & -\alpha\beta B^\top C & \mathbf{0} \\ -\alpha\beta C^\top B & \tau r\mathbf{I}-\alpha\beta C^\top C & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\beta}\mathbf{I} \end{bmatrix} \tag{36}$$

and

$$G = \begin{bmatrix} \multicolumn{2}{c}{G_{11}} & \begin{matrix} (1-\eta)\alpha B^\top \\ (1-\eta)(\alpha-1)C^\top \end{matrix} \\ (1-\eta)\alpha B & (1-\eta)(\alpha-1)C & \frac{2-\eta}{\beta}\mathbf{I} \end{bmatrix} \tag{37}$$

with

$$G_{11} = \begin{bmatrix} -[2-\eta+\eta\alpha]\alpha\beta B^\top B & -[2-\eta+\eta(\alpha-1)]\alpha\beta B^\top C \\ -[2-\eta+\eta(\alpha-1)]\alpha\beta C^\top B & (2-\eta)D_0-[2-\eta+\eta(\alpha-1)](\alpha-1)\beta C^\top C \end{bmatrix}.$$

Proposition 4.4. *For any $\beta > 0$ and*

$$0 < \eta < 2-\sqrt{-\alpha} \text{ and } \tau > \frac{1}{(2-\eta)^2+\alpha}, \quad \forall \alpha \leq 0, \tag{38}$$

the matrices H in (36) and G in (37) are symmetric positive definite.

Proof. Since $r > \beta\|C^\top C\|$, we can verify that for any $w \neq \mathbf{0}$, the following holds

$$w^\top H w > 0, \qquad \forall \alpha \leq 0, \tau > 0.$$

By some congruence transformations, the matrix G in (37) can be transformed into $\mathrm{diag}(\bar{G}, \frac{2-\eta}{\beta}\mathbf{I})$ with

$$\bar{G} = \begin{bmatrix} -\left(2-\eta+\frac{\alpha}{2-\eta}\right)\alpha\beta B^\top B & -\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)\alpha\beta B^\top C \\ -\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)\alpha\beta C^\top B & (2-\eta)D_0-\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)(\alpha-1)\beta C^\top C \end{bmatrix}$$

and moreover

$$\begin{aligned} \bar{G} &\succeq \begin{bmatrix} -\left(2-\eta+\frac{\alpha}{2-\eta}\right)\alpha\beta B^\top B & -\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)\alpha\beta B^\top C \\ -\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)\alpha\beta C^\top B & (2-\eta)(\tau-1)\beta C^\top C-\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)(\alpha-1)\beta C^\top C \end{bmatrix} \\ &= \beta \begin{bmatrix} B & \\ & C \end{bmatrix}^\top \begin{bmatrix} -\left(2-\eta+\frac{\alpha}{2-\eta}\right)\alpha & -\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)\alpha \\ -\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)\alpha & \left\{(2-\eta)(\tau-1)-\left(2-\eta+\frac{\alpha-1}{2-\eta}\right)(\alpha-1)\right\} \end{bmatrix} \begin{bmatrix} B & \\ & C \end{bmatrix}. \end{aligned}$$

For any $0 < \eta \leq 2-\sqrt{-\alpha}$ with $\alpha \leq 0$, we deduce $-\left(2-\eta+\frac{\alpha}{2-\eta}\right)\alpha \geq 0$. Moreover, for any $\tau > \frac{1}{(2-\eta)^2+\alpha}$ it holds that $[(2-\eta)^2+\alpha](\tau-1)+(2-\eta)^2+(\alpha-1)>0$, which indicates $\bar{G} \succeq \mathbf{0}$ and hence $G \succ \mathbf{0}$. Summarizing the above discussions, matrix $\boldsymbol{H}$ in (36) and matrix $\boldsymbol{G}$ in (37) are symmetric positive definite. ■

Notice that, the lower bound of τ in (38) is different from that specified in (9), but it can approximate the so-called 0.75 as long as we take $\eta = 2-\sqrt{\frac{4}{3}-\alpha}$. For $\eta \in (0, 2-\frac{2}{\sqrt{3}})$, we have $\frac{4}{3}-(2-\eta)^2 < \alpha \leq 0$ which implies $\frac{1}{(2-\eta)^2+\alpha} < 0.75$, and hence τ can be smaller than 0.75.

5. Concluding remarks

In this paper, we develop a new prediction-correction framework for solving linearly constrained convex optimization problems. We have also applied this new framework to reformulate some existing first-order methods to simplify their original convergence analysis. In practice, our prediction-correction framework can also be applied to reformulate other first-order methods for solving two-block or multi-block separable convex optimization problems, such as the Peaceman-Rachford splitting method [10], forward-backward algorithm [5], proximal point method [9], and their stochastic/inexact versions [2, 19, 25]. This is because these methods can be reformulated as similar prediction-correction iterations via variational analysis. These interesting topics may be explored in future work. In addition, whether our proposed prediction-correction framework can be applied to reformulate accelerated primal-dual algorithms with accelerated convergence rates [17, 20, 23] remains an open question.

CRediT authorship contribution statement Jianchao Bai: Writing - original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis. Shuang Rao: Writing - original draft, Methodology, Formal analysis.

Data availability No data was used for the research described in the article.

Acknowledgements This work was supported by the National Natural Science Foundation of China (12471298) and the Open Project Program of Center for Applied Mathematics of Fujian Province (S20250605).

References

[1] J. Bai, J. Li, F. Xu, H. Zhang, Generalized symmetric ADMM for separable convex optimization, Comput. Optim. Appl. 70 (2018) 129-170.

[2] J. Bai, F. Bian, X. Chang, L. Du, Accelerated stochastic Peaceman-Rachford method for empirical risk minimization, J. Oper. Res. Soc. China, 11 (2023) 783-807.

[3] J. Bai, L. Jia, Z. Peng, new insight on augmented Lagrangian method with applications in machine learning, J. Sci. Comput. 99 (2024) Article number 53.

[4] J. Bai, S. Rao, D. Han, R. Sun, Double-proximal augmented Lagrangian methods with improved convergence condition, Optimization Online, https://optimization-online.org/wp-content/uploads/2024/08/ALM-BRHSJan0.pdf, 2024.

[5] J. Bai, Y. Chen, X. Yu, H. Zhang, Generalized asymmetric forward-backward-adjoint algorithms for convex-concave saddle-point problem, J. Sci. Comput. 102 (2025) Article number 80.

[6] A. Beck, First-Order Methods in Optimization, Philadelphia, SIAM, 2017.

[7] A. Chambolle, T. Pock, A first-order primal-dual algorithms for convex problem with applications to imaging, J. Math. Imaging Vis. 40 (2011) 120-145.

[8] R. Glowinski, A. Marroco, Approximation parélements finis d'rdre un et résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires, Rev. Fr. Autom. Inform. Rech. Opér. Anal. Numér. 9 (1975) 41-76.

[9] G. Gu, B. He, X. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl. 59 (2014) 135-161.

[10] Y. Gu, B. Jiang, D. Han, An indefinite-proximal-based strictly contractive Peaceman-Rachford splitting method, J. Comput. Math. 41 (2023) 1017-1040.

[11] B. He, X. Yuan, On the $\mathcal{O}(1/n)$ convergence rate of the Douglas-Rachford alternating direction method, SIAM J. Numer. Anal. 50 (2012) 700-709.

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

[13] B. He, F. Ma, X. Yuan, Optimally linearizing the alternating direction method of multipliers for convex programming, Comput. Optim. Appl. 75 (2020) 361-388.

[14] B. He, F. Ma, X. Yuan, Optimal proximal augmented Lagrangian method and its applicationto fullJacobian splitting for multi-block separable convex minimization problems, IMA J.Numer. Anal. 40 (2020) 1188-1216,

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

[16] B. He, X. Yuan, On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization, Commun. Optim. Theory, 2023 (2023) Article number 28.

[17] X. He, N. Huang, Y. Fang, Non-ergodic convergence rate of an inertial accelerated primal-dual algorithm for saddle point problems, Commun. Nonlinear Sci. Numer. Simulat. 140 (2024) Article number 108289.

[18] M. Hestenes, Multiplier and gradient methods, J. Optim. Theory Appl. 4 (1969) 303-320.

[19] F. Jiang, X. Cai, D. Han, Inexact asymmetric forward-backward-adjoint splitting algorithms for saddle point problems, Numer. Algor. 94 (2023) 479-509.

[20] G. Luo, Q. Yang, A fast symmetric alternating direction method of multipliers, Numer. Math. Theor. Meth. Appl. 13 (2020) 200-219.

[21] F. Ma, M. Ni, A class of customized proximal point algorithms forlinearly constrained convex optimization, Comput. Appl. Math. 37 (2018) 896-911.

[22] F. Ma, Convergence study on the proximal alternating direction method with larger step size, Numer. Algor. 85 (2020) 399-425.

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

[24] M. Tao, Some parallel splitting methods for separable convex programming with the $\mathcal{O}(1/t)$ convergence rate, Pacific J. Optim. 10 (2012) 359-384.

[25] Z. Wu, Y. Song, F. Jiang, Inexact generalized ADMM with relative error criteria for linearly constrained convex optimization problems, Optim. Lett. 18 (2024) 447-470.

[26] Z. Zhu, F. Chen, J. Zhang, Z. Wen, A unified primal-dual algorithm framework for inequality constrained problems, J. Sci. Comput. 97 (2023) Article number 39.