# On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators

**Xingju Cai · Guoyong Gu · Bingsheng He**

**Abstract** Nemirovski's analysis (SIAM J. Optim. 15:229–251, 2005) indicates that the extragradient method has the $O(1/t)$ convergence rate for variational inequalities with Lipschitz continuous monotone operators. For the same problems, in the last decades, a class of Fejér monotone projection and contraction methods is developed. Until now, only convergence results are available to these projection and contraction methods, though the numerical experiments indicate that they always outperform the extragradient method. The reason is that the former benefits from the 'optimal' step size in the contraction sense. In this paper, we prove the convergence rate under a unified conceptual framework, which includes the projection and contraction methods as special cases and thus perfects the theory of the existing projection and contraction methods. Preliminary numerical results demonstrate that the projection and contraction methods converge twice faster than the extragradient method.

**Keywords** Variational inequality · Projection and contraction method · Convergence rate

X. Cai · G. Gu · B. He (✉)
Department of Mathematics, Nanjing University, Nanjing 210093, P.R. China
e-mail: hebma@nju.edu.cn

X. Cai
e-mail: caixingju@njnu.edu.cn

G. Gu
e-mail: ggu@nju.edu.cn

# 1 Introduction

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

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

Notice that VI$(\Omega, F)$ is invariant when $F$ is multiplied by a positive scalar $\beta > 0$. It is well known ([1], p. 267) that, for any $\beta > 0$,

$$u^* \text{ is a solution of VI}(\Omega, F) \quad \Longleftrightarrow \quad u^* = P_\Omega[u^* - \beta F(u^*)], \tag{1.2}$$

where $P_\Omega(\cdot)$ denotes the projection onto $\Omega$ with respect to the Euclidean norm, i.e.,

$$P_\Omega(v) = \text{argmin}\{\|u - v\| \mid u \in \Omega\}.$$

Throughout this paper we assume that the mapping $F$ is monotone and Lipschitz continuous, i.e.,

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

and there is a constant $L > 0$ (not necessary to know), such that

$$\|F(u) - F(v)\| \leq L\|u - v\|, \quad \forall u, v \in R^n.$$

Moreover, we assume that the solution set of VI$(\Omega, F)$, denoted by $\Omega^*$, is nonempty. The nonempty assumption of the solution set, together with the monotonicity assumption of $F$, implies that $\Omega^*$ is closed and convex (see p. 158 in [3]).

Among the algorithms for monotone variational inequalities, the extragradient (EG) method proposed by Korpelevich [12] is one of the simple and attractive methods. In fact, each iteration of the extragradient method can be divided into two steps. The $k$-th iteration of EG method begins with a given $u^k \in \Omega$, the first step produces a vector $\tilde{u}^k$ via a projection

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)], \tag{1.3a}$$

where $\beta_k > 0$ is selected to satisfy (see [13])

$$\beta_k \|F(u^k) - F(\tilde{u}^k)\| \leq \nu \|u^k - \tilde{u}^k\|, \quad \nu \in (0, 1). \tag{1.3b}$$

Since $\tilde{u}^k$ is not accepted as the new iterate, for designation convenience, we call it as a *predictor* and $\beta_k$ is named the *prediction step size*. The second step (correction step) of the $k$-th iteration updates the new iterate $u^{k+1}$ by

$$u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)], \tag{1.3c}$$

where $\beta_k$ and $\tilde{u}^k$ are given in (1.3a). The sequence $\{u^k\}$ generated by the extragradient method is Fejér monotone with respect to the solution set, namely,

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - (1 - \nu^2)\|u^k - \tilde{u}^k\|^2. \tag{1.4}$$

For a proof of the above contraction property, the readers may consult [3] (see pp. 1115–1118 therein). Notice that, in the extragradient method, the step size of the prediction (1.3a) and that of the correction (1.3c) are equal. Thus these two steps seem like 'symmetric'.

Because of its simple iterative forms, recently, the extragradient method has been applied to solve some large optimization problems in the area of information science, such as in machine learning [11, 14, 22, 23], optical network [16, 17] and speech recognition [18], etc. In addition, Nemirovski [15] and Tseng [24] proved the $O(1/t)$ convergence rate of the extragradient method. Both in the theoretical and practical aspects, the interest in the extragradient method becomes more active.

In the last decades, a class of projection and contraction (PC) methods for monotone variational inequalities [5, 6, 8, 19, 20] are developed. Similarly as in the extragradient method, each iteration of the PC methods consists of two steps. The prediction step of PC methods produces the predictor $\tilde{u}^k$ via (1.3a) just as in the extragradient method (but the condition (1.3b) is not necessary). The PC methods exploit a pair of geminate directions [7, 8] offered by the predictor, namely, they are

$$d\left(u^k, \tilde{u}^k\right) = \left(u^k - \tilde{u}^k\right) - \beta_k\left(F\left(u^k\right) - F\left(\tilde{u}^k\right)\right) \quad \text{and} \quad \beta_k F\left(\tilde{u}^k\right). \qquad (1.5)$$

Here, both the directions are ascent directions of the unknown distance function $\frac{1}{2}\|u - u^*\|^2$ at the point $u^k$. Based on such directions, the goal of the correction step is to generate a new iterate which is closer to the solution set. It leads to choosing the 'optimal' step length

$$\varrho_k = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}, \qquad (1.6)$$

and a relaxation factor $\gamma \in (0, 2)$, the second step (*correction step*) of the PC methods updates the new iterate $u^{k+1}$ by

$$u^{k+1} = u^k - \gamma \varrho_k d\left(u^k, \tilde{u}^k\right), \qquad (1.7)$$

or

$$u^{k+1} = P_\Omega\left[u^k - \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right)\right]. \qquad (1.8)$$

The PC methods (without line search) make one (or two) projection(s) on $\Omega$ at each iteration, and the distance of the iterates to the solution set monotonically converges to zero. According to the terminology in [2], these methods belong to the class of Fejér contraction methods. In fact, the only difference between the extragradient method and one of the PC methods is that they use different step sizes in the correction step (see (1.3c) and (1.8)). According to our numerical experiments [6, 8], the PC methods always outperform the extragradient methods.

Stimulated by the complexity statement of the extragradient method, this paper shows the $O(1/t)$ convergence rate of the projection and contraction methods for monotone VIs. Recall that $\Omega^*$ can be characterized as (see (2.3.2) in p. 159 of [3])

$$\Omega^* = \bigcap_{u \in \Omega}\left\{\tilde{u} \in \Omega : (u - \tilde{u})^T F(u) \geq 0\right\}.$$

This implies that $\tilde{u} \in \Omega$ is an approximate solution of VI$(\Omega, F)$ with the accuracy $\epsilon$ if it satisfies

$$\tilde{u} \in \Omega \quad \text{and} \quad \inf_{u \in \Omega} \left\{ (u - \tilde{u})^T F(u) \right\} \geq -\epsilon.$$

In this paper, we show that, for given $\epsilon > 0$, in $O(L/\epsilon)$ iterations the projection and contraction methods can find a $\tilde{u}$ such that

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u \in \mathcal{D}(\tilde{u})} \left\{ (\tilde{u} - u)^T F(u) \right\} \leq \epsilon, \tag{1.9}$$

where

$$\mathcal{D}(\tilde{u}) = \left\{ u \in \Omega \mid \|u - \tilde{u}\| \leq 1 \right\}.$$

As a byproduct of the complexity analysis, we find why taking a suitable relaxation factor $\gamma \in (1, 2)$ in the correction steps (1.7) and (1.8) of the PC methods can achieve the faster convergence.

The outline of this paper is as follows. Section 2 recalls some basic concepts in the projection and contraction methods. In Sect. 3, we investigate the geminate descent directions of the distance function. Section 4 shows the contraction property of the PC methods. In Sect. 5, we carry out the complexity analysis, which results in an $O(1/t)$ convergence rate and suggests using the large relaxation factor in the correction step of the PC methods. In Sect. 6, we present some numerical results to indicate the efficiency of the PC methods in comparison with the extragradient method. Finally, some conclusion remarks are addressed in the last section.

Throughout the paper, the following notational conventions are used. We use $u^*$ to denote a fixed but arbitrary point in the solution set $\Omega^*$. A superscript such as in $u^k$ refers to a specific vector and usually denotes an iteration index. For any real matrix $M$ and vector $v$, we denote the transpose by $M^T$ and $v^T$, respectively. The Euclidean norm will be denoted by $\| \cdot \|$.

## 2 Preliminaries

In this section, we summarize the basic concepts of the projection mapping and three fundamental inequalities for constructing the PC methods. Throughout this paper, we assume that the projection on $\Omega$ in the Euclidean-norm has a closed form and it is easy to be carried out. Since

$$P_\Omega(v) = \operatorname{argmin} \left\{ \frac{1}{2} \|u - v\|^2 \,\Big|\, u \in \Omega \right\},$$

according to the optimal solution of the convex minimization problem, we have

$$\left( v - P_\Omega(v) \right)^T \left( u - P_\Omega(v) \right) \leq 0, \quad \forall v \in R^n, \forall u \in \Omega. \tag{2.1}$$

Consequently, for any $u \in \Omega$, it follows from (2.1) that

$$\|u - v\|^2 = \big\|\big(u - P_\Omega(v)\big) - \big(v - P_\Omega(v)\big)\big\|^2$$

$$= \big\|u - P_\Omega(v)\big\|^2 - 2\big(v - P_\Omega(v)\big)^T\big(u - P_\Omega(v)\big) + \big\|v - P_\Omega(v)\big\|^2$$

$$\geq \big\|u - P_\Omega(v)\big\|^2 + \big\|v - P_\Omega(v)\big\|^2.$$

Therefore, we have

$$\big\|u - P_\Omega(v)\big\|^2 \leq \|u - v\|^2 - \big\|v - P_\Omega(v)\big\|^2, \quad \forall\, v \in R^n, \ \forall\, u \in \Omega. \qquad (2.2)$$

For given $u$ and $\beta > 0$, let $\tilde{u} = P_\Omega[u - \beta F(u)]$ be given via a projection. We say that $\tilde{u}$ is a test-vector of $\mathrm{VI}(\Omega, F)$ because

$$u = \tilde{u} \quad \Leftrightarrow \quad u \in \Omega^*.$$

Since $\tilde{u} \in \Omega$, it follows from (1.1) that

$$(\text{FI-1}) \quad \big(\tilde{u} - u^*\big)^T \beta F\big(u^*\big) \geq 0, \quad \forall u^* \in \Omega^*. \qquad (2.3)$$

Setting $v = u - \beta F(u)$ and $u = u^*$ in the inequality (2.1), we obtain

$$(\text{FI-2}) \quad \big(\tilde{u} - u^*\big)^T \big((u - \tilde{u}) - \beta F(u)\big) \geq 0, \quad \forall u^* \in \Omega^*. \qquad (2.4)$$

Under the assumption that $F$ is monotone we have

$$(\text{FI-3}) \quad \big(\tilde{u} - u^*\big)^T \beta\big(F(\tilde{u}) - F\big(u^*\big)\big) \geq 0, \quad \forall u^* \in \Omega^*. \qquad (2.5)$$

The inequalities (2.3), (2.4) and (2.5) play an important role in the projection and contraction methods. They were emphasized in [5] as *three fundamental inequalities* in the projection and contraction methods.

**Definition 2.1** (Ascent direction) For any but fixed $u^* \in \Omega^*$, a direction $d$ is called an ascent direction of $\frac{1}{2}\|u - u^*\|^2$ at $u$ if and only if the inner-product $(u - u^*)^T d > 0$.

**Definition 2.2** (Predictor in projection-type methods) For given $u^k$, $\tilde{u}^k$ given by (1.3a) is called the predictor.

**Definition 2.3** (Prediction step-size condition in EG method) The condition (1.3b) is called the prediction step-size condition in the extragradient method.

Indeed, the predictor $\tilde{u}^k$ in the projection and contraction methods [5, 6, 8, 19] is produced by (1.3a). Because the mapping $F$ is Lipschitz continuous (even if the constant $L > 0$ is unknown), without loss of generality, we can assume that $\inf_{k \geq 0}\{\beta_k\} \geq \beta_L > 0$ and $\beta_L = O(1/L)$. In practical computation, we can choose a constant $\nu \in (0, 1)$ and make an initial guess of $\beta = \nu/L$ and decrease $\beta$ by a constant factor and repeat the procedure whenever (1.3b) is violated.

To the direction $d(u^k, \tilde{u}^k)$ defined in (1.5), there is a correlative ascent direction $\beta_k F(\tilde{u}^k)$. Notice that the projection equation (1.3a) can be written as

$$\tilde{u}^k = P_\Omega \{ \tilde{u}^k - [\beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k)] \}, \tag{2.6}$$

where $d(u^k, \tilde{u}^k)$ is defined in (1.5). It follows from (1.2) that $\tilde{u}^k$ is a solution of VI$(\Omega, F)$ if and only if $d(u^k, \tilde{u}^k) = 0$. In [9, 10], $d(u^k, \tilde{u}^k)$ and $\beta_k F(\tilde{u}^k)$ in (2.6) are called a pair of geminate directions and denoted by $d_1(u^k, \tilde{u}^k)$ and $d_2(u^k, \tilde{u}^k)$, respectively. In this paper, we restrict $d_2(u^k, \tilde{u}^k)$ to be $F(\tilde{u}^k)$ times a positive scalar $\beta_k$. The projection and contraction methods considered in this paper belong to the *prox-like contraction methods* [9, 10]. Instead of the step-size condition in the extragradient method (see (1.3b)), we use the following general step-size conditions for the projection and contraction methods.

**Definition 2.4** (Prediction step-size conditions in PC methods) Let $c_1, c_2 > 0$ be given constants. For given $u^k$, let $\tilde{u}^k$ be given by (1.3a) and $d(u^k, \tilde{u}^k)$ be defined by (1.5). We say $\beta_k$ satisfies the prediction step-size conditions in the projection and contraction methods, if its related direction $d(u^k, \tilde{u}^k)$ satisfies

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \geq c_1 \| u^k - \tilde{u}^k \|^2 \tag{2.7}$$

and

$$\varrho_k := \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\| d(u^k, \tilde{u}^k) \|^2} \geq c_2. \tag{2.8}$$

## 3 The ascent directions

For any but fixed $u^* \in \Omega^*$, $(u - u^*)$ is the gradient of the unknown distance function $\frac{1}{2} \| u - u^* \|^2$ in the Euclidean-norm[1] at the point $u$.

### 3.1 Geminate ascent directions

The forthcoming analysis is based on the general conditions. Note that an equivalent expression of (2.6) is

$$\tilde{u}^k \in \Omega, \ (u - \tilde{u}^k)^T \{ \beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k) \} \geq 0, \quad \forall u \in \Omega, \tag{3.1}$$

and from (2.8) we have

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) = \varrho_k \| d(u^k, \tilde{u}^k) \|^2. \tag{3.2}$$

---

[1]For convenience, we only consider the distance function in the Euclidean-norm. All the results in this paper are easy to extended to the contraction of the distance function in $G$-norm where $G$ is a positive definite matrix.

The following lemmas tell us that both the direction $d(u^k, \tilde{u}^k)$ (for $u^k \in R^n$) and $F(\tilde{u}^k)$ (for $u^k \in \Omega$) are ascent directions of the function $\frac{1}{2}\|u - u^*\|^2$ whenever $u^k$ is not a solution point. The proof is similar to those in [7], for completeness sake of this paper, we restate the short proofs.

**Lemma 3.1** *Let $u^k$, $\tilde{u}^k$ and $d(u^k, \tilde{u}^k)$ be given by (2.6) and the general conditions (2.7) and (2.8) be satisfied. Then we have*

$$\left(u^k - u^*\right)^T d\left(u^k, \tilde{u}^k\right) \geq \varrho_k \left\|d\left(u^k, \tilde{u}^k\right)\right\|^2, \quad \forall u^k \in R^n, \; u^* \in \Omega^*. \qquad (3.3)$$

*Proof* Note that $u^* \in \Omega$. By setting $u = u^*$ in (3.1) (the equivalent expression of (2.6)), we get

$$\left(\tilde{u}^k - u^*\right)^T d\left(u^k, \tilde{u}^k\right) \geq \left(\tilde{u}^k - u^*\right)^T \beta_k F\left(\tilde{u}^k\right) \geq 0, \quad \forall u^* \in \Omega^*.$$

The last inequality follows from the monotonicity of $F$ and $(\tilde{u}^k - u^*)^T F(u^*) \geq 0$. Therefore,

$$\left(u^k - u^*\right)^T d\left(u^k, \tilde{u}^k\right) \geq \left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right), \quad \forall u^* \in \Omega^*.$$

The assertion (3.3) is followed from the above inequality and (3.2) directly. $\qquad \square$

**Lemma 3.2** *Let $u^k$, $\tilde{u}^k$ and $d(u^k, \tilde{u}^k)$ be given by (2.6) and the general conditions (2.7) and (2.8) be satisfied. If $u^k \in \Omega$, then we have*

$$\left(u^k - u^*\right)^T \beta_k F\left(\tilde{u}^k\right) \geq \varrho_k \left\|d\left(u^k, \tilde{u}^k\right)\right\|^2, \quad \forall u^* \in \Omega^*. \qquad (3.4)$$

*Proof* Since $(\tilde{u}^k - u^*)^T \beta_k F(\tilde{u}^k) \geq 0$, we have

$$\left(u^k - u^*\right)^T \beta_k F\left(\tilde{u}^k\right) \geq \left(u^k - \tilde{u}^k\right)^T \beta_k F\left(\tilde{u}^k\right), \quad \forall u^* \in \Omega^*.$$

Note that because $u^k \in \Omega$, by setting $u = u^k$ in (3.1), we get

$$\left(u^k - \tilde{u}^k\right)^T \beta_k F\left(\tilde{u}^k\right) \geq \left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right).$$

From the above two inequalities follows that

$$\left(u^k - u^*\right)^T \beta_k F\left(\tilde{u}^k\right) \geq \left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right), \quad \forall u^* \in \Omega^*.$$

The assertion (3.4) is followed from the above inequality and (3.2) directly. $\qquad \square$

We would like to emphasize that (3.3) holds for $u^k \in R^n$ while (3.4) is hold only for $u^k \in \Omega$.

### 3.2 Ascent directions in the extragradient method

The condition (1.3b) is necessary in the prediction step of the extragradient method. Setting $u = u^k$, $\tilde{u} = \tilde{u}^k$ and $\beta = \beta_k$ in the fundamental inequalities (2.3), (2.4) and (2.5), and adding them, we get

$$\left(\tilde{u}^k - u^*\right)^T d\left(u^k, \tilde{u}^k\right) \geq 0, \quad \forall u^* \in \Omega^*, \tag{3.5}$$

where $d(u^k, \tilde{u}^k)$ defined in (1.5). It follows from (3.5) that

$$\left(u^k - u^*\right)^T d\left(u^k, \tilde{u}^k\right) \geq \left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right). \tag{3.6}$$

Note that, under the condition (1.3b), it follows that

$$\left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right) = \left\|u^k - \tilde{u}^k\right\|^2 - \left(u^k - \tilde{u}^k\right)^T \beta_k\left(F\left(u^k\right) - F\left(\tilde{u}^k\right)\right)$$
$$\geq (1 - \nu)\left\|u^k - \tilde{u}^k\right\|^2. \tag{3.7}$$

This means that $d(u^k, \tilde{u}^k)$ is an ascent direction of the unknown distance function $\frac{1}{2}\|u - u^*\|^2$ at the point $u^k$, and the condition (2.7) is satisfied with $c_1 = 1 - \nu$. By using the definition of $d(u^k, \tilde{u}^k)$, we have

$$2\left(u^k - \tilde{u}^k\right)d\left(u^k, \tilde{u}^k\right) - \left\|d\left(u^k, \tilde{u}^k\right)\right\|^2$$
$$= d\left(u^k, \tilde{u}^k\right)^T \left\{2\left(u^k - \tilde{u}^k\right) - d\left(u^k, \tilde{u}^k\right)\right\}$$
$$= \left\{\left(u^k - \tilde{u}^k\right) - \beta_k\left(F\left(u^k\right) - F\left(\tilde{u}^k\right)\right)\right\}^T \left\{\left(u^k - \tilde{u}^k\right) + \beta_k\left(F\left(u^k\right) - F\left(\tilde{u}^k\right)\right)\right\}$$
$$= \left\|u^k - \tilde{u}^k\right\|^2 - \beta_k^2\left\|F\left(u^k\right) - F\left(\tilde{u}^k\right)\right\|^2$$
$$\geq \left(1 - \nu^2\right)\left\|u^k - \tilde{u}^k\right\|^2.$$

From the above inequality follows that the condition (2.8) is satisfied with $c_2 \geq \frac{1}{2}$. In other words, the step-size conditions in Definition 2.4 are satisfied if the condition (1.3b) is holds. Thus, the directions $d(u^k, \tilde{u}^k)$ and $\beta_k F(\tilde{u}^k)$ defined in (1.5) are a pair of geminate ascent directions. In other words, the step-size condition (1.3b) is sufficient for the one in the projection and contraction methods.

## 4 Corrector and the convergence in the contraction sense

The extragradient method uses (1.3c) to update the new iterate. Based on the pair of geminate ascent directions in (2.6), namely, $d(u^k, \tilde{u}^k)$ and $\beta_k F(\tilde{u}^k)$, in the projection and contraction methods, we use the one of the following corrector forms to update the new iterate $u^{k+1}$:

$$\text{(Correction of PC Method-I)} \quad u_I^{k+1} = u^k - \gamma \varrho_k d\left(u^k, \tilde{u}^k\right), \tag{4.1}$$

or

$$\text{(Correction of PC Method-II)} \quad u_{II}^{k+1} = P_\Omega\big[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)\big], \qquad (4.2)$$

where $\gamma \in (0, 2)$ and $\varrho_k$ is defined in (2.8). Note that the same step size length is used in (4.1) and (4.2) even if the search directions are different. Recall that $\tilde{u}^k$ is obtained via a projection, by using the correction form (4.2), we have to make an additional projection on $\Omega$ in the PC methods. Replacing $\gamma \varrho_k$ in (4.2) by 1 (and if the step size condition (1.3b) is satisfied), it reduces to the update form of the extragradient method (see (1.3c)). For any solution point $u^* \in \Omega^*$, we define

$$\vartheta_I(\gamma) = \big\|u^k - u^*\big\|^2 - \big\|u_I^{k+1} - u^*\big\|^2 \qquad (4.3)$$

and

$$\vartheta_{II}(\gamma) = \big\|u^k - u^*\big\|^2 - \big\|u_{II}^{k+1} - u^*\big\|^2, \qquad (4.4)$$

which measure the profit in the $k$-th iteration. The following theorem gives a lower bound of the profit function, the similar results were established in [6–8].

**Theorem 4.1** *For given $u^k$, let the general conditions (2.7) and (2.8) be satisfied. If the corrector is updated by (4.1) or (4.2), then for any $u^* \in \Omega^*$ and $\gamma > 0$, we have*

$$\vartheta_I(\gamma) \geq q(\gamma), \qquad (4.5)$$

*and*

$$\vartheta_{II}(\gamma) \geq q(\gamma) + \big\|u_I^{k+1} - u_{II}^{k+1}\big\|^2, \qquad (4.6)$$

*respectively, where*

$$q(\gamma) = \gamma(2 - \gamma)\varrho_k^2\big\|d\big(u^k, \tilde{u}^k\big)\big\|^2. \qquad (4.7)$$

*Proof* Using the definition of $\vartheta_I(\gamma)$ and $u_I^{k+1}$ (see (4.1)), we have

$$\begin{aligned}
\vartheta_I(\gamma) &= \big\|u^k - u^*\big\|^2 - \big\|u^k - u^* - \gamma \varrho_k d\big(u^k, \tilde{u}^k\big)\big\|^2 \\
&= 2\gamma \varrho_k \big(u^k - u^*\big)^T d\big(u^k, \tilde{u}^k\big) - \gamma^2 \varrho_k^2\big\|d\big(u^k, \tilde{u}^k\big)\big\|^2.
\end{aligned} \qquad (4.8)$$

Recalling (3.3), we obtain

$$2\gamma \varrho_k \big(u^k - u^*\big)^T d\big(u^k, \tilde{u}^k\big) \geq 2\gamma \varrho_k^2\big\|d\big(u^k, \tilde{u}^k\big)\big\|^2.$$

Substituting it in (4.8) and using the definition of $q(\gamma)$, we get $\vartheta_I(\gamma) \geq q(\gamma)$ and the first assertion is proved. Now, we turn to show the second assertion. Because

$$u_{II}^{k+1} = P_\Omega\big[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)\big],$$

and $u^* \in \Omega$, by setting $u = u^*$ and $v = u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)$ in (2.2), we have

$$\big\|u^* - u_{II}^{k+1}\big\|^2 \leq \big\|u^* - \big(u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)\big)\big\|^2 - \big\|u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k) - u_{II}^{k+1}\big\|^2. \quad (4.9)$$

Thus,

$$\vartheta_{II}(\gamma) = \left\| u^k - u^* \right\|^2 - \left\| u_{II}^{k+1} - u^* \right\|^2$$

$$\geq \left\| u^k - u^* \right\|^2 - \left\| \left( u^k - u^* \right) - \gamma \varrho_k \beta_k F\left( \tilde{u}^k \right) \right\|^2$$

$$+ \left\| \left( u^k - u_{II}^{k+1} \right) - \gamma \varrho_k \beta_k F\left( \tilde{u}^k \right) \right\|^2$$

$$= \left\| u^k - u_{II}^{k+1} \right\|^2 + 2\gamma \varrho_k \beta_k \left( u_{II}^{k+1} - u^* \right)^T F\left( \tilde{u}^k \right)$$

$$\geq \left\| u^k - u_{II}^{k+1} \right\|^2 + 2\gamma \varrho_k \beta_k \left( u_{II}^{k+1} - \tilde{u}^k \right)^T F\left( \tilde{u}^k \right). \qquad (4.10)$$

The last inequality in (4.10) follows from $(\tilde{u}^k - u^*)^T F(\tilde{u}^k) \geq 0$. Since $u_{II}^{k+1} \in \Omega$, by setting $u = u_{II}^{k+1}$ in (3.1), we get

$$\left( u_{II}^{k+1} - \tilde{u}^k \right)^T \left\{ \beta_k F\left( \tilde{u}^k \right) - d\left( u^k, \tilde{u}^k \right) \right\} \geq 0,$$

and consequently, substituting it in the right hand side of (4.10), we obtain

$$\vartheta_{II}(\gamma) \geq \left\| u^k - u_{II}^{k+1} \right\|^2 + 2\gamma \varrho_k \left( u_{II}^{k+1} - \tilde{u}^k \right)^T d\left( u^k, \tilde{u}^k \right)$$

$$= \left\| u^k - u_{II}^{k+1} \right\|^2 + 2\gamma \varrho_k \left( u^k - \tilde{u}^k \right)^T d\left( u^k, \tilde{u}^k \right)$$

$$- 2\gamma \varrho_k \left( u^k - u_{II}^{k+1} \right)^T d\left( u^k, \tilde{u}^k \right). \qquad (4.11)$$

To the two crossed term in the right hand side of (4.11), we have (by using (3.2))

$$2\gamma \varrho_k \left( u^k - \tilde{u}^k \right)^T d\left( u^k, \tilde{u}^k \right) = 2\gamma \varrho_k^2 \left\| d\left( u^k, \tilde{u}^k \right) \right\|^2,$$

and

$$-2\gamma \varrho_k \left( u^k - u_{II}^{k+1} \right)^T d\left( u^k, \tilde{u}^k \right)$$

$$= \left\| \left( u^k - u_{II}^{k+1} \right) - \gamma \varrho_k d\left( u^k, \tilde{u}^k \right) \right\|^2 - \left\| u^k - u_{II}^{k+1} \right\|^2 - \gamma^2 \varrho_k^2 \left\| d\left( u^k, \tilde{u}^k \right) \right\|^2,$$

respectively. Substituting them in the right hand side of (4.11) and using $u^k - \gamma \varrho_k d(u^k, \tilde{u}^k) = u_I^{k+1}$, we obtain

$$\vartheta_{II}(\gamma) \geq \gamma(2 - \gamma)\varrho_k^2 \left\| d\left( u^k, \tilde{u}^k \right) \right\|^2 + \left\| u_I^{k+1} - u_{II}^{k+1} \right\|^2$$

$$= q(\gamma) + \left\| u_I^{k+1} - u_{II}^{k+1} \right\|^2, \qquad (4.12)$$

and the proof is complete.                                                                                    □

Note that $q(\gamma)$ is a quadratic function of $\gamma$, it reaches its maximum at $\gamma^* = 1$. In practice, $\varrho_k$ is the 'optimal' step size in (4.1) and (4.2), $\gamma$ is a relaxation factor. Because $q(\gamma)$ is a lower bound of $\vartheta_I(\gamma)$ (resp. $\vartheta_{II}(\gamma)$), the desirable new iterate is updated by (4.1) (resp. (4.2)) with $\gamma \in [1, 2)$ and the reason is interpreted in Fig. 1.

**Fig. 1** The interpreting of $\gamma \in [1, 2)$ for more profit in each iteration
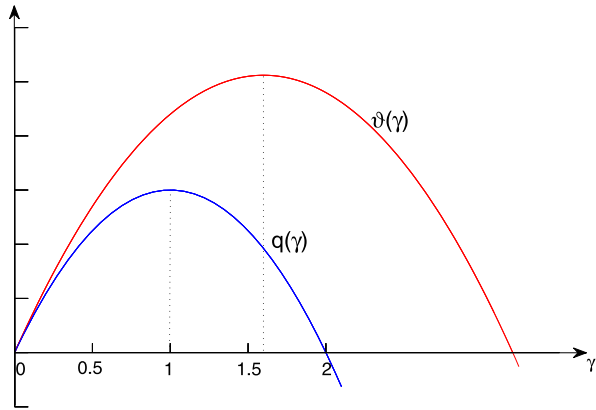
From Theorem 4.1 and the definition of $\varrho_k$ (see (2.8)), we obtain

$$\left\| u^{k+1} - u^* \right\|^2 \leq \left\| u^k - u^* \right\|^2 - \gamma(2 - \gamma)\varrho_k \left( u^k - \tilde{u}^k \right)^T d\left( u^k, \tilde{u}^k \right). \tag{4.13}$$

By using the general conditions (2.7) and (2.8), it follows from the above inequality that

$$\left\| u^{k+1} - u^* \right\|^2 \leq \left\| u^k - u^* \right\|^2 - \gamma(2 - \gamma)c_1 c_2 \left\| u^k - \tilde{u}^k \right\|^2. \tag{4.14}$$

Due to the property (4.14), we call the methods which use update forms (4.1) and (4.2) PC Method-I and PC Method II, respectively. Note that the assertion (4.14) is derived from the general conditions (2.7)–(2.8). From (4.14), the convergence result of the PC methods follows directly (just as the convergence of the extragradient method follows from (1.4), for details, see Theorem 2.1 in [6]).

## 5 Convergence rate of the PC methods

This section proves the convergence rate of the projection and contraction methods. Recall that the base of the complexity proof is (see (2.3.2) in p. 159 of [3])

$$\Omega^* = \bigcap_{u \in \Omega} \left\{ \tilde{u} \in \Omega : (u - \tilde{u})^T F(u) \geq 0 \right\}. \tag{5.1}$$

In the sequel, for given $\epsilon > 0$ and $\mathcal{D} \subset \Omega$, we focus our attention to find a $\tilde{u}$ such that

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u \in \mathcal{D}(\tilde{u})} (\tilde{u} - u)^T F(u) \leq \epsilon. \tag{5.2}$$

Although the PC Method I uses the update form (4.1) and it does not guarantee that $\{u^k\}$ belongs to $\Omega$, the sequence $\{\tilde{u}^k\} \subset \Omega$ in the PC methods with different corrector forms. Now, we prove the key inequality of the PC Method I for the complexity analysis.

**Lemma 5.1** *For given $u^k \in R^n$, let $\tilde{u}^k$ and $d(u^k, \tilde{u}^k)$ be given by (2.6) and the general conditions (2.7) and (2.8) be satisfied. If the new iterate $u^{k+1}$ is updated by (4.1) with any $\gamma > 0$, then we have*

$$\left(u - \tilde{u}^k\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right) + \frac{1}{2}\left(\|u - u^k\|^2 - \|u - u^{k+1}\|^2\right) \geq \frac{1}{2}q(\gamma), \quad \forall u \in \Omega, \quad (5.3)$$

*where $q(\gamma)$ is defined in (4.7).*

*Proof* Because (due to (3.1))

$$\left(u - \tilde{u}^k\right)^T \beta_k F\left(\tilde{u}^k\right) \geq \left(u - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right), \quad \forall u \in \Omega,$$

and (see (4.1))

$$\gamma \varrho_k d\left(u^k, \tilde{u}^k\right) = u^k - u^{k+1},$$

we need only to show that

$$\left(u - \tilde{u}^k\right)^T \left(u^k - u^{k+1}\right) + \frac{1}{2}\left(\|u - u^k\|^2 - \|u - u^{k+1}\|^2\right) \geq \frac{1}{2}q(\gamma), \quad \forall u \in \Omega. \quad (5.4)$$

To the crossed term in the left hand side of (5.4), namely $(u - \tilde{u}^k)^T (u^k - u^{k+1})$, using an identity

$$(a - b)^T (c - d) = \frac{1}{2}\left(\|a - d\|^2 - \|a - c\|^2\right) + \frac{1}{2}\left(\|c - b\|^2 - \|d - b\|^2\right),$$

we obtain

$$\left(u - \tilde{u}^k\right)^T \left(u^k - u^{k+1}\right) = \frac{1}{2}\left(\|u - u^{k+1}\|^2 - \|u - u^k\|^2\right)$$
$$+ \frac{1}{2}\left(\|u^k - \tilde{u}^k\|^2 - \|u^{k+1} - \tilde{u}^k\|^2\right). \quad (5.5)$$

By using $u^{k+1} = u^k - \gamma \varrho_k d(u^k, \tilde{u}^k)$ and (3.2), we get

$$\|u^k - \tilde{u}^k\|^2 - \|u^{k+1} - \tilde{u}^k\|^2 = \|u^k - \tilde{u}^k\|^2 - \|(u^k - \tilde{u}^k) - \gamma \varrho_k d\left(u^k, \tilde{u}^k\right)\|^2$$
$$= 2\gamma \varrho_k \left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right) - \gamma^2 \varrho_k^2 \|d\left(u^k, \tilde{u}^k\right)\|^2$$
$$= \gamma(2 - \gamma)\varrho_k^2 \|d\left(u^k, \tilde{u}^k\right)\|^2.$$

Substituting it in the right hand side of (5.5) and using the definition of $q(\gamma)$, we obtain (5.4) and the lemma is proved. □

The both sequences $\{\tilde{u}^k\}$ and $\{u^k\}$ in the PC method II belong to $\Omega$. In the following lemma we prove the same assertion for PC method II as in Lemma 5.1.

**Lemma 5.2** *For given $u^k \in \Omega$, let $\tilde{u}^k$ and $d(u^k, \tilde{u}^k)$ be given by (2.6) and the general conditions (2.7) and (2.8) be satisfied. If the new iterate $u^{k+1}$ is updated by (4.2) with any $\gamma > 0$, then we have*

$$\left(u - \tilde{u}^k\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right) + \frac{1}{2}\left(\left\|u - u^k\right\|^2 - \left\|u - u^{k+1}\right\|^2\right) \geq \frac{1}{2}q(\gamma), \quad \forall u \in \Omega,$$

(5.6)

*where $q(\gamma)$ is defined in (4.7).*

*Proof* For investigating $(u - \tilde{u}^k)^T \beta_k F(\tilde{u}^k)$, we divide it in the terms

$$\left(u^{k+1} - \tilde{u}^k\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right) \quad \text{and} \quad \left(u - u^{k+1}\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right).$$

(5.7)

First, we deal with the term $(u^{k+1} - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k)$. Since $u^{k+1} \in \Omega$, substituting $u = u^{k+1}$ in (3.1) we get

$$\left(u^{k+1} - \tilde{u}^k\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right) \geq \gamma \varrho_k \left(u^{k+1} - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right)$$

$$= \gamma \varrho_k \left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right)$$

$$- \gamma \varrho_k \left(u^k - u^{k+1}\right)^T d\left(u^k, \tilde{u}^k\right).$$

(5.8)

To the first crossed term of the right hand side of (5.8), using (3.2), we have

$$\gamma \varrho_k \left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right) = \gamma \varrho_k^2 \left\|d\left(u^k, \tilde{u}^k\right)\right\|^2.$$

To the second crossed term of the right hand side of (5.8), using the Cauchy-Schwarz Inequality, we get

$$-\gamma \varrho_k \left(u^k - u^{k+1}\right)^T d\left(u^k, \tilde{u}^k\right) \geq -\frac{1}{2}\left\|u^k - u^{k+1}\right\|^2 - \frac{1}{2}\gamma^2 \varrho_k^2 \left\|d\left(u^k, \tilde{u}^k\right)\right\|^2.$$

Substituting them in the right hand side of (5.8), we obtain

$$\left(u^{k+1} - \tilde{u}^k\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right) \geq \frac{1}{2}\gamma(2 - \gamma)\varrho_k^2 \left\|d\left(u^k, \tilde{u}^k\right)\right\|^2 - \frac{1}{2}\left\|u^k - u^{k+1}\right\|^2.$$

(5.9)

Now, we turn to treat of the term $(u - u^{k+1})^T \gamma \varrho_k \beta_k F(\tilde{u}^k)$ in (5.7). Since $u^{k+1}$ is updated by (4.2), $u^{k+1}$ is the projection of $(u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k))$ on $\Omega$, it follows from (2.1) that

$$\left\{\left(u^k - \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right)\right) - u^{k+1}\right\}^T \left(u - u^{k+1}\right) \leq 0, \quad \forall u \in \Omega,$$

and consequently

$$\left(u - u^{k+1}\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right) \geq \left(u - u^{k+1}\right)^T \left(u^k - u^{k+1}\right), \quad \forall u \in \Omega.$$

Using the identity $a^T b = \frac{1}{2}\{\|a\|^2 - \|a - b\|^2 + \|b\|^2\}$ to the right hand side of the last inequality, we obtain

$$\left(u - u^{k+1}\right)^T \gamma \varrho_k \beta_k F\left(\tilde{u}^k\right)$$
$$\geq \frac{1}{2}\left(\left\|u - u^{k+1}\right\|^2 - \left\|u - u^k\right\|^2\right) + \frac{1}{2}\left\|u^k - u^{k+1}\right\|^2. \tag{5.10}$$

Adding (5.9) and (5.10) and using the definition of $q(\gamma)$, we get (5.6) and the proof is complete. □

For the different projection and contraction methods, we have the same key inequality which is shown in Lemmas 5.1 and 5.2, respectively. By setting $u = u^*$ in (5.3) and (5.6), we get

$$\left\|u^k - u^*\right\|^2 - \left\|u^{k+1} - u^*\right\|^2 \geq 2\gamma \varrho_k \beta_k \left(\tilde{u}^k - u^*\right)^T F\left(\tilde{u}^k\right) + q(\gamma).$$

Because $(\tilde{u}^k - u^*)^T F(\tilde{u}^k) \geq (\tilde{u}^k - u^*)^T F(u^*) \geq 0$ and $q(\gamma) = \gamma(2 - \gamma)\varrho_k^2 \times \|d(u^k, \tilde{u}^k)\|^2$, it follows from the last inequality that

$$\left\|u^{k+1} - u^*\right\|^2 \leq \left\|u^k - u^*\right\|^2 - \gamma(2 - \gamma)\varrho_k^2 \left\|d\left(u^k, \tilde{u}^k\right)\right\|^2.$$

This is just the form (4.13) in Sect. 4. In other words, the contraction property (4.13) of PC methods is the consequent result of Lemmas 5.1 and 5.2, respectively.

For the convergence rate proof, we allow $\gamma \in (0, 2]$. In this case, we still have $q(\gamma) \geq 0$. By using the monotonicity of $F$, from (5.3) and (5.6) we get

$$\left(u - \tilde{u}^k\right)^T \varrho_k \beta_k F(u) + \frac{1}{2\gamma}\left\|u - u^k\right\|^2 \geq \frac{1}{2\gamma}\left\|u - u^{k+1}\right\|^2, \quad \forall u \in \Omega. \tag{5.11}$$

This inequality is essential for the convergence rate proofs.

**Theorem 5.1** *For any integer $t > 0$, we have a $\tilde{u}_t \in \Omega$ which satisfies*

$$(\tilde{u}_t - u)^T F(u) \leq \frac{1}{2\gamma \Upsilon_t}\left\|u - u^0\right\|^2, \quad \forall u \in \Omega, \tag{5.12}$$

*where*

$$\tilde{u}_t = \frac{1}{\Upsilon_t}\sum_{k=0}^{t}\varrho_k \beta_k \tilde{u}^k \quad and \quad \Upsilon_t = \sum_{k=0}^{t}\varrho_k \beta_k. \tag{5.13}$$

*Proof* Summing the inequality (5.11) over $k = 0, \ldots, t$, we obtain

$$\left(\left(\sum_{k=0}^{t} \varrho_k \beta_k\right)u - \sum_{k=0}^{t} \varrho_k \beta_k \tilde{u}^k\right)^T F(u) + \frac{1}{2\gamma}\left\|u - u^0\right\|^2 \geq 0, \quad \forall u \in \Omega.$$

Using the notations of $\Upsilon_t$ and $\tilde{u}_t$ in the above inequality, we derive

$$(\tilde{u}_t - u)^T F(u) \leq \frac{\|u - u^0\|^2}{2\gamma \Upsilon_t}, \quad \forall u \in \Omega.$$

Indeed, $\tilde{u}_t \in \Omega$ because it is a convex combination of $\tilde{u}^0, \tilde{u}^1, \ldots, \tilde{u}^t$. The proof is complete. □

If the step-size sequence $\{\beta_k\}$ in (1.3a) is low-bounded away from zero, i.e.

$$\inf_{k \geq 0}\{\beta_k\} \geq \beta_L > 0, \tag{5.14}$$

then it follows from (2.8) and (5.13) that

$$\Upsilon_t \geq (t + 1)c_2 \beta_L,$$

and thus the PC methods have $O(1/t)$ convergence rate. For any substantial set $\mathcal{D} \subset \Omega$, the PC methods reach

$$(\tilde{u}_t - u)^T F(u) \leq \epsilon, \quad \forall u \in \mathcal{D}, \text{ in at most } t = \left\lceil \frac{D^2}{2\gamma c_2 \beta_L \epsilon} \right\rceil$$

iterations, where $\tilde{u}_t$ is defined in (5.13) and $D = \sup\{\|u - u^0\| \mid u \in \mathcal{D}\}$. This convergence rate is in the ergodic sense, the statement (5.12) suggests us to take a larger parameter $\gamma \in (0, 2]$ in the correction steps of the PC methods. To use the $O(1/t)$ convergence rate, we need only to check that the sequence $\{\beta_k\}$ is low-bound away from zero and the general conditions in Definition 2.4 are satisfied.

## 6 Numerical experiments

This section is devoted to test the efficiency of the PC methods in comparison with the extragradient method [12]. In particular, we use (1.3a), (1.3b), (1.3c) as the recursion form of the extragradient method. All codes are written in Matlab and run on a Lenovo X200 Computer with 2.53 GHz.

### 6.1 Test examples of minimizing a sum of distances

The first part of the test examples is a min-max problem which can be formulated as a monotone linear variational inequality with skew-symmetric matrix. The problem is to find the shortest network in a given full Steiner topology (see Example 1 in [25]). The points-edges connections of the network are depicted in Fig. 2, where $P = \{b_{[1]}, \ldots, b_{[10]}\}$ are given points in $R^2$ (called regular points) and $x_{[1]}, \ldots, x_{[8]}$
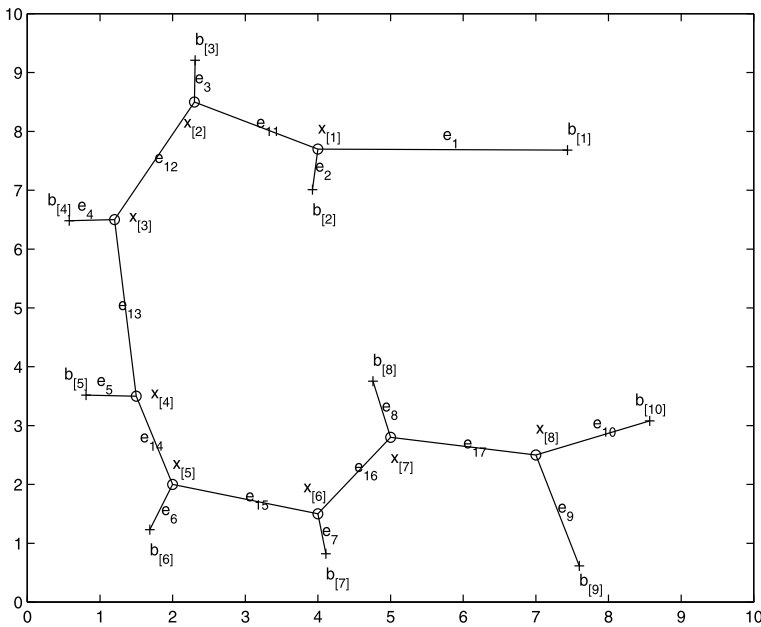
**Fig. 2** Original graph of the problem

**Table 1** The coordinates of the 10 regular points

|          | $x$-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       |

are the locations of the additional points (called Steiner points). The coordinates of the 10 regular points are given in Table 1. The mathematical form of the shortest sum of the distance under $p$-norm is

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

$$\left. + \sum_{j=1}^{7} \|x_{[j]} - x_{[j+1]}\|_p \right\}. \tag{6.1}$$

Mainly, we are interested in $p = 1, 2$ and $\infty$. For any $d \in R^2$ and $p = 1, 2$ and $\infty$, we have

$$\|d\|_p = \max_{\xi \in B_q} \xi^T d, \tag{6.2}$$

where

$$B_q = \left\{ \xi \in R^2 \mid \|\xi\|_q \leq 1 \right\} \quad \text{and} \quad q = \begin{cases} \infty & \text{if } p = 1, \\ 2 & \text{if } p = 2, \\ 1 & \text{if } p = \infty. \end{cases}$$

*Formulating the problem to a linear variational inequality*   The problem (6.1) is equivalent to the following min-max problem

$$\min_{x_{[j]} \in R^2} \max_{z_{[i]} \in \mathcal{B}} \left\{ z_{[1]}^T (x_{[1]} - b_{[1]}) + \sum_{j=1}^{8} z_{[j+1]}^T (x_{[j]} - b_{[j+1]}) \right.$$

$$\left. + z_{[10]}^T (x_{[8]} - b_{[10]}) + \sum_{j=1}^{7} z_{[j+10]}^T (x_{[j]} - x_{[j+1]}) \right\}, \quad (6.3)$$

where $z_{[i]}, i = 1, \ldots, 17$ are vectors in $R^2$. In fact, $z_{[i]}$ is the dual variable ordered to edge $e_i$. The compact form of (6.3) is

$$\min_{x \in \mathcal{R}} \max_{z \in \mathcal{B}} z^T (Ax - b) \qquad (6.4)$$

where

$$x^T = \left( x_{[1]}^T, \ldots, x_{[8]}^T \right)^T, \qquad z^T = \left( z_{[1]}^T, \ldots, z_{[17]}^T \right)^T,$$

$$\mathcal{R} = R^2 \times \cdots \times R^2, \qquad \mathcal{B} = B_q \times \cdots \times B_q, \qquad (6.5)$$

$A$ is block matrix which has form

$$A = \begin{pmatrix} I_2 & & & & \\ I_2 & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & I_2 & \\ & & & I_2 & \\ I_2 & -I_2 & & & \\ & \ddots & \ddots & & \\ & & I_2 & -I_2 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_{[1]} \\ b_{[2]} \\ \vdots \\ \vdots \\ b_{[9]} \\ b_{[10]} \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \qquad (6.6)$$

Let $(x^*, z^*) \in \mathcal{R} \times \mathcal{B}$ be any solution of (6.4), it follows that

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

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

$$x^* \in \mathcal{R}, z^* \in \mathcal{B}_2, \quad \begin{cases} (x - x^*)^T (A^T z^*) \geq 0, & \forall x \in \mathcal{R}, \\ (z - z^*)^T (-Ax^* + b) \geq 0, & \forall z \in \mathcal{B}. \end{cases} \qquad (6.7)$$

The compact form of (6.7) is the following linear variational inequality:

$$\text{LVI}(\Omega, M, q) \quad u^* \in \Omega, \quad (u - u^*)^T (Mu^* + q) \geq 0, \quad \forall u \in \Omega, \tag{6.8}$$

where

$$u = \begin{pmatrix} x \\ z \end{pmatrix}, \qquad M = \begin{pmatrix} 0 & A^T \\ -A & 0 \end{pmatrix}, \qquad q = \begin{pmatrix} 0 \\ b \end{pmatrix} \quad \text{and} \quad \Omega = \mathcal{R} \times \mathcal{B}. \tag{6.9}$$

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

*Solving the LVI by the extragradient method* By using the extragradient method (see (1.3a)–(1.3c)), for given $u^k$, to produce $\tilde{u}^k$ by (1.3a), the condition (1.3b) should be satisfied. Note that the mapping $F(u) = Mu + q$ in LVI$(\Omega, M, q)$ is Lipschitz continuous. For this example, one can calculate $\|M\| \approx 2.2089$. In order to ensure the condition (1.3b) to be satisfied, we should to take $\beta \in (0, 1/\|M\|)$. Since $1/\|M\| \approx 0.4527$, we use the extragradient method to solve the problem (6.8)–(6.9) with different $\beta = 0.30, 0.35, 0.40$ and $0.45$, respectively. The start point is $u^0 = 0$ and the iteration is stopped as soon as

$$\|u^k - \tilde{u}^k\| \leq 10^{-10}.$$

The optimal networks are depicted in Figs. 3, 4. The iteration numbers for the shortest distance problems under different norms are reported in Table 2. It seems
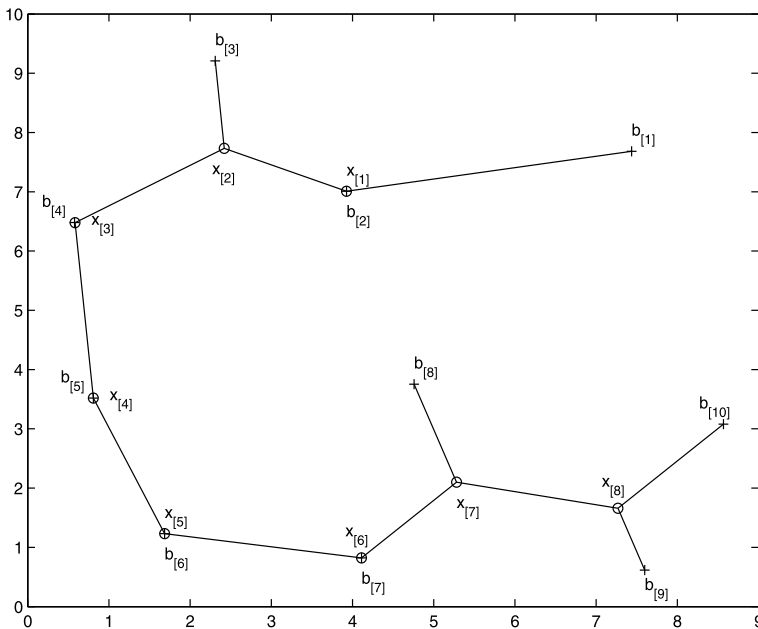


**Fig. 3** Optimal solution, $l_2$-norm
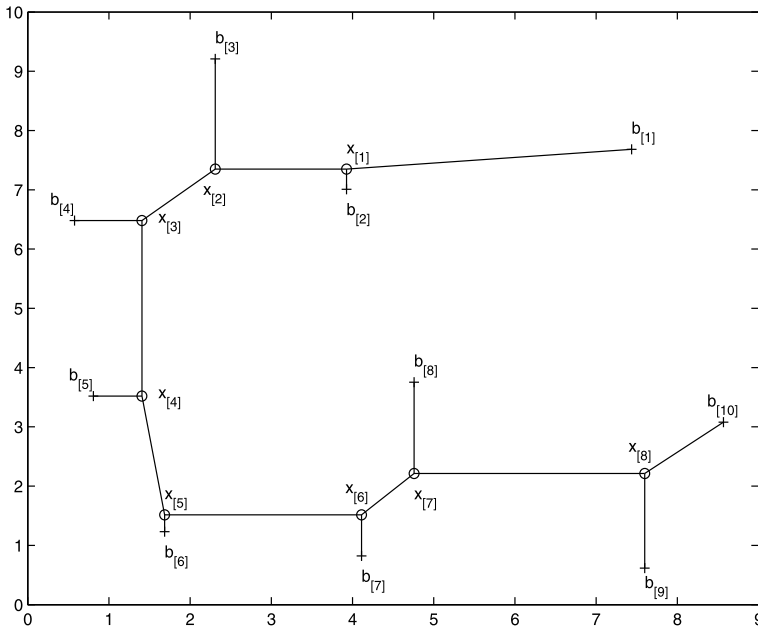
**Fig. 4** Optimal solution, $l_1$-norm

**Table 2** Iteration numbers of EG method by using different $\beta$'s

| Different norms | $\beta = 0.30$ | $\beta = 0.35$ | $\beta = 0.40$ | $\beta = 0.45$ |
|---|---|---|---|---|
| $l_1$ | 550 | 418 | 333 | 275 |
| $l_2$ | 500 | 380 | 303 | 250 |
| $l_\infty$ | 535 | 407 | 325 | 269 |

that, for fast convergence, the constant parameter $\beta \in (0, 1/\|M\|)$ should be closed to $1/\|M\|$.

*Solving the LVI by the PC methods* By using the PC methods, the condition (1.3b) is not necessary. We need only to ensure the general conditions in Definition 2.4 to be satisfied. Thus, we can take $\beta_k \equiv 1$ in (1.3a) and the predictor $\tilde{u}^k$ is given by
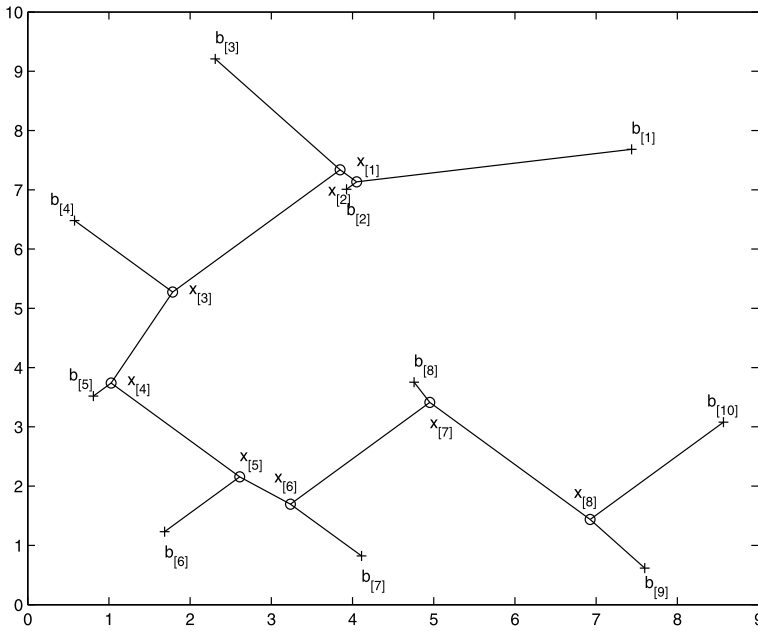
$$\tilde{u}^k = P_\Omega\left[u^k - \left(Mu^k + q\right)\right].$$

Because $M^T = -M$, the above projection mapping can be rewritten as

$$\tilde{u}^k = P_\Omega\left\{\tilde{u}^k - \left[\left(Mu^k + q\right) - \left(u^k - \tilde{u}^k\right)\right]\right\}$$
$$= P_\Omega\left\{\tilde{u}^k - \left[\left(M\tilde{u}^k + q\right) - \left(I + M^T\right)\left(u^k - \tilde{u}^k\right)\right]\right\}.$$

This is the form (2.6) with

$$\beta_k \equiv 1, \qquad F\left(\tilde{u}^k\right) = \left(M\tilde{u}^k + q\right) \quad \text{and} \quad d\left(u^k, \tilde{u}^k\right) = \left(I + M^T\right)\left(u^k - \tilde{u}^k\right).$$

**Fig. 5** Optimal solution, $l_\infty$-norm

**Table 3** Numerical results of PC methods and comparison with the extragradient method

| Different norms | PC method I | | PC method II | | Extragradient method | | Optimal objective value of the network (6.1) |
|---|---|---|---|---|---|---|---|
| | No. It | CPU sec | No. It | CPU sec | No. It | CPU sec | |
| $l_1$ | 156 | 0.021 | 81 | 0.011 | 275 | 0.048 | 28.665858000 |
| $l_2$ | 188 | 0.033 | 106 | 0.022 | 250 | 0.050 | 25.356067793 |
| $l_\infty$ | 144 | 0.021 | 84 | 0.013 | 269 | 0.058 | 21.112913500 |

For this $d(u^k, \tilde{u}^k)$, we have

$$\left(u^k - \tilde{u}^k\right)^T d\left(u^k, \tilde{u}^k\right) = \left\|u^k - \tilde{u}^k\right\|^2, \quad \text{(due to } M^T = -M)$$

and

$$\varrho_k = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2} = \frac{\|u^k - \tilde{u}^k\|^2}{\|(I + M^T)(u^k - \tilde{u}^k)\|^2} \geq \frac{1}{\|I + M^T\|^2}.$$

Thus the general conditions in Definition 2.4 are satisfied with $c_1 = 1$ and $c_2 = 1/\|I + M^T\|^2$. We use the PC Method-I (see (1.7)) or PC Method II (see (1.8)) to update the new iterate. The iteration numbers and the CPU times are listed in Table 3. For comparison, we also report the CPU times by using the extragradient method with $\beta = 0.45$ (the fastest case in Table 2) and the shortest distance of the network in different norms.

It is observed that

$$\frac{\text{Computational load of PC Method II}}{\text{Computational load of the extragradient method}} < 40\ \%.$$

Both the PC Methods converge much faster than the Extragradient method.

## 6.2 Test examples of nonlinear complementarity problems

We take the nonlinear complementarity problems

$$(\text{NCP}) \quad u \geq 0, \qquad F(u) \geq 0, \qquad u^T F(u) = 0,$$

as the second part of the test examples. Complementarity problem is a variational inequality $VI(\Omega, F)$ with $\Omega = R^n_+$, the non-negative orthant in $R^n$. In this case, $P_\Omega(v) = \max(v, 0)$.

*Constructing the test problems*   The mapping $F(u)$ in the tested NCP is given by

$$F(u) = D(u) + Mu + q, \tag{6.10}$$

where $D(u) : R^n \to R^n$ is the nonlinear part, $M$ is an $n \times n$ matrix, and $q \in R^n$ is a vector.

- In $D(u)$, the nonlinear part of $F(u)$, the components are

$$D_j(u) = d_j \cdot \arctan(a_j \cdot u_j),$$

  where $a$ and $d$ are random vectors[2] whose elements are in $(0, 1)$.
- The matrix $M$ in the linear part is given by $M = A^T A + B$. $A$ is an $n \times n$ matrix whose entries are randomly generated in the interval $(-5, +5)$, and $B$ is an $n \times n$ skew-symmetric random matrix $(B^T = -B)$ whose entries[3] are in the interval $(-5, +5)$.

It is clear that the mapping composed in this way is monotone. We construct the following 3 sets of test examples by choosing different vector $q$ in (6.10).

1. In the first set of the test examples, the elements of vector $q$ is generated from a uniform distribution in the interval $(-500, 500)$.
2. The second set[4] of the test examples is similar to the first set. Instead of $q \in (-500, 500)$, the vector $q$ is generated from a uniform distribution in the interval $(-500, 0)$.

---

[2]A similar type of (small) problems was tested in [21] where the components of the nonlinear mapping $D(u)$ are $D_j(u) = c \cdot \arctan(u_j)$.

[3]In the paper by Harker and Pang [4], the matrix $M = A^T A + B + D$, where $A$ and $B$ are the same matrices as what we use here, and $D$ is a diagonal matrix with uniformly distributed random entries $d_{jj} \in (0.0, 0.3)$.

[4]In [4], the similar problems in the first set are called easy problems while the 2-nd set problems are called hard problems.

3. The third set of test examples has a known solution $u^* \in R_+^n$. Let vector $p$ be generated from a uniform distribution in the interval $(-10, 10)$ and

$$u^* = \max(p, 0). \tag{6.11}$$

By setting

$$v = \max(-p, 0) \quad \text{and} \quad q = v - \left(D(u^*) + Mu^*\right),$$

we have $F(u^*) = D(u^*) + Mu^* + q = v = \max(-p, 0)$. Thus,

$$\left(u^*\right)^T F\left(u^*\right) = \left(\max(p, 0)\right)^T \left(\max(-p, 0)\right) = 0.$$

In this way we constructed a test NCP with a known solution $u^*$ described in (6.11).

*Implementation details* For given $u^k$, we use (1.3a) to produce $\tilde{u}^k$ which satisfies condition in (1.3b) with $\nu = 0.95$. Note that in this case the general conditions (2.7) and (2.8) are satisfied with $c_1 = (1 - \nu)$ and $c_2 = 1/2$.

We use a modified Armijo rule for choosing the parameter $\beta_k$. It should be mentioned, in practical computation, if $r_k := \beta_k \|F(u^k) - F(\tilde{u}^k)\| / \|u^k - \tilde{u}^k\|$ is too small, it will lead slow convergence. Therefore, if $r_k \leq \mu = 0.4$, the trial parameter $\beta_k$ will be enlarged for the next iteration. In the case that $F$ is Lipschitz continuous with constant $L > 0$, by using the above mentioned strategies for choosing $\beta_k$, we still have $\inf\{\beta_k\} \geq \beta_L = O(1/L)$. These 'refined' strategies are necessary for fast convergence. Algorithm 1 is the implementation details.

**Algorithm 1**
Step 0.    Set $\beta_0 = 1$, $u^0 \in \Omega$ and $k = 0$.
Step 1.    $\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)]$,
            $$r_k := \frac{\beta_k \|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$
            **while**    $r_k > \nu$
                $\beta_k := 0.7 * \beta_k * \min\{1, \frac{1}{r_k}\}$, $\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)]$
            **end(while)**
            Use different forms ((1.3c), (4.1) or (4.2)) to update $u^{k+1}$.
            **If** $r_k \leq \mu$ then $\beta_k := \beta_k * \nu * 0.9/r_k$, **end(if)**
Step 2.    $\beta_{k+1} = \beta_k$ and $k = k + 1$, go to Step 1.

The iterations begin with $u^0 = 0$, $\beta_0 = 1$ and stop as soon as

$$\frac{\|u^k - P_\Omega[u^k - F(u^k)]\|_\infty}{\|u^0 - P_\Omega[u^0 - F(u^0)]\|_\infty} \leq 10^{-6}. \tag{6.12}$$

Since both $F(u^k)$ and $F(\tilde{u}^k)$ are involved in those methods recursions, each iteration of the test methods needs at least 2 times of evaluations of the mapping $F$. We use No. It and No. $F$ to denote the numbers of iterations and the evaluations of the mapping $F$, respectively. The size of the tested problems is from 500 to 2000.

**Table 4** Numerical results of the first set examples

| Problem size $n$ | Extra-gradient method $u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)]$ | | | PC method I ($\gamma = 1.9$) $u^{k+1} = u^k - \gamma \varrho_k d(u^k, \tilde{u}^k)$ | | | PC method II ($\gamma = 1.9$) $u^{k+1} = P_\Omega[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)]$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | No. It | No. $F$ | CPU sec | No. It | No. $F$ | CPU sec | No. It | No. $F$ | CPU sec |
| 500 | 705 | 1448 | 0.27 | 449 | 943 | 0.15 | 372 | 791 | 0.13 |
| 1000 | 728 | 1495 | 2.48 | 477 | 999 | 1.66 | 412 | 857 | 1.43 |
| 2000 | 795 | 1633 | 10.68 | 512 | 1074 | 6.96 | 417 | 889 | 5.79 |
| 3000 | 763 | 1565 | 21.47 | 491 | 1029 | 14.13 | 431 | 882 | 12.17 |
| 4000 | 797 | 1635 | 40.05 | 509 | 1068 | 26.16 | 433 | 903 | 22.08 |
| 5000 | 825 | 1694 | 64.71 | 537 | 1125 | 42.98 | 436 | 988 | 35.49 |

**Table 5** Numerical results of the second set examples

| Problem size $n$ | Extra-gradient method $u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)]$ | | | PC method I ($\gamma = 1.9$) $u^{k+1} = u^k - \gamma \varrho_k d(u^k, \tilde{u}^k)$ | | | PC method II ($\gamma = 1.9$) $u^{k+1} = P_\Omega[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)]$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | No. It | No. $F$ | CPU sec | No. It | No. $F$ | CPU sec | No. It | No. $F$ | CPU sec |
| 500 | 1511 | 3102 | 0.57 | 910 | 1916 | 0.31 | 809 | 791 | 0.28 |
| 1000 | 1604 | 3296 | 5.55 | 968 | 2053 | 3.45 | 852 | 857 | 3.06 |
| 2000 | 1858 | 3817 | 24.13 | 981 | 2091 | 13.30 | 981 | 889 | 13.30 |
| 3000 | 1782 | 3655 | 50.87 | 1067 | 2253 | 31.36 | 977 | 882 | 28.55 |
| 4000 | 1771 | 3636 | 88.64 | 1060 | 2237 | 54.58 | 943 | 903 | 49.02 |
| 5000 | 1717 | 3526 | 134.76 | 1027 | 2167 | 83.49 | 907 | 988 | 74.67 |

**Table 6** Numerical results of the third set examples

| Problem size $n$ | Extra-gradient method $u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)]$ | | | PC method I ($\gamma = 1.9$) $u^{k+1} = u^k - \gamma \varrho_k d(u^k, \tilde{u}^k)$ | | | PC method II ($\gamma = 1.9$) $u^{k+1} = P_\Omega[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)]$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | No. It | No. $F$ | CPU sec | No. It | No. $F$ | CPU sec | No. It | No. $F$ | CPU sec |
| 500 | 1169 | 2398 | 0.42 | 715 | 1505 | 0.25 | 621 | 1319 | 0.22 |
| 1000 | 1551 | 3181 | 5.34 | 941 | 1981 | 3.33 | 827 | 1760 | 2.97 |
| 2000 | 1426 | 2931 | 18.59 | 853 | 1800 | 11.48 | 751 | 1597 | 10.17 |
| 3000 | 1685 | 3451 | 48.03 | 1005 | 2115 | 29.59 | 893 | 1908 | 26.27 |
| 4000 | 1604 | 3290 | 80.92 | 962 | 2022 | 49.68 | 849 | 1805 | 44.38 |
| 5000 | 1724 | 3656 | 135.12 | 1037 | 2187 | 83.45 | 918 | 1942 | 74.15 |

*Comparison between the extragradient method and the PC method II* In the case that the condition (1.3b) is satisfied, replacing $\gamma \varrho_k$ in (4.2) by 1, the PC method II becomes the extragradient method. According to the assertion in Theorems 4.1 and 5.1, we take the relaxation factor $\gamma = 1.9$ in the PC methods. The test results for the 3 sets of NCP are given in Tables 4, 5 and 6, respectively.

In the third test examples, as the stop criterion is satisfied, we have $\|u^k - u^*\|_\infty \approx 2 \times 10^{-4}$ by using the all three test methods. The PC method II and the extragradi-

ent method use the same direction but different step size in the correction step. The numerical results show that the PC method II is much efficient than the extragradient method. Even if the PC methods need to calculate the step size $\varrho_k$ in each iteration, while the computational load required by the additional effort is significantly less than the dominating task (the evaluations of $F(u^k)$ and $F(\tilde{u}^k)$). It is observed that

$$\frac{\text{Computational load of PC Method II}}{\text{Computational load of the extragradient method}} \le 55 \%.$$

The different PC methods use the one of the geminate directions but the same step size in their correction forms. Between the PC methods, PC method II needs fewer iterations than PC method I, this evidence coincides with the assertions in Theorem 4.1 (see (4.5) and (4.6)). Thus, we suggest to use PC method II when the projection on $\Omega$ is easy to be carried out. Otherwise (when the projection is the dominating task in the iteration), we use PC method I because its update form (4.1) does not contain the projection.

## 7 Conclusions

In a unified framework, we proved the $O(1/t)$ convergence rate of the projection and contraction methods for monotone variational inequalities. The convergence rate is the same as that for the extragradient method. In fact, our convergence rate include the extragradient method as a special case. The complexity analysis in this paper is based on the general conditions defined in Definition 2.4 and thus can be extended to a broaden class of similar contraction methods. Preliminary numerical results indicate that the PC methods do outperform the extragradient method.

## References

1. Bertsekas, D.P., Tsitsiklis, J.N.: Parallel and Distributed Computation, Numerical Methods. Prentice-Hall, Englewood Cliffs (1989)
2. Blum, E., Oettli, W.: Mathematische Optimierung: Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung. Springer, Berlin (1975)
3. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Vols. I and II. Springer Series in Operations Research. Springer, New York (2003)
4. Harker, P.T., Pang, J.S.: A damped-Newton method for the linear complementarity problem. Lect. Appl. Math. **26**, 265–284 (1990)
5. He, B.S.: A class of projection and contraction methods for monotone variational inequalities. Appl. Math. Optim. **35**, 69–76 (1997)
6. He, B.S., Liao, L.-Z.: Improvements of some projection methods for monotone nonlinear variational inequalities. J. Optim. Theory Appl. **112**, 111–128 (2002)
7. He, B.S., Xu, M.-H.: A general framework of contraction methods for monotone variational inequalities. Pac. J. Optim. **4**, 195–212 (2008)
8. He, B.S., Yuan, X.M., Zhang, J.J.Z.: Comparison of two kinds of prediction-correction methods for monotone variational inequalities. Comput. Optim. Appl. **27**, 247–267 (2004)

9. He, B.S., Liao, L.-Z., Wang, X.: Proximal-like contraction methods for monotone variational inequalities in a unified framework I: effective quadruplet and primary methods. Comput. Optim. Appl. **51**, 649–679 (2012)
10. He, B.S., Liao, L.-Z., Wang, X.: Proximal-like contraction methods for monotone variational inequalities in a unified framework II: general methods and numerical experiments. Comput. Optim. Appl. **51**, 681–708 (2012)
11. Howard, A.G.: Large margin, transformation learning. PhD Thesis, Graduate School of Arts and Science, Columbia University (2009)
12. Korpelevich, G.M.: The extragradient method for finding saddle points and other problems. Ekon. Mat. Metod. **12**, 747–756 (1976)
13. Khobotov, E.N.: Modification of the extragradient method for solving variational inequalities and certain optimization problems. USSR Comput. Math. Math. Phys. **27**, 120–127 (1987)
14. Lacoste-Julien, S.: Discriminative machine learning with structure. PhD Thesis, Computer Science, University of California, Berkeley (2009)
15. Nemirovski, A.: Prox-method with rate of convergence $O(1/t)$ for variational inequality with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. Optim. **15**, 229–251 (2005)
16. Pan, Y.: A game theoretical approach to constrained OSNR optimization problems in optical network. PhD Thesis, Electrical and Computer Engineering, University of Toronto (2009)
17. Pan, Y., Pavel, L.: Games with coupled propagated constraints in optical networks with multi-link topologies. Automatica **45**, 871–880 (2009)
18. Sha, F.: Large margin training of acoustic models for speech recognition. PhD Thesis, Computer and Information Science, University of Pennsylvania (2007)
19. Solodov, M.V., Tseng, P.: Modified projection-type methods for monotone variational inequalities. SIAM J. Control Optim. **34**, 1814–1830 (1996)
20. Sun, D.: A class of iterative methods for solving nonlinear projection equations. J. Optim. Theory Appl. **91**, 123–140 (1996)
21. Taji, K., Fukushima, M., Ibaraki, I.: A globally convergent Newton method for solving strongly monotone variational inequalities. Math. Program. **58**, 369–383 (1993)
22. Taskar, B., Lacoste-Julien, S., Jordan, M.I.: Structured prediction, dual extragradient and Bregman projections. J. Mach. Learn. Res. **7**, 1627–1653 (2006)
23. Taskar, B., Lacoste-Julien, S., Jordan, M.I.: Structured prediction via extragradient method. In: Weiss, Y., Schoelkopf, B., Platt, J. (eds.) Advances in Neural Information Processing Systems (NIPS), vol. 18 (2006)
24. Tseng, P.: On accelerated proximal gradient methods for convex-concave optimization. Department of Mathematics, University of Washington, Seattle, WA 98195, USA (2008)
25. Xue, G.L., Ye, Y.Y.: An efficient algorithm for minimizing a sum of Euclidean norms with applications. SIAM J. Optim. **7**, 1017–1036 (1997)