Journal of Nonlinear and Convex Analysis
Volume 16, Number 8, 2015, 1539–1556
Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

A PROJECTION-BASED SPLITTING METHOD FOR STRUCTURED VARIATIONAL INEQUALITIES

HONGJIN HE* AND HONG-KUN XU†

This paper is dedicated to Professor Wataru Takahashi on the occasion of his 70th birthday.

Abstract. This paper develops a projection-based splitting method for solving the *structured variational inequality* (SVI). The method is a two-stage method which consists of a prediction step and a correction step and which fully exploits the favorable structure of the problem. One important benefit of our method is that the prediction step can be implemented *simultaneously* and the correction step requires tiny computational cost. Thus, it gains eligibility for solving large-scale problems. We prove the global convergence of the new method under some mild assumptions. In addition, we formulate the *generalized split equality problem* and *network resource allocation problem* as special cases of the SVI and gainfully employ our new method to these problems.

1. Introduction

Let $F(\cdot)$ be a continuous mapping from $\mathbb{R}^n$ into itself, and let Ω be a closed convex subset of $\mathbb{R}^n$. The classical *variational inequality* (VI) can be characterized as finding a vector $\mathbf{x}^* \in \Omega$ such that

$$\langle \mathbf{x} - \mathbf{x}^*,\ F(\mathbf{x}^*)\rangle \geq 0, \quad \forall \mathbf{x} \in \Omega. \tag{1.1a}$$

In this paper, we are concerned with the linearly constrained VIP with *separable structure* in the sense that the underlying mapping $F(\cdot)$ consists of m individual parts. More specifically, the structured VIP is of the form:

$$\mathbf{x} := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad F(\mathbf{x}) := \begin{pmatrix} f_1(x_1) \\ f_2(x_2) \\ \vdots \\ f_m(x_m) \end{pmatrix}, \tag{1.1b}$$

and the set Ω is defined by

$$\Omega := \left\{ (x_1, x_2, \ldots, x_m) \;\middle|\; \sum_{i=1}^{m} A_i x_i = b,\ x_i \in \mathcal{X}_i \right\}. \tag{1.1c}$$

2010 *Mathematics Subject Classification.* 65K05, 90C25, 90C35.
Key words and phrases. Projection-based splitting method, structured variational inequality, split equality problem, split feasibility problem, convex feasibility problem, network resource allocation problem.
*Supported in part by National Natural Science Foundation of China (No. 11301123) and the Zhejiang Provincial NSF (No. LZ14A010003).
†Supported in part by NSC 102-2115-M-110-001-MY3; Corresponding author.

Here every $f_i(\cdot) : \mathcal{X}_i \to \mathbb{R}^{n_i}$ is a monotone operator; each $A_i \in \mathbb{R}^{l\times n_i}$ is a given matrix; $b \in \mathbb{R}^l$ is a given vector; each $\mathcal{X}_i \subset \mathbb{R}^{n_i}$ is a nonempty closed convex set, and $\sum_{i=1}^m n_i = n$. Throughout, the solution set of (1.1) is assumed to be nonempty. It is well documented in the literature [3, 12, 14, 36] that the variational inequality theory provides a powerful unifying methodology for the study of mathematical programming and equilibrium problems. In past several decades, the variational inequality theory and algorithms have been developed significantly, see the monograph [14]. Recently, many structured optimization problems arising from image processing, statistical learning and transportation can be formulated and studied as VIs, e.g., see [13, 18, 38, 40, 41] for certain applications; meanwhile, algorithms for (1.1) have received considerable attention in the literature; the reader is referred to [19, 21, 22, 24] for special cases where $m \leq 3$ of (1.1).

Among the existing algorithms for solving general VIs, the projection-like methods are the simplest, especially when the projection onto the feasible set can be represented explicitly. However, the mixed set Ω in (1.1c), which consists of linear constraints and simple convex sets, results in a big difficulty for computing the projection onto Ω, thereby making the traditional projection-like methods, such as the *extragradient method* [30], harder to be implemented. Indeed, the structured VI (1.1) has some favorable structures that could be utilized. By attaching a Lagrangian multiplier $\lambda \in \mathbb{R}^l$ to the linear constraints, the VI (1.1) can be recast as finding a vector $\mathbf{u}^* \in \mathcal{U}$ satisfying the property

$$\langle \mathbf{u} - \mathbf{u}^*,\ H(\mathbf{u}^*)\rangle \geq 0, \quad \forall \mathbf{u} \in \mathcal{U}, \tag{1.2a}$$

where

$$\mathbf{u} := \begin{pmatrix} \mathbf{x} \\ \lambda \end{pmatrix}, \quad H(\mathbf{u}) := \begin{pmatrix} f_1(x_1) - A_1^\top \lambda \\ \vdots \\ f_m(x_m) - A_m^\top \lambda \\ \sum_{i=1}^m A_i x_i - b \end{pmatrix} \quad \text{and} \quad \mathcal{U} := \prod_{i=1}^m \mathcal{X}_i \times \mathbb{R}^l. \tag{1.2b}$$

[Note: For the sake of notational simplicity, we shall, throughout the rest of this paper, write $(\mathbf{x}, \lambda)$ for $\mathbf{u}$.] Comparatively speaking, the set $\mathcal{U}$ in (1.2b) is easier than the set Ω in (1.1c), and some computational benefits may occur in algorithm implementation. In what follows, we will study the VIs in the form (1.2) which is denoted by SVI $(\mathcal{U}, H)$.

When considering a special case of (1.1) with $m = 2$, one of the most popular methods is the *alternating direction method of multipliers* (ADMM) introduced in [15, 16], which decomposes the subproblem of *augmented lagrangian method* (ALM) [25, **?**] into two smaller scaled sub-variational inequalities. Specifically, ADMM solves (1.2) via the following procedure:

Given $(x_2^k, \lambda^k) \in \mathcal{X}_2 \times \mathbb{R}^l$, find a point $x_1^{k+1} \in \mathcal{X}_1$ such that the following inequality holds for all $x_1 \in \mathcal{X}_1$

$$\left\langle x_1 - x_1^{k+1}, f_1(x_1^{k+1}) - A_1^\top \left[\lambda^k - \beta\left(A_1 x_1^{k+1} + A_2 x_2^k - b\right)\right]\right\rangle \geq 0. \tag{1.3a}$$

With (x_1^{k+1}, λ^k), then seek a vector $x_2^{k+1} \in \mathcal{X}_2$ satisfying, for all $x_2 \in \mathcal{X}_2$,

$$\left\langle x_2 - x_2^{k+1}, f_2(x_2^{k+1}) - A_2^\top \left[\lambda^k - \beta\left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\right)\right]\right\rangle \geq 0. \tag{1.3b}$$

Finally, update the Lagrangian multiplier λ^{k+1} *via*

$$\lambda^{k+1} = \lambda^k - \beta\left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\right). \tag{1.3c}$$

Here, $\beta > 0$ is a penalty parameter. Obviously, the ADMM fully exploits the separable structure of (1.1) and updates its iterates in a Gauss-Seidel order. In recent years, the successful applications of ADMM in the areas of signal/image processing and statistical learning have made it received a revived interest, we refer the reader to [4] for a comprehensive review on ADMM. Since ADMM outperforms many classical algorithms, a straightforward extension of ADMM (1.3) to handle multiple-block convex minimization problems attracts much attention. Furthermore, applying this extension to solve (1.2) immediately leads to the following iterative scheme:

Given $(x_1^{k+1},\ldots,x_{i-1}^{k+1},x_{i+1}^k,\ldots,x_m^k,\lambda^k)$, *sequentially find the new iterate* x_i^{k+1} *of the i-th variable so that* x_i^{k+1} *is a solution to the VIP:*

$$\left\langle x_i' - x_i,\ f_i(x_i) - A_i^\top\left[\lambda^k - \beta\left(A_i x_i + p^k\right)\right]\right\rangle \geq 0, \quad \forall x_i' \in \mathcal{X}_i. \tag{1.4a}$$

where

$$p^k := \sum_{j=1}^{i-1} A_j x_j^{k+1} + \sum_{j=i+1}^{m} A_j x_j^k - b.$$

Finally, update the Lagrangian multiplier λ^{k+1} *via*

$$\lambda^{k+1} = \lambda^k - \beta\left(\sum_{i=1}^{m} A_i x_i^{k+1} - b\right). \tag{1.4b}$$

Unfortunately, it was well demonstrated in [11] that the extended ADMM (1.4) is not necessarily convergent for all cases where $m \geq 3$, and this holds true even when we solve the subproblems (1.4a) simultaneously (i.e., the full Jacobian decomposition of ALM, see [20]). To ensure the global convergence, the most mature way is to update the output of (1.4) by adding a further correction step, e.g., see [18, 19, 20, 23]. However, taking a close look at the subproblems (1.3a) and (1.3b) (or (1.4a)), we observe that ADMM (or its extension (1.4)) still solves a series of *variational inequality subproblems* exactly, but only reducing the problem's scale. Actually, solving a VIP exactly is difficult, if not impossible in many cases. Moreover, some applications of ADMM empirically indicate that the difficulty of subproblems may significantly affect the efficiency of the method (see [4, 10, 17] for the numerical experiences).

In this paper, we use a projection step instead of solving a variational inequality subproblem, and then propose a two-stage method, which consists of a prediction step and a correction step, to solve the structured VI (1.1). A notable benefit is that the new method is more implementable than ADMM and its variants as long as the projections onto $\mathcal{X}_i$'s are easy enough. We shall mention that an additional correction step but with tiny computational cost is a must to guarantee the global convergence of our method. Since the new method fully exploits the separable structure by splitting VI (1.2) into $(m+1)$ individual parts and inherits the simplicity of projection-like methods, we call it *projection-based splitting method* and denote it by **ProjSM**. Another remarkable advantage of the ProjSM is that the simultaneous prediction step makes it gain eligibility to parallel implementation for large-scale problems. Finally, we introduce a *generalized split equality problem* (GSEP), which includes some well-known special cases, such as *split feasibility problem*, *split equality problem*, and *convex feasibility problem*. Then, we formulate the GSEP and the *network resource allocation problem* (NRAP) as special cases of VI (1.1) and demonstrate the applicability of the ProjSM to these problems.

The remainder of this paper is built up as follows. In Section 2, we summarize some notation, definitions, and well-known results that will be used in our later analysis. In Section 3, we describe the algorithmic framework of the proposed method. In Section 4, we establish the global convergence of the new method under some mild assumptions. The applicability of our new algorithm to GSEP and NRAP is demonstrated in Section 5. Finally, some concluding remarks are provided in Section 6.

2. Preliminaries

In this section, we summarize some basic notation, concepts and well known results that will play important roles in further discussions.

Let $\mathbb{R}^n$ be the n-dimensional Euclidean space and $^\top$ symbolize the transpose. For any two vectors $u, v \in \mathbb{R}^n$, we use $\langle u, v\rangle$ to denote the standard inner product. Furthermore, with a given symmetric positive definite matrix M, let $\|u\|_M = \sqrt{\langle u, Mu\rangle}$ represent the M-norm, and particularly, let $\|\cdot\|$ denote the standard Euclidean norm. For any matrix A, let $\|A\|$ be its matrix 2-norm.

Throughout, let $P_{\Omega,M}[\cdot]$ be the projection operator from $\mathbb{R}^n$ onto a nonempty closed convex set Ω under the M-norm, which is defined by

$$P_{\Omega,M}[v] := \arg\min\left\{\|u - v\|_M \mid u \in \Omega\right\}, \qquad v \in \mathbb{R}^n,$$

and particularly, we denote by $P_\Omega[\cdot]$ the projection operator under the Euclidean norm. It is well known from [3, p. 211] that the projection $P_{\Omega,M}[\cdot]$ can be characterized by the following inequality:

$$\langle w - P_{\Omega,M}[v],\ M\left(v - P_{\Omega,M}[v]\right)\rangle \le 0, \quad \forall v \in \mathbb{R}^n,\ \forall w \in \Omega. \tag{2.1}$$

Definition 2.1. An operator $F(\cdot)$ from Ω into $\mathbb{R}^n$ is said to be

a) monotone if

$$\langle u - v,\ F(u) - F(v)\rangle \ge 0, \quad \forall u,\ v \in \Omega.$$

b) Lipschitz continuous if there exists a constant $L > 0$ such that

$$\|F(u) - F(v)\| \le L\|u - v\|, \quad \forall u,\ v \in \Omega.$$

Throughout this paper, we assume that each set $\mathcal{X}_i$ $(i = 1, \ldots, m)$ is simple enough in the sense that the projection onto it is easy to compute. Moreover, every function $f_i(\cdot)$ $(i = 1, \ldots, m)$ is assumed to be monotone and Lipschitz continuous with constant $L_{f_i} > 0$.

Below, we give a well-known fixed point characterization of solutions of VI (1.1) and refer the reader to [3, p. 267] for its proof.

Lemma 2.2. *Let Ω be a closed convex set of $\mathbb{R}^n$ and let G be a symmetric positive definite matrix. Then, a point $\mathbf{x}^* \in \Omega$ is a solution of VI* (1.1) *if and only if*

$$\mathbf{x}^* = P_{\Omega,G}\left[\mathbf{x}^* - \beta G^{-1}F(\mathbf{x}^*)\right], \quad \forall \beta > 0.$$

Obviously, the above lemma can be restated as solving VI (1.1) being equivalent to finding a zero of the mapping:

$$E_{[\beta,G]}\left(\mathbf{x}, F, \Omega\right) := \mathbf{x} - P_{\Omega,G}\left[\mathbf{x} - \beta G^{-1}F(\mathbf{x})\right]. \tag{2.2}$$

Thus, $\|E_{[\beta,G]}\left(\mathbf{x}^*, F, \Omega\right)\| = 0$ means that $\mathbf{x}^*$ is a solution of VIP (1.1), and in practice, we could use $\|E_{[\beta,G]}\left(\mathbf{x}, F, \Omega\right)\| \le \mathrm{tol}$ to be a termination criteria with a preset stopping tolerance 'tol'.

3. The algorithm

In this section, we describe the *projection-based splitting method* (ProjSM) and give some important remarks on this method.

Before stating the algorithm, we first introduce some notation for simplicity. For the index i from 1 to m, let $I_{n_i} \in \mathbb{R}^{n_i \times n_i}$ be the identity matrix and denote the difference between $f_i(x_i)$ and $f_i(\widetilde{x}_i)$ as $\phi_i(x_i, \widetilde{x}_i) := f_i(x_i) - f_i(\widetilde{x}_i)$. Furthermore, we denote

$$M := \begin{pmatrix} r_1 I_{n_1} & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & r_m I_{n_m} & 0 \\ 0 & \dots & 0 & \frac{1}{\beta} I_l \end{pmatrix} \quad \text{and} \quad \Phi(\mathbf{x}, \widetilde{\mathbf{x}}) := \begin{pmatrix} \phi_1(x_1, \widetilde{x}_1) \\ \vdots \\ \phi_m(x_m, \widetilde{x}_m) \\ 0 \end{pmatrix}. \tag{3.1}$$

Note that for $\mathbf{u} := (\mathbf{x}, \lambda)$, we have

$$\|\mathbf{u}\|_M^2 = \langle \mathbf{u},\ M\mathbf{u}\rangle = \sum_{i=1}^m r_i \langle x_i, x_i\rangle + \frac{1}{\beta}\langle \lambda, \lambda\rangle = \sum_{i=1}^m r_i \|x_i\|^2 + \frac{1}{\beta}\|\lambda\|^2.$$

Since the proposed ProjSM is a two-stage method, we divide the description of our method into two parts. Below, we first present the details of the prediction step.

(**Initialization**). Given $\gamma \in (0,2)$, $\nu \in (0,1)$, $\beta > 0$, $\mathbf{u}^0 \in \mathcal{U}$.
(**Prediction step**). Generate a medium point $\widehat{\lambda}^k$ via

$$\widehat{\lambda}^k = \lambda^k - \beta\left(\sum_{i=1}^m A_i x_i^k - b\right). \tag{3.2a}$$

Then, obtain $(\widetilde{x}_1^k, \dots, \widetilde{x}_m^k)$ with appropriate $r_i > 0$ (simultaneously, if possible):

$$\widetilde{x}_i^k := P_{\mathcal{X}_i}\left\{x_i^k - \frac{1}{r_i}\left(f_i(x_i^k) - A_i^\top \widehat{\lambda}^k\right)\right\}, \quad (i = 1, \dots, m). \tag{3.2b}$$

Compute the predictor $\widetilde{\lambda}^k$ via

$$\widetilde{\lambda}^k = \lambda^k - \beta\left(\sum_{i=1}^m A_i \widetilde{x}_i^k - b\right). \tag{3.2c}$$

Remark 3.1. We easily observe that the projection step (3.2b) dominates the main computational task of the prediction step. Under the assumptions that all $\mathcal{X}_i$'s are simple convex sets, the prediction step is easily implemented. On the other hand, the projection step (3.2b) enjoys its simultaneous implementation on parallel computers. Hence, our method is fully eligible for solving large-scale problem.

Remark 3.2. For the choice of each r_i $(i = 1, \ldots, m)$, we shall seek an appropriate r_i by setting a constant or a line search procedure such that

$$(3.3)\qquad r_i\nu \left\|x_i^k - \widetilde{x}_i^k\right\|^2 \geq \frac{m+2}{4}\beta \left\|A_i x_i^k - A_i\widetilde{x}_i^k\right\|^2 + \left\langle x_i^k - \widetilde{x}_i^k,\ f_i(x_i^k) - f_i(\widetilde{x}_i^k)\right\rangle, \quad 1 \leq i \leq m,$$

where $\nu \in (0, 1)$ is a given constant. Using the Cauchy-Schwarz inequality together with the Lipschitz continuity of $f_i(\cdot)$, it is easy to derive that every

$$r_i \geq \left(L_{f_i} + \frac{m+2}{4}\beta\|A_i^\top A_i\|\right)/\nu$$

always ensures inequality (3.3). Notice that the choice of r_i (condition (3.3)) will play a key role in the coming convergence analysis.

Now, we describe the correction step of the new ProjSM as follows.

(**Correction step I**). Update $\mathbf{u}^{k+1} := (\mathbf{x}^{k+1}, \lambda^{k+1})$ via

$$(3.4a)\qquad \mathbf{u}^{k+1} := \mathbf{u}^k - \gamma\alpha_k d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k),$$

where the step size α_k is defined by

$$(3.4b)\qquad \alpha_k := \frac{\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)}{\|d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\|_M^2},$$

with

$$(3.4c)\qquad d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) := (\mathbf{u}^k - \widetilde{\mathbf{u}}^k) - M^{-1}\Phi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k),$$

and

$$(3.4d)\quad \varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) := \left\langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\rangle + \left\langle \lambda^k - \widetilde{\lambda}^k,\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k)\right\rangle.$$

Remark 3.3. Notice that the correction step (3.4) only involves some simple matrix-vector products, and the matrix M defined in (3.1) is a block scalar matrix so that it is trivial to compute its inverse. Thus, the computational cost of the correction step is relatively cheap. In addition, we can also get a variant of (3.4) without matrix inverse as follows:

(**Correction step II**). Update $\mathbf{u}^{k+1} := (\mathbf{x}^{k+1}, \lambda^{k+1})$ via

$$\mathbf{u}^{k+1} := \mathbf{u}^k - \gamma\widehat{\alpha}_k \widehat{d}(\mathbf{u}^k, \widetilde{\mathbf{u}}^k), \tag{3.5a}$$

where the step size $\widehat{\alpha}_k$ is defined by

$$\widehat{\alpha}_k := \frac{\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)}{\|\widehat{d}(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\|^2}, \tag{3.5b}$$

with $\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ given by (3.4d) and

$$\widehat{d}(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) := M(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) - \Phi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k). \tag{3.5c}$$

The main distinction between (3.4) and (3.5) comes from the definitions of $d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ and $\widehat{d}(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$. It is clear that $\widehat{d}(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) = Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$. Then, we can similarly prove the global convergence of the variant (3.5).

In summary, our ProjSM is computationally attractive due to its simplicity.

4. Convergence analysis

This section aims mainly at the establishment of the global convergence of the ProjSM (see (3.2) and (3.4)), leaving skipped the analogous analysis of its variant (see (3.2) and (3.5)). In the subsequent analysis, we further denote by

$$\mathbf{A} := (A_1, A_2, \ldots, A_m, 0) \in \mathbb{R}^{l\times(n+l)}$$

a block matrix for notational convenience.

Lemma 4.1. *Suppose that $\mathbf{u}^* := (\mathbf{x}^*, \lambda^*)$ is a solution of* SVI $(\mathcal{U}, H)$. *Then the sequences $\{\mathbf{u}^k\}$ and $\{\widetilde{\mathbf{u}}^k\}$ generated by the proposed* ProjSM *satisfy that*

$$\left\langle \mathbf{u}^k - \mathbf{u}^*,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\rangle \geq \varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k),$$

where $d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ and $\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ are given by (3.4c) *and* (3.4d)*, respectively.*

Proof. By setting $w := x_i^*$, $v := x_i^k - \frac{1}{r_i}\left(f_i(x_i^k) - A_i^\top\widehat{\lambda}^k\right)$, $M := I_{n_i}$ and $\Omega := \mathcal{X}_i$ in (2.1), it follows from (3.2b) that

$$\left\langle x_i^* - \widetilde{x}_i^k,\ x_i^k - \frac{1}{r_i}\left(f_i(x_i^k) - A_i^\top\widehat{\lambda}^k\right) - \widetilde{x}_i^k\right\rangle \leq 0.$$

Multiplying the above inequality by r_i and using the formulas (3.2a) and (3.2c) together with the notation in (3.1), we arrive at

$$\begin{aligned}&\left\langle x_i^* - \widetilde{x}_i^k,\ f_i(\widetilde{x}_i^k) - A_i^\top\widetilde{\lambda}^k + r_i(\widetilde{x}_i^k - x_i^k) + \phi_i(x_i^k, \widetilde{x}_i^k)\right\rangle \\ &\qquad - \left\langle x_i^* - \widetilde{x}_i^k,\ \beta A_i^\top\left(\textstyle\sum_{j=1}^m A_j(\widetilde{x}_j^k - x_j^k)\right)\right\rangle \geq 0.\end{aligned} \tag{4.1}$$

Additionally, reformulating (3.2c) as $\frac{1}{\beta}(\lambda^k - \widetilde{\lambda}^k) = \left(\sum_{i=1}^m A_i\widetilde{x}_i^k - b\right)$ immediately yields

$$\left\langle \lambda^* - \widetilde{\lambda}^k,\ \left(\sum_{i=1}^m A_i\widetilde{x}_i^k - b\right) + \frac{1}{\beta}\left(\widetilde{\lambda}^k - \lambda^k\right)\right\rangle = 0. \tag{4.2}$$

Consequently, upon summing up (4.1) from $i = 1$ to m and (4.2) and recalling the definitions of M and $\mathbf{A}$, we can rewrite the resulting inequality into a compact form as follows:

$$\left\langle \mathbf{u}^* - \widetilde{\mathbf{u}}^k,\ H(\widetilde{\mathbf{u}}^k) - Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) + \beta \mathbf{A}^\top \mathbf{A}(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \right\rangle \geq 0.$$

Equivalently,

$$\left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) \right\rangle \geq \left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ H(\widetilde{\mathbf{u}}^k) + \beta \mathbf{A}^\top \mathbf{A}(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \right\rangle. \tag{4.3}$$

Observe that the monotonicity of $f_i(\cdot)$ for $1 \leq i \leq m$ implies that the operator $H(\mathbf{u})$ defined in (1.2b) is also monotone. Therefore, for any solution $\mathbf{u}^*$ of SVI $(\mathcal{U}, H)$, it follows from the definition of VIP (1.2) that

$$\left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ H(\widetilde{\mathbf{u}}^k) \right\rangle \geq \left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ H(\mathbf{u}^*) \right\rangle \geq 0.$$

This together with (4.3) further implies that

$$\begin{aligned}
\left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) \right\rangle &\geq \left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ H(\widetilde{\mathbf{u}}^k) + \beta \mathbf{A}^\top \mathbf{A}(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \right\rangle \\
&= \left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ H(\widetilde{\mathbf{u}}^k) \right\rangle + \left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ \beta \mathbf{A}^\top \mathbf{A}(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \right\rangle \\
&\geq \left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ \beta \mathbf{A}^\top \mathbf{A}(\mathbf{u}^k - \widetilde{\mathbf{u}}^k) \right\rangle \\
&= \beta \left\langle \sum_{i=1}^m A_i(\widetilde{x}_i^k - x_i^*),\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k) \right\rangle \\
&= \left\langle \lambda^k - \widetilde{\lambda}^k,\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k) \right\rangle, \qquad (4.4)
\end{aligned}$$

where the third equality comes from the fact $\sum_{i=1}^m A_i x_i^* = b$ and (3.2c).

On the other hand, from the definition of $\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ and (4.4) it follows that

$$\begin{aligned}
&\left\langle \mathbf{u}^k - \mathbf{u}^*,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) \right\rangle \\
&= \left\langle \widetilde{\mathbf{u}}^k - \mathbf{u}^*,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) \right\rangle + \left\langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) \right\rangle \\
&\geq \left\langle \lambda^k - \widetilde{\lambda}^k,\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k) \right\rangle + \left\langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) \right\rangle \\
&= \varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k).
\end{aligned}$$

The assertion is proved. □

The following result, which plays a central role in the global convergence analysis, implies that $-d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ is a descent direction of the distance function $\frac{1}{2}\|\mathbf{u} - \mathbf{u}^*\|^2$ at $\mathbf{u}^k$, where $\mathbf{u}^*$ is a solution of SVI $(\mathcal{U}, H)$.

Lemma 4.2. *Suppose that $\nu \in (0, 1)$, $\beta > 0$ and each r_i from $i = 1$ to m satisfies condition* (3.3)*. Then, there exists a constant $c > 0$ such that*

$$\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) \geq c \left\| \mathbf{u}^k - \widetilde{\mathbf{u}}^k \right\|^2, \quad \forall k \geq 1. \tag{4.5}$$

Proof. First, an application of the well-known inequality

$$2\langle \mathbf{a},\ \mathbf{b}\rangle \geq -\tau\|\mathbf{a}\|^2 - \frac{1}{\tau}\|\mathbf{b}\|^2, \quad \forall \mathbf{a},\ \mathbf{b} \in \mathbb{R}^n,\ \tau > 0,$$

implies that

$$\left\langle \lambda^k - \widetilde{\lambda}^k,\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k) \right\rangle \geq -\sum_{i=1}^m \frac{(m+1)\beta}{4}\left\|A_i(x_i^k - \widetilde{x}_i^k)\right\|^2 - \frac{m\|\lambda^k - \widetilde{\lambda}^k\|^2}{(m+1)\beta}. \tag{4.6}$$

Consequently, it is immediately clear from condition (3.3) that

$$\begin{aligned}
&\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\\
&= \left\langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k,\ Md(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\rangle + \left\langle \lambda^k - \widetilde{\lambda}^k,\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k)\right\rangle\\
&= \left\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\right\|_M^2 - \sum_{i=1}^m \left\langle x_i^k - \widetilde{x}_i^k,\ \phi_i(x_i^k, \widetilde{x}_i^k)\right\rangle + \left\langle \lambda^k - \widetilde{\lambda}^k,\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k)\right\rangle\\
&\geq \sum_{i=1}^m (1-\nu) r_i \left\|x_i^k - \widetilde{x}_i^k\right\|^2 + \frac{1}{(m+1)\beta}\left\|\lambda^k - \widetilde{\lambda}^k\right\|^2 + \sum_{i=1}^m \frac{\beta}{4}\left\|A_i x_i^k - A_i \widetilde{x}_i^k\right\|^2\\
&\geq c\left\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\right\|^2,
\end{aligned}$$

where $c := \min\left\{\min_{1\leq i\leq m}\{(1-\nu)r_i\},\ \frac{1}{(m+1)\beta}\right\}$. Hence proved. ☐

The above assertion implies that $\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) > 0$ if $\mathbf{u}^k \neq \widetilde{\mathbf{u}}^k$. This together with Lemma 4.1 clearly shows that $-d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ is a descent direction at $\mathbf{u}^k$. Of course, $\mathbf{u}^k = \widetilde{\mathbf{u}}^k$ means that we have got a solution of SVI $(\mathcal{U}, H)$ and we can employ $\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\| \leq \text{tol}$ to be the termination criteria for the algorithm's implementation.

Next, we state the rationale for the choice of the step size α_k in (3.4b). We first denote by

$$\mathbf{u}^{k+1}(\alpha) := \mathbf{u}^k - \alpha d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$$

the function of step size α dependent on $\mathbf{u}^k$ and $\widetilde{\mathbf{u}}^k$. Let $\mathbf{u}^*$ be an arbitrary solution of SVI $(\mathcal{U}, H)$ and let

$$\Theta(\alpha) := \left\|\mathbf{u}^k - \mathbf{u}^*\right\|_M^2 - \left\|\mathbf{u}^{k+1}(\alpha) - \mathbf{u}^*\right\|_M^2$$

be a progress-function to measure the improvement obtained at the k-th iteration of the method. Clearly, large $\Theta(\alpha)$ means that more improvement is obtained. Therefore, we hopefully maximize $\Theta(\alpha)$ for seeking an optimal improvement. Unfortunately, an unknown $\mathbf{u}^*$ involved in $\Theta(\alpha)$ results in a difficult (even impossible) task to maximize $\Theta(\alpha)$ directly. Actually, by invoking the result of Lemma 4.1, it is easy to derive that

$$\Theta(\alpha) = \left\|\mathbf{u}^k - \mathbf{u}^*\right\|_M^2 - \left\|\mathbf{u}^{k+1}(\alpha) - \mathbf{u}^*\right\|_M^2$$

$$
\begin{aligned}
&= \left\|\mathbf{u}^k - \mathbf{u}^*\right\|_M^2 - \left\|\mathbf{u}^k - \alpha d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) - \mathbf{u}^*\right\|_M^2 \\
&= 2\alpha \left\langle \mathbf{u}^k - \mathbf{u}^*,\ M d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\rangle - \alpha^2 \left\|d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\|_M^2 \\
&\geq 2\alpha\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) - \alpha^2 \left\|d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\|_M^2 =: \psi(\alpha). \qquad (4.7)
\end{aligned}
$$

Note that $\psi(\alpha)$ in (4.7) is a quadratic function of α without the unknown $\mathbf{u}^*$. Therefore, we can alternatively maximize the lower bound function $\psi(\alpha)$ to find an optimal step size α. Obviously, $\psi(\alpha)$ achieves its maximum at

$$
\alpha_k = \frac{\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)}{\|d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\|_M^2},
$$

which is exactly (3.4b). Thanks to the inequality used in (4.7), it is natural to compensate $\Theta(\alpha)$ by introducing a relaxation factor γ for α_k, i.e.,

$$
\Theta(\gamma\alpha_k) \geq 2\gamma\alpha_k\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k) - \gamma^2\alpha_k^2 \left\|d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\|_M^2 = \gamma(2-\gamma)\alpha_k\varphi(\mathbf{u}^k, \widetilde{\mathbf{u}}^k).
$$

We shall restrict $\gamma \in (0, 2)$ so that the right-hand side in the above relation is positive, that is, an improvement can be obtained at each iteration. Below, we further prove that the optimal step size α_k is uniformly lower bounded away from a positive number.

Lemma 4.3. *Suppose that $\nu \in (0, 1)$, $\beta > 0$ and each r_i from $i = 1$ to m satisfies condition* (3.3)*. Then, the step size α_k defined by* (3.4b) *is lower bounded away from zero; that is, $\inf_{k\geq 1} \alpha_k \geq \alpha_{\min} > 0$ for some positive constant $\alpha_{\min}$.*

Proof. Recalling the definition of $d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)$ in (3.4c), it follows from the monotonicity of Φ and the Lipschitz continuity of each f_i that

$$
\begin{aligned}
\left\|d(\mathbf{u}^k, \widetilde{\mathbf{u}}^k)\right\|_M^2 &= \left\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\right\|_M^2 - 2\left\langle \mathbf{u}^k - \widetilde{\mathbf{u}}^k,\ \Phi(\mathbf{x}^k, \widetilde{\mathbf{x}}^k)\right\rangle + \left\|\Phi(\mathbf{x}^k, \widetilde{\mathbf{x}}^k)\right\|_{M^{-1}}^2 \\
&\leq \left\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\right\|_M^2 + \left\|\Phi(\mathbf{x}^k, \widetilde{\mathbf{x}}^k)\right\|_{M^{-1}}^2 \\
&\leq \sum_{i=1}^m \frac{(r_i^2 + L_{f_i}^2)}{r_i} \left\|x_i^k - \widetilde{x}_i^k\right\|^2 + \frac{1}{\beta}\left\|\lambda^k - \widetilde{\lambda}^k\right\|^2 \\
&\leq C\left\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\right\|^2,
\end{aligned}
$$

where

$$
C := \max\left\{\max_{1\leq i\leq m}\left\{\frac{(r_i^2 + L_{f_i}^2)}{r_i}\right\},\ \frac{1}{\beta}\right\}.
$$

Combining with the inequality (4.5), we immediately get $\alpha_k \geq \alpha_{\min} := c/C$. ◻

Theorem 4.4. *Assume $\nu \in (0, 1)$, $\beta > 0$ and condition* (3.3)*. Let $\mathbf{u}^* := (\mathbf{x}^*, \lambda^*)$ be an arbitrary solution of* SVI $(\mathcal{U}, H)$*. Then, the sequence $\{\mathbf{u}^k\}$ generated by the proposed* ProjSM *satisfies the property*

$$
\left\|\mathbf{u}^{k+1} - \mathbf{u}^*\right\|_M^2 \leq \left\|\mathbf{u}^k - \mathbf{u}^*\right\|_M^2 - \gamma(2-\gamma)c\alpha_{\min}\left\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\right\|^2. \qquad (4.8)
$$

Proof. By (3.5a) and Lemmas 4.1-4.3, we deduce that

$$\begin{aligned}
&\left\|\mathbf{u}^{k+1}-\mathbf{u}^*\right\|_M^2 \\
&=\left\|\mathbf{u}^k-\gamma\alpha_k d(\mathbf{u}^k,\widetilde{\mathbf{u}}^k)-\mathbf{u}^*\right\|_M^2 \\
&=\left\|\mathbf{u}^k-\mathbf{u}^*\right\|_M^2-2\gamma\alpha_k\left\langle \mathbf{u}^k-\mathbf{u}^*,\ Md(\mathbf{u}^k,\widetilde{\mathbf{u}}^k)\right\rangle+\gamma^2\alpha_k^2\left\|d(\mathbf{u}^k,\widetilde{\mathbf{u}}^k)\right\|_M^2 \\
&\le\left\|\mathbf{u}^k-\mathbf{u}^*\right\|_M^2-2\gamma\alpha_k\varphi(\mathbf{u}^k,\widetilde{\mathbf{u}}^k)+\gamma^2\alpha_k^2\left\|d(\mathbf{u}^k,\widetilde{\mathbf{u}}^k)\right\|_M^2 \\
&=\left\|\mathbf{u}^k-\mathbf{u}^*\right\|_M^2-\gamma(2-\gamma)\alpha_k\varphi(\mathbf{u}^k,\widetilde{\mathbf{u}}^k) \\
&\le\left\|\mathbf{u}^k-\mathbf{u}^*\right\|_M^2-\gamma(2-\gamma)c\alpha_{\min}\left\|\mathbf{u}^k-\widetilde{\mathbf{u}}^k\right\|^2.
\end{aligned}$$

The assertion of this lemma is proved. □

Remark 4.5. If we consider the variant correction step (3.5), we can similarly prove the Fejér monotonicity of the sequence $\{\mathbf{u}^k\}$ as follows:

$$\left\|\mathbf{u}^{k+1}-\mathbf{u}^*\right\|^2\le\left\|\mathbf{u}^k-\mathbf{u}^*\right\|^2-\gamma(2-\gamma)\hat{c}\widehat{\alpha}_{\min}\left\|\mathbf{u}^k-\widetilde{\mathbf{u}}^k\right\|^2.$$

Here $\hat{c}$ and $\widehat{\alpha}_{\min}$ are some positive constant, which can be deduced in similar ways used in Lemmas 4.2 and 4.3, respectively.

With the above results, we are now in a position to establish the global convergence of the proposed method.

Theorem 4.6. *Assume $\nu\in(0,1)$, $\beta>0$ and condition* (3.3). *Then the sequence $\{\mathbf{u}^k\}$ generated by the proposed* ProjSM *converges to a solution of* SVI $(\mathcal{U},H)$.

Proof. Let $\mathbf{u}^*$ be an arbitrary solution of SVI $(\mathcal{U},H)$. It is immediately clear from (4.8) that

$$\left\|\mathbf{u}^{k+1}-\mathbf{u}^*\right\|_M^2\le\left\|\mathbf{u}^k-\mathbf{u}^*\right\|_M^2\le\cdots\le\left\|\mathbf{u}^0-\mathbf{u}^*\right\|_M^2<\infty, \tag{4.9}$$

which implies that the sequence $\{\|\mathbf{u}^k-\mathbf{u}^*\|_M\}$ is decreasing, and particularly, it is bounded and moreover,

$$\lim_{k\to\infty}\|\mathbf{u}^k-\mathbf{u}^*\|_M \quad \text{exists.} \tag{4.10}$$

On the other hand, rewriting (4.8) as

$$\gamma(2-\gamma)c\alpha_{\min}\left\|\mathbf{u}^k-\widetilde{\mathbf{u}}^k\right\|^2\le\left\|\mathbf{u}^k-\mathbf{u}^*\right\|_M^2-\left\|\mathbf{u}^{k+1}-\mathbf{u}^*\right\|_M^2$$

and summing up over k, we arrive at

$$\begin{aligned}
\gamma(2-\gamma)c\alpha_{\min}\left(\sum_{k=0}^{\infty}\left\|\mathbf{u}^k-\widetilde{\mathbf{u}}^k\right\|^2\right)&\le\sum_{k=0}^{\infty}\left\{\left\|\mathbf{u}^k-\mathbf{u}^*\right\|_M^2-\left\|\mathbf{u}^{k+1}-\mathbf{u}^*\right\|_M^2\right\}\\
&\le\left\|\mathbf{u}^0-\mathbf{u}^*\right\|_M^2.
\end{aligned}$$

We immediately conclude that

$$\lim_{k\to\infty} \left\| \mathbf{u}^k - \widetilde{\mathbf{u}}^k \right\|^2 = 0.$$

It turns out that the sequences $\{\mathbf{u}^k\}$ and $\{\widetilde{\mathbf{u}}^k\}$ have the same cluster points. To prove the convergence of the sequence $\{\mathbf{u}^k\}$, let $\mathbf{u}^\infty$ be a cluster point of $\{\mathbf{u}^k\}$ and let $\{\mathbf{u}^{k_j}\}$ be a subsequence converging to $\mathbf{u}^\infty$; hence the subsequence $\{\widetilde{\mathbf{u}}^{k_j}\}$ of $\{\widetilde{\mathbf{u}}^k\}$ also converges to $\mathbf{u}^\infty$. Now, taking the limit as $j \to \infty$ over the subsequence $\{k_j\}$ in (3.2a)–(3.2c), we obtain

$$\begin{cases} x_i^\infty = P_{\mathcal{X}_i} \left\{ x_i^\infty - \frac{1}{r_i} \left(f_i(x_i^\infty) - A_i^\top \lambda^\infty \right) \right\} & (\text{ for } i = 1, \ldots, m), \\ \sum_{i=1}^m A_i x_i^\infty = b. \end{cases}$$

Equivalently, $E_{[1,M]}(\mathbf{u}^\infty, H, \mathcal{U}) = 0$ with the block diagonal matrix M given in (3.1). It then follows from Lemma (2.2) that $\mathbf{u}^\infty$ is a solution of SVI $(\mathcal{U}, H)$.

However, since $\mathbf{u}^* := (\mathbf{x}^*, \lambda^*)$ is an arbitrary solution of SVI $(\mathcal{U}, H)$, we can substitute $\mathbf{u}^\infty$ for $\mathbf{u}^*$ in (4.9) and (4.10) to result in

$$\lim_{k\to\infty} \|\mathbf{u}^k - \mathbf{u}^\infty\|_M = \lim_{j\to\infty} \|\mathbf{u}^{k_j} - \mathbf{u}^\infty\|_M = 0.$$

This shows that the full sequence $\{\mathbf{u}^k\}$ converges to $\mathbf{u}^\infty$, a solution of SVI $(\mathcal{U}, H)$. The proof is complete. □

5. Applications

As we have mentioned in Section 1, variational inequality problem is a fundamental model for treating a wide range of real world problems. In this section, we study some special cases of VIP (1.1) and demonstrate the applicability of the proposed ProjSM to these problems.

5.1. **Generalized split equality problem.** We first consider the *generalized split equality problem* (GSEP) which is stated as

$$\text{(5.1)} \qquad \text{finding} \quad \mathbf{x} := (x_1, \ldots, x_m) \in \prod_{i=1}^m \mathcal{X}_i \quad \text{satisfying} \quad \sum_{i=1}^m A_i x_i = b,$$

where each $\mathcal{X}_i$ $(i = 1, \ldots, m)$ is a nonempty closed convex subset of $\mathbb{R}^{n_i}$; $A_i : \mathcal{X}_i \to \mathbb{R}^l$ $(i = 1, \ldots, m)$ is a bounded linear operator and $b \in \mathbb{R}^l$ is a given vector. Note that the GESP is an extension of the recently introduced *split equality problem* (SEP) in [6, 34], and it has widespread applications in decomposition method in PDEs, game theory and intensity-modulated radiation therapy. On the other hand, the GSEP (5.1) further provides us with a model for unifying some classical problems. Here, we summarize several well-known examples as follows:

5.1.1. *Split feasibility problem.* The *split feasibility problem* (SFP), which was first introduced in [8] for modeling inverse problems arising from phase retrievals and medical image reconstruction [7], can be mathematically characterized as finding a point $x_1 \in \mathbb{R}^n$ with the property

$$\text{(5.2)} \qquad x_1 \in \mathcal{C} \quad \text{and} \quad Ax_1 \in \mathcal{Q},$$

where $\mathcal{C}$ and $\mathcal{Q}$ are nonempty closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and $A : \mathbb{R}^n \to \mathbb{R}^l$ is a bounded linear operator. In the past decades, the SFP (5.2) and its special case of *multiple-set split feasibility problem* have been received considerable attention in terms of numerical algorithm, e.g., see [5, 31, 39, 42, 43, 45, 46]. By introducing an auxiliary variable $x_2 \in \mathbb{R}^m$, the SFP (5.2) can be immediately recast as:

$$\text{(5.3)} \qquad \text{finding} \quad x_1 \in \mathcal{C},\ x_2 \in \mathcal{Q} \quad \text{such that} \quad Ax_1 = x_2,$$

which is an intuitive special case of GSEP (5.1) by taking $m = 2$, $\mathcal{X}_1 := \mathcal{C}$, $\mathcal{X}_2 := \mathcal{Q}$, $A_1 := A$, $A_2 := -I$, and $b := 0$.

5.1.2. *Split equality problem.* The SEP is a newly introduced model to find two points satisfying a linear equality, that is,

$$\text{(5.4)} \qquad \text{finding} \quad x_1 \in \mathcal{C},\ x_2 \in \mathcal{Q} \quad \text{such that} \quad Ax_1 = Bx_2.$$

Clearly, setting $B := I_n$ in (5.4) immediately yields SFP (5.3) as a special case of GSEP (5.1). Moreover, SEP (5.4) is also a specialization of GSEP (5.1) by taking the same settings in (5.3) except $A_2 := -B$.

5.1.3. *Convex feasibility problem.* The well known *convex feasibility problem* (CFP), which consists of finding a point in the intersection of convex sets, is one of the most classical problems in the communities of physical sciences, statistics and image reconstruction. Mathematically, it reads as follows:

$$\text{(5.5)} \qquad \text{finding a point} \quad x \in \bigcap_{i=1}^{m} \mathcal{X}_i \neq \varnothing,$$

where every $\mathcal{X}_i$ $(i = 1, \dots, m)$ is a closed convex subset of $\mathbb{R}^n$. To solve this problem, projection-like algorithms have been well developed in the literature, e.g., see [2] for a review on the CFP. Below, we show that CFP (5.5) can be viewed as a standard GSEP (5.1). By introducing m auxiliary variables x_i $(i = 1, \dots, m)$, then CFP (5.5) can be further recast as finding m points $(x_1, \dots, x_m)$ with the property

$$(x_1, \dots, x_m) \in \prod_{i=1}^{m} \mathcal{X}_i, \;\; x_1 - x_2 = 0,\ x_2 - x_3 = 0,\ \dots,\ x_m - x_1 = 0.$$

Equivalently, we further rewrite it into

$$\text{(5.6)} \qquad \text{finding} \quad \mathbf{x} := (x_1, \dots, x_m) \in \prod_{i=1}^{m} \mathcal{X}_i \quad \text{such that} \quad \sum_{i=1}^{m} A_i x_i = 0,$$

where A_i is the i-th column of the block matrix A defined as follows:

$$\text{(5.7)} \qquad A := (A_1, A_2, \dots, A_{m-1}, A_m) := \begin{pmatrix} I_n & -I_n & \dots & 0 & 0 \\ 0 & I_n & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_n & -I_n \\ -I_n & 0 & \dots & 0 & I_n \end{pmatrix}.$$

It is clear that CFP (5.6) is a specialization of GSEP (5.1) by taking $b := 0$. As shown in [9], certainly, CFP (5.5) can also be regarded as a special case of SFP

(5.2) by setting $x_1 := x$, $\mathcal{C} := \cap_{i=1}^m \mathcal{X}_i$, $\mathcal{Q} := \mathbb{R}^n$ and $A := I_n$. Compared to the latter case, a noteworthy benefit of (5.6) is that the new reformulation makes the projections onto $\mathcal{X}_i$ $(i = 1, \dots, m)$ are more implementable than the projection onto $\cap_{i=1}^m \mathcal{X}_i$.

5.1.4. *Implementation on GSEP.* From the above reformulations, it can be easily seen that GSEP (5.1) is an interesting model for unifying some classical problems. Indeed, we further see that GSEP (5.1) is a special case of VIP (1.1) by taking

$$f_1(x_1) = f_2(x_2) = \cdots = f_m(x_m) = 0.$$

Accordingly, when applying the proposed ProjSM to solve GSEP (5.1), we can simplify the iterative schemes (3.2) and (3.4) as follows:

(**Initialization**). Given $\gamma \in (0, 2)$, $\nu \in (0, 1)$, $\beta > 0$, $\mathbf{u}^0 \in \mathcal{U}$.
(**Prediction step**). Obtain $\widetilde{\mathbf{u}}^k = (\widetilde{\mathbf{x}}^k, \widetilde{\lambda}^k)$ with suitable $r_i > 0$ via:

$$\begin{cases} \widetilde{x}_i^k := P_{\mathcal{X}_i}\left\{x_i^k + \frac{1}{r_i}A_i^\top\left[\lambda^k - \beta\left(\sum_{i=1}^m A_i x_i^k - b\right)\right]\right\}, \ (i = 1, \dots, m), \\ \widetilde{\lambda}^k = \lambda^k - \beta\left(\sum_{i=1}^m A_i \widetilde{x}_i^k - b\right). \end{cases}$$

(**Correction step**). Update $\mathbf{u}^{k+1} := (\mathbf{x}^{k+1}, \lambda^{k+1})$ via

$$\mathbf{u}^{k+1} := \mathbf{u}^k - \gamma\alpha_k(\mathbf{u}^k - \widetilde{\mathbf{u}}^k), \tag{5.8}$$

where

$$\alpha_k := 1 + \frac{\left\langle \lambda^k - \widetilde{\lambda}^k,\ \sum_{i=1}^m A_i(x_i^k - \widetilde{x}_i^k)\right\rangle}{\|\mathbf{u}^k - \widetilde{\mathbf{u}}^k\|_M^2}.$$

From the correction step (5.8), we can see that the new iterate $\mathbf{u}^{k+1}$ essentially is an affine combination of the predictor $\widetilde{\mathbf{u}}^k$ and the last point $\mathbf{u}^k$, which potentially brings promising numerical performance. For the choice of r_i in the prediction step, condition (3.3) amounts to

$$r_i \geq \frac{m+2}{4\nu}\beta\|A_i^\top A_i\|, \quad (i = 1, \dots, m).$$

Below, we respectively give remarks to highlight the superiority of the proposed ProjSM.

- When considering SFP (5.2), we have $m = 1$. By setting $\nu = \frac{3}{4}$, we obtain
$$r_1 \geq \beta\|A_1^\top A_1\|,$$
and this condition is more flexible than the requirement of CQ algorithm in [5].
- When applying ProjSM to solve SEP (5.4), we take $m = 2$ and get
$$r_1 \geq \frac{\beta}{\nu}\|A_1^\top A_1\| \quad \text{and} \quad r_2 \geq \frac{\beta}{\nu}\|A_2^\top A_2\|,$$
which is also weaker than the requirements in [34, 35].
- To handle the CFP (5.5), one of the most popular methods is the *alternating projection algorithm* (APA) [2]. The APA is simple; however, its sequential projection philosophy may be time consuming if $\mathcal{X}_i$'s are not easy enough

to compute their projections. It is clear that our ProjSM could make full use of modern parallel computers to compute the projections onto $\mathcal{X}_i$'s simultaneously.

5.2. **Network resource allocation problem.** In this section, we consider the *network resource allocation problem* (NRAP) modeling in power [26, 33], channel [1, 29], bandwidth [37] and storage allocations [32], which can be characterized as

$$\max \left\{ \sum_{i=1}^{m} \theta_i(x) \;\Big|\; x \in \mathcal{X} := \bigcap_{i=1}^{m} \mathcal{X}_i \right\}, \tag{5.9}$$

where $\theta_i(\cdot) : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \ldots, m)$ are concave and continuously Fréchet differentiable utility functions; $\mathcal{X}_i \subset \mathbb{R}^n$ $(i = 1, \ldots, m)$ are nonempty closed convex sets. The NRAP often has been studied as a VI thanks to the rich set of efficient solvers for VIP. However, the direct application of traditional projection-like methods tailored for VIP to (5.9) fails to be implementable for the reason that it is not an easy task to compute the projection onto the set $\mathcal{X} := \cap_{i=1}^{m} \mathcal{X}_i$, even when $\mathcal{X}_i$'s are simple enough so that projections (i.e., $P_{\mathcal{X}_i}(\cdot)$'s) have explicit representations. To circumvent this difficulty, Iiduka [27, 28] judiciously proposed a series of *decentralized* projection algorithms for (5.9). However, these methods have strong requirements, that is, the utility functions (i.e., $\theta_i(\cdot)$'s) are *strongly concave* and their gradients are Lipschitz continuous. Such strong conditions may preclude the potential applications of these methods. Indeed, it is interesting to observe that NRAP (5.9) can be viewed as a generalization of CFP (5.5) with an addition objective function $\sum_{i=1}^{m} \theta_i(\cdot)$. Similarly, by introducing m variables x_i $(i = 1, \ldots, m)$, then the model (5.9) can be reformulated as

$$\min \left\{ \sum_{i=1}^{m} -\theta_i(x_i) \;\Big|\; \sum_{i=1}^{m} A_i x_i = 0;\; x_i \in \mathcal{X}_i,\; (i = 1, \ldots, m) \right\}, \tag{5.10}$$

where the A_i's are defined in (5.7). Since the $\theta_i(x_i)$'s are concave and differentiable, letting $f_i(x_i) := -\nabla\theta_i(x_i)$, $(i = 1, \ldots, m)$ in (1.1b), we immediately put (5.10) into a special case of VI (1.1). Hence, we can gainfully employ the proposed ProjSM ((3.2) and (3.4)) to solve (5.10) directly. Note that our ProjSM only requires that $\theta_i(\cdot)$'s be concave and their gradients be Lipschitz continuous, which are relatively weaker than the assumptions of the aforementioned methods. Another noteworthy feature of our ProjSM is that the new method can also be regarded as a *decentralized* algorithm.

6. Conclusions

We consider a well-structured VIP, where the underlying mapping is separable into m individual parts and the feasible set contains a linear constraint and m simple convex sets. To fully exploit the separable structure, we propose a projection-based splitting method, which consists of a parallel prediction step and a cheap correction step. The method is eligible for solving large-scale problems, due to the fact that its prediction step can be implemented simultaneously. Our new method is globally convergent under some mild assumptions. Finally, we introduce a GSEP, which includes some well-known problems such as SFP, SEP, and CFP as special cases. In addition, we formulate the NRAP as a specialization of the structured VI (1.1). The applicability of our method to these problems is demonstrated. When considering complicated sets $\mathcal{X}_i$'s in the sense that the projections onto them (i.e., $P_{\mathcal{X}_i}(\cdot)$'s) can not be explicitly calculated, we can adopt the strategies introduced in [31, 44] to overcome the difficulty, which is also our future work.

References

[1] R. Agrawal and V. Subramanian, *Optimality of certain channel aware scheduling policies*, in Proceedings of the Annual Allerton Conference on Communication Control and Computing, volume 40, 2002, pp. 1533–1542.

[2] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev. **38** (1996), 367–426.

[3] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation, Numerical Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[4] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn. **3** (2010), 1–122.

[5] C. L. Byrne, *Iterative oblique projection onto convex sets and the split feasibility problem*, Inverse Problems **18** (2002), 441–453.

[6] C. L. Byrne and A. Moudafi, *Extensions of the CQ algorithm for the split feasibility and split equality problems*, submitted to J. Nonlinear Convex Anal., 2013.

[7] Y. Censor, T. Bortfeld, B. Martin and A. Trofimov, *A unified approach for inversion problems in intensity-modulated radiation therapy*, Phys. Med. Biol. **51** (2006), 2353–2365.

[8] Y. Censor and T. Elfving, *A multiprojection algorithm using Bregman projections in product space*, Numer. Algor. **8** (1994), 221–239.

[9] Y. Censor, A. Gibali and S. Reich, *Algorithms for the split variational inequality problem*, Numer. Algor. **59** (2012), 301–323.

[10] R. H. Chan, J. F. Yang and X. M. Yuan, *Alternating direction method for image inpainting in wavelet domain*, SIAM J. Imaging Sci. **4** (2011), 807–826.

[11] 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 necessarily convergent*, Math. Program., DOI: 10.1007/s10107-014-0826-5.

[12] S. Dafermos, *Traffic equilibrium and variational inequalities*, Transport. Sci. **14** (1980), 42–54.

[13] Q. T. Dinh, I. Necoara, C. Savorgnan and M. Diehl, *An inexact perturbed path-following method for Lagrangian decomposition in large-scale separable convex optimization*, SIAM J. Optim. **23** (2013), 95–125.

[14] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[15] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems, : M. Fortin and R. Glowinski (eds), The Netherlands, 1983, pp. 299–331.

[16] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Comput. Math. Appl. **2** (1976), 16–40.

[17] D. R. Han, H. J. He, H. Yang and X. M. Yuan, *A customized Douglas-Rachford splitting algorithm for separable convex minimization with linear constraints*, Numer. Math. **127** (2014), 167–200.

[18] D. R. Han, X. M. Yuan and W. X. Zhang, *An augmented-Lagrangian-based parallel splitting method for separable convex minimization with applications to image processing*, Math. Comput. **83** (2014), 2263–2291.

[19] B. S. He, *Parallel splitting augmented Lagrangian methods for monotone structured variational inequalities*, Comput. Optim. Appl. **42** (2009), 195–212.

[20] B. S. He, L. S. Hou and X. M. Yuan, *On full Jacobian decomposition of the augmented Lagrangian method for separable convex programming*, available at Optimization Online, 2013.

[21] B. S. He, L. Z. Liao, D. R. Han and H. Yang, *A new inexact alternating direction method for monotone variational inequalities*, Math. Program. **92** (2002), 103–118.

[22] B. S. He, L. Z. Liao and M. J. Qian, *Alternating projection based prediction-correction methods for structured variational inequalities*, J. Comput. Math. **24** (2006), 693–710.

[23] B. S. He, M. Tao and X. M. Yuan, *Alternating direction method with Gaussian-back substitution for separable convex programming*, SIAM J. Optimiz. **22** (2012), 313–340.

[24] H. J. He, D. R. Han, W. Y. Sun and Y. N. Chen, *A hybrid splitting method for variational inequality problems with separable structure*, Optim. Method Softw. **28** (2013), 725–742.

[25] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl. **4** (1969), 303–320.

[26] H. Iiduka, *Fixed point optimization algorithm and its application to power control in cdma data networks*, Math. Program. **133** (2012), 227–242.

[27] H. Iiduka, *Unicast decentralized algorithm for solving centralized optimization problems in network resource allocation*, J. Nonlinear Convex Anal. **13** (2012), 31–48.

[28] H. Iiduka, *Multicast decentralized optimization algorithm for network resource allocation problems*, J. Nonlinear Convex Anal. **14** (2013), 817–839.

[29] F. P. Kelly, A. K. Maulloo and D. K. H. Tan, *Rate control for communication networks: shadow prices, proportional fairness and stability*, J. Oper. Res. Soc. **49** (1998), 237–252.

[30] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems*, Matecon **12** (1976), 747–756.

[31] G. López, V. Martin-Marquez, F. H. Wang and H. K. Xu, *Solving the split feasibility problem without prior knowledge of matrix norms*, Inverse Problems **28** (2012), 085004(18pp).

[32] P. Maillé and L. Toka, *Managing a peer-to-peer data storage system in a selfish society*, IEEE J. Sel. Areas Commun. **26** (2008), 1295–1301.

[33] F. Meshkati, H. V. Poor, S. C. Schwartz and N. B. Mandayam, *An energy-efficient approach to power control and receiver design in wireless data networks*, IEEE Trans. Commun **53** (2005), 1885–1894.

[34] A. Moudafi, *A relaxed alternating CQ-algorithms for convex feasibility problems*, Nonlinear Anal. **79** (2013), 117–121.

[35] A. Moudafi and E. Al-Shemas, *Simultaneous iterative methods for split equality problem*, Trans. Math. Program. Appl **1** (2013), 1–11.

[36] A. Nagurney, *Network Economics: A Variational Inequality Approach*, Advances in Computational Economics, Kluwer Academic, Boston, 1993.

[37] A. Nedić and A. Ozdaglar, *Cooperative distributed multi-agent optimization*, in Convex Optimization in Signal Processing and Communications, D.P. Palomar and Y.C. Eldar (eds.), Cambridge University Press, New York, 2010, pp. 340–386.

[38] O. Oreifej, X. Li and M. Shah, *Simultaneous video stabilization and moving object detection in turbulence*, IEEE Trans. Pattern Anal. Mach. Intell. **35** (2013), 450–462.

[39] B. Qu and N. H. Xiu, *A note on the CQ algorithm for the split feasibility problem*, Inverse Problems **21** (2005), 1655–1665.

[40] M. Tao and X.M. Yuan, *Recovering low-rank and sparse components of matrices from incomplete and noisy observations*, SIAM J. Optim. **21** (2011), 57–81.

[41] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu and K. Knight, *Sparsity and smoothness via the fused LASSO*, J. R. Stat. Soc. Ser. B **67** (2005), 91–108.

[42] H. K. Xu, *A variable Krasnosel'skii-Mann algorithm and the multiple-set split feasibility problem*, Inverse Problems **22** (2006), 2021–2034.

[43] H. K. Xu, *Iterative methods for the split feasibility problem in infinite dimensional Hilbert spaces*, Inverse Problems **26** (2010), 105018(17pp).

[44] Q. Z. Yang, *The relaxed CQ algorithm solving the split feasibility problem*, Inverse Problems **20** (2004), 1261–1266.

[45] W. X. Zhang, D. R. Han and Z. B. Li, *A self-adaptive projection method for solving the multiple-sets split feasibility problem*, Inverse Problems **25** (2009), 115001.

[46] W. X. Zhang, D. R. Han and X. M. Yuan, *An efficient simultaneous method for the constrained multiple-sets split feasibility problem*, Comput. Optim. Appl. **52** (2012), 825–843.

Manuscript received May 10, 2014
revised July 27, 2014

H. He
Department of Mathematics, School of Sciences, Hangzhou Dianzi University, 310018, Hangzhou, China
E-mail address: hehjmath@hdu.edu.cn

H. K. Xu
Department of Mathematics, School of Science, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China; Department of Applied Mathematics, National Sun Yat-Sen University, Kaohsiung, Taiwan 80424
E-mail address: xuhk@math.nsysu.edu.tw