Proximal alternating direction-based contraction methods for separable linearly constrained convex optimization

Bingsheng He* Zheng Peng and Xiang-feng Wang

Department of Mathematics, Nanjing University, Nanjing, 210093, China.

Abstract

Alternating direction method (ADM) has been well studied in the context of linearly constrained convex programming problems. Recently, because of its significant efficiency and easy implementation in novel applications, ADM is extended to the case where the number of separable parts is a finite number. The algorithmic framework of the extended method consists of two phases. On each iteration it first produces a trial point by using the usual alternating direction scheme. Then the next iterate is updated by using a distance-descent direction offered by the trial point. The generated sequence approaches the solution set monotonically in the Fejér sense, and the method is called alternating direction-based contraction (ADBC) method. In this paper, in order to simplify the subproblems in the first phase, we add a proximal term to the objective function of the minimization subproblems. The resultant algorithm is called proximal alternating direction-based contraction (PADBC) methods. In addition, we present different linearized versions of the PADBC methods which substantially broaden the applicable scope of the ADBC method. All the presented algorithms are guided by a general framework of the contraction methods for monotone variational inequalities and thus the convergence follows straightforwardly.

Keywords: alternating direction method, linearly constrained convex programming, separable structure, contraction method

1 Introduction

The linearly constrained convex programming problem in the sense that both the objective function and the constraints are separable into finitely many of parts has the following mathematical form:

$$\min\{\sum_{i=1}^{m}\theta_i(x_i) \mid \sum_{i=1}^{m}A_ix_i = b, \;\; x_i \in \mathcal{X}_i, \; i = 1, 2, \ldots, m\}, \tag{1.1}$$

where $\theta_i : \Re^{n_i} \to \Re$ $(i = 1, 2, \cdots, m)$ are closed proper convex functions (not necessarily smooth); $A_i \in \Re^{l\times n_i}$ $(i = 1, 2, \cdots, m)$; $\mathcal{X}_i \subseteq \Re^{n_i}$ $(i = 1, 2, \cdots, m)$ are closed convex sets; and $\sum_{i=1}^m n_i = n$. We shall denote $x = (x_1, \ldots, x_m)$ and $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m$, where x_i (resp. $\mathcal{X}_i$) is the i-th part of x (resp. $\mathcal{X}$). It is assumed that the solution set of (1.1), denoted by $\mathcal{X}^*$, is not empty. Let $\lambda \in \Re^l$ denote the Lagrangian multiplier for the linear constraints $\sum_{i=1}^m A_ix_i = b$ and Λ^* be

*This author was supported by the NSFC Grant 10971095, the NSF of Jiangsu Province Grant BK2008255 and and the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (No.708044). Email: hebma@nju.edu.cn

the set of the related optimal multipliers. For convenience, we denote the augmented Lagrangian function of (1.1) by

$$\mathcal{L}_A(x_1,\ldots,x_m,\lambda) := \textstyle\sum_{i=1}^m \theta_i(x_i) - \lambda^T \left(\sum_{i=1}^m A_i x_i - b\right) + \frac{1}{2}\left\|\sum_{i=1}^m A_i x_i - b\right\|_H^2, \tag{1.2}$$

where $H \in \Re^{l\times l}$ is a positive definite matrix. In the following

$$\mathcal{W} = \mathcal{X} \times \Re^l \qquad \text{and} \qquad \mathcal{W}^* = \mathcal{X}^* \times \Lambda^*.$$

For given $w^k = (x_1^k, x_2^k, \ldots, x_m^k, \lambda^k)$, let $\tilde{w}^k = (\tilde{x}_1^k, \tilde{x}_2^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ be generated by the following alternating direction (AD) scheme:

$$\begin{cases} \tilde{x}_i^k = \arg\min\{\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) \mid x_i \in \mathcal{X}_i\}, \;\; i = 1, \ldots, m, \\ \tilde{\lambda}^k = \lambda^k - H(\sum_{i=1}^m A_i \tilde{x}_i^k - b). \end{cases} \tag{1.3}$$

In the i-th sub-problem of (1.3), since $(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i^k, x_{i+1}^k, \ldots, x_m^k, \lambda^k)$ are available, the variable in

$$\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k)$$

is x_i (see (1.2)) and the to be minimized nonlinear functions are

$$\theta_i(x_i) \qquad \text{and} \qquad \frac{1}{2}\|\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b\|_H^2.$$

Therefore, the scheme (1.3) can be viewed as a practical and structured-exploited variant (split form or relaxed form) of Gauss-Seidel form, with the adaptation of minimizing the involved separable variables $\tilde{x}_i$ separably in an alternating order.

It is well known that for $m \le 2$, one can directly take $\tilde{w}^k$ by (1.3) as the next iterate, and the resultant methods become some classical methods for constrained convex optimization.

- For $m = 1$, taking $w^{k+1} = \tilde{w}^k$, the alternating directions scheme (1.3) is the recursion of Augmented Lagrangian Method (ALM) for problem

  $$\min\{\theta_1(x_1) \mid A_1 x_1 = b, \; x_1 \in \mathcal{X}_1\}. \tag{1.4}$$

  The sequence $\{\lambda^k\}$ generated by ALM satisfies

  $$\|\lambda^{k+1} - \lambda^*\|_{H^{-1}}^2 \le \|\lambda^k - \lambda^*\|_{H^{-1}}^2 - \|\lambda^k - \lambda^{k+1}\|_{H^{-1}}^2, \; \forall\, \lambda^* \in \Lambda^*. \tag{1.5}$$

  We refer to, e.g. [1, 25], for the intensive study of ALM for minimization problem with nonlinear constraints.

- For $m = 2$, taking $w^{k+1} = \tilde{w}^k$, the alternating directions scheme is the recursion of Alternating Directions Method (ADM) for problem

  $$\min\{\theta_1(x_1) + \theta_2(x_2) \mid A_1 x_1 + A_2 x_2 = b, \; x_1 \in \mathcal{X}_1, \; x_2 \in \mathcal{X}_2\}. \tag{1.6}$$

  The sequence $\{(x_2^k, \lambda^k)\}$ generated by ADM satisfies (see [20])

  $$\begin{aligned} \big(\|A_2(x_2^{k+1} - x_2^*)\|_H^2 + \|\lambda^{k+1} - \lambda^*\|_{H^{-1}}^2\big) &\le \big(\|A_2(x_2^k - x_2^*)\|_H^2 + \|\lambda^k - \lambda^*\|_{H^{-1}}^2\big) \\ &\quad - \big(\|A_2(x_2^k - x_2^{k+1})\|_H^2 + \|\lambda^k - \lambda^{k+1}\|_{H^{-1}}^2\big), \quad \forall\,(x^*, \lambda^*) \in \mathcal{W}^*. \end{aligned} \tag{1.7}$$

Alternating directions method dates back to [11]. It is perhaps one of the most popular methods for solving (1.6). Because of its significant efficiency and easy implementation, ADM has attracted wide attention of many authors in various areas, see e.g. [2, 4, 6, 7, 10, 12, 13, 15, 20, 22, 31]. In particular, some novel and attractive applications of ADM have been discovered very recently, e.g. the total-variation problem in Image Processing [8, 24, 30], the covariance selection problem and semidefinite least square problem in Statistics [19, 32], the semidefinite programming problems [27, 29], the sparse and low-rank recovery problem in Engineering [33]. It is clear that some special minimization problems such as

$$\min\{\theta_0(z) \mid z \in \mathcal{Z}_1 \cap \mathcal{Z}_2 \cap \ldots \cap \mathcal{Z}_m\}$$

and

$$\min\{\theta_1(z) + \theta_2(z) + \cdots + \theta_m(z) \mid z \in \mathcal{Z}\}$$

can be converted to the form of the problem (1.1) in which A_j is the j-th column of the following matrix

$$A = \begin{pmatrix} I & -I & 0 & \cdots & 0 \\ 0 & I & -I & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I & -I \\ -I & 0 & \cdots & 0 & I \end{pmatrix}.$$

However, for $m > 2$, there is no convergence statement if one directly takes $\tilde{w}^k$ by (1.3) as the next iterate. Thus, we say, $\tilde{w}^k$ is a trial point. In practice, $w^k = (x_1^k, x_2^k, \ldots, x_m^k, \lambda^k) \in \mathcal{W}^*$ if and only if $w^k = \tilde{w}^k$, and the distance $\|w^k - \tilde{w}^k\|$ measures how much w^k fails to be in $\mathcal{W}^*$. Fortunately, in the case that $w^k \notin \mathcal{W}^*$, the trial point $\tilde{w}^k$ offers us a descent direction to the solution set. Along this direction, with a determinate step length size, the new iterate w^{k+1} is updated and closer to $\mathcal{W}^*$. As demonstrated in [17], the resultant method falls into the frameworks of both the extended AD scheme (1.3), and the contraction-type methods (according to the definition in [3]) in the sense that the sequence of iterates is Fejér monotone with respect to $\mathcal{W}^*$. Therefore, the method is called Alternating Direction-Based Contraction (ADBC) method [17].

In order to simplify the subproblems in (1.3), in this paper, we present some similar methods to the ADBC method but use different schemes to produce the trial point $\tilde{w}^k$. In detail, instead of solving the sub-problems

$$\min\{\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) \mid x_i \in \mathcal{X}_i\}, \quad i = 1, \ldots, m, \tag{1.8}$$

in the scheme (1.3), we solve the subproblems by adding a quadratic proximal term $\frac{r_i}{2}\|x_i - x_i^k\|^2$ ($r_i \geq 0$ is a constant) to the objective $\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k)$ and linearizing its nonlinear term

$$\theta_i(x_i) \qquad \text{or/and} \qquad \frac{1}{2}\|\sum_{j=1}^{i-1} A_j \tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b\|_H^2.$$

The resulting scheme for producing the trial point $\tilde{w}^k$ is called the Linearized versions of the Proximal Alternating Directions Scheme. Again, in the case that $w^k \notin \mathcal{W}^*$, the trial point $\tilde{w}^k$ offers us the search directions which satisfy the conditions in the general framework [18] and the new iterate is closer to $\mathcal{W}^*$. For this reason, the proposed methods are named Proximal Alternating Directions Based Contraction (PADBC) methods.

In the following, H is a given $l \times l$ positive definite matrix and $\nu > 0$ is a constant. In the case that we need linearize $\theta_i(x_i)$ in $\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k)$, we assume that each

$\theta_i(x_i), i=1,\ldots,m$ is differentiable and $f_i(x_i)=\nabla\theta_i(x_i)$ is Lipschitz continuous, *i.e.*, there is $l_i>0$, such that

$$\|f_i(x_i)-f_i(x_i')\|\le l_i\|x_i-x_i'\|,\quad \forall\, x_i,x_i'\in\Re^{n_i},\quad i=1,\ldots,m. \tag{1.9}$$

The rest of this paper is organized as follows: In Section 2, as a preparation for the rest of the analysis, we review some basic properties of the projection mapping and variational inequalities. A useful general framework (including search directions and error measure function) for contraction methods of variational inequalities is quoted in this section. Section 3 presents the proximal alternating direction scheme for producing the trial vector, and offers the distance-decent directions and error measure function which satisfy the conditions of the general framework. In section 4, using the rationale of the general framework of contraction methods, we give the updating forms and prove the convergence of the resulting methods. From Section 5 to Section 7, we present different linearized versions of the proximal alternating direction scheme and their related requirements on the proximal coefficients. The parallel analysis validates the conditions in the general framework and thus the resultant contraction methods are convergent. Finally, some conclusions are drawn in Section 8.

2 Preliminaries

For convergence analysis, it is useful to characterize the first order optimal condition of the constrained optimization problem by a variational inequality. In this section, we summarize some basic properties and the related definitions which will be used in the analysis and discussions.

2.1 VI Characterization of the first order optimal condition

Let $f_i(x_i)\in\partial(\theta_i(x_i))$ for $i=1,2,\cdots,m$. It is clear that the first order optimal condition of the generally separable linearly constrained convex programming problem (1.1) is characterized by the following variational inequality: Find $w^*=(x_1^*,...,x_m^*,\lambda^*)\in\mathcal{W}:=\mathcal{X}_1\times\mathcal{X}_2\times\cdots\times\mathcal{X}_m\times\Re^l$ such that

$$\left\{\begin{array}{l}(x_1'-x_1^*)^T\{f_1(x_1^*)-A_1^T\lambda^*\}\ge 0,\\ \qquad\vdots\\ (x_m'-x_m^*)^T\{f_m(x_m^*)-A_m^T\lambda^*\}\ge 0,\\ (\lambda'-\lambda^*)^T(\sum_{i=1}^m A_ix_i^*-b)\ge 0,\end{array}\right.\qquad \forall\, w'=(x_1',x_2',\cdots,x_m',\lambda')\in\mathcal{W}. \tag{2.1}$$

In the more compact form:

$$(w'-w^*)^TF(w^*)\ge 0,\quad \forall\, w'\in\mathcal{W}, \tag{2.2}$$

where

$$w=\begin{pmatrix}x_1\\x_2\\\vdots\\x_m\\\lambda\end{pmatrix}\qquad\text{and}\qquad F(w)=\begin{pmatrix}f_1(x_1)-A_1^T\lambda\\f_2(x_2)-A_2^T\lambda\\\vdots\\f_m(x_m)-A_m^T\lambda\\\sum_{i=1}^m A_ix_i-b\end{pmatrix}. \tag{2.3}$$

We denote problem (2.2)-(2.3) by VI$(\mathcal{W},F)$. Note that $F(w)$ defined in (2.3) is monotone whenever each f_i is monotone. Recall that the solution set of VI$(\mathcal{W},F)$ is denoted by $\mathcal{W}^*$. The matrix

M is defined by

$$M = \begin{pmatrix} A_1^T H A_1 & 0 & \cdots & \cdots & 0 \\ A_2^T H A_1 & A_2^T H A_2 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ A_m^T H A_1 & A_m^T H A_2 & \cdots & A_m^T H A_m & 0 \\ 0 & 0 & \cdots & 0 & H^{-1} \end{pmatrix} + \begin{pmatrix} r_1 I_{n_1} & 0 & \cdots & \cdots & 0 \\ 0 & r_2 I_{n_2} & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & r_m I_{n_m} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}. \tag{2.4}$$

Note that M ($r_i \geq 0$) is positive semi-definite (but not symmetric).

2.2 Variational inequality and projection mapping

This subsection summarizes preliminaries of the n-dimensional variational inequalities. Let Ω be a nonempty subset of $\Re^n$, F be a continuous mapping from $\Re^n$ to itself. The variational inequality problem, denoted by $\mathrm{VI}(\Omega, F)$, is to find a vector $w^* \in \Omega$ such that

$$\mathrm{VI}(\Omega, F) \qquad (w - w^*)^T F(w^*) \geq 0, \quad \forall\, w \in \Omega.$$

Let G be a $n \times n$ positive definite matrix, we denote $\|w\|_G = \sqrt{w^T G w}$ as the G-norm of vector $w \in \Re^n$. The projection under G-norm will be denoted by $P_{\Omega,G}(\cdot)$. In other words, for given v,

$$P_{\Omega,G}(v) = \operatorname{argmin}\{\|v - w\|_G \mid w \in \Omega\}.$$

From the above definition, it follows that

$$(v - P_{\Omega,G}(v))^T G (w - P_{\Omega,G}(v)) \leq 0, \quad \forall\, v \in \Re^n, \forall w \in \Omega. \tag{2.5}$$

Consequently, we have

$$\|P_{\Omega,G}(v) - P_{\Omega,G}(w)\|_G \leq \|v - w\|_G, \qquad \forall\, v, w \in \Re^n, \tag{2.6}$$

and

$$\|P_{\Omega,G}(v) - w\|_G^2 \leq \|v - w\|_G^2 - \|v - P_{\Omega,G}(v)\|_G^2, \qquad \forall v \in \Re^n, \forall\, w \in \Omega. \tag{2.7}$$

Definition 2.1. a). F is said to be monotone with respect to Ω if

$$(v - w)^T (F(v) - F(w)) \geq 0, \quad \forall\, v, w \in \Omega.$$

b). F is strongly monotone with respect to Ω if there exists a constant $\kappa > 0$ such that

$$(v - w)^T (F(v) - F(w)) \geq \kappa \|v - w\|^2, \quad \forall\, v, w \in \Omega.$$

We say $VI(\Omega, F)$ is monotone if the mapping F is monotone.

Lemma 2.2. *Let $G \in \Re^{n\times n}$ be any positive definite matrix. Then w^* is a solution of $VI(\Omega, F)$ if and only if*

$$w^* = P_{\Omega,G}[w^* - \alpha_k G^{-1} F(w^*)], \qquad \forall\, \alpha > 0. \tag{2.8}$$

Proof. See ([2], pp. 267). □

According to Lemma 2.2, for any positive definite matrix $G \in \Re^{n\times n}$, $p \in \Re^n$ and $\alpha > 0$,

$$w^* = P_{\Omega,G}[w^* - \alpha G^{-1} p] \qquad \Leftrightarrow \qquad w^* \in \Omega, \quad (w - w^*)^T p \geq 0, \quad \forall\, w \in \Omega. \tag{2.9}$$

The solution set of a monotone variational inequality is convex (see Theorem 2.3.5 in [9]). For monotone $\mathrm{VI}(\Omega, F)$, we have

$$(\tilde{w} - w^*)^T F(\tilde{w}) \geq (\tilde{w} - w^*)^T F(w^*), \quad \forall\, \tilde{w} \in \Omega,\ w^* \in \Omega^*.$$

Consequently, because $(\tilde{w} - w^*)^T F(w^*) \geq 0$, we obtain

$$(\tilde{w} - w^*)^T F(\tilde{w}) \geq 0, \quad \forall\, \tilde{w} \in \Omega. \tag{2.10}$$

2.3 A general framework of some contraction methods

We use the rationale of the general framework of some contraction methods in [18] for the finite-dimensional monotone variational inequalities to prove the convergence of the proposed methods. For given w, a vector $\tilde{w}$ generated by certain well-defined scheme is called a trial point. A trial point $\tilde{w} \in \Omega$ is said to be a test vector of w if and only if

$$w = \tilde{w} \quad \Rightarrow \quad w \in \Omega^*,$$

where Ω^* is the solution set of $VI(\Omega, F)$.

The general framework. For given w, let $\tilde{w} \in \Omega$ be a test vector of w. For this pair of w and $\tilde{w}$, we find $d_1(w,\tilde{w}),\ d_2(w,\tilde{w}) \in \Re^n$ and $\varphi(w,\tilde{w}) \in \Re$ which satisfy following conditions:

1. It holds that

$$\tilde{w} \in \Omega, \quad (w' - \tilde{w})^T \big(d_2(w,\tilde{w}) - d_1(w,\tilde{w})\big) \geq 0, \quad \forall\, w' \in \Omega. \tag{2.11a}$$

2. There is a constant $K > 0$ such that

$$\|d_1(w,\tilde{w})\| \leq K\|w - \tilde{w}\|. \tag{2.11b}$$

3. For any $w^* \in \Omega^*$,

$$(\tilde{w} - w^*)^T d_2(w,\tilde{w}) \geq \varphi(w,\tilde{w}) - (w - \tilde{w})^T d_1(w,\tilde{w}). \tag{2.11c}$$

4. $\varphi(w,\tilde{w})$ is an *error measure function* of $\mathrm{VI}(\Omega, F)$, i.e., there is a constant $\tau > 0$, such that

$$\varphi(w,\tilde{w}) \geq \tau\|w - \tilde{w}\|^2 \qquad \& \qquad \varphi(w,\tilde{w}) = 0 \ \Leftrightarrow\ w = \tilde{w}. \tag{2.11d}$$

In [18], (2.11a) was stated in its equivalent form (see Lemma 2.2)

$$\tilde{w} = P_\Omega\{\tilde{w} - [d_2(w,\tilde{w}) - d_1(w,\tilde{w})]\}. \tag{2.12}$$

For any $w^* \in \Omega^*$, $(w - w^*)$ is the gradient of the unknown distance function $\frac{1}{2}\|w - w^*\|^2$ at the point w. A direction d is called a distance-descent direction if and only if the inner-product $\big\langle w - w^*, d\big\rangle < 0$. It was proved (see Lemmas 3.2 and 3.3 in [18]) that the directions $-d_1(w,\tilde{w})$ and $-d_2(w,\tilde{w})$ in the general framework are descent directions of the unknown distance function $\|w - w^*\|^2$ when w is not a solution point. Only for completeness, we quote the short proofs.

Lemma 2.3. *For $w \in \Re^n$, if $d_1(w,\tilde{w}),\ d_2(w,\tilde{w})$ and $\varphi(w,\tilde{w})$ satisfy the conditions* (2.11a) *and* (2.11c) *in the general framework, then we have*

$$(w - w^*)^T d_1(w,\tilde{w}) \geq \varphi(w,\tilde{w}), \quad \forall w^* \in \Omega^*. \tag{2.13}$$

Proof. Because $w^* \in \Omega$, (2.11a) implies

$$(\tilde{w} - w^*)^T d_1(w,\tilde{w}) \geq (\tilde{w} - w^*)^T d_2(w,\tilde{w}). \tag{2.14}$$

Combining the (2.14) with (2.11c), we obtain

$$(\tilde{w} - w^*)^T d_1(w,\tilde{w}) \geq \varphi(w,\tilde{w}) - (w - \tilde{w})^T d_1(w,\tilde{w})$$

and consequently (2.13). The lemma is proved. $\square$

Lemma 2.4. *For $w \in \Omega$, if $d_1(w, \tilde{w})$, $d_2(w, \tilde{w})$ and $\varphi(w, \tilde{w})$ satisfy the conditions* (2.11a) *and* (2.11c) *in the general framework, then we have*

$$(w - w^*)^T d_2(w, \tilde{w}) \geq \varphi(w, \tilde{w}), \quad \forall\, w^* \in \Omega^*. \tag{2.15}$$

Proof. Since $w \in \Omega$, set $w' = u$ in (2.12) we get

$$(w - \tilde{w})^T d_2(w, \tilde{w}) \geq (w - \tilde{w})^T d_1(w, \tilde{w}). \tag{2.16}$$

Add (2.16) and (2.11c) we obtain (2.15) and the lemma is proved. □

Note that condition (2.11b) in the general framework means that $\|d_1(w, \tilde{w})\| \to 0$ as $\|w - \tilde{w}\| \to 0$. However, the framework does not claim the same requirement for $\|d_2(w, \tilde{w})\|$. Along $-d_1(w, \tilde{w})$ (resp. $-d_2(w, \tilde{w})$) we can reach a point which is closer to the solution set and thus we name it the framework of contraction methods. Due to the properties of (2.11d), $\varphi(w, \tilde{w})$ can be viewed as an error measure function; it measures how much w fails to be in Ω^*.

2.4 Some contraction methods based on the general framework

The general framework offers us a powerful tool for constructing contraction methods. To see this importance, we give the following simple examples:

The proximal point algorithms. Applying the proximal point algorithm [23, 26] to VI(Ω, F), for given $w^k \in \Omega$ and $\beta > 0$, the sub-problem in k-th iteration is

$$\text{(PPA)} \qquad w \in \Omega, \qquad (w' - w)^T F_k(w) \geq 0, \quad \forall\, w' \in \Omega, \tag{2.17a}$$

where

$$F_k(w) = \beta F(w) + (w - w^k). \tag{2.17b}$$

Let $\tilde{w}^k$ be the solution of (2.17), then we have

$$\tilde{w}^k \in \Omega, \quad (w' - \tilde{w}^k)^T \big(d_2(w^k, \tilde{w}^k) - d_1(w^k, \tilde{w}^k)\big) \geq 0, \quad \forall w' \in \Omega, \tag{2.18a}$$

where

$$d_1(w^k, \tilde{w}^k) = w^k - \tilde{w}^k \qquad \text{and} \qquad d_2(w^k, \tilde{w}^k) = \beta F(\tilde{w}^k). \tag{2.18b}$$

It follows from (2.18) that the conditions (2.11a) and (2.11b) are satisfied. Since $\tilde{w}^k \in \Omega$ and $w^* \in \Omega^*$, it follows from (2.10) that

$$(\tilde{w} - w^*)^T d_2(w, \tilde{w}) = \beta(\tilde{w} - w^*)^T F(\tilde{w}) \geq 0, \quad \forall\, w^* \in \Omega^*.$$

By setting

$$\varphi(w^k, \tilde{w}^k) = \|w^k - \tilde{w}^k\|^2, \tag{2.19}$$

it follows that

$$(\tilde{w} - w^*)^T d_2(w, \tilde{w}) \geq \varphi(w^k, \tilde{w}^k) - (w^k - \tilde{w}^k)^T d_1(w^k, \tilde{w}^k), \tag{2.20}$$

and thus conditions (2.11c) and (2.11d) are fulfilled. In the classical PPA , the solution of (2.17), $\tilde{w}^k$, is taken as the next iterate w^{k+1}. Since $d_1(w^k, \tilde{w}^k)$, $d_2(w^k, \tilde{w}^k)$ and $\varphi(w^k, \tilde{w}^k)$ satisfy the general framework, we can use the updating form

$$w^{k+1} = w^k - \alpha d_1(w^k, \tilde{w}^k), \quad \alpha \in (0, 2), \tag{2.21}$$

to produce the next iterate. It follows from (2.13) that

$$\|w^{k+1} - w^*\|^2 \leq \|w^k - w^*\|^2 - \alpha(2 - \alpha)\|w^k - \tilde{w}^k\|^2, \quad \forall\, w^* \in \Omega^*, \tag{2.22}$$

and the sequence $\{w^k\}$ is Fejér monotone with respect to Ω^*.

Projection and contraction method. To solve VI(Ω, F), for given $w^k \in \Omega$ and $\beta > 0$, let

$$\tilde{w}^k = P_\Omega[w^k - \beta F(w^k)], \tag{2.23}$$

where $\beta > 0$ is selected (under the condition that F is Lipschitz continuous) to satisfy

$$\beta\|F(w) - F(\tilde{w})\| \le \mu\|w - \tilde{w}\|, \quad \mu \in (0,1). \tag{2.24}$$

By setting

$$d_1(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k) - \beta(F(w^k) - F(\tilde{w}^k)) \tag{2.25}$$

and

$$d_2(w^k, \tilde{w}^k) = \beta F(\tilde{w}^k), \tag{2.26}$$

equation (2.23) can be written as

$$\tilde{w}^k = P_\Omega\{\tilde{w}^k - [d_2(w^k, \tilde{w}^k) - d_1(w^k, \tilde{w}^k)]\},$$

which is an equivalent expression of the condition (2.11a) (see (2.12)). It follows from (2.24) and (2.25) that

$$\|d_1(w^k, \tilde{w}^k)\| \le (1+\mu)\|w^k - \tilde{w}^k\|$$

and thus (2.11b) is satisfied. Again, since $(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) = \beta(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \ge 0$ (see (2.10)), by setting

$$\varphi(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k)^T d_1(w^k, \tilde{w}^k), \tag{2.27}$$

we get condition (2.11c). Using the Cauchy-Schwarz Inequality, it follows from (2.24) that

$$\begin{aligned} \varphi(w^k, \tilde{w}^k) &= (w^k - \tilde{w}^k)^T d_1(w^k, \tilde{w}^k) \\ &= \|w^k - \tilde{w}^k\|^2 - (w^k - \tilde{w}^k)^T \beta\big(F(w^k) - F(\tilde{w}^k)\big) \\ &\ge (1-\mu)\|w^k - \tilde{w}^k\|^2, \end{aligned} \tag{2.28}$$

and thus condition (2.11d) is satisfied. The directions $d_1(w^k, \tilde{w}^k)$, $d_2(w^k, \tilde{w}^k)$ and the error measure function $\varphi(w^k, \tilde{w}^k)$ satisfy the general framework. If we use the updating form

$$w^{k+1} = w^k - d_1(w^k, \tilde{w}^k), \tag{2.29}$$

it follows from (2.13) and (2.27) that

$$\begin{aligned} \|w^{k+1} - w^*\|^2 &= \|w^k - w^*\|^2 - 2(w^k - w^*)^T d_1(w^k, \tilde{w}^k) + \|d_1(w^k, \tilde{w}^k)\|^2 \\ &\le \|w^k - w^*\|^2 - 2\varphi(w^k, \tilde{w}^k) + \|d_1(w^k, \tilde{w}^k)\|^2 \\ &= \|w^k - w^*\|^2 - d_1(w^k, \tilde{w}^k)^T\big(2(w^k - \tilde{w}^k) - d_1(w^k, \tilde{w}^k)\big) \\ &= \|w^k - w^*\|^2 - \big(\|w^k - \tilde{w}^k\|^2 - \beta^2\|F(w^k) - F(\tilde{w}^k)\|^2\big) \\ &\le \|w^k - w^*\|^2 - (1-\mu^2)\|w^k - \tilde{w}^k\|^2. \end{aligned} \tag{2.30}$$

The last inequality of (2.30) follows from (2.24) and the sequence $\{w^k\}$ is Fejér monotone with respect to Ω^*. The directions $d_1(w^k, \tilde{w}^k)$ (resp. $d_2(w^k, \tilde{w}^k)$) defined in (2.25) (resp. (2.26)) were used in [14] (resp. in [21]) and the resulting methods are called projection and contraction methods.

In the above mentioned examples, the update form for the new iterate is

$$w^{k+1} = w^k - \alpha d_1(w^k, \tilde{w}^k)$$

with $\alpha \in (0,2)$ (in (2.21)) or $\alpha = 1$ (in (2.29)). Indeed, by using the framework, we get (2.22) and (2.30). These inequalities are principal for the convergence. Therefore, in the following sections, for the trial point $\tilde{w}^k$ produced by the proximal alternating direction scheme and its different linearized versions, we concentrate our effort to find $d_1(w^k, \tilde{w}^k)$, $d_2(w^k, \tilde{w}^k)$ and $\varphi(w^k, \tilde{w}^k)$ which satisfy the general framework.

3 Proximal alternating direction scheme

In this section, we use the proximal alternating direction scheme in the first phase to produce the trial point. In other words, we only add a quadratic proximal term $\frac{r_i}{2}\|x_i - x_i^k\|^2$ to the objective $\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k)$. We abuse the notations that have been used in Section 2 without ambiguity. The subscript such as in x_i of $x = (x_1, \ldots, x_m)$ denotes the i-th part of the vector x. A superscript such as in w^k refers to a specific vector and usually denotes an iteration index.

The proximal alternating direction (AD) scheme:
For $i = 1, \ldots, m$, with available $\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i^k, x_{i+1}^k, \ldots, x_m^k$ and λ^k, solve the minimization problem

$$\min\{\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k) + \frac{r_i}{2}\|x_i - x_i^k\|^2 \mid x_i \in \mathcal{X}_i\}, \tag{3.1a}$$

and denote its solution by $\tilde{x}_i^k$. Then set

$$\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^m A_j\tilde{x}_j^k - b). \tag{3.1b}$$

Note that the variable in the i-th sub-problem of (3.1a) is x_i. Since $f_i(x_i) \in \partial\big(\theta_i(x_i)\big)$, the first order optimal condition of the i-th sub-problem (see $\mathcal{L}_A(x_1, \ldots, x_m, \lambda)$ in (1.2)) can be characterized as the following variational inequality: $\tilde{x}_i^k \in \mathcal{X}_i$,

$$(x_i' - \tilde{x}_i^k)^T\big\{f_i(\tilde{x}_i^k) - A_i^T\lambda^k + A_i^T H\big(\sum_{j=1}^i A_j\tilde{x}_j^k + \sum_{j=i+1}^m A_j x_j^k - b\big) + r_i(\tilde{x}_i^k - x_i^k)\big\} \ge 0, \ \forall\ x_i' \in \mathcal{X}_i. \tag{3.2}$$

Using $\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^m A_j\tilde{x}_j^k - b)$ and by a manipulation, (3.2) can be rewritten as

$$\begin{aligned}\tilde{x}_i^k \in \mathcal{X}_i, \quad & (x_i' - \tilde{x}_i^k)^T\big\{f_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k + A_i^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big)\big\} \\ \ge \quad & (x_i' - \tilde{x}_i^k)^T\big\{A_i^T H\big(\sum_{j=1}^i A_j(x_j^k - \tilde{x}_j^k)\big) + r_i(x_i^k - \tilde{x}_i^k)\big\} \ge 0, \ \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m.\end{aligned} \tag{3.3}$$

If $w^k = \tilde{w}^k$, it follows from (3.3) and (3.1b) that

$$\begin{cases} \tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i' - \tilde{x}_i^k)^T\big\{f_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k\big\} \ge 0, \quad \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m, \\ \qquad\qquad \sum_{j=1}^m A_j\tilde{x}_j^k - b = 0. \end{cases}$$

Thus, the point $\tilde{w}^k$ generated by the proximal alternating direction scheme (3.1) is a test vector for given w^k. In the following we will find $d_1(w, \tilde{w}),\ d_2(w, \tilde{w}) \in \Re^n$ and $\varphi(w, \tilde{w}) \in \Re$ which satisfy conditions (2.11) of the framework in Subsection 2.3.

Lemma 3.1. *Let $\tilde{w}^k$ be generated by the proximal alternating direction scheme* (3.1) *from the given vector w^k. Then, we have*

$$\tilde{w}^k \in \mathcal{W}, \quad (w' - \tilde{w}^k)^T\big(d_2(w^k, \tilde{w}^k) - d_1(w^k, \tilde{w}^k)\big) \ge 0, \qquad \forall\ w' \in \mathcal{W}, \tag{3.4}$$

where

$$d_1(w^k, \tilde{w}^k) = M(w^k - \tilde{w}^k), \tag{3.5}$$

M is defined in (2.4) *and*

$$d_2(w^k,\tilde w^k) = F(\tilde w^k) + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H\Big(\sum_{j=1}^m A_j(x_j^k - \tilde x_j^k)\Big). \tag{3.6}$$

Proof. First, it follows from (3.3) that $\tilde x^k \in \mathcal{X}$ and

$$\begin{pmatrix} x_1' - \tilde x_1^k \\ x_2' - \tilde x_2^k \\ \vdots \\ x_m' - \tilde x_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(\tilde x_1^k) - A_1^T \tilde\lambda^k \\ f_2(\tilde x_2^k) - A_2^T \tilde\lambda^k \\ \vdots \\ f_m(\tilde x_m^k) - A_m^T \tilde\lambda^k \end{pmatrix} + \begin{pmatrix} A_1^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde x_j^k)\big) \\ A_2^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde x_j^k)\big) \\ \vdots \\ A_m^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde x_j^k)\big) \end{pmatrix} \right\}$$

$$\geq \begin{pmatrix} x_1' - \tilde x_1^k \\ x_2' - \tilde x_2^k \\ \vdots \\ x_m' - \tilde x_m^k \end{pmatrix}^T \left\{ \begin{pmatrix} A_1^T H\big(\sum_{j=1}^1 A_j(x_j^k - \tilde x_j^k)\big) \\ A_2^T H\big(\sum_{j=1}^2 A_j(x_j^k - \tilde x_j^k)\big) \\ \vdots \\ A_m^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde x_j^k)\big) \end{pmatrix} + \begin{pmatrix} r_1(x_1^k - \tilde x_1^k) \\ r_2(x_2^k - \tilde x_2^k) \\ \vdots \\ r_m(x_m^k - \tilde x_m^k) \end{pmatrix} \right\} \tag{3.7}$$

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

$$(\lambda' - \tilde\lambda^k)^T \Big(\sum_{i=1}^m A_i \tilde x_i^k - b\Big) = (\lambda' - \tilde\lambda^k)^T H^{-1}(\lambda^k - \tilde\lambda^k)$$

to (3.7), we obtain $\tilde w^k \in \mathcal{W}$ and

$$\begin{pmatrix} x_1' - \tilde x_1^k \\ x_2' - \tilde x_2^k \\ \vdots \\ x_m' - \tilde x_m^k \\ \lambda' - \tilde\lambda^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(\tilde x_1^k) - A_1^T \tilde\lambda^k \\ f_2(\tilde x_2^k) - A_2^T \tilde\lambda^k \\ \vdots \\ f_m(\tilde x_m^k) - A_m^T \tilde\lambda^k \\ \sum_{i=1}^m A_i \tilde x_i^k - b \end{pmatrix} + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H\Big(\sum_{j=1}^m A_j(x_j^k - \tilde x_j^k)\Big) \right\}$$

$$\geq \begin{pmatrix} x_1' - \tilde x_1^k \\ x_2' - \tilde x_2^k \\ \vdots \\ x_m' - \tilde x_m^k \\ \lambda' - \tilde\lambda^k \end{pmatrix}^T \left\{ \begin{pmatrix} A_1^T H\big(\sum_{j=1}^1 A_j(x_j^k - \tilde x_j^k)\big) \\ A_2^T H\big(\sum_{j=1}^2 A_j(x_j^k - \tilde x_j^k)\big) \\ \vdots \\ A_m^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde x_j^k)\big) \\ H^{-1}(\lambda^k - \tilde\lambda^k) \end{pmatrix} + \begin{pmatrix} r_1(x_1^k - \tilde x_1^k) \\ r_2(x_2^k - \tilde x_2^k) \\ \vdots \\ r_m(x_m^k - \tilde x_m^k) \\ 0 \end{pmatrix} \right\},$$

for all $w' \in \mathcal{W}$. Using the notations of $d_1(w^k, \tilde w^k)$ and $d_2(w^k, \tilde w^k)$ the above inequality is

$$\tilde w^k \in \mathcal{W}, \quad (w' - \tilde w^k)^T d_2(w^k, \tilde w^k) \geq (w' - \tilde w^k)^T d_1(w^k, \tilde w^k), \quad \forall\, w' \in \mathcal{W},$$

and the assertion of this lemma is proved. □

Note that the assertion of Lemma 3.1 is the condition (2.11a) of the framework in Subsection 2.3. Indeed, $\|d_1(w^k, \tilde w^k)\| \leq \|M\| \cdot \|w^k - \tilde w^k\|$ (see (3.5)), thus the condition (2.11b) is also satisfied.

Lemma 3.2. *Let $\tilde{w}^k$ be generated by the proximal AD scheme* (3.1) *from the given vector w^k. Then, we have*

$$(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big), \quad \forall\, w^* \in \mathcal{W}^*. \tag{3.8}$$

where $d_2(w^k, \tilde{w}^k)$ is defined in (3.6).

Proof. Using the definition of $d_2(w^k, \tilde{w}^k)$, we get

$$(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) = (\tilde{w}^k - w^*)^T F(\tilde{w}^k) + \big(\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^*)\big)^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big). \tag{3.9}$$

Since $\tilde{w}^k \in \mathcal{W}$ and $w^* \in \mathcal{W}^*$, it follows from (2.10) that $(\tilde{w}^k - w^*)^T F(\tilde{w}^k) \geq 0$. Substituting it in (3.9), we obtain

$$(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) \geq \big(\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^*)\big)^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big). \tag{3.10}$$

Because

$$\sum_{j=1}^m A_j x_j^* = b \qquad \text{and} \qquad H\big(\sum_{j=1}^m A_j \tilde{x}_j^k - b\big) = \lambda^k - \tilde{\lambda}^k,$$

we obtain

$$\big(\sum_{j=1}^m A_j(\tilde{x}_j^k - x_j^*)\big)^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) = (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big). \tag{3.11}$$

Combining (3.10) and (3.11), the lemma is proved. $\square$

By defining

$$\varphi(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k)^T d_1(w^k, \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big), \tag{3.12}$$

it immediately follows from (3.8) that

$$(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k) - (w^k - \tilde{w}^k)^T d_1(w^k, \tilde{w}^k), \quad \forall\, w^* \in \mathcal{W}^*, \tag{3.13}$$

where $d_1(w^k, \tilde{w}^k)$ and $d_2(w^k, \tilde{w}^k)$ are defined in (3.5) and (3.6), respectively. Note that (3.13) is the condition (2.11c) of the framework. The remainder is to show the condition (2.11d).

Lemma 3.3. *Let $\tilde{w}^k$ be generated by the proximal alternating direction scheme* (3.1) *from the given vector w^k. Then, we have*

$$\varphi(w^k, \tilde{w}^k) \geq \frac{1}{2}\big(\sum_{j=1}^m \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2\big) + \sum_{j=1}^m r_j \|x_j^k - \tilde{x}_j^k\|^2, \tag{3.14}$$

where $\varphi(w^k, \tilde{w}^k)$ is defined by (3.12).

Proof. Recall that (see (2.4))

$$
\begin{array}{l}
(w^k-\tilde{w}^k)^T M (w^k-\tilde{w}^k) \\
= \begin{pmatrix} x_1^k-\tilde{x}_1^k \\ x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T
\begin{pmatrix} A_1^T H A_1 & & & & \\ A_2^T H A_1 & A_2^T H A_2 & & & \\ \vdots & \vdots & \ddots & & \\ A_m^T H A_1 & A_m^T H A_2 & \cdots & A_m^T H A_m & \\ 0 & 0 & \cdots & 0 & H^{-1} \end{pmatrix}
\begin{pmatrix} x_1^k-\tilde{x}_1^k \\ x_2^k-\tilde{x}_2^k \\ \vdots \\ x_m^k-\tilde{x}_m^k \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \\
\quad + \sum_{j=1}^m r_j \|x_j^k-\tilde{x}_j^k\|^2 \\
= \frac{1}{2} \begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_m(x_m^k-\tilde{x}_m^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T
\begin{pmatrix} 2H & H & \cdots & H & 0 \\ H & 2H & \cdots & H & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H & H & \cdots & 2H & 0 \\ 0 & 0 & \cdots & 0 & 2H^{-1} \end{pmatrix}
\begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_m(x_m^k-\tilde{x}_m^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \\
\quad + \sum_{j=1}^m r_j \|x_j^k-\tilde{x}_j^k\|^2.
\end{array} \tag{3.15}
$$

By manipulations, we have

$$
\begin{array}{l}
(\lambda^k-\tilde{\lambda}^k)^T \big(\sum_{j=1}^m A_j(x_j^k-\tilde{x}_j^k)\big) \\
= \frac{1}{2} \begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_m(x_m^k-\tilde{x}_m^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T
\begin{pmatrix} 0 & 0 & \dots & 0 & I \\ 0 & 0 & \dots & 0 & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & I \\ I & I & \dots & I & 0 \end{pmatrix}
\begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_m(x_m^k-\tilde{x}_m^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}.
\end{array} \tag{3.16}
$$

Adding (3.15) and (3.16), it follows that

$$
\begin{aligned}
\varphi(w^k,\tilde{w}^k) &= (w^k-\tilde{w}^k)^T M (w^k-\tilde{w}^k) + (\lambda^k-\tilde{\lambda}^k)^T \big(\sum_{j=1}^m A_j(x_j^k-\tilde{x}_j^k)\big) \\
&= \frac{1}{2} \begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_m(x_m^k-\tilde{x}_m^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix}^T
\begin{pmatrix} H & H & \dots & H & I \\ H & H & \dots & H & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H & H & \dots & H & I \\ I & I & \dots & I & H^{-1} \end{pmatrix}
\begin{pmatrix} A_1(x_1^k-\tilde{x}_1^k) \\ A_2(x_2^k-\tilde{x}_2^k) \\ \vdots \\ A_m(x_m^k-\tilde{x}_m^k) \\ \lambda^k-\tilde{\lambda}^k \end{pmatrix} \\
&\quad + \frac{1}{2}\sum_{j=1}^m \|A_j(x_j^k-\tilde{x}_j^k)\|_H^2 + \frac{1}{2}\|\lambda^k-\tilde{\lambda}^k\|_{H^{-1}}^2 + \sum_{j=1}^m r_j\|x_j^k-\tilde{x}_j^k\|^2 \\
&\geq \frac{1}{2}\big(\sum_{j=1}^m \|A_j(x_j^k-\tilde{x}_j^k)\|_H^2 + \|\lambda^k-\tilde{\lambda}^k\|_{H^{-1}}^2\big) + \sum_{j=1}^m r_j\|x_j^k-\tilde{x}_j^k\|^2.
\end{aligned} \tag{3.17}
$$

The last inequality in (3.17) is true because the matrix

$$\begin{pmatrix} H & H & \dots & H & I \\ H & H & \dots & H & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ H & H & \dots & H & I \\ I & I & \dots & I & H^{-1} \end{pmatrix}$$

is positive semi-definite. The lemma is proved. $\Box$

In order to guarantee $\varphi(w^k, \tilde{w}^k)$ to satisfy the condition (2.11d) of the framework in Subsection 2.3, we need the following requirement on r_i in (3.1).

Requirement on $\boldsymbol{r_i}$ in (3.1).

$$r_i\|x_i^k - \tilde{x}_i^k\|^2 \geq \nu\|x_i^k - \tilde{x}_i^k\|^2 - \frac{1}{2}\|A_i(x_i^k - \tilde{x}_i^k)\|_H^2, \quad i = 1, \dots, m. \tag{3.18}$$

Although r_i in (3.18) is not necessary non-negative, we suggest to take $r_i \geq \nu$ for $i = 1, \dots, m$, where $\nu > 0$ is a small scalar.

Lemma 3.4. *Let $\tilde{w}^k$ be generated by the proximal alternating direction scheme* (3.1) *from the given vector w^k. In addition, if the requirement* (3.18) *is satisfied, then*

$$\varphi(w^k, \tilde{w}^k) \geq \nu \sum_{j=1}^{m} \|x_j^k - \tilde{x}_j^k\|^2 + \frac{1}{2}\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2. \tag{3.19}$$

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

Proof. We get the assertion (3.19) from (3.14) and (3.18) directly. If $\varphi(w^k, \tilde{w}^k) = 0$, it follows from (3.19) that

$$x_i^k - \tilde{x}_i^k = 0, \quad i = 1, \dots, m, \quad \text{and} \quad \|\lambda^k - \tilde{\lambda}^k\| = 0.$$

Substituting it in (3.3), we get

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i - \tilde{x}_i^k)^T\{f_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k\} \geq 0, \quad \forall\, x_i^k \in \mathcal{X}_i, \;\; i = 1, \dots, m. \tag{3.20}$$

In addition, we have

$$\sum_{j=1}^{m} A_j\tilde{x}_j^k - b = H^{-1}(\lambda^k - \tilde{\lambda}^k) = 0. \tag{3.21}$$

Combining (3.20) and (3.21) we get

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

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

For given w^k, the trial point $\tilde{w}^k$ produced by the proximal alternating direction scheme (3.1) is a test vector of w^k. Under the requirement (3.18), the analysis in this section proves that $d_1(w^k, \tilde{w}^k)$, $d_2(w^k, \tilde{w}^k)$ (defined in (3.5) and (3.6), respectively) and $\varphi(w^k, \tilde{w}^k)$ (defined in (3.12)) satisfy conditions (2.11) of the framework in Subsection 2.3.

4 PADBC methods and the convergence

Since $d_1(w^k, \tilde{w}^k)$, $d_2(w^k, \tilde{w}^k)$ and $\varphi(w^k, \tilde{w}^k)$ offered by the alternating direction scheme satisfy the conditions (2.11) of the framework of the contraction methods in Subsection 2.3, as pointed in Lemma 2.3 and Lemma 2.4, we have

$$(w^k - w^*)^T d_1(w^k, \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k), \quad \forall w^k \in \Re^{n+l}, w^* \in \Omega^*, \tag{4.1}$$

and

$$(w^k - w^*)^T d_2(w^k, \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k), \quad \forall\, w^k \in \Omega, w^* \in \Omega^*. \tag{4.2}$$

This means that $-G^{-1}d_1(w^k, \tilde{w}^k)$ (resp. $-G^{-1}d_2(w^k, \tilde{w}^k)$) is a descent direction of the unknown distance function $\frac{1}{2}\|w - w^*\|_G^2$ at the point w^k (resp. $w^k \in \mathcal{W}$). Therefore, using the proximal alternating direction scheme outputs as the foot-stone as in [17, 18], we construct the the proximal alternating direction-based contraction method.

Proximal alternating direction-based contraction (PADBC) methods:

Step 0. Given positive definite matrix $G \in \Re^{(n+l)\times(n+l)}$, $\epsilon > 0$ and $w^0 \in \mathcal{W}$.

For $k = 0, 1, \ldots,$ do

Step 1. Produce a trial point $\tilde{w}^k \in \mathcal{W}$ via the proximal alternating direction scheme (3.1).

Step 2. If $\|w^k - \tilde{w}^k\| \leq \epsilon$, then stop, $\tilde{w}^k$ is an approximate solution of (2.1); else set

$$w^{k+1} = w^k - \alpha_k G^{-1} d_1(w^k, \tilde{w}^k) \tag{4.3a}$$

or

$$w^{k+1} = P_{\mathcal{W},G}[w^k - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k)], \tag{4.3b}$$

where

$$\alpha_k = \gamma \alpha_k^*, \qquad \alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|G^{-1}d_1(w^k, \tilde{w}^k)\|_G^2} \qquad \text{and} \qquad \gamma \in (0, 2). \tag{4.4}$$

The sequence $\{w^k\}$ generated by (4.3a) is not necessary contained in $\mathcal{W}$ while the updating form (4.3b) produces a sequence in $\mathcal{W}$. Note that the step size α_k is only dependent on $\varphi(w^k, \tilde{w}^k)$, $d_1(w^k, \tilde{w}^k)$ and γ, even if the search direction in the updating form (4.3b) is $d_2(w^k, \tilde{w}^k)$. The proposed methods use different search directions but the same step length sizes! According to our numerical experience in the similar research [16], the updating form (4.3b) usually outperforms (4.3a). However, we use (4.3b) to get the new iterate only when the projection $P_{\mathcal{W},G}(\cdot)$ is easy to be carried out, and thus the computational load of the updating forms is minor. Since both $d_1(w^k, \tilde{w}^k)$ and $d_2(w^k, \tilde{w}^k)$ are offered by the proximal alternating direction scheme, they are called the proximal AD scheme-based search directions.

The new iterate w^{k+1} is determined by the chosen positive definite matrix G and the step size α_k in the updating form (4.3a) or (4.3b). In order to explain how to determine the step size α_k in (4.3), we define the step-size-dependent new iterate by

$$w_{I}(\alpha_k) = w^k - \alpha_k G^{-1} d_1(w^k, \tilde{w}^k) \tag{4.5}$$

and

$$w_{II}(\alpha_k) = P_{\mathcal{W},G}[w^k - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k)]. \tag{4.6}$$

In this way,

$$\vartheta_{I}(\alpha_k) = \|w^k - w^*\|_G^2 - \|w_{I}(\alpha_k) - w^*\|_G^2 \tag{4.7}$$

and

$$\vartheta_{II}(\alpha_k) = \|w^k - w^*\|_G^2 - \|w_{II}(\alpha_k) - w^*\|_G^2 \tag{4.8}$$

are the distance decrease functions in the k-th iteration by using updating form (4.5) and (4.6), respectively. Since $w^* \in \mathcal{W}^*$ is unknown, we cannot maximize $\vartheta(\alpha_k)$ directly. The following theorem introduces a tight lower bound of $\vartheta_I(\alpha_k)$ and $\vartheta_{II}(\alpha_k)$, namely $q(\alpha_k)$, which does not include the unknown vector w^*.

Theorem 4.1. *For any* $w^* \in \mathcal{W}^*$ *and* $\alpha_k \geq 0$*, we have*

$$\vartheta_I(\alpha_k) \geq q(\alpha_k) \tag{4.9}$$

and

$$\vartheta_{II}(\alpha_k) \geq q(\alpha_k), \tag{4.10}$$

where

$$q(\alpha_k) = 2\alpha_k \varphi(w^k, \tilde{w}^k) - \alpha_k^2 \|G^{-1} d_1(w^k, \tilde{w}^k)\|_G^2. \tag{4.11}$$

Proof. From (4.5) and (4.7) we have

$$\begin{aligned}
\vartheta_I(\alpha_k) &= \|w^k - w^*\|_G^2 - \|w^k - w^* - \alpha_k G^{-1} d_1(w^k, \tilde{w}^k)\|_G^2 \\
&= 2\alpha_k (w^k - w^*)^T d_1(w^k, \tilde{w}^k) - \alpha_k^2 \|G^{-1} d_1(w^k, \tilde{w}^k)\|_G^2 \\
&\geq 2\alpha_k \varphi(w^k, \tilde{w}^k) - \alpha_k^2 \|G^{-1} d_1(w^k, \tilde{w}^k)\|_G^2.
\end{aligned}$$

The last inequality follows from (4.1) and the right hand side of the above inequality is $q(\alpha_k)$. Now we turn to prove $\vartheta_{II}(\alpha_k) \geq q(\alpha_k)$. Since $w^* \in \mathcal{W}$, we have (see (2.7))

$$\|P_{\mathcal{W},G}(v) - w^*\|_G^2 \leq \|v - w^*\|_G^2 - \|v - P_{\mathcal{W},G}(v)\|_G^2, \quad \forall v \in \Re^{n+l}.$$

Setting $v = w^k - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k)$ and $w_{II}(\alpha_k) = P_{\mathcal{W},G}(v)$ in the above inequality, we get

$$\begin{aligned}
\|w_{II}(\alpha_k) - w^*\|_G^2 \leq\; & \|w^k - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k) - w^*\|_G^2 \\
& - \|w^k - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k) - w_{II}(\alpha_k)\|_G^2.
\end{aligned}$$

Substituting it in (4.8), we obtain

$$\begin{aligned}
\vartheta_{II}(\alpha_k) &\geq \|w^k - w^*\|_G^2 - \|w^k - w^* - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k)\|_G^2 \\
&\quad + \|w^k - w_{II}(\alpha_k) - \alpha_k G^{-1} d_2(w^k, \tilde{w}^k)\|_G^2 \\
&= \|w^k - w_{II}(\alpha_k)\|_G^2 + 2\alpha_k (w_{II}(\alpha_k) - w^*)^T d_2(w^k, \tilde{w}^k).
\end{aligned} \tag{4.12}$$

Since $w_{II}(\alpha_k) \in \mathcal{W}$, it follows from (3.4) that

$$(w_{II}(\alpha_k) - \tilde{w}^k)^T d_2(w^k, \tilde{w}^k) \geq (w_{II}(\alpha_k) - \tilde{w}^k)^T d_1(w^k, \tilde{w}^k). \tag{4.13}$$

Adding (see (3.13))

$$(\tilde{w}^k - w^*)^T d_2(w^k, \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k) - (w^k - \tilde{w}^k)^T d_1(w^k, \tilde{w}^k)$$

to (4.13), we get

$$(w_{II}(\alpha_k) - w^*)^T d_2(w^k, \tilde{w}^k) \geq \varphi(w^k, \tilde{w}^k) + (w_{II}(\alpha_k) - w^k)^T d_1(w^k, \tilde{w}^k). \tag{4.14}$$

Substituting (4.14) in the right hand side of (4.12), we obtain

$$\begin{aligned}
\vartheta_{II}(\alpha_k) &\geq \|w^k - w_{II}(\alpha_k)\|_G^2 + 2\alpha \varphi(w^k, \tilde{w}^k) + 2\alpha_k (w_{II}(\alpha_k) - w^k)^T d_1(w^k, \tilde{w}^k) \\
&= \|w^k - w_{II}(\alpha_k) - \alpha_k G^{-1} d_1(w^k, \tilde{w}^k)\|_G^2 + 2\alpha_k \varphi(w^k, \tilde{w}^k) - \alpha_k^2 \|G^{-1} d_1(w^k, \tilde{w}^k)\|_G^2 \\
&\geq 2\alpha_k \varphi(w^k, \tilde{w}^k) - \alpha_k^2 \|G^{-1} d_1(w^k, \tilde{w}^k)\|_G^2.
\end{aligned}$$

This completes the proof. $\square$

By defining

$$\vartheta(\alpha_k) = \min\{\vartheta_{I}(\alpha_k), \vartheta_{II}(\alpha_k)\}, \tag{4.15}$$

it follows from Theorem 4.1 that

$$\vartheta(\alpha_k) \geq q(\alpha_k). \tag{4.16}$$

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

$$\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k)}{\|G^{-1}d_1(w^k, \tilde{w}^k)\|_G^2}, \tag{4.17}$$

and this is just the same as defined in (4.4). Moreover, due to (2.11b) and (2.11d), $\alpha_k^* > 0$ is bounded away from zero. In other words, there is a constant $c_0 > 0$ such that

$$\alpha_k^* \geq c_0, \quad \forall\, k \geq 0. \tag{4.18}$$

From (4.9), (4.11) and (4.17) follows that

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

Because some inequalities are used in the proof of (4.9) and (4.10), in practical computation, taking a relaxed factor $\gamma > 1$ is useful for fast convergence. By using (4.11) and (4.17), we have

$$\begin{aligned} q(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\varphi(w^k, \tilde{w}^k) - (\gamma\alpha_k^*)^2\|G^{-1}d_1(w^k, \tilde{w}^k)\|_G^2 \\ &= \gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k). \end{aligned} \tag{4.19}$$

In order to guarantee that the right hand side of (4.19) is positive, we take $\gamma \in [1, 2)$. The following theorem points out that the sequence $\{w^k\}$ generated by the proposed method is Fejér monotone with respect to $\mathcal{W}^*$.

Theorem 4.2. *For any $w^* \in \mathcal{W}^*$, the sequence $\{w^k\}$ generated by the proposed method with contraction form* (4.3a) *or* (4.3b) *satisfies*

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k), \quad \forall\, w^* \in \mathcal{W}^*. \tag{4.20}$$

Proof. Because the contraction form is (see (4.3a))

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

or

$$w^{k+1} = P_{\mathcal{W}, G}[w^k - \gamma\alpha_k^* G^{-1} d_2(w^k, \tilde{w}^k)],$$

it follows from (4.9) and (4.10) that

$$\vartheta(\gamma\alpha_k^*) \geq q(\gamma\alpha_k^*). \tag{4.21}$$

Therefore,

$$\|w^{k+1} - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - q(\gamma\alpha_k^*). \tag{4.22}$$

The result of this theorem follows from (4.19) directly. $\square$

Theorem 4.2 tells us that the generated sequence $\{w^k\}$ approaches the solution set monotonically in the Fejér sense. Thus, according to [3], the proposed methods belong to the contraction methods. Thus, they are called Proximal Alternating Direction Based Contraction Methods.

Theorem 4.3. *Let $\{w^k\}$ and $\{\tilde{w}^k\}$ be the sequences generated by the proposed PADBC method for (1.1). Then, we have:*

1. *The sequence $\{w^k\}$ is bounded.*
2. $\lim_{k\to\infty}\{\|x_1^k - \tilde{x}_1^k\|^2 + \cdots + \|x_m^k - \tilde{x}_m^k\|^2 + \|\lambda^k - \tilde{\lambda}^k\|^2\} = 0.$
3. *The sequence $\{\tilde{w}^k\}$ converges to some $w^\infty \in \mathcal{W}^*$.*

Proof. The first assertion follows from (4.20) directly. Moreover, since $\alpha_k^* \geq c_0$, from (4.20) we get

$$\sum_{k=0}^{\infty} \gamma(2-\gamma)c_0\varphi(w^k, \tilde{w}^k) \leq \|w^0 - w^*\|_G^2$$

and thus

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

Consequently, it follows from Lemma 3.4 that

$$\lim_{k\to\infty} \|x_i^k - \tilde{x}_i^k\| = 0, \quad i = 1, \ldots, m \quad \text{and} \quad \lim_{k\to\infty} \|\lambda^k - \tilde{\lambda}^k\| = 0.$$

The second assertion is proved. Substituting it in (3.3), we get

$$\tilde{x}_i^k \in \mathcal{X}_i, \quad \lim_{k\to\infty} (x_i - \tilde{x}_i^k)^T\{f_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k\} \geq 0, \quad \forall x_i \in \mathcal{X}_i, \;\; i = 1, \ldots, m. \tag{4.23}$$

In addition, we have

$$\lim_{k\to\infty} (\sum_{j=1}^m A_j\tilde{x}_j^k - b) = \lim_{k\to\infty} H^{-1}(\lambda^k - \tilde{\lambda}^k) = 0. \tag{4.24}$$

Combining (4.23) and (4.24) we get

$$\tilde{w}^k \in \mathcal{W}, \quad \lim_{k\to\infty} (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq 0, \quad \forall\, w \in \mathcal{W} \tag{4.25}$$

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

Let w^∞ be a cluster point of $\{\tilde{w}^k\}$ and the subsequence $\{\tilde{w}^{k_j}\}$ converges to w^∞. It follows from (4.25) that

$$\tilde{w}^{k_j} \in \mathcal{W}, \quad \lim_{k\to\infty} (w - \tilde{w}^{k_j})^T F(\tilde{w}^{k_j}) \geq 0, \quad \forall\, w \in \mathcal{W} \tag{4.26}$$

and consequently

$$\tilde{w}^\infty \in \mathcal{W}, \quad \lim_{k\to\infty} (w - \tilde{w}^\infty)^T F(\tilde{w}^\infty) \geq 0, \quad \forall\, w \in \mathcal{W}.$$

This means that $w^\infty \in \mathcal{W}^*$. Since $\{w^k\}$ is Fejér monotone and $\lim_{k\to\infty} \|w^k - \tilde{w}^k\| = 0$, the sequence $\{\tilde{w}^k\}$ cannot have other cluster point and $\{\tilde{w}^k\}$ converges to $w^\infty \in \mathcal{W}^*$. □

In the proposed methods, H and G are given positive definite matrices, in practical computation, we can take H (resp. G) as a scalar matrix (resp. identity matrix). The convergence proofs are based on the general framework in Subsection 2.3 and thus similar as those in [17, 18]. Since the practical objective of adding the proximal term is to simplify the optimization subproblems (1.8) for producing the trial vector $\tilde{w}^k$, in the following we consider linearized versions of the proximal alternating direction schemes. For the validity to use the general framework, we need only to show that the directions $d_1(w^k, \tilde{w}^k)$ and $d_2(w^k, \tilde{w}^k)$ and error measure function $\varphi(w^k, \tilde{w}^k)$ offered by different linearized versions satisfy the conditions (2.11).

5 Linearized version A of the proximal AD scheme

In the linearized version A of the proximal alternating direction scheme, we only linearize the part of the quadratic function in $\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k)$ of (3.1a) at x_i^k. Dropping the constant term

$$\frac{1}{2}\|\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b\|_H^2,$$

the quadratic function

$$\frac{1}{2}\|\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b\|_H^2$$

is replaced by

$$(x_i - x_i^k)^T A_i^T H(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b).$$

The linearized version A of the proximal alternating direction scheme:
For $i = 1, \ldots, m$, with available $\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i^k, x_{i+1}^k, \ldots, x_m^k$ and λ^k, we solve the minimization problem

$$\min\left\{ \begin{array}{l} \theta_i(x_i) - (\lambda^k)^T(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b) \\ +(x_i - x_i^k)^T A_i^T H(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b) + \frac{r_i}{2}\|x_i - x_i^k\|^2 \end{array} \middle| \, x_i \in \mathcal{X}_i \right\}, \quad (5.1a)$$

and denote the solution by $\tilde{x}_i^k$. Then set

$$\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^{m} A_j\tilde{x}_j^k - b). \quad (5.1b)$$

The analysis is parallel to that in Section 3. The variable in (5.1a) is x_i and its first order optimal condition can be characterized as the following variational inequality: $\tilde{x}_i^k \in \mathcal{X}_i$,

$$(x_i' - \tilde{x}_i^k)^T\{f_i(\tilde{x}_i^k) - A_i^T\lambda^k + A_i^T H(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b) + r_i(\tilde{x}_i^k - x_i^k)\} \ge 0, \ \forall\, x_i' \in \mathcal{X}_i. \quad (5.2)$$

Using $\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^{m} A_j\tilde{x}_j^k - b)$ and by a manipulation, (5.3) can be rewritten as

$$\begin{array}{ll} \tilde{x}_i^k \in \mathcal{X}_i, & (x_i' - \tilde{x}_i^k)^T\{f_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k + A_i^T H\big(\sum_{j=1}^{m} A_j(x_j^k - \tilde{x}_j^k)\big)\} \\ \ge & (x_i' - \tilde{x}_i^k)^T\{A_i^T H\big(\sum_{j=1}^{i-1} A_j(x_j^k - \tilde{x}_j^k)\big) + r_i(x_i^k - \tilde{x}_i^k)\} \ge 0, \ \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m. \end{array} \quad (5.3)$$

Note that (5.3) is obtained by substituting

$$A_i^T H\big(\sum_{j=1}^{i} A_j(x_j^k - \tilde{x}_j^k)\big) \quad \text{in the right hand side of (3.3) by} \quad A_i^T H\big(\sum_{j=1}^{i-1} A_j(x_j^k - \tilde{x}_j^k)\big).$$

If $w^k = \tilde{w}^k$, it follows from (5.3) and (5.1b) that

$$\begin{cases} \tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i' - \tilde{x}_i^k)^T \{ f_i(\tilde{x}_i^k) - A_i^T \tilde{\lambda}^k \} \geq 0, \quad \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m, \\ \qquad \sum_{j=1}^m A_j \tilde{x}_j^k - b = 0. \end{cases}$$

Thus, $\tilde{w}^k$ generated by the linearized version A of the proximal alternating direction scheme (5.1) is a test vector for given w^k. In the following we will find $d_1^A(w, \tilde{w})$, $d_2^A(w, \tilde{w}) \in \Re^n$ and $\varphi^A(w, \tilde{w}) \in \Re$ which satisfy Conditions (2.11) of the framework in Subsection 2.3.

Lemma 5.1. *Let $\tilde{w}^k$ be generated by the linearized version A of the proximal AD scheme* (5.1) *from the given vector w^k. Then, we have*

$$\tilde{w}^k \in \mathcal{W}, \quad (w' - \tilde{w}^k)^T \big(d_2^A(w^k, \tilde{w}^k) - d_1^A(w^k, \tilde{w}^k) \big) \geq 0, \qquad \forall\, w' \in \mathcal{W}, \tag{5.4}$$

where

$$d_1^A(w^k, \tilde{w}^k) = M(w^k - \tilde{w}^k) - \begin{pmatrix} A_1^T H A_1 (x_1^k - \tilde{x}_1^k) \\ A_2^T H A_2 (x_2^k - \tilde{x}_2^k) \\ \vdots \\ A_m^T H A_m (x_m^k - \tilde{x}_m^k) \\ 0 \end{pmatrix}, \tag{5.5}$$

and M is defined in (2.4). *And*

$$d_2^A(w^k, \tilde{w}^k) = F(\tilde{w}^k) + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H \big(\sum_{j=1}^m A_j (x_j^k - \tilde{x}_j^k) \big) \tag{5.6}$$

is the same $d_2(w^k, \tilde{w}^k)$ defined in (3.6).

Proof. The proof is similar as those of Lemma 3.1. In comparison (3.3) and (5.3) we obtain $\tilde{w}^k \in \mathcal{W}$ and

$$\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_m' - \tilde{x}_m^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(\tilde{x}_1^k) - A_1^T \tilde{\lambda}^k \\ f_2(\tilde{x}_2^k) - A_2^T \tilde{\lambda}^k \\ \vdots \\ f_m(\tilde{x}_m^k) - A_m^T \tilde{\lambda}^k \\ \sum_{i=1}^m A_i \tilde{x}_i^k - b \end{pmatrix} + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H \big(\sum_{j=1}^m A_j (x_j^k - \tilde{x}_j^k) \big) \right\}$$

$$\geq \begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_m' - \tilde{x}_m^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} A_1^T H \big(\sum_{j=1}^0 A_j (x_j^k - \tilde{x}_j^k) \big) \\ A_2^T H \big(\sum_{j=1}^1 A_j (x_j^k - \tilde{x}_j^k) \big) \\ \vdots \\ A_m^T H \big(\sum_{j=1}^{m-1} A_j (x_j^k - \tilde{x}_j^k) \big) \\ H^{-1} (\lambda^k - \tilde{\lambda}^k) \end{pmatrix} + \begin{pmatrix} r_1 (x_1^k - \tilde{x}_1^k) \\ r_2 (x_2^k - \tilde{x}_2^k) \\ \vdots \\ r_m (x_m^k - \tilde{x}_m^k) \\ 0 \end{pmatrix} \right\}, \tag{5.7}$$

for all $w' \in \mathcal{W}$. Note that (5.7) is obtained by substituting

$$A_i^T H \big(\sum_{j=1}^i A_j (x_j^k - \tilde{x}_j^k) \big)$$

in the right hand side of (3.7) by

$$A_i^T H\big(\sum_{j=1}^{i-1} A_j(x_j^k - \tilde{x}_j^k)\big).$$

In addition, the left hand side of (5.7) is

$$(w' - \tilde{w}^k)^T d_2^A(w^k, \tilde{w}^k).$$

Using the notation of $d_1^A(w^k, \tilde{w}^k)$ to the right hand side of (5.7), it can be expressed as

$$\tilde{w}^k \in \mathcal{W}, \quad (w' - \tilde{w}^k)^T d_2^A(w^k, \tilde{w}^k) \ge (w' - \tilde{w}^k)^T d_1^A(w^k, \tilde{w}^k), \quad \forall\, w' \in \mathcal{W}.$$

The assertion of this lemma is proved. □

Note that the assertion of Lemma 5.1 is the condition (2.11a) of the framework in Subsection 2.3. Indeed, there is a constant $K > 0$ such that $\|d_1^A(w^k, \tilde{w}^k)\| \le K\|w^k - \tilde{w}^k\|$ (see (5.5)). Thus, the condition (2.11b) is also satisfied. Since the updating form of $\tilde{\lambda}^k$ in this section are the same as in Section 3 and definition of $d_2^A(w^k, \tilde{w}^k) = d_2(w^k, \tilde{w}^k)$, we have:

Lemma 5.2. *Let $\tilde{w}^k$ be generated by the linearized version A of the proximal AD scheme* (5.1) *from the given vector w^k. Then, we have*

$$(\tilde{w}^k - w^*)^T d_2^A(w^k, \tilde{w}^k) \ge (\lambda^k - \tilde{\lambda}^k)^T\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \quad \forall\, w^* \in \mathcal{W}^*. \tag{5.8}$$

Moreover, by defining

$$\varphi^A(w^k, \tilde{w}^k) := (w^k - \tilde{w}^k)^T d_1^A(w^k, \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big), \tag{5.9}$$

we obtain

$$(\tilde{w}^k - w^*)^T d_2^A(w^k, \tilde{w}^k) \ge \varphi^A(w^k, \tilde{w}^k) - (w^k - \tilde{w}^k)^T d_1^A(w^k, \tilde{w}^k) \quad \forall\, w^* \in \mathcal{W}^*, \tag{5.10}$$

where $d_1^A(w^k, \tilde{w}^k)$ and $d_2^A(w^k, \tilde{w}^k)$ are defined in (5.5) and (5.6), respectively. Note that (5.10) is the condition (2.11c) of the framework and the remainder is to show the condition (2.11d).

Lemma 5.3. *Let $\tilde{w}^k$ be generated by the linearized version A of the proximal AD scheme* (5.1) *from the given vector w^k. Then, we have*

$$\begin{aligned}\varphi^A(w^k, \tilde{w}^k) \quad \ge \quad & \frac{1}{2}\big(\sum_{j=1}^m \|A_j(x_j^k - \tilde{x}_j^k)\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2\big) \\ & + \sum_{j=1}^m \Big(r_j\|x_j^k - \tilde{x}_j^k\|^2 - (x_j^k - \tilde{x}_j^k)^T A_j^T H A_j(x_j^k - \tilde{x}_j^k)\Big),\end{aligned} \tag{5.11}$$

where $\varphi^A(w^k, \tilde{w}^k)$ is defined by (5.9).

Proof. The proof is similar as that in Lemma 3.3 and thus omitted. □

In order to guarantee $\varphi^A(w^k, \tilde{w}^k)$ to satisfy the condition (2.11d) of the framework in Subsection 2.3, we need the following requirement on r_i in (5.1).

Requirement on r_i in (5.1).

$$r_i\|x_i^k - \tilde{x}_i^k\|^2 \ge \nu\|x_i^k - \tilde{x}_i^k\|^2 + \frac{1}{2}\|A_i(x_i^k - \tilde{x}_i^k)\|_H^2. \tag{5.12}$$

This requirement is satisfied when $r_i \ge \nu + \frac{1}{2}\|A_i^T H A_i\|$.

Lemma 5.4. *Let $\tilde{w}^k$ be generated by the linearized version A of the proximal AD scheme* (5.1) *from the given vector w^k. If the requirement* (5.12) *is satisfied, then*

$$\varphi^A(w^k, \tilde{w}^k) \geq \nu \sum_{j=1}^{m} \|x_j^k - \tilde{x}_j^k\|^2 + \frac{1}{2}\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2. \tag{5.13}$$

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

Proof. We get the assertion (5.13) from (5.11) and (5.12) directly. The reminder of the proof is the same as that in Theorem 3.4. □

For given w^k, the trial point $\tilde{w}^k$ produced by the linearized version A of the proximal alternating direction scheme (5.1) is a test vector of w^k. Under the requirement (5.12), the analysis in this section proves that $d_1^A(w^k, \tilde{w}^k)$, $d_2^A(w^k, \tilde{w}^k)$ (defined in (5.5) and (5.6), respectively) and $\varphi^A(w^k, \tilde{w}^k)$ (defined in (5.9)) satisfy conditions (2.11) of the framework in Subsection 2.3. Since $d_2^A(w^k, \tilde{w}^k) = d_2(w^k, \tilde{w}^k)$, substituting $d_1(w^k, \tilde{w}^k)$ and $\varphi(w^k, \tilde{w}^k)$ in the updating forms (4.3) by $d_1^A(w^k, \tilde{w}^k)$ and $\varphi^A(w^k, \tilde{w}^k)$, respectively, the resultant method is a contraction method and its convergence proofs are same as those in Section 4.

6 Linearized version B of the proximal AD scheme

In the linearized version B of the proximal alternating direction scheme, we linearize the function $\theta_i(x_i)$ in $\mathcal{L}_A(\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i, x_{i+1}^k, \ldots, x_m^k, \lambda^k)$ of (3.1a) at x_i^k. In this case we assume that each $\theta_i(x_i), i = 1, \ldots, m$ is differentiable and $f_i(x_i) = \nabla\theta_i(x_i)$ is Lipschitz continuous. Thus, $\theta(x_i)$ is replaced (dropping the constant term $\theta_i(x_i^k)$) by

$$f_i(x_i^k)^T(x_i - x_i^k).$$

The linearized version B of the proximal alternating direction scheme:
For $i = 1, \ldots, m$, with available $\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i^k, x_{i+1}^k, \ldots, x_m^k$ and λ^k, we solve the minimization problem

$$\min \left\{ \begin{array}{l} f_i(x_i^k)^T(x_i - x_i^k) - (\lambda^k)^T(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b) \\ +\frac{1}{2}\|\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b\|_H^2 + \frac{r_i}{2}\|x_i - x_i^k\|^2 \end{array} \middle| \, x_i \in \mathcal{X}_i \right\}, \tag{6.1a}$$

and denote the solution by $\tilde{x}_i^k$. Then set

$$\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^{m} A_j\tilde{x}_j^k - b). \tag{6.1b}$$

The analysis is parallel to that in Section 3. Again, the variable in (6.1a) is x_i and the first order optimal condition of the i-th sub-problem can be characterized as the following variational inequality: $\tilde{x}_i^k \in \mathcal{X}_i$,

$$(x_i' - \tilde{x}_i^k)^T\big\{f_i(x_i^k) - A_i^T\lambda^k + A_i^T H\big(\sum_{j=1}^{i} A_j\tilde{x}_j^k + \sum_{j=i+1}^{m} A_j x_j^k - b\big) + r_i(\tilde{x}_i^k - x_i^k)\big\} \geq 0, \ \forall\, x_i' \in \mathcal{X}_i. \tag{6.2}$$

Using $\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^m A_j\tilde{x}_j^k - b)$ and by a manipulation, (6.2) can be rewritten as

$$\begin{aligned}
&\tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i' - \tilde{x}_i^k)^T\big\{f_i(x_i^k) - A_i^T\tilde{\lambda}^k + A_i^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big)\big\} \\
&\geq \quad (x_i' - \tilde{x}_i^k)^T\big\{A_i^T H\big(\sum_{j=1}^i A_j(x_j^k - \tilde{x}_j^k)\big) + r_i(x_i^k - \tilde{x}_i^k)\big\} \geq 0, \ \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m. \quad (6.3)
\end{aligned}$$

Note that (6.3) is obtained by substituting

$$f_i(\tilde{x}_i^k) \qquad \text{in the left hand side of (3.3) by} \qquad f_i(x_i^k).$$

If $w^k = \tilde{w}^k$, it follows from (6.3) and (6.1b) that

$$\begin{cases} \tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i' - \tilde{x}_i^k)^T\{f_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k\} \geq 0, \quad \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m, \\ \qquad\qquad \sum_{j=1}^m A_j\tilde{x}_j^k - b = 0. \end{cases}$$

Thus, $\tilde{w}^k$ generated by the linearized version B of the proximal alternating direction scheme (6.1) is a test vector for given w^k. In the following we will find $d_1^B(w, \tilde{w})$, $d_2^B(w, \tilde{w}) \in \Re^n$ and $\varphi^B(w, \tilde{w}) \in \Re$ which satisfy conditions (2.11) of the framework in Subsection 2.3.

Lemma 6.1. *Let $\tilde{w}^k$ be generated by the linearized version B of the proximal AD scheme* (6.1) *from the given vector w^k. Then, we have*

$$\tilde{w}^k \in \mathcal{W}, \quad (w' - \tilde{w}^k)^T\big(d_2^B(w^k, \tilde{w}^k) - d_1^B(w^k, \tilde{w}^k)\big) \geq 0, \quad \forall\, w' \in \mathcal{W}, \tag{6.4}$$

where

$$d_1^B(w^k, \tilde{w}^k) = M(w^k - \tilde{w}^k) - \begin{pmatrix} f_1(x_1^k) - f_1(\tilde{x}_1^k) \\ f_2(x_2^k) - f_2(\tilde{x}_2^k) \\ \vdots \\ f_m(x_m^k) - f_m(\tilde{x}_m^k) \\ 0 \end{pmatrix}, \tag{6.5}$$

and M is defined in (2.4). *And*

$$d_2^B(w^k, \tilde{w}^k) = F(\tilde{w}^k) + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \tag{6.6}$$

is the same $d_2(w^k, \tilde{w}^k)$ defined in (3.6).

Proof. The proof is similar as that of Lemma 3.1. In comparison (3.3) and (6.3) we obtain $\tilde{w}^k \in \mathcal{W}$ and

$$\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_m' - \tilde{x}_m^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(x_1^k) - A_1^T \tilde{\lambda}^k \\ f_2(x_2^k) - A_2^T \tilde{\lambda}^k \\ \vdots \\ f_m(x_m^k) - A_m^T \tilde{\lambda}^k \\ \sum_{i=1}^m A_i \tilde{x}_i^k - b \end{pmatrix} + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \right\}$$

$$\geq \begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_m' - \tilde{x}_m^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} A_1^T H\big(\sum_{j=1}^1 A_j(x_j^k - \tilde{x}_j^k)\big) \\ A_2^T H\big(\sum_{j=1}^2 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ A_m^T H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \\ H^{-1}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} + \begin{pmatrix} r_1(x_1^k - \tilde{x}_1^k) \\ r_2(x_2^k - \tilde{x}_2^k) \\ \vdots \\ r_m(x_m^k - \tilde{x}_m^k) \\ 0 \end{pmatrix} \right\}, \tag{6.7}$$

for all $w' \in \mathcal{W}$. Note that the right hand side of (6.7) is $(w' - \tilde{w}^k)^T M(w^k - \tilde{w}^k)$. Adding

$$\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_m' - \tilde{x}_m^k \\ 0 \end{pmatrix}^T \begin{pmatrix} f_1(\tilde{x}_1^k) - f_1(x_1^k) \\ f_2(\tilde{x}_2^k) - f_2(x_2^k) \\ \vdots \\ f_m(\tilde{x}_m^k) - f_m(x_m^k) \\ 0 \end{pmatrix}$$

to the both sides of (6.7) and using the notations of $d_1^B(w^k, \tilde{w}^k)$ and $d_2^B(w^k, \tilde{w}^k)$, the above inequality is

$$\tilde{w}^k \in \mathcal{W}, \quad (w' - \tilde{w}^k)^T d_2^B(w^k, \tilde{w}^k) \geq (w' - \tilde{w}^k)^T d_1^B(w^k, \tilde{w}^k), \quad \forall\, w' \in \mathcal{W}$$

and the assertion of this lemma is proved. $\square$

Note that the assertion of Lemma 6.1 is the condition (2.11a) of the framework in Subsection 2.3 for the problem $VI(\mathcal{W}, F)$. Under the assumption that f_i is Lipschitz continuous, there is a constant $K > 0$ such that $\|d_1^B(w^k, \tilde{w}^k)\| \leq K\|w^k - \tilde{w}^k\|$ (see (6.5)). Thus, the condition (2.11b) is also satisfied. Since the updating form of $\tilde{\lambda}^k$ in this section are the same as in Section 3 and $d_2^B(w^k, \tilde{w}^k) = d_2(w^k, \tilde{w}^k)$, we have

Lemma 6.2. *Let $\tilde{w}^k$ be generated by the linearized version B of the proximal AD scheme* (6.1) *from the given vector w^k. Then, we have*

$$(\tilde{w}^k - w^*)^T d_2^B(w^k, \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \quad \forall\, w^* \in \mathcal{W}^*. \tag{6.8}$$

Moreover, by defining

$$\varphi^B(w^k, \tilde{w}^k) = (w^k - \tilde{w}^k)^T d_1^B(w^k, \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T \big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big), \tag{6.9}$$

we have

$$(\tilde{w}^k - w^*)^T d_2^B(w^k, \tilde{w}^k) \geq \varphi^B(w^k, \tilde{w}^k) - (w^k - \tilde{w}^k)^T d_1^B(w^k, \tilde{w}^k) \quad \forall\, w^* \in \mathcal{W}^*, \tag{6.10}$$

$d_1^B(w^k, \tilde{w}^k)$ and $d_2^B(w^k, \tilde{w}^k)$ are defined in (6.5) and (6.6), respectively. Note that (6.10) is the condition (2.11c) of the framework and the remainder is to show the condition (2.11d).

Lemma 6.3. *Let $\tilde{w}^k$ be generated by the linearized version B of the proximal AD scheme* (6.1) *from the given vector w^k. Then, we have*

$$\begin{aligned}\varphi^B(w^k,\tilde{w}^k) \quad \geq \quad & \frac{1}{2}\Big(\sum_{j=1}^{m}\|A_j(x_j^k-\tilde{x}_j^k)\|_H^2+\|\lambda^k-\tilde{\lambda}^k\|_{H^{-1}}^2\Big) \\ & +\sum_{j=1}^{m}\Big(r_j\|x_j^k-\tilde{x}_j^k\|^2-(x_j^k-\tilde{x}_j^k)^T(f(x_j^k)-f(\tilde{x}_j^k)\Big), \end{aligned} \qquad (6.11)$$

where $\varphi^B(w^k,\tilde{w}^k)$ is defined by (6.9).

Proof. The proof is similar as those in Lemma 3.3 and thus omitted. □

In order to guarantee $\varphi^B(w^k,\tilde{w}^k)$ to satisfy the condition (2.11d) of the framework in Subsection 2.3, we need the following requirement on r_i in (6.1).

Requirement on r_i in (6.1).

$$r_i\|x_i^k-\tilde{x}_i^k\|^2 \geq \nu\|x_i^k-\tilde{x}_i^k\|^2-\frac{1}{2}\|A_i(x_i^k-\tilde{x}_i^k)\|_H^2+(x_i^k-\tilde{x}_i^k)^T\big(f(x_i^k)-f(\tilde{x}_i^k)\big). \qquad (6.12)$$

Obviously, when $r_i \geq \nu + l_i$ (see (1.9)), the requirement is satisfied.

Lemma 6.4. *Let $\tilde{w}^k$ be generated by the linearized version B of the proximal AD scheme* (6.1) *from the given vector w^k. If the requirement* (6.12) *is satisfied, then*

$$\varphi^B(w^k,\tilde{w}^k) \geq \nu\sum_{j=1}^{m}\|x_j^k-\tilde{x}_j^k\|^2+\frac{1}{2}\|\lambda^k-\tilde{\lambda}^k\|_{H^{-1}}^2. \qquad (6.13)$$

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

Proof. We get the assertion (6.13) from (6.11) and (6.12) directly. The reminder of the proof is the same as that in Lemma 3.4. □

For given w^k, the trial point $\tilde{w}^k$ produced by the linearized version B of the proximal alternating direction scheme (6.1) is a test vector of w^k. Under the requirement (6.12), the analysis in this section proves that $d_1^B(w^k,\tilde{w}^k)$, $d_2^B(w^k,\tilde{w}^k)$ (defined in (6.5) and (6.6), respectively) and $\varphi^B(w^k,\tilde{w}^k)$ (defined in (6.9)) satisfy conditions (2.11) of the framework in Subsection 2.3. Since $d_2^B(w^k,\tilde{w}^k)=d_2(w^k,\tilde{w}^k)$, substituting $d_1(w^k,\tilde{w}^k)$ and $\varphi(w^k,\tilde{w}^k)$ in the updating forms (4.3) by $d_1^B(w^k,\tilde{w}^k)$ and $\varphi^B(w^k,\tilde{w}^k)$, respectively, the resultant method is a contraction method and its convergence proofs are same as those in Section 4.

7 Linearized version C of the proximal AD scheme

In the linearized version C of the proximal alternating direction scheme, we linearize both

$$\theta_i(x_i) \qquad \text{and} \qquad \frac{1}{2}\|\sum_{j=1}^{i-1}A_j\tilde{x}_j^k+A_ix_i+\sum_{j=i+1}^{m}A_jx_j^k-b\|_H^2$$

in (3.1a) at x_i^k. Dropping the constant term

$$\theta_i(x_i^k) \qquad \text{and} \qquad \frac{1}{2}\|\sum_{j=1}^{i-1}A_j\tilde{x}_j^k+\sum_{j=i}^{m}A_jx_j^k-b\|_H^2,$$

the function

$$\theta_i(x_i) + \frac{1}{2}\|\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b\|_H^2$$

is replaced by

$$f(x_i^k)^T(x_i - x_i^k) + (x_i - x_i^k)^T A_i^T H\big(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b\big).$$

For the given $w^k = (x_1^k, \cdots, x_m^k, \lambda^k)$, the task of the alternating directions step of the k-th iteration generates a $\tilde{w}^k = (\tilde{x}_1^k, \cdots, \tilde{x}_m^k, \tilde{\lambda}^k)$ via the following procedure:

The linearized version C of the proximal alternating direction scheme:
For $i = 1, \ldots, m$, with available $\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i^k, x_{i+1}^k, \ldots, x_m^k$ and λ^k, we solve the minimization problem

$$\min \left\{ \begin{array}{l} f_i(x_i^k)^T(x_i - x_i^k) - (\lambda^k)^T(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + A_i x_i + \sum_{j=i+1}^{m} A_j x_j^k - b) \\ +(x_i - x_i^k)^T A_i^T H(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b) + \frac{r_i}{2}\|x_i - x_i^k\|^2 \end{array} \middle| \, x_i \in \mathcal{X}_i \right\}. \quad (7.1a)$$

and denote the solution by $\tilde{x}_i^k$. Then set

$$\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^{m} A_j\tilde{x}_j^k - b). \quad (7.1b)$$

Note that the variable in (7.1a) is x_i and its first order optimal condition can be characterized as the following variational inequality: $\tilde{x}_i^k \in \mathcal{X}_i$,

$$(x_i' - \tilde{x}_i^k)^T\big\{f_i(x_i^k) - A_i^T[\lambda^k - H(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b)] + r_i(\tilde{x}_i^k - x_i^k)\big\} \geq 0, \ \forall\, x_i' \in \mathcal{X}_i. \quad (7.2)$$

Using $\tilde{\lambda}^k = \lambda^k - H(\sum_{j=1}^{m} A_j\tilde{x}_j^k - b)$ and by a manipulation, (7.2) can be rewritten as

$$\begin{aligned} &\tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i' - \tilde{x}_i^k)^T\big\{f_i(x_i^k) - A_i^T\tilde{\lambda}^k + A_i^T H\big(\sum_{j=1}^{m} A_j(x_j^k - \tilde{x}_j^k)\big)\big\} \\ &\geq \quad (x_i' - \tilde{x}_i^k)^T\big\{A_i^T H\big(\sum_{j=1}^{i-1} A_j(x_j^k - \tilde{x}_j^k)\big) + r_i(x_i^k - \tilde{x}_i^k)\big\} \geq 0, \ \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m. \quad (7.3) \end{aligned}$$

Note that (7.3) is obtained by substituting

$$f_i(\tilde{x}_i^k) \qquad \text{in the left hand side of (3.3) by} \qquad f_i(x_i^k)$$

and

$$A_i^T H\big(\sum_{j=1}^{i} A_j(x_j^k - \tilde{x}_j^k)\big) \quad \text{in the right hand side of (3.3) by} \quad A_i^T H\big(\sum_{j=1}^{i-1} A_j(x_j^k - \tilde{x}_j^k)\big).$$

If $w^k = \tilde{w}^k$, it follows from (7.3) and (7.1b) that

$$\left\{ \begin{array}{l} \tilde{x}_i^k \in \mathcal{X}_i, \quad (x_i' - \tilde{x}_i^k)^T\big\{f_i(\tilde{x}_i^k) - A_i^T\tilde{\lambda}^k\big\} \geq 0, \quad \forall x_i' \in \mathcal{X}_i, \ i = 1, \ldots, m, \\ \qquad\qquad \sum_{j=1}^{m} A_j\tilde{x}_j^k - b = 0. \end{array} \right.$$

Thus, $\tilde{w}^k$ generated by the linearized version C of the proximal alternating direction scheme (7.1) is a test vector for given w^k. In the following we will find $d_1^C(w,\tilde{w})$, $d_2^C(w,\tilde{w}) \in \Re^n$ and $\varphi^C(w,\tilde{w}) \in \Re$ which satisfy Conditions (2.11) of the framework in Subsection 2.3.

Lemma 7.1. *Let $\tilde{w}^k$ be generated by the linearized version C of the proximal AD scheme* (7.1) *from the given vector w^k. Then, we have*

$$\tilde{w}^k \in \mathcal{W}, \quad (w' - \tilde{w}^k)^T \big(d_2^C(w^k,\tilde{w}^k) - d_1^C(w^k,\tilde{w}^k) \big) \ge 0, \quad \forall\, w' \in \mathcal{W}, \tag{7.4}$$

where

$$d_1^C(w^k,\tilde{w}^k) = M(w^k - \tilde{w}^k) - \begin{pmatrix} (f_1(x_1^k) - f_1(\tilde{x}_1^k)) + A_1^T H A_1(x_1^k - \tilde{x}_1^k) \\ (f_2(x_2^k) - f_2(\tilde{x}_2^k)) + A_2^T H A_2(x_2^k - \tilde{x}_2^k) \\ \vdots \\ (f_m(x_m^k) - f_m(\tilde{x}_m^k)) + A_m^T H A_m(x_m^k - \tilde{x}_m^k) \\ 0 \end{pmatrix}, \tag{7.5}$$

and M is defined in (2.4)*, and*

$$d_2^C(w^k,\tilde{w}^k) = F(\tilde{w}^k) + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \tag{7.6}$$

is the same $d_2(w^k,\tilde{w}^k)$ defined in (3.6).

Proof. The proof is similar as those of Lemma 3.1. In comparison (3.3) and (7.3) we obtain $\tilde{w}^k \in \mathcal{W}$ and

$$\begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_m' - \tilde{x}_m^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} f_1(x_1^k) - A_1^T\tilde{\lambda}^k \\ f_2(x_2^k) - A_2^T\tilde{\lambda}^k \\ \vdots \\ f_m(x_m^k) - A_m^T\tilde{\lambda}^k \\ \sum_{i=1}^m A_i\tilde{x}_i^k - b \end{pmatrix} + \begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_m^T \\ 0 \end{pmatrix} H\big(\sum_{j=1}^m A_j(x_j^k - \tilde{x}_j^k)\big) \right\}$$
$$\ge \begin{pmatrix} x_1' - \tilde{x}_1^k \\ x_2' - \tilde{x}_2^k \\ \vdots \\ x_m' - \tilde{x}_m^k \\ \lambda' - \tilde{\lambda}^k \end{pmatrix}^T \left\{ \begin{pmatrix} A_1^T H\big(\sum_{j=1}^0 A_j(x_j^k - \tilde{x}_j^k)\big) \\ A_2^T H\big(\sum_{j=1}^1 A_j(x_j^k - \tilde{x}_j^k)\big) \\ \vdots \\ A_m^T H\big(\sum_{j=1}^{m-1} A_j(x_j^k - \tilde{x}_j^k)\big) \\ H^{-1}(\lambda^k - \tilde{\lambda}^k) \end{pmatrix} + \begin{pmatrix} r_1(x_1^k - \tilde{x}_1^k) \\ r_2(x_2^k - \tilde{x}_2^k) \\ \vdots \\ r_m(x_m^k - \tilde{x}_m^k) \\ 0 \end{pmatrix} \right\}, \tag{7.7}$$

for all $w' \in \mathcal{W}$. Note that (7.7) is obtained by substituting

$$f_i(\tilde{x}_i^k) \quad \text{in the left hand side of (3.8) by} \quad f_i(x_i^k)$$

and

$$A_i^T H\big(\textstyle\sum_{j=1}^i A_j(x_j^k - \tilde{x}_j^k)\big) \quad \text{in the right hand side of (3.8) by} \quad A_i^T H\big(\textstyle\sum_{j=1}^{i-1} A_j(x_j^k - \tilde{x}_j^k)\big).$$

Using the notations of $d_1^C(w^k,\tilde{w}^k)$ and $d_2^C(w^k,\tilde{w}^k)$ the above inequality is

$$\tilde{w}^k \in \mathcal{W}, \quad (w'-\tilde{w}^k)^T d_2^C(w^k,\tilde{w}^k) \geq (w'-\tilde{w}^k)^T d_1^C(w^k,\tilde{w}^k), \quad \forall\, w' \in \mathcal{W}$$

and the assertion of this lemma is proved. □

Note that the assertion of Lemma 7.1 is the condition (2.11a) of the framework in Subsection 2.3 for the problem $VI(\mathcal{W},F)$. Under the assumption that f_i is Lipschitz continuous, there is a constant $K>0$ such that $\|d_1^C(w^k,\tilde{w}^k)\| \leq K\|w^k-\tilde{w}^k\|$ (see (7.5)). Thus, the condition (2.11b) is also satisfied. Since the updating form of $\tilde{\lambda}^k$ in this section are the same as in Section 3 and definition of $d_2^C(w^k,\tilde{w}^k)=d_2(w^k,\tilde{w}^k)$, we have:

Lemma 7.2. *Let $\tilde{w}^k$ be generated by the linearized version C of the proximal AD scheme* (7.1) *from the given vector w^k. Then, we have*

$$(\tilde{w}^k-w^*)^T d_2^C(w^k,\tilde{w}^k) \geq (\lambda^k-\tilde{\lambda}^k)^T\big(\sum_{j=1}^m A_j(x_j^k-\tilde{x}_j^k)\big) \quad \forall\, w^* \in \mathcal{W}^*. \tag{7.8}$$

Moreover, by defining

$$\varphi^C(w^k,\tilde{w}^k) = (w^k-\tilde{w}^k)^T d_1^C(w^k,\tilde{w}^k) + (\lambda^k-\tilde{\lambda}^k)^T\big(\sum_{j=1}^m A_j(x_j^k-\tilde{x}_j^k)\big), \tag{7.9}$$

we have

$$(\tilde{w}^k-w^*)^T d_2^C(w^k,\tilde{w}^k) \geq \varphi^C(w^k,\tilde{w}^k) - (w^k-\tilde{w}^k)^T d_1^C(w^k,\tilde{w}^k) \quad \forall\, w^* \in \mathcal{W}^*, \tag{7.10}$$

$d_1^C(w^k,\tilde{w}^k)$ and $d_2^C(w^k,\tilde{w}^k)$ are defined in (7.5) and (7.6), respectively. Note that (7.10) is the condition (2.11c) of the framework and the remainder is to show the condition (2.11d).

Lemma 7.3. *Let $\tilde{w}^k=(\tilde{x}_1^k,\ldots,\tilde{x}_m^k,\tilde{\lambda}^k)$ be generated by the proximal alternating direction scheme* (7.1) *from the given vector $w^k=(x_1^k,\ldots,x_m^k,\lambda^k)$. Then, we have*

$$\begin{aligned}\varphi^C(w^k,\tilde{w}^k) \quad \geq \quad & \frac{1}{2}\big(\sum_{j=1}^m \|A_j(x_j^k-\tilde{x}_j^k)\|_H^2 + \|\lambda^k-\tilde{\lambda}^k\|_{H^{-1}}^2\big) + \sum_{j=1}^m r_i\|x_j^k-\tilde{x}_j^k\|^2 \\ & -\sum_{j=1}^m \Big((x_j^k-\tilde{x}_j^k)^T\big(f_j(x_j^k)-f_j(\tilde{x}_j^k)\big) + \|A_j(x_j^k-\tilde{x}_j^k)\|_H^2\Big),\end{aligned} \tag{7.11}$$

where $\varphi^C(w^k,\tilde{w}^k)$ is defined by (7.9).

Proof. The proof is similar as that in Lemma 3.3 and thus omitted. □

In order to guarantee $\varphi^C(w^k,\tilde{w}^k)$ to satisfy the condition (2.11d) of the framework in Subsection 2.3, we need the following requirement on r_i in (7.1).

Requirement on r_i in (7.1).

$$r_i\|x_i^k-\tilde{x}_i^k\|^2 \geq \nu\|x_i^k-\tilde{x}_i^k\|^2 + \frac{1}{2}\|A_i(x_i^k-\tilde{x}_i^k)\|_H^2 + (x_i^k-\tilde{x}_i^k)^T\big(f_i(x_i^k)-f_i(\tilde{x}_i^k)\big). \tag{7.12}$$

This requirement is satisfied when $r_i \geq \nu + \frac{1}{2}\|A_i^T H A_i\| + l_i$ (see (1.9)).

Lemma 7.4. *Let $\tilde{w}^k = (\tilde{x}_1^k, \ldots, \tilde{x}_m^k, \tilde{\lambda}^k)$ be generated by the proximal alternating direction scheme (7.1) from the given vector $w^k = (x_1^k, \ldots, x_m^k, \lambda^k)$. If the requirement (7.12) is satisfied, then*

$$\varphi^C(w^k, \tilde{w}^k) \geq \nu \sum_{j=1}^{m} \|x_j^k - \tilde{x}_j^k\|^2 + \frac{1}{2}\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2. \tag{7.13}$$

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

Proof. We get the assertion (7.13) from (7.11) and (7.12) directly. The reminder of the proof is the same as that in Lemma 3.4. □

For given w^k, the trial point $\tilde{w}^k$ produced by the linearized version C of the proximal alternating direction scheme (7.1) is a test vector of w^k. Under the requirement (7.12), the analysis in this section proves that $d_1^C(w^k, \tilde{w}^k)$, $d_2^C(w^k, \tilde{w}^k)$ (defined in (7.5) and (7.6), respectively) and $\varphi^C(w^k, \tilde{w}^k)$ (defined in (6.9)) satisfy conditions (2.11) of the framework in Subsection 2.3. Since $d_2^C(w^k, \tilde{w}^k) = d_2(w^k, \tilde{w}^k)$, substituting $d_1(w^k, \tilde{w}^k)$ and $\varphi(w^k, \tilde{w}^k)$ in the updating forms (4.3) by $d_1^C(w^k, \tilde{w}^k)$ and $\varphi^C(w^k, \tilde{w}^k)$, respectively, the resultant method is a contraction method and its convergence proofs are same as those in Section 4.

8 Conclusions

In order to simplify the minimization sub-problems in the first phase of the alternating direction-based contraction method [17], this paper presents some modified methods by adding the proximal terms to the sub-problems. Use the linearized versions, the subproblems become easy and the method is more practical. For the convergence property of the proposed methods, it has various restriction on the proximal parameters r_i $(i = 1, 2, ..., m)$ due to various version of the proximal alternating direction scheme. In summary, by defining

$$p_i(x_i^k, \tilde{x}_i^k) = \nu\|x_i^k - \tilde{x}_i^k\|^2 - \frac{1}{2}\|A_i(x_i^k - \tilde{x}_i^k)\|_H^2,$$

we list the requirement on $r_i, i = 1, \ldots, m$ in the following table:

Scheme	Requirement on r_i, $i = 1, \ldots, m$.
Proximal alternating direction scheme	$r_i\|x_i^k - \tilde{x}_i^k\|^2 \;\geq\; p_i(x_i^k, \tilde{x}_i^k).$
Linearized version A (Linearizing the quadratic function)	$r_i\|x_i^k - \tilde{x}_i^k\|^2 \;\geq\; p_i(x_i^k, \tilde{x}_i^k) + \|A_i(x_i^k - \tilde{x}_i^k)\|_H^2.$
Linearized version B (Linearizing the function θ_i)	$r_i\|x_i^k - \tilde{x}_i^k\|^2 \;\geq\; p_i(x_i^k, \tilde{x}_i^k) + (x_i^k - \tilde{x}_i^k)^T(f_i(x_i^k) - f_i(\tilde{x}_i^k)).$
Linearized version C (Linearizing both the function θ_i and the quadratic function)	$r_i\|x_i^k - \tilde{x}_i^k\|^2 \;\geq\; p_i(x_i^k, \tilde{x}_i^k) + \begin{pmatrix} (x_i^k - \tilde{x}_i^k)^T(f_i(x_i^k) - f_i(\tilde{x}_i^k)) \\ +\|A_i(x_i^k - \tilde{x}_i^k)\|_H^2. \end{pmatrix}.$

Indeed, by scaling a positive factor $1/r_i$, the first order optimal condition of (7.1a) (see (7.2)) can be rewritten as

$$(x_i' - \tilde{x}_i^k)^T \Big\{ \frac{1}{r_i}\Big(f_i(x_i^k) - A_i^T[\lambda^k - H(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b)]\Big) + (\tilde{x}_i^k - x_i^k)\Big\} \ge 0, \ \forall\, x_i' \in \mathcal{X}_i.$$

According to Lemma 2.2, this is equivalent to the projection equation

$$\tilde{x}_i^k = P_{\mathcal{X}_i}\Big\{ x_i^k - \frac{1}{r_i}\Big(f_i(x_i^k) - A_i^T\big[\lambda^k - H(\sum_{j=1}^{i-1} A_j\tilde{x}_j^k + \sum_{j=i}^{m} A_j x_j^k - b)\big]\Big)\Big\}.$$

Since $\tilde{x}_1^k, \ldots, \tilde{x}_{i-1}^k, x_i^k, \ldots, x_m^k, \lambda^k$ in the right hand side of the above equation are available, the solution of (7.1a) can be obtained by a projection on $\mathcal{X}_i$ under Euclidean-norm. Therefore, the linearized versions of the proximal alternating direction-based contraction methods substantially broaden the applicable scope of the alternating direction-based contraction method.

The convergence of the different versions of the proposed method follows from the general framework of the contraction methods [18]. In practice, the general framework guides us to construct the first order algorithms for solving the problems of separable linearly constrained convex optimization and simplifies the convergence proofs.

References

[1] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier methods*, Academic Press, 1982.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[3] E. Blum and W. Oettli, *Mathematische Optimierung, Econometrics and Operations Research XX*, Springer Verlag, 1975.

[4] G. Chen, and M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Mathematical Programming*, 64, pp. 81-101, 1994.

[5] J. Douglas and H. H. Rachford, On the numerical solution of the heat conduction problem in 2 and 3 space variables, *Transactions of the American Mathematical Society*, 82 (1956), 421-439.

[6] J. Eckstein, Some saddle-function splitting methods for convex programming, *Optimization Methods and Software*, 4, pp.75-83, 1994.

[7] J. Eckstein and M. Fukushima, Some reformulation and applications of the alternating direction method of multipliers, *Large Scale Optimization: State of the Art*, W. W. Hager *et al* eds., Kluwer Academic Publishers, pp. 115-134, 1994.

[8] E. Esser, Applications of Lagrangian-Based alternating direction methods and connections to split Bregman, UCLA CAM Report 09-31, 2009.

[9] F. Facchinei and J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity problems, Volume I, Springer Series in Operations Research, Springer-Verlag, 2003.

[10] M. Fukushima, Application of the alternating direction method of multipliers to separable convex programming problems, *Computational Optimization and Applications*, 2, pp. 93-111, 1992.

[11] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite-element approximations, *Computer and Mathematics with Applications*, 2, pp. 17-40, 1976.

[12] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.

[13] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM Studies in Applied Mathematics, Philadelphia, PA, 1989.

[14] B.S. He, A class of projection and contraction methods for monotone variational inequalities. Appl. Math. Optim. 35, 69–76 (1997).

[15] B. S. He, L.-Z. Liao, D. Han and H. Yang, A new inexact alternating directions method for monontone variational inequalities, *Mathematical Programming*, 92, pp. 103-118, 2002.

[16] B. S. He, L.-Z. Liao and M.-J. Qian, Alternating projection based prediction-correction methods for structured variational inequalities, *Journal of Computational Mathematics*, 24, pp. 693-710, 2006.

[17] B. S. He, M. Tao, M. H. Xu and X. M. Yuan, Alternating directions based contraction method for generally separable linearly constrained convex programming problems, 2009, http://www.optimization-online.org/DB-HTML/2009/11/2465.html

[18] B. S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, *Pacific J. Optimization*, 4, pp. 195-212, 2008.

[19] B. S. He, M. H. Xu and X. M. Yuan, Solving large-scale least squares covariance matrix problems by alternating direction methods, manuscript, 2009.

[20] B. S. He and H. Yang, Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities, *Operations Research Letters*, 23, pp. 151-161, 1998.

[21] B.S. He, X.M. Yuan and J.Z. Zhang, Comparison of two kinds of prediction-correction methods for monotone variational inequalities. Comput. Optim. Appl. 27, 247–267 (2004).

[22] S. Kontogiorgis and R. R. Meyer, A variable-penalty alternating directions method for convex optimization, *Mathematical Programming*, 83, pp. 29-53, 1998.

[23] B. Martinet, Regularization d'inequations variationelles par approximations sucessives, Revue Francaise d'Informatique et de RechercheOpérationelle 4 (1970), 154–159.

[24] M. Ng, P. A. Weiss and X. M. Yuan, Solving constrained total-variation problems via alternating direction methods, Manuscript, 2009, available on Optimization Online.

[25] J. Nocedal and S. J. Wright, *Numerical Optimization.* Second Edition, Spriger Verlag, 2006.

[26] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Con. Optim.*, 14, pp. 877-898, 1976.

[27] J. Sun and S. Zhang, A modified alternating direction method for convex quadratically constrained quadratic semidefinite programs, manuscript, 2009.

[28] M. Teboulle, Convergence of proximal-like algorithms, *SIAM J. Optim.*, 7, pp. 1069-1083, 1997.

[29] Z. Wen, D. Goldfarb and W. Yin, Alternating direction augmented Lagrangian methods for semidefinite programming, manuscript, 2009.

[30] J. Yang and Y. Zhang, Alternating direction algorithms for l_1-problems in compressive sensing, TR09-37, CAAM, Rice University.

[31] C. H. Ye and X. M. Yuan, A descent method for structured monotone variational inequalities, *Optimization Methods and Software*, 22, pp. 329-338, 2007.

[32] X. M. Yuan, Alternating direction methods for sparse covariance selection, manuscript, 2009, available on Optimizaton Online.

[33] X. M. Yuan and J. Yang, Sparse and low rank sparse matrix decomposition via alternating directions method, manuscript, 2009, available on Optimization Online.