J Optim Theory Appl (2014) 163:105–129
DOI 10.1007/s10957-013-0489-z

Inexact Alternating-Direction-Based Contraction Methods for Separable Linearly Constrained Convex Optimization

Guoyong Gu · Bingsheng He · Junfeng Yang

Received: 19 February 2013 / Accepted: 17 November 2013 / Published online: 13 December 2013

Abstract Alternating direction method of multipliers has been well studied in the context of linearly constrained convex optimization. In the last few years, we have witnessed a number of novel applications arising from image processing, compressive sensing and statistics, etc., where the approach is surprisingly efficient. In the early applications, the objective function of the linearly constrained convex optimization problem is separable into two parts. Recently, the alternating direction method of multipliers has been extended to the case where the number of the separable parts in the objective function is finite. However, in each iteration, the subproblems are required to be solved exactly. In this paper, by introducing some reasonable inexactness criteria, we propose two inexact alternating-direction-based contraction methods, which substantially broaden the applicable scope of the approach. The convergence and complexity results for both methods are derived in the framework of variational inequalities.

Keywords Alternating direction method of multipliers · Separable linearly constrained convex optimization · Contraction method · Complexity

Communicated by Panos M. Pardalos.

G. Gu (✉) · B. He · J. Yang
Department of Mathematics, Nanjing University, Nanjing, 210093, China
e-mail: ggu@nju.edu.cn

B. He
e-mail: hebma@nju.edu.cn

J. Yang
e-mail: jfyang@nju.edu.cn

1 Introduction

Because of its significant efficiency and easy implementation, the alternating direction method of multipliers (ADMM) has attracted wide attention in various areas [1, 2]. In particular, some novel and attractive applications of the ADMM have been discovered very recently; e.g., total-variation regularization problems in image processing [3], ℓ_1-norm minimization in compressive sensing [4], semidefinite optimization problems [5], the covariance selection problem and semidefinite least squares problem in statistics [6, 7], the sparse and low-rank recovery problem in engineering [8], etc.

The paper is organized as follows. As a preparation, in the next section, we give a brief review of ADMM, which serves as a motivation of our inexact alternating-direction-based contraction methods. Some basic properties of the projection mappings and variational inequalities are recalled as well. Then, in Sect. 3 we present two contraction methods, which are based on two descent directions generated from an inexact alternating minimization of $\mathcal{L}_A$ defined in (4). The rationale of the two descent search directions follows in Sect. 4, and the convergence and complexity results are proved in Sects. 5 and 6, respectively. Finally, some conclusions are drawn in Sect. 7.

2 Preliminaries

Before introducing the ADMM, we briefly review the augmented Lagrangian method (ALM), for which we consider the following linearly constrained convex optimization problem:

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

Here $\theta(x) : \mathbb{R}^n \to \mathbb{R}$ is a convex function (not necessarily smooth), $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $\mathcal{X} \subseteq \mathbb{R}^n$ is a closed and convex set. For solving problem (1), the classical ALM generates a sequence of iterates via the following scheme:

$$\begin{cases} x^{k+1} = \arg\min\{\theta(x) - \langle\lambda^k, Ax - b\rangle + \frac{1}{2}\|Ax - b\|_H^2 \mid x \in \mathcal{X}\}, \\ \lambda^{k+1} = \lambda^k - H(Ax^{k+1} - b), \end{cases}$$

where $H \in \mathbb{R}^{m\times m}$ is a positive definite scaling matrix penalizing the violation of the linear constraints, and $\lambda^k \in \mathbb{R}^m$ is the associated Lagrange multiplier; see, e.g., [9, 10] for more details.

An important special scenario of (1), which captures concrete applications in many fields [6, 11–13], is the following case, where the objective function is separable into two parts:

$$\min\big\{\theta_1(x_1) + \theta_2(x_2) \mid A_1x_1 + A_2x_2 = b, x_i \in \mathcal{X}_i, i = 1, 2\big\}. \tag{2}$$

Here $n_1 + n_2 = n$, and, for $i = 1, 2$, $\theta_i : \mathbb{R}^{n_i} \to \mathbb{R}$ are convex functions (not necessarily smooth), $A_i \in \mathbb{R}^{m\times n_i}$, and $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ are closed and convex sets. For solving (2), the ADMM, which dates back to [14] and is closely related to the Douglas–Rachford operator splitting method [15], is perhaps one of the most popular methods. Given (x_2^k, λ^k), the ADMM generates $(x_1^{k+1}, x_2^{k+1}, \lambda^{k+1})$ via the following scheme:

$$\begin{cases} x_1^{k+1} = \arg\min\{\theta_1(x_1) - \langle \lambda^k, A_1x_1 + A_2x_2^k - b\rangle + \frac{1}{2}\|A_1x_1 + A_2x_2^k - b\|_H^2 \mid \\ \qquad\qquad x_1 \in \mathcal{X}_1\}, \\ x_2^{k+1} = \arg\min\{\theta_2(x_2) - \langle \lambda^k, A_1x_1^{k+1} + A_2x_2 - b\rangle + \frac{1}{2}\|A_1x_1^{k+1} + A_2x_2 - b\|_H^2 \mid \\ \qquad\qquad x_2 \in \mathcal{X}_2\}, \\ \lambda^{k+1} = \lambda^k - H(A_1x_1^{k+1} + A_2x_2^{k+1} - b). \end{cases}$$

Therefore, ADMM can be viewed as a practical and structure-exploiting variant (in a split or relaxed form) of the classical ALM for solving the separable problem (2), with the adaptation of minimizing the involved separable variables separately in an alternating order.

In this paper, we consider a more general separable case of (1) in the sense that the objective function is separable into finitely many parts:

$$\min\left\{\sum_{i=1}^N \theta_i(x_i) \,\middle|\, \sum_{i=1}^N A_ix_i = b, x_i \in \mathcal{X}_i, i = 1, \ldots, N\right\}, \tag{3}$$

where $\sum_{i=1}^N n_i = n$, and, for $i = 1, \ldots, N$, $\theta_i : \mathbb{R}^{n_i} \to \mathbb{R}$ are convex functions (not necessarily smooth), $A_i \in \mathbb{R}^{m\times n_i}$, $b \in \mathbb{R}^m$, and $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ are closed and convex sets. Without loss of generality, we assume that the solution set of (3) is nonempty. Recently, the ADMM was extended to handle (3) by He et al. [16]. For convenience, we denote the augmented Lagrangian function of (3) by

$$\mathcal{L}_A(x_1, \ldots, x_N, \lambda) := \sum_{i=1}^N \theta_i(x_i) - \left\langle \lambda, \sum_{i=1}^N A_ix_i - b\right\rangle + \frac{1}{2}\left\|\sum_{i=1}^N A_ix_i - b\right\|_H^2, \tag{4}$$

where $H \in \mathbb{R}^{m\times m}$ is a symmetric positive definite scaling matrix, and $\lambda \in \mathbb{R}^m$ is the Lagrangian multiplier. Given $(x_1^k, \ldots, x_N^k, \lambda^k)$, the algorithm in [16] generates the next iterate $(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1})$ in two steps. First, the algorithm produces a trial point $(\tilde{x}_1^k, \ldots, \tilde{x}_N^k, \tilde{\lambda}^k)$ by the following alternating minimization scheme:

$$\begin{cases} \tilde{x}_i^k = \arg\min\{\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_N^k, \lambda^k) \mid x_i \in \mathcal{X}_i\}, \quad i = 1, \ldots, N, \\ \tilde{\lambda}^k = \lambda^k - H(\sum_{i=1}^N A_i\tilde{x}_i^k - b). \end{cases} \tag{5}$$

Then, the next iterate $(x_1^{k+1}, \ldots, x_N^{k+1}, \lambda^{k+1})$ is obtained from $(x_1^k, \ldots, x_N^k, \lambda^k)$ and the trial point $(\tilde{x}_1^k, \ldots, \tilde{x}_N^k, \tilde{\lambda}^k)$. To the best of our knowledge, it is the first time that tractable algorithms for (3) based on the full utilization of their separable structure have been developed. As demonstrated in [16], the resulting method falls into the frameworks of both the descent-like methods in the sense that the iterates generated by the extended ADMM scheme (5) can be used to construct descent directions and the contraction-type methods (according to the definition in [17]) as the distance between the iterates and the solution set of (3) is monotonically decreasing. Therefore, the method is called the alternating-direction-based contraction method.

From a practical point of view, however, in many cases, solving each subproblem in (5) accurately is either expensive or even impossible. On the other hand, there seems to be little justification of the effort required to calculate the accurate solutions at each iteration. In this paper, we develop two inexact methods for solving problem (3), in which the subproblems in (5) are solved inexactly. So, the methods presented in this paper are named inexact alternating-direction-based contraction methods.

In the sequel, we briefly review some basic properties and related definitions that will be used in the forthcoming analysis. We consider the generally separable linearly constrained convex optimization problem (3). For $i = 1, \dots, N$, we let $f_i(x_i) \in \partial(\theta_i(x_i))$, where $\partial(\theta_i(x_i))$ denotes the subdifferential of θ_i at x_i. Moreover, we let $\mathcal{W} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \times \mathbb{R}^m$. Then, it is evident that the first-order optimality condition of (3) is equivalent to the following variational inequality for finding $(x_1^*, \dots, x_N^*, \lambda^*) \in \mathcal{W}$:

$$\begin{cases} \langle x_i - x_i^*, f_i(x_i^*) - A_i^T \lambda^* \rangle \ge 0, & i = 1, \dots, N, \\ \langle \lambda - \lambda^*, \sum_{i=1}^N A_i x_i^* - b \rangle \ge 0, \end{cases} \quad \forall (x_1, \dots, x_N, \lambda) \in \mathcal{W}. \tag{6}$$

By defining

$$w := \begin{pmatrix} x_1 \\ \vdots \\ x_N \\ \lambda \end{pmatrix} \quad \text{and} \quad F(w) := \begin{pmatrix} f_1(x_1) - A_1^T \lambda \\ \vdots \\ f_N(x_N) - A_N^T \lambda \\ \sum_{i=1}^N A_i x_i - b \end{pmatrix}$$

the problem (6) can be rewritten in a more compact form as

$$\langle w' - w^*, F(w^*) \rangle \ge 0 \quad \forall w' \in \mathcal{W},$$

which we denote by VI$(\mathcal{W}, F)$. Recall that $F(w)$ is said to be monotone iff

$$\langle u - v, F(u) - F(v) \rangle \ge 0 \quad \forall u, v \in \mathcal{W}.$$

One can easily verify that $F(w)$ is monotone whenever all f_i, $i = 1, \dots, N$, are monotone. Under the nonempty assumption of the solution set of (3), the solution set, $\mathcal{W}^*$, of VI$(\mathcal{W}, F)$ is nonempty and convex.

Given a positive definite matrix G of size $(n+m) \times (n+m)$, we define the G-norm of $u \in \mathbb{R}^{n+m}$ as $\|u\|_G = \sqrt{\langle u, Gu \rangle}$. The projection onto $\mathcal{W}$ under the G-norm is defined as

$$P_{\mathcal{W},G}(v) := \arg\min \{ \|v - w\|_G \mid w \in \mathcal{W} \}.$$

From the above definition it follows that

$$\langle v - P_{\mathcal{W},G}(v), G(w - P_{\mathcal{W},G}(v)) \rangle \le 0 \quad \forall v \in \mathbb{R}^{n+m}, \forall w \in \mathcal{W}. \tag{7}$$

Consequently, we have

$$\| P_{\mathcal{W},G}(u) - P_{\mathcal{W},G}(v) \|_G \le \|u - v\|_G \quad \forall u, v \in \mathbb{R}^{n+m}$$

and

$$\left\| P_{\mathcal{W},G}(v) - w \right\|_G^2 \le \|v - w\|_G^2 - \left\| v - P_{\mathcal{W},G}(v) \right\|_G^2 \quad \forall v \in \mathbb{R}^{n+m}, \forall w \in \mathcal{W}. \tag{8}$$

An important property of the projection is contained in the following lemma, for which the omitted proof can be found in [18, pp. 267].

Lemma 2.1 *Let $\mathcal{W}$ be a closed convex set in $\mathbb{R}^{n+m}$, and let G be any $(n+m) \times (n+m)$ positive definite matrix. Then w^* is a solution of VI$(\mathcal{W}, F)$ if and only if*

$$w^* = P_{\mathcal{W},G}\left[w^* - \alpha G^{-1} F\left(w^*\right)\right] \quad \forall \alpha > 0.$$

In other words, we have

$$w^* = P_{\mathcal{W},G}\left[w^* - \alpha G^{-1} F\left(w^*\right)\right] \quad \Leftrightarrow \quad w^* \in \mathcal{W}, \quad \left\langle w - w^*, F\left(w^*\right)\right\rangle \ge 0 \quad \forall w \in \mathcal{W}. \tag{9}$$

3 Two Contraction Methods

In this section, we present two algorithms, each of which consists of a search direction and a step length. The search directions are based on the inexact alternating minimization of $\mathcal{L}_A$, while both algorithms use the same step length.

3.1 Search Directions

For given $w^k = (x_1^k, \dots, x_N^k, \lambda^k)$, the alternating direction scheme of the kth iteration generates a trial point $\tilde{w}^k = (\tilde{x}_1^k, \dots, \tilde{x}_N^k, \tilde{\lambda}^k)$ via the following procedure:

For $i = 1, \dots, N$, $\tilde{x}_i^k$ is computed as

$$\tilde{x}_i^k = P_{\mathcal{X}_i}\left\{ \tilde{x}_i^k - \left[f_i\left(\tilde{x}_i^k\right) - A_i^T \lambda^k + A_i^T H \left(\sum_{j=1}^{i} A_j \tilde{x}_j^k + \sum_{j=i+1}^{N} A_j x_j^k - b \right) + \xi_i^k \right] \right\}, \tag{10}$$

where

$$\|\xi_i^k\| \le \left\| A_i\left(x_i^k - \tilde{x}_i^k\right) \right\|_H \quad \text{and} \quad \left|\left\langle x_i^k - \tilde{x}_i^k, \xi_i^k \right\rangle\right| \le \frac{1}{4} \left\| A_i\left(x_i^k - \tilde{x}_i^k\right) \right\|_H^2. \tag{11}$$

Then, we set

$$\tilde{\lambda}^k = \lambda^k - H\left(\sum_{j=1}^{N} A_j \tilde{x}_j^k - b \right). \tag{12}$$

We claim that the above approximation scheme is appropriate for inexactly solving the minimization subproblems, i.e.,

$$\min\left\{ \mathcal{L}_A\left(\tilde{x}_1^k, \dots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \dots, x_N^k, \lambda^k\right) \mid x_i \in \mathcal{X}_i \right\}. \tag{13}$$

Assume that $\bar{x}_i^k$ is an optimal solution of (13). By the definition of $\mathcal{L}_A$ in (4), the first-order optimality condition of (13) reduces to finding $\bar{x}_i^k \in \mathcal{X}_i$ such that

$$\left\langle x_i' - \bar{x}_i^k, f_i(\bar{x}_i^k) - A_i^T \left[\lambda^k - H\left(\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + A_i \bar{x}_i^k + \sum_{j=i+1}^{N} A_j x_j^k - b\right)\right]\right\rangle \geq 0 \quad \forall x_i' \in \mathcal{X}_i$$

where $f_i(\bar{x}_i^k)$ is a subgradient of $\theta_i(x_i)$ at $\bar{x}_i^k$, i.e., $f_i(\bar{x}_i^k) \in \partial(\theta_i(\bar{x}_i^k))$. According to Lemma 2.1, the above variational inequality is equivalent to

$$\bar{x}_i^k = P_{\mathcal{X}_i}\left\{\bar{x}_i^k - \left[f_i(\bar{x}_i^k) - A_i^T \lambda^k + A_i^T H\left(\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + A_i \bar{x}_i^k + \sum_{j=i+1}^{N} A_j x_j^k - b\right)\right]\right\}.$$

If $\tilde{x}_i^k$ is an optimal solution of (13), then by setting $\xi_i^k = 0$ in (10) the approximation conditions in (11) are satisfied. In fact, if $\|A_i(x_i^k - \bar{x}_i^k)\|_H = 0$, then the approximation conditions in (11) require that the corresponding subproblem (10) be solved exactly. Thus, without loss of generality, we may assume that

$$\left\|A_i(x_i^k - \bar{x}_i^k)\right\|_H \neq 0.$$

Suppose that $\hat{x}_i^k$ is an approximate solution of (13), i.e.,

$$\hat{x}_i^k \approx P_{\mathcal{X}_i}\left\{\hat{x}_i^k - \left[f_i(\hat{x}_i^k) - A_i^T \lambda^k + A_i^T H\left(\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + A_i \hat{x}_i^k + \sum_{j=i+1}^{N} A_j x_j^k - b\right)\right]\right\}.$$

By setting

$$\tilde{x}_i^k = P_{\mathcal{X}_i}\left\{\hat{x}_i^k - \left[f_i(\hat{x}_i^k) - A_i^T \lambda^k + A_i^T H\left(\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + A_i \hat{x}_i^k + \sum_{j=i+1}^{N} A_j x_j^k - b\right)\right]\right\}$$

simple manipulation shows that (10) holds with

$$\xi_i^k = (\tilde{x}_i^k - \hat{x}_i^k) - (f(\tilde{x}_i^k) - f(\hat{x}_i^k)) - A_i^T H A_i(\tilde{x}_i^k - \hat{x}_i^k).$$

When $\hat{x}_i^k$ is sufficiently close to the exact solution $\bar{x}_i^k$, the relations between $\bar{x}_i^k$, $\hat{x}_i^k$, and $\tilde{x}_i^k$ ensure that

$$\hat{x}_i^k \approx \tilde{x}_i^k \quad \text{and} \quad \left\|A_i(x_i^k - \tilde{x}_i^k)\right\|_H \approx \left\|A_i(x_i^k - \bar{x}_i^k)\right\|_H.$$

Therefore, for a suitable approximate solution $\hat{x}_i^k$, we can define an $\tilde{x}_i^k$ such that the inexactness conditions (10) and (11) are satisfied.

Instead of accepting $\tilde{w}^k$ as a new iterate, we use it to generate descent search directions. In the sequel, we define the matrix M and the vector ξ^k as

$$M := \begin{pmatrix} A_1^T H A_1 & 0 & \cdots & 0 & 0 \\ A_2^T H A_1 & A_2^T H A_2 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ A_N^T H A_1 & A_N^T H A_2 & \cdots & A_N^T H A_N & 0 \\ 0 & 0 & \cdots & 0 & H^{-1} \end{pmatrix} \quad \text{and} \quad \xi^k := \begin{pmatrix} \xi_1^k \\ \xi_2^k \\ \vdots \\ \xi_N^k \\ 0 \end{pmatrix}. \tag{14}$$

Then, our two search directions are, respectively, given by

$$d_1\bigl(w^k, \tilde{w}^k, \xi^k\bigr) = M\bigl(w^k - \tilde{w}^k\bigr) - \xi^k \tag{15}$$

and

$$d_2\bigl(w^k, \tilde{w}^k, \xi^k\bigr) = F\bigl(\tilde{w}^k\bigr) + \begin{pmatrix} A_1^T \\ \vdots \\ A_N^T \\ 0 \end{pmatrix} H\Biggl(\sum_{j=1}^N A_j\bigl(x_j^k - \tilde{x}_j^k\bigr)\Biggr). \tag{16}$$

Based on w^k, $\tilde{w}^k$, and ξ^k, we define

$$\varphi\bigl(w^k, \tilde{w}^k, \xi^k\bigr) = \bigl\langle w^k - \tilde{w}^k, d_1\bigl(w^k, \tilde{w}^k, \xi^k\bigr)\bigr\rangle + \Biggl\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=1}^N A_j\bigl(x_j^k - \tilde{x}_j^k\bigr)\Biggr\rangle. \tag{17}$$

The function $\varphi(w^k, \tilde{w}^k, \xi^k)$ is a key component in analyzing the proposed methods. In the subsequent section, we prove (in Theorem 4.1) that

$$\begin{cases} \varphi(w^k, \tilde{w}^k, \xi^k) \geq \frac{1}{4}(\sum_{j=1}^N \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2), \\ \varphi(w^k, \tilde{w}^k, \xi^k) = 0 \quad \Leftrightarrow \quad \tilde{w}^k \in \mathcal{W}^*. \end{cases} \tag{18}$$

Hence, $\varphi(w^k, \tilde{w}^k, \xi^k)$ can be viewed as an error measuring function, which measures how much w^k fails to be in $\mathcal{W}^*$. Furthermore, by utilizing $\varphi(w^k, \tilde{w}^k, \xi^k)$ we will prove in Theorems 4.2 and 4.3 that, for any given positive definite matrix $G \in \mathbb{R}^{(n+m)\times(n+m)}$, both $-G^{-1}d_1(w^k, \tilde{w}^k, \xi^k)$ and $-G^{-1}d_2(w^k, \tilde{w}^k, \xi^k)$ are descent directions with respect to the unknown distance function $\|w - w^*\|_G^2$.

3.2 The Step Length

In this subsection, we provide a step length for the search directions $d_1(w^k, \tilde{w}^k, \xi^k)$ and $d_2(w^k, \tilde{w}^k, \xi^k)$. Later, we will justify the choice of the step length to be defined and show that the step length has a positive lower bound.

Given a positive definite matrix $G \in \mathbb{R}^{(n+m)\times(n+m)}$, the new iterate w^{k+1} is generated as

$$w^{k+1} = w^k - \alpha_k G^{-1} d_1\big(w^k, \tilde{w}^k, \xi^k\big) \tag{19}$$

or

$$w^{k+1} = P_{\mathcal{W},G}\big[w^k - \alpha_k G^{-1} d_2\big(w^k, \tilde{w}^k, \xi^k\big)\big], \tag{20}$$

where

$$\alpha_k = \gamma \alpha_k^*, \quad \alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k, \xi^k)}{\|G^{-1} d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2}, \quad \text{and} \quad \gamma \in (0, 2). \tag{21}$$

The sequence $\{w^k\}$ generated by (19) is not necessarily contained in $\mathcal{W}$, while the sequence produced by (20) lies in $\mathcal{W}$. Note that the step length α_k in both (19) and (20) depends merely on $\varphi(w^k, \tilde{w}^k, \xi^k)$, $d_1(w^k, \tilde{w}^k, \xi^k)$, and γ. The proposed methods utilize different search directions but the same step length. According to our numerical experiments in [19], the update form (20) usually outperforms (19), provided that the projection onto $\mathcal{W}$ can easily be carried out.

We mention that the proposed inexact ADMMs (19) and (20) are different from those in the literature [19, 20]. In fact, the inexact methods proposed in [19, 20] belong to the proximal-type methods [21–24], while the ADMM subproblems in this paper do not include proximal terms of any form.

4 Rationale of the Two Directions

To derive the convergence of the proposed methods, we use similar arguments as those in the general framework proposed in [25]. By Lemma 2.1 the equality in (10) is equivalent to

$$\left\langle x_i' - \tilde{x}_i^k, f_i\big(\tilde{x}_i^k\big) - A_i^T\left[\lambda^k - H\left(\sum_{j=1}^{i} A_j \tilde{x}_j^k + \sum_{j=i+1}^{N} A_j x_j^k - b\right)\right] + \xi_i^k \right\rangle \ge 0 \quad \forall x_i' \in \mathcal{X}_i.$$

By substituting $\tilde{\lambda}^k$ given in (12), the above inequality can be rewritten as

$$\left\langle x_i' - \tilde{x}_i^k, f_i\big(\tilde{x}_i^k\big) - A_i^T \tilde{\lambda}^k + A_i^T H\left(\sum_{j=i+1}^{N} A_j\big(x_j^k - \tilde{x}_j^k\big)\right) + \xi_i^k \right\rangle \ge 0 \quad \forall x_i' \in \mathcal{X}_i. \tag{22}$$

According to the general framework [25], for the pair $(w^k, \tilde{w}^k)$, the following conditions are required to guarantee the convergence:

$$\tilde{w}^k = P_{\mathcal{W}}\big\{\tilde{w}^k - \big[d_2\big(w^k, \tilde{w}^k, \xi^k\big) - d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big]\big\}, \tag{23}$$

$$\big\langle \tilde{w}^k - w^*, d_2\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle \ge \varphi\big(w^k, \tilde{w}^k, \xi^k\big) - \big\langle w^k - \tilde{w}^k, d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle, \tag{24}$$

for which we give proofs in Lemmas 4.1 and 4.2, respectively.

Lemma 4.1 *Let $\tilde{w}^k = (\tilde{x}_1^k, \dots, \tilde{x}_N^k, \tilde{\lambda}^k)$ be generated by the inexact alternating direction scheme (10)–(12) from the given vector $w^k = (x_1^k, \dots, x_N^k, \lambda^k)$. Then, we have $\tilde{w}^k \in \mathcal{W}$ and*

$$\langle w' - \tilde{w}^k, d_2(w^k, \tilde{w}^k, \xi^k) - d_1(w^k, \tilde{w}^k, \xi^k)\rangle \geq 0 \quad \forall w' \in \mathcal{W},$$

where $d_1(w^k, \tilde{w}^k, \xi^k)$ and $d_2(w^k, \tilde{w}^k, \xi^k)$ are defined in (15) and (16), respectively.

Proof Denote $\tilde{x}^k := (\tilde{x}_1^k, \dots, \tilde{x}_N^k)$ and $\mathcal{X} := \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$. It is clear from (10) that $\tilde{x}^k \in \mathcal{X}$. From (22), for all $x' \in \mathcal{X}$, we have

$$\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_N' - \tilde{x}_N^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(\tilde{x}_1^k) - A_1^T \tilde{\lambda}^k \\ f_2(\tilde{x}_2^k) - A_2^T \tilde{\lambda}^k \\ \vdots \\ f_N(\tilde{x}_N^k) - A_N^T \tilde{\lambda}^k \end{pmatrix} + \begin{pmatrix} A_1^T H(\sum_{j=2}^N A_j(x_j^k - \tilde{x}_j^k)) \\ A_2^T H(\sum_{j=3}^N A_j(x_j^k - \tilde{x}_j^k)) \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \xi_1^k \\ \xi_2^k \\ \vdots \\ \xi_N^k \end{pmatrix} \right\}$$
$$\geq 0. \tag{25}$$

By adding

$$\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_N' - \tilde{x}_N^k \end{pmatrix}^T \begin{pmatrix} A_1^T H(\sum_{j=1}^1 A_j(x_j^k - \tilde{x}_j^k)) \\ A_2^T H(\sum_{j=1}^2 A_j(x_j^k - \tilde{x}_j^k)) \\ \vdots \\ A_N^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \end{pmatrix}$$

to both sides of (25), we get

$$\begin{aligned} &\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_N' - \tilde{x}_N^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(\tilde{x}_1^k) - A_1^T \tilde{\lambda}^k + A_1^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \\ f_2(\tilde{x}_2^k) - A_2^T \tilde{\lambda}^k + A_2^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \\ \vdots \\ f_N(\tilde{x}_N^k) - A_N^T \tilde{\lambda}^k + A_N^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \end{pmatrix} + \begin{pmatrix} \xi_1^k \\ \xi_2^k \\ \vdots \\ \xi_N^k \end{pmatrix} \right\} \\ &\geq \begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_N' - \tilde{x}_N^k \end{pmatrix}^T \begin{pmatrix} A_1^T H(\sum_{j=1}^1 A_j(x_j^k - \tilde{x}_j^k)) \\ A_2^T H(\sum_{j=1}^2 A_j(x_j^k - \tilde{x}_j^k)) \\ \vdots \\ A_N^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \end{pmatrix} \\ &= \begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_N' - \tilde{x}_N^k \end{pmatrix}^T \begin{pmatrix} A_1^T H A_1 & 0 & \cdots & 0 \\ A_2^T H A_1 & A_2^T H A_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ A_N^T H A_1 & A_N^T H A_2 & \cdots & A_N^T H A_N \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_N^k - \tilde{x}_N^k \end{pmatrix} \end{aligned} \tag{26}$$

for all $x' \in \mathcal{X}$. Since $\sum_{i=1}^N A_i\tilde{x}_i^k - b = H^{-1}(\lambda^k - \tilde{\lambda}^k)$, by embedding the equality

$$\left\langle \lambda' - \tilde{\lambda}^k, \sum_{i=1}^N A_i\tilde{x}_i^k - b\right\rangle = \left\langle \lambda' - \tilde{\lambda}^k, H^{-1}\left(\lambda^k - \tilde{\lambda}^k\right)\right\rangle$$

into (26) we obtain $\tilde{w}^k \in \mathcal{W}$ and

$$\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_N' - \tilde{x}_N^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(\tilde{x}_1^k) - A_1^T\tilde{\lambda}^k + A_1^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \\ f_2(\tilde{x}_2^k) - A_2^T\tilde{\lambda}^k + A_2^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \\ \vdots \\ f_N(\tilde{x}_N^k) - A_N^T\tilde{\lambda}^k + A_N^T H(\sum_{j=1}^N A_j(x_j^k - \tilde{x}_j^k)) \\ \sum_{i=1}^N A_i\tilde{x}_i^k - b \end{pmatrix} + \begin{pmatrix} \xi_1^k \\ \xi_2^k \\ \vdots \\ \xi_N^k \\ 0 \end{pmatrix} \right\}$$

$$\geq \begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_N' - \tilde{x}_N^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} A_1^T HA_1 & 0 & \cdots & 0 & 0 \\ A_2^T HA_1 & A_2^T HA_2 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ A_N^T HA_1 & A_N^T HA_2 & \cdots & A_N^T HA_N & 0 \\ 0 & 0 & \cdots & 0 & H^{-1} \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_N^k - \tilde{x}_N^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix}$$

for all $w' \in \mathcal{W}$. Recalling the definitions of $d_2(w^k, \tilde{w}^k, \xi^k)$ and M (see (16) and (14)), the last inequality can be rewritten in the following compact form:

$$\tilde{w}^k \in \mathcal{W}, \quad \left\langle w' - \tilde{w}^k, d_2\left(w^k, \tilde{w}^k, \xi^k\right) + \xi^k\right\rangle \geq \left\langle w' - \tilde{w}^k, M\left(w^k - \tilde{w}^k\right)\right\rangle \quad \forall w' \in \mathcal{W}.$$

The assertion of this lemma follows directly from the above inequality and the definition of $d_1(w^k, \tilde{w}^k, \xi^k)$ in (15). □

According to (9), the assertion in Lemma 4.1 is equivalent to (23).

Lemma 4.2 *Let $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_N^k, \tilde{\lambda}^k)$ be generated by the inexact alternating direction scheme* (10)–(12) *from the given vector $w^k = (x_1^k, \ldots, x_N^k, \lambda^k)$. Then, we have*

$$\left\langle \tilde{w}^k - w^*, d_2\left(w^k, \tilde{w}^k, \xi^k\right)\right\rangle \geq \varphi\left(w^k, \tilde{w}^k, \xi^k\right) - \left\langle w^k - \tilde{w}^k, d_1\left(w^k, \tilde{w}^k, \xi^k\right)\right\rangle \quad \forall w^* \in \mathcal{W}^* \tag{27}$$

where $d_1(w^k, \tilde{w}^k, \xi^k)$, $d_2(w^k, \tilde{w}^k, \xi^k)$, and $\varphi(w^k, \tilde{w}^k, \xi^k)$ are defined in (15), (16), *and* (17).

Proof According to (16), we have

$$\begin{aligned} &\left\langle \tilde{w}^k - w^*, d_2\left(w^k, \tilde{w}^k, \xi^k\right)\right\rangle \\ &\quad = \left\langle \tilde{w}^k - w^*, F\left(\tilde{w}^k\right)\right\rangle + \left\langle \sum_{j=1}^N A_j\left(\tilde{x}_j^k - x_j^*\right), H\left(\sum_{j=1}^N A_j\left(x_j^k - \tilde{x}_j^k\right)\right)\right\rangle. \end{aligned} \tag{28}$$

From the monotonicity of $F(w)$ and the fact that $\tilde{w}^k \in \mathcal{W}$ it follows that

$$\langle \tilde{w}^k - w^*, F(\tilde{w}^k)\rangle \geq \langle \tilde{w}^k - w^*, F(w^*)\rangle \geq 0.$$

Substituting the above inequality into (28), we obtain

$$\langle \tilde{w}^k - w^*, d_2(w^k, \tilde{w}^k, \xi^k)\rangle \geq \left\langle \sum_{j=1}^{N} A_j(\tilde{x}_j^k - x_j^*), H\left(\sum_{j=1}^{N} A_j(x_j^k - \tilde{x}_j^k)\right)\right\rangle.$$

Since $\sum_{j=1}^N A_j x_j^* = b$ and $H(\sum_{j=1}^N A_j \tilde{x}_j^k - b) = \lambda^k - \tilde{\lambda}^k$, the above inequality becomes

$$\langle \tilde{w}^k - w^*, d_2(w^k, \tilde{w}^k, \xi^k)\rangle \geq \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=1}^{N} A_j(x_j^k - \tilde{x}_j^k)\right\rangle \quad \forall w^* \in \mathcal{W}^*.$$

Inequality (27) follows immediately by further considering the definition of $\varphi(w^k, \tilde{w}^k, \xi^k)$. □

Lemma 4.3 *Let $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_N^k, \tilde{\lambda}^k)$ be generated by the inexact alternating direction scheme* (10)–(12) *from the given vector $w^k = (x_1^k, \ldots, x_N^k, \lambda^k)$. Then, we have*

$$\begin{aligned}&\langle w^k - \tilde{w}^k, M(w^k - \tilde{w}^k)\rangle + \left\langle \lambda^k - \tilde{\lambda}^k, \sum_{j=1}^{N} A_j(x_j^k - \tilde{x}_j^k)\right\rangle \\ &\quad = \frac{1}{2}\left(\sum_{j=1}^{N} \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 + \left\|\sum_{j=1}^{N} A_j x_j^k - b\right\|_H^2\right).\end{aligned}$$

Proof From the definition of the matrix M in (14) we have

$$M(w^k - \tilde{w}^k) = \begin{pmatrix} A_1^T H A_1 & 0 & \cdots & 0 & 0 \\ A_2^T H A_1 & A_2^T H A_2 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ A_N^T H A_1 & A_N^T H A_2 & \cdots & A_N^T H A_N & 0 \\ 0 & 0 & \cdots & 0 & H^{-1} \end{pmatrix} \begin{pmatrix} x_1^k - \tilde{x}_1^k \\ x_2^k - \tilde{x}_2^k \\ \vdots \\ x_N^k - \tilde{x}_N^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix},$$

from which we can easily verify the following equality:

$$\begin{aligned}&\langle w^k - \tilde{w}^k, M(w^k - \tilde{w}^k)\rangle \\ &\quad = \frac{1}{2}\left(\sum_{j=1}^{N} \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2\right)\end{aligned}$$

$$+\frac{1}{2}\begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_N(x_N^k-\tilde{x}_N^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} H & H & \cdots & H & 0 \\ H & H & \cdots & H & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H & H & \cdots & H & 0 \\ 0 & 0 & \cdots & 0 & H^{-1} \end{pmatrix} \begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_N(x_N^k-\tilde{x}_N^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}. \tag{29}$$

Simple calculations show that

$$\begin{aligned} &\left\langle \lambda^k-\tilde{\lambda}^k, \sum_{j=1}^N A_j\left(x_j^k-\tilde{x}_j^k\right)\right\rangle \\ &= \frac{1}{2}\begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_N(x_N^k-\tilde{x}_N^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} 0 & 0 & \cdots & 0 & I \\ 0 & 0 & \cdots & 0 & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & I \\ I & I & \ldots & I & 0 \end{pmatrix} \begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_N(x_N^k-\tilde{x}_N^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}. \end{aligned} \tag{30}$$

The addition of (29) and (30) yields

$$\begin{aligned} &\left\langle w^k-\tilde{w}^k, M\left(w^k-\tilde{w}^k\right)\right\rangle + \left\langle \lambda^k-\tilde{\lambda}^k, \sum_{j=1}^N A_j\left(x_j^k-\tilde{x}_j^k\right)\right\rangle \\ &= \frac{1}{2}\begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_N(x_N^k-\tilde{x}_N^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T \begin{pmatrix} H & H & \ldots & H & I \\ H & H & \ldots & H & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H & H & \cdots & H & I \\ I & I & \cdots & I & H^{-1} \end{pmatrix} \begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_N(x_N^k-\tilde{x}_N^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \\ &\quad + \frac{1}{2}\left(\sum_{j=1}^N \left\|A_j\left(x_j^k-\tilde{x}_j^k\right)\right\|_H^2 + \left\|\lambda^k-\tilde{\lambda}^k\right\|_{H^{-1}}^2\right) \\ &= \frac{1}{2}\left\|\sum_{j=1}^N A_j\left(x_j^k-\tilde{x}_j^k\right) + H^{-1}\left(\lambda^k-\tilde{\lambda}^k\right)\right\|_H^2 \\ &\quad + \frac{1}{2}\left(\sum_{j=1}^N \left\|A_j\left(x_j^k-\tilde{x}_j^k\right)\right\|_H^2 + \left\|\lambda^k-\tilde{\lambda}^k\right\|_{H^{-1}}^2\right) \\ &= \frac{1}{2}\left\|\sum_{j=1}^N A_j x_j^k - b\right\|_H^2 + \frac{1}{2}\left(\sum_{j=1}^N \left\|A_j\left(x_j^k-\tilde{x}_j^k\right)\right\|_H^2 + \left\|\lambda^k-\tilde{\lambda}^k\right\|_{H^{-1}}^2\right), \end{aligned}$$

where the last equality follows from (12). The lemma is proved. □

Now, we prove (18) and the descent properties of

$$-d_1(w^k, \tilde{w}^k, \xi^k) \quad \text{and} \quad -d_2(w^k, \tilde{w}^k, \xi^k).$$

Theorem 4.1 *Let $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_N^k, \tilde{\lambda}^k)$ be generated by the inexact alternating direction scheme (10)–(12) from the given vector $w^k = (x_1^k, \ldots, x_N^k, \lambda^k)$. If the inexactness criteria (11) are satisfied, then*

$$\varphi\bigl(w^k, \tilde{w}^k, \xi^k\bigr) \geq \frac{1}{4}\Biggl(\sum_{j=1}^{N}\bigl\|A_j\bigl(x_j^k - \tilde{x}_j^k\bigr)\bigr\|_H^2 + \bigl\|\lambda^k - \tilde{\lambda}^k\bigr\|_{H^{-1}}^2\Biggr). \tag{31}$$

In addition, if $\varphi(w^k, \tilde{w}^k, \xi^k) = 0$, then $\tilde{w}^k \in \mathcal{W}^*$ is a solution of VI$(\mathcal{W}, F)$.

Proof First, it follows from (15), (17), and Lemma 4.3 that

$$\begin{aligned}\varphi\bigl(w^k, \tilde{w}^k, \xi^k\bigr) = {} & \frac{1}{2}\Biggl(\sum_{j=1}^{N}\bigl\|A_j\bigl(x_j^k - \tilde{x}_j^k\bigr)\bigr\|_H^2 + \bigl\|\lambda^k - \tilde{\lambda}^k\bigr\|_{H^{-1}}^2\Biggr) \\ & + \frac{1}{2}\Biggl\|\sum_{j=1}^{N} A_j x_j^k - b\Biggr\|_H^2 - \bigl\langle w^k - \tilde{w}^k, \xi^k\bigr\rangle.\end{aligned} \tag{32}$$

From the inexactness criteria (11) we have

$$-\bigl\langle w^k - \tilde{w}^k, \xi^k\bigr\rangle \geq -\frac{1}{4}\sum_{j=1}^{N}\bigl\|A_j\bigl(x_j^k - \tilde{x}_j^k\bigr)\bigr\|_H^2.$$

Substituting the above inequality into (32), the first part of the theorem follows immediately. Consequently, if $\varphi(w^k, \tilde{w}^k, \xi^k) = 0$, it follows that

$$A_j\bigl(x_j^k - \tilde{x}_j^k\bigr) = 0, \quad j = 1, \ldots, N, \quad \text{and} \quad \lambda^k = \tilde{\lambda}^k.$$

Moreover, it follows from (11) that $\xi_j^k = 0$, $j = 1, \ldots, N$. By substituting the above equations into (22) we get

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad \bigl\langle x_i - \tilde{x}_i^k, f_i\bigl(\tilde{x}_i^k\bigr) - A_i^T\tilde{\lambda}^k\bigr\rangle \geq 0 \quad \forall x_i^k \in \mathcal{X}_i, i = 1, \ldots, N, \tag{33}$$

and

$$\sum_{j=1}^{N} A_j\tilde{x}_j^k - b = H^{-1}\bigl(\lambda^k - \tilde{\lambda}^k\bigr) = 0. \tag{34}$$

Combining (33) and (34), we get

$$\tilde{w}^k \in \mathcal{W}, \quad \bigl\langle w - \tilde{w}^k, F\bigl(\tilde{w}^k\bigr)\bigr\rangle \geq 0 \quad \forall w \in \mathcal{W},$$

and thus $\tilde{w}^k$ is a solution of VI$(\mathcal{W}, F)$. □

Theorem 4.2 *Let $\tilde{w}^k = (\tilde{x}_1^k, \dots, \tilde{x}_N^k, \tilde{\lambda}^k)$ be generated by the inexact alternating direction scheme (10)–(12) from the given vector $w^k = (x_1^k, \dots, x_N^k, \lambda^k)$. Then, we have*

$$\langle w^k - w^*, d_1(w^k, \tilde{w}^k, \xi^k)\rangle \geq \varphi(w^k, \tilde{w}^k, \xi^k) \quad \forall w^* \in \mathcal{W}^*, \tag{35}$$

where $d_1(w^k, \tilde{w}^k, \xi^k)$ and $\varphi(w^k, \tilde{w}^k, \xi^k)$ are defined in (15) and (17), respectively.

Proof First, it follows from Lemma 4.1 that

$$\langle \tilde{w}^k - w^*, d_1(w^k, \tilde{w}^k, \xi^k)\rangle \geq \langle \tilde{w}^k - w^*, d_2(w^k, \tilde{w}^k, \xi^k)\rangle \quad \forall w^* \in \mathcal{W}.$$

Combining with Lemma 4.2, we have

$$\langle \tilde{w}^k - w^*, d_1(w^k, \tilde{w}^k, \xi^k)\rangle \geq \varphi(w^k, \tilde{w}^k, \xi^k) - \langle w^k - \tilde{w}^k, d_1(w^k, \tilde{w}^k, \xi^k)\rangle \quad \forall w^* \in \mathcal{W}^*$$

which implies (35). □

Recall that the sequence $\{w^k\}$ generated by (19) is not necessarily contained in $\mathcal{W}$. Therefore, the w^k in Theorem 4.2 must be allowed to be any point in $\mathbb{R}^{m+n}$. In contrast, the sequence produced by (20) lies in $\mathcal{W}$, and thus it is required in the next theorem that w^k belongs to $\mathcal{W}$.

Theorem 4.3 *Let $\tilde{w}^k = (\tilde{x}_1^k, \dots, \tilde{x}_N^k, \tilde{\lambda}^k)$ be generated by the inexact alternating directions scheme (10)–(12) from the given vector $w^k = (x_1^k, \dots, x_N^k, \lambda^k)$. If $w^k \in \mathcal{W}$, then*

$$\langle w^k - w^*, d_2(w^k, \tilde{w}^k, \xi^k)\rangle \geq \varphi(w^k, \tilde{w}^k, \xi^k) \quad \forall w^* \in \mathcal{W}^*, \tag{36}$$

where $d_2(w^k, \tilde{w}^k, \xi^k)$ and $\varphi(w^k, \tilde{w}^k, \xi^k)$ are defined in (16) and (17), respectively.

Proof Since $w^k \in \mathcal{W}$, it follows from Lemma 4.1 that

$$\langle w^k - \tilde{w}^k, d_2(w^k, \tilde{w}^k, \xi^k)\rangle \geq \langle w^k - \tilde{w}^k, d_1(w^k, \tilde{w}^k, \xi^k)\rangle. \tag{37}$$

Then, the addition of (27) to both sides of (37) yields (36). □

5 Convergence

Using the directions $d_1(w^k, \tilde{w}^k, \xi^k)$ and $d_2(w^k, \tilde{w}^k, \xi^k)$ offered by (15) and (16), the new iterate w^{k+1} is determined by the positive definite matrix G and the step length α_k, see (19) and (20). In order to explain how to determine the step length, we define the new step-length-dependent iterate by

$$w_1^{k+1}(\alpha_k) = w^k - \alpha_k G^{-1} d_1(w^k, \tilde{w}^k, \xi^k) \tag{38}$$

and

$$w_2^{k+1}(\alpha_k) = P_{\mathcal{W},G}[w^k - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k, \xi^k)]. \tag{39}$$

In this way,

$$\vartheta_1(\alpha_k) = \|w^k - w^*\|_G^2 - \|w_1^{k+1}(\alpha_k) - w^*\|_G^2 \tag{40}$$

and

$$\vartheta_2(\alpha_k) = \|w^k - w^*\|_G^2 - \|w_2^{k+1}(\alpha_k) - w^*\|_G^2 \tag{41}$$

measure the improvement in the kth iteration by using updating forms (38) and (39), respectively. Since the optimal solution $w^* \in \mathcal{W}^*$ is unknown, it is generally infeasible to maximize the improvement directly. The following theorem introduces a tight lower bound on $\vartheta_1(\alpha_k)$ and $\vartheta_2(\alpha_k)$, which does not depend on the unknown vector w^*.

Theorem 5.1 *For any $w^* \in \mathcal{W}^*$ and $\alpha_k \geq 0$, we have*

$$\vartheta_1(\alpha_k) \geq q(\alpha_k) \quad \text{and} \quad \vartheta_2(\alpha_k) \geq q(\alpha_k),$$

where

$$q(\alpha_k) = 2\alpha_k \varphi\big(w^k, \tilde{w}^k, \xi^k\big) - \alpha_k^2 \big\|G^{-1} d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\|_G^2. \tag{42}$$

Proof From (38) and (40) we have

$$\begin{aligned}
\vartheta_1(\alpha_k) &= \|w^k - w^*\|_G^2 - \big\|w^k - w^* - \alpha_k G^{-1} d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\|_G^2 \\
&= 2\big\langle w^k - w^*, \alpha_k d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle - \alpha_k^2 \big\|G^{-1} d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\|_G^2 \\
&\geq 2\alpha_k \varphi\big(w^k, \tilde{w}^k, \xi^k\big) - \alpha_k^2 \big\|G^{-1} d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\|_G^2 = q(\alpha_k),
\end{aligned}$$

where the last inequality follows from (35). Hence, the first assertion of this theorem is proved. By setting $v = w^k - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k, \xi^k)$ and $u = w^*$ in (8) we get

$$\begin{aligned}
\big\|w_2^{k+1}(\alpha_k) - w^*\big\|_G^2 &\leq \big\|w^k - \alpha_k G^{-1} d_2\big(w^k, \tilde{w}^k, \xi^k\big) - w^*\big\|_G^2 \\
&\quad - \big\|w^k - \alpha_k G^{-1} d_2\big(w^k, \tilde{w}^k, \xi^k\big) - w_2^{k+1}(\alpha_k)\big\|_G^2.
\end{aligned}$$

Substituting the above inequality into (41), we obtain

$$\begin{aligned}
\vartheta_2(\alpha_k) &\geq \big\|w^k - w^*\big\|_G^2 - \big\|w^k - w^* - \alpha_k G^{-1} d_2\big(w^k, \tilde{w}^k, \xi^k\big)\big\|_G^2 \\
&\quad + \big\|w^k - w_2^{k+1}(\alpha_k) - \alpha_k G^{-1} d_2\big(w^k, \tilde{w}^k, \xi^k\big)\big\|_G^2 \\
&= \big\|w^k - w_2^{k+1}(\alpha_k)\big\|_G^2 + 2\big\langle w_2^{k+1}(\alpha_k) - w^*, \alpha_k d_2\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle.
\end{aligned} \tag{43}$$

Since $w_2^{k+1}(\alpha_k) \in \mathcal{W}$, it follows from Lemma 4.1 that

$$\big\langle w_2^{k+1}(\alpha_k) - \tilde{w}^k, d_2\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle \geq \big\langle w_2^{k+1}(\alpha_k) - \tilde{w}^k, d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle. \tag{44}$$

The addition of (27) to both sides of (44) yields

$$\begin{aligned}
&\big\langle w_2^{k+1}(\alpha_k) - w^*, d_2\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle \\
&\quad \geq \varphi\big(w^k, \tilde{w}^k, \xi^k\big) + \big\langle w_2^{k+1}(\alpha_k) - w^k, d_1\big(w^k, \tilde{w}^k, \xi^k\big)\big\rangle.
\end{aligned} \tag{45}$$

Substituting (45) into the right-hand side of (43), we obtain

$$\begin{aligned}\vartheta_2(\alpha_k) &\geq \|w^k - w_2^{k+1}(\alpha_k)\|_G^2 + 2\alpha_k\varphi(w^k, \tilde{w}^k, \xi^k) \\ &\quad + 2\alpha_k(w_2^{k+1}(\alpha_k) - w^k)^T d_1(w^k, \tilde{w}^k, \xi^k) \\ &= \|w^k - w_2^{k+1}(\alpha_k) - \alpha_k G^{-1} d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2 - \alpha_k^2 \|G^{-1} d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2 \\ &\quad + 2\alpha_k\varphi(w^k, \tilde{w}^k, \xi^k) \\ &\geq 2\alpha_k\varphi(w^k, \tilde{w}^k, \xi^k) - \alpha_k^2\|G^{-1}d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2 = q(\alpha_k).\end{aligned}$$

Hence, the proof is completed. □

Denote

$$\vartheta(\alpha_k) := \min\{\vartheta_1(\alpha_k), \vartheta_2(\alpha_k)\} \geq q(\alpha_k). \tag{46}$$

Note that $q(\alpha_k)$ is a quadratic function of α_k, and it reaches its maximum at

$$\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k, \xi^k)}{\|G^{-1}d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2}, \tag{47}$$

which is just the same as defined in (21). Consequently, it follows from Theorem 5.1 that

$$\vartheta(\alpha_k^*) \geq q(\alpha_k^*) = \alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k).$$

Since inequality (35) is used in the proof of Theorem 5.1, in practice, multiplication of the "optimal" step length α_k^* by a factor $\gamma > 1$ may result in faster convergence. By using (42) and (47) we have

$$\begin{aligned}q(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k) - (\gamma\alpha_k^*)^2\|G^{-1}d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2 \\ &= \gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k).\end{aligned} \tag{48}$$

In order to guarantee that the right-hand side of (48) is positive, we choose $\gamma \in [1, 2)$. The following theorem shows that the sequence $\{w^k\}$ generated by the proposed method is Fejèr monotone with respect to $\mathcal{W}^*$.

Theorem 5.2 *For any $w^* \in \mathcal{W}^*$, the sequence $\{w^k\}$ generated by each of the proposed methods (with update form (19) or (20)) satisfies*

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k) \quad \forall w^* \in \mathcal{W}^*. \tag{49}$$

Proof It follows from Theorem 5.1 that $\vartheta(\gamma\alpha_k^*) \geq q(\gamma\alpha_k^*)$, which is equivalent to

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - q(\gamma\alpha_k^*).$$

Then, the result of this theorem directly follows from (48). □

Theorem 5.2 indicates that the sequence $\{w^k\}$ converges to the solution set monotonically in the Fejèr sense. Thus, according to [16, 17], the proposed method belongs to the class of contraction methods. In the following, we show that $\alpha_k^* > 0$ has a positive lower bound.

Lemma 5.1 *For any given (but fixed) positive definite matrix G, there exists a constant $c_0 > 0$ such that $\alpha_k^* \geq c_0$ for all $k > 0$.*

Proof From the definition of M in (14) it is easy to show that

$$M(w^k - \tilde{w}^k) = \begin{pmatrix} A_1^T H^{1/2} & 0 & \cdots & 0 & 0 \\ A_2^T H^{1/2} & A_2^T H^{1/2} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ A_N^T H^{1/2} & A_N^T H^{1/2} & \cdots & A_N^T H^{1/2} & 0 \\ 0 & 0 & \cdots & 0 & H^{-1/2} \end{pmatrix} \times \begin{pmatrix} H^{1/2} A_1 (x_1^k - \tilde{x}_1^k) \\ H^{1/2} A_2 (x_2^k - \tilde{x}_2^k) \\ \vdots \\ H^{1/2} A_N (x_N^k - \tilde{x}_N^k) \\ H^{-1/2} (\lambda^k - \tilde{\lambda}^k) \end{pmatrix}.$$

Therefore, there exists a constant $K > 0$ such that

$$\|M(w^k - \tilde{w}^k)\|^2 \leq K \left(\sum_{j=1}^N \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \right).$$

Moreover, according to the inexactness criteria (11), we have

$$\|\xi^k\|^2 \leq \sum_{j=1}^N \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2.$$

Since $d_1(w^k, \tilde{w}^k, \xi^k) = M(w^k - \tilde{w}^k) - \xi^k$, we have

$$\begin{aligned} \|d_1(w^k, \tilde{w}^k, \xi^k)\|^2 &\leq 2\|M(w^k - \tilde{w}^k)\|^2 + 2\|\xi^k\|^2 \\ &\leq 2(K+1) \left(\sum_{j=1}^N \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \right). \end{aligned}$$

Using (31) and recalling that any two norms in a finite-dimensional space are equivalent, we have

$$\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k, \xi^k)}{\|d_1(w^k, \tilde{w}^k, \xi^k)\|_{G^{-1}}^2} \geq \frac{c'\varphi(w^k, \tilde{w}^k, \xi^k)}{\|d_1(w^k, \tilde{w}^k, \xi^k)\|^2} \geq c_0 := \frac{c'}{8(K+1)} > 0, \tag{50}$$

where $c' > 0$ is a constant. This completes the proof of the lemma. □

Theorem 5.3 *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the proposed alternating-direction-based contraction methods* (19) *and* (20) *for problem* (3). *Then, we have*

(i) *The sequence $\{w^k\}$ is bounded.*
(ii) $\lim_{k\to\infty}\{\sum_{j=1}^N \|A_j(x_j^k - \tilde{x}_j^k)\|^2 + \|\lambda^k - \tilde{\lambda}^k\|^2\} = 0$.
(iii) *Any accumulation point of $\{\tilde{w}^k\}$ is a solution of* (3).
(iv) *When A_i, $i = 1, \dots, N$, are full column rank matrices, the sequence $\{\tilde{w}^k\}$ converges to a unique $w^\infty \in \mathcal{W}^*$.*

Proof The first assertion follows from (49). Moreover, by combining the recursion of (49) and the fact that $\alpha_k^* \geq c_0 > 0$ it is easy to show that

$$\lim_{k\to\infty} \varphi\big(w^k, \tilde{w}^k, \xi^k\big) = 0.$$

Consequently, it follows from Theorem 4.1 that

$$\lim_{k\to\infty} \big\|A_i\big(x_i^k - \tilde{x}_i^k\big)\big\| = 0, \quad i = 1, \dots, N, \quad \text{and} \quad \lim_{k\to\infty} \big\|\lambda^k - \tilde{\lambda}^k\big\| = 0, \tag{51}$$

and the second assertion is proved. Using similar arguments as in the proof of Theorem 4.1, we obtain

$$\tilde{w}^k \in \mathcal{W}, \quad \lim_{k\to\infty} \big\langle w - \tilde{w}^k, F\big(\tilde{w}^k\big)\big\rangle \geq 0 \quad \forall w \in \mathcal{W}, \tag{52}$$

and thus any accumulation point of $\{\tilde{w}^k\}$ is a solution of VI($\mathcal{W}$, F), i.e., a solution of (3). If all A_i are full column rank matrices, it follows from the first assertion and (51) that $\{\tilde{w}^k\}$ is also bounded. Let w^∞ be an accumulation point of $\{\tilde{w}^k\}$. Then, there exists a subsequence $\{\tilde{w}^{k_j}\}$ that converges to w^∞. It follows from (52) that

$$\tilde{w}^{k_j} \in \mathcal{W}, \quad \lim_{k\to\infty} \big\langle w - \tilde{w}^{k_j}, F\big(\tilde{w}^{k_j}\big)\big\rangle \geq 0 \quad \forall w \in \mathcal{W},$$

and consequently, we have

$$w^\infty \in \mathcal{W}, \quad \big\langle w - w^\infty, F\big(w^\infty\big)\big\rangle \geq 0 \quad \forall w \in \mathcal{W},$$

which implies that $w^\infty \in \mathcal{W}^*$. Since $\{w^k\}$ is Fejèr monotone and $\lim_{k\to\infty} \|w^k - \tilde{w}^k\| = 0$, the sequence $\{\tilde{w}^k\}$ cannot have any other accumulation point and thus must converge to w^∞. □

6 Complexity

The analysis in this section is inspired by [26]. It is based on a key inequality (see Lemmas 6.1 and 6.2) that is similar to that in [26]. In the current framework of variational inequalities, the analysis becomes much simpler and more elegant. As a preparation for the proof of the complexity result, we first give an alternative characterization of the optimal solution set $\mathcal{W}^*$, namely,

$$\mathcal{W}^* = \bigcap_{w \in \mathcal{W}} \{w^* \in \mathcal{W} : \langle w - w^*, F(w) \rangle \geq 0\}.$$

For a proof of this characterization, we refer to Theorem 2.3.5 in the book [27]. According to the above alternative characterization, we have that $\hat{w} \in \mathcal{W}$ is an ϵ-optimal solution of $\mathrm{VI}(\mathcal{W}, F)$ if it satisfies

$$\hat{w} \in \mathcal{W} \quad \text{and} \quad \sup_{w \in \mathcal{W}} \{\langle \hat{w} - w, F(w) \rangle\} \leq \epsilon. \tag{53}$$

In general, our complexity analysis follows the lines of [28], but instead of using $\tilde{w}^k$ directly, we need to introduce an auxiliary vector, namely,

$$\hat{w}^k = \begin{pmatrix} \tilde{x}_1^k \\ \vdots \\ \tilde{x}_N^k \\ \hat{\lambda}^k \end{pmatrix}, \quad \text{where } \hat{\lambda}^k = \tilde{\lambda}^k - H \sum_{j=1}^N A_j (x_j^k - \tilde{x}_j^k).$$

Since $\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^N A_j \tilde{x}_j^k - b)$ (see (12)), we have $\hat{\lambda}^k = \lambda^k - H(\sum_{j=1}^N A_j x_j^k - b)$ or, equivalently,

$$H^{-1}(\lambda^k - \tilde{\lambda}^k) = \sum_{j=1}^N A_j \tilde{x}_j^k - b = -\sum_{j=1}^N A_j (x_j^k - \tilde{x}_j^k) + H^{-1}(\lambda^k - \hat{\lambda}^k).$$

As a consequence, we may rewrite the two descent directions separately as

$$d_1(w^k, \tilde{w}^k, \xi^k) = M(w^k - \tilde{w}^k) - \xi^k = \hat{M}(w^k - \hat{w}^k) - \xi^k =: \hat{d}_1(w^k, \hat{w}^k, \xi^k), \tag{54}$$

where

$$\hat{M} := \begin{pmatrix} A_1^T H A_1 & 0 & \cdots & 0 & 0 \\ A_2^T H A_1 & A_2^T H A_2 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & 0 & 0 \\ A_N^T H A_1 & A_N^T H A_2 & \cdots & A_N^T H A_N & 0 \\ -A_1 & -A_2 & \cdots & -A_N & H^{-1} \end{pmatrix}$$

and

$$d_2(w^k, \tilde{w}^k, \xi^k) = F(\tilde{w}^k) + \begin{pmatrix} A_1^T \\ \vdots \\ A_N^T \\ 0 \end{pmatrix} H \left(\sum_{j=1}^N A_j (x_j^k - \tilde{x}_j^k) \right) = F(\hat{w}^k). \tag{55}$$

In fact, we constructed $\hat{w}^k$ and $\hat{M}$ such that $\hat{d}_1(w^k, \hat{w}^k, \xi^k)$ is exactly the direction $d_1(w^k, \tilde{w}^k, \xi^k)$ and $F(\hat{w}^k)$ is exactly the direction $d_2(w^k, \tilde{w}^k, \xi^k)$. Moreover, the assertion in Lemma 4.1 can be rewritten accordingly as

$$\langle w - \hat{w}^k, F(\hat{w}^k) - \hat{d}_1(w^k, \hat{w}^k, \xi^k)\rangle \geq 0 \quad \forall w \in \mathcal{W}. \tag{56}$$

Now we are ready to prove the key inequality for both algorithms, which is given in the following two lemmas.

Lemma 6.1 *If the new iterate w^{k+1} is updated by* (19), *then we have*

$$\langle w - \hat{w}^k, \gamma\alpha_k^* F(\hat{w}^k)\rangle + \frac{1}{2}\big(\|w - w^k\|_G^2 - \|w - w^{k+1}\|_G^2\big) \geq 0 \quad \forall w \in \mathcal{W}.$$

Proof Due to (56), we have

$$\langle w - \hat{w}^k, \gamma\alpha_k^* F(\hat{w}^k)\rangle \geq \langle w - \hat{w}^k, \gamma\alpha_k^* \hat{d}_1(w^k, \hat{w}^k, \xi^k)\rangle \quad \forall w \in \mathcal{W}. \tag{57}$$

In addition, we have, by using (54) and (19),

$$\gamma\alpha_k^* \hat{d}_1(w^k, \hat{w}^k, \xi^k) = \gamma\alpha_k^* d_1(w^k, \tilde{w}^k, \xi^k) = G(w^k - w^{k+1}).$$

Substitution of the last equation into (57) yields

$$\langle w - \hat{w}^k, \gamma\alpha_k^* F(\hat{w}^k)\rangle \geq \langle w - \hat{w}^k, G(w^k - w^{k+1})\rangle.$$

Thus, it suffices to show that

$$\langle w - \hat{w}^k, G(w^k - w^{k+1})\rangle + \frac{1}{2}\big(\|w - w^k\|_G^2 - \|w - w^{k+1}\|_G^2\big) \geq 0 \quad \forall w \in \mathcal{W}. \tag{58}$$

Applying the equality

$$\langle a - b, G(c - d)\rangle + \frac{1}{2}\big(\|a - c\|_G^2 - \|a - d\|_G^2\big) = \frac{1}{2}\big(\|c - b\|_G^2 - \|d - b\|_G^2\big),$$

we derive that

$$\begin{aligned}
&\langle w - \hat{w}^k, G(w^k - w^{k+1})\rangle + \frac{1}{2}\big(\|w - w^k\|_G^2 - \|w - w^{k+1}\|_G^2\big) \\
&\quad = \frac{1}{2}\big(\|w^k - \hat{w}^k\|_G^2 - \|w^{k+1} - \hat{w}^k\|_G^2\big).
\end{aligned} \tag{59}$$

In view of (19), we have

$$\begin{aligned}
&\|w^k - \hat{w}^k\|_G^2 - \|w^{k+1} - \hat{w}^k\|_G^2 \\
&\quad = \|w^k - \hat{w}^k\|_G^2 - \|w^k - \hat{w}^k - \gamma\alpha_k^* G^{-1} d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2 \\
&\quad = 2\langle w^k - \hat{w}^k, \gamma\alpha_k^* d_1(w^k, \tilde{w}^k, \xi^k)\rangle - \gamma^2(\alpha_k^*)^2 \|G^{-1} d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2.
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
&\langle w^k-\hat{w}^k, d_1(w^k,\tilde{w}^k,\xi^k)\rangle \\
&\quad=\langle w^k-\tilde{w}^k, d_1(w^k,\tilde{w}^k,\xi^k)\rangle+\langle \tilde{w}^k-\hat{w}^k, d_1(w^k,\tilde{w}^k,\xi^k)\rangle \\
&\quad=\langle w^k-\tilde{w}^k, d_1(w^k,\tilde{w}^k,\xi^k)\rangle+\left\langle H\sum_{j=1}^N A_j(x_j^k-\tilde{x}_j^k), H^{-1}(\lambda^k-\tilde{\lambda}^k)\right\rangle \\
&\quad=\langle w^k-\tilde{w}^k, d_1(w^k,\tilde{w}^k,\xi^k)\rangle+\left\langle \lambda^k-\tilde{\lambda}^k, \sum_{j=1}^N A_j(x_j^k-\tilde{x}_j^k)\right\rangle \\
&\quad=\varphi(w^k,\tilde{w}^k,\xi^k),
\end{aligned}\tag{60}$$

where the second equality follows from the definitions of $\tilde{w}^k$, $\hat{w}^k$, and $d_1(w^k,\tilde{w}^k,\xi^k)$. By combining the last two equations and using (21) we obtain

$$\begin{aligned}
&\|w^k-\hat{w}^k\|_G^2-\|w^{k+1}-\hat{w}^k\|_G^2 \\
&\quad=2\gamma\alpha_k^*\varphi(w^k,\tilde{w}^k,\xi^k)-\gamma^2(\alpha_k^*)^2\|G^{-1}d_1(w^k,\tilde{w}^k,\xi^k)\|_G^2 \\
&\quad=\gamma(2-\gamma)\alpha_k^*\varphi(w^k,\tilde{w}^k,\xi^k).
\end{aligned}$$

Substituting the last equation into (59) yields

$$\begin{aligned}
&\langle w-\hat{w}^k, G(w^k-w^{k+1})\rangle+\frac{1}{2}(\|w-w^k\|_G^2-\|w-w^{k+1}\|_G^2) \\
&\quad=\frac{1}{2}\gamma(2-\gamma)\alpha_k^*\varphi(w^k,\tilde{w}^k,\xi^k)\ge 0,
\end{aligned}$$

which is just (58). Thus, the proof is complete. □

Lemma 6.2 *If the new iterate w^{k+1} is updated by* (20), *then we have*

$$\langle w-\hat{w}^k, \gamma\alpha_k^*F(\hat{w}^k)\rangle+\frac{1}{2}(\|w-w^k\|_G^2-\|w-w^{k+1}\|_G^2)\ge 0\quad \forall w\in\mathcal{W}.$$

Proof To begin with, we separate the term $(w-\hat{w}^k)^T\gamma\alpha_k^*F(\hat{w}^k)$ into two as

$$\langle w-\hat{w}^k, \gamma\alpha_k^*F(\hat{w}^k)\rangle=\langle w^{k+1}-\hat{w}^k, \gamma\alpha_k^*F(\hat{w}^k)\rangle+\langle w-w^{k+1}, \gamma\alpha_k^*F(\hat{w}^k)\rangle.\tag{61}$$

In the sequel, we will deal with the above two terms separately.

Since $w^{k+1}\in\mathcal{W}$, we have, by substituting w with w^{k+1} in (56),

$$\begin{aligned}
&\langle w^{k+1}-\hat{w}^k, \gamma\alpha_k^*F(\hat{w}^k)\rangle \\
&\quad\ge\langle w^{k+1}-\hat{w}^k, \gamma\alpha_k^*\hat{d}_1(w^k,\hat{w}^k,\xi^k)\rangle \\
&\quad=\langle w^{k+1}-\hat{w}^k, \gamma\alpha_k^*d_1(w^k,\tilde{w}^k,\xi^k)\rangle \\
&\quad=\langle w^k-\hat{w}^k, \gamma\alpha_k^*d_1(w^k,\tilde{w}^k,\xi^k)\rangle-\langle w^k-w^{k+1}, \gamma\alpha_k^*d_1(w^k,\tilde{w}^k,\xi^k)\rangle.
\end{aligned}\tag{62}$$

Recall that the first term on the right-hand side of (62) was calculated in (60). As to the second term, we have

$$\begin{aligned}&\langle w^k - w^{k+1}, \gamma\alpha_k^* d_1(w^k, \tilde{w}^k, \xi^k)\rangle \\ &\quad = \langle w^k - w^{k+1}, G(\gamma\alpha_k^* G^{-1} d_1(w^k, \tilde{w}^k, \xi^k))\rangle \\ &\quad \le \frac{1}{2}\|w^k - w^{k+1}\|_G^2 + \frac{1}{2}\gamma^2(\alpha_k^*)^2 \|G^{-1} d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2.\end{aligned}$$

Thus, we obtain

$$\begin{aligned}&\langle w^{k+1} - \hat{w}^k, \gamma\alpha_k^* F(\hat{w}^k)\rangle \\ &\quad \ge \gamma\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k) - \frac{1}{2}\gamma^2(\alpha_k^*)^2 \|G^{-1} d_1(w^k, \tilde{w}^k, \xi^k)\|_G^2 - \frac{1}{2}\|w^k - w^{k+1}\|_G^2 \\ &\quad = \frac{1}{2}\gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k) - \frac{1}{2}\|w^k - w^{k+1}\|_G^2,\end{aligned} \tag{63}$$

where the equality follows from (21).

Now, we turn to consider the second term $\langle w - w^{k+1}, \gamma\alpha_k^* F(\hat{w}^k)\rangle$ in (61). Since w^{k+1} is updated by (20), w^{k+1} is the projection of $[w^k - \gamma\alpha_k^* G^{-1} F(\hat{w}^k)]$ onto $\mathcal{W}$ under the G-norm. It follows from (7) that

$$\langle [w^k - \gamma\alpha_k^* G^{-1} F(\hat{w}^k)] - w^{k+1}, G(w - w^{k+1})\rangle \le 0 \quad \forall w \in \mathcal{W}.$$

As a consequence, we have

$$\langle w - w^{k+1}, \gamma\alpha_k^* F(\hat{w}^k)\rangle \ge \langle w - w^{k+1}, G(w^k - w^{k+1})\rangle.$$

By applying the formula $\langle a, Gb\rangle = \frac{1}{2}(\|a\|_G^2 - \|a-b\|_G^2 + \|b\|_G^2)$ to the right-hand side of the last inequality we derive that

$$\langle w - w^{k+1}, \gamma\alpha_k^* F(\hat{w}^k)\rangle \ge \frac{1}{2}(\|w - w^{k+1}\|_G^2 - \|w - w^k\|_G^2) + \frac{1}{2}\|w^k - w^{k+1}\|_G^2. \tag{64}$$

By incorporating inequalities (63) and (64) into Eq. (61), the assertion of Lemma 6.2 follows. □

Having the same key inequality for both methods, the $O(1/t)$ rate of convergence (in an ergodic sense) can be obtained easily.

Theorem 6.1 *For any integer $t > 0$, we have a $\hat{w}_t \in \mathcal{W}$ that satisfies*

$$\langle \hat{w}_t - w, F(w)\rangle \le \frac{1}{2\gamma\Upsilon_t}\|w - w^0\|_G^2 \quad \forall w \in \mathcal{W},$$

where

$$\hat{w}_t = \frac{1}{\Upsilon_t}\sum_{k=0}^{t} \alpha_k^* \hat{w}^k \quad \text{and} \quad \Upsilon_t = \sum_{k=0}^{t} \alpha_k^*.$$

Proof In Lemmas 6.1 and 6.2, we have proved the same key inequality for both methods, namely,

$$\big\langle w-\hat{w}^k, \gamma\alpha_k^* F\big(\hat{w}^k\big)\big\rangle + \frac{1}{2}\big(\|w-w^k\|_G^2 - \|w-w^{k+1}\|_G^2\big) \ge 0 \quad \forall w \in \mathcal{W}.$$

Since F is monotone, we have

$$\big\langle w-\hat{w}^k, \gamma\alpha_k^* F(w)\big\rangle + \frac{1}{2}\big(\|w-w^k\|_G^2 - \|w-w^{k+1}\|_G^2\big) \ge 0 \quad \forall w \in \mathcal{W},$$

or, equivalently,

$$\big\langle \hat{w}^k - w, \gamma\alpha_k^* F(w)\big\rangle + \frac{1}{2}\big(\|w-w^{k+1}\|_G^2 - \|w-w^k\|_G^2\big) \le 0 \quad \forall w \in \mathcal{W},$$

When taking the sum of the above inequalities over $k = 0, \dots, t$, we obtain

$$\left\langle \sum_{k=0}^{t} \alpha_k^* \hat{w}^k - \left(\sum_{k=0}^{t} \alpha_k^*\right) w, F(w) \right\rangle + \frac{1}{2\gamma}\big(\|w^{t+1} - w^0\|_G^2 - \|w-w^0\|_G^2\big) \le 0.$$

By dropping the term $\|w^{t+1} - w^0\|_G^2$ and incorporating Υ_t and $\hat{w}_t$ into the above inequality, we have

$$\big\langle \hat{w}_t - w, F(w)\big\rangle \le \frac{\|w-w^0\|_G^2}{2\gamma\Upsilon_t} \quad \forall w \in \mathcal{W}.$$

Hence, the proof is complete. □

Since it follows from (50) that

$$\Upsilon_t = \sum_{k=0}^{t} \alpha_k^* \ge (t+1)c_0,$$

we have, by Theorem 6.1,

$$\big\langle \hat{w}_t - w, F(w)\big\rangle \le \frac{1}{2\gamma\Upsilon_t}\|w-w^0\|_G^2 \le \frac{\|w-w^0\|_G^2}{2\gamma c_0(t+1)} \quad \forall w \in \mathcal{W}.$$

According to (53), the above inequality implies the $O(1/t)$-rate of convergence immediately. We emphasize that our convergence rate is in the ergodic sense. From a theoretical point of view, this suggests to use a larger parameter $\gamma \in (0, 2)$ in implementations.

7 Conclusions

Attracted by the practical efficiency of the alternating direction method of multipliers, an alternating-direction-based contraction method was proposed in [16]. The new method deals with the general separable and linearly constrained convex optimization problem, where the objective function is separable into finitely many parts. However, the new method requires the exact solution of ADMM subproblems, which limits its applicability. To overcome this limitation, this paper presents two inexact alternating-direction-based contraction methods. These methods are practically more viable since the subproblems are solved inexactly. The convergence properties and complexity results ($O(1/t)$ rate of convergence) of the proposed methods were derived. We emphasize that even for the simplest case, where the objective function is separable into two parts, our methods are different from the common inexact methods in the literature [19, 20], as our subproblems for computing search directions do not include proximal terms of any form. In addition, the complexity results, which are proved in the framework of variational inequalities, are new for this kind of inexact ADMs.

Acknowledgements Authors wish to thank Prof. Panos M. Pardalos (the associate editor) and Prof. Cornelis Roos (Delft University of Technology) for useful comments and suggestions.

G. Gu was supported by the NSFC grant 11001124.

B. He was supported by the NSFC grant 91130007.

J. Yang was supported by the NSFC grant 11371192.

References

1. Floudas, C.A., Pardalos, P.M. (eds.): Encyclopedia of Optimization, 2nd edn. Springer, Berlin (2009)
2. Pardalos, P.M., Resende, M.G.C. (eds.): Handbook of Applied Optimization. Oxford University Press, Oxford (2002)
3. Ng, M.K., Weiss, P., Yuan, X.: Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods. SIAM J. Sci. Comput. **32**(5), 2710–2736 (2010)
4. Yang, J., Zhang, Y.: Alternating direction algorithms for ℓ_1-problems in compressive sensing. SIAM J. Sci. Comput. **33**(1), 250–278 (2011)
5. Wen, Z., Goldfarb, D., Yin, W.: Alternating direction augmented Lagrangian methods for semidefinite programming. Math. Program. Comput. **2**(3–4), 203–230 (2010)
6. He, B., Xu, M., Yuan, X.: Solving large-scale least squares semidefinite programming by alternating direction methods. SIAM J. Matrix Anal. Appl. **32**(1), 136–152 (2011)
7. Yuan, X.: Alternating direction method for covariance selection models. J. Sci. Comput. **51**(2), 261–273 (2012)
8. Yuan, X., Yang, J.: Sparse and low-rank matrix decomposition via alternating direction method. Pac. J. Optim. **9**(1), 167–180 (2013)
9. Bertsekas, D.P.: Constrained Optimization and Lagrange Multiplier Methods. Computer Science and Applied Mathematics. Academic Press/Harcourt Brace Jovanovich Publishers, New York (1982)
10. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer Series in Operations Research and Financial Engineering. Springer, New York (2006)
11. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
12. Chan, R.H., Yang, J., Yuan, X.: Alternating direction method for image inpainting in wavelet domains. SIAM J. Imaging Sci. **4**(3), 807–826 (2011)
13. Goldstein, T., Osher, S.: The split Bregman method for $L1$-regularized problems. SIAM J. Imaging Sci. **2**(2), 323–343 (2009)
14. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comput. Math. Appl. **2**(1), 17–40 (1976)
15. Douglas, J. Jr., Rachford, H.H. Jr.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. **82**, 421–439 (1956)
16. He, B., Tao, M., Xu, M., Yuan, X.: An alternating direction-based contraction method for linearly constrained separable convex programming problems. Optimization **62**(4), 573–596 (2013)

17. Blum, E., Oettli, W.: Grundlagen und Verfahren. Springer, Berlin (1975). Grundlagen und Verfahren, Mit einem Anhang "Bibliographie zur Nichtlinearer Programmierung", Ökonometrie und Unternehmensforschung, No. XX
18. Bertsekas, D.P., Tsitsiklis, J.N.: Parallel and Distributed Computation: Numerical Methods. Prentice-Hall, Upper Saddle River (1989)
19. He, B., Liao, L., Qian, M.: Alternating projection based prediction-correction methods for structured variational inequalities. J. Comput. Math. **24**(6), 693–710 (2006)
20. He, B., Liao, L., Han, D., Yang, H.: A new inexact alternating directions method for monotone variational inequalities. Math. Program., Ser. A **92**(1), 103–118 (2002)
21. Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. Math. Program., Ser. A **64**(1), 81–101 (1994)
22. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**(5), 877–898 (1976)
23. Teboulle, M.: Convergence of proximal-like algorithms. SIAM J. Optim. **7**(4), 1069–1083 (1997)
24. Tseng, P.: Alternating projection-proximal methods for convex programming and variational inequalities. SIAM J. Optim. **7**(4), 951–965 (1997)
25. He, B., Xu, M.: A general framework of contraction methods for monotone variational inequalities. Pac. J. Optim. **4**(2), 195–212 (2008)
26. Nemirovski, A.: Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. Optim. **15**(1), 229–251 (2004) (Electronic)
27. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I. Springer Series in Operations Research. Springer, New York (2003)
28. Cai, X., Gu, G., He, B.: On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators. Comput. Optim. Appl. (2013). doi:10.1007/s10589-013-9599-7