Journal of Computational Mathematics, Vol.21, No.4, 2003, 495–504.

A MODIFIED VARIABLE-PENALTY ALTERNATING DIRECTIONS METHOD FOR MONOTONE VARIATIONAL INEQUALITIES *1)

Bing-sheng He Sheng-li Wang
(*Department of Mathematics, Nanjing University, Nanjing 210093, China*)
Hai Yang
(*Department of Civil Engineering, The Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong*)

Abstract

Alternating directions method is one of the approaches for solving linearly constrained separate monotone variational inequalities. Experience on applications has shown that the number of iteration significantly depends on the penalty for the system of linearly constrained equations and therefore the method with variable penalties is advantageous in practice. In this paper, we extend the Kontogiorgis and Meyer method [12] by removing the monotonicity assumption on the variable penalty matrices. Moreover, we introduce a self-adaptive rule that leads the method to be more efficient and insensitive for various initial penalties. Numerical results for a class of Fermat-Weber problems show that the modified method and its self-adaptive technique are proper and necessary in practice.

Key words: Monotone variational inequalities, Alternating directions method, Fermat-Weber problem.

1. Introduction

The mathematical form of variational inequalities consists of finding a vector $u^* \in \Omega$ such that

$$\mathrm{VI}(\Omega, F) \qquad (u-u^*)^T F(u^*) \geq 0, \qquad \forall\ u \in \Omega, \tag{1}$$

where Ω is a nonempty, closed convex subset of $\mathcal{R}^l$, F is a continuous mapping from $\mathcal{R}^l$ to itself. In practice, many VI problems have the following separable structure, namely (e.g., [14]),

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad F(u) = \begin{pmatrix} f(x) \\ g(y) \end{pmatrix}, \tag{2}$$

$$\Omega = \{(x,y) | x \in \mathcal{X}, y \in \mathcal{Y}, Ax + By = b\}, \tag{3}$$

where $\mathcal{X} \subset \mathcal{R}^n$ and $\mathcal{Y} \subset \mathcal{R}^m$ are given closed convex sets, $f: \mathcal{X} \to \mathcal{R}^n$, $g: \mathcal{Y} \to \mathcal{R}^m$ are given monotone operators, $A \in \mathcal{R}^{r\times n}, B \in \mathcal{R}^{r \times m}$ are given matrices, and $b \in \mathcal{R}^r$ is a given vector.

By attaching a Lagrange multiplier vector $\lambda \in \mathcal{R}^r$ to the linear constraints $Ax + By = b$, the problem under consideration can be explained as a *mixed variational inequality* (VI with equality restriction $Ax + By = b$ and unrestricted variable λ):

$$\text{Find} \quad w^* \in \mathcal{W}, \quad \text{such that} \quad (w-w^*)^T Q(w^*) \geq 0, \quad \forall\, w \in \mathcal{W}, \tag{4}$$

* Received March 25, 2001.
1) The first author was supported by the NSFC grant 10271054, the third author was supported in part by the Hong Kong Research Grants Council through a RGC-CERG Grant (HKUST6203/99E).

where

$$w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \qquad Q(w) = \begin{pmatrix} f(x) - A^T\lambda \\ g(y) - B^T\lambda \\ Ax + By - b \end{pmatrix}, \qquad \mathcal{W} = \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^r. \tag{5}$$

Problem (4)-(5) is denoted as MVI($\mathcal{W}, Q$) and will be concerned in this paper. It has been well known (e.g., see [14]) that solving MVI($\mathcal{W}, Q$) is equivalent to finding a zero point of

$$e(w) := w - P_{\mathcal{W}}[w - Q(w)], \tag{6}$$

where $P_{\mathcal{W}}(\cdot)$ denotes the projection on $\mathcal{W}$. $\|e(w)\|$ can be viewed as a '*error bound*' that measures how much w fails to be a solution of MVI($\mathcal{W}, Q$).

As a tool for solving MVI($\mathcal{W}, Q$) problems, the alternating directions method was originally proposed by Gabay [5] and Gabay and Mercier [4]. At each iteration of this method, the new iterate $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^r$ is generated from a given triple $w^k = (x^k, y^k, \lambda^k) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{R}^r$ by the following procedure: First, x^{k+1} is obtained (with y^k and λ^k held fixed) by solving

$$(x' - x^{k+1})^T\Big(f(x^{k+1}) - A^T[\lambda^k - \beta(Ax^{k+1} + By^k - b)]\Big) \geq 0, \qquad \forall\ x' \in \mathcal{X}, \tag{7}$$

and then y^{k+1} is produced (with x^{k+1} and λ^k held fixed) by solving

$$(y' - y^{k+1})^T\Big(g(y^{k+1}) - B^T[\lambda^k - \beta(Ax^{k+1} + By^{k+1} - b)]\Big) \geq 0, \qquad \forall\ y' \in \mathcal{Y}. \tag{8}$$

Finally, the multipliers are updated by

$$\lambda^{k+1} = \lambda^k - \gamma\beta(Ax^{k+1} + By^{k+1} - b), \tag{9}$$

where $\gamma \in (0, \frac{1+\sqrt{5}}{2})$ and $\beta > 0$ are given constants. This method is referred to as a *method of multiplier* in the literature [5], and the convergence proof can be found in [4, 6] (for $B = I$) and [13, 15] (for general B). Further studies and applications of such methods can be found in Glowinski [6], Glowinski and Le Tallec [7], Eckstein and Fukushima [1] and He and Yang [9].

Experience on applications [2, 3, 12] has shown that if the fixed penalty β is chosen too small or too large the solution time can significantly increase. In order to improve such methods, recently, Kontogiorgis and Meyer [12] presented a more general alternating directions method, in which they took a sequence of symmetric positive definite (spd) penalty matrices $\{H_k\}$ instead of the constant penalty β. The convergence of their method was proved under the assumption that the eigenvalues of $\{H_k\}$ are uniformly bounded from below away from zero, and, with finitely many exceptions, the eigenvalues of $H_k - H_{k+1}$ are nonnegative.

In this paper, we continue the Kontogiorgis and Meyer's research [12] and present a modified variable-penalty alternating directions method that allows the eigenvalues of $\{H_k\}$ either to increase or to decrease in each iteration. This can be beneficial in applications. In addition, similarly as in [10], we propose a self-adaptive adjusting rule that leads the method to be more advantageous in practice.

The following notation is used in this paper. We denote by $I_{n\times n}$ the identity matrix in $\mathcal{R}^{n\times n}$. For any real matrix M and vector v, we denote the transposition by M^T and v^T, respectively. The notation $M \succeq 0$ means that M is a positive semi-definite matrix, and $M \succ 0$ means that M is a positive definite matrix. Superscripts such as in v^k refer to specific vectors and are usually iteration indices. The Euclidean norm of vector z will be denoted by $\|z\|$, i.e., $\|z\| = \sqrt{z^T z}$.

2. The General Structure of the Modified Method

Throughout this paper, we call the method by Kontogiorgis and Meyer [12] and our modified method ADM method and MADM method, respectively. To describe the MADM method, we need a non-negative sequence $\{\eta_k\}$ that satisfies $\sum_{k=0}^{\infty} \eta_k < \infty$.

Modified Variable-penalty Alternating Directions Method (short **MADM**)

Step 0. *Given* $\varepsilon > 0$, $\gamma \in (0, \frac{1+\sqrt{5}}{2})$, *a non-negative sequence* $\{\eta_k\}$ *satisfying* $\sum_{k=0}^{\infty} \eta_k < \infty$, *a symmetric positive definite matrix* (spd) H_0, $y^0 \in \mathcal{Y}$ *and* $\lambda^0 \in \mathcal{R}^r$. *Set* $k = 0$.

Step 1. *Convergence verification* ($k \geq 1$)
If $\|e(w^k)\|_\infty < \varepsilon$, stop;

Step 2. *Find* $x^{k+1} \in \mathcal{X}$(*with fixed* y^k *and* λ^k), *such that*

$$(x' - x^{k+1})^T f_k(x^{k+1}) \geq 0, \quad \forall\ x' \in \mathcal{X}. \tag{10}$$

where

$$f_k(x) = f(x) - A^T[\lambda^k - H_k(Ax + By^k - b)]. \tag{11}$$

Step 3. *Find* $y^{k+1} \in \mathcal{Y}$(*with fixed* x^{k+1} *and* λ^k), *such that*

$$(y' - y^{k+1})^T g_k(y^{k+1}) \geq 0, \quad \forall\ y' \in \mathcal{Y}. \tag{12}$$

where

$$g_k(y) = g(y) - B^T[\lambda^k - H_k(Ax^{k+1} + By - b)]. \tag{13}$$

Step 4. *Update*

$$\lambda^{k+1} = \lambda^k - \gamma H_k(Ax^{k+1} + By^{k+1} - b). \tag{14}$$

Step 5. *Adjust the penalty matrix* H_k ($k \geq 1$) *such that,*

$$\frac{1}{1+\eta_k} H_k \preceq H_{k+1} \preceq (1+\eta_k) H_k. \tag{15}$$

Set $k := k+1$, *and go to* Step 1.

Remark 1. In the ADM method [12], the restriction on matrix sequence $\{H_k\}$ can be equivalently described as

$$\frac{1}{1+\eta_k} H_k \preceq H_{k+1} \preceq H_k, \quad \forall k \geq k_0 \quad (\text{with} \quad \eta_k \geq 0 \quad \text{and} \quad \sum_{k=1}^{\infty} \eta_k < \infty). \tag{16}$$

In comparison with (15) and (16), the MADM method does not need the monotonicity assumption and allows the sequence $\{H_k\}$ be more flexible. Furthermore, instead of $\gamma \equiv 1$ in [12], we relax $\gamma \in (0, \frac{1+\sqrt{5}}{2})$.

Remark 2. There are various ways to construct such spd matrix H_k satisfying (15) and we will show some of them in Section 5. Since η_i is non-negative and $\sum_{i=0}^{\infty} \eta_i < \infty$, $\prod_{i=1}^{\infty}(1+\eta_i)$ is convergent and greater than zero. It follows form (15) that the matrix sequence $\{H_k\}$ is both upper and below (from away from zero) bounded. Also from (15) we have

$$\frac{1}{1+\eta_k}\|\cdot\|_{H_k}^2 \leq \|\cdot\|_{H_{k+1}}^2 \leq (1+\eta_k)\|\cdot\|_{H_k}^2, \qquad \frac{1}{1+\eta_k}\|\cdot\|_{H_k^{-1}}^2 \leq \|\cdot\|_{H_{k+1}^{-1}}^2 \leq (1+\eta_k)\|\cdot\|_{H_k^{-1}}^2.$$

Remark 3. For bounded $\{H_k\}$, it can be shown there is a constant $c_0 > 0$, such that

$$\|e(w^{k+1})\|^2 \leq c_0\Big(\|Ax^{k+1} + By^{k+1} - b\|^2 + \|B(y^k - y^{k+1})\|^2\Big). \tag{17}$$

Since solving MVI($\mathcal{W}, Q$) is equivalent to finding a zero point of $e(w)$, we need only to prove that

$$\lim_{k\to\infty} (\|Ax^{k+1} + By^{k+1} - b\|^2 + \|B(y^k - y^{k+1})\|^2) = 0.$$

In fact, it is easy from (10) -(14) to check, if $Ax^{k+1} + By^{k+1} - b = 0$ and $B(y^k - y^{k+1}) = 0$, then $w^{k+1} = (x^{k+1}, y^{k+1}, \lambda^{k+1})$ is a solution of MVI($\mathcal{W}, Q$).

For convenience, we make some basic assumptions to guarantee that the problem under consideration is solvable and the MADM method is well defined.

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

Assumption B. Problems (10) and (12) are solvable.

3. Some Preparations for the Convergence Analysis

In this section, we do some preparations for the convergence analysis. We will investigate the difference between

$$\|\lambda^k - \lambda^*\|^2_{H_k^{-1}} + \gamma\|B(y^k - y^*)\|^2_{H_k} \qquad \text{and} \qquad \|\lambda^{k+1} - \lambda^*\|^2_{H_k^{-1}} + \gamma\|B(y^{k+1} - y^*)\|^2_{H_k}.$$

Now, let us first observe the difference of $\|\lambda^k - \lambda^*\|^2_{H_k^{-1}}$ and $\|\lambda^{k+1} - \lambda^*\|^2_{H_k^{-1}}$. Using (14) and the identity

$$\|\lambda^k - \lambda^*\|^2_{H_k^{-1}} \equiv \|\lambda^{k+1} - \lambda^*\|^2_{H_k^{-1}} - \|\lambda^k - \lambda^{k+1}\|^2_{H_k^{-1}} + 2(\lambda^k - \lambda^*)^T H_k^{-1}(\lambda^k - \lambda^{k+1}),$$

we get

$$\begin{aligned} \|\lambda^k - \lambda^*\|^2_{H_k^{-1}} &= \|\lambda^{k+1} - \lambda^*\|^2_{H_k^{-1}} - \gamma^2\|Ax^{k+1} + By^{k+1} - b\|^2_{H_k} \\ &\quad + 2\gamma(\lambda^k - \lambda^*)^T(Ax^{k+1} + By^{k+1} - b). \end{aligned} \tag{18}$$

The following lemma provides a desirable property of the last term of (18).

Lemma 1. *For any $w^* = (x^*, y^*, \lambda^*) \in \mathcal{W}^*$, we have*

$$\begin{aligned} &(\lambda^k - \lambda^*)^T(Ax^{k+1} + By^{k+1} - b) \\ &\geq \|Ax^{k+1} + By^{k+1} - b\|^2_{H_k} + (Ax^{k+1} - Ax^*)^T H_k(By^k - By^{k+1}). \end{aligned} \tag{19}$$

Proof. Since $w^* \in \mathcal{W}^*$, $x^{k+1} \in \mathcal{X}$ and $y^{k+1} \in \mathcal{Y}$, we have

$$(x^{k+1} - x^*)^T(f(x^*) - A^T\lambda^*) \geq 0 \tag{20}$$

and

$$(y^{k+1} - y^*)^T(g(y^*) - B^T\lambda^*) \geq 0. \tag{21}$$

On the other hand, from (10) and (12), it follows that

$$(x^* - x^{k+1})^T\Big(f(x^{k+1}) - A^T[\lambda^k - H_k(Ax^{k+1} + By^k - b)]\Big) \geq 0, \tag{22}$$

and

$$(y^* - y^{k+1})^T\Big(g(y^{k+1}) - B^T[\lambda^k - H_k(Ax^{k+1} + By^{k+1} - b)]\Big) \geq 0. \tag{23}$$

Adding (20) and (22), and using the monotonicity of operator f, we get

$$(x^{k+1} - x^*)^T\Big(A^T[(\lambda^k - \lambda^*) - H_k(Ax^{k+1} + By^k - b)]\Big) \geq 0. \tag{24}$$

Similarly, adding (21) and (23), and using the monotonicity of operator g, it follows that

$$(y^{k+1} - y^*)^T\Big(B^T[(\lambda^k - \lambda^*) - H_k(Ax^{k+1} + By^{k+1} - b)]\Big) \geq 0. \tag{25}$$

Combining (24) and (25) and using $Ax^* + By^* = b$, we get the assertion of this lemma.

From Lemma 1, we have

$$\begin{aligned} \|\lambda^k - \lambda^*\|^2_{H_k^{-1}} &= \|\lambda^{k+1} - \lambda^*\|^2_{H_k^{-1}} + \gamma(2 - \gamma)\|Ax^{k+1} + By^{k+1} - b\|^2_{H_k} \\ &\quad + 2\gamma(Ax^{k+1} - Ax^*)^T H_k(By^k - By^{k+1}). \end{aligned} \tag{26}$$

In addition, we have the identity

$$\gamma\|B(y^k-y^*)\|^2_{H_k} \equiv \gamma\|B(y^{k+1}-y^*)\|^2_{H_k}+\gamma\|By^k-By^{k+1}\|^2_{H_k} + 2\gamma(By^{k+1}-By^*)^T H_k(By^k-By^{k+1}). \quad (27)$$

Thus, combining (26) and (27) and using $Ax^*+By^*=b$ we get

$$\begin{aligned}&\|\lambda^k-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^k-y^*)\|^2_{H_k}\\ &=\|\lambda^{k+1}-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^{k+1}-y^*)\|^2_{H_k}\\ &\quad+\gamma(2-\gamma)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}+\gamma\|B(y^k-y^{k+1})\|^2_{H_k}\\ &\quad+2\gamma(Ax^{k+1}+By^{k+1}-b)^T H_k(By^k-By^{k+1}).\end{aligned} \quad (28)$$

In the following lemma, we observe the last term in (28).

Lemma 2. *For $k\ge 1$ we have*

$$\begin{aligned}&(Ax^{k+1}+By^{k+1}-b)^T H_k(By^k-By^{k+1})\\ &\quad\ge(1-\gamma)(Ax^k+By^k-b)^T H_{k-1}(By^k-By^{k+1}).\end{aligned} \quad (29)$$

Proof. By setting $y'=y^k$ in (12) we get

$$(y^k-y^{k+1})^T\Big(g(y^{k+1})-B^T[\lambda^k-H_k(Ax^{k+1}+By^{k+1}-b)]\Big)\ge 0. \quad (30)$$

Similarly, taking $k:=k-1$ and $y'=y^{k+1}$ in (12) we have

$$(y^{k+1}-y^k)^T\Big(g(y^k)-B^T[\lambda^{k-1}-H_k(Ax^k+By^k-b)]\Big)\ge 0. \quad (31)$$

By adding (30) and (31) and using the monotonicity of operator g, we obtain

$$(y^{k+1}-y^k)^T B^T\Big([\lambda^k-H_k(Ax^{k+1}+By^{k+1}-b)]-[\lambda^{k-1}-H_{k-1}(Ax^k+By^k-b)]\Big)\ge 0. \quad (32)$$

Substituting $\lambda^k=\lambda^{k-1}-\gamma H_{k-1}(Ax^k+By^k-b)$ in (32), the assertion of this lemma follows immediately.

4. The Main Theorem and the Convergence Proof

Remember that we restrict $\gamma\in(0,\frac{1+\sqrt{5}}{2})$ and hence $1+\gamma-\gamma^2>0$. Let

$$T=2-\frac{1}{3}(1+\gamma-\gamma^2), \quad (33)$$

and then we have

$$2-T=\frac{1}{3}(1+\gamma-\gamma^2)>0 \qquad\text{and}\qquad T-\gamma=\frac{1}{3}(\gamma^2-4\gamma+5)=\frac{1}{3}[(\gamma-2)^2+1]>\frac{1}{3}.$$

Now we are in the stage to prove the main theorem of this paper.

Theorem 1. *Let $w^*=(x^*,y^*,\lambda^*)\in\mathcal{W}^*$ be a solution point of MVI($\mathcal{W},Q$). Let $\{w^k\}=\{(x^k,y^k,\lambda^k)\}$ be the sequence generated by the MADM method, then there is a $k_0\ge 0$, such that*

$$\begin{aligned}&\|\lambda^{k+1}-\lambda^*\|^2_{H_{k+1}^{-1}}+\gamma\|B(y^{k+1}-y^*)\|^2_{H_{k+1}}+\gamma(T-\gamma)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}\\ &\le(1+\eta_k)\Big(\|\lambda^k-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^k-y^*)\|^2_{H_k}+\gamma(T-\gamma)\|Ax^k+By^k-b\|^2_{H_{k-1}}\Big)\\ &\quad-\gamma(2-T)\Big(\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}+\|B(y^k-y^{k+1})\|^2_{H_k}\Big),\qquad\forall k\ge k_0.\end{aligned} \quad (34)$$

Proof. It follows from (28) and (29) that

$$\begin{aligned}&\|\lambda^k-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^k-y^*)\|^2_{H_k}\\ &\geq \|\lambda^{k+1}-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^{k+1}-y^*)\|^2_{H_k}\\ &\quad+\gamma(2-\gamma)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}+\gamma\|B(y^k-y^{k+1})\|^2_{H_k}\\ &\quad+2\gamma(1-\gamma)(Ax^k+By^k-b)^TH_{k-1}(By^k-By^{k+1}).\end{aligned}\tag{35}$$

Using Cauchy-Schwarz inequality, we have

$$\begin{aligned}&2\gamma(1-\gamma)(Ax^k+By^k-b)^TH_{k-1}(By^k-By^{k+1})\\ &\geq -\gamma(T-\gamma)\|Ax^k+By^k-b\|^2_{H_{k-1}}-\frac{\gamma(1-\gamma)^2}{T-\gamma}\|B(y^k-y^{k+1})\|^2_{H_{k-1}}\\ &\geq -\gamma(T-\gamma)\|Ax^k+By^k-b\|^2_{H_{k-1}}-\frac{\gamma(1-\gamma)^2}{T-\gamma}(1+\eta_{k-1})\|B(y^k-y^{k+1})\|^2_{H_k}.\end{aligned}\tag{36}$$

Substituting (36) in (35), we derive

$$\begin{aligned}&\|\lambda^{k+1}-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^{k+1}-y^*)\|^2_{H_k}+\gamma(T-\gamma)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}\\ &\leq \|\lambda^k-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^k-y^*)\|^2_{H_k}+\gamma(T-\gamma)\|Ax^k+By^k-b\|^2_{H_{k-1}}\\ &\quad-\gamma(2-T)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}-\gamma\delta_k\|B(y^k-y^{k+1})\|^2_{H_k},\end{aligned}\tag{37}$$

where

$$\delta_k := 1-\frac{(1-\gamma)^2}{(T-\gamma)}(1+\eta_{k-1}).$$

Note that

$$\delta_k = 1-\frac{(1-\gamma)^2}{(T-\gamma)}-\frac{\eta_{k-1}(1-\gamma)^2}{(T-\gamma)}=\frac{2(1+\gamma-\gamma^2)}{\gamma^2-4\gamma+5}-\frac{\eta_{k-1}(1-\gamma)^2}{(T-\gamma)},\tag{38}$$

the second equality of (38) is obtained by using (33). From $\eta_k\geq 0$ and $\sum_{k=0}^{\infty}\eta_k<\infty$, we have $\lim_{k\to\infty}\eta_k=0$. Then there exists a positive constant k_0 such that

$$0\leq\frac{\eta_{k-1}(1-\gamma)^2}{(T-\gamma)}<\frac{1}{15}(1+\gamma-\gamma^2),\qquad \forall k\geq k_0.$$

Since $\sup_{\gamma\in(0,\frac{1+\sqrt{5}}{2})}\{\gamma^2-4\gamma+5\}=5$, it follows from (38) that for $k\geq k_0$

$$\delta_k\geq\Big(\frac{2}{5}-\frac{1}{15}\Big)(1+\gamma-\gamma^2)=\frac{1}{3}(1+\gamma-\gamma^2)=(2-T).\tag{39}$$

In addition, we have

$$\begin{aligned}&\|\lambda^{k+1}-\lambda^*\|^2_{H_{k+1}^{-1}}+\gamma\|B(y^{k+1}-y^*)\|^2_{H_{k+1}}+\gamma(T-\gamma)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}\\ &\leq(1+\eta_k)\Big(\|\lambda^{k+1}-\lambda^*\|^2_{H_k^{-1}}+\gamma\|B(y^{k+1}-y^*)\|^2_{H_k}+\gamma(T-\gamma)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}\Big)\end{aligned}\tag{40}$$

Substituting (39) and (40) in (37), we obtain the conclusion of the theorem immediately.

Using Theorem 1, we can prove the convergence of our method as follows.

Theorem 2. *Let $\{(x^k,y^k,\lambda^k)\}$ be the sequence generated by the modified variable-penalty alternating directions method for MVI$(\mathcal{W},Q)$. We have*

$$\lim_{k\to\infty}\big(\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}+\|B(y^k-y^{k+1})\|^2_{H_k}\big)=0.\tag{41}$$

Proof. Since $\{\eta_k\}$ is nonnegative and $\sum_{i=1}^{\infty}\eta_i < \infty$, it follows that $\prod_{i=1}^{\infty}(1+\eta_i)$ is bounded. Denote

$$C_s := \sum_{i=1}^{\infty}\eta_i, \qquad C_p := \prod_{i=1}^{\infty}(1+\eta_i). \tag{42}$$

From Theorem 1, we have for all $k \ge k_0$ that

$$\|\lambda^{k+1}-\lambda^*\|^2_{H^{-1}_{k+1}} + \gamma\|B(y^{k+1}-y^*)\|^2_{H_{k+1}} + \gamma(T-\gamma)\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k}$$
$$\le C_p\Big(\|\lambda^{k_0}-\lambda^*\|^2_{H^{-1}_{k_0}} + \gamma\|B(y^{k_0}-y^*)\|^2_{H_{k_0}} + \gamma(T-\gamma)\|Ax^{k_0}+By^{k_0}-b\|^2_{H_{k_0-1}}\Big).$$

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

$$\|\lambda^{k}-\lambda^*\|^2_{H^{-1}_{k}} + \gamma\|B(y^{k}-y^*)\|^2_{H_{k}} + \gamma(T-\gamma)\|Ax^{k}+By^{k}-b\|^2_{H_{k-1}} \le C, \qquad \forall k\ge 0. \tag{43}$$

From Theorem 1 and (43) we get

$$\sum_{i=k_0}^{\infty}\gamma(2-T)\big(\|Ax^{i+1}+By^{i+1}-b\|^2_{H_i} + \|B(y^i-y^{i+1})\|^2_{H_i}\big) \le (1+C_s)C \tag{44}$$

and hence

$$\lim_{k\to\infty}\gamma(2-T)\big(\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k} + \|B(y^k-y^{k+1})\|^2_{H_k}\big) = 0.$$

Since $\gamma\in(0,\frac{1+\sqrt{5}}{2})$ and $2-T=\frac{1}{3}(1+\gamma-\gamma^2)>0$, it follows that

$$\lim_{k\to\infty}\big(\|Ax^{k+1}+By^{k+1}-b\|^2_{H_k} + \|B(y^k-y^{k+1})\|^2_{H_k}\big) = 0$$

and thus Theorem 2 is proved.

Remember that the sequence $\{H_k\}$ is bounded, it follows from Theorem 2 that

$$\lim_{k\to\infty}\big(\|Ax^{k+1}+By^{k+1}-b\|^2 + \|B(y^k-y^{k+1})\|^2\big) = 0$$

and the MADM method is convergent.

5. Numerical Experiments

Let $w^{k+1}=(x^{k+1},y^{k+1},\lambda^{k+1})\in\mathcal{X}\times\mathcal{Y}\times\mathcal{R}^r$ be generated from a given triple $w^k=(x^k,y^k,\lambda^k)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{R}^r$ by (10)-(14) with $\gamma=1$. In this case, it follows that $e_y(w^{k+1})=0$ and thus

$$\|e(w^{k+1})\|^2 = \|e_x(w^{k+1})\|^2 + \|e_\lambda(w^{k+1})\|^2,$$

where

$$e_x(w) = x - P_{\mathcal{X}}\{x-[f(x)-A^T\lambda]\} \qquad \text{and} \qquad e_\lambda(w) = Ax+By-b.$$

For the sake of balance, in [10], the authors adjusted the penalty parameter β such that $\|e_x(w)\|\approx\|e_\lambda(w)\|$. For VI problem (1)-(3) has block form

$$x=(x_1,x_2,\cdots,x_l)^T, \qquad y=(y_1,y_2,\cdots,y_l)^T, \tag{45}$$

$$f(x)) = \big(f_1(x_1),f_2(x_2),\cdots,f_l(x_l)\big)^T, \qquad g(y)=\big(g_1(y_1),g_2(y_2),\cdots,g_l(y_l)\big)^T \tag{46}$$

and

$$\Omega=\{u=(x,y)|x_i\in\mathcal{X}_i, y_i\in\mathcal{Y}_i, A_ix_i+B_iy_i=b_i,\ \forall i=1,\cdots,l.\} \tag{47}$$

we suggest the similar strategy as in [10] for adjusting the penalty matries:

A simple strategy for Adjusting penalty matrix H_k in MADM method.

$$H_{k+1}^{(i)} = \begin{cases} (1+\tau_k)H_k^{(i)}, & \text{if} \quad \|x_i^k - P_{\mathcal{X}_i}[x_i^k - (f_i(x_i^k) - A_i^T\lambda_i^k)]\| < \mu\|A_i x_i^k + B_i y_i^k - b_i\| \ , \\ \dfrac{1}{1+\tau_k}H_k^{(i)}, & \text{if} \quad \mu\|x_i^k - P_{\mathcal{X}_i}[x_i^k - (f_i(x_i^k) - A_i^T\lambda_i^k)]\| > \|A_i x_i^k + B_i y_i^k - b_i\| \ , \\ H_k^{(i)}, & \text{otherwise.} \end{cases}$$

with a symmetric matrix $H_0 = \mathrm{diag}\{H_0^{(1)}, H_0^{(2)}, \cdots, H_0^{(s)}\}$ and $H_0^{(i)} \succ 0$, $i = 1, \cdots, l$.

In the numerical tests, we consider the Fermat-Weber problem [12]

$$\min_{y \in R^n} \sum_{i=1}^{l} a_i \|y - b_{[i]}\| \tag{48}$$

in which the vectors $b_{[i]}$ and the weights $a_i > 0$ are given. For $n = 2$ the problem has a single-facility location interpretation: $b_{[i]}$ are shipment centers, represented as points in the plane; the sought minimizer is the location of the facility to be built, such that the sum of the transportation costs between the centers and the facility is minimized, where each cost is proportional to the Euclidean distance. Introducing auxiliary vectors of $x_{[1]}, \cdots, x_{[l]}$, Problem (48) can be rewritten as

$$\min \{ \sum_{i=1}^{l} a_i \|x_{[i]}\| \mid x_{[i]} = y - b_{[i]}, \ i = 1, \cdots, l \}. \tag{49}$$

Then it can be formulated into the form (45) – (47) with

$$x_i = x_{[i]}, \quad y_i = y, \quad f_i(x_i) = a_i \tfrac{x_i}{\|x_i\|}, \quad g_i(y_i) = 0,$$
$$A_i = I_{n\times n}, \quad B_i = -I_{n\times n}, \quad b_i = -b_{[i]}, \quad \mathcal{X}_i = \mathcal{Y}_i = \mathcal{R}^n, \quad i = 1, \cdots, l.$$

Thus, under the non-degeneracy assumption, we can apply the alternating directions method with self-adaptive block diagonal penalty matrix to solve Problem (49), in which we take $\gamma = 1$, $H_0 = \mathrm{diag}\{\beta_{[1]}^0 I_n, \beta_{([2]}^0 I_n, \cdots, \beta_{(l)}^0 I_n\}$ with $\beta_{[i]}^0 > 0, i = 1, \cdots, l$ and

$$\eta_k = \min\{1, (\max\{1, k - 100\})^{-2}\} = \{1, 1, \ldots, 1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \ldots\}.$$

In particular implementation, we have

$$\beta_{[i]}^{k+1} = \begin{cases} (1+\eta_k)\beta_{[i]}^k & \text{if} \quad \|a_i \frac{x_{[i]}^k}{\|x_{[i]}^k\|} - \lambda_{[i]}^k\| < 0.1\|x_{[i]}^k - y^k + b_{[i]}\|, \\ \beta_{[i]}^k/(1+\eta_k) & \text{if} \quad 0.1\|a_i \frac{x_{[i]}^k}{\|x_{[i]}^k\|} - \lambda_{[i]}^k\| > \|x_{[i]}^k - y^k + b_{[i]}\|, \\ \beta_{[i]}^k & \text{otherwise} \end{cases}$$

and

$$x_{[i]}^{k+1} = \Big(1 - \frac{a_i}{\|\theta_{[i]}^k\|}\Big)\frac{\theta_{[i]}^k}{\beta_{[i]}^k}, \qquad \theta_{[i]}^k = \lambda_{[i]}^k + \beta_{[i]}^k y^k - \beta_{[i]}^k b_{[i]}, \quad i = 1, \cdots, l,$$
$$y^{k+1} = \Big(\sum_{i=1}^{l} \beta_i^k\Big)^{-1} \sum_{i=1}^{l} \Big(\beta_{[i]}^k x_{[i]}^{k+1} + \beta_{[i]}^k b_{[i]} - \lambda_{[i]}^k\Big),$$
$$\lambda_{[i]}^{k+1} = \lambda_{[i]}^k - \beta_i^k(x_{[i]}^{k+1} - y^{k+1} + b_{[i]}), \quad i = 1, \cdots, l.$$

In [12], the author used ADM method for Problem (49) by taking

$$H_0 = \mathrm{diag}\Big\{\frac{2a_1}{\|b_{[1]}\|}I_n, \frac{2a_2}{\|b_{[2]}\|}I_n, \cdots, \frac{2a_l}{\|b_{[l]}\|}I_n\Big\}, \qquad L = \frac{0.075}{nl}\sum_{i=1}^{l} a_i,$$

which are denoted by $H_0[12]$ and $L[12]$ respectively in the sequel, and

$$H_k = \text{diag}\{\beta^k_{[1]} I_n, \beta^k_{[2]} I_n, \cdots, \beta^k_{[l]} I_n\},$$

$$\beta^k_{[i]} = \begin{cases} 1.05\beta^{k-T}_{[i]}, & \text{if } \beta^{k-T}_{[i]} < L, \\ \max\{0.98\beta^{k-T}_{[i]}, L\}, & \text{otherwise}, \end{cases} \quad \text{for } i = 1, \cdots, l,$$

where the penalties were updated every $T(=10)$ iterations.

To assess the impact of the penalty value on performance, we generated 12 classes of data, with the number of points l in $\{25, 50, 75\}$ and the dimension n in $\{2, 4, 8, 16\}$. For each class the problem is generated randomly. The weights a_i were uniformly distributed in [1, 10], while the components of b are uniformly distributed in [10, 100]. The stopping test was $\|e(w^k)\|_\infty \le 10^{-6}$. Table 1 reports the computational results by applying the proposed MADM method and the ADM method in [12] for solving Problem (49), respectively.

Table 1. Number of iterations for Fermat-Weber problems

n	l	$H_0 = 10^{-2}I$		$H_0 = 10^{-1}I$		$H_0 = I$		$H_0 = 10I$		$H_0 = 10^2 I$		$H_0 = H_0[12]$	
		ADM	MADM	ADM	MADM	ADM	MADM	ADM	MADM	ADM	MADM	ADM	MADM
2	25	2044	113	210	63	179	86	1793	97	>10000	101	99	69
	50	2809	55	301	58	149	60	1443	58	>10000	66	141	48
	75	4411	136	464	75	134	65	1294	66	>10000	74	222	67
4	25	678	49	76	38	187	66	1814	66	>10000	77	57	48
	50	1334	52	146	57	171	56	1656	60	>10000	61	97	64
	75	461	52	59	36	187	65	1807	71	>10000	71	56	40
8	25	356	67	47	42	236	69	2298	72	>10000	70	43	38
	50	256	63	37	42	254	72	2473	75	>10000	75	36	38
	75	270	68	38	43	250	72	2436	79	>10000	77	39	37
16	25	182	56	42	57	338	80	3317	84	>10000	78	35	40
	50	162	53	41	55	326	77	3190	78	>10000	78	37	39
	75	168	53	42	58	331	80	3249	81	>10000	82	36	41

It seems that the solution time of the proposed MADM method is much less than that of ADM method. We note that, for the same problem, the iteration numbers of ADM method are significantly depends on the initial penalty. Although the ADM method performs as efficiently as (or somewhat more efficiently than) MADM method when the initial penalty matrix H_0 is selected carefully as in [12], it is inconvenient to choose a proper initial penalty in ADM method for individual problems. Another difficulty encountered by the ADM method is how to choose a proper lower boundary L. As we have seen that $L = \frac{0.075}{nl}\sum_{i=1}^{l} a_i$ in [12] is not an obvious one.

The iteration numbers of the proposed MADM method are insensitive to the initial penalty. To demonstrate it more clearly, we test the Fermat-Weber problem with $n = 16$ and $l = 75$ (the largest problem in Table 1 and 2). The initial penalty matrix H_0 is a positive diagonal matrix whose elements are $\beta^0_{[i]} I_{16}, i = 1, \cdots, 75$. We let $\beta^0_{[i]}$ be uniformly distributed in $(10^{-p}, 10^p)$ and list the iteration numbers in Table 2.

Table 2. Number of iterations of the proposed method for Fermat-Weber problems

$\beta^0_{[i]} \in (10^{-p}, 10^p)$, $p =$	1	2	3	4	5	6	7	8	9	10
Number of iterations	85	89	90	92	91	95	102	105	110	111

Conclusion. In this paper, we proposed a modified variable-penalty alternating directions method. The presented MADM method extends the ADM method by allowing the penalty matrix to vary more flexible. The preliminary numerical tests show that the proposed method with self-adaptive technique is more preferable in practice.

References

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

[2] M. Fortin and R. Glowinski, Eds., Augmented Lagrangian Methods: Applications to the solution of Boundary-Valued Problems, North-Holland, Amsterdam, (1983).

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

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

[5] D. Gabay, Applications of the method of multipliers to variational inequalities, Augmented Lagrange Methods: Applications to the Solution of Boundary-valued Problems, M. Fortin and R. Glowinski, eds., North Holland, Amsterdam, The Netherlands, (1983), 299-331.

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

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

[8] P.T. Harker and J.S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: A Survey of theory, algorithms and applications, *Mathematical Programming*, **48** (1990), 161-220.

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

[10] B.S. He, H. Yang and S.L. Wang, Alternating directions method with self-adaptive parameter for monotone variational inequalities, *JOTA*, **106** (2000), 349-368.

[11] D. Kinderlehhrer and G. Stampacchia, An Introduction to Variational Inequalities and Their Applications, Academic Press, New York, (1980).

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

[13] P.L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numerical Analysis*, **16** (1979), 964-979.

[14] A. Nagurney, Network Economics, A Variational Inequality Approach, Kluwer Academic Publishers, Dordrecht, Boston, London, (1993).

[15] P. Tseng, Applications of splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.,* **29** (1991), 119-138.