Math. Program., Ser. A 92: 103–118 (2002)

Digital Object Identifier (DOI) 10.1007/s101070100280

Bingsheng He · Li-Zhi Liao · Deren Han · Hai Yang

A new inexact alternating directions method for monotone variational inequalities

Received: January 4, 2000 / Accepted: October 2001
Published online February 14, 2002 –

Abstract. The alternating directions method (ADM) is an effective method for solving a class of variational inequalities (VI) when the proximal and penalty parameters in sub-VI problems are properly selected. In this paper, we propose a new ADM method which needs to solve two strongly monotone sub-VI problems in each iteration approximately and allows the parameters to vary from iteration to iteration. The convergence of the proposed ADM method is proved under quite mild assumptions and flexible parameter conditions.

Key words. variational inequality – alternating directions method – inexact method

1. Introduction

A variational inequality problem 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 $\mathcal{R}^n$, and F is a mapping from $\mathcal{R}^n$ into itself. In this paper, we consider the VI problem with the following structure:

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad F(u) = \begin{pmatrix} f(x) \\ g(y) \end{pmatrix}, \tag{1.2}$$

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

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

B.S. He: Department of Mathematics, Nanjing University, Nanjing, 210093, P.R. China, e-mail: hebma@nju.edu.cn. This author was supported by the NSFC Grant 19971040 and NSF Jiangsu Grant BK99041.

L.-Z. Liao: Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, P.R. China, e-mail: liliao@hkbu.edu.hk. This author was supported in part by Grant FRG/98-99/II-29 of Hong Kong Baptist University.

D. Han: Department of Mathematics, Nanjing University, Nanjing, 210093, P.R. China,
e-mail: handr@263.net

H. Yang: Department of Civil Engineering, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong, P.R. China. e-mail: cehyang@ust.hk. This author was supported in part by UGC Grant RI94/95. EG01.

Mathematics Subject Classification (1991): 90C30, 90C33, 65K05

general VI problem, it has numerous important applications, especially in economics and transportation equilibrium problems [1,4,9,11,21–23]. For solving structured problem (1.2)–(1.3), a number of decomposition methods have been suggested in the literature, such as [3,8,9,11,14,15,27,28].

By attaching a Lagrange multiplier vector $\lambda \in \mathcal{R}^l$ to the linear constraint $Ax+By=b$, one obtains an equivalent form of problem (1.2)–(1.3):

$$z^* \in \mathcal{Z}, \quad (z - z^*)^T Q(z^*) \geq 0, \quad \forall z \in \mathcal{Z}, \tag{1.4}$$

where

$$z = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad Q(z) = \begin{pmatrix} f(x) - A^T\lambda \\ g(y) - B^T\lambda \\ Ax + By - b \end{pmatrix}, \quad \mathcal{Z} = \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^l. \tag{1.5}$$

In the following, we denote VI problem (1.4)–(1.5) by MVI($\mathcal{Z}$, Q). For MVI($\mathcal{Z}$, Q) problem, Gabay [12] and Gabay and Mercier [13] proposed the following ADM method (alternating directions method of multipliers). In their method, the new iterate $\tilde{z} = (\tilde{x}, \tilde{y}, \tilde{\lambda}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^l$ is generated from a given triplet $z = (x, y, \lambda) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^l$ via the following procedure:

Given $(x, y, \lambda) \in \mathcal{Z}$*, find* $\tilde{x} \in \mathcal{X}$ *such that*

$$(x' - \tilde{x})^T \{f(\tilde{x}) - A^T[\lambda - \beta(A\tilde{x} + By - b)]\} \geq 0, \quad \forall x' \in \mathcal{X}, \tag{1.6}$$

and then find $\tilde{y} \in \mathcal{Y}$ *such that*

$$(y' - \tilde{y})^T \{g(\tilde{y}) - B^T[\lambda - \beta(A\tilde{x} + B\tilde{y} - b)]\} \geq 0, \quad \forall y' \in \mathcal{Y}, \tag{1.7}$$

finally, update λ *via*

$$\tilde{\lambda} = \lambda - \beta(A\tilde{x} + B\tilde{y} - b), \tag{1.8}$$

where $\beta > 0$ is a given constant penalty parameter of the linear constraint.

The ADM method has been studied extensively in the theoretical frameworks of both Lagrangian functions [10] and maximal monotone operators [8,12]. Most of the existing ADM methods in the literature require that the sub-VI problems (6)–(7) can be solved exactly in each iteration. Eckstein and Bertsekas [8] first constructed an ADM method which allows the inexact computation for sub-VI problems, and then Chen and Teboulle [3] introduced another inexact one. Since the efficiency of the method is dependent on the penalty parameter, most recently, He and Yang [19] extended the basic ADM method (1.6)–(1.8) by allowing the penalty parameter β to vary monotonically (non-increasingly or non-decreasingly). In [20], Kontogiorgis and Meyer gave another form of ADM methods which allows the penalty parameter to be a positive matrix and vary non-increasingly.

Note that the basic ADM method (1.6)–(1.8) has to solve two monotone sub-VI problems in each iteration. In many cases, solving these problems are quite difficult. In [3], Chen and Teboulle gave an ADM method which solves the original VI problem via solving a series of strongly monotone sub-VI problems. That is, for given triplet $z = (x, y, \lambda) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^l$, they obtained the new $(\tilde{x}, \tilde{y}) \in \mathcal{X} \times \mathcal{Y}$ via solving the following two strongly monotone sub-VI problems:

$$(x' - \tilde{x})^T \left\{ \frac{1}{\beta}(\tilde{x} - x) + f(\tilde{x}) - A^T[\lambda - \beta(Ax + By - b)] \right\} \geq 0, \quad \forall\, x' \in \mathcal{X}, \quad (1.9)$$

$$(y' - \tilde{y})^T \left\{ \frac{1}{\beta}(\tilde{y} - y) + g(\tilde{y}) - B^T[\lambda - \beta(Ax + By - b)] \right\} \geq 0, \quad \forall\, y' \in \mathcal{Y}. \quad (1.10)$$

Then the same update for λ is adopted

$$\tilde{\lambda} = \lambda - \beta(A\tilde{x} + B\tilde{y} - b). \quad (1.11)$$

It is clear that the original alternating directions method (1.6)–(1.8) adopts the new information in the iteration whenever possible. The sub-VI problems (1.9) and (1.10) in the method of Chen and Teboulle [3] are uniformly strongly monotone and easier to be solved than sub-VI problems (1.6) and (1.7). Combining the advantages of both methods, a new method, which generates the new triplet $\tilde{z} = (\tilde{x}, \tilde{y}, \tilde{\lambda})$ from $z = (x, y, \lambda)$ can be given by

$$(x' - \tilde{x})^T \{r(\tilde{x} - x) + f(\tilde{x}) - A^T[\lambda - \beta(A\tilde{x} + By - b)]\} \geq 0, \quad \forall\, x' \in \mathcal{X}, \quad (1.12)$$

$$(y' - \tilde{y})^T \{s(\tilde{y} - y) + g(\tilde{y}) - B^T[\lambda - \beta(A\tilde{x} + B\tilde{y} - b)]\} \geq 0, \quad \forall\, y' \in \mathcal{Y}, \quad (1.13)$$

and

$$\tilde{\lambda} = \lambda - \beta(A\tilde{x} + B\tilde{y} - b), \quad (1.14)$$

where $\beta,\ r,\ s$ are given positive constants. The algorithm of [7] is in fact only a special case of this method, but its extension to (1.12)–(1.14) is straightforward. The method (1.12)–(1.14) can be viewed as a proximal point algorithm to the alternating directions method of Gabay [12] and Gabay and Mercier [13] or a decomposition method to the proximal point method of augmented Lagrange function [24]. Further properties of this method can be found in [2,25].

Some applications [11,20] of method (1.6)–(1.8) have shown that the solution time is significantly dependent on the choice of penalty parameter β. Besides, the choice of proximal parameters r and s could improve the condition of Problem (1.12) and (1.13). Herefore, to enhance the efficiency of method (1.12)–(1.14), in this paper, we present a modified one that allows the penalty and proximal parameters to vary from iteration to iteration. Namely, the constant penalty parameter β and the proximal parameters r and s are replaced by some sequences of positive definite matrices $\{H_k\}$, $\{R_k\}$ and $\{S_k\}$, respectively.

This paper is organized as follows. Some preliminaries of variational inequalities are provided in Sect. 2. In Sect. 3, we present a new proximal ADM method and analyze the main convergence properties of the method under certain flexible conditions on the variable parameters. In Sect. 4, an inexact proximal ADM method is derived from the new proximal ADM method. The convergence of the new inexact proximal ADM method is proved. Finally, some concluding remarks are drawn in Sect. 5.

2. Preliminaries

In this section, we summarize some basic properties and related definitions which will be used in the following discussions.

First, we denote $\|x\| = \sqrt{x^T x}$ as the Euclidean-norm and $\|A\|$ as the matrix norm of A. Let Ω be a nonempty closed convex subset of $\mathcal{R}^l$ and let $P_\Omega(\cdot)$ denote the projection mapping from $\mathcal{R}^l$ onto Ω. A basic property of the projection mapping $P_\Omega(\cdot)$ is

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

For any $\alpha > 0$, it is well known [6] that the VI problem (1.1) is equivalent to the projection equation

$$u = P_\Omega[u - \alpha F(u)].$$

Let

$$E_{[\Omega,\alpha F]}(u) := u - P_\Omega[u - \alpha F(u)]$$

denote the residual function of the equation, then the VI problem (1.1) is equivalent to finding a zero point of $E_{[\Omega,F]}(u)$. Setting $v := u - \alpha F(u)$ in (2.1) we get

$$\alpha(E_{[\Omega,\alpha F]}(u))^T F(u) \ge \|E_{[\Omega,\alpha F]}(u)\|^2, \quad \forall u \in \Omega. \tag{2.2}$$

Definitions. 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

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

The following Lemma 1 can be viewed as a corollary of Proposition 3.4 in [26].

Lemma 1. *Let F be strongly monotone on Ω with modulus $\mu > 0$ and u^* be the unique solution of the strongly monotone VI problem (1.1). Then we have*

$$2\alpha(E_{[\Omega,\alpha F]}(u))^T F(u) - \|E_{[\Omega,\alpha F]}(u)\|^2 \ge \|u - u^*\|^2, \quad \forall u \in \Omega \quad \textit{and} \quad \alpha \ge \mu^{-1}. \tag{2.3}$$

Proof. For any fixed $u \in \Omega$ and $\alpha > 0$, we define

$$h(u, v, \alpha) := \alpha^2\|F(u)\|^2 - \|u - \alpha F(u) - v\|^2 = 2\alpha(u - v)^T F(u) - \|u - v\|^2$$

and it follows that

$$h(u, P_\Omega[u - \alpha F(u)], \alpha) = \max\{h(u, v, \alpha) \mid v \in \Omega\}.$$

Since $u^* \in \Omega$ we have

$$h(u, P_\Omega[u - \alpha F(u)], \alpha) \ge h(u, u^*, \alpha),$$

and this yields

$$2\alpha(E_{[\Omega,\alpha F]}(u))^T F(u) - \|E_{[\Omega,\alpha F]}(u)\|^2 \ge 2\alpha(u - u^*)^T F(u) - \|u - u^*\|^2. \tag{2.4}$$

On the other hand, since u^* is the solution and $u \in \Omega$, we have $(u - u^*)^T F(u^*) \geq 0$ and consequently

$$(u - u^*)^T F(u) \geq (u - u^*)^T (F(u) - F(u^*)) \geq \mu \|u - u^*\|^2. \tag{2.5}$$

Then the assertion follows from (2.4) and (2.5) directly.

□

For MVI($\mathcal{Z}$, Q) problem (1.4)–(1.5) considered in this paper, the equivalent task is to find a zero point of

$$E_{[\mathcal{Z},Q]}(z) := \begin{pmatrix} x - P_{\mathcal{X}}\{x - [f(x) - A^T\lambda]\} \\ y - P_{\mathcal{Y}}\{y - [g(y) - B^T\lambda]\} \\ Ax + By - b \end{pmatrix}. \tag{2.6}$$

We make the following standard assumptions:

Assumption A. A1. $\mathcal{X}$ and $\mathcal{Y}$ are nonempty and closed convex sets. $f(x)$ is continuous and monotone with respect to $\mathcal{X}$ and $g(y)$ is continuous and monotone with respect to $\mathcal{Y}$.

A2. The solution set of MVI($\mathcal{Z}$, Q), denoted by $\mathcal{Z}^*$, is nonempty.

Because f and g are monotone and $\mathcal{X}$ and $\mathcal{Y}$ are closed convex, the solution set $\mathcal{Z}^*$ of MVI($\mathcal{Z}$, Q) is closed and convex. For any $z \in \mathcal{Z}$, we let

$$\text{dist}(z, \mathcal{Z}^*) := \min\{\|z - z^*\| \mid z^* \in \mathcal{Z}^*\}$$

denote the Euclidean distance from z to $\mathcal{Z}^*$. It is clear that

$$\text{dist}(z, \mathcal{Z}^*) = 0 \quad \Longleftrightarrow \quad E_{[\mathcal{Z},Q]}(z) = 0.$$

In the literature, $\|E_{[\mathcal{Z},Q]}(z)\|$ is referred to as the error bound that measures how much z fails to be in $\mathcal{Z}^*$. In this paper, our analysis is based on the error bound $E_{[\mathcal{Z},Q]}(\cdot)$ and that the functions f and g must be continuous (and hence single-valued). Therefore, our approach is different from the monotonicity-based analyses like [7,8,20,24,25] that allow for set-valued monotone f and g, and provide stronger convergence guarantees.

3. A new ADM method and its main properties

Based on the discussion in Sect. 1, we now formally present our new proximal ADM method.

The new proximal ADM method: Starting with an initial arbitrary triplet $(x^0, y^0, \lambda^0) \in \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R}^l$, a sequence $\{(x^k, y^k, \lambda^k)\} \subset \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R}^l$, $k \geq 0$, is successively generated by the following steps:

Step 1. *Find* $x^{k+1} \in \mathcal{X}$ *such that*

$$(x' - x^{k+1})^T f_k(x^{k+1}) \geq 0, \quad \forall x' \in \mathcal{X}. \tag{3.1}$$

Step 2. *Find* $y^{k+1} \in \mathcal{Y}$ *such that*

$$(y' - y^{k+1})^T g_k(y^{k+1}) \geq 0, \quad \forall y' \in \mathcal{Y}. \tag{3.2}$$

Step 3. *Update* λ^{k+1} *via*

$$\lambda^{k+1} = \lambda^k - H_k(Ax^{k+1} + By^{k+1} - b). \tag{3.3}$$

Here

$$f_k(x) = f(x) - A^T[\lambda^k - H_k(Ax + By^k - b)] + R_k(x - x^k), \tag{3.4}$$

$$g_k(y) = g(y) - B^T[\lambda^k - H_k(Ax^{k+1} + By - b)] + S_k(y - y^k), \tag{3.5}$$

and $\{H_k\}$, $\{R_k\}$ and $\{S_k\}$ are sequences of both lower and upper bounded symmetric positive definite matrices. We say that a sequence of positive definite matrices $\{H_k\}$ is both lower and upper bounded if

$$\inf_k \{\xi_k \mid \xi_k \text{ is the smallest eigenvalue of matrix } H_k\} = \xi_{\min} > 0$$

and

$$\sup_k \{\zeta_k \mid \zeta_k \text{ is the largest eigenvalue of matrix } H_k\} = \zeta_{\max} < +\infty.$$

The notation $H' \succ (\succeq) H$ means that $H' - H$ is positive definite (positive semi-definite).

Remark 1. If $R_k = S_k \equiv 0$ and $H_k \equiv \beta I$, then the new proximal ADM method reduces to the basic ADM method (1.6)–(1.8). Since $R_k \succ 0$ (resp. $S_k \succ 0$), the operator $f_k(x)$ (resp. $g_k(y)$) is strongly monotone whenever $f(x)$ (resp. $g(y)$) is monotone. This guarantees that the sub-VI problems (3.1) and (3.2) are solvable and have unique solutions. For such 'easy' VI problems, there are a number of solution methods, such as projection methods [1,17] and Newton-type methods [16,26].

In the remaining of this section, let $z^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ denote the triplet generated by the new ADM method from a given $(x^k, y^k, \lambda^k) \in \mathcal{Z}$. We now investigate the main properties of the exact iterations. For convenience, we denote

$$u = \begin{pmatrix} x \\ y \end{pmatrix} \qquad \text{and} \qquad D_k = \begin{pmatrix} R_k & 0 \\ 0 & S_k \end{pmatrix}.$$

First, we have

Lemma 2. *If* $Ax^{k+1} + By^k - b = 0$, $x^{k+1} = x^k$ *and* $y^{k+1} = y^k$, *then* (x^k, y^k, λ^k) *is a solution of MVI*($\mathcal{Z}, Q$).

Proof. From (3.1)–(3.3) and the assumptions of this lemma, it follows that

$$(x' - x^{k+1})^T\{f(x^{k+1}) - A^T\lambda^k\} \geq 0, \quad \forall x' \in \mathcal{X},$$
$$(y' - y^{k+1})^T\{g(y^{k+1}) - B^T\lambda^k\} \geq 0, \quad \forall y' \in \mathcal{Y},$$

and

$$(x^{k+1}, y^{k+1}, \lambda^{k+1}) = (x^k, y^k, \lambda^k).$$

Hence, (x^k, y^k, λ^k) satisfies Eqs.(1.4)–(1.5), and is a solution of MVI($\mathcal{Z}, Q$). □

Lemma 2 indicates that the proposed method will terminate if $(\|u^{k+1} - u^k\|^2 + \|Ax^{k+1} + By^k - b\|^2) = 0$. Recall that solving problem MVI($\mathcal{Z}$, Q) is equivalent to finding a zero point of $E_{[\mathcal{Z},Q]}(z)$. Then, the following lemma implies that in order to show the convergence of the new proximal ADM method, we only need to verify

$$\lim_{k\to\infty}\left(\|u^{k+1} - u^k\|^2_{D_k} + \|Ax^{k+1} + By^k - b\|^2_{H_k}\right) = 0.$$

Lemma 3. *There exists a constant $\mu > 0$ such that for all $k \geq 0$,*

$$\mu\left\|E_{[\mathcal{Z},Q]}(z^{k+1})\right\|^2 \leq \left(\|u^{k+1} - u^k\|^2_{D_k} + \|Ax^{k+1} + By^k - b\|^2_{H_k}\right). \tag{3.6}$$

Proof. Note that

$$E_{[\mathcal{Z},Q]}(z^{k+1}) = \begin{pmatrix} x^{k+1} - P_{\mathcal{X}}\{x^{k+1} - [f(x^{k+1}) - A^T\lambda^{k+1}]\} \\ y^{k+1} - P_{\mathcal{Y}}\{y^{k+1} - [g(y^{k+1}) - B^T\lambda^{k+1}]\} \\ Ax^{k+1} + By^{k+1} - b \end{pmatrix}. \tag{3.7}$$

Based on the equivalence of the solutions of the variational inequality and the projection equation, the x^{k+1} in (3.1) and y^{k+1} in (3.2) satisfy

$$x^{k+1} = P_{\mathcal{X}}[x^{k+1} - f_k(x^{k+1})]$$

and

$$y^{k+1} = P_{\mathcal{Y}}[y^{k+1} - g_k(y^{k+1})],$$

respectively. Replacing the first x^{k+1} in (3.7) by $P_{\mathcal{X}}[x^{k+1} - f_k(x^{k+1})]$, y^{k+1} by $P_{\mathcal{Y}}[y^{k+1} - g_k(y^{k+1})]$ and using $\|P_{\mathcal{Z}}(z) - P_{\mathcal{Z}}(z')\| \leq \|z - z'\|$, we get

$$\begin{aligned}\left\|E_{[\mathcal{Z},Q]}(z^{k+1})\right\| &= \left\|\begin{matrix} P_{\mathcal{X}}\,[x^{k+1} - f_k(x^{k+1})] - P_{\mathcal{X}}\{x^{k+1} - [f(x^{k+1}) - A^T\lambda^{k+1}]\} \\ P_{\mathcal{Y}}[y^{k+1} - g_k(y^{k+1})] - P_{\mathcal{Y}}\{y^{k+1} - [g(y^{k+1}) - B^T\lambda^{k+1}]\} \\ Ax^{k+1} + By^{k+1} - b \end{matrix}\right\| \\ &\leq \left\|\begin{matrix} -f_k(x^{k+1}) + f(x^{k+1}) - A^T\lambda^{k+1} \\ -g_k(y^{k+1}) + g(y^{k+1}) - B^T\lambda^{k+1} \\ Ax^{k+1} + By^{k+1} - b \end{matrix}\right\|.\end{aligned}$$

Furthermore, using (3.4) and (3.5) and substituting $\lambda^k = \lambda^{k+1} + H_k(Ax^{k+1} + By^{k+1} - b)$ into the above inequality, we get

$$\begin{aligned}\left\|E_{[\mathcal{Z},Q]}(z^{k+1})\right\| &\leq \left\|\begin{matrix} A^T H_k B(y^{k+1} - y^k) - R_k(x^{k+1} - x^k) \\ -S_k(y^{k+1} - y^k) \\ (Ax^{k+1} + By^k - b) + B(y^{k+1} - y^k) \end{matrix}\right\| \\ &\leq \|R_k\|\cdot\|x^{k+1} - x^k\| + \left(\|A^T H_k B\| + \|S_k\| + \|B\|\right)\cdot\|y^{k+1} - y^k\| \\ &\quad + \|Ax^{k+1} + By^k - b\| \\ &\leq c_k\cdot\left(\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\| + \|Ax^{k+1} + By^k - b\|\right), \end{aligned} \tag{3.8}$$

where $c_k = \max(\|R_k\|, \|A^T H_k B\| + \|S_k\| + \|B\|, 1)$. Since $\{H_k\}$, $\{R_k\}$ and $\{S_k\}$ are upper bounded, it follows from the above inequality that there exists a constant $\mu > 0$ such that

$$\mu \left\| E_{[\mathcal{Z},Q]}(z^{k+1}) \right\|^2 \le \left(\|u^{k+1} - u^k\|_{D_k}^2 + \|Ax^{k+1} + By^k - b\|_{H_k}^2 \right).$$

□

Theorem 1. *Let $z^* = (x^*, y^*, \lambda^*) \in \mathcal{Z}^*$ be any solution point of MVI$(\mathcal{Z}, Q)$. Then*

$$\begin{aligned} &\|u^{k+1} - u^*\|_{D_k}^2 + \|\lambda^{k+1} - \lambda^*\|_{H_k^{-1}}^2 + \|B(y^{k+1} - y^*)\|_{H_k}^2 \\ &\le \|u^k - u^*\|_{D_k}^2 + \|\lambda^k - \lambda^*\|_{H_k^{-1}}^2 + \|B(y^k - y^*)\|_{H_k}^2 \\ &\quad - \left(\|u^{k+1} - u^k\|_{D_k}^2 + \|Ax^{k+1} + By^k - b\|_{H_k}^2 \right). \end{aligned} \tag{3.9}$$

Theorem 1, which provides the fundamental result in the convergence analysis of the new ADM method, is the main theorem of this paper. When $H_k \equiv H$ and $D_k \equiv D$, it yields from (3.9) that the sequence $\{\|u^k - u^*\|_D^2 + \|\lambda^k - \lambda^*\|_{H^{-1}}^2 + \|B(y^k - y^*)\|_H^2\}$ is strictly monotonically decreasing and $\lim_{k\to\infty}(\|u^{k+1} - u^k\|_D^2 + \|Ax^{k+1} + By^k - b\|_H^2) = 0$, and thus the convergence is a direct result of (3.9). For variable penalty and proximal parameters, from (3.9) we can derive different conditions for sequences $\{H_k\}$ and $\{D_k\}$ that should be obeyed to guarantee the convergence. The following lemma is devoted to prove this theorem.

Lemma 4. *For any $z^* = (x^*, y^*, \lambda^*) \in \mathcal{Z}^*$, it holds*

$$\begin{aligned} &(u^{k+1} - u^*)^T D_k (u^{k+1} - u^k) + (\lambda^{k+1} - \lambda^*)^T H_k^{-1} (\lambda^{k+1} - \lambda^k) \\ &\le (Ax^{k+1} - Ax^*)^T H_k (By^{k+1} - By^k). \end{aligned}$$

Proof. Since z^* is a solution of MVI$(\mathcal{Z}, Q)$ and $x^{k+1} \in \mathcal{X}$, $y^{k+1} \in \mathcal{Y}$, we have

$$(x^{k+1} - x^*)^T \{f(x^*) - A^T \lambda^*\} \ge 0 \tag{3.10}$$

and

$$(y^{k+1} - y^*)^T \{g(y^*) - B^T \lambda^*\} \ge 0. \tag{3.11}$$

On the other hand, from (3.1)–(3.2), $x^* \in \mathcal{X}$ and $y^* \in \mathcal{Y}$, it follows that

$$(x^* - x^{k+1})^T \left\{ f(x^{k+1}) - A^T \lambda^{k+1} + R_k(x^{k+1} - x^k) + A^T H_k (By^k - By^{k+1}) \right\} \ge 0, \tag{3.12}$$

and

$$(y^* - y^{k+1})^T \left\{ g(y^{k+1}) - B^T \lambda^{k+1} + S_k (y^{k+1} - y^k) \right\} \ge 0. \tag{3.13}$$

Adding Eqs. (3.10) and (3.12) and using the monotonicity of f, we have

$$\begin{aligned} &(x^{k+1} - x^*)^T R_k (x^{k+1} - x^k) + (Ax^{k+1} - Ax^*)^T (\lambda^* - \lambda^{k+1}) \\ &\le (Ax^{k+1} - Ax^*)^T H_k (By^{k+1} - By^k). \end{aligned} \tag{3.14}$$

In a similar way, adding Eqs. (3.11) and (3.13) and using the monotonicity of g, we have

$$(y^{k+1} - y^*)^T S_k(y^{k+1} - y^k) + (By^{k+1} - By^*)^T(\lambda^* - \lambda^{k+1}) \le 0. \tag{3.15}$$

Adding (3.14) and (3.15) and using $Ax^* + By^* = b$, it follows that

$$\begin{aligned}&(u^{k+1} - u^*)^T D_k(u^{k+1} - u^k) + (Ax^{k+1} + By^{k+1} - b)^T(\lambda^* - \lambda^{k+1})\\ &\quad \le (Ax^{k+1} - Ax^*)^T H_k(By^{k+1} - By^k).\end{aligned}$$

Using $(Ax^{k+1} + By^{k+1} - b) = H_k^{-1}(\lambda^k - \lambda^{k+1})$, the assertion of this lemma follows directly.

□

Proof of Theorem 1. Using identity

$$\|a+b\|^2 = \|a\|^2 - \|b\|^2 + 2(a+b)^T b,$$

we get

$$\begin{aligned}&\|u^{k+1} - u^*\|^2_{D_k} + \|\lambda^{k+1} - \lambda^*\|^2_{H_k^{-1}} + \|B(y^{k+1} - y^*)\|^2_{H_k}\\ &= \|u^k - u^*\|^2_{D_k} + \|\lambda^k - \lambda^*\|^2_{H_k^{-1}} + \|B(y^k - y^*)\|^2_{H_k}\\ &\quad -\Big(\|u^{k+1} - u^k\|^2_{D_k} + \|\lambda^{k+1} - \lambda^k\|^2_{H_k^{-1}} + \|B(y^{k+1} - y^k)\|^2_{H_k}\Big)\\ &\quad +2(u^{k+1} - u^*)^T D_k(u^{k+1} - u^k) + 2(\lambda^{k+1} - \lambda^*)^T H_k^{-1}(\lambda^{k+1} - \lambda^k)\\ &\quad +2(By^{k+1} - By^*)^T H_k(By^{k+1} - By^k).\end{aligned} \tag{3.16}$$

Now we deal with the crossing terms in (3.16).

$$\begin{aligned}&2(u^{k+1} - u^*)^T D_k(u^{k+1} - u^k) + 2(\lambda^{k+1} - \lambda^*)^T H_k^{-1}(\lambda^{k+1} - \lambda^k)\\ &\qquad +2(By^{k+1} - By^*)^T H_k(By^{k+1} - By^k) \qquad \text{(use Lemma 4)}\\ &\le 2(Ax^{k+1} - Ax^*)^T H_k(By^{k+1} - By^k)\\ &\qquad +2(By^{k+1} - By^*)^T H_k(By^{k+1} - By^k) \qquad \text{(use } Ax^* + By^* = b\text{)}\\ &= 2(Ax^{k+1} + By^{k+1} - b)^T H_k(By^{k+1} - By^k) \qquad \text{(use (3.3))}\\ &= -2(\lambda^{k+1} - \lambda^k)^T(By^{k+1} - By^k).\end{aligned} \tag{3.17}$$

It follows from (3.16) and (3.17) that

$$\begin{aligned}&\|u^{k+1} - u^*\|^2_{D_k} + \|\lambda^{k+1} - \lambda^*\|^2_{H_k^{-1}} + \|B(y^{k+1} - y^*)\|^2_{H_k}\\ &\le \|u^k - u^*\|^2_{D_k} + \|\lambda^k - \lambda^*\|^2_{H_k^{-1}} + \|B(y^k - y^*)\|^2_{H_k}\\ &\quad -\Big(\|u^{k+1} - u^k\|^2_{D_k} + \|\lambda^{k+1} - \lambda^k + H_k B(y^{k+1} - y^k)\|^2_{H_k^{-1}}\Big).\end{aligned} \tag{3.18}$$

Note that (see (3.3))

$$\lambda^{k+1} - \lambda^k + H_k B(y^{k+1} - y^k) = H_k(Ax^{k+1} + By^k - b). \tag{3.19}$$

Substituting (3.19) into (3.18), we get the assertion of this theorem.

□

Theorem 1 may lead to many ways to change the penalty and proximal parameters from iteration to iteration. In the following, we propose one flexible condition which is not monotone on $\{H_k\}$ and $\{D_k\}$.

Condition C. Let $\{\eta_k\}_0^\infty$ be a non-negative sequence with $\sum_{k=0}^\infty \eta_k < +\infty$. The sequences $\{H_k\}$ and $\{D_k\}$ satisfy the following conditions.

i) $H_0 \succ 0, \quad \frac{1}{1+\eta_k} H_k \preceq H_{k+1} \preceq (1+\eta_k) H_k, \quad$ for all $k \geq 0$,
ii) $D \succ 0, \quad D \preceq D_{k+1} \preceq (1+\eta_k) D_k, \quad$ for all $k \geq 0$.

Note that the assumption $\sum_{k=1}^\infty \eta_k < +\infty$ yields $\prod_{k=1}^\infty (1+\eta_k) < +\infty$. We denote

$$C_S := \sum_{k=1}^\infty \eta_k \quad \text{and} \quad C_P := \prod_{k=1}^\infty (1+\eta_k)$$

and consequently have

$$C_P^{-1} H_0 \preceq H_k \preceq C_P H_0, \qquad \forall\, k \geq 0$$

and

$$D \preceq D_k \preceq C_P D_0, \qquad \forall\, k \geq 0.$$

In other words, the two positive sequences are lower and upper bounded.

Now, we discuss the property of our new ADM method under Condition C.

Theorem 2. *Let $\{H_k\}$ and $\{D_k\}$ be the sequences satisfying Condition C and let $z^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ be the triplet generated by the exact iteration from a given $z^k = (x^k, y^k, \lambda^k)$. Denote*

$$G_k := \begin{pmatrix} D_k & 0 & 0 \\ 0 & H_k^{-1} & 0 \\ 0 & 0 & H_k \end{pmatrix} \qquad \text{and} \qquad w := \begin{pmatrix} u \\ \lambda \\ By \end{pmatrix}. \tag{3.20}$$

Then we have

$$\begin{aligned} \|w^{k+1} - w^*\|_{G_{k+1}}^2 \leq{} & (1+\eta_k)\|w^k - w^*\|_{G_k}^2 \\ & - \left(\|u^{k+1} - u^k\|_{D_k}^2 + \|Ax^{k+1} + By^k - b\|_{H_k}^2\right). \end{aligned} \tag{3.21}$$

Proof. Condition C implies that

$$0 \prec G_{k+1} \preceq (1+\eta_k) G_k$$

and thus

$$\|w^{k+1} - w^*\|_{G_{k+1}}^2 \leq (1+\eta_k)\|w^{k+1} - w^*\|_{G_k}^2. \tag{3.22}$$

On the other hand, using the notation in (3.20), the result in Theorem 1 (see (3.9)) can be written as

$$\|w^{k+1} - w^*\|_{G_k}^2 \leq \|w^k - w^*\|_{G_k}^2 - \left(\|u^{k+1} - u^k\|_{D_k}^2 + \|Ax^{k+1} + By^k - b\|_{H_k}^2\right). \tag{3.23}$$

The assertion of this theorem follows from (3.22) and (3.23) immediately.

□

The result in Theorem 2 is obtained under the assumption that sub-VI problems (3.1) and (3.2) are solved exactly (the method marked as an exact alternating directions method). However, we are interested in the convergence when problems (3.1) and (3.2) are solved inexactly. For the desirable convergence $\lim_{k\to\infty} \operatorname{dist}(z^k, \mathcal{Z}^*) = 0$, we postpone the discussion until Theorem 4 which is true for both exact and inexact methods in the next section.

4. An inexact method and its convergence

In many cases, solving sub-VI problems (3.1) and (3.2) exactly is either impossible or expensive. On the other hand, there seems to be little justification on the effort of obtaining the accurate solutions of the sub-VI problems in each iteration. In fact, many inexact methods and approximate rules have been proposed in proximal point algorithms and other fields [3, 8, 18]. Inspired by these results, we extend our proximal ADM method to an inexact one, which solves the sub-VI problems approximately. Namely, instead of Step 1 and Step 2 in the exact method, we adopt the following inexact Step 1' and Step 2', respectively.

The approximation rule in the new proximal ADM method:

Step 1'. *Find $x^{k+1} \in \mathcal{X}$ such that*

$$\|x^{k+1} - \tilde{x}^{k+1}\| \le \nu_k. \tag{4.1}$$

Step 2'. *Find $y^{k+1} \in \mathcal{Y}$ such that*

$$\|y^{k+1} - \tilde{y}^{k+1}\| \le \nu_k, \tag{4.2}$$

where $\{\nu_k\}$ is a non-negative sequence satisfying $\sum_{k=1}^{\infty} \nu_k < +\infty$, and $\tilde{x}^{k+1}$ and $\tilde{y}^{k+1}$ are the exact solutions of (3.1) and (3.2), respectively.

Remark 2. The requirements of (4.1) and (4.2) are achievable. These can be justified as follows. Since $f_k(x)$ is strongly monotone, say with modulus $r_k > 0$, according to Lemma 1, we have

$$\|x^{k+1} - \tilde{x}^{k+1}\|^2 \le 2\alpha\big(E_{[\mathcal{X},\alpha f_k]}(x^{k+1})\big)^T f_k(x^{k+1}) - \big\|E_{[\mathcal{X},\alpha f_k]}(x^{k+1})\big\|^2, \quad \forall\, \alpha \ge r_k^{-1}$$

where

$$E_{[\mathcal{X},\alpha f_k]}(x) = x - P_{\mathcal{X}}[x - \alpha f_k(x)].$$

Recall that

$$E_{[\mathcal{X},\alpha f_k]}(\tilde{x}^{k+1}) = 0, \qquad \alpha\big(E_{[\mathcal{X},\alpha f_k]}(\tilde{x}^{k+1})\big)^T f_k(\tilde{x}^{k+1}) - \big\|E_{[\mathcal{X},\alpha f_k]}(\tilde{x}^{k+1})\big\|^2 = 0$$

and (see 2.2))

$$2\alpha\big(E_{[\mathcal{X},\alpha f_k]}(x^{k+1})\big)^T f_k(x^{k+1}) - \big\|E_{[\mathcal{X},\alpha f_k]}(x^{k+1})\big\|^2 \ge \big\|E_{[\mathcal{X},\alpha f_k]}(x^{k+1})\big\|^2.$$

We can take an $\alpha \ge r_k^{-1}$ and find an x^{k+1} such that

$$2\alpha\big(E_{[\mathcal{X},\alpha f_k]}(x^{k+1})\big)^T f_k(x^{k+1}) - \big\| E_{[\mathcal{X},\alpha f_k]}(x^{k+1})\big\|^2 \le \nu_k^2. \tag{4.3}$$

This guarantees that $\|x^{k+1} - \tilde{x}^{k+1}\| \le \nu_k$. Note that there is no $\tilde{x}^{k+1}$ in (4.3). Inequality (4.3) provides a practical and achievable condition of satisfying (4.1). Following the same discussion, a similar condition of satisfying (4.2) can be established.

Since $\{H_k\}$ and $\{D_k\}$ are bounded and $\lambda^{k+1} - \tilde{\lambda}^{k+1} = H_k A(x^{k+1} - \tilde{x}^{k+1}) + H_k B(y^{k+1} - \tilde{y}^{k+1})$, according to the approximation rule in (4.1) and (4.2), there exists a constant $c > 0$ such that

$$\|w^{k+1} - \tilde{w}^{k+1}\|_{G_{k+1}} \le c\nu_k. \tag{4.4}$$

Theorem 3. *Let $\{z^k\} = \{(x^k, y^k, \lambda^k)\}$ be the sequence generated from the inexact ADM method and $\tilde{z}^{k+1} = (\tilde{x}^{k+1}, \tilde{y}^{k+1}, \tilde{\lambda}^{k+1})$ be the triplet generated by the related exact iteration from $z^k = (x^k, y^k, \lambda^k)$. Then we have*

$$\lim_{k\to\infty}\Big(\|\tilde{u}^{k+1} - u^k\|_{D_k}^2 + \|A\tilde{x}^{k+1} + By^k - b\|_{H_k}^2\Big) = 0. \tag{4.5}$$

Proof. Since $\{\nu_k\}$ is summable, so is $\{\nu_k^2\}$. Denote

$$E_1 := \sum_{k=0}^{\infty} \nu_k \qquad \text{and} \qquad E_2 := \sum_{k=0}^{\infty} \nu_k^2$$

and recall

$$C_S := \sum_{k=1}^{\infty} \eta_k \qquad \text{and} \qquad C_P := \prod_{k=1}^{\infty} (1+\eta_k).$$

According to Theorem 2 (in fact, replacing w^{k+1} in (3.21) by $\tilde{w}^{k+1}$), we have

$$\begin{aligned}\|\tilde{w}^{k+1} - w^*\|_{G_{k+1}}^2 \le{}& (1+\eta_k)\|w^k - w^*\|_{G_k}^2 \\ &- \Big(\|\tilde{u}^{k+1} - u^k\|_{D_k}^2 + \|A\tilde{x}^{k+1} + By^k - b\|_{H_k}^2\Big).\end{aligned} \tag{4.6}$$

It follows from (4.4) that

$$\begin{aligned}\|w^{k+1} - w^*\|_{G_{k+1}} &\le \|\tilde{w}^{k+1} - w^*\|_{G_{k+1}} + \|w^{k+1} - \tilde{w}^{k+1}\|_{G_{k+1}} \\ &\le \|\tilde{w}^{k+1} - w^*\|_{G_{k+1}} + c\nu_k.\end{aligned} \tag{4.7}$$

From (4.6) and (4.7) we get

$$\|w^{k+1} - w^*\|_{G_{k+1}} \le (1+\eta_k)^{\frac{1}{2}}\|w^k - w^*\|_{G_k} + c\nu_k$$

and consequently for all k,

$$\begin{aligned}\|w^{k+1} - w^*\|_{G_{k+1}} &\le (1+\eta_k)^{\frac{1}{2}}\Big((1+\eta_{k-1})^{\frac{1}{2}}\|w^{k-1} - w^*\|_{G_{k-1}} + c\nu_{k-1}\Big) + c\nu_k \\ &\le \Big(\Pi_{i=0}^k(1+\eta_i)\Big)^{\frac{1}{2}} \cdot \|w^0 - w^*\|_{G_0} + \Big(\Pi_{i=1}^k(1+\eta_i)\Big)^{\frac{1}{2}} c\nu_0 \\ &\quad + \cdots + (1+\eta_k)^{\frac{1}{2}} c\nu_{k-1} + c\nu_k \\ &\le C_P^{1/2} \cdot \big(\|w^0 - w^*\|_{G_0} + cE_1\big) := C_w.\end{aligned} \tag{4.8}$$

Therefore, the sequence is bounded. Furthermore

$$\begin{aligned}
&\|w^{k+1} - w^*\|^2_{G_{k+1}} \\
&\quad = \|\tilde{w}^{k+1} - w^* + (w^{k+1} - \tilde{w}^{k+1})\|^2_{G_{k+1}} \qquad \text{(Cauchy-Schwarz)} \\
&\quad \le \|\tilde{w}^{k+1} - w^*\|^2_{G_{k+1}} + 2\|\tilde{w}^{k+1} - w^*\|_{G_{k+1}} \cdot \|w^{k+1} - \tilde{w}^{k+1}\|_{G_{k+1}} \\
&\qquad + \|w^{k+1} - \tilde{w}^{k+1}\|^2_{G_{k+1}} \qquad \text{(use (4.6) and (4.4))} \\
&\quad \le (1+\eta_k)\|w^k - w^*\|^2_{G_k} + 2(1+\eta_k)^{\frac{1}{2}}\|w^k - w^*\|_{G_k} \cdot c\nu_k + c^2\nu_k^2 \\
&\qquad - \big(\|\tilde{u}^{k+1} - u^k\|^2_{D_k} + \|A\tilde{x}^{k+1} + By^k - b\|^2_{H_k}\big) \qquad \text{(use (4.8))} \\
&\quad \le \|w^k - w^*\|^2_{G_k} + C_w^2\eta_k + 2C_P C_w c\nu_k + c^2\nu_k^2 \\
&\qquad - \big(\|\tilde{u}^{k+1} - u^k\|^2_{D_k} + \|A\tilde{x}^{k+1} + By^k - b\|^2_{H_k}\big).
\end{aligned}$$

It follows that for all k,

$$\begin{aligned}
\|w^{k+1} - w^*\|^2_{G_{k+1}} \le \|w^0 - w^*\|^2_{G_0} &+ C_w^2 C_S + 2C_P C_w c E_1 + c^2 E_2 \\
&- \Sigma_{i=0}^k \big(\|\tilde{u}^{i+1} - u^i\|^2_{D_i} + \|A\tilde{x}^{i+1} + By^i - b\|^2_{H_i}\big).
\end{aligned}$$

Let $k \to \infty$, we have

$$\lim_{k\to\infty} \big(\|\tilde{u}^{k+1} - u^k\|^2_{D_k} + \|A\tilde{x}^{k+1} + By^k - b\|^2_{H_k}\big) = 0.$$

□

Theorem 4. *Let $\{z^k\}$ be the sequence generated from the inexact ADM method. Then the method is convergent in the sense*

$$\lim_{k\to\infty} dist(z^k, \mathcal{Z}^*) = 0.$$

Proof. Let $\tilde{z}^{k+1} = (\tilde{x}^{k+1}, \tilde{y}^{k+1}, \tilde{\lambda}^{k+1})$ be the triplet generated from the related exact iteration with a given $z^k = (x^k, y^k, \lambda^k)$. It follows from Theorem 3 and Lemma 3 that we have

$$\lim_{k\to\infty} \left\| E_{[\mathcal{Z},Q]}(\tilde{z}^k) \right\|^2 = 0.$$

From (4.8) we know that the sequence $\{w^k\}$ is bounded and so are $\{z^k\}$ and $\{\tilde{z}^k\}$. Then there exists a bounded closed set, say $S(z^0)$, such that $\{\tilde{z}^k\} \in S(z^0)$. Since

$$\lim_{k\to\infty} \|z^k - \tilde{z}^k\| = 0,$$

$\lim_{k\to\infty} \mathrm{dist}(\tilde{z}^k, \mathcal{Z}^*) = 0$ is an equivalent statement of this theorem. If we have

$$\lim\sup_{k\to\infty} \mathrm{dist}(\tilde{z}^k, \mathcal{Z}^*) = \delta > 0,$$

then

$$\{\tilde{z}^k\} \subset S = S(z^0) \cap \{z \mid \mathrm{dist}(z, \mathcal{Z}^*) \ge \delta/2\}.$$

Since $S \cap \mathcal{Z}^* = \emptyset$, then $E_{[\mathcal{Z},Q]}(z) \neq 0$ for all $z \in S$. Because S is compact and $E_{[\mathcal{Z},Q]}(z)$ is continuous on S, we have

$$\min_{z \in S} \left\{ \| E_{[\mathcal{Z},Q]}(z) \|^2 \right\} = \varepsilon > 0. \tag{4.9}$$

This contradicts the fact

$$\{\tilde{z}^k\} \subset S \qquad \text{and} \qquad \lim_{k \to \infty} \left\| E_{[\mathcal{Z},Q]}(\tilde{z}^k) \right\|^2 = 0.$$

Therefore the proof is complete.

□

5. Concluding remarks

The proposed inexact ADM method in this paper extends the original one [7] by allowing the penalty and proximal parameters to vary from iteration to iteration. In addition, it enables us to take the flexible matrix sequences $\{R_k\}$, $\{S_k\}$ and $\{H_k\}$ for r, s and β, respectively. However, how to choose and/or adjust such matrices is still an interesting research topic. From our quantitative convergence analysis in this paper, the following principles should be observed in adjusting these matrices.

The proximal term. The objective of introducing the proximal terms, $R_k(x - x^k)$ in (3.4) and $S_k(x - x^k)$ in (3.5), is to improve the condition of sub-problems (3.1) and (3.2), respectively. It should be noted that the trade-off between the cost per iteration and the total number of iterations should be always balanced. In fact, large $R_k \succ 0$ (resp. $S_k \succ 0$) will lead to an easy solution for sub-problem (3.1) (resp. sub-problem (3.2)), but the number of outer-iterations will be increased. Therefore, for sub-problems which are not extremely ill-posed, the proximal parameters should be small.

The penalty term. Recall that solving MVI($\mathcal{Z}$, Q) is equivalent to finding a zero point of $E_{[\mathcal{Z},Q]}(z)$ and

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

Notice that (see the first inequality in (3.8))

$$\left\| E_{[\mathcal{Z},Q]}(z^{k+1}) \right\| = \left\| \begin{matrix} e_x(z^{k+1}) \\ e_y(z^{k+1}) \\ e_\lambda(z^{k+1}) \end{matrix} \right\| \leq \left\| \begin{matrix} A^T H_k B(y^{k+1} - y^k) - R_k(x^{k+1} - x^k) \\ -S_k(y^{k+1} - y^k) \\ Ax^{k+1} + By^{k+1} - b \end{matrix} \right\|, \tag{5.1}$$

therefore for small $S_k \succ 0$,

$$\left\| E_{[\mathcal{Z},Q]}(z^{k+1}) \right\|^2 \approx \|e_x(z^{k+1})\|^2 + \|e_\lambda(z^{k+1})\|^2.$$

For the sake of balance, we suggest to adjust the penalty parameter matrix $H_k \succ 0$ so that $\|e_x(z)\| \approx \|e_\lambda(z)\|$. Based on (5.1), it seems that the following consideration is reasonable. For an iterate $z = (x, y, \lambda) \in \mathcal{Z}$, if $\|e_x(z)\| \ll \|e_\lambda(z)\|$, we should increase H in the next iteration, conversely, we should decrease H when $\|e_x(z)\| \gg \|e_\lambda(z)\|$. For example, we can take $H_k = \beta_k I$ and adjust β_k according to the following formulas:

$$\beta_{k+1} = \begin{cases} (1+\eta_k)\beta_k & \text{if} \quad \|e_x(z^k)\| < \frac{1}{4}\|e_\lambda(z^k)\|, \\ \beta_k/(1+\eta_k) & \text{if} \quad \|e_x(z^k)\| > 4\|e_\lambda(z^k)\|, \\ \beta_k & \text{otherwise,} \end{cases} \tag{5.2}$$

where

$$\eta_k = \min\left\{1, \frac{1}{(\max\{1, k-l\})^2}\right\} = \left\{1, 1, \dots, 1, \frac{1}{4}, \frac{1}{9}, \dots\right\}$$

and

$$l = \text{dimension of } \lambda.$$

Our preliminary numerical experiments indicate that the adjustment for $\beta_k > 0$ (resp. $H_k \succ 0$) is necessary. The method with a fixed penalty parameter $\beta > 0$ (resp. $H \succ 0$) converges extremely slowly when the parameter is either too large or too small, even if for some toy examples. The adjustment strategy (5.2) for the penalty parameter always leads to some improvement.

Many optimization problems in real application can be converted into a MVI($\mathcal{Z}$, Q) problem with special structures. To develop the parameter adjusting rules in the proposed method for structured MVI($\mathcal{Z}$, Q) is one of our on-going research topics.

Acknowledgements. The authors would like to thank the two anonymous referees for their careful reading of an earlier version of this paper and constructive suggestions. In particular, the suggestions from one referee enable us to investigate more flexible parameter conditions and simplify the proof of the inexact alternating direction method significantly. The authors are also very grateful to the Associate Editor for the AE's help in finishing this final version.

References

1. Bertsekas, D.P., Gafni, E. M. (1982): Projection method for variational inequalities with applications to the traffic assignment problem. Mathematical Programming Study **17**, 139–159
2. Burachik, R.S., Scheimberg, S., Svaiter, B.F. (1999): Robustness of the extragradient-proximal point algorithm. Working Paper, Instituto de Mathematica Pura e Aplicada (IMPA), Rio de Janeiro, Brazil
3. Chen, G., Teboulle, M. (1994): A proximal-based decomposition method for convex minimization problems. Mathematical Programming **64**, 81–101
4. Dafermos, S. (1980): Traffic equilibrium and variational inequalities. Transportation Science **14**, 42–54
5. Drezner, Z. (ed.) (1995): Facility Location: A survey of applications and methods. Springer-Verlag, New York
6. Eaves, B.C. (1971): On the basic theorem of complementarity. Mathematical Programming **1**, 68–75
7. Eckstein, J. (1994): Some saddle-function splitting methods for convex programming. Optimization Methods and Software **4**, 75–83
8. Eckstein, J., Bertsekas, D.P. (1992): On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming **55**, 293–318
9. Eckstein, J., Fukushima, M. (1994): Some reformulation and applications of the alternating directions method of multipliers. In: Hager, W.W. et al., eds., Large Scale Optimization: State of the Art, pp. 115–134. Kluwer Academic Publishers
10. Fortin, M., Glowinski, R. (Eds.) (1983): Augmented Lagrangian methods: Applications to the numerical solution of Boundary-Value Problems. North-Holland, Amsterdam
11. Fukushima, M. (1992): Application of the alternating directions method of multipliers to separable convex programming problems. Computational Optimization and Applications **2**, 93–111

12. Gabay, D. (1983): Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R., eds., Augmented Lagrangian methods: Applications to the numerical solution of Boundary-Value Problems, pp. 299–331. North-Holland, Amsterdam, The Netherlands
13. Gabay, D., Mercier, B. (1976): A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Computer and Mathematics with Applications **2**, 17–40
14. Glowinski, R. (1984): Numerical Methods for Nonlinear Variational Problems. Springer-Verlag, New York
15. Glowinski, R., Le Tallec, P. (1989): Augmented Lagrangian and Operator-Splitting Methods. In: Nonlinear Mechanics, SIAM Studies in Applied Mathematics, Philadelphia, PA
16. Harker, P.T., Pang, J.S. (1990): Finite dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and Applications. Mathematical Programming **48**, 161–220
17. He, B.S. (1997): A class of new methods for monotone variational inequalities. Applied Mathematics and Optimization **35**, 69–76
18. He, B.S. (1999): Inexact implicit methods for monotone general variational inequalities. Mathematical Programming **86**, 199–217
19. He, B.S., Yang, H. (1998): Some convergence properties of a method of multipliers for linearly constrained monotone variational inequalities. Operations Research Letters **23**, 151–161
20. Kontogiorgis, S., Meyer, R.R. (1998): A variable-penalty alternating directions method for convex optimization. Mathematical Programming **83**, 29–53
21. Nagurney, A. (1993): Network economics, A variational inequality approach. Kluwer Academics Publishers, Dordrecht
22. Nagurney, A., Ramanujam, P. (1996): Transportation network policy modeling with goal targets and generalized penalty functions. Transportation Science **30**, 3–13
23. Nagurney, A., Thore, S., Pan, J. (1996): Spatial market policy modeling with goal targets. Operations Research **44**, 393–406
24. Rockafellar, R.T. (1976): Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Mathematics of Operations Research **1**, 97–116
25. Solodov, M., Svaiter, B.F. (1999): A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. Set-Valued Analysis **7**, 117–132
26. Taji, K., Fukushima, M., Ibaraki, T. (1993): A globally convergent Newton method for solving strongly monotone variational inequalities. Mathematical Programming **58**, 369–383
27. Tseng, P. (1991): Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM J. Control and Optimization **29**, 119–138
28. Tseng, P. (1997): Alternating projection-proximal methods for convex programming and variational inequalities. SIAM J. Optimization **7**, 951–965