Computational Optimization and Applications, 27, 247–267, 2004

Comparison of Two Kinds of Prediction-Correction Methods for Monotone Variational Inequalities

BINGSHENG HE hebma@nju.edu.cn
Department of Mathematics, Nanjing University, Nanjing 210093, People's Republic of China

XIAOMING YUAN
JASON J.Z. ZHANG
Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong

Received October 24, 2001; Revised September 18, 2002

Abstract. In this paper, we study the relationship between the forward-backward splitting method and the extra-gradient method for monotone variational inequalities. Both of the methods can be viewed as prediction-correction methods. The only difference is that they use different search directions in the correction-step. Our analysis explains theoretically why the extra-gradient methods usually outperform the forward-backward splitting methods. We suggest some modifications for the two methods and numerical results are given to verify the superiority of the modified methods.

Keywords: Monotone variational inequalities, forward-backward splitting methods, extra-gradient methods, prediction-correction methods

1. Introduction

Let Ω be a nonempty closed convex subset of R^n and F be a continuous monotone mapping from R^n into itself. A variational inequality problem, denoted by VI(Ω, F), is to determine a vector $u^* \in \Omega$ such that

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

VI(Ω, F) problem includes nonlinear complementarity problems (when $\Omega = R^n_+$) and system of nonlinear equations (when $\Omega = R^n$), and thus it has many important applications [5, 7].

Among the existing methods for nonlinear variational inequality problems, the simplest one is the Goldstein-Levitin-Polyak projection method [6, 17], which starts with any $u^0 \in \Omega$, and iteratively updates u^{k+1} according to the formula

$$\text{(Explicit Method)} \quad u^{k+1} = P_\Omega[u^k - \beta_k F(u^k)], \tag{1.2}$$

where β_k is a chosen positive step size and $P_\Omega(v)$ denotes the projection of v onto Ω. This projection method can be viewed as an explicit method because u^{k+1} occurs only on the left-hand side of the Eq. (1.2). Under suitable assumptions, e.g., F is Lipschitz continuous (with a constant $L > 0$):

$$\|F(u) - F(v)\| \leq L\|u - v\|$$

and uniformly strongly monotone (with a constant modulus $\tau > 0$)

$$(u - v)^T(F(u) - F(v)) \geq \tau\|u - v\|^2,$$

and the step size β_k satisfies

$$0 < \beta_L \leq \beta_k \leq \beta_U < \frac{2\tau}{L^2}, \tag{1.3}$$

the explicit method (1.2) is globally and linearly convergent. However, the efficiency of this method depends on the estimations of the Lipschitz constant L and the uniform strong monotone modulus τ. It is very expensive to estimate the modulus τ and the Lipschitz constant L, even if F is an affine mapping. Hence, in practice, the explicit method (1.2) is used only for well conditioned problems and, in general, uniformly strong monotonicity is a necessary condition.

The forward-backward splitting method [27] and the extra-gradient method [15, 16] are considerably simple projection-type methods in the literature that can overcome the drawback of the Goldstein-Levitin-Polyak method. They are applicable for solving monotone variational inequalities (not necessary strongly monotone). For a given $u \in \Omega$, let

$$p \stackrel{\text{def}}{=} P_\Omega[u - \beta F(u)]. \tag{1.4}$$

Under the assumption

$$\text{(A)} \quad \beta\|F(u) - F(p)\| \leq \nu\|u - p\|, \quad \nu \in (0, 1), \tag{1.5}$$

they both take p as a predictor. Then, the forward-backward splitting method generates the new iterate via

$$\text{(FB)} \quad u_{\text{FB}}^+ = P_\Omega[p + \beta(F(u) - F(p))], \tag{1.6}$$

while the extra-gradient method [16] produces the new iterate by

$$\text{(EG)} \quad u_{\text{EG}}^+ = P_\Omega[u - \beta F(p)]. \tag{1.7}$$

The forward-backward splitting method (1.6) is a special case of the method given in [27] by setting $J_{\beta A} = P_\Omega$. The monotonicity-based analysis [27] can be used to discuss set-valued operator F, provides more general methods and guarantees stronger convergence.

Our interest in this paper, however, is only to compare the efficiencies of the forward-backward splitting method (1.6) and the extra-gradient method (1.7). For any solution point $u^* \in \Omega^*$, let

$$\Theta_{\mathrm{FB}} := \|u - u^*\|^2 - \|u_{\mathrm{FB}}^+ - u^*\|^2. \tag{1.8}$$

and

$$\Theta_{\mathrm{EG}} := \|u - u^*\|^2 - \|u_{\mathrm{EG}}^+ - u^*\|^2. \tag{1.9}$$

We will prove that for two suitably introduced amounts Υ_{FB} and Υ_{EG},

$$\Theta_{\mathrm{FB}} \geq \Upsilon_{\mathrm{FB}} := \Upsilon + \|p + \beta(F(u) - F(p)) - u_{\mathrm{FB}}^+\|^2, \tag{1.10}$$
$$\Theta_{\mathrm{EG}} \geq \Upsilon_{\mathrm{EG}} := \Upsilon + \|p + \beta(F(u) - F(p)) - u_{\mathrm{EG}}^+\|^2, \tag{1.11}$$

and

$$\Upsilon_{\mathrm{EG}} \geq \Upsilon_{\mathrm{FB}} + \|u_{\mathrm{EG}}^+ - u_{\mathrm{FB}}^+\|^2, \tag{1.12}$$

where

$$\Upsilon = \|u - p\|^2 - \beta^2\|F(u) - F(p)\|^2.$$

Moreover, it will be shown by an example that both the inequalities (1.10) and (1.11) are tight. The main result (1.12) indicates that it is likely the extra-gradient method would be better than the forward-backward splitting method.

The paper is organized as follows. In Section 2, we summarize some basic concepts and the consequent results. Sections 3 and 4 study convergence behaviours of the forward-backward splitting method and the extra-gradient method, respectively. Based on the analysis in Sections 3 and 4, the main theoretical result is given in Section 5. In Section 6, we suggest some improvements for both methods. In Section 7, we present some numerical results to indicate that the improvements in the modified methods are significant. Some concluding remarks are addressed in Section 8.

Throughout this paper we assume that the operator F is monotone and Lipschitz continuous on Ω, and the solution set of VI(Ω, F), denoted by Ω^*, is nonempty.

2. Preliminaries

In the following, we state some basic concepts for the variational inequality, which are useful in the following analysis. For convenience, we only consider the projection under the Euclidean norm and do not consider the projection under the general G-norm. However, the general case can be easily extended once the basic ideas are clear. Let F be a mapping on Ω. F is said to be monotone on Ω if

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

Notice that the variational inequality VI(Ω, F) is invariant when we multiply F by some positive scalar β. Thus VI(Ω, F) is equivalent to the following projection equation (see [4])

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

Using (1.4), solving VI(Ω, F) is equivalent to finding a zero point of the residue function

$$e(u, \beta) := u - p. \tag{2.1}$$

Another popular reformulation for VI(Ω, F) is the multi-valued equation

$$0 \in T(u) =: F(u) + N_\Omega(u), \tag{2.2}$$

where $N_\Omega(\cdot)$ is the normal cone operator with respect to Ω, i.e.,

$$N_\Omega(u) := \begin{cases} \{w \mid (v-u)^T w \le 0, \forall\, v \in \Omega\} & \text{if } u \in \Omega, \\ \emptyset & \text{otherwise.} \end{cases} \tag{2.3}$$

Note that $N_\Omega(u)$ is a cone and hence $\beta N_\Omega(u) = N_\Omega(u)$ for all $u \in R^n$ and $\beta > 0$.

A basic property of the projection mapping is

$$(v - P_\Omega(v))^T (P_\Omega(v) - u) \ge 0, \quad \forall\, v \in R^n, \quad \forall\, u \in \Omega. \tag{2.4}$$

It follows that

$$\|P_\Omega(v) - u\|^2 \le \|v - u\|^2 - \|v - P_\Omega(v)\|^2, \quad \forall\, v \in R^n, u \in \Omega. \tag{2.5}$$

Let $u^* \in \Omega^*$ be a solution. For any $u \in R^n$, since $p = P_\Omega[u - \beta F(u)] \in \Omega$ (see (1.4)), it follows from (1.1) that

$$\text{(FI1)} \quad \beta F(u^*)^T \{p - u^*\} \ge 0, \quad \forall\, u \in R^n. \tag{2.6}$$

Setting $v = u - \beta F(u)$ in inequality (2.4) and using (1.4), we obtain

$$\text{(FI2)} \quad \{(u - \beta F(u)) - p\}^T \{p - u^*\} \ge 0, \quad \forall\, u \in R^n. \tag{2.7}$$

Under the assumption that F is monotone we have

$$\text{(FI3)} \quad \{\beta F(p) - \beta F(u^*)\}^T \{p - u^*\} \ge 0, \quad \forall\, u \in R^n. \tag{2.8}$$

The above three fundamental inequalities (2.6)–(2.8) play a very important role in the convergence analysis of projection type methods [9–13, 22–25]. Let

$$g(u, \beta) := \beta F(p) \tag{2.9}$$

and

$$d(u, \beta) := (u - p) - \beta(F(u) - F(p)). \tag{2.10}$$

Adding FI1 and FI3 and using the notation of $e(u, \beta)$, we get

$$(u - u^*)^T g(u, \beta) \geq (u - p)^T g(u, \beta). \tag{2.11}$$

Adding FI1, FI2 and FI3, we have

$$(u - u^*)^T d(u, \beta) \geq (u - p)^T d(u, \beta). \tag{2.12}$$

Moreover, the recursion of the forward-backward splitting method (1.6) can be written as

$$u^+_{\mathrm{FB}} = P_\Omega[u - d(u, \beta)] \tag{2.13}$$

while the formula of the extra-gradient method (1.7) can be written as

$$u^+_{\mathrm{EG}} = P_\Omega[u - g(u, \beta)]. \tag{2.14}$$

We define an amount which is useful in the coming analysis:

$$\Upsilon(\alpha) := 2\alpha e(u, \beta)^T d(u, \beta) - \alpha^2 \|d(u, \beta\|^2. \tag{2.15}$$

3. The general forward-backward splitting method

Using the formulation of the multi-valued equation for VI(Ω, F), the iterative scheme of the Douglas-Rachford operator splitting method [3, 18] is

$$u^{k+1}_{\mathrm{DR}} = (I + \beta F)^{-1}\{(I + \beta N_\Omega)^{-1}[I - \beta F] + \beta F\}u^k. \tag{3.1}$$

Since $N_\Omega(\cdot)$ is the normal cone operator to Ω, $(I + \beta N_\Omega)^{-1}$ is just the orthogonal projection operator onto Ω. Formula (3.1) can be written as

$$u^{k+1} = P_\Omega[u^k - \beta F(u^k)] + \beta(F(u^k) - F(u^{k+1})). \tag{3.2}$$

Because u^{k+1} occurs on both sides of Eq. (3.2), using the terminology in numerical analysis, the Douglas-Rachford operator splitting method can be viewed as an implicit method. Note that if we use the Goldstein-Levitin-Polyak method (1.2) to make a prediction

$$\bar{u} = P_\Omega[u - \beta F(u)] = p \tag{3.3}$$

and then use the Douglas-Rachford operator splitting method to make a correction

$$u^+ = p + \beta(F(u) - F(p)), \tag{3.4}$$

then the resulting formula of *the Prediction-Correction method* can be written as (see notation $d(u, \beta)$ in (2.10))

$$u_{\mathrm{PC}}^{+} = u - d(u, \beta). \tag{3.5}$$

Since

$$u_{\mathrm{FB}}^{+} = P_{\Omega}[u - d(u, \beta)],$$

the forward-backward splitting method can be viewed as a prediction-correction method. Let us study the more general forward-backward splitting method of the following form

$$u_{\mathrm{FB}}^{+}(\alpha) = P_{\Omega}[u - \alpha d(u, \beta)]. \tag{3.6}$$

For convenience, we introduce the notation

$$u_{\mathrm{PC}}^{+}(\alpha) = u - \alpha d(u, \beta). \tag{3.7}$$

Theorem 1. *Let*

$$\Theta_{\mathrm{FB}}(\alpha) := \|u - u^*\|^2 - \|u_{\mathrm{FB}}^{+}(\alpha) - u^*\|^2 \tag{3.8}$$

and

$$\Upsilon_{\mathrm{FB}}(\alpha) := \Upsilon(\alpha) + \|u_{\mathrm{PC}}^{+}(\alpha) - u_{\mathrm{FB}}^{+}(\alpha)\|^2, \tag{3.9}$$

where $\Upsilon(\alpha)$ *is defined in* (2.15). *Then we have*

$$\Theta_{\mathrm{FB}}(\alpha) \geq \Upsilon_{\mathrm{FB}}(\alpha). \tag{3.10}$$

Proof: Since $u_{\mathrm{FB}}^{+}(\alpha) = P_{\Omega}(u_{\mathrm{PC}}^{+}(\alpha))$ (see (3.6) and (3.7)) and $u^* \in \Omega$, it follows from (2.5) that

$$\|u_{\mathrm{FB}}^{+}(\alpha) - u^*\|^2 \leq \|u_{\mathrm{PC}}^{+}(\alpha) - u^*\|^2 - \|u_{\mathrm{PC}}^{+}(\alpha) - u_{\mathrm{FB}}^{+}(\alpha)\|^2, \tag{3.11}$$

and obtains consequently from (3.8) that

$$\Theta_{\mathrm{FB}}(\alpha) \geq \|u - u^*\|^2 - \|u_{\mathrm{PC}}^{+}(\alpha) - u^*\|^2 + \|u_{\mathrm{PC}}^{+}(\alpha) - u_{\mathrm{FB}}^{+}(\alpha)\|^2. \tag{3.12}$$

Using (3.7) and (2.12) we get

$$\|u^k - u^*\|^2 - \|u_{\mathrm{PC}}^{+}(\alpha) - u^*\|^2 \geq 2\alpha e(u, \beta)^T d(u, \beta) - \alpha^2 \|d(u, \beta)\|^2,$$

and the theorem is proved. □

4. The general extra-gradient method

A classical method to solve the multi-valued Eq. (2.2) for VI(Ω, F) is the proximal point algorithm [21], which starts with any vector $u^0 \in \Omega$, and iteratively updates u^{k+1} satisfying the following requirement

$$0 \in (u^{k+1} - u^k) + \beta_k T(u^{k+1}).$$

This is equivalent to

$$-((u^{k+1} - u^k) + \beta_k F(u^{k+1})) \in N_\Omega(u^{k+1}).$$

In other words, for given $u^k \in \Omega$, the new iterate u^{k+1} is obtained via finding

$$u \in \Omega, \quad (u' - u)^T((u - u^k + \beta_k F(u)) \geq 0, \quad \forall u' \in \Omega. \tag{4.1}$$

It means that u^{k+1} is the solution of

$$u = P_\Omega\{u - [(u - u^k) + \beta_k F(u)]\}$$

and thus

$$\text{(Proximal point algorithm)} \quad u^{k+1} = P_\Omega[u^k - \beta_k F(u^{k+1})]. \tag{4.2}$$

Because u^{k+1} occurs on both sides of Eq. (4.2), the proximal point algorithm can also be viewed as an implicit method. Note that if we use the Goldstein-Levitin-Polyak method (1.2) to make a prediction

$$\bar{u} = P_\Omega[u - \beta F(u)] \tag{4.3}$$

and then use the proximal point algorithm to make a correction

$$u^+ = P_\Omega[u - \beta F(\bar{u})], \tag{4.4}$$

the resulting formula is just the extra-gradient recursion

$$u^+_{\mathrm{EG}} = P_\Omega[u - g(u, \beta)].$$

Therefore, the extra-gradient method can also be viewed as a prediction-correction method. Let us study the convergence of the general extra-gradient method of the following form

$$u^+_{\mathrm{EG}}(\alpha) = P_\Omega[u - \alpha g(u, \beta)]. \tag{4.5}$$

Theorem 2. *Let $u^+_{\mathrm{PC}}(\alpha)$ be defined in* (3.7),

$$\Theta_{\mathrm{EG}}(\alpha) := \|u - u^*\|^2 - \|u^+_{\mathrm{EG}}(\alpha) - u^*\|^2 \tag{4.6}$$

and

$$\Upsilon_{\mathrm{EG}}(\alpha) := \Upsilon(\alpha) + \|u^{+}_{\mathrm{PC}}(\alpha) - u^{+}_{\mathrm{EG}}(\alpha)\|^2, \tag{4.7}$$

where $\Upsilon(\alpha)$ *is defined in* (2.15). *Then we have*

$$\Theta_{\mathrm{EG}}(\alpha) \geq \Upsilon_{\mathrm{EG}}(\alpha). \tag{4.8}$$

Proof: Since $u^{+}_{\mathrm{EG}}(\alpha) = P_{\Omega}[u - \alpha g(u, \beta)]$ and $u^* \in \Omega$, it follows from (2.5) that

$$\|u^{+}_{\mathrm{EG}}(\alpha) - u^*\|^2 \leq \|u - \alpha g(u, \beta) - u^*\|^2 - \|u - \alpha g(u, \beta) - u^{+}_{\mathrm{EG}}(\alpha)\|^2, \tag{4.9}$$

and consequently we get

$$\Theta_{\mathrm{EG}}(\alpha) \geq \|u - u^*\|^2 - \|u - u^* - \alpha g(u, \beta)\|^2 + \|u - u^{+}_{\mathrm{EG}}(\alpha) - \alpha g(u, \beta)\|^2.$$

By using (2.11) and a simple manipulation we obtain

$$\Theta_{\mathrm{EG}}(\alpha) \geq \|u - u^{+}_{\mathrm{EG}}(\alpha)\|^2 + 2\alpha e(u, \beta)^T g(u, \beta) - 2\alpha(u - u^{+}_{\mathrm{EG}}(\alpha))^T g(u, \beta). \tag{4.10}$$

Using $g(u, \beta) = d(u, \beta) - \{e(u, \beta) - \beta F(u)\}$, it follows that

$$\begin{aligned}\Theta_{\mathrm{EG}}(\alpha) \geq\; & 2\alpha e(u, \beta)^T g(u, \beta) + 2\alpha(u - u^{+}_{\mathrm{EG}}(\alpha))^T \{e(u, \beta) - \beta F(u)\} \\ & - \alpha^2 \|d(u, \beta)\|^2 + \|(u - u^{+}_{\mathrm{EG}}(\alpha)) - \alpha d(u, \beta)\|^2,\end{aligned} \tag{4.11}$$

which can be rewritten as

$$\begin{aligned}\Theta_{\mathrm{EG}}(\alpha) \geq\; & 2\alpha e(u, \beta)^T d(u, \beta) - \alpha^2 \|d(u, \beta)\|^2 + \|(u - u^{+}_{\mathrm{EG}}(\alpha)) - \alpha d(u, \beta)\|^2 \\ & + 2\alpha\{u - u^{+}_{\mathrm{EG}}(\alpha) - e(u, \beta)\}^T \{e(u, \beta) - \beta F(u)\} \\ =\; & \Upsilon_{\mathrm{EG}}(\alpha) + 2\alpha\{u - u^{+}_{\mathrm{EG}}(\alpha) - e(u, \beta)\}^T \{e(u, \beta) - \beta F(u)\}.\end{aligned} \tag{4.12}$$

Now we consider the last term on the right-hand-side of (4.12). Notice that

$$u - u^{+}_{\mathrm{EG}}(\alpha) - e(u, \beta) = P_{\Omega}[u - \beta F(u)] - u^{+}_{\mathrm{EG}}(\alpha). \tag{4.13}$$

Setting $v := u - \beta F(u)$ and $u := u^{+}_{\mathrm{EG}}(\alpha)$ in the basic inequality (2.4) of the projection mapping, we get

$$\{e(u, \beta) - \beta F(u)\}^T \{P_{\Omega}[u - \beta F(u)] - u^{+}_{\mathrm{EG}}(\alpha)\} \geq 0,$$

and therefore

$$\{u - u^{+}_{\mathrm{EG}}(\alpha) - e(u, \beta)\}^T \{e(u, \beta) - \beta F(u)\} \geq 0. \tag{4.14}$$

Substituting (4.14) into (4.12), it follows that

$$\Theta_{\mathrm{EG}}(\alpha) \geq \Upsilon_{\mathrm{EG}}(\alpha),$$

and the theorem is proved. □

5. The main theoretical result

The assertions of Theorems 1 and 2 are similar. Since

$$\Upsilon(\alpha) := 2\alpha e(u, \beta)^T d(u, \beta) - \alpha^2 \|d(u, \beta)\|^2,$$

it follows from Theorems 1 and 2 that both the general forward-backward splitting method (3.6) and the general extra-gradient method (4.5) are contraction methods for any

$$\alpha \in \left(0, \frac{2e(u, \beta)^T d(u, \beta)}{\|d(u, \beta)\|^2}\right). \tag{5.1}$$

Note that under Assumption (**A**) we have

$$\begin{aligned} e(u, \beta)^T d(u, \beta) &= \|u - p\|^2 - (u - p)^T \{\beta F(u) - \beta F(p)\} \\ &\geq (1 - \upsilon)\|u - p\|^2. \end{aligned} \tag{5.2}$$

In addition, for any $u \notin \Omega^*$, since

$$\begin{aligned} e(u, \beta)^T d(u, \beta) &= \|u - p\|^2 - (u - p)^T \{\beta F(u) - \beta F(p)\} \\ &> \frac{1}{2}\|u - p\|^2 - (u - p)^T \{\beta F(u) - \beta F(p)\} + \frac{1}{2}\|\beta F(u) - \beta F(p)\|^2 \\ &= \frac{1}{2}\|d(u, \beta)\|^2, \end{aligned}$$

we have

$$\tau(u, \beta) := \frac{e(u, \beta)^T d(u, \beta)}{\|d(u, \beta)\|^2} > \frac{1}{2}. \tag{5.3}$$

The following theorem gives a common result for both the general forward-backward splitting method (3.6) and the general extra-gradient method (4.5).

Theorem 3. *Let $d(u, \beta)$ and $g(u, \beta)$ be given by* (2.10) *and* (2.9), *respectively, and*

$$\tau(u, \beta) = \frac{e(u, \beta)^T d(u, \beta)}{\|d(u, \beta)\|^2}, \quad \alpha(u, \beta) = \gamma\tau(u, \beta) \quad \textit{and} \quad \gamma \in (0, 2). \tag{5.4}$$

For given $u^k \in \Omega$, β is chosen such that the assumption (A) is satisfied. Whenever the new iterate u^{k+1} is generated by

$$u^{k+1} = P_\Omega[u^k - \alpha(u^k,\beta)d(u^k,\beta)] \quad \text{or} \quad u^{k+1} = P_\Omega[u^k - \alpha(u^k,\beta)g(u^k,\beta)],$$

we have

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{\gamma(2-\gamma)(1-\nu)}{2}\|e(u^k,\beta)\|^2, \quad \forall u^* \in \Omega^*. \tag{5.5}$$

Proof: From Theorems 1 and 2 we have

$$\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 \ge \Upsilon(\alpha(u^k,\beta)).$$

Using (2.15), (5.4), (5.3) and (5.2) we obtain

$$\begin{aligned}\Upsilon(\alpha(u^k,\beta)) &= 2\alpha(u^k,\beta)e(u^k,\beta)^T d(u^k,\beta) - (\alpha(u^k,\beta))^2\|d(u^k,\beta)\|^2\\ &= 2\alpha(u^k,\beta)e(u^k,\beta)^T d(u^k,\beta) - (\alpha(u^k,\beta)\gamma e(u^k,\beta)^T d(u^k,\beta)\\ &= \tau(u^k,\beta)\gamma(2-\gamma)e(u^k,\beta)^T d(u^k,\beta)\\ &\ge \frac{1}{2}\gamma(2-\gamma)(1-\upsilon)\|e(u^k,\beta)\|^2,\end{aligned}$$

and the assertion is proved. □

In general, Inequality (3.10) in Theorem 1 (resp. (4.8) in Theorem 2) is tight. This can be seen from the following example. Let us consider a VI(Ω, F) with

$$\Omega = R^2, \quad F(u) = Mu \quad \text{and} \quad M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

This variational inequality is monotone and has a unique solution $u^* = 0$. Note that

$$M^2 = -I, \quad \text{and} \quad M^T M = I.$$

For any $u \in R^2$ and $\beta \in (0, 1)$ we have

$$e(u,\beta) = \beta M u \quad \text{and} \quad d(u,\beta) = (I - \beta M)\beta M u = \beta^2 u + \beta M u.$$

Using $u^T M u = 0$ and $\|Mu\| = \|u\|$ we get

$$e(u,\beta)^T d(u,\beta) = \beta^2\|u\|^2 \quad \text{and} \quad \|d(u,\beta)\|^2 = \beta^2(1+\beta^2)\|u\|^2.$$

When the problem is solved by the general forward-backward splitting method (3.6) with $\alpha \in (0, \frac{2}{1+\beta^2})$, we have

$$\|u_{\mathrm{FB}}^+(\alpha)\|^2 = \|(1-\alpha\beta^2)u - \alpha\beta M u\|^2 = (1 - 2\alpha\beta^2 + \alpha^2\beta^2(1+\beta^2))\|u\|^2$$

and

$$\Theta_{FB}(\alpha) = (2\alpha\beta^2 - \alpha^2\beta^2(1+\beta^2))\|u\|^2 = \Upsilon_{FB}(\alpha).$$

For the general extra-gradient method, since in this special example

$$e(u,\beta) - \beta F(u) = 0 \quad \text{and hence} \quad g(u,\beta) = d(u,\beta),$$

we have also $\Theta_{FG}(\alpha) = \Upsilon_{EG}(\alpha)$, which means that Inequality (4.8) is tight in this example.

Nevertheless, the following theorem indicates that in each iterative step, we may expect the general extra-gradient method (4.5) to get more progress than the general forward-backward splitting method (3.6).

Theorem 4. *Let $\Upsilon_{EG}(\alpha)$ and $\Upsilon_{FB}(\alpha)$ be defined as in Theorems* 1 *and* 2. *We have*

$$\Upsilon_{EG}(\alpha) - \Upsilon_{FB}(\alpha) \geq \|u^+_{EG}(\alpha) - u^+_{FB}(\alpha)\|^2. \tag{5.6}$$

Proof: It follows from (3.9) and (4.7) that

$$\Upsilon_{EG}(\alpha) - \Upsilon_{FB}(\alpha) = \|u^+_{PC}(\alpha) - u^+_{EG}(\alpha)\|^2 - \|u^+_{PC}(\alpha) - u^+_{FB}(\alpha)\|^2. \tag{5.7}$$

Note that (see (3.6) and (3.7))

$$u^+_{FB}(\alpha) = P_\Omega(u^+_{PC}(\alpha))$$

and $u^+_{FB}(\alpha) \in \Omega$. Setting $v = u^+_{PC}(\alpha)$ and $u = u^+_{EG}(\alpha)$ in (2.5) we obtain

$$\|u^+_{FB}(\alpha) - u^+_{EG}(\alpha)\|^2 \leq \|u^+_{PC}(\alpha) - u^+_{EG}(\alpha)\|^2 - \|u^+_{PC}(\alpha) - u^+_{FB}(\alpha)\|^2. \tag{5.8}$$

The assertion of this theorem follows directly from (5.7) and (5.8) □

6. Modifications of the both methods

From the analysis of Sections 3 and 4, both the general forward-backward splitting method and the general extra-gradient method can be viewed as prediction-correction methods that use Goldstein-Levitin-Polyak formula to make a prediction. The difference is only in the correction step, in which the forward-backward splitting method utilizes the Douglas-Rachford operator splitting formula [3, 18] while the extra-gradient method uses the proximal point formula [21]. Nevertheless, both methods use the same step size in the correction step. In this section, we discuss how to improve efficiency of such methods via choosing suitable $\{\beta_k\}$ in the prediction step and $\{\alpha_k\}$ in the correction step. The modifications are based on a practical code of the extra-gradient method by Khobotov [15]. This technology provides an acceptable β_k satisfying Assumption (A) after reducing its value for a finite number of times whenever the operator F is Lipschitz continuous. In each iteration, the final accepted step size is less than or equal to the probe step size. Moreover, the accepted step size β_k in the k-th iteration is taken as the probe step size in the $(k+1)$-th iteration. Hence, the sequence $\{\beta_k\}$ of the accepted step sizes is monotonically non-increasing. Since the sequence $\{\beta_k\}$ generated by the basic extra-gradient method also satisfies the step rule in [27], our discussion about the forward-backward splitting methods is obtained by substituting $g(u, \beta)$ in EG method by $d(u, \beta)$.

6.1. *Modifications under Assumption (A)*

The analysis of this section is similar to which in [14] for extra-gradient method. It seems that the step size α_k in both (3.6) and (4.5) should depend on the current point u^k and the step size β_k in the prediction step. This motivates our following analysis which aims at finding a 'best' α for given u^k and β_k. The analysis is based on

$$\begin{aligned} \Theta_{\mathrm{FB}}(\alpha) &= \|u-u^*\|^2 - \|u^+_{\mathrm{FB}}(\alpha)-u^*\|^2 \geq \Upsilon_{\mathrm{FB}}(\alpha),\\ \Theta_{\mathrm{EG}}(\alpha) &= \|u-u^*\|^2 - \|u^+_{\mathrm{EG}}(\alpha)-u^*\|^2 \geq \Upsilon_{\mathrm{EG}}(\alpha),\\ &\Upsilon_{\mathrm{EG}}(\alpha) \geq \Upsilon_{\mathrm{FB}}(\alpha) \geq \Upsilon(\alpha), \end{aligned}$$

and recall that

$$\Upsilon(\alpha) = 2\alpha e(u,\beta)^T d(u,\beta) - \alpha^2\|d(u,\beta)\|^2. \tag{6.1}$$

The right-hand-side of (6.1) is a quadratic function of α and it reaches its maximum at

$$\alpha^* = \tau(u,\beta) \quad \text{where } \tau(u,\beta) = \frac{e(u,\beta)^T d(u,\beta)}{\|d(u,\beta)\|^2}. \tag{6.2}$$

Thus from (6.1) and (6.2) we have the maximum value of $\Upsilon(\alpha)$:

$$\Upsilon(\alpha^*) = \alpha^* e(u,\beta)^T d(u,\beta) = \tau(u,\beta)\cdot e(u,\beta)^T d(u,\beta). \tag{6.3}$$

Based on the above analysis, for given u and β, the 'ideal' step size α^*_k in the correction step is given by (6.2). However, in the above considerations, we ignored the part $\Theta_k(FB_\alpha) - \Upsilon_k(\alpha)$ (resp. $\Theta_k(EG_\alpha) - \Upsilon_k(\alpha)$). To ensure a faster convergence, we use a relaxation factor $\gamma \in [1, 2)$ but close to 2. Thus, in the correction step we take the step size

$$\text{(Modification 1)} \quad \alpha_k = \gamma\tau(u^k,\beta_k), \quad \gamma \in [1,2).$$

Instead of $\alpha_k \equiv 1$ in the basic FB and EG methods, we determine α_k via *Modification 1* and this requires only $O(n)$ extra flops.

The sequence $\{\beta_k\}$ in the FB and EG methods is monotonically non-increasing. However, this may cause a slow convergence if

$$r_k := \frac{\beta_k\|F(u^k) - F(P_\Omega[u^k - \beta_k F(u^k)])\|}{\|u^k - P_\Omega[u^k - \beta_k F(u^k)]\|}$$

is too small. In this situation, enlarging the probe step size β for the next iteration is necessary. Therefore, when we start the $(k+1)$-th iteration, we take

$$\text{(Modification 2)} \quad \beta_{k+1} := \begin{cases} \dfrac{0.8\nu}{r_k}\beta_k & \text{if} \quad r_k \le \mu, \\ \beta_k & \text{otherwise} \end{cases}$$

where $\mu \in (0, 0.5)$ is a constant.

6.2. *Improvements under new conditions*

In fact, the method is well defined when $\alpha^* = \tau(u, \beta) > 0$ (see (6.2) or [25]). In order to guarantee $\tau(u, \beta) > 0$ we need only to ensure $(u-p)^T\beta(F(u) - F(p)) < \|u-p\|^2$. In other words, Assumption (A) (i. e., $\beta\|F(u) - F(p)\| < \|u - p\|$, see (1.5)) is not necessary for convergence. However, from numerical point of view, $\beta\|F(u) - F(p)\| \approx \|u - p\|$ is favorable for fast convergence. According to our numerical experiences, instead of Assumption (A), we suggest to take the following condition (C):

$$\text{(C)} \quad \frac{(u-p)^T\beta(F(u) - F(p))}{\|u-p\|^2} \le \frac{2}{3} \quad \text{and} \quad \frac{\beta\|F(u) - F(p)\|}{\|u-p\|} \le \frac{3}{2}. \tag{6.4}$$

By considering the above improvements, we obtain the following improved prediction-correction methods.

The improved prediction-correction methods under Condition (C)

Step 0. Let $\beta_0 > 0$, $0 < \mu < 1 < \upsilon$, $u^0 \in \Omega$, $\gamma \in [1, 2)$ and $k = 0$.
Step 1. Set $\bar{u}^k = P_\Omega[u^k - \beta_k F(u^k)]$.

Step 2. If $\boxed{\frac{(u^k-\bar{u}^k)^T\beta_k(F(u^k)-F(\bar{u}^k))}{\|u^k-\bar{u}^k\|^2} \le \frac{2}{3} \quad \text{and} \quad r_k := \frac{\beta_k\|F(u^k)-F(\bar{u}^k)\|}{\|u^k-\bar{u}^k\|} \le \nu}$

then set $\quad e(u^k, \beta_k) = u^k - \bar{u}^k, \quad g(u^k, \beta_k) = \beta_k F(\bar{u}^k),$
$d(u^k, \beta_k) = e(u^k, \beta_k) - \beta_k F(u^k) + g(u^k, \beta_k),$

$\boxed{\tau_k = \frac{e(u^k,\beta_k)^T d(u^k,\beta_k)}{\|d(u^k,\beta_k)\|^2}, \quad \alpha_k = \gamma\tau_k,}$ *Modification* 1

$u^{k+1} = P_\Omega[u^k - \alpha_k d(u^k, \beta_k)],$ (in the forward-backward splitting method)

or
$u^{k+1} = P_\Omega[u^k - \alpha_k g(u^k, \beta_k)],$ (in the extra-gradient method).

$\boxed{\beta_k := \begin{cases} \frac{0.8\upsilon}{r_k}\beta_k & \text{if}\, r_k \le \mu, \\ \beta_k & \text{otherwise,} \end{cases}}$ *Modification* 2

$\beta_{k+1} = \beta_k \quad \text{and} \quad k := k+1, \quad$ go to Step 1.

Step 3. Reduce the value of β_k by $\beta_k := \frac{3}{4}\beta_k * \min\{1, \frac{v}{r_k}\}$,

$$\text{set } \bar{u}^k = P_\Omega[u^k - \beta_k F(u^k)] \quad \text{and go to Step 2.}$$

Remark. Modification 2 is to avoid $\beta\|F(u) - F(p)\| \ll \|u - p\|$ while $r_k \le v$ is to avoid $\beta\|F(u) - F(p)\| \gg \|u - p\|$. In this way we balance $\beta\|F(u) - F(p)\|$ and $\|u - p\|$ in the iteration process.

7. Numerical experiments

The purpose of this section is to verify the theoretical assertions via numerical experiments.

7.1. The first set of test examples

In order to see the effects of Modifications 1 and 2 under Assumption (A), we test a set of nonlinear complementarity problems

$$u \ge 0, \quad F(u) \ge 0, \quad u^T F(u) = 0 \tag{7.1}$$

with the following different forms of the improved methods.

Methods of FB type	Methods of EG type
• FBA (FB method under Assum. (A))	• EGA (EG method under Assum. (A))
• FBA_1 (FBA with Modification 1)	• EGA_1 (EGA with Modification 1)
• FBA_2 (FBA with Modification 2)	• EGA_2 (EGA with Modification 2)
• FBA_{12} (FBA with Modifications 1 & 2)	• EGA_{12} (EGA with Modifications 1 & 2)

Note that in each iteration the computational costs of these methods are almost equal.

In our test problems we take

$$F(u) = D(u) + Mu + q, \tag{7.2}$$

where $D(u)$ and $Mu + q$ are the nonlinear part and the linear part of $F(u)$, respectively.

We form the linear part in the test problems similarly as in [8].[1] The matrix $M = A^T A + B$, where A is an $n \times n$ matrix whose entries are randomly generated in the interval $(-5, +5)$ and a skew-symmetric matrix B is generated in the same way. The vector q is generated from a uniform distribution in the interval $(-500, 500)$. In $D(u)$, the nonlinear part of $F(u)$, the components are $D_j(u) = d_j * \arctan(u_j)$, where d_j is a random variable in $(0, 1)$. A similar type of (small) problems was tested in [20] and [26].[2] It is easy to see that the Jacobian matrix of mapping F is positive semidefinite (not necessary symmetric) and hence the problem is monotone. In details, by using the pseudo random numbers, we form A, B, q and d via the following Matlab code:

```
A=zeros(n,n); t=0; for i=1:n  for j=1:n  t=mod(t*31416+13846,46261);
  A(i,j)=t*(10/46261)-5; end; end;
                 %% A is a random matrix and A_{ij} is in (-5,5) %%
B=zeros(n,n); t=0; for i=1:n  for j=i+1:n  t=mod(t*42108+13846,46273);
  B(i,j)=t*10/46273-5; B(j,i)= - B(i,j); end; end;
                 %% B is skew-symmetric and B_{ij} is in (-5,5) %%
M= A'*A+B; t=0;
q=zeros(n,1); for j=1:n  t=mod(t*45278+13846,46219); q(j)=t; end;
  q=(q/46219 -0.5)*1000;
d=zeros(n,1); for j=1:n  t=mod(t*45278+13846,46219); d(j)=t; end;
  d=d/46219;
```

All codes are written in Matlab and run on a PIII-600 Acer notebook computer. We took $\nu = 0.9$ in Assumption (A), $\gamma = 1.8$ in Modification 1 and $\mu = 0.3$ in Modification 2. The computation begins with $u^0 = 0$, $\beta_0 = 1$ and stop as soon as $\|e(u^k, 1)\|_\infty \leq 10^{-7}$. We test the problems with $n = 100$, 200 and 500. The test results with methods of the FB type and the EG type are reported in Tables 1 and 2, respectively.

First, in each case, the method of EG type needs fewer iterations than the corresponding method of FB type, and this advantage of the EG method is even more obvious when modifications 1 and/or 2 are taken. Second, we can see from Tables 1 and 2 that the modifications 1 and 2 in both FB type methods and EG type methods are effective. They reduce the numbers of iteration remarkably and lead to a rapid convergence. In fact, the numerical results agree with the theoretical analysis.

Table 1. Numerical results for methods of the FB type.

	Method FBA		Method FBA_1		Method FBA_2		Method FBA_{12}	
n	No. of it.	CPU (sec.)	No. of it.	CPU (sec.)	No. of it.	CPU (sec.)	No. of it.	CPU (sec.)
100	737	1.21	488	0.83	670	1.10	357	0.66
200	1226	6.76	636	3.52	904	4.89	502	2.80
500	1158	30.60	671	17.91	983	26.47	534	14.56

Table 2. Numerical results for methods of the EG type.

	Method EGA		Method EGA_1		Method EGA_2		Method EGA_{12}	
n	No. of it.	CPU (sec.)	No. of it.	CPU (sec.)	No. of it.	CPU (sec.)	No. of it.	CPU (sec.)
100	731	1.21	383	0.72	562	0.94	305	0.55
200	844	4.61	460	2.53	804	4.40	438	2.48
500	1131	29.99	467	12.42	849	23.01	476	13.13

Table 3. The coordinates of the 10 regular points in [28].

	x-coordinate	y-coordinate		x-coordinate	y-coordinate
$b_{[1]}$	7.436490	7.683284	$b_{[6]}$	1.685912	1.231672
$b_{[2]}$	3.926097	7.008798	$b_{[7]}$	4.110855	0.821114
$b_{[3]}$	2.309469	9.208211	$b_{[8]}$	4.757506	3.753666
$b_{[4]}$	0.577367	6.480938	$b_{[9]}$	7.598152	0.615836
$b_{[5]}$	0.808314	3.519062	$b_{[10]}$	8.568129	3.079179

7.2. *The second test example*

The purpose of our second test example is to indicate that the improved extra-gradient method is applicable for some scientific problems. The problem is Example 1 in [28] for finding the shortest network in a given full Steiner topology. In this example, $P = \{b_{[1]}, \ldots, b_{[10]}\}$ are given points in R^2 (called regular points) whose coordinates are given in Table 3. $x_{[1]}, \ldots, x_{[8]}$ are the locations of the additional points (called Steiner points). The points-edges connections of the network are depicted in figure 1.

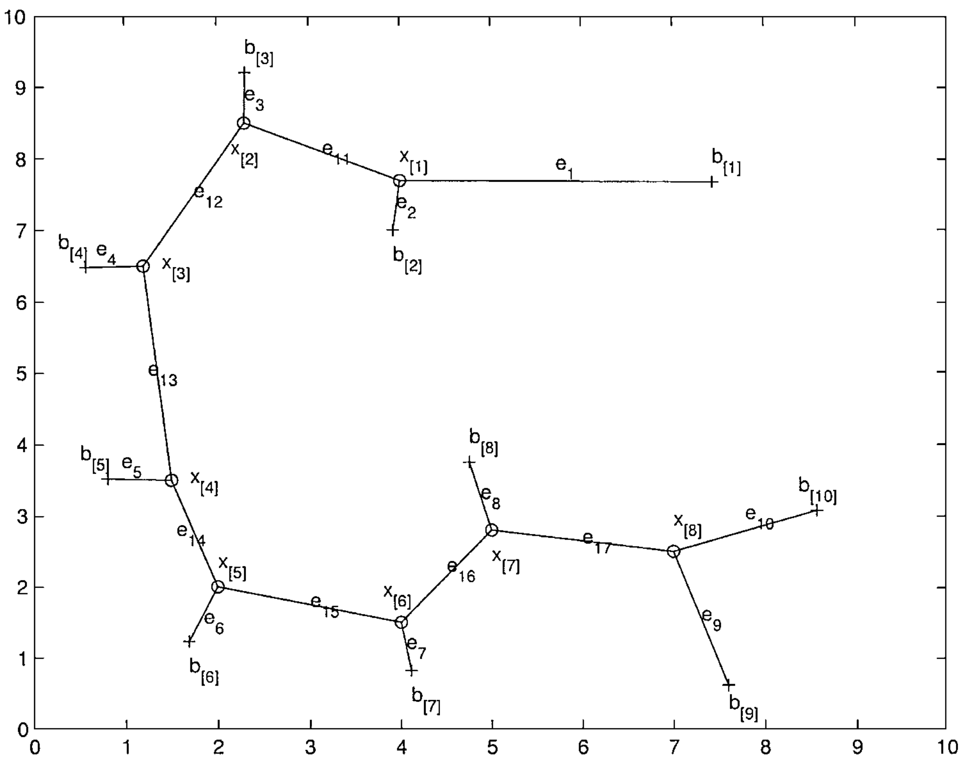

Figure 1. The points-edges connections of Example 1 in [28].

The mathematical form of the problem is

$$\min_{x_{[j]}\in R^2}\left\{\|x_{[1]}-b_{[1]}\|_2+\sum_{j=1}^{8}\|x_{[j]}-b_{[j+1]}\|_2+\|x_{[8]}-b_{[10]}\|_2\right.$$
$$\left.+\sum_{j=1}^{7}\|x_{[j]}-x_{[j+1]}\|_2\right\} \tag{7.3}$$

For any $d \in \mathcal{R}^2$, since

$$\|d\|_2=\max_{\xi\in B_2}\xi^T d, \quad \text{where } B=\{\xi\in\mathcal{R}^2 \mid \|\xi\|_2\le 1\}, \tag{7.4}$$

problem (7.3) is equivalent to the following min-max problem

$$\min_{x_{[j]}\in R^2}\max_{z_{[i]}\in B}\left\{\begin{array}{l} z_{[1]}^T\big(x_{[1]}-b_{[1]}\big)+\sum_{j=1}^{8} z_{[j+1]}^T\big(x_{[j]}-b_{[j+1]}\big) \\ \quad +z_{[10]}^T\big(x_{[8]}-b_{[10]}\big)+\sum_{j=1}^{7} z_{[j+10]}^T\big(x_{[j]}-x_{[j+1]}\big)\end{array}\right\}, \tag{7.5}$$

where $z_{[i]}$, $i=1,\dots,17$ are vectors in $\mathcal{R}^2$. In fact, $z_{[i]}$ is the dual variable ordered to edge e_i. The compact form of (7.5) is

$$\min_{x\in\mathcal{R}}\max_{z\in\mathcal{B}} z^T(Ax-b) \tag{7.6}$$

where

$$\begin{array}{ll} x^T=\big(x_{[1]}^T,\dots,x_{[8]}^T\big)^T, & z^T=\big(z_{[1]}^T,\dots,z_{[17]}^T\big)^T \\ \mathcal{R}=R^2\times\cdots\times R^2, & \mathcal{B}=B\times\cdots\times B, \end{array} \tag{7.7}$$

A is block matrix which has form

$$A=\begin{pmatrix} I_2 & & & \\ I_2 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & I_2 \\ & & & I_2 \\ I_2 & -I_2 & & \\ & \ddots & \ddots & \\ & & I_2 & -I_2 \end{pmatrix} \quad \text{and} \quad b=\begin{pmatrix} b_{[1]} \\ b_{[2]} \\ \vdots \\ \vdots \\ b_{[9]} \\ b_{[10]} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{7.8}$$

Let $(x^*, z^*) \in \mathcal{X} \times \mathcal{B}$ be any solution of (7.6), it follows that

$$z^T(Ax^* - b) \le z^{*T}(Ax^* - b) \le z^{*T}(Ax - b), \quad \forall x \in \mathcal{R}, z \in \mathcal{B}.$$

Thus, (x^*, z^*) is a solution of the following variational inequality:

$$x^* \in \mathcal{R}, z^* \in \mathcal{B}_2, \quad \begin{cases} (x - x^*)^T(A^T z^*) \ge 0, & \forall x \in \mathcal{R}, \\ (z - z^*)^T(-Ax^* + b) \ge 0, & \forall z \in \mathcal{B}, \end{cases} \tag{7.9}$$

The compact form of (7.9) is the following linear variational inequality:

$$\mathrm{LVI}(\Omega, M, q) \quad u^* \in \Omega, \quad (u - u^*)^T(Mu^* + q) \ge 0, \quad \forall u \in \Omega, \tag{7.10}$$

Table 4. Part of the output of the improved extra-gradient method with $x^0 = 0$ for Example 1 in [28].

Iteration	Network-cost	$\|e(u)\|$	Iteration	Network-cost	$\|e(u)\|$
1	67.4046273974	3.2e + 000	80	25.3560680147	2.1e–007
20	25.6180798312	2.3e–001	100	25.3560677830	2.8e–009
40	25.3585237874	1.4e–003	106	25.3560677800	7.7e–010
60	25.3560897957	1.6e–005	116	25.3560677794	9.3e–011

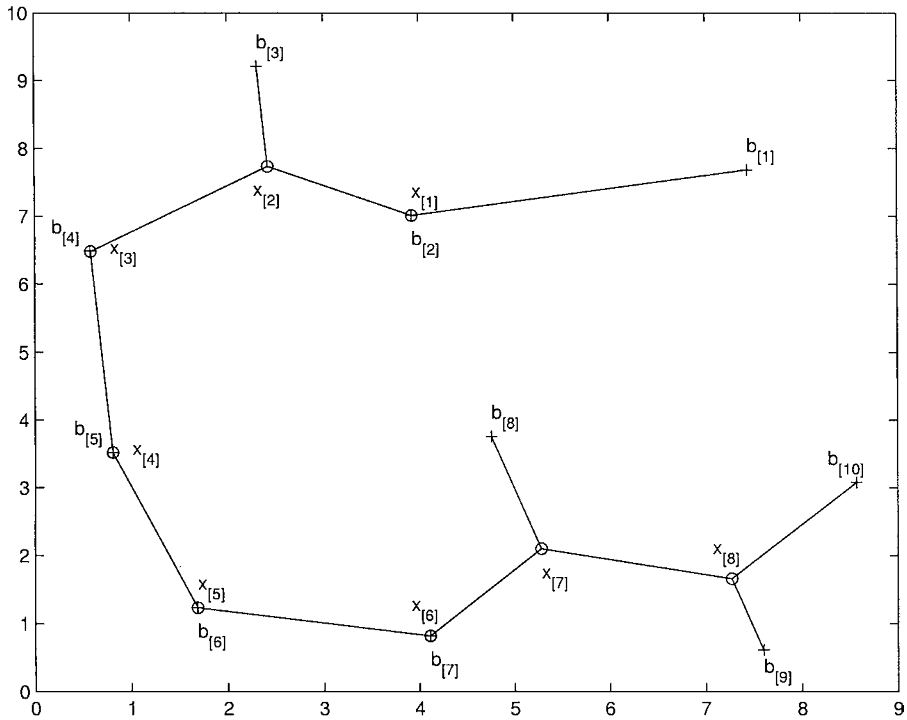

Figure 2. The shortest network of Example 1 in [28] under l_2-norm.

where

$$u = \begin{pmatrix} x \\ z \end{pmatrix}, \quad M = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}, \quad q = \begin{pmatrix} 0 \\ b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{R} \times \mathcal{B}. \tag{7.11}$$

Note that M is skew-symmetric and thus the linear variational inequality is monotone.

The improved method under Condition (C) was applied to solve the resulting LVI. We took $\nu = \frac{3}{2}$ in Condition (C), $\gamma = 1.5$ in Modification 1 and $\mu = 0.6$ in Modification 2. By taking $\beta_0 = 1$ and $u^0 = 0$, the improved algorithm of EG type reached $\|e(u)\|_2 \leq 10^{-10}$ in a total 116 iterations.

Since some Steiner points in the shortest network coincide with regular points, the problem is degenerate. Table 4 shows a part of the computer output in this experiment. For the same problem, the start point in [28] was $x^0 = 0$ and thus we have the same start network-cost at the beginning of computation. The algorithm in [28] reached the final cost 25.3560677802 in 23 iterations, while our algorithm reaches this cost in 106 iterations (see

Table 5. Number of iterations of the improved EG and FB methods under conditions (C) for Example 2 with different β_0.

β_0	0.0001	0.001	0.01	0.1	1	10	100	1000	10000
Method EG type	128	128	128	126	116	127	127	127	127
Method FB type	143	143	143	138	146	145	145	145	145

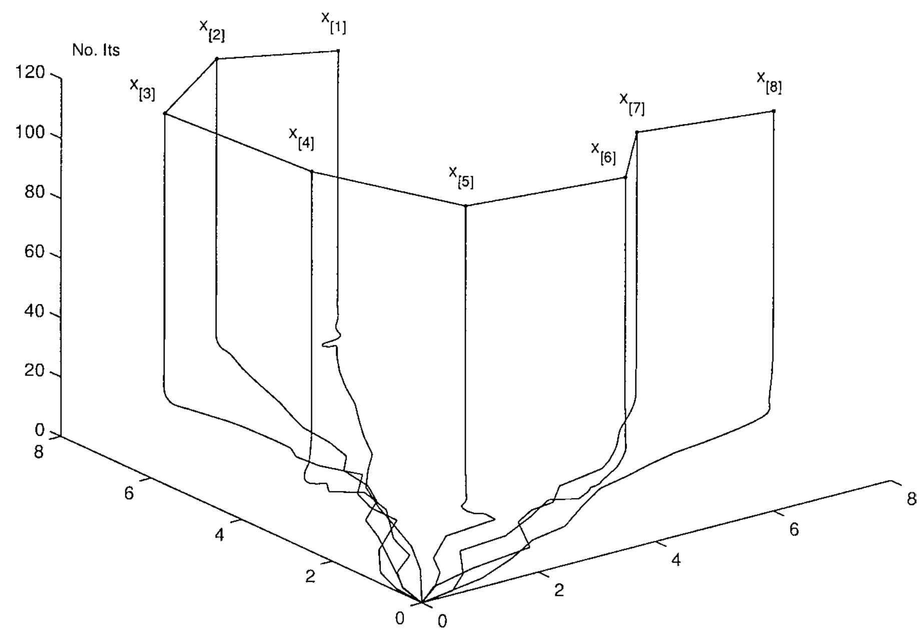

Figure 3. Convergence tendency of the Steiner points.

Table 4). It seems that we need about 4.5 times of the iterations that the interior point algorithm needs. However, the cost of each iteration in our algorithm is much lower. So, it shows that the method is applicable to some scientific problems.

The shortest network under l_2-norm is depicted in figure 2. The convergence tendency of the Steiner points of the iterates with the start point $x^0 = 0$ is given in figure 3.

Similarly, under condition (C), the method of EG type needs fewer iterations than the method of FB type. We tested this example with different start scaling parameters β_0, and the number of iterations of different methods are given in Table 5.

However, without the balancing strategy described in Section 6, the methods converge extremely slowly when the starting parameter β_0 is too large or too small.

8. Concluding remarks

We have investigated the relationship between the forward-backward splitting methods and the extra-gradient methods for monotone variational inequalities. For a given $u^k \in \Omega$, both methods take the same prediction step. Once the same prediction step size is determined, the two methods have the same range (5.1) for the correction step size. The only difference is that they use different search directions in the correction step. The computational costs of the two methods in each iteration are almost equal. The proof of Theorem 4 (for the general extra-gradient methods) is a little more complicated than the proof of Theorem 3 (for the general forward-backward splitting methods). Both the theoretical analysis and the numerical experience show that in general we can not expect the forward-backward splitting methods to have better performance than the extra-gradient methods. The preliminary numerical experiments show that the improved extra-gradient method is applicable for some scientific problems.

Acknowledgment

The authors thank the referees for their valuable comments and suggestions which helped us improve the presentation of this paper.

B.S. He was supported by the NSFC grant 10271054 and Jiangsu NSF grant BK2002075. The research of J.J.Z. Zhang is partially supported by City U Strategic Research Grant No. 7001258.

Notes

1. In the paper by Harker and Pang [8], the matrix $M = A^T A + B + D$, where A and B are the same matrices as what we used here, and D is a diagonal matrix with uniformly distributed random variable $d_{jj} \in (0.0, 0.3)$.
2. In [20] and [26], the components of the nonlinear mapping $D(u)$ are $D_j(u) = constant * \arctan(u_j)$.

References

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

2. R.W. Cottle, J.S. Pang, and R.E. Stone, The Linear Complementarity Problem: Academic Press: New York, 1992.
3. J. Douglas and H.H. Rachford, "On the numerical solution of the heat conduction problem in 2 and 3 space variables," Trans. Amer. Math. Soc., vol. 82, pp. 421–439, 1956.
4. B.C. Eaves, "On the basic theorem of complementarity," Mathematical Programming, vol. 1, pp. 68–75, 1971.
5. R. Glowinski, Numerical Methods for Nonlinear Variational Problems. Springer-Verlag: New York, Berlin, Heidelberg, Tokyo, 1984.
6. A.A. Goldstein, "Convex programming in Hilbert space," Bulletin of American Mathematical Society, vol. 70, pp. 709–710, 1964.
7. P.T. Harker and J.S. Pang, "Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory algorithms and applications," Mathematical Programming, vol. 48, pp. 161–220, 1990.
8. P.T. Harker and J.S. Pang, "A Damped-Newton method for the linear complementarity problem," Lectures in Applied Mathematics, vol. 26, pp. 265–284, 1990.
9. B.S. He, "A projection and contraction method for a class of linear complementarity problem and its application in convex quadratic programming," Applied Mathematics and Optimization, vol. 25, pp. 247–262, 1992.
10. B.S. He, "A new method for a class of linear variational inequalities," Mathematical Programming, vol. 66, pp. 137–144, 1994.
11. B.S. He, "Solving a class of linear projection equations," Numerische Mathematik, vol. 68, pp. 71–80, 1994.
12. B.S. He, "A class of projection and contraction methods for monotone variational inequalities," Applied Mathematics & Optimization, vol. 35, pp. 69–76, 1997.
13. B.S. He, "Inexact implicit methods for monotone general variational inequalities," Mathematical Programming, vol. 86, pp. 199–216, 1999.
14. B.S. He and L.Z. Liao, "Improvements of some projection methods for monotone nonlinear variational inequalities," J. of Optimization Theory and Applications, vol. 112, pp. 111–128, 2002.
15. E.N. Khobotov, "Modification of the extragradient method for solving variational inequalities and certain optimization problems," U.S.S.R. Comput. Math. Phys., vol. 27, pp. 120–127, 1987.
16. G.M. Korpelevich, "The extragradient method for finding saddle points and other problems," Ekonomika i Matematchskie Metody, vol. 12, pp. 747–756, 1976.
17. E.S. Levitin and B.T. Polyak, "Constrained minimization problems," USSR Computational Mathematics and Mathematical Physics, vol. 6, pp. 1–50, 1966.
18. P.L. Lions and B. Mercier, "Splitting algorithms for the sum of two nonlinear operators," SIAM J. Numerical Analysis, vol. 16, pp. 964–979, 1979.
19. P. Marcotte, "Application of Khobotov's algorithm to variational inequalities and network equilibrium problems," Inf. Syst. Oper. Res., vol. 29, pp. 258–270, 1984.
20. P. Marcotee and J.P. Dussault, "A note on a globally convergent newton method for solving variational inequalities," Operation Research Letters, vol. 6, pp. 35–42, 1987.
21. R.T. Rockafellar, "Augmented lagrangians and applications of the proximal point algorithm in convex programming," Mathematics of Operations Research, vol. 1, pp. 97–116, 1976.
22. M.V. Solodov and P. Tseng, "Modified projection-type methods for monotone variational inequalities," SIAM J. Control Optim, vol. 34, pp. 1814–1830, 1996.
23. D. Sun, "A projection and contraction method for the no-linear complementary problem and its extensions," Mathematica Numerica Sinica, vol. 16, pp. 183–194, 1994.
24. D. Sun, "A new stepsize skill for solving a class of nonlinear projection equations," Journal of Computational Mathematics, vol. 13, pp. 357–368, 1995.
25. D. Sun, "A class of iterative methods for solving nonlinear projection equations," Journal of Optimization Theory and Application, vol. 91, pp. 123–140, 1996.
26. K. Taji, M. Fukushima, and T. Ibaraki, "A globally convergent newton method for solving strongly monotone variational inequalities," Mathematical Programming, vol. 58, pp. 369–383, 1993.
27. P. Tseng, "A modified forward-backword splitting method for maximal monotone mappings," SIAM. J. Control. Optim., vol. 38, pp. 431–446, 2000.
28. G.L. Xue and Y.Y. Ye, "An efficient algorithm for minimizing a sum of Euclidean norms with applications," SIAM J. Optim., vol. 7, pp. 1017–1036, 1997.