GENERALIZED AFBA ALGORITHM FOR SADDLE-POINT PROBLEMS *

JIANCHAO BAI† AND YANG CHEN‡

Abstract. The convex-concave minimax problem, also known as the saddle-point problem, has been extensively studied from various aspects including the algorithm design, convergence condition and complexity. In this paper, we propose a generalized asymmetric forward-backward-adjoint (G-AFBA) algorithm to solve such a problem by utilizing both the proximal techniques and the interactive information of the primal-dual updates. Except enjoying proximal subproblems, G-AFBA exploits a more relaxed convergence condition, namely, more flexible and larger proximal stepsizes, which could result in significant improvements in numerical performance. We establish the global convergence and $O(1/T)$ convergence rate of G-AFBA in the ergodic and pointwise senses based on an equivalent prediction-correction interpretation, where T denotes the iteration number. The Q-linear and R-linear convergence rates of G-AFBA are also obtained under a specific calmness condition. Additionally, we propose a faster G-AFBA with $O(1/T^2)$ pointwise convergence rate by modifying the quadratic terms of the subproblems in G-AFBA as well as utilizing an adaptive convex combination technique. The relationship between our basic G-AFBA and several well-established methods are also analyzed concisely.

Key words. Saddle-point problem, forward-backward-adjoint splitting method, convergence and complexity

AMS subject classifications. 65K10, 65Y20, 90C25, 94A08

1. Introduction.

Consider the following generic saddle-point problem

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} \mathcal{L}(x,y) := f(x) + \langle Kx, y\rangle - g(y), \tag{1.1}$$

where $f : \mathcal{X} \to (-\infty, \infty]$ and $g : \mathcal{Y} \to (-\infty, \infty]$ are proper lower semicontinuous convex functions (not necessarily smooth), $\mathcal{X}$ and $\mathcal{Y}$ are finite-dimensional real Euclidean spaces, $K : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator. If K^* denotes the adjoint operator of K, f^* and g^* denote the Fenchel conjugate [33] of f and g respectively, then (1.1) amounts to the following primal/dual problem:

$$\min_{x\in\mathcal{X}} f(x) + g^*(Kx) \qquad \text{and} \qquad \min_{y\in\mathcal{Y}} f^*(-K^*y) + g(y).$$

Due to these certain relationships, the problem (1.1) has covered a wide range of practical applications, including machine learning, signal and image processing, economics, statistics, see e.g. [6, 9, 15, 19, 22, 41, 42] and the references therein. As said by He et al. [19] that "the saddle point problem (1.1) includes many special cases such as the optimality condition of the canonical convex programming problem with linear constraints, scientific computing problems, and particularly a number of variational models arising in image reconstruction problems." In this paper, we will study a generalized asymmetric forward-backward-adjoint (G-AFBA) algorithm for solving (1.1), and we assume that the solution set of (1.1) is nonempty.

*Authors are ordered alphabetically. This research was supported by the National Natural Science Foundation of China (12001430), Guangdong Basic and Applied Basic Research Foundation (2023A1515012405), Shaanxi Fundamental Science Research Project for Mathematics and Physics (22JSQ001).

† Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen 518057, P.R. China; School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, P.R. China (jianchaobai@nwpu.edu.cn).

‡ School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, P.R. China (cy1202208@163.com).

1.1. Notation. Let $\mathbb{R}^n$ be the set of n-dimensional Euclidean space equipped with an inner product $\langle\cdot,\cdot\rangle$ and Euclidean norm $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$. Let $\mathbf{I}$ be the identity matrix and $\mathbf{0}$ be zero matrix/vector. Given a self-adjoint (not necessarily positive definite) linear operator H, we denote $\langle x, Hx\rangle$ by $\|x\|_H^2$, that is, $\|x\|_H^2 = x^\top Hx$ with the superscript $^\top$ denoting the transpose. Denote the Euclidean distance from $x \in \mathcal{C}$ to the closed convex set $\mathcal{C}$ by $\operatorname{dist}(x,\mathcal{C}) = \min_{y\in\mathcal{C}} \|x-y\|$, and the G-weighted distance by $\operatorname{dist}_G(x,\mathcal{C}) = \min_{y\in\mathcal{C}} \|x-y\|_G$ where G is a self-adjoint and positive definite linear operator. The symbol $\rho(G)$ denotes the spectral radius of G, while $\lambda_{\min}(G)$ and $\lambda_{\max}(G)$ denote the minimum and maximum eigenvalue of G, respectively.

1.2. Related work. The separable structure of f and g inspires researchers to treat them individually in the algorithm design so as to make full use of the properties of each component objective function. A very earlier and simpler approach for solving (1.1) is the Arrow-Hurwicz method [1] that involves sequential iterations:

$$\text{(PDHG)} \quad \begin{cases} x^{k+1} = \arg\min\left\{\mathcal{L}(x,y^k) + \dfrac{1}{2\tau}\left\|x - x^k\right\|^2 \mid x \in \mathcal{X}\right\}, \\ y^{k+1} = \arg\max\left\{\mathcal{L}(x^{k+1},y) - \dfrac{1}{2\sigma}\left\|y - y^k\right\|^2 \mid y \in \mathcal{Y}\right\}, \end{cases} \tag{1.2}$$

where the parameters $\tau, \sigma > 0$ are often regarded as the corresponding stepsizes of subproblems. The famous Arrow-Hurwicz method was also called primal-dual hybrid gradient method (PDHG) due to the seminal work [42], and it was described [43] as a proximal version of the traditional augmented Lagrangian method (ALM) for some canonical convex programming problems. O'Connor and Vandenberghe [30] showed that PDHG can be viewed as a special case of the Douglas-Rachford splitting algorithm [29] from the perspective of solving the monotone inclusion problem. Another related algorithm based on (1.2) and the popular extrapolation technique (see e.g. [32]) dates back to Chambolle-Pock [6]:

$$\begin{cases} x^{k+1} = \arg\min\left\{\mathcal{L}(x,y^k) + \dfrac{1}{2\tau}\left\|x - x^k\right\|^2 \mid x \in \mathcal{X}\right\}, \\ \overline{x}^{k+1} = x^{k+1} + \alpha(x^{k+1} - x^k), \\ y^{k+1} = \arg\max\left\{\mathcal{L}(\overline{x}^{k+1},y) - \dfrac{1}{2\sigma}\left\|y - y^k\right\|^2 \mid y \in \mathcal{Y}\right\}. \end{cases} \tag{1.3}$$

Here, $\alpha \in [0,1]$ is an extrapolation step, and clearly (1.3) reduces to (1.2) when $\alpha = 0$. It was shown in [6] that (1.3) is closely related to the existing extrapolational gradient method [26] and a preconditioned version of the alternating direction method of multipliers (ADMM, [14]). The connection between (1.3) and the forward-backward splitting method [29] can be found in e.g. [35]. Although the scheme (1.3) enjoys a proximal technique, some counter-examples provided in [20] highlighted that it is not necessarily convergent (when $\alpha = 0$). Moreover, the convergence of (1.3) with $\alpha \in (0,1)$ remains unknown, although its global convergence with $\alpha = 0$ had been established by He et al. [16] by assuming strong convexity on one of the objective functions. So far, the widely used scheme of (1.3) is the case with $\alpha = 1$:

$$\text{(CP-PPA)} \quad \begin{cases} x^{k+1} = \arg\min\left\{\mathcal{L}(x,y^k) + \dfrac{1}{2\tau}\left\|x - x^k\right\|^2 \mid x \in \mathcal{X}\right\}, \\ \overline{x}^{k+1} = 2x^{k+1} - x^k, \\ y^{k+1} = \arg\max\left\{\mathcal{L}(\overline{x}^{k+1},y) - \dfrac{1}{2\sigma}\left\|y - y^k\right\|^2 \mid y \in \mathcal{Y}\right\}. \end{cases} \tag{1.4}$$

The stepsize parameters τ and s should satisfy

$$\frac{1}{\tau\sigma} > L \quad \text{with} \quad L = \rho(K^*K) \tag{1.5}$$

to ensure the global convergence of CP-PPA. To relax the above condition so as to alleviate the burden of choosing stepsize parameters, He et al. [19] recently extended the aforementioned CP-PPA (1.4) to the following generalized version:

$$\text{(GCP-PPA)} \begin{cases} x^{k+1} = \arg\min\left\{\mathcal{L}(x, y^k) + \frac{1}{2\tau}\|x - x^k\|^2 \mid x \in \mathcal{X}\right\}, \\ \overline{x}^{k+1} = x^{k+1} + \alpha(x^{k+1} - x^k), \\ \overline{y}^{k+1} = \arg\max\left\{\mathcal{L}(\overline{x}^{k+1}, y) - \frac{1}{2\sigma}\|y - y^k\|^2 \mid y \in \mathcal{Y}\right\}, \\ y^{k+1} = \overline{y}^{k+1} - (1-\alpha)\sigma A(x^{k+1} - x^k), \end{cases} \tag{1.6}$$

in which $\alpha \in [0, 1]$ and the condition (1.5) was improved to

$$\frac{1}{\tau\sigma} > (1 - \alpha + \alpha^2)L. \tag{1.7}$$

Obviously, the scheme (1.6) reduces to CP-PPA when $\alpha = 1$ and hence it looks more general and flexible than the previous. Especially, if $\alpha = 0.5$, then the condition (1.7) becomes $\frac{1}{\tau\sigma} > 0.75L$, meaning that the lower bound is improved by 25% compared to (1.5). Except the above deterministic methods, some stochastic and accelerated first-order methods have been well-studied when the objective functions of (1.1) have special structures, see e.g. [8, 9, 22, 27, 37, 39, 45] to list a few.

As a generation of (1.3), the Condat-Vũ scheme proposed independently by Condat [10] and Vũ [35] have attracted much attention in recent years and its convergence can be proved by casting the scheme in a forward-backward splitting method. However, the condition of involved parameters is more restrictive than that of PDHG due to the presence of an additional smooth function. A powerful method related to the Condat-Vũ scheme is the asymmetric forward-backward-adjoint splitting (AFBA) method [27] which is used to solve monotone inclusion problems involving three terms, a maximally monotone, a cocoercive and a bounded linear operator. As commented by Jiang-Cai-Han [24], “AFBA has a conservative stepsize range that is even more restrictive than that of the Condat-Vũ scheme.” Hence, they proposed an inexact AFBA with absolute error criteria and achieved the global $O(1/T^2)$ convergence rate under the assumption that a part of (both) the underlying functions are strongly convex, where T denotes the iteration number. To our regret, the parameter condition in [24, Algorithm 2] is even strict than (1.7). For a recent survey on some proximal splitting algorithms, we refer to [11] for more details.

1.3. The algorithm and contribution. Notice that, the update of y^{k+1} in (1.6) uses its previous iteration and the iterative residual of x-variable. A natural question is that can we correct the $\overline{x}^{k+1}$ based on the iterative residual of y-variable to ensure convergence but under a more relaxed condition? Since the numerical performance of PDHG-type methods depend on the choice of the stepsize parameters (τ, σ), motivated by the simplicity of the AFBA method and the mentioned question to employ even larger stepsizes, we propose the following generalized AFBA framework:

$$\text{(G-AFBA)} \quad \begin{cases} \overline{x}^{k+1} = \arg\min_{x\in\mathcal{X}} \Big\{ f(x) + \frac{1}{2\tau}\big\|x - x^k + \tau K^* y^k\big\|^2 \Big\}, \\ \overline{y}^{k+1} = \arg\min_{y\in\mathcal{Y}} \Big\{ g(y) + \frac{1}{2\sigma}\big\|y - y^k - \sigma K[\overline{x}^{k+1} + \alpha(\overline{x}^{k+1} - x^k)]\big\|^2 \Big\}, \\ x^{k+1} = \overline{x}^{k+1} - (1-\alpha)\mu\, \tau K^*(\overline{y}^{k+1} - y^k), \\ y^{k+1} = \overline{y}^{k+1} + (1-\alpha)(1-\mu)\, \sigma K(\overline{x}^{k+1} - x^k), \end{cases} \tag{1.8}$$

where $\alpha, \mu \in [0,1]$, and the positive parameters (τ, σ) satisfy

$$\frac{1}{\tau\sigma} > \frac{\alpha - (-1+\mu-\mu^2)(1-\alpha)^2 + \sqrt{[(-1+\mu-\mu^2)(1-\alpha)^2+\alpha]^2 + 4\alpha(1-\alpha)^2}}{2} L. \tag{1.9}$$

Contributions of our work are summarized the following three aspects:

♠ **Flexibility of the algorithm.** Table 1.1 shows certain connections of our proposed G-AFBA (1.8) with several well-established methods in terms of the convergence rate and region of stepsizes parameters (τ, σ), which indicates that G-AFBA is more general and more flexible than some state-of-the-art methods. We refer to Section 5 for detailed analysis on the relationship between G-AFBA and existing methods, as well as the application to multi-block separable convex programming. Actually, the main difference between G-AFBA (1.8) and existing PDHG-type methods lies in the last two steps in (1.8), and these steps can be regarded as a correction step as in (3.2) which uses interactive information of each update. Based on the main subproblems of G-AFBA, we also discuss a faster G-AFBA, an adaptive G-AFBA and a linearized G-AFBA in Section 4 and Section 5, respectively.

Cases	Algorithms	Region of (τ,σ)	Pointwise rate	Linear rate
$\alpha = 1$	CP-PPA [6] & Reduced ALM	(1.5)	×	×
$(\alpha,\mu) = (0,1)$	Exact version of Algorithm 2 [24]	(1.5)	×	×
$\mu = 0$	GCP-PPA [19]	(1.7)	✓	×
$\alpha,\mu \in [0,1]$	G-AFBA(ours)	(1.9)	✓	✓
$\alpha = 0$	G-AFBA1(ours)	(5.4)	✓	✓

Table 1.1: Relationship between G-AFBA (1.8) and several existing methods.

♠ **Larger stepsize parameters.** Figure 1.1 visualizes the region of $\frac{1}{\tau\sigma L}$ in (1.9) and (1.7), where Figure 1.1(b) is the same as Figure 1.1(a) but at different azimuth and elevation angles. As shown in Figure 1.1, the region of $\frac{1}{\tau\sigma L}$ in (1.9) is larger than that in (1.7), hence the current 0.75 in e.g. [19, 25, 28] is *no longer the tight upper bound* for the stepsizes (τ, σ), that is to say, users have the flexibility on choosing larger stepsizes when applying our proposed G-AFBA (1.8). For example, by setting $(\alpha, \mu) = (0.33, 0.47)$, the condition (1.9) reduces to $\frac{1}{\tau\sigma} > 0.7185L$. From the convergence analysis of G-AFBA, it seems that the stepsizes can be enlarged by exploiting additional correction step based on interactive information, and this technique may be also used to investigate larger parameters in [2, 17, 18].

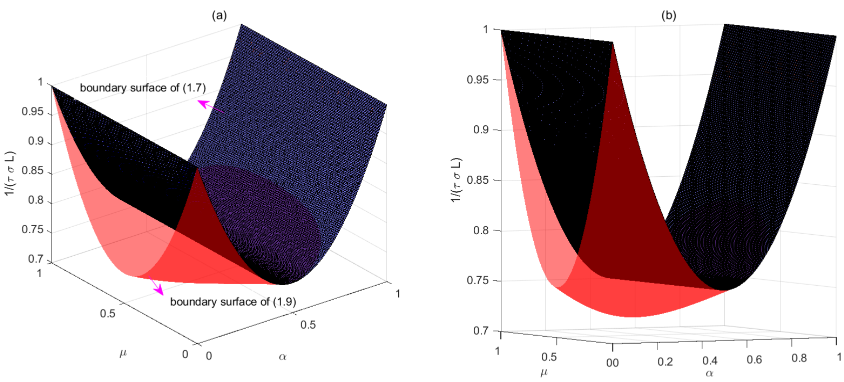

Fig. 1.1: Visualization on the region of $\frac{1}{\tau\sigma L}$ in (1.9) and (1.7).

♠ **Global convergence and various convergence rates.** To simplify the convergence of G-AFBA (1.8), we interpreted it as a prediction-correction framework and characterize its generated sequence as a similar variational inequality to the characterization of the saddle-point of (1.1). In addition to establishing the familiar global convergence of G-AFBA (1.8) and its $O(1/T)$ ergodic convergence rate, we also establish its $O(1/T)$ convergence rate in the pointwise sense and linear convergence rate under proper assumption. Besides, we also discuss a faster G-AFBA with $O(1/T^2)$ convergence rate.

1.4. Organization of the paper. In Section 2, we prepare some preliminaries that are used to analyze the convergence of the proposed method. Section 3 is dedicated to analyzing the global convergence and sublinear/linear convergence rate based on an equivalent prediction-correction interpretation for the proposed G-AFBA (1.8). Section 4 presents a modified G-AFBA with faster convergence guaranteed. In section 5, we analyze the relationship of our G-AFBA (1.8) and some existing methods, extend G-AFBA to an adaptive version with more relaxed condition and the multi-block separable convex optimization problem.

2. Preliminaries.

In this section, we provide a variational characterization for the saddle-point of (1.1) and some block-matrices with nice properties. These preliminaries are prepared for showing the convergence of G-AFBA.

2.1. Variational characterization. We call a point $(x^*, y^*) \in \Omega := \mathcal{X} \times \mathcal{Y}$ the saddle-point of (1.1) if it satisfies

$$\mathcal{L}_{y\in\mathcal{Y}}(x^*, y) \leq \mathcal{L}(x^*, y^*) \leq \mathcal{L}_{x\in\mathcal{X}}(x, y^*),$$

that is,

$$\begin{cases} f(x) - f(x^*) + \left\langle x - x^*, K^* y^* \right\rangle \geq 0, & \forall x \in \mathcal{X}, \\ g(y) - g(y^*) + \left\langle y - y^*, -Kx^* \right\rangle \geq 0, & \forall y \in \mathcal{Y}. \end{cases} \tag{2.1}$$

These inequalities can be expressed as the following mixed variational inequality

$$\mathrm{VI}(\theta, \mathcal{J}, \Omega): \quad \theta(u) - \theta(u^*) + \left\langle u - u^*, \mathcal{J}(u^*) \right\rangle \geq 0, \quad \forall u \in \Omega, \tag{2.2}$$

in which

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad \theta(u) = f(x) + g(y), \quad \mathcal{J}(u) = \begin{pmatrix} K^*y \\ -Kx \end{pmatrix}. \tag{2.3}$$

Notice that the above operator $\mathcal{J}(u)$ is monotone and satisfies

$$\big\langle u - v, \mathcal{J}(u) - F(v)\big\rangle \equiv 0, \qquad \forall u, v \in \Omega.$$

In the convex optimization, u^* satisfies (2.2) if and only if u^* is a primal-dual solution of the problem (1.1). Because of the nonempty assumption on the solution set of (1.1), the solution set of $\mathrm{VI}(\theta, \mathcal{J}, \Omega)$, denoted by Ω^*, is also nonempty.

2.2. Some matrices and properties. In order to simplify and concisely analyze convergence of G-AFBA, we introduce the following matrices

$$Q = \begin{bmatrix} \frac{1}{\tau}\mathbf{I} & -K^* \\ -\alpha K & \frac{1}{\sigma}\mathbf{I} \end{bmatrix}, \quad M = \begin{bmatrix} \mathbf{I} & -(1-\alpha)\mu\tau K^* \\ (1-\alpha)(1-\mu)\sigma K & \mathbf{I} \end{bmatrix}. \tag{2.4}$$

The matrix M is nonsingular if $1 + (1-\alpha)^2(1-\mu)\mu\tau\sigma L > 0$, which has been ensured by the condition (1.9) [see also the sequel (2.8)]. Given these matrices, we define

$$H := QM^{-1} \quad \text{and} \quad G := Q^\top + Q - M^\top HM. \tag{2.5}$$

and show the following properties under mild conditions.

Proposition 2.1. *The matrices H and G defined in (2.5) are positive definite under the condition (1.9).*

Proof. Followed by the surprising techniques [21, Section 3.1] to ensure the positive definiteness of H and G, we just need to analyze the conditions to ensure the positive definiteness of

$$D = Q^\top M \quad \text{and} \quad G = Q^\top + Q - D. \tag{2.6}$$

In practice, if H is a positive definite matrix, then there exists a positive definite matrix D such that

$$H = QD^{-1}Q^\top,$$

and vice versa when D is positive definite. Compare this relationship with the definition of H in (2.5) to immediately obtain $M^{-1} = D^{-1}Q^\top$ and hence to confirm the relationships in (2.6).

Next, we analyze the conditions to ensure the positive definiteness of the two matrices given in (2.6). It follows from (2.4) and (2.6) that

$$D = \begin{bmatrix} \frac{1}{\tau}\mathbf{I} - \alpha(1-\alpha)(1-\mu)\sigma K^*K & -\big[\alpha + (1-\alpha)\mu\big]K^* \\ -\big[\alpha + (1-\alpha)\mu\big]K & \frac{1}{\sigma}\mathbf{I} + (1-\alpha)\mu\tau KK^* \end{bmatrix}$$

and

$$G = \begin{bmatrix} \frac{1}{\tau}\mathbf{I} + \alpha(1-\alpha)(1-\mu)\sigma K^*K & \big[(1-\alpha)\mu - 1\big]K^* \\ \big[(1-\alpha)\mu - 1\big]K & \frac{1}{\sigma}\mathbf{I} - (1-\alpha)\mu\tau KK^* \end{bmatrix}.$$

For any $\tau, \sigma > 0$ and $L = \rho(K^*K) > 0$, the matrix D is positive definite if

$$\Big(\frac{1}{\tau} - \alpha(1-\alpha)(1-\mu)\sigma L\Big)\Big(\frac{1}{\sigma} + (1-\alpha)\mu\tau L\Big) - \Big[\alpha + (1-\alpha)\mu\Big]^2 L > 0 \quad (2.7)$$

$$\Longleftrightarrow \Big\{(1-\alpha)\mu - (1-\alpha)(1-\mu)\alpha - \big[\alpha^2 + 2\alpha(1-\alpha)\mu + (1-\alpha)^2\mu^2\big]\Big\}\frac{L}{\tau\sigma} + \frac{1}{(\tau\sigma)^2} - (1-\alpha)^2(1-\mu)\mu\alpha L^2 > 0,$$

$$\Longleftrightarrow \quad \frac{1}{(\tau\sigma)^2} + \big[(1-\mu)\mu(1-\alpha)^2 - \alpha\big]\frac{L}{\tau\sigma} - (1-\alpha)^2(1-\mu)\mu\alpha L^2 > 0,$$

$$\Longleftrightarrow \quad \Big[\frac{1}{\tau\sigma} + (1-\alpha)^2(1-\mu)\mu L\Big]\Big[\frac{1}{\tau\sigma} - \alpha L\Big] > 0. \quad (2.8)$$

In fact, the condition (2.7) is a sufficient condition to ensure the positive definiteness of D. Suppose K is a $m \times n (m \le n)$ dimensional operator matrix and let $K = S\Sigma U^\top$ be the singular value decomposition of K, where both $S \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\Sigma = (\Sigma_m, \mathbf{0})$ is a diagonal matrix with $\Sigma_m = \mathrm{diag}(s_1, s_2, \cdots, s_m) \in \mathbb{R}^{m\times m}$ and $s_i (i = 1, 2, \cdots, m)$ being the nonzero singular values of K. Then, we have

$$K^*K = U\begin{bmatrix} \Sigma_m^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}U^\top \quad \textit{and} \quad KK^* = S\Sigma_m^2 S^\top,$$

and hence

$$D = \begin{bmatrix} U & \mathbf{0} \\ \mathbf{0} & S\end{bmatrix}\underbrace{\begin{bmatrix} \frac{1}{\tau}\mathbf{I} - \alpha(1-\alpha)(1-\mu)\sigma\Sigma_m^2 & \mathbf{0} & -\big[\alpha + (1-\alpha)\mu\big]\Sigma_m \\ \mathbf{0} & \frac{1}{\tau}\mathbf{I} & \mathbf{0} \\ -\big[\alpha + (1-\alpha)\mu\big]\Sigma_m & \mathbf{0} & \frac{1}{\sigma}\mathbf{I} + (1-\alpha)\mu\tau\Sigma_m^2 \end{bmatrix}}_{P}\begin{bmatrix} U & \mathbf{0} \\ \mathbf{0} & S\end{bmatrix}^\top$$

Similar way to the above analysis can be found in e.g. [36, Page 16]. Since $\rho(K^*K) = \rho(KK^*) = \max_i s_i^2 > 0$ and $\tau > 0$, we have by the condition (2.7) that the matrix P is clearly positive definite.

Analogously, we can derive the following sufficient condition to ensure the positive definiteness of the matrix G:

$$\frac{1}{\tau\sigma} > \frac{\alpha - (-1+\mu-\mu^2)(1-\alpha)^2 + \sqrt{[(-1+\mu-\mu^2)(1-\alpha)^2 + \alpha]^2 + 4\alpha(1-\alpha)^2}}{2}L,$$

in other words,

$$\frac{1}{(\tau\sigma)^2} + \Big[(-1+\mu-\mu^2)(1-\alpha)^2 - \alpha\Big]\frac{L}{\tau\sigma} - (1-\alpha)^2(1-\mu)\mu\alpha L^2 > 0$$

$$\Longleftrightarrow \Big[\frac{1}{\tau\sigma} + (1-\alpha)^2(1-\mu)\mu L\Big]\Big[\frac{1}{\tau\sigma} - \alpha L\Big] > (1-\alpha)^2\frac{L}{\tau\sigma}.$$

Notice that, if the last inequality holds, then the condition (2.8) also holds. Namely, if G is positive definite, then D is also positive definite. The proof is completed. □

3. Convergence analysis. In this section, we first reformulate G-AFBA (1.8) as the following prediction-correction framework by the previous notations u, M and the proximal operator of τf defined as

$$\mathrm{prox}_{\tau f}(y) := \arg\min_{x\in\mathcal{X}}\Big\{f(x) + \frac{1}{2\tau}\|x-y\|^2\Big\}, \quad \forall \tau > 0,$$

Then, we establish the global convergence of G-AFBA and its sublinear convergence rate in both the ergodic and pointwise senses. Finally, the linear convergence rate of G-AFBA is investigated under proper assumptions.

A prediction-correction reformulation of G-AFBA for solving (1.1).

Prediction Step:

$$\widetilde{x}^k = \mathrm{prox}_{\tau f}\big(x^k - \tau K^* y^k\big); \tag{3.1a}$$

$$\widetilde{y}^k = \mathrm{prox}_{\sigma g}\big(y^k + \sigma K[\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)]\big); \tag{3.1b}$$

Correction Step:

$$u^{k+1} = u^k - M(u^k - \widetilde{u}^k). \tag{3.2}$$

3.1. Global convergence. The global convergence of G-AFBA will be analyzed based on the above prediction-correction reformulation.

LEMMA 3.1. *Let $\{\widetilde{u}^k = (\widetilde{x}^k; \widetilde{y}^k)\}$ be the predictor generated by (3.1a)-(3.1b) and $\{u^{k+1} = (x^{k+1}; y^{k+1})\}$ be the corrector generated by (3.2). Then, for any $u \in \Omega$, the following inequality*

$$\mathcal{L}(x, \widetilde{y}^k) - \mathcal{L}(\widetilde{x}^k, y) \geq (u - \widetilde{u}^k)^\top Q(u^k - \widetilde{u}^k) \tag{3.3}$$

holds with the matrix Q given by (2.4).

Proof. We can derive from the first-order optimality condition of (3.1a) that

$$f(x) - f(\widetilde{x}^k) + \big\langle x - \widetilde{x}^k, K^* y^k + \frac{1}{\tau}(\widetilde{x}^k - x^k)\big\rangle \geq 0, \quad \forall x \in \mathcal{X}.$$

Rearranging the last inequality to obtain

$$f(x) - f(\widetilde{x}^k) + \big\langle x - \widetilde{x}^k, K^* \widetilde{y}^k\big\rangle \geq \Big\langle x - \widetilde{x}^k, \frac{1}{\tau}(x^k - \widetilde{x}^k) - K^*(y^k - \widetilde{y}^k)\Big\rangle \tag{3.4}$$

for any $x \in \mathcal{X}$. Similarly, we have from (3.1b) that

$$g(y) - g(\widetilde{y}^k) + \big\langle y - \widetilde{y}^k, -K[\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)] + \frac{1}{\sigma}(\widetilde{y}^k - y^k)\big\rangle \geq 0, \quad \forall y \in \mathcal{Y},$$

which can be equivalently rewritten as

$$g(y) - g(\widetilde{y}^k) + \big\langle y - \widetilde{y}^k, -K\widetilde{x}^k\big\rangle \geq \Big\langle y - \widetilde{y}^k, -\alpha K(x^k - \widetilde{x}^k) + \frac{1}{\sigma}(y^k - \widetilde{y}^k)\Big\rangle \tag{3.5}$$

for any $y \in \mathcal{Y}$. Combining (3.4) and (3.5) completes[1] the proof of (3.3). □

The following lemma provides a more detailed analysis on the crossing term $(u - \widetilde{u}^k)^\top Q(u^k - \widetilde{u}^k)$ in some quadratic terms, and finally shows that the sequences $\{u^* - u^k\}$ is contractive under the weighted norm $\|u\|_H = \sqrt{u^\top H u}$.

[1]We can also get $\theta(u) - \theta(\widetilde{u}^k) + \langle u - \widetilde{u}^k, \mathcal{J}(\widetilde{u}^k)\rangle \geq (u - \widetilde{u}^k)^\top Q(u^k - \widetilde{u}^k)$, which is similar to the variational inequality in (2.2). Anyway, the iterative sequence generated by G-AFBA can be characterized as a similar characterization form for the saddle-point of (1.1).

LEMMA 3.2. *Under the condition (1.9), the iterates $\widetilde{u}^k$ and u^{k+1} satisfy*

$$\mathcal{L}(x,\widetilde{y}^k)-\mathcal{L}(\widetilde{x}^k,y)\geq\frac{1}{2}\big(\|u-u^{k+1}\|_H^2-\|u-u^k\|_H^2\big)+\frac{1}{2}\|u^k-\widetilde{u}^k\|_G^2 \tag{3.6}$$

for any $u\in\Omega$, where H and G are defined in (2.5). Moreover, we have

$$\|u^*-u^k\|_H^2\geq\|u^*-u^{k+1}\|_H^2+\|u^k-\widetilde{u}^k\|_G^2,\quad\forall u^*\in\Omega^*. \tag{3.7}$$

Proof. According to (3.2) and the definition of H in (2.5), we have

$$(u-\widetilde{u}^k)^\top Q(u^k-\widetilde{u}^k)=(u-\widetilde{u}^k)^\top H(u^k-u^{k+1}). \tag{3.8}$$

Then, applying the identity

$$(a-b)^\top H(c-d)=\frac{1}{2}\Big\{\|a-d\|_H^2-\|a-c\|_H^2\Big\}+\frac{1}{2}\Big\{\|c-b\|_H^2-\|d-b\|_H^2\Big\}$$

with $a=u$, $b=\widetilde{u}^k$, $c=u^k$ and $d=u^{k+1}$ to the right-hand side of (3.8) gives

$$\begin{aligned}
&(u-\widetilde{u}^k)^\top H(u^k-u^{k+1})-\frac{1}{2}\Big\{\|u-u^{k+1}\|_H^2-\|u-u^k\|_H^2\Big\}\\
=&\frac{1}{2}\Big\{\|u^k-\widetilde{u}^k\|_H^2-\|u^{k+1}-\widetilde{u}^k\|_H^2\Big\}\\
=&\frac{1}{2}\Big\{\|u^k-\widetilde{u}^k\|_H^2-\|u^{k+1}-u^k+(u^k-\widetilde{u}^k)\|_H^2\Big\}\\
\overset{(3.2)}{=}&\frac{1}{2}\Big\{\|u^k-\widetilde{u}^k\|_H^2-\|(u^k-\widetilde{u}^k)-M(u^k-\widetilde{u}^k)\|_H^2\Big\}\\
=&\frac{1}{2}\Big\{(u^k-\widetilde{u}^k)^\top(Q^\top+Q-M^\top HM)(u^k-\widetilde{u}^k)\Big\}\overset{(2.5)}{=}\frac{1}{2}\|u^k-\widetilde{u}^k\|_G^2,
\end{aligned} \tag{3.9}$$

where the fourth equality exploits the relation $Q=HM$ and its symmetric property. Then, substituting (3.8) and (3.9) into (3.3) confirms the assertion (3.6).

Set $u=u^*$ in (3.6) and use (2.1) with $(x,y)=(\widetilde{x}^k,\widetilde{y}^k)$ to obtain

$$\|u^*-u^k\|_H^2-\|u^*-u^{k+1}\|_H^2-\|u^k-\widetilde{u}^k\|_G^2\geq 2\big[\mathcal{L}(\widetilde{x}^k,y^*)-\mathcal{L}(x^*,\widetilde{y}^k)\big]\geq 0.$$

That is, the assertion (3.7) follows directly. □

In what follows, we are ready to prove the global convergence of the proposed G-AFBA based on the key inequality (3.7) in Lemma 3.2.

THEOREM 3.3. *The sequence $\{u^{k+1}\}$ generated by G-AFBA under the condition (1.9) converges to a solution point of (1.1).*

Proof. On the one hand, it follows from (3.7) and the positive definiteness of G and H that the sequence $\{u^k\}$ is bounded

$$\lim_{k\to\infty}\|u^k-\widetilde{u}^k\|=0. \tag{3.10}$$

As a result, the sequence $\{\widetilde{u}^k\}$ is also bounded and has at least one cluster point. Let u^∞ be a cluster point of $\{\widetilde{u}^k\}$ and $\{\widetilde{u}^{k_j}\}$ be a subsequence converging to u^∞. Then, it follows from (3.3) that

$$\theta(u)-\theta(\widetilde{u}^{k_j})+\langle u-\widetilde{u}^{k_j},\mathcal{J}(\widetilde{u}^{k_j})\rangle\geq(u-\widetilde{u}^{k_j})^\top Q(u^{k_j}-\widetilde{u}^{k_j}),\quad\forall u\in\Omega,$$

which, together with (3.10) and the continuity of $\theta(u)$ and $\mathcal{J}(u)$, implies

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

That is to say, u^∞ is a solution point of (2.2) and hence is a solution point of (1.1).

On the other hand, using ((3.10) and $\lim_{j\to\infty} u^{k_j}=u^\infty$, the sequence u^{k_j} also converges to u^∞. For any $k>k_j$, we deduce from (3.7) that

$$\big\|u^\infty-u^{k_j}\big\|_H\geq\big\|u^\infty-u^k\big\|_H$$

So, the sequence $\{u^k\}$ converges to u^∞. The proof is now complete. ∎

3.2. Sublinear rate of convergence. In this section, we aim at analyzing the worst-case $\mathcal{O}(1/T)$ convergence rate in the ergodic and pointwise senses for G-AFBA, where T denotes the number of iterations.

It is evident that (2.1) can be also expressed as

$$\mathcal{L}(x,y^*)-\mathcal{L}(x^*,y)\geq 0,\quad \forall (x,y)\in\Omega.$$

So, we can define $\overline{u}=(\overline{x};\overline{y})$ as an ϵ-approximate solution to (1.1) if it holds $\mathcal{L}(\overline{x},y)-\mathcal{L}(x,\overline{y})\leq\epsilon$ for any $u\in\mathcal{T}_{\overline{u}}=\{u\in\Omega^*\mid \|u-\overline{u}\|\leq 1\}$ and given accuracy $\epsilon>0$. In the following, we will demonstrate that, after T iterations, we are able to find a point $\overline{u}$ generated by G-AFBA such that

$$\sup_{u\in\mathcal{T}_{\overline{u}}}\big\{\mathcal{L}(\overline{x},y)-\mathcal{L}(x,\overline{y})\big\}\leq\epsilon:=\mathcal{O}(1/T).$$

THEOREM 3.4. *Let $\{\widetilde{u}^k\}$ be the predictor generated by (3.1a)-(3.1b) and $\{u^{k+1}\}$ be the corrector generated by (3.2). For any integer $T>0$ and $\kappa\geq 0$, let*

$$x_T=\frac{1}{T}\sum_{k=\kappa}^{T+\kappa}\widetilde{x}^k \quad\text{and}\quad y_T=\frac{1}{T}\sum_{k=\kappa}^{T+\kappa}\widetilde{y}^k. \tag{3.11}$$

Then, under the condition (1.9) we have

$$\mathcal{L}(x_T,y)-\mathcal{L}(x,y_T)\leq\frac{1}{2(T+1)}\big\|u-u^\kappa\big\|_H^2,\quad \forall u\in\Omega. \tag{3.12}$$

Proof. The inequality (3.6) together with the positive definiteness of G implies

$$\mathcal{L}(\widetilde{x}^k,y)-\mathcal{L}(x,\widetilde{y}^k)\leq\frac{1}{2}\big\{\big\|u-u^k\big\|_H^2-\big\|u-u^{k+1}\big\|_H^2\big\}$$

for any $u\in\Omega$. Sum the last inequality over $k=\kappa,1,\cdots,T+\kappa$ to obtain

$$\sum_{k=\kappa}^{T+\kappa}\big[\mathcal{L}(\widetilde{x}^k,y)-\mathcal{L}(x,\widetilde{y}^k)\big]\leq\frac{1}{2}\big\|u-u^\kappa\big\|_H^2,$$

which, by the convexity of f,g and the notations in (3.11), shows

$$(T+1)\big[\mathcal{L}(x_T,y)-\mathcal{L}(x,y_T)\big]\leq\frac{1}{2}\big\|u-u^\kappa\big\|_H^2.$$

Then, the assertion (3.12) is confirmed. ∎

A similar result to (3.12) in the sense of expectation can be found in [4]. Theorem 3.4 show that under a more flexible condition (1.9), the primal-dual gap of the ergodic sequence generated by G-AFBA converges to zero with the worst-case $\mathcal{O}(1/T)$ convergence rate. Next, we will establish the worst-case $\mathcal{O}(1/T)$ convergence rate measured by the iteration complexity for G-AFBA based on the monotonicity of the sequence $\{\|u^k - u^{k+1}\|^2\}$ as follows.

THEOREM 3.5. *Under the condition (1.9), the predictor $\{\widetilde{u}^k\}$ generated by (3.1a)-(3.1b) and the corrector $\{u^{k+1}\}$ generated by (3.2) satisfy*

$$\left\|u^k - \widetilde{u}^k\right\|^2 \geq \left\|u^{k+1} - \widetilde{u}^{k+1}\right\|^2. \tag{3.13}$$

Proof. It follows from (3.3) with $u = \widetilde{u}^{k+1}$ that

$$\mathcal{L}(\widetilde{x}^{k+1}, \widetilde{y}^k) - \mathcal{L}(\widetilde{x}^k, \widetilde{y}^{k+1}) \geq (\widetilde{u}^{k+1} - \widetilde{u}^k)^\top Q(u^k - \widetilde{u}^k). \tag{3.14}$$

Similarly, (3.3) is true for the $(k+1)$-th iteration, that is,

$$\mathcal{L}(x, \widetilde{y}^{k+1}) - \mathcal{L}(\widetilde{x}^{k+1}, y) \geq (u - \widetilde{u}^{k+1})^\top Q(u^{k+1} - \widetilde{u}^{k+1})$$

which, by setting $u = \widetilde{u}^k$, results in

$$\mathcal{L}(\widetilde{x}^k, \widetilde{y}^{k+1}) - \mathcal{L}(\widetilde{x}^{k+1}, \widetilde{y}^k) \geq (\widetilde{u}^k - \widetilde{u}^{k+1})^\top Q(u^{k+1} - \widetilde{u}^{k+1}). \tag{3.15}$$

Combine (3.14) and (3.15) to have

$$(\widetilde{u}^k - \widetilde{u}^{k+1})^\top Q\big\{(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\} \geq 0. \tag{3.16}$$

Then, adding the equality

$$\begin{aligned} &\big\{(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\}^\top Q\big\{(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\} \\ =&\frac{1}{2}\big\|u^k - \widetilde{u}^k - (u^{k+1} - \widetilde{u}^{k+1})\big\|^2_{(Q^\top + Q)} \end{aligned} \tag{3.17}$$

to both sides of (3.16) leads to

$$\begin{aligned} &\frac{1}{2}\big\|u^k - \widetilde{u}^k - (u^{k+1} - \widetilde{u}^{k+1})\big\|^2_{(Q^\top + Q)} \\ \leq&(u^k - u^{k+1})^\top Q\big\{(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\} \\ \overset{(3.2)}{=}&(u^k - \widetilde{u}^k)^\top M^\top Q\big\{(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\} \\ \overset{(2.4)}{=}&(u^k - \widetilde{u}^k)^\top M^\top HM\big\{(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\}. \end{aligned}$$

Using this relationship, we deduce by the following identity $\|a\|_H^2 - \|b\|_H^2 = 2a^\top H(a - b) - \|a-b\|_H^2$ with $a = M(u^k - \widetilde{u}^k)$ and $b = M(u^{k+1} - \widetilde{u}^{k+1})$ that

$$\begin{aligned} &\big\|M(u^k - \widetilde{u}^k)\big\|_H^2 - \big\|M(u^{k+1} - \widetilde{u}^{k+1})\big\|_H^2 \\ =&2(u^k - \widetilde{u}^k)^\top M^\top HM\big\{(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\} \\ &- \big\|M(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\|_H^2 \\ \geq&\big\|u^k - \widetilde{u}^k - (u^{k+1} - \widetilde{u}^{k+1})\big\|^2_{(Q^\top + Q)} - \big\|M(u^k - \widetilde{u}^k) - (u^{k+1} - \widetilde{u}^{k+1})\big\|_H^2 \\ \overset{(2.5)}{=}&\big\|u^k - \widetilde{u}^k - (u^{k+1} - \widetilde{u}^{k+1})\big\|_G^2 \geq 0. \end{aligned}$$

This confirms the assertion in (3.13) based on the nonsingularity of M and the positive definiteness of H. □

THEOREM 3.6. *For any integer $T > 0$ and $\kappa \geq 0$, under the condition (1.9) the sequence $\{u^{k+1}\}$ generated by G-AFBA satisfy*

$$\left\|u^{T+\kappa} - u^{T+\kappa+1}\right\|_H^2 \leq \frac{1}{(T+1)c_0}\left\|u^\kappa - u^*\right\|_H^2, \quad \forall u^* \in \Omega^*. \tag{3.18}$$

Proof. Firstly, by previous inequality (3.7) and the positive definiteness of G, there exists a constant $c_0 > 0$ such that

$$\left\|u^{k+1} - u^*\right\|_H^2 \leq \left\|u^k - u^*\right\|_H^2 - c_0\left\|u^k - u^{k+1}\right\|_H^2. \tag{3.19}$$

The monotonicity of $\{\|u^k - u^{k+1}\|^2\}$ shows

$$\sum_{k=\kappa}^{T+\kappa} \left\|u^k - u^{k+1}\right\|_H^2 \geq (1+T)\left\|u^{T+\kappa} - u^{T+\kappa+1}\right\|_H^2.$$

Then, sum (3.19) over $k = \kappa, 1, \cdots, T+\kappa$ together with the last inequality to have

$$\left\|u^\kappa - u^*\right\|_H^2 \geq \sum_{k=\kappa}^{T+\kappa} c_0\left\|u^k - u^{k+1}\right\|_H^2 \geq (1+T)c_0\left\|u^{T+\kappa} - u^{T+\kappa+1}\right\|_H^2$$

for any $u^* \in \Omega^*$ and hence confirms the assertion in (3.18) immediately. □

For any given $\epsilon > 0$, Theorem 3.6 shows that the proposed G-AFBA (1.8) needs at most $[c/\epsilon]$ iterations to ensure $\left\|u^k - u^{k+1}\right\|_H^2 \leq \epsilon$, where $c = \inf_{u^* \in \Omega^*} \|u^0 - u^*\|_H^2/c_0$. Recall that u^{k+1} is a solution point of $\mathrm{VI}(\theta, \mathcal{J}, \Omega)$ if and only if $\|u^k - u^{k+1}\| = 0$. So, the worst-case $O(1/T)$ convergence rate in a pointwise sense is established for G-AFBA (1.8). Besides, Theorem 3.6 also indicates that $\left\|u^k - u^{k+1}\right\| \leq \epsilon$ can be utilized as a stopping criterion for implementing G-AFBA (1.8).

3.3. Linear rate of convergence. We define the KKT mapping as

$$R(u) := \begin{pmatrix} x - \mathrm{prox}_f\,(x - K^*y) \\ y - \mathrm{prox}_g(y + Kx) \end{pmatrix}, \quad \forall u = (x; y) \in \Omega, \tag{3.20}$$

and it is continuous on Ω because the proximal operator of a proper convex function is Lipschitz continuous with unit Lipschitz constant. For any $u \in \Omega$, $u \in \Omega^*$ if and only if $R(u) = 0$ and hence $\Omega^* = \{u \in \Omega \mid R(u) = 0\}$.

In the following, under the calmness condition (3.21), we will establish the Q-linear and R-linear convergence rate of G-AFBA in terms of $\mathrm{dist}_H(u^k, \Omega^*)$ and u^k. Similar condition had been used for the linear rate of the ADMM and the inexact primal-dual algorithm, see e.g. [3, 23].

THEOREM 3.7. *Let $\{\widetilde{u}^k\}$ be the predictor generated by (3.1a)-(3.1b) and $\{u^{k+1}\}$ be the corrector generated by (3.2). Then, we have the following properties:*

(i) There exists a saddle-point $u^\infty = (x^\infty; y^\infty) \in \Omega^*$ such that

$$\lim_{k\to\infty} \widetilde{u}^k = \lim_{k\to\infty} u^{k+1} = u^\infty.$$

(ii) If R^{-1} is calm at the origin for u^∞ with modulus $\theta > 0$, that is,

$$dist(u, \Omega^*) \leq \theta \|R(u)\|, \quad \forall u \in \{u \in \Omega^* \big| \|u - u^\infty\| \leq r\}, \tag{3.21}$$

for some $r > 0$, then there exists a positive number

$$\xi = \sqrt{1 - \frac{1}{(1+\theta\sqrt{\kappa_1})}\sqrt{\frac{\lambda_{\min}(G)}{\kappa_2}}} \in (0,1)$$

such that

$$\mathrm{dist}_H(u^{k+1}, \Omega^*) \leq \xi \mathrm{dist}_H(u^k, \Omega^*), \tag{3.22}$$

for all $k \geq 0$, where κ_1 and κ_2 are given by (3.26) and (3.28) respectively.

(iii) The sequence $\{u^k\}$ converges to u^∞ R-linearly.

Proof. The first assertion holds clearly by Theorem 3.3. So, there exists a positive integer $\overline{k}$ such that

$$\|u^k - u^\infty\| \leq r, \quad k \geq \overline{k}. \tag{3.23}$$

Notice that, the optimality conditions of (3.1a)-(3.1b) can be characterized as

$$\begin{cases} \widetilde{x}^k = \mathrm{prox}_f\Big[\widetilde{x}^k - \Big(\frac{1}{\tau}(\widetilde{x}^k - x^k) + K^* y^k\Big)\Big], \\ \widetilde{y}^k = \mathrm{prox}_g\Big[\widetilde{y}^k - \Big(\frac{1}{\sigma}(\widetilde{y}^k - y^k) - K\big(\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)\big)\Big)\Big]. \end{cases} \tag{3.24}$$

Combine (3.24) and the definition of $R(\cdot)$ in (3.20) to obtain

$$\begin{aligned} \|R(\widetilde{u}^k)\|^2 =& \big\|\widetilde{x}^k - \mathrm{prox}_f(\widetilde{x}^k - K^*\widetilde{y}^k)\big\|^2 + \big\|\widetilde{y}^k - \mathrm{prox}_g(\widetilde{y}^k + K\widetilde{x}^k)\big\|^2 \\ \leq& \big\| - \frac{1}{\tau}(\widetilde{x}^k - x^k) + K^*(\widetilde{y}^k - y^k)\big\|^2 + \big\|\alpha K(\widetilde{x}^k - x^k) + \frac{1}{\sigma}(\widetilde{y}^k - y^k)\big\|^2 \\ \leq& 2\big(\alpha L + \frac{1}{\tau^2}\big)\|x^k - \widetilde{x}^k\|^2 + 2\big(L + \frac{1}{\sigma^2}\big)\|y^k - \widetilde{y}^k\|^2 \\ \leq& \kappa_1 \|u^k - \widetilde{u}^k\|^2, \end{aligned} \tag{3.25}$$

where first inequality uses the nonexpansive property of $\mathrm{prox}_f(\cdot)$ and $\mathrm{prox}_g(\cdot)$, and

$$\kappa_1 = 2\max\left\{\alpha L + \frac{1}{\tau^2}, L + \frac{1}{\sigma^2}\right\}. \tag{3.26}$$

Now, it follows from (3.25) and (3.21) that for all $k \geq \overline{k}$,

$$\mathrm{dist}(\widetilde{u}^k, \Omega^*) \leq \theta\sqrt{\kappa_1}\|u^k - \widetilde{u}^k\| \leq \theta\sqrt{\frac{\kappa_1}{\lambda_{\min}(G)}}\|u^k - \widetilde{u}^k\|_G. \tag{3.27}$$

Let

$$\kappa_2 = \max\{\lambda_{\min}(G), \lambda_{\max}(H)\}. \tag{3.28}$$

Then, we can deduce from the triangle inequality $\|u^* - \widetilde{u}^k\| \geq \|u^* - u^k\| - \|u^k - \widetilde{u}^k\|$ that for $k \geq \overline{k}$,

$$\begin{aligned} \mathrm{dist}(\widetilde{u}^k, \Omega^*) \geq& \mathrm{dist}(u^k, \Omega^*) - \|u^k - \widetilde{u}^k\| \\ \geq& \frac{1}{\sqrt{\kappa_2}}\mathrm{dist}_H(u^k, \Omega^*) - \frac{1}{\sqrt{\lambda_{\min}(G)}}\|u^k - \widetilde{u}^k\|_G. \end{aligned} \tag{3.29}$$

So, combine (3.27) and (3.29) to yield

$$\|u^k - \widetilde{u}^k\|_G \geq \frac{1}{(1+\theta\sqrt{\kappa_1})}\sqrt{\frac{\lambda_{\min}(G)}{\kappa_2}}\mathrm{dist}_H(u^k, \Omega^*).$$

Consequently, it follows from (3.7) and the last inequality that for $k \geq \overline{k}$,

$$\begin{aligned}\mathrm{dist}_H^2(u^{k+1}, \Omega^*) \leq &\mathrm{dist}_H^2(u^k, \Omega^*) - \|u^k - \widetilde{u}^k\|_G^2 \\ &\leq \left[1 - \frac{1}{(1+\theta\sqrt{\kappa_1})}\sqrt{\frac{\lambda_{\min}(G)}{\kappa_2}}\right]\mathrm{dist}_H^2(u^k, \Omega^*).\end{aligned}$$

So, the assertion (3.22) holds, that is, the Q-linear convergence rate is established.

Finally, we prove the assertion (iii). Select $u_k^* = \in \Omega^*$ such that $\mathrm{dist}_H(u^k, \Omega^*) = \|u^k - u_k^*\|_H$ and let

$$d^{k+1} = u^{k+1} - u^k.$$

Then, if follows from (3.7) that $\left\|u_k^* - u^k\right\|_H \geq \left\|u_k^* - u^{k+1}\right\|_H$, which implies

$$\begin{aligned}\left\|d^{k+1}\right\|_H &= \left\|u^{k+1} - u^k\right\|_H \\ &\leq \left\|u_k^* - u^k\right\|_H + \left\|u_k^* - u^{k+1}\right\|_H \\ &\leq 2\left\|u_k^* - u^k\right\|_H = 2\mathrm{dist}_H(u^k, \Omega^*) \overset{(3.22)}{\leq} 2\xi^k \mathrm{dist}_H(u^0, \Omega^*).\end{aligned}$$

By the first assertion in Theorem 3.7, we have $u^\infty = u^k + \sum_{j=k}^\infty d^j$. So,

$$\begin{aligned}\left\|u^k - u^\infty\right\|_H &\leq \sum_{j=k}^\infty \left\|d^j\right\|_H \leq 2\mathrm{dist}_H(u^0, \Omega^*)\sum_{j=k}^\infty \xi^j \\ &= 2\mathrm{dist}_H(u^0, \Omega^*)\xi^k \sum_{j=0}^\infty \xi^j = \xi^k\left(2\mathrm{dist}_H(u^0, \Omega^*)\frac{1}{1-\xi}\right),\end{aligned}$$

that is, sequence $\{u^k\}$ converge to u^∞ R-linearly. □

Generally speaking, it is not easy to evaluate the constant θ in the condition (3.21), so the ratio ξ is not explicitly known either. However, our convergence results are more general and much stronger than those in [7, 8] which uses strong convexity assumption on the objective function. In fact, if we assume the mapping R defined by (3.20) is piecewise polyhedral, equivalently, R^{-1} is piecewise polyhedral, then it follows from [34] that there exist two constants $\theta, \eta > 0$ such that

$$\mathrm{dist}(u, \Omega^*) \leq \theta\|R(u)\|, \quad \forall u \in \left\{u \in \Omega^* \middle| \|R(u)\| \leq \eta\right\}. \tag{3.30}$$

By Theorem 3.3 again, (3.23) holds. When $R(u^k) > \eta$, we have

$$\mathrm{dist}(u^k, \Omega^*) \leq \|u^k - u^\infty\| \leq r < \frac{r}{\eta}\|R(u^k)\|. \tag{3.31}$$

Combine (3.30) and (3.31) to have a similar condition to (3.21), that is, $\mathrm{dist}(u, \Omega^*) \leq \max\{\theta, r/\eta\}\|R(u)\|$. As a result, similar to the proof of Theorem 3.7, we immediately have the following corollary.

Corollary 3.8. *Let $\{\widetilde{u}^k\}$ be the predictor generated by (3.1a)-(3.1b) and $\{u^{k+1}\}$ be the corrector generated by (3.2). Assume the mapping R defined by (3.20) is piecewise polyhedral. Then, the following properties hold:*

(i) There exists a constant $\overline{\theta} > 0$ *such that for all* $k \geq 0$, $\mathrm{dist}(\widetilde{u}^k, \Omega^*) \leq \overline{\theta}\|R(\widetilde{u}^k)\|$.

(ii) For all $k \geq 0$, *there exists a positive number*

$$\widetilde{\xi} = \sqrt{1 - \frac{1}{(1+\widetilde{\theta}\sqrt{\kappa_1})}\sqrt{\frac{\lambda_{\min}(G)}{\kappa_2}}} \in (0,1)$$

such that

$$\mathrm{dist}_H(u^{k+1}, \Omega^*) \leq \widetilde{\xi}\mathrm{dist}_H(u^k, \Omega^*).$$

Moreover, the sequence $\{u^k\}$ *converges to* u^∞ *R-linearly.*

4. A faster G-AFBA with accelerated rate. In this section, we show the $O(1/T^2)$ pointwise convergence rate of a faster G-AFBA (fG-AFBA) based on the prediction-correction framework at the beginning of Section 3 as well as some modification techniques for both the involved quadratic terms and the correction step.

---

fG-AFBA: A faster G-AFBA for solving (1.1).

Prediction Step:

$$\begin{cases} \breve{x}^k = \arg\min_{x\in\mathcal{X}} \Big\{ f(x) + \dfrac{\nu^k}{2\tau}\Big\|\Big(\dfrac{1}{\nu^k}x - \dfrac{1-\nu^k}{\nu^k}\breve{x}^{k-1}\Big) - x^k + \tau K^* y^k\Big\|^2\Big\}, \\ \breve{y}^k = \arg\min_{y\in\mathcal{Y}} \Big\{ g(y) + \dfrac{\nu^k}{2\sigma}\Big\|\Big(\dfrac{1}{\nu^k}y - \dfrac{1-\nu^k}{\nu^k}\breve{y}^{k-1}\Big) - y^k - \sigma K[\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)]\Big\|^2\Big\}, \\ \quad \text{where } \widetilde{x}^k \text{ is updated by (4.2).} \end{cases} \tag{4.1}$$

Correction Step:

$$u^{k+1} = u^k - M(u^k - \widetilde{u}^k),$$

where $\widetilde{u}^k = (\widetilde{x}^k; \widetilde{y}^k)$ is updated by

$$\widetilde{u}^k = \frac{1}{\nu^k}\breve{u}^k - \frac{1-\nu^k}{\nu^k}\breve{u}^{k-1} \tag{4.2}$$

with $\breve{u}^k = (\breve{x}^k; \breve{y}^k)$ and the sequence $\{\nu^k\}(k = -1, 0, 1, \cdots)$ satisfies

$$1/\nu^{k-1} = (1-\nu^k)/\nu^k, \quad \nu^{-1} \in (0,1). \tag{4.3}$$

LEMMA 4.1. *The iterates generated by (4.1) satisfy*

$$\theta(u) - \theta(\breve{u}^k) + \big\langle u - \breve{u}^k, \mathcal{J}(\widetilde{u}^k)\big\rangle \geq (u - \breve{u}^k)^\top Q(u^k - \widetilde{u}^k), \quad \forall u \in \Omega,$$

where Q *is given by (2.4),* θ *and* $\mathcal{J}$ *are defined by (2.3).*

Proof. It follows from the optimality condition of the $\breve{x}^k$-subproblem in (4.1) and the way of generating $\widetilde{u}^k$ in (4.2) that

$$f(x) - f(\breve{x}^k) + \big\langle x - \breve{x}^k, K^*y^k + \frac{1}{\tau}(\widetilde{x}^k - x^k)\big\rangle \geq 0, \quad \forall x \in \mathcal{X}.$$

Namely,

$$f(x) - f(\breve{x}^k) + \big\langle x - \breve{x}^k, K^*\widetilde{y}^k\big\rangle \geq \Big\langle x - \breve{x}^k, \frac{1}{\tau}(x^k - \widetilde{x}^k) - K^*(y^k - \widetilde{y}^k)\Big\rangle. \tag{4.4}$$

Similarly, for any $y \in \mathcal{Y}$ we have

$$g(y) - g(\breve{y}^k) + \langle y - \breve{y}^k, -K\widetilde{x}^k \rangle \geq \Big\langle y - \breve{y}^k, -\alpha K(x^k - \widetilde{x}^k) + \frac{1}{\sigma}(y^k - \widetilde{y}^k) \Big\rangle. \tag{4.5}$$

Combining (4.4)-(4.5) and the matrix Q given by (2.4) ends the proof. ◻

Lemma 4.1 implies that fG-AFBA is a special case of the framework [44, section 3] and hence have the $O(1/T^2)$ convergence rate as the following.

THEOREM 4.2. *Let H be defined by (2.5) and the condition (1.9) holds. Then, for any integer $T > 0$, the the iterates generated by fG-AFBA satisfy*

$$\|\breve{u}^T - \breve{u}^{T-1}\|_H^2 \leq O(1/T^2).$$

Proof. The proof is similar to [44, Theorem 3.2], because our fG-AFBA is a special case of the framework [44, section 3], the matrix H and G defined by (2.5) are positive definite and the matrix M is nonsingular under the condition (1.9). ◻

5. Relationship between (1.8) and other methods. In this section, we analyze the relationship between G-AFBA (1.8) and some well-established methods, and extend G-AFBA (1.8) to a multi-block separable convex optimization problem.

- **Case 1 (CP-PPA in [6] and a reduced ALM).** When $\alpha = 1$, the proposed G-AFBA (1.8) reduces to

$$\begin{cases} x^{k+1} = \arg\min\limits_{x \in \mathcal{X}} \Big\{ f(x) + \dfrac{1}{2\tau} \big\| x - x^k + \tau K^* y^k \big\|^2 \Big\}, \\ y^{k+1} = \arg\min\limits_{y \in \mathcal{Y}} \Big\{ g(y) + \dfrac{1}{2\sigma} \big\| y - y^k - \sigma K[2\widetilde{x}^k - x^k] \big\|^2 \Big\}, \end{cases}$$

which is clearly the same as CP-PPA [6]. As analyzed in [19, Section 3.2], the condition (1.5) can be relaxed as

$$\frac{1}{\tau\sigma} > 0.75L. \tag{5.1}$$

to ensure convergence of CP-PPA, and the constant 0.75 can not be replaced by any other smaller positive number. Besides, when $\alpha = 1$ and $g = 0$, the problem (1.1) is equivalent to the following minimization problem

$$\min_{x \in \mathcal{X}} f(x) \qquad \text{s.t. } Kx = \mathbf{0}.$$

and G-AFBA (1.8) recovers a ALM-type method

$$\begin{cases} x^{k+1} = \arg\min\limits_{x \in \mathcal{X}} \Big\{ f(x) + \dfrac{1}{2\tau} \big\| x - x^k + \tau K^* \lambda^k \big\|^2 \Big\}, \\ \lambda^{k+1} = \lambda^{k+1} + \sigma K(2x^{k+1} - x^k). \end{cases}$$

Note that, here two different parameters τ and σ are exploited, which is different from the traditional augmented Lagrangian method.

- **Case 2 (Exact version of [24, Algorithm 2]).** When $(\alpha, \mu) = (0, 1)$, the proposed G-AFBA reduces to

$$\begin{cases} \overline{x}^{k+1} = \arg\min\limits_{x \in \mathcal{X}} \Big\{ f(x) + \dfrac{1}{2\tau} \big\| x - x^k + \tau K^* y^k \big\|^2 \Big\}, \\ y^{k+1} = \arg\min\limits_{y \in \mathcal{Y}} \Big\{ g(y) + \dfrac{1}{2\sigma} \big\| y - y^k - \sigma K \overline{x}^{k+1} \big\|^2 \Big\}, \\ x^{k+1} = \overline{x}^{k+1} - \tau K^*(y^{k+1} - y^k), \end{cases} \tag{5.2}$$

which is the exact version of [24, Algorithm 2] with iterative relative error being zero. For this case, the previous convergence condition (1.9) reduces to (1.5). Assuming that a part of the (both) underlying functions are strongly convex, similar $O(1/T^2)$ and linear convergence rate can be obtained, see [24] for more details. However, as discussed in Section 4, the $O(1/T^2)$ convergence rate can be also obtained for the general problem (1.1).

- **Case 3.** If we just set $\alpha = 0$, then G-AFBA reduces to

$$\text{(G-AFBA1)}\quad \begin{cases} \overline{x}^{k+1} = \arg\min_{x\in\mathcal{X}} \Big\{ f(x) + \frac{1}{2\tau}\big\|x - x^k + \tau K^* y^k\big\|^2 \Big\}, \\ \overline{y}^{k+1} = \arg\min_{y\in\mathcal{Y}} \Big\{ g(y) + \frac{1}{2\sigma}\big\|y - y^k - \sigma K \overline{x}^{k+1}\big\|^2 \Big\}, \\ x^{k+1} = \overline{x}^{k+1} - \mu\tau K^*(\overline{y}^{k+1} - y^k), \\ y^{k+1} = \overline{y}^{k+1} + (1-\mu)\sigma K(\overline{x}^{k+1} - x^k), \end{cases} \tag{5.3}$$

which is similar to the scheme (5.2) and the difference is just the last two iterations. Moreover, one may treat (5.3) as an extension of (5.2), since an extra extrapolation step with inertia parameter $(1-\mu)\sigma \in (0,1)$ is exploited for the y-iterate while the x^{k+1}-iterate in (5.3) can be written as

$$x^{k+1} = \overline{x}^{k+1} - \tau K^*(\overline{y}^{k+1} - y^k) + (1-\mu)\tau K^*(y^{k+1} - y^k).$$

For the case (5.3), a surprising observation is that its convergence condition will be relaxed as

$$\frac{1}{\tau\sigma} > (1 - \mu + \mu^2)L, \tag{5.4}$$

Clearly, $(1-\mu+\mu^2) \le 1$ for any $\mu \in [0,1]$ and its optimal lower value is 0.75 when $\mu = 0.5$. Hence, set $\mu = 0.5$ in (5.4) to obtain the same condition as (5.1). In other words, $0.75L$ is the the optimal lower bound of the condition (5.4). Although we obtain the same tight convergence condition (5.1) as that in [19], our method is different to GCP-PPA (1.6). The forthcoming case shows that our G-AFBA can also reduce to GCP-PPA directly.

- **Case 4 (GCP-PPA [19]).** When $\mu = 0$, G-AFBA reduces to

$$\begin{cases} x^{k+1} = \arg\min_{x\in\mathcal{X}} \Big\{ f(x) + \frac{1}{2\tau}\big\|x - x^k + \tau K^* y^k\big\|^2 \Big\}, \\ \overline{y}^{k+1} = \arg\min_{y\in\mathcal{Y}} \Big\{ g(y) + \frac{1}{2\sigma}\big\|y - y^k - \sigma K[x^{k+1} + \alpha(x^{k+1} - x^k)]\big\|^2 \Big\}, \\ y^{k+1} = \overline{y}^{k+1} + (1-\alpha)\sigma K(x^{k+1} - x^k), \end{cases} \tag{5.5}$$

which is the method (1.6) proposed in [19]. As pointed in [19], applying (5.5) to the convex programming $\min\{f(x) \mid Ax = b, x \in \mathcal{X}\}$ is equivalent to CP-PPA. For the case (5.5), the previous condition (1.9) becomes

$$\frac{1}{\tau\sigma} > (1 - \alpha + \alpha^2)L.$$

It is similar to (5.4) and can reduce to (5.1) by setting $\alpha = 0.5$.

- **Case 5 (Connection of an extended G-AFBA with [38]).** Now, we assume

$$\frac{1}{\tau\sigma} > \frac{(1+\alpha)^2}{4}L, \quad \forall\alpha, \tag{5.6}$$

which indicates $\frac{1}{\tau\sigma} > \alpha L$ due to $\frac{(1+\alpha)^2}{4} > \alpha$ and hence both $\frac{1}{\tau\sigma}\mathbf{I} - \alpha K^*K$ and $\frac{1}{\tau\sigma}\mathbf{I} - \alpha KK^*$ are invertible. If we modify the previous correction step (3.2) as the following Newton-like correction step

$$u^{k+1} = u^k - \gamma\beta^k Q^{-\top}(u^k - \widetilde{u}^k), \tag{5.7}$$

where $\gamma \in (0, 2)$ is a relaxation factor and

$$Q^{-\top} = \begin{bmatrix} (\frac{1}{\tau\sigma}\mathbf{I} - \alpha K^*K)^{-1} & \mathbf{0} \\ \mathbf{0} & (\frac{1}{\tau\sigma}\mathbf{I} - \alpha KK^*)^{-1} \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma}\mathbf{I} & \alpha K^* \\ K & \frac{1}{\tau}\mathbf{I} \end{bmatrix}, \tag{5.8}$$

then an extended G-AFBA (eG-AFBA) follows

$$\begin{cases} \widetilde{u}^k = \begin{pmatrix} \widetilde{x}^k = \arg\min_{x\in\mathcal{X}} \left\{ f(x) + \frac{1}{2\tau}\|x - x^k + \tau K^* y^k\|^2 \right\} \\ \widetilde{y}^k = \arg\min_{y\in\mathcal{Y}} \left\{ g(y) + \frac{1}{2\sigma}\|y - y^k - \sigma K[\widetilde{x}^k + \alpha(\widetilde{x}^k - x^k)]\|^2 \right\} \end{pmatrix}, \\ u^{k+1} = u^k - \gamma\beta^k Q^{-\top}(u^k - \widetilde{u}^k), \quad \gamma \in (0, 2). \end{cases} \tag{5.9}$$

By (5.7), the previous equality (3.8) holds with $H = \frac{1}{\gamma\beta^k}QQ^\top$. Similar to way in [40], by updating β^k dynamically in the form of

$$\beta^k = \frac{\|u^k - \widetilde{u}^k\|^2_{Q^\top + Q}}{2\|u^k - \widetilde{u}^k\|_2^2}, \tag{5.10}$$

we have

$$\begin{aligned} &\|u^k - u^*\|^2_{QQ^\top} - \|u^{k+1} - u^*\|^2_{QQ^\top} \\ =&\|u^k - u^*\|^2_{QQ^\top} - \|u^k - u^* - \gamma\beta^k Q^{-\top}(u^k - \widetilde{u}^k)\|^2_{QQ^\top} \\ =& -(\gamma\beta^k)^2\|u^k - \widetilde{u}^k\|_2^2 + \gamma\beta^k\langle\nabla(\|u^k - u\|^2)\mid_{u=u^*}, Q(u^k - \widetilde{u}^k)\rangle \\ \geq& -(\gamma\beta^k)^2\|u^k - \widetilde{u}^k\|_2^2 + \gamma\beta^k\|u^k - \widetilde{u}^k\|^2_{Q^\top + Q} \\ =&\frac{\gamma(2-\gamma)}{2}\beta^k\|u^k - \widetilde{u}^k\|^2_{Q^\top + Q}. \end{aligned} \tag{5.11}$$

It is not difficult to verify that under the condition (5.6), the matrix $Q^\top + Q$ is symmetric positive definite. Hence, (5.11) implies the contraction of $\{u^k - u^*\}$, namely, the scheme (5.9) is globally convergent. When $\alpha = 0$, the same condition in [38], that is $\frac{1}{\tau\sigma} > \frac{1}{4}L$, is obtained, which implies that the lower bound is refined by 50% compared to (5.1); If we choose $\alpha = -1$, we have $\frac{1}{\tau\sigma} > 0$ which is more relax than that in the literature.

- **Case 6 (Connection of a linearized G-AFBA with [13]).** When the function f is smooth and its gradient is Lipschitz continuous, then we can linearize the $\overline{x}^{k+1}$-subproblem of G-AFBA as

$$\overline{x}^{k+1} = \arg\min_{x\in\mathcal{X}} \left\{ f(x^k) + \langle\nabla f(x^k), x\rangle + \frac{1}{2\tau}\left\|x - x^k + \tau K^* y^k\right\|^2 \right\}, \tag{5.12}$$

that is, a closed-form solution is obtained:

$$\overline{x}^{k+1} = x^k - \tau\big[\nabla f(x^k) + K^* y^k\big]. \tag{5.13}$$

For this case, the updates (5.13) and (3.1b) with $\alpha = 0$ amount to the first two updates of the following proximal alternating predictor corrector [13] with involved matrix $S = 1/\sigma\mathbf{I}$:

$$\text{(PAPC)} \quad \begin{cases} p^{k+1} = x^k - \tau\big[K^* y^k + \nabla f(x^k)\big], \\ y^{k+1} = \arg\min_{y\in\mathcal{Y}} \Big\{ g(y) + \dfrac{1}{2\sigma}\big\|y - y^k - \sigma K p^{k+1}\big\|^2 \Big\}, \\ x^{k+1} = x^k - \tau\big[K^* y^{k+1} + \nabla f(x^k)\big]. \end{cases} \tag{5.14}$$

Notice that the update of x^{k+1} in PAPC, which costs much computing time, is different from the correction step in our G-AFBA. For the G-AFBA with $\overline{x}^{k+1}$-subproblem linearized as in (5.12), its convergence can be analyzed in a similar way to [5, Remark 2.4] and hence is omitted.

- **Case 7 (Application of G-AFBA to separable convex optimization).** Consider the following multi-block separable convex optimization problem

$$\min \left\{ \sum_{i=1}^{q} f_i(x_i) \,\Big|\, \sum_{i=1}^{q} A_i x_i = b,\ x_i \in \mathbb{R}^{n_i} \right\}, \tag{5.15}$$

where $A_i \in \mathbb{R}^{m\times n_i} (i = 1, \cdots, q)$ are given matrices, each f_i is a proper lower semicontinuous convex function. Then its dual problem is

$$\min_{x\in\mathbb{R}^n} \max_{\lambda\in\mathbb{R}^m} \mathcal{L}(x, \lambda) := \sum_{i=1}^{q} f_i(x_i) + \langle Kx, \lambda\rangle - \langle b, \lambda\rangle,$$

where

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix},\ K = (A_1, \cdots, A_q),\ n = n_1 + \cdots + n_q.$$

The dual problem is a special case of (1.1). Hence, applying G-AFBA (1.8) to the dual form directly, we derive the following operator splitting method:

$$\begin{cases} \overline{x}_i^{k+1} = \arg\min_{x_i\in\mathbb{R}^{n_i}} \Big\{ f_i(x_i) + \dfrac{1}{2\tau}\big\|x_i - x_i^k + \tau A_i^\top \lambda^k\big\|^2 \Big\},\ i = 1, \cdots, q, \\ \overline{\lambda}^{k+1} = \lambda^k + \sigma \sum_{i=1}^{q} A_i\big[\overline{x}_i^{k+1} + \alpha(\overline{x}_i^{k+1} - x_i^k)\big] - b, \\ x_i^{k+1} = \overline{x}_i^{k+1} - (1-\alpha)\mu\ \tau A_i^\top(\overline{\lambda}^{k+1} - \lambda^k),\ i = 1, \cdots, q, \\ \lambda^{k+1} = \overline{\lambda}^{k+1} + (1-\alpha)(1-\mu)\ \sigma \sum_{i=1}^{q} A_i(\overline{x}_i^{k+1} - x_i^k). \end{cases} \tag{5.16}$$

Notice that all primal subproblems are updated paralelly and they admit a unique global solution. The $\overline{x}_i^{k+1}$-subproblem is equivalent to

$$\overline{x}_i^{k+1} = \arg\min_{x_i\in\mathbb{R}^{n_i}} \Big\{ f_i(x_i) + \langle\lambda^k, A_i x_i\rangle + \dfrac{1}{2\tau}\big\|x_i - x_i^k\big\|^2 \Big\},$$

or

$$\min_{x_i\in\mathbb{R}^{n_i}} \left\{f_i(x_i)+\left\langle\lambda^k, A_ix_i+\sum_{l=1,l\neq i}^{q} A_lx_l^k\right\rangle+\frac{0}{2}\left\|A_ix_i+\sum_{l=1,l\neq i}^{q} A_lx_l^k\right\|^2+\frac{1}{2\tau}\|x_i-x_i^k\|^2\right\},$$

which is different from the involved subproblem in the proximal ADMM [12]. The problem (5.15) also covers the following composite optimization problem

$$\min_{x\in\mathbb{R}^{n_1}} h_1(x)+h_2(x)+h_3(A_1x-b),$$

equivalently, $\min_{x\in\mathbb{R}^{n_1},z\in\mathbb{R}^m} \{h_1(x)+h_2(x)+h_3(z)|\ A_1x-z=b\}$. So, the our scheme (5.16) is applicable to such a problem.

REFERENCES

[1] K. Arrow, L. Hurwicz, H. Uzawa, *Studies in Linear and Non-linear Programming*, Stanford Mathematical Studies in the Social Sciences, vol. II, Stanford University Press, Stanford (1958).

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

[3] J. Bai, X. Chang, J. Li, F. Xu, *Convergence revisit on generalized symmetric ADMM*, Optimization, 70 (2021), pp. 149–168.

[4] J. Bai, W. Hager, H. Zhang, *An inexact accelerated stochastic ADMM for separable convex optimization*, Comput. Optim. Appl., 81 (2022), pp. 479–518.

[5] J. Bai, *A new insight on augmented Lagrangian method with applications in machine learning*, Optimization Online, (2022), https://optimization-online.org/2021/08/8559/.

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

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

[8] A. Chambolle, M. Ehrhardt, P. Richtarik, C. Schonlieb, *Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications*, SIAM J. Optim., 28 (2018), pp. 2783–2808.

[9] X. Chang, J. Yang, H. Zhang, *Golden ratio primal-dual algorithm with linesearch*, SIAM J. Optim., 32 (2022), pp. 1584–1613.

[10] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479.

[11] L. Condat, D. Kitahara, A. Contreras, A. Hirabayashi, *proximal splitting algorithms for convex optimization: a tour of recent advances, with new twists*, SIAM Review, 65 (2023), pp. 375–435.

[12] W. Deng, W. Yin, *On the global and linear convergence of the generalized alternating direction method of multipliers*, J. Sci. Comput. 66 (2016), pp. 889–916.

[13] Y. Drori, S. Sabach, M. Teboulle, *A simple algorithm for a class of nonsmooth convex-concave saddle-point problems*, Oper. Res. Lett., 43 (2015), pp. 209–214.

[14] D. Gabay, B. Mercier, *A dual algorithm for the solution of nonlinear variational problems viafinite-element approximations*, Comput. Math. Appl., 2 (1976), pp. 17–40.

[15] S. Hayden, O. Stanley, *A low patch-rank interpretation of texture*, SIAM J. Imaging Sci., 6 (2013), pp. 226–262.

[16] B. He, Y. You, X. Yuan, *On the convergence of primal-dual hybrid gradient algorithm*, SIAM J. Imaging Sci., 7 (2014), pp. 2526–2537.

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

[18] B. He, F. Ma 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.

[19] 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), pp. 1157–1183.

[20] B. He, S. Xu, X. Yuan, *On convergence of the Arrow-Hurwicz method for saddle point problems*, J. Math. Imaging Vision, 64 (2022), pp. 662–671.

[21] B. He, X. Yuan, *On construction of splitting contraction algorithms in a prediction-correction framework for separable convex optimization*, Commun. Optim. Theory, 2023 (2023), Artical ID 28, pp. 1–23.
[22] F. Jiang, X. Cai, Z. Wu, D. Han, *Approximate first-order primal-dual algorithms for saddle point problems*, Math. Comput., 90 (2021), pp. 1227–1262.
[23] F. Jiang, Z. Wu, X. Cai, H. Zhang, *A first-order inexact primal-dual algorithm for a class of convex-concave saddle point problems*, Numer. Algor., 88 (2021), pp. 1109–1136.
[24] F. Jiang, X. Cai, D. Han, *Inexact asymmetric forward-backward-adjoint splitting algorithms for saddle point problems*, Numer. Algor., (2023), https://doi.org/10.1007/s11075-023-01509-w.
[25] F. Jiang, Z. Zhang, H. He, *Solving saddle point problems: a landscape of primal-dual algorithm with larger stepsizes*, J. Global Optim., 85 (2023), :821–846.
[26] G. Korpelevič, *An extragradient method for finding saddle points and for other problems*, Èkon. Mat. Metody, 12 (1976), pp. 747–756.
[27] P. Latafat, P. Patrinos, *Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators*, Comput. Optim. Appl., 68 (2017), pp. 57–93.
[28] Z. Li, M. Yan, *New convergence analysis of a primal-dual algorithm with large stepsizes*, Adv. Comput. Math., 47 (2021), pp. 1-20.
[29] P. Lions, B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., 16 (1979), pp. 964–979.
[30] D. O'Connor, L. Vandenberghe, *On the equivalence of the primal-dual hybrid gradient method and Douglas-Rachford splitting*, Math. Program., 179 (2020), pp. 85–108.
[31] N. Parikh, S. Boyd, *Proximal algorithms*, Foundations and Trends in Optimization, 1 (2014), pp. 127– 239.
[32] B. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phy., 4 (1964), pp. 1–17.
[33] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, (2015).
[34] S. Robinson, *Some Continuity Properties of Polyhedral Multifunctions*, Mathematical Programming at Oberwolfach, Springer, Berlin, (1981), pp. 206–214.
[35] B. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681.
[36] N. Wang, J. Li, *A class of preconditioners based on symmetric-triangular decomposition and matrix splitting for generalized saddle point problems*, IMA J. Numer. Anal., (2022), https://doi.org/10.1093/imanum/drac061.
[37] W. Xian, F. Huang, Y. Zhang, H. Huang. *A faster decentralized algorithm for nonconvex minimax problems*, NeurIPS 2021, https://openreview.net/forum?id=rjIjkiyAJao.
[38] S. Xu. *A search direction inspired primal-dual method for saddle point problems*, Optimization Online, https://optimization-online.org/wp-content/uploads/2019/11/7491.pdf
[39] Z. Xu, H. Zhang, Y. Xu, G. Lan, *A unified single-loop alternating gradient projection algorithm for nonconvex-concave and convex-nonconcave minimax problems*, Math. Program., 201 (2023), pp. 635–706.
[40] C. Ye, X. Yuan, *A descent method for structured monotone variational inequalities*, Optim. Methods Softw., 22 (2007), pp. 329–338.
[41] J. Yang, Y. Zhang, *Alternating direction algorithms for ℓ_1-problems in compressive sensing*, SIAM J. Sci. Comput., 33 (2011), pp. 250–278.
[42] M. Zhu, T. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, CAM Report 08-34, UCLA, Los Angeles, CA, (2008).
[43] X. Zhang, M. Burger, S. Osher, *A unified primal-dual algorithm framework based on Bregman iteration*, J. Sci. Comput., 46 (2011), pp. 20–46.
[44] T. Zhang, Y. Xa, S. Li, *A faster prediction-correction framework for solving convex optimization problems*, (2023), arXiv: 2304.02465v1.
[45] Y. Zhu, D. Liu, Q. Dinh, *New primal-dual algorithms for a class of nonsmooth and nonlinear convex-concave minimax problems*, SIAM J, Optim., 32 (2022), pp. 2580–2611.