Analysis in Theory and Applications
Anal. Theory Appl., Vol. **36**, No. 3 (2020), pp. 262-282

DOI: 10.4208/ata.OA-SU13

Study on the Splitting Methods for Separable Convex Optimization in a Unified Algorithmic Framework

Bingsheng He[1,2,*]

[1] *Department of Mathematics, Southern University of Science and Technology, Shenzhen, Guangdong 510085, China*
[2] *Department of Mathematics, Nanjing University, Nanjing, Jiangsu 210093, China*

Received 5 July 2020; Accepted (in revised version) 11 July 2020

Dedicated to Professor Weiyi Su on the occasion of her 80th birthday

Abstract. It is well recognized the convenience of converting the linearly constrained convex optimization problems to a monotone variational inequality. Recently, we have proposed a unified algorithmic framework which can guide us to construct the solution methods for solving these monotone variational inequalities. In this work, we revisit two full Jacobian decomposition of the augmented Lagrangian methods for separable convex programming which we have studied a few years ago. In particular, exploiting this framework, we are able to give a very clear and elementary proof of the convergence of these solution methods.

Key Words: Convex programming, augmented Lagrangian method, multi-block separable model, Jacobian splitting, unified algorithmic framework.

AMS Subject Classifications: 90C25, 90C30, 90C33

1 Introduction

In this paper, we consider the generic convex minimization model with linear constraints:

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & \sum\limits_{i=1}^{m} \theta_i(x_i) \\ \text{s.t.} & \sum\limits_{i=1}^{m} A_i x_i = b; \\ & x_i \in \mathcal{X}_i, \quad i = 1, \cdots, m, \end{array} \tag{1.1}$$

*Corresponding author. *Email address:* hebma@nju.edu.cn (B. S. He)

where $\theta_i : \Re^{n_i} \to \Re$ $(i = 1, \cdots, m)$ are closed proper convex functions and they are not necessarily smooth; $\mathcal{X}_i \subseteq \Re^{n_i}$ $(i = 1, \cdots, m)$ are closed convex sets; $A_i \in \Re^{l\times n_i}$ $(i = 1, \cdots, m)$ are given matrices; $b \in \Re^l$ is a given vector; and $\sum_{i=1}^m n_i = n$. The solution set of (1.1) is assumed to be nonempty throughout our discussion. The Lagrangian function of the problem (1.1) is

$$L(x_1, x_2, \cdots, x_m, \lambda) = \sum_{i=1}^m \theta_i(x_i) - \lambda^T \Big(\sum_{i=1}^m A_i x_i - b \Big), \tag{1.2}$$

in which $\lambda \in \Re^\ell$ is the Lagrange multiplier. By adding a penalty term to the Lagrangian function (1.2), we obtain its augmented Lagrangian function

$$\mathcal{L}_\beta(x_1, \cdots, x_m, \lambda) = \sum_{i=1}^m \theta_i(x_i) - \lambda^T \Big(\sum_{i=1}^m A_i x_i - b \Big) + \frac{\beta}{2} \Big\| \sum_{i=1}^m A_i x_i - b \Big\|^2, \tag{1.3}$$

where $\beta > 0$ is the penalty parameter for the linear constraints of (1.1). The augmented Lagrangian method (ALM) originally proposed in [11,13] for the problem (1.1) reads as

$$\begin{cases} (x_1^{k+1}, \cdots, x_m^{k+1}) = \arg\min\{\mathcal{L}_\beta(x_1, \cdots, x_m, \lambda^k) \mid x_i \in X_i,\ i = 1, \cdots, m\}, \\ \lambda^{k+1} = \lambda^k - \beta\Big(\sum_{i=1}^m A_i x_i^{k+1} - b \Big). \end{cases} \tag{1.4}$$

The ALM plays a significant role in both theoretical study and algorithmic design for various convex programming models. ALM scheme (1.4) is indeed an application of the well-known proximal point algorithm (PPA) that can date back to the seminal work [12, 14, 15] to the dual problem of (1.1). Throughout, we call $(x_1, \cdots, x_m)$ and λ the primal and dual variables, respectively.

It is well known that ADMM [3] is powerful for the problem (1.1) when $m = 2$. In order to use the separability of the problem, one considers to use the direct extension of ADMM [3] to solve (1.1) for $m \geq 3$. It leads to the following recursion:

1.1 The direct extension of ADMM

The k-th iteration begins with a given $(x_2^k, \cdots, x_m^k, \lambda^k)$, then

$$\begin{cases} x_1^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1, x_2^k, \cdots, x_m^k, \lambda^k) \mid x_1 \in \mathcal{X}_1\}, \\ x_2^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1^{k+1}, x_2, \cdots, x_m^k, \lambda^k) \mid x_2 \in \mathcal{X}_2\}, \\ \qquad \vdots \\ x_i^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1^{k+1}, \cdots, x_{i-1}^{k+1}, x_i, x_{i+1}^k, \cdots, x_m^k, \lambda^k) \mid x_i \in \mathcal{X}_i\}, \\ \qquad \vdots \\ x_m^{k+1} \in \arg\min\{\mathcal{L}_\beta(x_1^{k+1}, \cdots, x_{m-1}^{k+1}, x_m, \lambda^k) \mid x_m \in \mathcal{X}_m\}, \\ \lambda^{k+1} = \lambda^k - \beta\Big(\sum_{i=1}^m A_i x_i^{k+1} - b \Big). \end{cases} \tag{1.5}$$

Unfortunately, for $m \geq 3$, this direct extension of ADMM is not necessarily convergent, we can see a counterexample in [2].

1.2 Full Jacobian splitting prediction-correction method

It is interesting to consider decomposing the primal ALM subproblem in (1.4) by the Jacobian manner so that the resulting subproblems can be solved in parallel [6]. More precisely, applying the full Jacobian splitting to the primal subproblem in (1.4). The k-th iteration begins with given $(x_1^k, \cdots, x_m^k, \lambda^k)$, it produces the predictor by

$$\begin{cases} \begin{cases} x_1^{k+\frac{1}{2}} \in \arg\min\left\{\mathcal{L}_\beta(x_1, x_2^k, \cdots, x_m^k, \lambda^k) \mid x_1 \in \mathcal{X}_1\right\}, \\ \qquad\vdots \\ x_i^{k+\frac{1}{2}} \in \arg\min\left\{\mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^k, \cdots, x_m^k, \lambda^k) \mid x_i \in \mathcal{X}_i\right\}, \\ \qquad\vdots \\ x_m^{k+\frac{1}{2}} \in \arg\min\left\{\mathcal{L}_\beta(x_1^k, \cdots, x_{m-1}^k, x_m, \lambda^k) \mid x_m \in \mathcal{X}_m\right\}, \end{cases} & (1.6a) \\ \lambda^{k+\frac{1}{2}} = \lambda^k - \beta\Big(\sum_{i=1}^m A_i x_i^{k+\frac{1}{2}} - b\Big), & (1.6b) \end{cases}$$

and then, the new iterate is given by

$$\begin{pmatrix} x_1^{k+1} \\ \vdots \\ x_m^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x_1^{k} \\ \vdots \\ x_m^{k} \\ \lambda^{k} \end{pmatrix} - \alpha \begin{pmatrix} x_1^k - x_1^{k+\frac{1}{2}} \\ \vdots \\ x_m^k - x_m^{k+\frac{1}{2}} \\ \lambda^k - \lambda^{k+\frac{1}{2}} \end{pmatrix}. \tag{1.7}$$

We call (1.6)-(1.7) the full Jacobian splitting version of ALM [6] for the multi-block separable convex minimization (1.1) and the step size α in (1.7) will be discussed later. This method enjoys the feature that all the x_i-subproblems can be solved in parallel, and this is an important feature when large-scale data is under consideration and parallel computing infrastructures are available.

If we directly take the output of (1.6) as the new iterate (in other words, setting $\alpha = 1$ in (1.7)), the algorithm does not ensure the convergence. To construct a splitting method for solving the problem (1.1) without correction, it is necessary to add a regularization term $\frac{\tau\beta}{2}\|A_i(x_i - x_i^k)\|^2$ to the objective function of each primal subproblem. With given $(x_1^k, \cdots, x_m^k, \lambda^k)$, the k-th iteration generates the new iterate by the following scheme:

$$\left\{\begin{array}{l}\left\{\begin{array}{l} x_1^{k+1} \in \arg\min\Big\{\mathcal{L}_\beta(x_1, x_2^k, \cdots, x_m^k, \lambda^k) + \frac{\tau\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \;\Big|\; x_1 \in X_1\Big\}, \\ \qquad\qquad \vdots \\ x_i^{k+1} \in \arg\min\Big\{\mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^k, \cdots, x_m^k, \lambda^k) + \frac{\tau\beta}{2}\|A_i(x_i - x_i^k)\|^2 \;\Big|\; x_i \in X_i\Big\}, \\ \qquad\qquad \vdots \\ x_m^{k+1} \in \arg\min\Big\{\mathcal{L}_\beta(x_1^k, \cdots, x_{m-1}^k, x_m, \lambda^k) + \frac{\tau\beta}{2}\|A_m(x_m - x_m^k)\|^2 \;\Big|\; x_m \in X_m\Big\}, \end{array}\right. \quad (1.8a) \\ \lambda^{k+1} = \lambda^k - \beta\Big(\sum_{i=1}^m A_i x_i^{k+1} - b\Big), \quad (1.8b)\end{array}\right.$$

where $\tau > m-1$ is a constant. The method is a variant version of the one published in [7].

On the other hand, in the past few years, developing from [4, 5], we published a unified algorithmic framework for convex optimization [16] in the frame of variational inequality. Besides simplifying the convergence analysis, the framework provides us guidance for constructing solution methods [4]. In this paper, we investigates the full splitting methods (1.6)-(1.7) and (1.8) in such a framework.

The rest of this paper is organized as follows. We recall some preliminaries in Section 2. Section 3 focuses on the algorithmic framework for the variational inequality which arising from the problem (1.1). In Sections 4 and 5, by defining suitable prediction and correction steps, we interpret the method (1.6)-(1.7) and (1.8), respectively, in the unified algorithmic framework. Finally, we make some conclusions in Section 6.

2 Variational inequality characterization and preliminaries

2.1 Variational inequality characterization of (2.1)

In order to simply describing the problem, we give some compact notations. By denoting

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad \boldsymbol{\theta}(\boldsymbol{x}) = \sum_{i=1}^m \theta_i(x_i), \quad \mathcal{A} = (A_1, A_2, \cdots, A_m),$$

and

$$\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_m,$$

the problem (1.1) thus enjoys a compact form

$$\min\{\boldsymbol{\theta}(\boldsymbol{x}) \mid \mathcal{A}\boldsymbol{x} = b,\ \boldsymbol{x} \in \mathcal{X}\}. \tag{2.1}$$

Its corresponding Lagrangian function and augmented Lagrangian function can be written as

$$L(\boldsymbol{x},\lambda) = \boldsymbol{\theta}(\boldsymbol{x}) - \lambda^T(\mathcal{A}\boldsymbol{x} - b), \tag{2.2a}$$

$$\mathcal{L}_\beta(\boldsymbol{x},\lambda) = \boldsymbol{\theta}(\boldsymbol{x}) - \lambda^T(\mathcal{A}\boldsymbol{x} - b) + \frac{\beta}{2}\|\mathcal{A}\boldsymbol{x} - b\|^2, \tag{2.2b}$$

respectively. A pair $(\boldsymbol{x}^*,\lambda^*)$ defined on $\mathcal{X} \times \Re^\ell$ is called a saddle point of the Lagrangian function (1.2) if it satisfies the inequalities

$$L_{\lambda\in\Re^\ell}(\boldsymbol{x}^*,\lambda) \le L(\boldsymbol{x}^*,\lambda^*) \le L_{\boldsymbol{x}\in\mathcal{X}}(\boldsymbol{x},\lambda^*).$$

Alternatively, we can rewrite these inequalities as the variational inequalities:

$$\begin{cases} \boldsymbol{x}^* \in \mathcal{X}, & \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\boldsymbol{x}^*) + (\boldsymbol{x} - \boldsymbol{x}^*)^T(-\mathcal{A}^T\lambda^*) \ge 0, \quad \forall \boldsymbol{x} \in \mathcal{X}, \\ \lambda^* \in \Re^\ell, & (\lambda - \lambda^*)^T(\mathcal{A}\boldsymbol{x}^* - b) \ge 0, \quad \forall \lambda \in \Re^\ell, \end{cases} \tag{2.3}$$

or in a compact form VI(Ω, F):

$$\boldsymbol{w}^* \in \boldsymbol{\Omega}, \quad \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\boldsymbol{x}^*) + (\boldsymbol{w} - \boldsymbol{w}^*)^T\boldsymbol{F}(\boldsymbol{w}^*) \ge 0, \quad \forall \boldsymbol{w} \in \boldsymbol{\Omega}, \tag{2.4a}$$

where

$$\boldsymbol{w} = \begin{pmatrix} \boldsymbol{x} \\ \lambda \end{pmatrix}, \quad \boldsymbol{F}(\boldsymbol{w}) = \begin{pmatrix} -\mathcal{A}^T\lambda \\ \mathcal{A}\boldsymbol{x} - b \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Omega} = \mathcal{X} \times \Re^\ell. \tag{2.4b}$$

We denote the variational inequality (2.4) by VI$(\boldsymbol{\Omega}, \boldsymbol{F}, \boldsymbol{\theta})$. Note that for the operator $\boldsymbol{F}$ defined in (2.4b) is affine with a skew-symmetric matrix, and thus we have

$$(\boldsymbol{w} - \tilde{\boldsymbol{w}})^T(\boldsymbol{F}(\boldsymbol{w}) - \boldsymbol{F}(\tilde{\boldsymbol{w}})) \equiv 0. \tag{2.5}$$

The function $\boldsymbol{\theta}$ is convex and the operator $\boldsymbol{F}$ is still monotone. We denote the solution set of the variational inequality (2.4) by $\boldsymbol{\Omega}^*$.

2.2 A basic lemma

The following lemma is basic and will be frequently used in our analysis. Its proof is elementary and thus omitted.

Lemma 2.1. *Let $\mathcal{Z} \subset \Re^n$ be a closed convex set, $\theta(z)$ and $f(z)$ be convex functions. If f is differentiable on a open set which contains $\mathcal{Z}$, and the solution set of the minimization problem* $\min\{\theta(z) + f(z) \,|\, z \in \mathcal{Z}\}$ *is nonempty, then*

$$z^* \in \arg\min\{\theta(z) + f(z) \,|\, z \in \mathcal{Z}\} \tag{2.6a}$$

if and only if

$$z^* \in \mathcal{Z}, \quad \theta(z) - \theta(z^*) + (z - z^*)^T\nabla f(z^*) \ge 0, \quad \forall z \in \mathcal{Z}. \tag{2.6b}$$

2.3 A unified algorithmic framework for variational inequalities

In the review paper [16], the optimal condition of the linearly constrained convex optimization is described in a variational inequality:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^T F(w^*) \geq 0, \quad \forall w \in \Omega. \tag{2.7}$$

For solving the variational inequality (2.7), we have proposed an algorithmic framework which consists of a prediction and a correction.

The Algorithmic framework:
The prediction step begins with a given v^k, finds a vector $\tilde{w}^k \in \Omega$ which satisfying

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^T F(\tilde{w}^k) \geq (v - \tilde{v}^k)^T Q(v^k - \tilde{v}^k), \quad \forall w \in \Omega, \tag{2.8a}$$

where Q is not necessarily symmetric, but $Q^T + Q$ is positive definite.
The correction step updates the new iterate by

$$v^{k+1} = v^k - \alpha M(v^k - \tilde{v}^k). \tag{2.8b}$$

The vector u generally is the vector $\boldsymbol{x}$ defined in (2.4), the vector v can be a part of elements of the vector w. If the matrix Q itself is a symmetric positive definite matrix, by setting $M = I$ and $\alpha = 1$, it leads a proximal point algorithm and widely applied in imaging science [1,9,10].

Convergence conditions of the algorithm (2.8) :
For the matrices Q and M, and the step size α given in (2.8), the matrices

$$H = QM^{-1}, \tag{2.9a}$$
$$G = Q^T + Q - \alpha M^T H M, \tag{2.9b}$$

are positive definite.

For more details of this framework, the reader is referred to [8,16].

3 Unified algorithmic framework for variational inequality (2.4)

In this section, we interpret the unified algorithmic framework (2.8)-(2.9) to solve the problem VI$(\boldsymbol{\Omega}, \boldsymbol{F}, \boldsymbol{\theta})$ (2.4). Because we introduce an additional algorithm which uses the calculated step size in its correction step, we rewrite the framework in the following wise.

Prediction of the algorithms for solving the variational inequality (2.4):
With a given $\boldsymbol{w}^k$, find a vector $\tilde{\boldsymbol{w}}^k$, which satisfying

$$\begin{aligned} \tilde{\boldsymbol{w}}^k \in \Omega, \quad \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k) + (\boldsymbol{w} - \tilde{\boldsymbol{w}}^k)^T \boldsymbol{F}(\tilde{\boldsymbol{w}}^k) \\ \geq (\boldsymbol{w} - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k), \quad \forall \boldsymbol{w} \in \boldsymbol{\Omega}, \end{aligned} \tag{3.1}$$

where the matrix Q is not necessarily symmetric, but

$$Q^T + Q \text{ must be positive definite when each } A_i \text{ is full column rank matrix.} \tag{3.2}$$

The two classes of algorithms use the same predictor $\tilde{\boldsymbol{w}}^k$ provided by the prediction (3.1) whose matrix Q satisfies the condition (3.2).

3.1 Algorithm I using the constant correction step

The Algorithm I utilizes the predictor offered by (3.1). Its correction step updates the new iterate by the following formula which uses a constant step size.

Correction of the Algorithm I for solving the variational inequality (2.4):
The new iterate $\boldsymbol{w}^{k+1}$ is updated by

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \alpha M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k), \tag{3.3}$$

where the matrix M and the step-size α satisfy the following conditions:

1. there is a positive matrix H such that

$$HM = Q. \tag{3.4a}$$

2. the matrix G, defined by

$$G = Q^T + Q - \alpha M^T H M \quad \text{is positive definite,} \tag{3.4b}$$

when each A_i is full column rank matrix.

In principal, Algorithm I belongs to the unified algorithmic framework (2.8)-(2.9). For completeness, we show the key-inequality for the convergence in the following theorem.

Theorem 3.1. *For solving the variational inequality* (2.4), *Algorithm I uses* (3.1) *and* (3.3) *to produce the predictor and corrector, respectively. If the conditions* (3.2) *and* (3.4) *are satisfied, then the generated sequence* $\{\boldsymbol{w}^k\}$ *satisfies*

$$\|\boldsymbol{w}^{k+1} - \boldsymbol{w}^*\|_H^2 \leq \|\boldsymbol{w}^k - \boldsymbol{w}^*\|_H^2 - \alpha \|\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k\|_G^2, \quad \forall \boldsymbol{w}^* \in \boldsymbol{\Omega}^*. \tag{3.5}$$

Proof. Using $Q=HM$ (see (3.4a)) and the correction update form (3.3), it follows from (3.1) that

$$\begin{aligned}\tilde{\boldsymbol{w}}^k\in\boldsymbol{\Omega},\quad &2\alpha\big\{\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k)+(\boldsymbol{w}-\tilde{\boldsymbol{w}}^k)^T\boldsymbol{F}(\tilde{\boldsymbol{w}}^k)\big\}\\ &\geq 2(\boldsymbol{w}-\tilde{\boldsymbol{w}}^k)^TH(\boldsymbol{w}^k-\boldsymbol{w}^{k+1}),\quad \forall\boldsymbol{w}\in\boldsymbol{\Omega}.\end{aligned}\tag{3.6}$$

Applying the identity

$$2(a-b)^TH(c-d)=\big(\|a-d\|_H^2-\|a-c\|_H^2\big)+\big(\|c-b\|_H^2-\|d-b\|_H^2\big),$$

to the right-hand side of (3.6) with

$$a=\boldsymbol{w},\quad b=\tilde{\boldsymbol{w}}^k,\quad c=\boldsymbol{w}^k\quad\text{and}\quad d=\boldsymbol{w}^{k+1},$$

we obtain

$$\begin{aligned}&2(\boldsymbol{w}-\tilde{\boldsymbol{w}}^k)^TH(\boldsymbol{w}^k-\boldsymbol{w}^{k+1})\\ &\quad=\big(\|\boldsymbol{w}-\boldsymbol{w}^{k+1}\|_H^2-\|\boldsymbol{w}-\boldsymbol{w}^k\|_H^2\big)+\big(\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_H^2-\|\boldsymbol{w}^{k+1}-\tilde{\boldsymbol{w}}^k\|_H^2\big).\end{aligned}\tag{3.7}$$

For the last term on the right hand side of (3.7), we have

$$\begin{aligned}&\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_H^2-\|\boldsymbol{w}^{k+1}-\tilde{\boldsymbol{w}}^k\|_H^2\\ &\quad=\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_H^2-\|(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)-(\boldsymbol{w}^k-\boldsymbol{w}^{k+1})\|_H^2\\ &\quad\overset{(3.3)}{=}\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_H^2-\|(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)-\alpha M(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)\|_H^2\\ &\quad=2\alpha(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)^THM(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)-\alpha^2(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)^TM^THM(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)\\ &\quad\overset{(3.4a)}{=}\alpha(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)^T(Q^T+Q-\alpha M^THM)(\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k)\\ &\quad\overset{(3.4b)}{=}\alpha\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_G^2.\end{aligned}\tag{3.8}$$

Substituting (3.7) and (3.8) in (3.6), we get

$$\begin{aligned}&2\alpha\big\{\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k)+(\boldsymbol{w}-\tilde{\boldsymbol{w}}^k)^TF(\tilde{\boldsymbol{w}}^k)\big\}\\ &\qquad\geq\big(\|\boldsymbol{w}-\boldsymbol{w}^{k+1}\|_H^2-\|\boldsymbol{w}-\boldsymbol{w}^k\|_H^2\big)+\alpha\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_G^2,\quad\forall\boldsymbol{w}\in\boldsymbol{\Omega}.\end{aligned}$$

Taking $\boldsymbol{w}=\boldsymbol{w}^*$ in the above inequality, it follows that

$$\begin{aligned}&\|\boldsymbol{w}^k-\boldsymbol{w}^*\|_H^2-\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\|_H^2\\ &\qquad\geq\alpha\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_G^2+2\alpha\big\{\boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k)-\boldsymbol{\theta}(\boldsymbol{x}^*)+(\tilde{\boldsymbol{w}}^k-\boldsymbol{w}^*)^T\boldsymbol{F}(\tilde{\boldsymbol{w}}^k)\big\}.\end{aligned}\tag{3.9}$$

By using $(\tilde{\boldsymbol{w}}^k-\boldsymbol{w}^*)^T\boldsymbol{F}(\tilde{\boldsymbol{w}}^k)=(\tilde{\boldsymbol{w}}^k-\boldsymbol{w}^*)^T\boldsymbol{F}(\boldsymbol{w}^*)$ and the optimality of $\boldsymbol{w}^*$, we have

$$\boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k)-\boldsymbol{\theta}(\boldsymbol{x}^*)+(\tilde{\boldsymbol{w}}^k-\boldsymbol{w}^*)^T\boldsymbol{F}(\tilde{\boldsymbol{w}}^k)=\boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k)-\boldsymbol{\theta}(\boldsymbol{x}^*)+(\tilde{\boldsymbol{w}}^k-\boldsymbol{w}^*)^T\boldsymbol{F}(\boldsymbol{w}^*)\geq 0.$$

Finally, it follows from (3.9) that

$$\|\boldsymbol{w}^k-\boldsymbol{w}^*\|_H^2-\|\boldsymbol{w}^{k+1}-\boldsymbol{w}^*\|_H^2\geq\alpha\|\boldsymbol{w}^k-\tilde{\boldsymbol{w}}^k\|_G^2$$

and the assertion (3.5) of this theorem is proved. □

3.2 Algorithm II using the calculated correction step

The Algorithm II utilizes the predictor offered by (3.1). Its correction step updates the new iterate by the following formula which uses a calculated step size.

Correction of the Algorithm II for solving the variational inequality (2.4):
The new iterate $\boldsymbol{w}^{k+1}$ is updated by

$$\boldsymbol{w}^{k+1} = \boldsymbol{w}^k - \alpha_k M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k), \tag{3.10}$$

where the matrix M and the step-size α_k satisfy the following conditions:

1. there is a positive matrix H such that

$$HM = Q. \tag{3.11a}$$

2. the step-size α_k given by

$$\alpha_k = \gamma\alpha_k^*, \qquad \alpha_k^* = \frac{(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)}{\|M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)\|_H^2}, \qquad \gamma \in (0,2). \tag{3.11b}$$

Similarly as for the Algorithm I, we have the following theorem which contains the key-inequality for the convergence of the Algorithm II.

Theorem 3.2. *For solving the variational inequality* (2.4)*, the Algorithm II uses* (3.1) *and* (3.10) *to produce the predictor and to update the new iterate, respectively. If the conditions* (3.2) *and* (3.11) *are satisfied, then the generated sequence* $\{\boldsymbol{w}^k\}$ *satisfies*

$$\|\boldsymbol{w}^{k+1} - \boldsymbol{w}^*\|_H^2 \le \|\boldsymbol{w}^k - \boldsymbol{w}^*\|_H^2 - \frac{\gamma(2-\gamma)}{2}\alpha_k^*\|\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k\|_{(Q^T+Q)}^2, \quad \forall \boldsymbol{w}^* \in \boldsymbol{\Omega}^*. \tag{3.12}$$

Proof. Setting an arbitrarily but fixed $\boldsymbol{w}^* \in \Omega^*$ in (3.1), it follows that

$$(\tilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k) \ge \boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k) - \boldsymbol{\theta}(\boldsymbol{x}^*) + (\tilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^T \boldsymbol{F}(\tilde{\boldsymbol{w}}^k), \quad \forall \boldsymbol{w}^* \in \boldsymbol{\Omega}^*.$$

By using $(\tilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^T \boldsymbol{F}(\tilde{\boldsymbol{w}}^k) = (\tilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^T \boldsymbol{F}(\boldsymbol{w}^*)$ and the optimality of $\boldsymbol{w}^*$, we have

$$(\tilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^T F(\tilde{\boldsymbol{w}}^k) + \boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k) - \boldsymbol{\theta}(\boldsymbol{x}^*) \ge (\tilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^T F(\boldsymbol{w}^*) + \boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k) - \boldsymbol{\theta}(\boldsymbol{x}^*) \ge 0,$$

and thus

$$(\tilde{\boldsymbol{w}}^k - \boldsymbol{w}^*)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k) \ge 0.$$

Using $Q = HM$, from the above inequality follows that

$$(\boldsymbol{w}^k - \boldsymbol{w}^*)^T HM(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k) \ge (\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k), \quad \forall \boldsymbol{w}^* \in \boldsymbol{\Omega}^*. \tag{3.13}$$

The right hand side of the above inequality equals $\frac{1}{2}\|\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k\|^2_{(Q^T+Q)}$ and thus is positive whenever $\boldsymbol{w}^k \neq \tilde{\boldsymbol{w}}^k$. For any step-size dependent new iterate

$$\boldsymbol{w}^{k+1}(\alpha) = \boldsymbol{w}^k - \alpha M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k), \tag{3.14}$$

we investigate the profit the k-th iteration, namely,

$$\vartheta(\alpha) = \|\boldsymbol{w}^k - \boldsymbol{w}^*\|^2_H - \|\boldsymbol{w}^{k+1}(\alpha) - \boldsymbol{w}^*\|^2_H. \tag{3.15}$$

By a manipulation, we get

$$\begin{aligned}\vartheta(\alpha) &= \|\boldsymbol{w}^k - \boldsymbol{w}^*\|^2_H - \|\boldsymbol{w}^{k+1}(\alpha) - \boldsymbol{w}^*\|^2_H \\ &= \|\boldsymbol{w}^k - \boldsymbol{w}^*\|^2_H - \|(\boldsymbol{w}^k - \boldsymbol{w}^*) - \alpha M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)\|^2_H \\ &= 2\alpha(\boldsymbol{w}^k - \boldsymbol{w}^*)^T HM(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k) - \alpha^2\|M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)\|^2_H \\ &\geq 2\alpha(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k) - \alpha^2\|M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)\|^2_H \\ &=: q(\alpha).\end{aligned} \tag{3.16}$$

Notice that $q(\alpha)$ is a quadratic function of α which reaches its maximum at

$$\alpha_k^* = \frac{(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)}{\|M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)\|^2_H},$$

which is just the same one defined in (3.4b). By using $\alpha_k^*\|M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)\|^2_H = (\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)$, we get

$$\begin{aligned}\vartheta(\alpha_k) &\geq q(\alpha_k) = q(\gamma\alpha_k^*) \\ &= 2\gamma\alpha_k^*(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k) - \gamma^2\alpha_k^*\big(\alpha_k^*\|M(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)\|^2_H\big) \\ &= \gamma(2-\gamma)\alpha_k^*(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k).\end{aligned} \tag{3.17}$$

Thus

$$\|\boldsymbol{w}^k - \boldsymbol{w}^*\|^2_H - \|\boldsymbol{w}^{k+1} - \boldsymbol{w}^*\|^2_H \geq \frac{\gamma(2-\gamma)}{2}\alpha_k^*\|\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k\|^2_{(Q^T+Q)}$$

and the assertion (3.12) is proved. □

For given matrices Q, M, H, the step size α_k in (3.11b) is low bounded away from zero. Thus, (3.12) is the key-inequality for the convergence of Algorithm II. Even though the correction of the Algorithm II needs to calculate the step size, usually, it converges faster than Algorithm I.

4 Splitting method (1.6)-(1.7) in the unified framework

Now, we are in the stage to prove the convergence of the considered splitting methods in Section 1 by using the unified algorithmic framework in Section 3. This section deals with the splitting method (1.6)-(1.7). We will interpret the method in a suitable prediction-correction algorithm and check the conditions (3.2) and (3.4).

Prediction: For given $\boldsymbol{w}^k = (\boldsymbol{x}^k, \lambda^k)$, we define the predictor by

$$\begin{cases} \begin{cases} \tilde{x}_1^k \in \arg\min\big\{\mathcal{L}_\beta(x_1, x_2^k, \cdots, x_m^k, \lambda^k) \,\big|\, x_1 \in X_1\big\}, \\ \qquad\vdots \\ \tilde{x}_i^k \in \arg\min\big\{\mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^k, \cdots, x_m^k, \lambda^k) \,\big|\, x_i \in X_i\big\}, \\ \qquad\vdots \\ \tilde{x}_m^k \in \arg\min\big\{\mathcal{L}_\beta(x_1^k, \cdots, x_{m-1}^k, x_m, \lambda^k) \,\big|\, x_m \in X_m\big\}, \end{cases} & (4.1a) \\ \tilde{\lambda}^k = \lambda^k - \beta\Big(\sum_{i=1}^m A_i \tilde{x}_i^k - b\Big) + 2\beta\Big(\sum_{i=1}^m A_i(\tilde{x}_i^k - x_i^k)\Big). & (4.1b) \end{cases}$$

Lemma 4.1. *For solving the variational inequality* (2.4)*, the predictor* $\tilde{\boldsymbol{w}}^k = (\tilde{\boldsymbol{x}}^k, \tilde{\lambda}^k)$ *offered by* (4.1) *satisfies*

$$\tilde{\boldsymbol{w}}^k \in \Omega, \quad \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k) + (\boldsymbol{w} - \tilde{\boldsymbol{w}}^k)^T \boldsymbol{F}(\tilde{\boldsymbol{w}}^k) \geq (\boldsymbol{w} - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k), \quad \forall \boldsymbol{w} \in \boldsymbol{\Omega}, \tag{4.2}$$

where

$$Q = \begin{pmatrix} \beta\mathcal{A}^T\mathcal{A} + \beta\mathcal{D} & 0 \\ -2\mathcal{A} & \dfrac{1}{\beta} I \end{pmatrix} \tag{4.3}$$

and $\mathcal{D} = diag(A_1^T A_1, A_2^T A_2, \cdots, A_m^T A_m)$.

Proof. In the prediction step (4.1a), $\tilde{x}_i^k$ is defined by

$$\tilde{x}_i^k \in \arg\min\big\{\mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^k, \cdots, x_m^k, \lambda^k) \,\big|\, x_i \in X_i\big\}.$$

Ignoring the constant term of the objective function (see (1.3)), it leads that

$$\tilde{x}_i^k \in \arg\min\bigg\{\theta_i(x_i) - x_i^T A_i^T \lambda^k + \frac{\beta}{2}\|A_i(x_i - x_i^k) + (\mathcal{A}\boldsymbol{x}^k - b)\|^2 \,\bigg|\, x_i \in X_i\bigg\}.$$

According to Lemma 2.1, using the optimal condition of the above problem, we get

$$\begin{aligned} \tilde{x}_i^k \in \mathcal{X}_i, \quad & \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T \\ & \big\{-A_i^T \lambda^k + \beta A_i^T\big[A_i(\tilde{x}_i^k - x_i^k) + (\mathcal{A}\boldsymbol{x}^k - b)\big]\big\} \geq 0, \quad \forall x_i \in \mathcal{X}_i. \end{aligned} \tag{4.4}$$

Using (4.1b), we have

$$\tilde{\lambda}^k = \lambda^k - \beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) + 2\beta\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k), \tag{4.5}$$

and thus

$$\begin{aligned}\lambda^k =&\tilde{\lambda}^k + \beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) - 2\beta\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k)\\ =&\tilde{\lambda}^k + \beta(\mathcal{A}\boldsymbol{x}^k - b) - \beta\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k).\end{aligned}$$

Substituting it in (4.4), we obtain

$$\begin{aligned}&\tilde{x}_i^k \in \mathcal{X}_i, \quad \theta_i(x_i) - \theta_i(\tilde{x}_i^k) + (x_i - \tilde{x}_i^k)^T\\ &\qquad \left\{-A_i^T\tilde{\lambda}^k + \beta A_i^T\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k) + \beta A_i^T A_i(\tilde{x}_i^k - x_i^k)\right\} \geq 0, \quad \forall x_i \in \mathcal{X}_i.\end{aligned} \tag{4.6}$$

Notice that

$$\mathcal{A}^T\mathcal{A} = \begin{pmatrix} A_1^T\mathcal{A} \\ A_2^T\mathcal{A} \\ \vdots \\ A_m^T\mathcal{A}\end{pmatrix} \quad \text{and} \quad \mathcal{D} = \begin{pmatrix} A_1^TA_1 & 0 & \cdots & 0 \\ 0 & A_2^TA_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_m^TA_m \end{pmatrix}.$$

Summing the inequality (4.6) from $i = 1$ to $i = m$, we get

$$\begin{aligned}&\tilde{\boldsymbol{x}}^k \in \boldsymbol{\mathcal{X}}, \quad \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k) + (\boldsymbol{x} - \tilde{\boldsymbol{x}}^k)^T\\ &\qquad \left\{-\mathcal{A}^T\tilde{\lambda}^k + \beta\mathcal{A}^T\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k) + \beta\mathcal{D}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k)]\right\} \geq 0, \quad \forall \boldsymbol{x} \in \boldsymbol{\mathcal{X}}.\end{aligned} \tag{4.7}$$

Notice that (4.5) can be written as

$$(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) - 2\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) = 0$$

and thus we have

$$\tilde{\lambda}^k \in \Re^l, \quad (\lambda - \tilde{\lambda}^k)^T\left\{(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) - 2\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k)\right\} \geq 0, \quad \forall \lambda \in \Re^l. \tag{4.8}$$

Combining (4.7) and (4.8) and using the notation $\boldsymbol{F}(\tilde{\boldsymbol{w}}^k)$, the assertion of this lemma follows immediately. □

It is clear that the matrix Q satisfies the condition (3.2) in the prediction step (3.1). Now, we give the related correction form.

Lemma 4.2. *Let $\mathcal{I} = diag(I_{n_1}, I_{n_2}, \cdots, I_{n_m})$. For given $\boldsymbol{w}^k = (\boldsymbol{x}^k, \lambda^k)$, if the predictor $\tilde{\boldsymbol{w}}^k = (\tilde{\boldsymbol{x}}^k, \tilde{\lambda}^k)$ is generated by* (4.1)*, then the correction*

$$\begin{pmatrix} \boldsymbol{x}^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^k \\ \lambda^k \end{pmatrix} - \alpha \begin{pmatrix} \mathcal{I} & 0 \\ -2\beta\mathcal{A} & I \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \tag{4.9}$$

produces the same new iterate $\boldsymbol{w}^{k+1}$ as the same one by (1.7).

Proof. For the same $\boldsymbol{w}^k = (\boldsymbol{x}^k, \lambda^k)$, comparing (1.6a) and (4.1a), we have

$$\tilde{x}_i^k = x_i^{k+\frac{1}{2}}, \quad i = 1, \cdots, m.$$

Thus, the $\boldsymbol{x}$-part of the correction (1.7) can be written as

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha(\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k). \tag{4.10}$$

By using $\boldsymbol{x}^{k+\frac{1}{2}} = \tilde{\boldsymbol{x}}^k$, it follows from (1.6b) that

$$\lambda^k - \lambda^{k+\frac{1}{2}} = \beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b).$$

On the other hand, from (4.1b) we have

$$\beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) = \lambda^k - \tilde{\lambda}^k - 2\beta\mathcal{A}(\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k).$$

From the above two equations we get

$$\lambda^k - \lambda^{k+\frac{1}{2}} = \lambda^k - \tilde{\lambda}^k - 2\beta\mathcal{A}(\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k).$$

Substituting it in (1.7), the λ-part correction can be written as

$$\begin{aligned} \lambda^{k+1} =& \lambda^k - \alpha(\lambda^k - \lambda^{k+\frac{1}{2}}) \\ =& \lambda^k - \alpha[-2\beta\mathcal{A}(\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k) + (\lambda^k - \tilde{\lambda}^k)]. \end{aligned} \tag{4.11}$$

Combining (4.10) and (4.11), we get (4.9), and the assertion in proved. □

The splitting method (1.6)-(1.7) has been interpreted in the Algorithm I whose prediction and correction steps are (4.1) and (4.9), respectively. Note that in the correction step,

$$M = \begin{pmatrix} \mathcal{I} & 0 \\ -2\beta\mathcal{A} & I \end{pmatrix}. \tag{4.12}$$

Indeed, we have the positive definite matrix

$$H = \begin{pmatrix} \beta(\mathcal{A}^T\mathcal{A} + \mathcal{D}) & 0 \\ 0 & \dfrac{1}{\beta}I \end{pmatrix}, \tag{4.13}$$

which satisfies the condition (3.4a), namely, $HM = Q$ (also see the matrix Q in (4.3)).

The remained question is to give the range of the constant α in (3.4b). Since $Q^T + Q$ is positive definite, there is a positive scalar $\alpha_{\max}$ which satisfies

$$\alpha_{\max} = \arg\max_{\alpha}\{G = Q^T + Q - \alpha M^T H M \,|\, G \succeq 0\}.$$

We list the result in the following lemma.

Lemma 4.3. *For the given matrices Q, H and M in* (4.3), (4.12) *and* (4.13), *respectively, the matrix*

$$G=(Q^T+Q)-\alpha M^T HM \text{ is positive definite for any } \alpha\in\Big(0,2\big(1-\sqrt{\tfrac{m}{m+1}}\,\big)\Big). \tag{4.14}$$

Proof. Using $Q=HM$, we write the matrix G in the following form

$$\begin{aligned}(Q^T+Q)-\alpha M^T HM =&M^T(HM^{-1}+M^{-T}H)M-\alpha M^T HM\\ =&M^T[(HM^{-1}+M^{-T}H)-\alpha H]M.\end{aligned}$$

In order to show the assertion (4.14), we need only to find the largest α such that

$$\text{Find the largest } \alpha \text{ such } \ (HM^{-1}+M^{-T}H)-\alpha H\succeq 0. \tag{4.15}$$

Using the expressions of M and H (see (4.12) and (4.13)), we have

$$M^{-1}=\begin{pmatrix}\mathcal{I} & 0\\ 2\beta\mathcal{A} & I\end{pmatrix}$$

and thus

$$HM^{-1}=\begin{pmatrix}\beta(\mathcal{A}^T\mathcal{A}+\mathcal{D}) & 0\\ 2\mathcal{A} & \dfrac{1}{\beta}I\end{pmatrix}.$$

Notice that

$$HM^{-1}+M^{-T}H=2\begin{pmatrix}\beta(\mathcal{A}^T\mathcal{A}+\mathcal{D}) & \mathcal{A}^T\\ \mathcal{A} & \dfrac{1}{\beta}I\end{pmatrix}. \tag{4.16}$$

For discussing the positivity of the matrix $[(HM^{-1}+M^{-T}H)-\alpha H]$ in (4.15), we define

$$\mathcal{P}=\mathrm{diag}\Big(\sqrt{\beta}A_1,\cdots,\sqrt{\beta}A_m,\frac{1}{\sqrt{\beta}}I\Big),$$

then we have (see (4.16) and (4.13))

$$\begin{aligned}&(HM^{-1}+M^{-T}H)-\alpha H\\ =&\mathcal{P}^T\left\{2\begin{pmatrix}2I_l & I_l & \cdots & I_l & I_l\\ I_l & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & I_l & \vdots\\ I_l & \cdots & I_l & 2I_l & I_l\\ I_l & \cdots & \cdots & I_l & I_l\end{pmatrix}-\alpha\begin{pmatrix}2I_l & I_l & \cdots & I_l & 0\\ I_l & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & I_l & \vdots\\ I_l & \cdots & I_l & 2I_l & 0\\ 0 & \cdots & \cdots & 0 & I_l\end{pmatrix}\right\}_{(m+1)\times(m+1)}\mathcal{P}.\end{aligned} \tag{4.17}$$

In this way, in order to show (4.14), we need only to find the largest α such that the $(m+1)\times(m+1)$ symmetric matrix

$$T=2\begin{pmatrix} I_m+ee^T & e \\ e^T & 1 \end{pmatrix}-\alpha\begin{pmatrix} I_m+ee^T & 0 \\ 0 & 1 \end{pmatrix}\succeq 0,$$

where e is an m-vector whose each element equals 1. Let $Tz=\nu z$, where ν is the eigenvalue of T and z is the related eigenvector. In the following we try to find the largest α such that all the eigenvalue of T are non-negative. Note that

$$T=\begin{pmatrix} (2-\alpha)(I_m+ee^T) & 2e \\ 2e^T & 2-\alpha \end{pmatrix}. \tag{4.18}$$

Without loss of generality, we assume that the eigenvectors of T have forms

$$z=\begin{pmatrix} y \\ 0 \end{pmatrix} \quad\text{and}\quad z=\begin{pmatrix} y \\ 1 \end{pmatrix},$$

where $y\in\Re^m$. In the first case, $z^T=(y^T,0)$, it follows from $Tz=\nu z$ and (4.18) that

$$\begin{cases} (2-\alpha)y+(2-\alpha)(e^Ty)e=\nu y, \\ e^Ty=0, \end{cases} \Rightarrow \begin{cases} (2-\alpha)y=\nu y, \\ e^Ty=0. \end{cases}$$

Therefore, we have $(m-1)$ linear independent vectors, y^i, $i=1,\cdots,m-1$, in the orthogonal subspace to e, and

$$z^i=\begin{pmatrix} y^i \\ 0 \end{pmatrix}, \quad i=1,\cdots,m-1,$$

are eigenvectors of T and the related eigenvalues are

$$\nu_1=\nu_2=\cdots=\nu_{m-1}=(2-\alpha).$$

In the second case, $z^T=(y^T,1)$, from $Tz=\nu z$ and (4.18) we have

$$\begin{cases} (2-\alpha)y+((2-\alpha)e^Ty+2)e=\nu y, \\ 2e^Ty+(2-\alpha)=\nu. \end{cases} \tag{4.19}$$

Left-multiplying the first equation of (4.19) by e^T and then using the second equation of (4.19) and $e^Te=m$, we derive that

$$\nu^2-(m+2)(2-\alpha)\nu+[(m+1)(2-\alpha)^2-4m]=0.$$

Thus, the small root of the above equation is

$$\nu_{\min}(T)=\frac{(m+2)(2-\alpha)-\sqrt{(m+2)^2(2-\alpha)^2-4[(m+1)(2-\alpha)^2-4m]}}{2}.$$

To ensure $\nu_{\min}(T) \geq 0$, we should find α such that

$$[(m+1)(2-\alpha)^2 - 4m] \geq 0.$$

The equivalent expression is

$$\alpha \leq 2\Big(1 - \sqrt{\frac{m}{m+1}}\Big).$$

The assertion of this lemma is proved. □

Due to the correctness of Lemma 4.1, Lemma 4.2 and Lemma 4.3, we have proved the convergence of the splitting method (1.6)-(1.7). we have proved the splitting method (1.6)-(1.7) can be interpreted as a convergent method which belongs to Algorithm I. Nevertheless, with the same predictor offered by (4.1), and the matrices Q and M given by (4.3) and (4.12), respectively, we suggest to use Algorithm II which needs to calculate the step size by (3.11b) in its correction step.

5 Splitting method (1.8) in the unified framework

In this section, we study the splitting method (1.8) in the frame of the unified algorithmic framework introduced in Section 3. For this purpose, the method (1.8) should be rewritten to a prediction-correction form.

Prediction: For given $\boldsymbol{w}^k = (\boldsymbol{x}^k, \lambda^k)$, we define the predictor by

$$\begin{cases} \begin{cases} \tilde{x}_1^k \in \arg\min\Big\{\mathcal{L}_\beta(x_1, x_2^k, \cdots, x_m^k, \lambda^k) + \frac{\tau\beta}{2}\|A_1(x_1 - x_1^k)\|^2 \,\Big|\, x_1 \in X_1\Big\}, \\ \qquad\vdots \\ \tilde{x}_i^k \in \arg\min\Big\{\mathcal{L}_\beta(x_1^k, \cdots, x_{i-1}^k, x_i, x_{i+1}^k, \cdots, x_m^k, \lambda^k) + \frac{\tau\beta}{2}\|A_i(x_i - x_i^k)\|^2 \,\Big|\, x_i \in X_i\Big\}, \\ \qquad\vdots \\ \tilde{x}_m^k \in \arg\min\Big\{\mathcal{L}_\beta(x_1^k, \cdots, x_{m-1}^k, x_m, \lambda^k) + \frac{\tau\beta}{2}\|A_m(x_m - x_m^k) \,\Big|\, x_m \in X_m\Big\}, \end{cases} & (5.1a) \\ \tilde{\lambda}^k = \lambda^k - \beta\Big(\sum_{i=1}^m A_i\tilde{x}_i^k - b\Big) + \beta\Big(\sum_{i=1}^m A_i(\tilde{x}_i^k - x_i^k)\Big). & (5.1b) \end{cases}$$

Notice that both the x-part and λ-part of this predictor are different from the one provided by (4.1) in the last section.

Lemma 5.1. *For solving the variational inequality* (2.4), *the predictor* $\tilde{\boldsymbol{w}}^k = (\tilde{\boldsymbol{x}}^k, \tilde{\lambda}^k)$ *offered by* (5.1) *satisfies*

$$\tilde{\boldsymbol{w}}^k \in \boldsymbol{\Omega}, \quad \boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k) + (\boldsymbol{w} - \tilde{\boldsymbol{w}}^k)^T \boldsymbol{F}(\tilde{\boldsymbol{w}}^k) \geq (\boldsymbol{w} - \tilde{\boldsymbol{w}}^k)^T Q(\boldsymbol{w}^k - \tilde{\boldsymbol{w}}^k), \quad \forall \boldsymbol{w} \in \boldsymbol{\Omega}, \quad (5.2)$$

where

$$Q = \begin{pmatrix} (\tau+1)\beta\mathcal{D} & 0 \\ -\mathcal{A} & \dfrac{1}{\beta}I \end{pmatrix} \tag{5.3}$$

and $\mathcal{D} = diag(A_1^TA_1, A_2^TA_2, \cdots, A_m^TA_m)$.

Proof. In the prediction step (5.1a), $\tilde{x}_i^k$ is defined by

$$\tilde{x}_i^k \in \arg\min\Big\{\mathcal{L}_\beta(x_1^k,\cdots,x_{i-1}^k,x_i,x_{i+1}^k,\cdots,x_m^k,\lambda^k) + \frac{\tau\beta}{2}\|A_i(x_i-x_i^k)\|^2 \,\Big|\, x_i\in X_i\Big\}.$$

Ignoring the constant term of the objective function (see (1.3)), it leads that

$$\begin{aligned}\tilde{x}_i^k \in \arg\min \Big\{&\theta_i(x_i) - x_i^TA_i^T\lambda^k + \frac{\beta}{2}\|A_i(x_i-x_i^k) + (\mathcal{A}\boldsymbol{x}^k - b)\|^2 \\ &+ \frac{\tau\beta}{2}\|A_i(x_i-x_i^k)\|^2 \,\Big|\, x_i\in X_i\Big\}.\end{aligned} \tag{5.4}$$

According to Lemma 2.1, using the optimal condition of the above problem, we get

$$\begin{aligned}&\tilde{x}_i^k\in\mathcal{X}_i, \quad \theta_i(x_i)-\theta_i(\tilde{x}_i^k)+(x_i-\tilde{x}_i^k)^T \\ &\quad \big\{-A_i^T\lambda^k+\beta A_i^T\big[A_i(\tilde{x}_i^k-x_i^k)+(\mathcal{A}\boldsymbol{x}^k-b)\big]+\tau\beta A_i(\tilde{x}_i^k-x_i^k)\big\}\ge 0, \quad \forall x_i\in\mathcal{X}_i.\end{aligned} \tag{5.5}$$

Using (5.1b), we have

$$\tilde{\lambda}^k = \lambda^k - \beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) + \beta\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k), \tag{5.6}$$

and thus

$$\lambda^k = \tilde{\lambda}^k + \beta(\mathcal{A}\boldsymbol{x}^k - b).$$

Substituting it in (5.5), we obtain

$$\begin{aligned}\tilde{x}_i^k\in\mathcal{X}_i, \quad &\theta_i(x_i)-\theta_i(\tilde{x}_i^k)+(x_i-\tilde{x}_i^k)^T \\ &\big\{-A_i^T\tilde{\lambda}^k+(\tau+1)\beta A_i^TA_i(\tilde{x}_i^k-x_i^k)\big\}\ge 0, \quad \forall x_i\in\mathcal{X}_i.\end{aligned} \tag{5.7}$$

Summing the inequality (4.6) from $i=1$ to $i=m$, we get

$$\begin{aligned}\tilde{\boldsymbol{x}}^k\in\boldsymbol{\mathcal{X}}, \quad &\boldsymbol{\theta}(\boldsymbol{x})-\boldsymbol{\theta}(\tilde{\boldsymbol{x}}^k)+(\boldsymbol{x}-\tilde{\boldsymbol{x}}^k)^T \\ &\big\{-\mathcal{A}^T\tilde{\lambda}^k+\beta(\tau+1)\mathcal{D}(\tilde{\boldsymbol{x}}^k-\boldsymbol{x}^k)]\big\}\ge 0, \quad \forall\boldsymbol{x}\in\boldsymbol{\mathcal{X}}.\end{aligned} \tag{5.8}$$

Notice that (5.6) can be written as

$$(\mathcal{A}\tilde{\boldsymbol{x}}^k-b)-\mathcal{A}(\tilde{\boldsymbol{x}}^k-\boldsymbol{x}^k)+\frac{1}{\beta}(\tilde{\lambda}^k-\lambda^k)=0,$$

and hence its corresponding variational inequality has a form

$$\tilde{\lambda}^k \in \Re^l, \quad (\lambda - \tilde{\lambda}^k)^T \Big\{ (\mathcal{A}\tilde{\boldsymbol{x}}^k - b) - \mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k) + \frac{1}{\beta}(\tilde{\lambda}^k - \lambda^k) \Big\} \geq 0, \quad \forall \lambda \in \Re^l. \tag{5.9}$$

Combining (5.8) and (5.9) and using the notation $\boldsymbol{F}(\tilde{\boldsymbol{w}}^k)$, the assertion of this lemma follows immediately. □

Lemma 5.2. *Let $\mathcal{I} = diag(I_{n_1}, I_{n_2}, \cdots, I_{n_m})$. For given $\boldsymbol{w}^k = (\boldsymbol{x}^k, \lambda^k)$, if the predictor $\tilde{\boldsymbol{w}}^k = (\tilde{\boldsymbol{x}}^k, \tilde{\lambda}^k)$ is generated by* (5.1)*, then the correction*

$$\begin{pmatrix} \boldsymbol{x}^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}^k \\ \lambda^k \end{pmatrix} - \begin{pmatrix} \mathcal{I} & 0 \\ -\beta\mathcal{A} & I \end{pmatrix} \begin{pmatrix} \boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k \\ \lambda^k - \tilde{\lambda}^k \end{pmatrix} \tag{5.10}$$

produces the same new iterate $\boldsymbol{w}^{k+1}$ as the one by (1.8)*.*

Proof. For the same $\boldsymbol{w}^k = (\boldsymbol{x}^k, \lambda^k)$, comparing (1.8a) and (5.1a), we have

$$\tilde{x}_i^k = x_i^{k+1}, \quad i = 1, \cdots, m.$$

Thus, the $\boldsymbol{x}$-part of the correction (1.7) can be written as

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - (\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k). \tag{5.11}$$

By using $\boldsymbol{x}^{k+1} = \tilde{\boldsymbol{x}}^k$, it follows from (1.8b)) that

$$\lambda^{k+1} = \lambda^k - \beta(\mathcal{A}\boldsymbol{x}^{k+1} - b) = \lambda^k - \beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b). \tag{5.12}$$

Since (see (5.1b))

$$\tilde{\lambda}^k = \lambda^k - \beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) + \beta\mathcal{A}(\tilde{\boldsymbol{x}}^k - \boldsymbol{x}^k),$$

we have

$$\beta(\mathcal{A}\tilde{\boldsymbol{x}}^k - b) = \lambda^k - \tilde{\lambda}^k - \beta\mathcal{A}(\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k).$$

Substituting it in (5.12), the λ-part of the correction (1.7) can be written as

$$\lambda^{k+1} = \lambda^k - [-\beta\mathcal{A}(\boldsymbol{x}^k - \tilde{\boldsymbol{x}}^k) + (\lambda^k - \tilde{\lambda}^k)]. \tag{5.13}$$

Combining (5.11) and (5.13), we get (5.10), and the assertion in proved. □

The splitting method (1.8) has been interpreted in the Algorithm I whose prediction and correction steps are (5.1) and (5.10), respectively. Note that in the correction step,

$$M = \begin{pmatrix} \mathcal{I} & 0 \\ -\beta\mathcal{A} & I \end{pmatrix}. \tag{5.14}$$

Indeed, we have positive definite matrix

$$H=\begin{pmatrix}(\tau+1)\beta\mathcal{D} & 0\\ 0 & \dfrac{1}{\beta}I\end{pmatrix}, \tag{5.15}$$

which satisfies $HM=Q$ (see the matrix Q in (5.3)). The remained task is to check the positivity of the matrix G when $\tau>m-1$.

Lemma 5.3. *For the given matrices Q, M and H (see* (5.3), (5.14) *and* (5.15)*), and any positive constant $\tau>m-1$, we have*

$$G=(Q^T+Q)-M^THM\succ 0. \tag{5.16}$$

Proof. Indeed, by a manipulation, we have

$$\begin{aligned}G=&(Q^T+Q)-M^THM=(Q^T+Q)-Q^TM\\ =&\begin{pmatrix}2(\tau+1)\beta\mathcal{D} & -\mathcal{A}^T\\ -\mathcal{A} & \dfrac{2}{\beta}I\end{pmatrix}-\begin{pmatrix}(\tau+1)\beta\mathcal{D} & -\mathcal{A}^T\\ 0 & \dfrac{1}{\beta}I\end{pmatrix}\begin{pmatrix}\mathcal{I} & 0\\ -\beta\mathcal{A} & I\end{pmatrix}\\ =&\begin{pmatrix}2(\tau+1)\beta\mathcal{D} & -\mathcal{A}^T\\ -\mathcal{A} & \dfrac{2}{\beta}I\end{pmatrix}-\begin{pmatrix}(\tau+1)\beta\mathcal{D}+\beta\mathcal{A}^T\mathcal{A} & -\mathcal{A}^T\\ -\mathcal{A} & \dfrac{1}{\beta}I\end{pmatrix}\\ =&\begin{pmatrix}(\tau+1)\beta\mathcal{D}-\beta\mathcal{A}^T\mathcal{A} & 0\\ 0 & \dfrac{1}{\beta}I\end{pmatrix}.\end{aligned}$$

By using

$$\mathcal{P}=\mathrm{diag}\Big(\sqrt{\beta}A_1,\cdots,\sqrt{\beta}A_m,\frac{1}{\sqrt{\beta}}I\Big),$$

we have

$$G=\mathcal{P}^T\begin{pmatrix}\tau I_l & -I_l & \cdots & -I_l & 0\\ -I_l & \ddots & \ddots & \vdots & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots\\ -I_l & \cdots & -I_l & \tau I_l & 0\\ 0 & \cdots & \cdots & 0 & I_l\end{pmatrix}\mathcal{P}. \tag{5.17}$$

In this way, we need only to consider the $m\times m$ symmetric matrix

$$T_0=(\tau+1)I_m-ee^T,$$

where e is an a m-vector whose each element equals 1. Indeed, $T_0\succ 0$ if and only if $\tau>m-1$. The lemma is proved. □

Due to the correctness of Lemma 5.1, Lemma 5.2 and Lemma 5.3, we have proved the convergence of the splitting method (1.8). we have proved the splitting method (1.8) can be interpreted as a convergent method which belongs to Algorithm I. Nevertheless with the same predictor offered by (5.1), the matrix Q and M given by (5.3) and (5.14), respectively, we suggest to use Algorithm II which needs to calculate the step size by (3.11b) in its correction step.

6 Conclusion remarks

For solving the convex minimization problem with linear constraints and an objective function in form of the sum of m functions without coupled variables, the straightforward splitting version of ALM with full Jacobian decomposition could be divergent. As a remedy which enjoys the feature that all the x_i-subproblems can be solved in parallel, we have proposed the methods (1.6)-(1.7) and (1.8). In this paper, by suitable rewriting, we let such methods fit into the unified algorithmic framework, and thus the convergence of the different previous methods follows immediately. Indeed, for linearly constrained convex optimization, the unified algorithmic frameworks can guide us to construct suitable solution methods. In addition, it is also powerful and flexible for convergence analysis.

Acknowledgements

The author was supported by the NSFC Grant No. 11871029.

References

[1] A. Chambolle, and T. Pock, On the ergodic convergence rates of a first-order primal-dual algorithm, Math. Program., 59 (2016), 253–287.
[2] C. H. Chen, B. S. He, Y. Y. Ye and X. M. Yuan, The direct extension of ADMM for multi-block convex minimization problems is not necessary convergent, Math. Program., 155 (2016), 57–79.
[3] R. Glowinski, Numerical Methods for Nonlinear Variational Problems, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1984.
[4] G. Y. Gu, B. S. He and X. M. Yuan, Customized proximal point algorithms for linearly constrained convex minimization and saddle-point problems: a unified approach, Comput. Optim. Appl., 59 (2014), 135–161.
[5] B. S. He, PPA-like contraction methods for convex optimization: a framework using variational inequality approach, J. Oper. Res. Soc. China, 3 (2015), 391–420.
[6] B. S. He, L. S. Hou and X. M. Yuan, On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming, SIAM J. Optim., 25 (2015), 2274–2312.
[7] B. S. He, M. Tao and X. M. Yuan, A splitting method for separable convex programming, IMA J. Numer. Anal., 31 (2015), 394–426.

[8] B. S. He, and X. M. Yuan, A class of ADMM-based algorithms for three-block separable convex programming, Comput. Optim. Appl., 70 (2018), 791–826.
[9] B. S. He and X. M. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective, SIAM J. Imag. Sci., 5 (2012), 119–149.
[10] B. S. He, X. M. Yuan and W. X. Zhang, A customized proximal point algorithm for convex minimization with linear constraints, Comput. Optim. Appl., 56 (2013), 559–572.
[11] M. R. Hestenes, Multiplier and gradient methods, J. Optim. Theory Appl., 4 (1969), 303–320.
[12] B. Martinet, Regularisation, d'inéquations variationelles par approximations succesives, Rev. Francaise d'Inform. Recherche Oper., 4 (1970), 154–159.
[13] M. J. D. Powell, A Method for Nonlinear Constraints in Minimization Problems, in Optimization edited by R. Fletcher, Academic Press, New York, 1969, 283–298.
[14] R. T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Con. Optim., 14 (1976), 877–898.
[15] R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. Oper. Res., 1 (1976), 97–116.
[16] B. S. He, My 20 years research on alternating directions method of multipliers, Oper. Res. Trans., 22 (2018), 1–31.