Journal of Computational Mathematics, Vol. 24, No. 6, 2006, 693–710.

ALTERNATING PROJECTION BASED PREDICTION-CORRECTION METHODS FOR STRUCTURED VARIATIONAL INEQUALITIES *1)

Bing-sheng He
(*Department of Mathematics, Nanjing University, Nanjing 210093, China*)
Li-zhi Liao
(*Department of Mathematics, Hong Kong Baptist University, Hong Kong, China*)
Mai-jian Qian
(*Department of Mathematics, California State University, Fullerton CA 92834, USA*)

Abstract

The monotone variational inequalities VI(Ω, F) have vast applications, including optimal controls and convex programming. In this paper we focus on the VI problems that have a particular splitting structure and in which the mapping F does not have an explicit form, therefore only its function values can be employed in the numerical methods for solving such problems. We study a set of numerical methods that are easily implementable. Each iteration of the proposed methods consists of two procedures. The first (prediction) procedure utilizes alternating projections to produce a predictor. The second (correction) procedure generates the new iterate via some minor computations. Convergence of the proposed methods is proved under mild conditions. Preliminary numerical experiments for some traffic equilibrium problems illustrate the effectiveness of the proposed methods.

Mathematics subject classification: 65K10, 90C25, 90C30.
Key words: Structured variational inequality, Monotonicity, Prediction-correction method.

1. Introduction

A variational inequality problem, denoted by VI(Ω, F), is to find a vector $u^* \in \Omega$ such that

$$(u - u^*)^T F(u^*) \geq 0, \quad \forall\, u \in \Omega, \tag{1.1}$$

where Ω is a nonempty closed convex subset of $\mathbb{R}^l$, and F is a mapping from $\mathbb{R}^l$ into itself. In this paper, we consider the VI problem with the following structure:

$$(x^*, y^*) \in \mathcal{D}, \quad \left\{ \begin{array}{l} (x - x^*)^T f(x^*) \geq 0, \\ (y - y^*)^T g(y^*) \geq 0, \end{array} \right. \qquad \forall\, (x, y) \in \mathcal{D}, \tag{1.2}$$

where

$$\mathcal{D} = \{(x, y) | x \in \mathcal{X},\ y \in \mathcal{Y},\ Ax + By = b\}, \tag{1.3}$$

$\mathcal{X}$ and $\mathcal{Y}$ are given nonempty closed convex subsets of $\mathbb{R}^n$ and $\mathbb{R}^p$, respectively, $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times p}$ are given matrices, $b \in \mathbb{R}^m$ is a given vector, $f : \mathcal{X} \to \mathbb{R}^n$ and $g : \mathcal{Y} \to \mathbb{R}^p$ are monotone operators. Problem (1.2)-(1.3) is a special case of the general VI problem (1.1), which

* Received December 9, 2005.
1) The research of the first author is supported by NSFC Grant 10571083. The research of the second author is supported in part by grants from Hong Kong Baptist University and the Research Grant Council of Hong Kong.

has numerous important applications, including applications in the fields of optimal controls and convex programming (see [1, 6, 7]).

Since in practice such problems usually involve large number of variables, numerical methods that can make use of the decomposed structure of problem (1.2)-(1.3) can greatly save computer storage as well as computing time. A number of decomposition methods have been proposed, for examples, see [3, 4, 5, 7, 8, 9, 15].

In many applications, the mapping f (resp. g) cannot be expressed explicitly and for a given $x \in \mathcal{X}$ (resp. $y \in \mathcal{Y}$), the function value $f(x)$ (resp. $g(y)$) can only be obtained via certain procedures. Given a variable value, the evaluation of f or g can be costly and time-consuming, and sometimes may pose social or political impact (such as posing toll charges to evaluate the traffic flow), therefore should not be taken lightly. In such applications, efficient numerical methods which only employ function values are highly desired.

Among all the existing decomposition methods which achieve linear convergence, in each iteration a subproblem equivalent to an implicit projection calls to be solved, as illustrated below. Solving each subproblem usually requires numerous function evaluations. In this paper we present a set of decomposition methods that involve only explicit projections, therefore require only one function evaluation in each iteration, yet they also yield linear convergence. The numerical experiments presented in Section 6 illustrate the effectiveness of the methods.

The proposed methods are motivated by the existing proximal alternating directions methods (abbreviated as PADMs) proposed in [15]. We briefly describe the PADMs as follows: First, by attaching a Lagrange multiplier vector $\lambda \in \mathbb{R}^m$ to the linear constraint $Ax + By = b$, the VI problem (1.2)-(1.3) is converted into the following equivalent non-constrained form:

$$(x^*, y^*, \lambda^*) \in \mathcal{W}, \quad \begin{cases} (x - x^*)^T (f(x^*) - A^T\lambda^*) \geq 0, \\ (y - y^*)^T (g(y^*) - B^T\lambda^*) \geq 0, \\ (\lambda - \lambda^*)^T (Ax^* + By^* - b) \geq 0, \end{cases} \quad \forall\, (x, y, \lambda) \in \mathcal{W} \tag{1.4}$$

where

$$\mathcal{W} = \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m. \tag{1.5}$$

We denote VI problem (1.4)-(1.5) by VI$(\mathcal{W}, Q)$, where

$$Q(w) = Q(x, y, \lambda) = \begin{pmatrix} f(x) - A^T\lambda \\ g(y) - B^T\lambda \\ Ax + By - b \end{pmatrix}. \tag{1.6}$$

Given a triplet $w^k = (x^k, y^k, \lambda^k) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$, the PADMs generate a new iterate $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$ via the following general procedure:
Given $(x^k, y^k, \lambda^k) \in \mathcal{W}$, first find an $\tilde{x}^k \in \mathcal{X}$ such that

$$(x' - \tilde{x}^k)^T \{f(\tilde{x}^k) - A^T[\lambda^k - \beta(A\tilde{x}^k + By^k - b)] + r(\tilde{x}^k - x^k)\} \geq 0, \quad \forall\, x' \in \mathcal{X}. \tag{1.7}$$

Then find a $\tilde{y}^k \in \mathcal{Y}$ such that

$$(y' - \tilde{y}^k)^T \{g(\tilde{y}^k) - B^T[\lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b)] + s(\tilde{y}^k - y^k)\} \geq 0, \quad \forall\, y' \in \mathcal{Y}. \tag{1.8}$$

Finally, update $\tilde{\lambda}^k$ via

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \tag{1.9}$$

Here $\beta > 0$ is a given *penalty parameter* of the linear constraint $Ax + By - b = 0$. The coefficients $r > 0$ and $s > 0$ in formulas (1.7) and (1.8) respectively are referred to as *proximal parameters*. The method is convergent by taking $w^{k+1} = \tilde{w}^k$ (for a proof see [12]).

Note that the solutions of subproblems (1.7) and (1.8) are equivalent to the solutions of the following projection equations (for details see Lemma 2.1 in Section 2):

$$\tilde{x}^k := P_{\mathcal{X}}\{x^k - \frac{1}{r}\big[f(\tilde{x}^k) - A^T[\lambda^k - \beta(A\tilde{x}^k + By^k - b)]\big]\} \tag{1.10}$$

and

$$\tilde{y}^k := P_{\mathcal{Y}}\{y^k - \frac{1}{s}\big[g(\tilde{y}^k) - B^T[\lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b)]\big]\}. \tag{1.11}$$

Solving the projections (1.10)-(1.11) exactly is not an easy task, since each is an implicit projection.

For those problems in which only the function value is available, in order to overcome the difficulty of solving the projections (1.10)-(1.11) directly, a natural simple idea is to apply a Gauss-Seidel type of step. By setting the known vectors in the unknown seat of the right-hand-sides of the (1.10)-(1.11), one obtains the new iteration triplet $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$ via the following alternating procedure with explicit projections:

$$\tilde{x}^k := P_{\mathcal{X}}\{x^k - \frac{1}{r}\big[f(x^k) - A^T[\lambda^k - \beta(Ax^k + By^k - b)]\big]\}, \tag{1.12}$$

$$\tilde{y}^k := P_{\mathcal{Y}}\{y^k - \frac{1}{s}\big[g(y^k) - B^T[\lambda^k - \beta(A\tilde{x}^k + By^k - b)]\big]\} \tag{1.13}$$

and

$$\tilde{\lambda}^k = \lambda^k - \beta(A\tilde{x}^k + B\tilde{y}^k - b). \tag{1.14}$$

Unfortunately, this method does not guarantee convergence if one simply takes $w^{k+1} = \tilde{w}^k$ as in some inexact methods [11]. In this paper, we present a set of methods that take $\tilde{w}^k$ produced by (1.12)-(1.14) as a predictor. The new iterate $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ is then generated by a minor correction to the predictor. Since the main labor of the proposed methods is the alternating projection, they are referred to as *alternating projection based prediction-correction methods*.

The rest of this paper is organized as follows: In Section 2 we summarize some preliminaries of variational inequalities. In Section 3 we present our methods and illustrate that our methods can be easily implemented. The main theorem of the proposed methods is proved in Section 4. In Section 5 we investigate some contractive properties of the iterates and prove the convergence. Preliminary numerical results for network equilibrium problems are reported in Section 6. Finally, some concluding remarks are drawn in Section 7.

2. Preliminaries

In this section, we summarize some basic properties and related definitions that will be used in the following discussions. Let G be a positive definite matrix, we denote $\|v\|_G = \sqrt{v^T G v}$ as the G-norm of vector v. Let Ω be a nonempty closed convex subset of $\mathbb{R}^l$. The projection under G-norm will be denoted by $P_{\Omega,G}(\cdot)$, i.e.,

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

From the above definition, it follows that

$$(v - P_{\Omega,G}(v))^T G(u - P_{\Omega,G}(v)) \le 0, \quad \forall v \in \mathbb{R}^l, \forall u \in \Omega. \tag{2.1}$$

Consequently, we have

$$\|P_{\Omega,G}(v) - P_{\Omega,G}(w)\|_G \le \|v-w\|_G, \qquad \forall v, w \in \mathbb{R}^l \tag{2.2}$$

and

$$\|u - P_{\Omega,G}(v)\|_G^2 \le \|v - u\|_G^2 - \|v - P_{\Omega,G}(v)\|_G^2, \qquad \forall v \in \mathbb{R}^l, \forall u \in \Omega. \tag{2.3}$$

Definition 2.1. *a). F is said to be monotone if*

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

b). F is strongly monotone if there exists a constant $\mu > 0$ such that

$$(u - v)^T (F(u) - F(v)) \ge \mu \|u - v\|^2, \quad \forall u, v \in \Omega.$$

c). f is Lipschitz continuous with respect to $\mathcal{X}$ if there exists a constant $L_f > 0$ such that

$$\|f(x) - f(\tilde{x})\| \le L_f \|x - \tilde{x}\|, \quad \forall\, x, \tilde{x} \in \mathcal{X}.$$

Lemma 2.1. *Let Ω be a closed convex set in $\mathbb{R}^l$ and G be any positive definite matrix. Then u^* is a solution of VI(Ω, F) if and only if*

$$u^* = P_{\Omega,G}[u^* - \alpha G^{-1} F(u^*)], \qquad \forall\, \alpha > 0. \tag{2.4}$$

Proof. See ([2], p. 267).

According to Lemma 2.1, for any positive definite matrix $G \in \mathbb{R}^{l \times l}$, $p \in \mathbb{R}^l$ and $\alpha > 0$,

$$u^* = P_{\Omega,G}[u^* - \alpha G^{-1} p] \qquad \text{is equivalent to} \qquad u^* \in \Omega, \quad (u - u^*)^T p \ge 0, \quad \forall\, u \in \Omega. \tag{2.5}$$

Moreover,

$$\text{if} \quad \tilde{u} = P_\Omega[u - p], \quad \text{then} \quad \tilde{u} = P_\Omega\{\tilde{u} - [(\tilde{u} - u) + p]\} = P_{\Omega,G}\{\tilde{u} - G^{-1}[(\tilde{u} - u) + p]\}. \tag{2.6}$$

Lemma 2.2. *The VI$(\mathcal{W}, Q)$ problem (1.4)-(1.5) can be equivalently solved by seeking a zero point of the mapping*

$$e(w, \mathcal{W}, Q) := \begin{pmatrix} e_x(w) \\ e_y(w) \\ e_\lambda(w) \end{pmatrix} = \begin{pmatrix} x - P_{\mathcal{X}}\{x - [f(x) - A^T\lambda]\} \\ y - P_{\mathcal{Y}}\{y - [g(y) - B^T\lambda]\} \\ Ax + By - b \end{pmatrix}. \tag{2.7}$$

Throughout this paper, we make the following standard assumptions:

Assumption A:

A1. $\mathcal{X}$ and $\mathcal{Y}$ are *simple* closed convex sets. Here a set is said to be simple means that the projection onto the set is simple to carry out, for example, the nonnegative orthant, a ball or a box.

A2. $f(x)$ (resp. $g(y)$) is monotone and Lipschitz continuous with respect to $\mathcal{X}$ (resp. $\mathcal{Y}$). L_f (resp. L_g) is the Lipschitz constant of mapping f (resp. g).

A3. The solution set of VI$(\mathcal{W}, Q)$, denoted by $\mathcal{W}^*$, is nonempty.

Because f and g are monotone mappings and $\mathcal{X}$ and $\mathcal{Y}$ are closed convex sets, the mapping $Q(w)$ is monotone on $\mathcal{W}$ and the solution set $\mathcal{W}^*$ of VI$(\mathcal{W}, Q)$ is closed and convex. For any $w \in \mathcal{W}$, we denote the Euclidean distance from w to $\mathcal{W}^*$ by

$$\text{dist}(w, \mathcal{W}^*) := \min\{\|w - w^*\| \mid w^* \in \mathcal{W}^*\}\ .$$

It is clear that

$$\text{dist}(w, \mathcal{W}^*) = 0 \quad \Longleftrightarrow \quad e(w, \mathcal{W}, Q) = 0.$$

Historically, the term $\|e(w, \mathcal{W}, Q)\|$ is referred to as the error bound of VI$(\mathcal{W}, Q)$, since it measures the magnitude of w being away from the solution set $\mathcal{W}^*$.

3. The Framework of the Proposed Methods

Each iteration of the proposed methods consists of two procedures. Given $(x^k, y^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ in the k-th iteration, the first procedure is a Gauss-Seidel type explicit projection which generates a predictor $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$. The second procedure produces the new iterate $(x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ via some simple computations. We let $\nu \in (0,1)$ be a given constant, H be a given proper symmetric positive definite matrix.

3.1. Prediction

Given triplet $(x^k, y^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, the predictor $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$ is generated by the following procedure:

Step 1. *Set*

$$\tilde{x}^k := P_{\mathcal{X}}\Big\{x^k - \frac{1}{r_k}\Big(f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)]\Big)\Big\}, \tag{3.1}$$

where $r_k > 0$ is a chosen parameter such that

$$\|\xi_x^k\| \le \nu r_k \|x^k - \tilde{x}^k\|, \qquad \xi_x^k := f(x^k) - f(\tilde{x}^k) + A^T HA(x^k - \tilde{x}^k). \tag{3.2}$$

Step 2. *Set*

$$\tilde{y}^k := P_{\mathcal{Y}}\Big\{y^k - \frac{1}{s_k}\Big(g(y^k) - B^T[\lambda^k - H(A\tilde{x}^k + By^k - b)]\Big)\Big\} \tag{3.3}$$

where $s_k > 0$ is a chosen parameter such that

$$\|\xi_y^k\| \le \nu s_k \|y^k - \tilde{y}^k\|, \qquad \xi_y^k := g(y^k) - g(\tilde{y}^k) + B^T HB(y^k - \tilde{y}^k). \tag{3.4}$$

Step 3. *Update $\tilde{\lambda}^k$ via*

$$\tilde{\lambda}^k = \lambda^k - H(A\tilde{x}^k + B\tilde{y}^k - b). \tag{3.5}$$

The details for finding suitable sequences $\{r_k\}$ and $\{s_k\}$ will be discussed later. Since the predictor $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathcal{W}$ is produced in the following order,

- obtain $\tilde{x}^k \in \mathcal{X}$ from given $(x^k, y^k, \lambda^k) \in \mathcal{W}$;
- obtain $\tilde{y}^k \in \mathcal{Y}$ from given $(\tilde{x}^k, y^k, \lambda^k) \in \mathcal{W}$;
- update $\tilde{\lambda}^k \in \mathbb{R}^m$ from given $(\tilde{x}^k, \tilde{y}^k, , \lambda^k) \in \mathcal{W}$,

this prediction procedure adopts the new information whenever possible. This step somewhat resembles the projection step proposed in [3]. However, the projection step in [3] can be viewed as a Jacobi type, while the prediction process proposed here can be viewed as a Gauss-Seidel type. The main task in the prediction step is to obtain the definitive projections. This process only requires the function values $f(x^k)$ and $g(y^k)$.

Remark 3.1. With a proper large scalar r (resp. s), we can obtain a pair of $\tilde{x}^k$ and ξ_x^k (resp. $\tilde{y}^k$ and ξ_y^k) to satisfy (3.2) (resp. (3.4)). In fact, for any

$$r \ge \frac{L_f + \|A^T HA\|}{\nu}, \tag{3.6}$$

it follows that

$$\|\xi_x^k\| \overset{(3.2)}{\le} (L_f + \|A^T HA\|)\|x^k - \tilde{x}^k\| \overset{(3.6)}{\le} \nu r \|x^k - \tilde{x}^k\|.$$

Similarly, for any

$$s \ge \frac{L_g + \|B^T HB\|}{\nu}, \tag{3.7}$$

we have

$$\|\xi_y^k\| \overset{(3.4)}{\le} (L_g + \|B^T HB\|)\|y^k - \tilde{y}^k\| \overset{(3.7)}{\le} \nu s\|y^k - \tilde{y}^k\|.$$

To simplify our following analysis, we denote $R_k = r_k I$, $S_k = s_k I$,

$$M_k = S_k + B^T HB, \qquad G_k = \begin{pmatrix} R_k & 0 & 0 \\ 0 & M_k & 0 \\ 0 & 0 & H^{-1} \end{pmatrix} \qquad \text{and} \qquad \xi^k = \begin{pmatrix} \xi_x^k \\ \xi_y^k \\ 0 \end{pmatrix}. \tag{3.8}$$

Now, ignoring the index k in the matrices R_k, M_k and G_k, we adopt a compact form for the predictor $\tilde{w}^k$. Note that

$$\begin{aligned} \tilde{x}^k &\overset{(3.1)}{=} P_{\mathcal{X},R}\big\{x^k - R^{-1}\big(f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)]\big)\big\} \\ &\overset{(3.5)}{=} P_{\mathcal{X},R}\{x^k - R^{-1}[f(x^k) - A^T\tilde{\lambda}^k + A^T H(A(x^k - \tilde{x}^k) + B(y^k - \tilde{y}^k))]\} \\ &\overset{(3.2)}{=} P_{\mathcal{X},R}\{x^k - R^{-1}[f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T HB(y^k - \tilde{y}^k) + \xi_x^k]\}. \end{aligned} \tag{3.9}$$

Since $M = S + B^T HB$, it follows that

$$\begin{aligned} \tilde{y}^k &\overset{(3.3)}{=} P_{\mathcal{Y},S}\big\{y^k - S^{-1}\big[g(y^k) - B^T[\lambda^k - H(A\tilde{x}^k + By^k - b)]\big]\big\} \\ &\overset{(3.5)}{=} P_{\mathcal{Y},S}\big\{\tilde{y}^k - S^{-1}\big[g(y^k) - B^T\tilde{\lambda}^k + B^T HB(y^k - \tilde{y}^k) + s(\tilde{y}^k - y^k)\big]\big\} \\ &\overset{(2.6)}{=} P_{\mathcal{Y},M}\big\{\tilde{y}^k - M^{-1}\big[g(y^k) - B^T\tilde{\lambda}^k + 2B^T HB(y^k - \tilde{y}^k) + M(\tilde{y}^k - y^k)\big]\big\} \\ &\overset{(3.4)}{=} P_{\mathcal{Y},M}\{y^k - M^{-1}[g(\tilde{y}^k) - B^T\tilde{\lambda}^k + B^T HB(y^k - \tilde{y}^k) + \xi_y^k]\}. \end{aligned} \tag{3.10}$$

Using the notation of $Q(w)$ (see (1.6)), it follows from (3.9), (3.10), and (3.5) that the predictor $\tilde{w}^k$ satisfies

$$\tilde{w}^k = P_{\mathcal{W},G}\{w^k - G^{-1}[Q(\tilde{w}^k) + (A, B, 0)^T HB(y^k - \tilde{y}^k) + \xi^k]\}. \tag{3.11}$$

Denote

$$q(w^k, \tilde{w}^k) := Q(\tilde{w}^k) + (A, B, 0)^T HB(y^k - \tilde{y}^k), \tag{3.12}$$

then equation (3.11) can be written as

$$\tilde{w}^k = P_{\mathcal{W},G}\{w^k - G^{-1}[q(w^k, \tilde{w}^k) + \xi^k]\}. \tag{3.13}$$

Before ending this subsection, we introduce another useful notation in the coming analysis

$$d(w^k, \tilde{w}^k, \xi^k) := (w^k - \tilde{w}^k) - G^{-1}\xi^k. \tag{3.14}$$

Moreover, we notice that

$$(\xi_y^k)^T S^{-1}\xi_y^k \ge (\xi_y^k)^T M^{-1}\xi_y^k \tag{3.15}$$

and thus

$$\|S^{-1}\xi_y^k\|_S^2 \ge \|M^{-1}\xi_y^k\|_M^2. \tag{3.16}$$

3.2. Correction

The task of the correction step is to produce the new iterate based on the predictor. We suggest to use the following correction forms:

$$\textbf{(Correction-I)} \qquad w^{k+1} := w_I^{k+1} = w^k - \alpha_k d(w^k, \tilde{w}^k, \xi^k) \tag{3.17}$$

or

$$\textbf{(Correction-II)} \qquad w^{k+1} := w_{II}^{k+1} = P_{\mathcal{W},G}\{w^k - \alpha_k G^{-1} q(w^k, \tilde{w}^k)\}. \tag{3.18}$$

Both correction forms have the same step-size

$$\alpha_k = \gamma \alpha_k^*, \tag{3.19}$$

where

$$\gamma \in (0,2) \qquad \text{and} \qquad \alpha_k^* = \frac{(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) + (w^k - \tilde{w}^k)^T G d(w^k, \tilde{w}^k, \xi^k)}{\|d(w^k, \tilde{w}^k, \xi^k)\|_G^2}. \tag{3.20}$$

Note that the computational load of setting the new iterate by (3.17) (resp. (3.18)) is insignificant. For example, when we use Correction-II (3.18), it follows that

$$\begin{aligned}
x^{k+1} &\overset{(3.18)}{=} P_{\mathcal{X},R}\{x^k - \alpha_k R^{-1}[f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T HB(y^k - \tilde{y}^k)]\} \\
&= P_{\mathcal{X},R}\{\tilde{x}^k - \alpha_k R^{-1}[f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T HB(y^k - \tilde{y}^k) + R(\tilde{x}^k - x^k)/\alpha_k]\} \\
&\overset{(2.5)}{=} P_{\mathcal{X}}\{\tilde{x}^k - \alpha_k[f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T HB(y^k - \tilde{y}^k) + R(\tilde{x}^k - x^k)/\alpha_k]\} \\
&= P_{\mathcal{X}}\{[Rx^k + (I-R)\tilde{x}^k] - \alpha_k[f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T HB(y^k - \tilde{y}^k)]\}.
\end{aligned} \tag{3.21}$$

A similar manipulation indicates that y^{k+1} and λ^{k+1} in (3.18) can be obtained by

$$y^{k+1} = P_{\mathcal{Y}}\{[My^k + (I-M)\tilde{y}^k] - \alpha_k[g(\tilde{y}^k) - B^T\tilde{\lambda}^k + B^T HB(y^k - \tilde{y}^k)]\} \tag{3.22}$$

and

$$\lambda^{k+1} = \lambda^k - \alpha_k H(A\tilde{x}^k + B\tilde{y}^k - b), \tag{3.23}$$

respectively. Since $\tilde{x}^k$, $\tilde{y}^k$, $\tilde{\lambda}^k$, $f(\tilde{x}^k)$, and $g(\tilde{y}^k)$ are known from the prediction procedure, the computational load in correction procedure (3.18) is insignificant.

Finally, it is worth mentioning that the Correction II (3.18) is different from the usual correction step in the hybrid proximal point method (HPPA). The Correction II is

$$w^{k+1} := P_{\mathcal{W},G}\{w^k - \alpha_k G^{-1} q(w^k, \tilde{w}^k)\},$$

while the correction step in the HPPA can be expressed as

$$w^{k+1} := P_{\mathcal{W},G}\{w^k - \alpha_k G^{-1} Q(\tilde{w}^k)\}.$$

The difference between q and Q in the two equations is listed in the form (3.12).

4. Contractive Properties and the Optimal Step Length

Let $w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*$ be any solution point for VI($\mathcal{W}, Q$). In the case that $\mathcal{W}^*$ is not a singleton, for any given w, we denote

$$\|w - w^*\|_G := \inf\{\|w - w^*\|_G \mid w^* \in \mathcal{W}^*\}.$$

Throughout this section, we let $w^k = (x^k, y^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ be a given vector, $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$ be the predictor generated by the prediction step, and w^k, $\tilde{w}^k$ and ξ^k satisfy (3.2) and (3.4) in Section 3.1. In order to investigate the convergence behavior for any $\alpha > 0$ in the correction forms, we denote the step-size dependent new iterate in both the correction forms I and II by $w^{k+1}(\alpha)$. Namely,

$$w_I^{k+1}(\alpha) = w^k - \alpha d(w^k, \tilde{w}^k, \xi^k) \qquad \text{and} \qquad w_{II}^{k+1}(\alpha) = P_{\mathcal{W},G}\{w^k - \alpha G^{-1} q(w^k, \tilde{w}^k)\}.$$

Let

$$\Theta_k(\alpha) := \|w^k - w^*\|_G^2 - \|w^{k+1}(\alpha) - w^*\|_G^2 \tag{4.1}$$

be referred as a *profit-function* of the proposed methods, since it measures the improvement obtained in the k-th iteration of the methods. Note that the progress $\Theta_k(\alpha)$ is a function of the step length α. It is natural to consider maximizing this function by choosing an optimal parameter α. However, since w^* is the solution point and thus is unknown, we can not maximize $\Theta_k(\alpha)$ directly. The first part of this section aims at providing a lower bound for $\Theta_k(\alpha)$, called $\Psi_k(\alpha)$, which does not include the unknown solution w^*. The theorem hereby converts the task of maximizing the function $\Theta_k(\alpha)$ to maximizing the function $\Psi_k(\alpha)$.

4.1. A lower-bound of the progress function

Theorem 4.1. *Let $w^{k+1}(\alpha)$ be the step-size dependent new iterate in the correction form I or II and $\Theta_k(\alpha)$ be defined in (4.1), then we have*

$$\Theta_k(\alpha) \geq \Psi_k(\alpha), \tag{4.2}$$

where

$$\Psi_k(\alpha) = 2\alpha\varphi(w^k, \tilde{w}^k, \xi^k) - \alpha^2\|d(w^k, \tilde{w}^k, \xi^k)\|_G^2 \tag{4.3}$$

and

$$\varphi(w^k, \tilde{w}^k, \xi^k) = (\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k) + (w^k - \tilde{w}^k)^T Gd(w^k, \tilde{w}^k, \xi^k). \tag{4.4}$$

The assertion of this theorem provides the fundamental result in the convergence analysis of the proposed methods. This result will be proved for Correction forms I and II in Subsections 4.1.1 and 4.1.2, respectively.

4.1.1. Proof of Theorem 4.1 for correction form I

First, we prove a proposition which is devoted to prove Theorem 4.1 for correction form I.

Proposition 4.1. *For correction form I, we have*

$$(w^k - w^*)^T Gd(w^k, \tilde{w}^k, \xi^k) \geq \varphi(w^k, \tilde{w}^k, \xi^k). \tag{4.5}$$

Proof. Since w^* is a solution of $\mathrm{VI}(\mathcal{W}, Q)$ and $\tilde{x}^k \in \mathcal{X}$, $\tilde{y}^k \in \mathcal{Y}$, we have

$$(\tilde{x}^k - x^*)^T\{f(x^*) - A^T\lambda^*\} \geq 0 \tag{4.6}$$

and

$$(\tilde{y}^k - y^*)^T\{g(y^*) - B^T\lambda^*\} \geq 0. \tag{4.7}$$

Setting $v = x^k - \frac{1}{r_k}\big(f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)]\big)$ and $u = x^*$ in (2.1), it follows from (3.1) that

$$\big\{x^k - \frac{1}{r_k}\big(f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)]\big) - \tilde{x}^k\big\}^T(x^* - \tilde{x}^k) \leq 0.$$

Using (3.5) and the notation of ξ_x^k (see (3.2)), it follows that

$$(x^* - \tilde{x}^k)^T\{f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T H(By^k - B\tilde{y}^k) + R(\tilde{x}^k - x^k) + \xi_x^k\} \geq 0. \tag{4.8}$$

Similarly, setting $v = y^k - \frac{1}{s_k}\big(g(y^k) - B^T[\lambda^k - H(A\tilde{x}^k + By^k - b)]\big)$ and $u = y^*$ in (2.1), it follows from (3.3) that

$$\big\{y^k - \frac{1}{s_k}\big(g(y^k) - B^T[\lambda^k - H(A\tilde{x}^k + By^k - b)]\big) - \tilde{y}^k\big\}^T(y^* - \tilde{y}^k) \leq 0.$$

Using (3.5) and the notation of ξ_y^k (see (3.4)), it follows that

$$(y^* - \tilde{y}^k)^T\{g(\tilde{y}^k) - B^T\tilde{\lambda}^k + S(\tilde{y}^k - y^k) + \xi_y^k\} \ge 0. \tag{4.9}$$

Since f is monotone, adding (4.6) and (4.8), we have

$$(\tilde{x}^k - x^*)^T\{R(x^k - \tilde{x}^k) - \xi_x^k\} + (A\tilde{x}^k - Ax^*)^T(\tilde{\lambda}^k - \lambda^*) \ge (A\tilde{x}^k - Ax^*)^T H(By^k - B\tilde{y}^k). \tag{4.10}$$

Similarly, adding inequalities (4.7) and (4.9) and utilizing the monotonicity of g, we obtain

$$(\tilde{y}^k - y^*)^T\{S(y^k - \tilde{y}^k) - \xi_y^k\} + (B\tilde{y}^k - By^*)^T(\tilde{\lambda}^k - \lambda^*) \ge 0.$$

Adding $(B\tilde{y}^k - By^*)^T H(By^k - B\tilde{y}^k)$ to both sides of the above inequality and using $M = S + B^T HB$, we get

$$(\tilde{y}^k - y^*)^T\{M(y^k - \tilde{y}^k) - \xi_y^k\} + (B\tilde{y}^k - By^*)^T(\tilde{\lambda}^k - \lambda^*) \ge (B\tilde{y}^k - By^*)^T H(By^k - B\tilde{y}^k). \tag{4.11}$$

Now adding inequalities (4.10) and (4.11), using $Ax^* + By^* = b$ and $A\tilde{x}^k + B\tilde{y}^k - b = H^{-1}(\lambda^k - \tilde{\lambda}^k)$ (see (3.5)), we obtain

$$\begin{aligned}&(\tilde{x}^k - x^*)^T\{R(x^k - \tilde{x}^k) - \xi_x^k\} + (\tilde{y}^k - y^*)^T\{M(y^k - \tilde{y}^k) - \xi_y^k\} + (\tilde{\lambda}^k - \lambda^*)^T H^{-1}(\lambda^k - \tilde{\lambda}^k)\\ &\quad \ge (\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k).\end{aligned} \tag{4.12}$$

Using the notation of G and $d(w^k, \tilde{w}^k, \xi^k)$, (4.12) can be written as

$$(\tilde{w}^k - w^*)^T G d(w^k, \tilde{w}^k, \xi^k) \ge (\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k)$$

and it follows that

$$(w^k - w^*)^T G d(w^k, \tilde{w}^k, \xi^k) \ge (\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k) + (w^k - \tilde{w}^k)^T G d(w^k, \tilde{w}^k, \xi^k). \tag{4.13}$$

The right-hand-side of (4.13) is $\varphi(w^k, \tilde{w}^k, \xi^k)$ and then the assertion of this proposition is proved.

Proof of Theorem 4.1. By a straightforward manipulation we have

$$\begin{aligned}\Theta_k(\alpha) &\overset{(4.1)}{=} \|w^k - w^*\|_G^2 - \|w^{k+1}(\alpha) - w^*\|_G^2\\ &\overset{(3.17)}{=} \|w^k - w^*\|_G^2 - \|w^k - \alpha d(w^k, \tilde{w}^k, \xi^k) - w^*\|_G^2\\ &= 2\alpha(w^k - w^*)^T G d(w^k, \tilde{w}^k, \xi^k) - \alpha^2\|d(w^k, \tilde{w}^k, \xi^k)\|_G^2\\ &\overset{(4.5)}{\ge} 2\alpha\varphi(w^k, \tilde{w}^k, \xi^k) - \alpha^2\|d(w^k, \tilde{w}^k, \xi^k)\|_G^2\\ &\overset{(4.3)}{=} \Psi_k(\alpha).\end{aligned} \tag{4.14}$$

The proof of Theorem 4.1 for correction form I is completed.

4.1.2. Proof of Theorem 4.1 for correction form Ⅱ

The following proposition is devoted to prove Theorem 4.1 for correction form Ⅱ.

Proposition 4.2. *For correction form Ⅱ, we have*

$$(w - w^*)^T q(w^k, \tilde{w}^k) \ge (w - \tilde{w}^k)^T q(w^k, \tilde{w}^k) + (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k), \quad \forall w \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m.$$

Proof. Since $\tilde{w}^k \in \mathcal{W}$ and $w^* \in \mathcal{W}^*$ is a solution of VI$(\mathcal{W}, Q)$, we have

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

Using the monotonicity of Q it follows that

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

Because $Q(\tilde{w}^k) = q(w^k, \tilde{w}^k) - (A, B, 0)^T HB(y^k - \tilde{y}^k)$ (see (3.12)), from the above inequality we obtain

$$(\tilde{w}^k - w^*)^T q(w^k, \tilde{w}^k) \geq (\tilde{w}^k - w^*)^T (A, B, 0)^T HB(y^k - \tilde{y}^k). \tag{4.15}$$

Using $(A, B, 0)(\tilde{w}^k - w^*) = A(\tilde{x}^k - x^*) + B(\tilde{y}^k - y^*) = A\tilde{x}^k + B\tilde{y}^k - b = H^{-1}(\lambda^k - \tilde{\lambda}^k)$, it follows from (4.15) that

$$(\tilde{w}^k - w^*)^T q(w^k, \tilde{w}^k) \geq (\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k)$$

and consequently we derive the assertion of this proposition immediately.

Proof of Theorem 4.1. Since $w^* \in \mathcal{W}$ and $w^{k+1}(\alpha) = P_{\mathcal{W},G}[w^k - \alpha G^{-1} q(w^k, \tilde{w}^k)]$, it follows from (2.3) that

$$\|w^{k+1}(\alpha) - w^*\|_G^2 \leq \|w^k - \alpha G^{-1} q(w^k, \tilde{w}^k) - w^*\|_G^2 - \|w^k - \alpha G^{-1} q(w^k, \tilde{w}^k) - w^{k+1}(\alpha)\|_G^2. \tag{4.16}$$

Consequently, we get

$$\begin{aligned} & \|w^k - w^*\|_G^2 - \|w^{k+1}(\alpha) - w^*\|_G^2 \\ \geq\ & \|w^k - w^*\|_G^2 + \|w^k - w^{k+1}(\alpha) - \alpha G^{-1} q(w^k, \tilde{w}^k)\|_G^2 - \|w^k - w^* - \alpha G^{-1} q(w^k, \tilde{w}^k)\|_G^2 \\ =\ & \|w^k - w^{k+1}(\alpha)\|_G^2 + 2\alpha \{w^{k+1}(\alpha) - w^*\}^T q(w^k, \tilde{w}^k). \end{aligned} \tag{4.17}$$

Applying the result of Proposition 4.2 to the last term in the right-hand-side of (4.17) and using the notation of $\Theta_k(\alpha)$, we obtain

$$\Theta_k(\alpha) \geq \|w^k - w^{k+1}(\alpha)\|_G^2 + 2\alpha\{w^{k+1}(\alpha) - \tilde{w}^k\}^T q(w^k, \tilde{w}^k) + 2\alpha(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \tag{4.18}$$

Since $\tilde{w}^k = P_{\mathcal{W},G}\{w^k - G^{-1}[q(w^k, \tilde{w}^k) + \xi^k]\}$ and $w^{k+1}(\alpha) \in \mathcal{W}$, it follows from (2.1) that for any $\alpha > 0$,

$$0 \geq 2\alpha\{w^{k+1}(\alpha) - \tilde{w}^k\}^T G\{[w^k - G^{-1} q(w^k, \tilde{w}^k) - G^{-1}\xi^k] - \tilde{w}^k\}. \tag{4.19}$$

Adding (4.18) and (4.19) and using $d(w^k, \tilde{w}^k, \xi^k) = (w^k - \tilde{w}^k) - G^{-1}\xi^k$ (see (3.14)), we obtain

$$\begin{aligned} \Theta_k(\alpha) \quad \geq \quad & \|w^k - w^{k+1}(\alpha)\|_G^2 + 2\alpha\{w^{k+1}(\alpha) - \tilde{w}^k\}^T G d(w^k, \tilde{w}^k, \xi^k) \\ & + 2\alpha(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k). \end{aligned} \tag{4.20}$$

By regrouping the right-hand-side of (4.20), we obtain

$$\begin{aligned} \Theta_k(\alpha) \quad & \geq \quad \|(w^k - w^{k+1}(\alpha)) - \alpha d(w^k, \tilde{w}^k, \xi^k)\|_G^2 - \alpha^2 \|d(w^k, \tilde{w}^k, \xi^k)\|_G^2 \\ & \qquad + 2\alpha\{(\lambda^k - \tilde{\lambda}^k)^T B(y^k - \tilde{y}^k) + (w^k - \tilde{w}^k)^T G d(w^k, \tilde{w}^k, \xi^k)\} \\ & \overset{(4.4)}{\geq} \quad 2\alpha\varphi(w^k, \tilde{w}^k, \xi^k) - \alpha^2 \|d(w^k, \tilde{w}^k, \xi^k)\|_G^2. \\ & \overset{(4.3)}{=} \quad \Psi_k(\alpha). \end{aligned}$$

and the proof is completed.

4.2. The step-size in the correction step

Based on the the result in Theorem 4.1 we get

$$\|w^{k+1}(\alpha) - w^*\|_G^2 \leq \|w^k - w^*\|_G^2 - \Psi_k(\alpha). \tag{4.21}$$

It is natural to maximize $\Psi_k(\alpha)$ in each iteration. Note that $\Psi_k(\alpha)$ is a quadratic function of α (see (4.3)) and it reaches its maximum at

$$\alpha_k^* = \frac{\varphi(w^k, \tilde{w}^k, \xi^k)}{\|d(w^k, \tilde{w}^k, \xi^k)\|_G^2}, \qquad \text{(due to (4.4), this is just the same as defined in (3.20))} \tag{4.22}$$

with

$$\Psi_k(\alpha_k^*) = \alpha_k^* \varphi(w^k, \tilde{w}^k, \xi^k). \tag{4.23}$$

To achieve the faster convergence, we propose a relaxation factor $\gamma \in [1, 2)$ and set the step-size α_k by $\alpha_k = \gamma\alpha_k^*$. By simple manipulations, we obtain

$$\begin{aligned} \Psi_k(\gamma\alpha_k^*) &\overset{(4.3)}{=} 2\gamma\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k) - (\gamma^2\alpha_k^*)(\alpha_k^*\|d(w^k, \tilde{w}^k, \xi^k)\|_G^2) \\ &\overset{(4.22)}{=} (2\gamma\alpha_k^* - \gamma^2\alpha_k^*)\varphi(w^k, \tilde{w}^k, \xi^k) \\ &\overset{(4.23)}{=} \gamma(2-\gamma)\Psi_k(\alpha_k^*). \end{aligned} \tag{4.24}$$

It follows from Theorem 4.1 that

$$\|w^{k+1} - w^*\|_G^2 \le \|w^k - w^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\varphi(w^k, \tilde{w}^k, \xi^k). \tag{4.25}$$

Proposition 4.3. *Under the same notations, we have*

$$\varphi(w^k, \tilde{w}^k, \xi^k) > \frac{1}{2}\big(\|A\tilde{x}^k + By^k - b\|_H^2 + \|d(w^k, \tilde{w}^k, \xi^k)\|_G^2\big). \tag{4.26}$$

Proof. It follows from (4.4) and (3.14) that

$$\begin{aligned} \varphi(w^k, \tilde{w}^k, \xi^k) &= (\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k) + \|w^k - \tilde{w}^k\|_G^2 - (w^k - \tilde{w}^k)^T\xi^k \\ &= (\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k) + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 \\ &\quad + \|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T\xi_x^k + \|y^k - \tilde{y}^k\|_M^2 - (y^k - \tilde{y}^k)^T\xi_y^k \\ &= (\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k) + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 + \|B(y^k - \tilde{y}^k)\|_H^2 \\ &\quad + \|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T\xi_x^k + \|y^k - \tilde{y}^k\|_S^2 - (y^k - \tilde{y}^k)^T\xi_y^k. \end{aligned} \tag{4.27}$$

Using $\lambda^k - \tilde{\lambda}^k = H(A\tilde{x}^k + B\tilde{y}^k - b)$ (see (3.5)), we have

$$\begin{aligned} &(\lambda^k - \tilde{\lambda}^k)^T(By^k - B\tilde{y}^k) + \frac{1}{2}\big(\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 + \|By^k - B\tilde{y}^k\|_H^2\big) \\ &= (A\tilde{x}^k + B\tilde{y}^k - b)^T H(By^k - B\tilde{y}^k) + \frac{1}{2}\big(\|A\tilde{x}^k + B\tilde{y}^k - b\|_H^2 + \|By^k - B\tilde{y}^k\|_H^2\big) \\ &= \frac{1}{2}\|A\tilde{x}^k + By^k - b\|_H^2. \end{aligned}$$

Substituting this into (4.27), it follows that

$$\begin{aligned} \varphi(w^k, \tilde{w}^k, \xi^k) &= \frac{1}{2}\big(\|A\tilde{x}^k + By^k - b\|_H^2 + \|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2 + \|By^k - B\tilde{y}^k\|_H^2\big) \\ &\quad + \|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T\xi_x^k + \|y^k - \tilde{y}^k\|_S^2 - (y^k - \tilde{y}^k)^T\xi_y^k. \end{aligned} \tag{4.28}$$

Using (3.2) we obtain

$$\begin{aligned} \|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T\xi_x^k &> \frac{1}{2}\|x^k - \tilde{x}^k\|_R^2 - (x^k - \tilde{x}^k)^T\xi_x^k + \frac{1}{2}\|R^{-1}\xi_x^k\|_R^2 \\ &= \frac{1}{2}\|x^k - \tilde{x}^k - R^{-1}\xi_x^k\|_R^2. \end{aligned}$$

Similarly, using (3.4), we obtain

$$\|y^k - \tilde{y}^k\|_S^2 - (y^k - \tilde{y}^k)^T\xi_y^k > \frac{1}{2}\|y^k - \tilde{y}^k - S^{-1}\xi_y^k\|_S^2.$$

Therefore, it follows from (4.28) that

$$\begin{aligned} \varphi(w^k, \tilde{w}^k, \xi^k) &> \frac{1}{2}\|A\tilde{x}^k + By^k - b\|_H^2 + \frac{1}{2}\|x^k - \tilde{x}^k - R^{-1}\xi_x^k\|_R^2 \\ &\quad + \frac{1}{2}\big(\|By^k - B\tilde{y}^k\|_H^2 + \|y^k - \tilde{y}^k - S^{-1}\xi_y^k\|_S^2\big) + \frac{1}{2}\|\lambda^k - \tilde{\lambda}^k\|_{H^{-1}}^2. \end{aligned} \tag{4.29}$$

By a manipulation, we have

$$\begin{aligned}
&\|By^k - B\tilde{y}^k\|_H^2 + \|y^k - \tilde{y}^k - S^{-1}\xi_y^k\|_S^2 \\
&\quad = \quad \|y^k - \tilde{y}^k\|_{(S+B^THB)}^2 - 2(y^k - \tilde{y}^k)^T\xi_y^k + \|S^{-1}\xi_y^k\|_S^2 \\
&\quad \overset{(3.16)}{\geq} \quad \|y^k - \tilde{y}^k\|_M^2 - 2(y^k - \tilde{y}^k)^T\xi_y^k + \|M^{-1}\xi_y^k\|_M^2 \\
&\quad = \quad \|y^k - \tilde{y}^k - M^{-1}\xi_y^k\|_M^2.
\end{aligned} \tag{4.30}$$

Substituting (4.30) into (4.29) and using the notation of $d(w^k, \tilde{w}^k, \xi^k)$ and G, the assertion of this proposition is proved.

From (4.22), (4.25), and (4.26) we get the following corollaries directly.

Corollary 4.1. *$\alpha_k^* > \frac{1}{2}$ for all $k \geq 0$.*

Corollary 4.2. *The sequence $\{w^k\}$ generated by the proposed methods satisfies*

$$\|w^{k+1} - w^*\|_G^2 \quad \leq \quad \|w^k - w^*\|_G^2 - \frac{\gamma(2-\gamma)}{4}\big(\|A\tilde{x}^k + By^k - b\|_H^2 + \|d(w^k, \tilde{w}^k, \xi^k)\|_G^2\big).$$

The above inequality tells us that the proposed methods belong to the projection and contraction methods [10].

5. Practical Implementation and Convergence

For a given $w^k = (x^k, y^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, $\tilde{x}^k$ is given by (3.1). The pair of $\tilde{x}^k$ and ξ_x^k with $\|\xi_x^k\| \leq \nu r_k \|x^k - \tilde{x}^k\|$ is generated by choosing a suitable large $r_k > 0$. In practical computation, a self-adaptive scheme is adopted to find such a suitable $r_k > 0$. For the fixed $p_x^k = f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)]$ and a trial $r_k > 0$, we set

$$\tilde{x}^k := P_{\mathcal{X}}\big[x^k - p_x^k/r_k\big],$$

and calculate

$$\xi_x^k := f(x^k) - f(\tilde{x}^k) + A^THA(x^k - \tilde{x}^k)$$

and

$$\nu_k := \|\xi_x^k\|/\big(r_k\|x^k - \tilde{x}^k\|\big).$$

If $\nu_k \leq \nu$, the trial $\tilde{x}^k$ is accepted; otherwise, r_k is increased by $r_k := r_k * \nu_k * 1.25$, this procedure is repeated. Since f is Lipschitz continuous, this process will generate an $r_k \geq (L_f + \|A^THA\|)/\nu$, the related $\tilde{x}^k$ satisfying condition (3.2). The same technique is used to obtain s_k and $\tilde{y}^k$ satisfying condition (3.4). In this way, the sequences $\{r_k\}$ and $\{s_k\}$ are monotonically non-decreasing and finally bounded above. The following is a detailed proposed method using Correction-II.

A self-adaptive approximate PPA based prediction-correction method

Step 0. Let $H \succ 0$, $\nu = 0.9$, $r_0 = s_0 = 1$, $w^0 = (x^0, y^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, $\gamma \in [1, 2)$. For $k = 0, 1, \ldots$ do:

Step 1. Calculate the predictor $\tilde{w}^k = (\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k)$:

Step 1.1. Calculate $\tilde{x}^k$:

1) Set $p_x^k := f(x^k) - A^T[\lambda^k - H(Ax^k + By^k - b)]$.

2) $\tilde{x}^k := P_{\mathcal{X}}\big[x^k - p_x^k/r_k\big]$;
$\xi_x^k := f(x^k) - f(\tilde{x}^k) + A^T HA(x^k - \tilde{x}^k)$;
$\nu_k := \|\xi_x^k\|/\big(r_k\|x^k - \tilde{x}^k\|\big)$.

3) If $\nu_k > \nu$, then increase r_k by $\quad r_k := r_k * \nu_k * 1.25 \quad$ and go to 2).

4) Prepare (limit the number) a reduced r for the next iteration if ν_k is too small:
$$r_{k+1} := \begin{cases} r_k * \nu_k * 1.25 & \text{if } \nu_k \le 0.5, \\ r_k & \text{otherwise.} \end{cases}$$

Step 1.2. Calculate $\tilde{y}^k$:

1) Set $p_y^k := g(y^k) - B^T[\lambda^k - H(A\tilde{x}^k + By^k - b)]$.

2) $\tilde{y}^k := P_{\mathcal{Y}}\big[y^k - p_y^k/s_k\big]$;
$\xi_y^k := g(y^k) - g(\tilde{y}^k) + B^T HB(y^k - \tilde{y}^k)$;
$\nu_k := \|\xi_y^k\|/\big(s_k\|y^k - \tilde{y}^k\|\big)$.

3) If $\nu_k > \nu$, then increase s_k by $\quad s_k := s_k * \nu_k * 1.25 \quad$ and go to 2).

4) Prepare (limit the number) a reduced s for the next iteration if ν_k is too small:
$$s_{k+1} := \begin{cases} s_k * \nu_k * 1.25 & \text{if } \nu_k \le 0.5, \\ s_k & \text{otherwise.} \end{cases}$$

Step 1.3. Calculate $\tilde{\lambda}^k$: set $p_\lambda^k := (A\tilde{x}^k + B\tilde{y}^k - b)$;
$\tilde{\lambda}^k := \lambda^k - Hp_\lambda^k$.

Step 2. Calculate the search direction in Correction-II:

Set $q_x^k := f(\tilde{x}^k) - A^T\tilde{\lambda}^k + A^T HB(y^k - \tilde{y}^k)$;
$q_y^k := g(\tilde{y}^k) - B^T\tilde{\lambda}^k + B^T HB(y^k - \tilde{y}^k)$;
$q_\lambda^k := H(A\tilde{x}^k + B\tilde{y}^k - b)$.

Step 3. Calculate the step-size in the correction step:

Set $d_x^k = x^k - \tilde{x}^k$; $\quad d_y^k = y^k - \tilde{y}^k$, $\quad d_\lambda^k = \lambda^k - \tilde{\lambda}^k$;
$$\alpha_k^* = \frac{(d_\lambda^k)^T Bd_y^k + (d_x^k)^T(r_k d_x^k - \xi_x^k) + (d_y^k)^T(Md_y^k - \xi_y^k) + (d_\lambda^k)^T p_\lambda^k}{(r_k d_x^k - \xi_x^k)^T(d_x^k - r_k^{-1}\xi_x^k) + (Md_y^k - \xi_y^k)^T(d_y^k - M^{-1}\xi_y^k) + (d_\lambda^k)^T p_\lambda^k};$$
$\alpha_k = \gamma\alpha_k^*$, $\quad$ (the formula of α_k^* see (3.20)).

Step 4. Calculate the new iterate $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$:

$x^{k+1} = P_{\mathcal{X}}\{[r_k x^k + (1 - r_k)\tilde{x}^k] - \alpha_k q_x^k\}$; $\qquad$ (see (3.21))
$y^{k+1} = P_{\mathcal{Y}}\{[My^k + (I - M)\tilde{y}^k] - \alpha_k q_y^k\}$; $\qquad$ (see (3.22))
$\lambda^{k+1} = \lambda^k - \alpha_k q_\lambda^k$; $\qquad$ (see (3.23))
$k := k + 1$, $\quad$ go to Step 1.

Remark 5.1. Numerical experiments indicate that, if r_k and s_k are chosen too large, then the convergence becomes very slow. Therefore, in Step 1, when ν_k is less than 0.5, r_k and s_k are reduced. However, the number of times that r_k and s_k are allowed to reduce is limited (usually up to 20). Therefore eventually $\{r_k\}$ and $\{s_k\}$ become non-decreasing and finally constant.

It is clear that the implementation of the proposed method is simple and well defined. The main computational load is the evaluation of $f(x)$ and $g(y)$.

Recall that solving problem VI$(\mathcal{W}, Q)$ is equivalent to finding a zero point of $e_G(w, \mathcal{W}, Q)$ (see (2.4)). In order to simplify the convergence proof for the proposed methods, we first introduce the following result:

Proposition 5.1. *For any given triplet* $(x^k, y^k, \lambda^k) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, *let the temporal point* $(\tilde{x}^k, \tilde{y}^k, \tilde{\lambda}^k) \in \mathcal{X} \times \mathcal{Y} \times \mathbb{R}^m$ *be generated by Procedure I in Section 3. Then there exists a constant* $a > 0$ *such that for all* $k \ge 0$,

$$\|e_G(\tilde{w}^k, \mathcal{W}, Q)\|_G \le a\|w^k - \tilde{w}^k\|_G \tag{5.1}$$

where

$$e_G(\tilde{w}^k, \mathcal{W}, Q) = \tilde{w}^k - P_{\mathcal{W},G}\{\tilde{w}^k - G^{-1}Q(\tilde{w}^k)\}. \tag{5.2}$$

Proof. Replacing the first $\tilde{w}^k$ in the right-hand-side of (5.2) by $P_{\mathcal{W},G}\{w^k - G^{-1}[q(w^k, \tilde{w}^k) + \xi^k]\}$ (see (3.13)), we get

$$\begin{array}{rcl} \|e_G(\tilde{w}^k, \mathcal{W}, Q)\|_G & = & \|P_{\mathcal{W},G}\{w^k - G^{-1}[q(w^k, \tilde{w}^k) + \xi^k]\} - P_{\mathcal{W},G}\{\tilde{w}^k - G^{-1}Q(\tilde{w}^k)\}\|_G \\ & \overset{(2.2)}{\le} & \|(w^k - \tilde{w}^k) - G^{-1}[q(w^k, \tilde{w}^k) - Q(\tilde{w}^k)] - G^{-1}\xi^k\|_G \\ & \overset{(3.12)}{\le} & \|w^k - \tilde{w}^k\|_G + \|G^{-1}(A, B, 0)^T HB(y^k - \tilde{y}^k)\|_G + \|G^{-1}\xi^k\|_G. \end{array}$$

Notice that under the conditions (3.2) and (3.4)

$$\begin{array}{rcl} \|G^{-1}\xi^k\|_G^2 & \overset{\text{def}}{=} & \|R^{-1}\xi_x^k\|_R^2 + \|M^{-1}\xi_y^k\|_M^2 \\ & \overset{(3.16)}{\le} & \|R^{-1}\xi_x^k\|_R^2 + \|S^{-1}\xi_y^k\|_S^2 \\ & \overset{(3.2,3.4)}{\le} & \nu^2\big(\|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_S^2\big) \\ & \le & \nu^2\big(\|x^k - \tilde{x}^k\|_R^2 + \|y^k - \tilde{y}^k\|_{(S+B^THB)}^2\big) \\ & \overset{\text{def}}{\le} & \nu^2\|w^k - \tilde{w}^k\|_G^2. \end{array} \tag{5.3}$$

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

$$\|e_G(\tilde{w}^k, \mathcal{W}, Q)\|_G \le a\|w^k - \tilde{w}^k\|_G$$

and the proposition is proved.

Theorem 5.1. *The sequence* $\{w^k\}$ *generated by the proposed methods converges to some* w^∞ *which is a solution of VI*$(\mathcal{W}, Q)$.

Proof. Since

$$\begin{array}{rcl} \|d(w^k, \tilde{w}^k, \xi^k)\|_G^2 & \overset{(3.14)}{=} & \|(w^k - \tilde{w}^k) - G^{-1}\xi^k\|_G^2 \\ & = & \|w^k - \tilde{w}^k\|_G^2 - 2(w^k - \tilde{w}^k)^T\xi^k + \|G^{-1}\xi^k\|_G^2 \\ & \ge & \|w^k - \tilde{w}^k\|_G^2 - 2\|w^k - \tilde{w}^k\|_G \cdot \|G^{-1}\xi^k\|_G + \|G^{-1}\xi^k\|_G^2 \\ & = & \big(\|w^k - \tilde{w}^k\|_G - \|G^{-1}\xi^k\|_G\big)^2 \\ & \overset{(5.3)}{\ge} & (1-\nu)^2\|w^k - \tilde{w}^k\|_G^2, \end{array}$$

from Corollary 4.2 we get

$$\|w^{k+1} - w^*\|_G^2 \le \|w^k - w^*\|_G^2 - \frac{\gamma(2-\gamma)}{4}(1-\nu)^2\|w^k - \tilde{w}^k\|_G^2. \tag{5.4}$$

It follows from (5.4) that

$$\|w^{k+1} - w^*\|_G^2 \le \|w^0 - w^*\|_G^2, \quad \forall k \ge 0 \tag{5.5}$$

and thus $\{w^k\}$ is bounded. Moreover, we have

$$\frac{\gamma(2-\gamma)}{4}(1-\nu)^2\sum_{k=0}^{\infty}\|w^k-\tilde{w}^k\|_G^2 \le \|w^0-w^*\|_G^2.$$

Therefore, we have

$$\lim_{k\to\infty}\|w^k-\tilde{w}^k\|_G^2 = 0 \tag{5.6}$$

and thus $\{\tilde{w}^k\}$ is also bounded and (due to Proposition 5.1)

$$\lim_{k\to\infty}\|e_G(\tilde{w}^k,\mathcal{W},Q)\|_G = 0.$$

Let w^∞ be a cluster point of $\{\tilde{w}^k\}$ and the subsequence $\{\tilde{w}^{k_j}\}$ converges to w^∞. Since $e_G(w,\mathcal{W},Q)$ is continuous function of w, it follows that

$$e_G(w^\infty,\mathcal{W},Q) = \lim_{j\to\infty} e_G(\tilde{w}^{k_j},\mathcal{W},Q) = 0 \qquad \text{and thus} \qquad e(w^\infty,\mathcal{W},Q)=0.$$

From Lemma 2.2, w^∞ is a solution point of VI$(\mathcal{W},Q)$.

Since $\lim_{k\to\infty}\|\tilde{w}^k-w^k\|=0$ and $\{\tilde{w}^{k_j}\}\to w^\infty$, for any given $\varepsilon>0$, there exists an $l>0$ such that

$$\|w^{k_l}-\tilde{w}^{k_l}\|_G < \frac{\varepsilon}{2}, \quad \text{and} \quad \|\tilde{w}^{k_l}-w^\infty\|_G < \frac{\varepsilon}{2}. \tag{5.7}$$

Therefore, for any $k\ge k_l$, it follows from (5.5) and (5.7) that

$$\|w^k-w^\infty\|_G \le \|w^{k_l}-w^\infty\|_G \le \|w^{k_l}-\tilde{w}^{k_l}\|_G + \|\tilde{w}^{k_l}-w^\infty\|_G < \varepsilon$$

and the sequence $\{w^k\}$ converges to w^∞.

6. Numerical Experiment for Traffic Equilibrium Problems

In this section, we present some preliminary numerical results on some capacity traffic equilibrium problems.

6.1. Traffic equilibrium problem with link capacity bound

The test problems are modified from the problems in [14] which were described in Section 5.2 of [13] as Example 1 and 2. The network equilibrium problem in [13] is a nonlinear complementarity problem of the traffic path-flow x. The modification is additionally to require the link flow $f\le b$, where b is the link capacity vector. In [13], because A is the path-link incidence matrix, the link-flow vector f is given by $f=A^Tx$. Therefore, the network equilibrium problem with link-flow restriction is a variational inequality

$$x\in S, \quad (x'-x)^TF(x)\ge 0, \quad \forall\, x'\in S$$

where

$$S=\{x\in\mathbb{R}^n \mid A^Tx\le b,\ x\ge 0\}$$

and F is described as in Section 5.2 of [13]. By introducing a positive slack variable $y\ge 0$ and setting $g(y)=0$, the problem can be converted into a structured variational inequality of form (1.2)-(1.3). In the test problems, for the reason of simulation, we give explicit forms of function $t(f)$ and $\lambda_w(d)$ (see the details in Section 5.2 of [13]) and thus $F(x)$ can be calculated. In the computational process, we restrict us only using the function value $F(x)$ for given x.

We use the proposed method with Correction-II and take $w^0 = 0$ as the starting point. Since $\|e_x(w^0)\|_\infty > 10^2$, instead of $e(w, \mathcal{W}, Q)$ (see (2.7)), we take

$$\max\left\{\frac{\|e_x(w^k)\|_\infty}{\|e_x(w^0)\|_\infty}, \|e_y(w^k)\|_\infty, \|e_\lambda(w^k)\|_\infty\right\} \le \varepsilon \tag{6.1}$$

as the stopping criterion.

6.2. Numerical results for problems with link capacity bound

In capacitated traffic equilibrium problems, every link flow up-bound is assigned by 30 or 40. We use the proposed method with Correction-II to solve the test problems. Since in these test problems $g(y) = 0$ and $B = I$, if we take $H = \beta I$, the subproblem (3.3) can be directly obtained by

$$\begin{aligned} \tilde{y}^k &\overset{(3.3)}{=} P_{\mathcal{Y}}\{y^k - [g(y^k) - B^T[\lambda^k - H(A\tilde{x}^k + By^k - b)]]/s\} \\ &= P_{\mathcal{Y}}\{y^k - [\beta(A\tilde{x}^k + y^k - b) - \lambda^k]/s\} \\ &= P_{\mathcal{Y}}\{[\lambda^k - \beta(A\tilde{x}^k - b) + (s-\beta)y^k]/s\} \end{aligned} \tag{6.2}$$

and the condition (3.4) is satisfied for any $s = \beta/\nu$. We use the self-adaptive method described in Section 5 to solve the test problems, instead of Step 1.2, we use (6.2) with $s = \beta/\nu$ to obtain $\tilde{y}^k$. In the test, we take $H = 5I$ and $w^0 = 0$.

We report the number of iterations, the mapping evaluations, and the CPU time for different capacities ($b = 30$ and $b = 40$) and various ε in Table 6.1.

Table 6.1. Numerical results for various ε in (6.1).

Examples	Link flow capacity	No. of iterations				No. of F evaluations				CPU-time $\varepsilon = 10^{-6}$
		$\varepsilon =$	10^{-4}	10^{-5}	10^{-6}	$\varepsilon =$	10^{-4}	10^{-5}	10^{-6}	
Example 1	30		132	159	199		296	350	430	0.06 Sec.
	40		150	189	220		329	399	471	0.08 Sec.
Example 2	30		168	200	229		366	399	498	0.10 Sec.
	40		203	288	333		440	621	716	0.13 Sec.

The solutions are obtained in a moderate number of iterations and the number of mapping F evaluations per iteration is approximately 2.

As illustrated in Section 6.1, the output vector x is the path-flow and the link flow vector is $A^T x$. In fact, λ^* in the output is referred as the toll charge on the congested link. For the two examples with link capacities $b = 40$ we list the optimal link flow and the toll charge in Table 6.2. and Table 6.3., respectively. Indeed, the link toll charge is greater than zero if and only if the link flow reaches the capacity.

Table 6.2.The optimal link flow and the toll charge on the link of Example 1 with $b = 40$.

Link	Flow	Charge	Link	Flow	Charge	Link	Flow	Charge	Link	Flow	Charge
1	0	0	8	32.90	0	15	27.06	0	22	33.95	0
2	12.94	0	9	0	0	16	5.27	0	23	0	0
3	40.00	25.2	10	0	0	17	1.83	0	24	12.94	0
4	12.94	0	11	0	0	18	32.90	0	25	40.00	124.6
5	0	0	12	33.95	0	19	0	0	26	32.33	0
6	40.00	125.4	13	27.06	0	20	0	0	27	34.16	0
7	34.73	0	14	12.94	0	21	0	0	28	0	0

Table 6.3. The optimal link flow and the toll charge on the link of Example 2 with $b = 40$.

Link	Flow	Charge	Link	Flow	Charge	Link	Flow	Charge	Link	Flow	Charge
1	40.00	4.3	11	1.85	0	21	40.00	1.1	31	11.96	0
2	38.15	0	12	11.96	0	22	40.00	136.6	32	40.00	164.2
3	40.00	163.2	13	26.19	0	23	26.19	0	33	40.00	135.7
4	13.81	0	14	13.81	0	24	0	0	34	26.19	0
5	0	0	15	0	0	25	0	0	35	28.04	0
6	0	0	16	0	0	26	0	0	36	40.00	301.3
7	0	0	17	0	0	27	0	0	37	0	0
8	0	0	18	0	0	28	0	0	−−	−−	−−
9	0	0	19	0	0	29	26.19	0	−−	−−	−−
10	40.00	1.1	20	40.00	1.8	30	1.85	0	−−	−−	−−

Remark 6.1. It is worth mentioning that, from the authors' observation, up to date the articles, in that the existing decomposition types of methods were proposed, do not attempt to derive the numerical results. As has been indicated in the introduction, in each iteration, implementing these methods involves solving a subproblem equivalent to an implicit projection, hence yields high computation complexity. Since the methods presented here require significantly reduced computations, we are able to produce numerical experiments.

7. Conclusion

In this paper we present some alternating projection based prediction-correction methods for solving monotone variational inequality problems with a special structure. Comparing with the existing alternating directions methods, we use some corrections which only require the insignificant amount of additional computations. The implementation is carried out by a simple projection. Preliminary numerical results with traffic equilibrium problems indicate that the proposed methods are effective in practice.

References

[1] D.P. Bertsekas and E.M. Gafni, Projection method for variational inequalities with applications to the traffic assignment problem, *Math. Programming Stud.*, **17** (1987), 139-159.

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

[3] G. Chen and M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Math. Programming*, **64** (1994), 81-101.

[4] J. Eckstein, Some saddle-function splitting methods for convex programming, *Optim. Methods Softw.*, **4** (1994), 75-83.

[5] J. Eckstein and D.P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Programming*, **55** (1992), 293-318.

[6] M.C. Ferris and J.S. Pang, Engineering and economic applications of complementarity problems, *SIAM Rev.*, **39** (1997), 669-713.

[7] M. Fukushima, Application of the alternating directions method of multipliers to separable convex programming problems, *Comput. Optim. Appl.*, **2** (1992), 93-111.

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

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

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

[11] B.S. He, Inexact implicit methods for monotone general variational inequalities, *Math. Programming*, **86** (1999), 199-217.

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

[13] B.S. He, L.-Z. Liao and X.-M. Yuan, A LQP based interior prediction-correction method for nonlinear complementarity problems, *J. Comput. Math.*, **24** (2006), 33-44.

[14] A. Nagurney and D. Zhang, Projected Dynamical Systems and Variational Inequalities with Applications, Kluwer Academic Publishers, Boston, Dordrecht, London, 1996.

[15] P. Tseng, Alternating projection-proximal methods for convex programming and variational inequalities, *SIAM J. Optim.*, **7** (1997), 951-965.