凸优化和单调变分不等式的收缩算法

第二讲：三个基本不等式和
变分不等式的投影收缩算法

Three fundamental inequalities and the projection
and contraction methods for variational inequalities

南京大学数学系　　何炳生
hebma@nju.edu.cn

The context of this lecture is mainly based on the publicatios [4, 6]

1 Basic properties of Projection Mapping and Variational Inequality

Let $\Omega \subset \Re^n$ be a closed convex set, F be a mapping from $\Re^n$ to itself . We investigate the solution methods for monotone variational inequality

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

We say the variational inequalit VI(Ω, F) is monotone, when its operator F is monotone. In other words, it satisfies

$$(u - v)^T (F(u) - F(v)) \geq 0, \quad \forall\, u,\, v \in \Re^n(\text{or } \Omega).$$

If $F(u) = Mu + q$, where M is n by n matrix and $q \in \Re^n$, F is an affine operator and the problem is a linear variational inequality (abbreviated to LVI). A LVI is monotone when

$$u^T M u \geq 0, \quad \forall u \in \Re^n.$$

In the above case, although M is not symmetric, we say M is positive semidefinite because the symmetric matrix $M^T + M$ is positive semidefinite.

Especially, when M is skew-symmetric, *i. e.*, $M^T = -M$, then $u^T M u \equiv 0$. In this case, the affine operator $F(u) = Mu + q$ is monotone.

Let $\Omega \subset \Re^n$ be a convex closed set and f be a convex function on Ω. Assume that f is differentiable on a open set that contains Ω. Then f is convex if and only if

$$f(y) - f(x) \geq \nabla f(x)^T (y - x), \quad \forall x, y \in \Omega. \tag{1.2}$$

This assertion can be found in

- R. Fletcher, Practical Methods of Optimization, Second Edition, §9.4. pp. 214–215, John Wiley & Sons, 1987.

Exchange the positions of x and y in (1.2), we get

$$f(x) - f(y) \geq \nabla f(y)^T (x - y), \quad \forall x, y \in \Omega. \tag{1.3}$$

Adding (1.2) and (1.3), it follows that

$$(y - x)^T (\nabla f(y) - \nabla f(x)) \geq 0, \quad \forall x, y \in \Omega. \tag{1.4}$$

Thus, the gradient operator of the differentiable convex function is monotone.

1.1 Basic properties of the projection mapping

We use $P_\Omega(\cdot)$ to denote the projection on Ω in Euclidean-norm, *i. e.* ,

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

An equivalent expression is $P_\Omega(v) = \arg\min\{\frac{1}{2}\|u-v\|^2 \mid u \in \Omega\}$. When $\Omega = \Re^n_+$ (the nonnegative orthant in $\Re^n$), each element of $P_\Omega(v)$ is given by

$$(P_\Omega(v))_j = \begin{cases} v_j, & \text{if} \quad v_j \ge 0; \\ 0, & \text{otherwise.} \end{cases}$$

When $\Omega = B(c,r) = \{c + ru \mid \|u\| \le 1\}$, a ball in $\Re^n$ with radius r centered on c, then

$$P_\Omega(v) = \begin{cases} \frac{r(v-c)}{\|v-c\|} + c, & \text{if} \quad \|v-c\| \ge r; \\ v, & \text{otherwise.} \end{cases}$$

The unit ball in l_∞ and l_1 norm centered on the origin are denoted by

$$B_\infty = \{u \in \Re^n | \ \|u\|_\infty \le 1\} \quad \text{and} \quad B_1 = \{u \in \Re^n | \ \|u\|_1 \le 1\},$$

respectively. The projection on B_∞ and B_1 are depicted in the following figures:

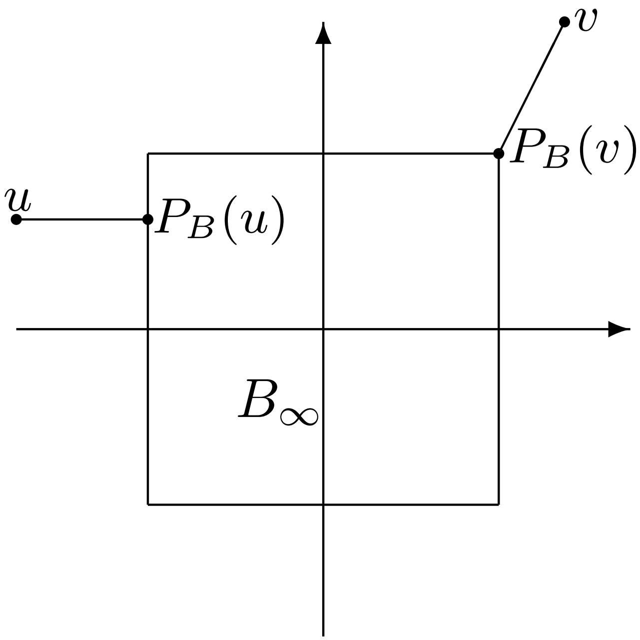

Fig.1 Projection on B_∞

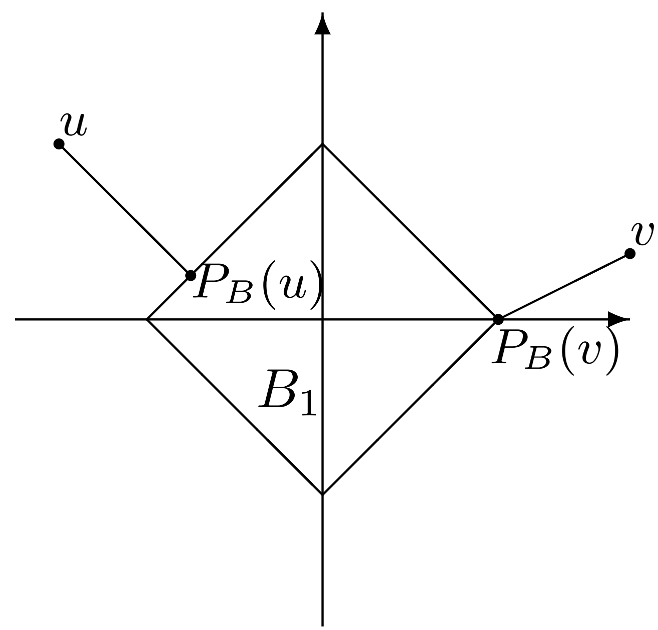

Fig.2 Projection on B_1

The following lemma illustrates an important property of the projection mapping.

Lemma 1.1 *Let $\Omega \subset \Re^n$ be a closed convex set and $P_\Omega(\cdot)$ be the projection on Ω. It holds that*

$$(v - P_\Omega(v))^T(u - P_\Omega(v)) \le 0, \quad \forall v \in \Re^n, \forall u \in \Omega. \tag{1.5}$$

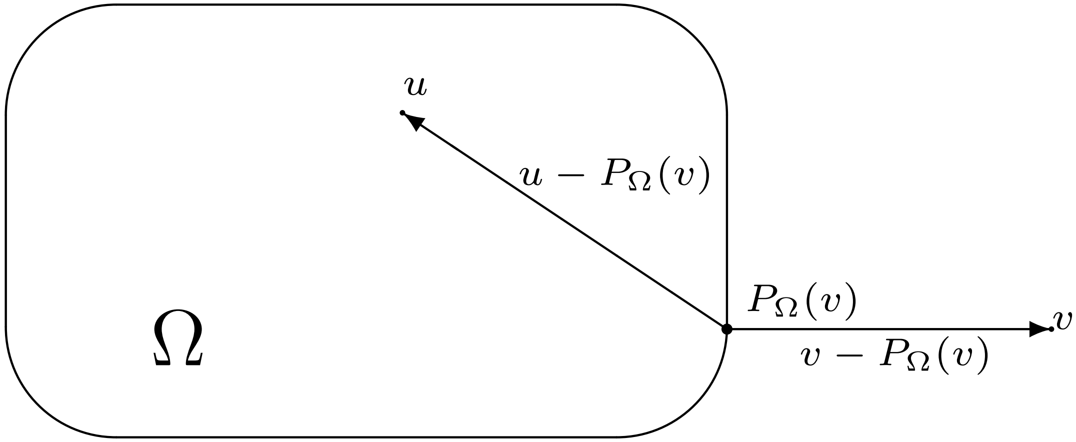

Fig.3 Geometric interpretation of the inequality (1.5)

Proof. First, since $P_\Omega(v) = \arg\min\{\|u-v\| \mid u \in \Omega\}$, we have

$$\|v - P_\Omega(v)\| \le \|v - w\|, \quad \forall w \in \Omega. \tag{1.6}$$

Because $P_\Omega(v) \in \Omega$ and $\Omega \subset \Re^n$ is closed and convex, then for any $u \in \Omega$ and $\theta \in (0,1)$, it holds that

$$w := \theta u + (1-\theta)P_\Omega(v) = P_\Omega(v) + \theta(u - P_\Omega(v)) \in \Omega.$$

For this w, by using (1.6), it follows that

$$\|v - P_\Omega(v)\|^2 \le \|v - w\|^2 = \|v - P_\Omega(v) - \theta(u - P_\Omega(v))\|^2.$$

Expanding the last inequality, for any $u \in \Omega$ and $\theta \in (0,1)$, we have

$$[v - P_\Omega(v)]^T [u - P_\Omega(v)] \le \frac{\theta}{2} \|u - P_\Omega(v)\|^2.$$

Let $\theta \to 0_+$, we get the assertion (1.5). $\square$

In the analysis of the projection and contraction methods, the inequality (1.5) is most important and useful. We call it as the **Tool inequality of the projection operator**. By using (1.5), it is easy to prove the following lemma.

Lemma 1.2 *Let $\Omega \subset \Re^n$ be a closed convex set, we have*

$$\|P_\Omega(v) - P_\Omega(u)\| \le \|v - u\|, \qquad \forall u, v \in \Re^n. \tag{1.7}$$

$$\|P_\Omega(v) - u\| \le \|v - u\|, \qquad \forall v \in \Re^n, u \in \Omega. \tag{1.8}$$

$$\|P_\Omega(v) - u\|^2 \le \|v - u\|^2 - \|v - P_\Omega(v)\|^2, \ \forall v \in \Re^n, u \in \Omega. \tag{1.9}$$

We leave the proofs to the reader.

1.2 The equivalent projection equation of the variational inequality

Assume that the solution set of (1.1), denoted by Ω^*, is nonempty. The solution set of monotone variational inequality is convex, its proof can be found in Theorem 2.3.5 [1]. We use u^* to denote any fixed point in Ω^*. For any scalar $\beta > 0$, the following statement is true.

$$u \in \Omega^* \quad \Leftrightarrow \quad u = P_\Omega[u - \beta F(u)].$$

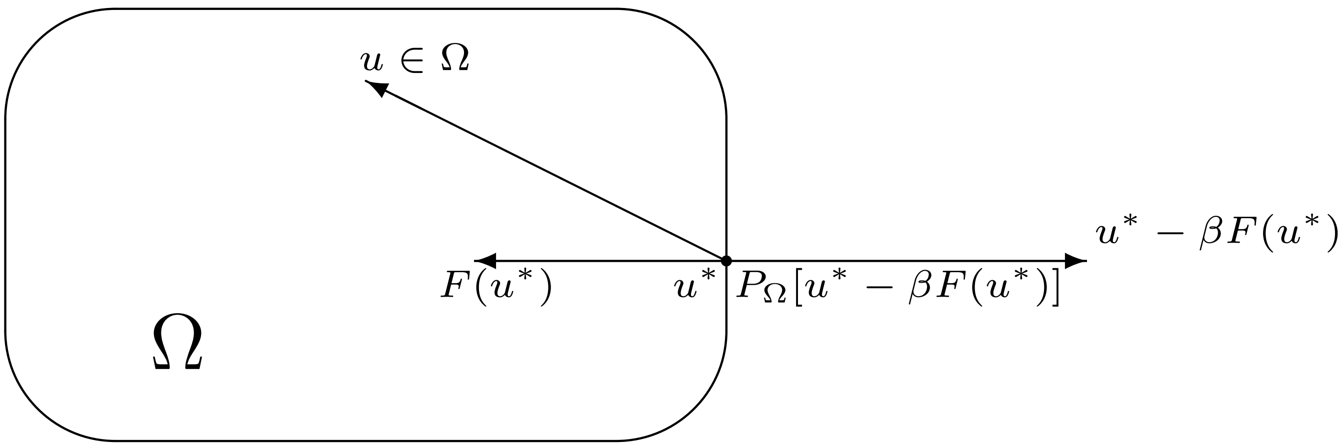

Fig.4 Geometric interpretation of $u^* \in \Omega^* \Leftrightarrow u^* = P_\Omega[u^* - \beta F(u^*)]$

In other words, solving the variational inequality (1.1) is equivalent to finding a zero point of $e(u, \beta)$, where

$$e(u, \beta) = u - P_{\Omega}[u - \beta F(u)] \tag{1.10}$$

The proof will be given in Theorem 1.1. Thus, for a given $\beta > 0$, $\|e(u, \beta)\|$ can be viewed as the error which measure how much u fails to be a solution point.

Theorem 1.1 *For given $\beta > 0$, u^* is a solution point od VI(Ω, F) if and only if $e(u^*, \beta) = 0$.*

Proof. " $\Rightarrow$ " Let u^* be a solution point of VI(Ω, F), we show $e(u^*, \beta) = 0$. Since $u^* \in \Omega$. By using the tool inequality (1.5), we get

$$\big(v - P_{\Omega}(v)\big)^T \big(u^* - P_{\Omega}(v)\big) \le 0, \quad \forall v \in \Re^n.$$

Setting $v = u^* - \beta F(u^*)$ in the above inequality and using the notation of $e(u, \beta)$, it follows from the last inequality that $(e(u^*, \beta) - \beta F(u^*))^T e(u^*, \beta) \le 0$, and thus

$$\|e(u^*, \beta)\|^2 \le \beta e(u^*, \beta)^T F(u^*). \tag{1.11}$$

On the other hand, because $P_{\Omega}[u^* - \beta F(u^*)] \in \Omega$ and u^* is a solution of the VI, it follows from (1.1) that

$$\{P_\Omega[u^* - \beta F(u^*)] - u^*\}^T F(u^*) \geq 0,$$

and thus

$$e(u^*, \beta)^T F(u^*) \leq 0. \tag{1.12}$$

From (1.11) and (1.12), we get $e(u^*, \beta) = 0$.

" $\Leftarrow$ ". If $e(u^*, \beta) = 0$, we show $u^* \in \Omega^*$. Taking $v = u^* - \beta F(u^*)$ in (1.5) and using the notation $e(u^*, \beta)$, we have

$$\{e(u^*, \beta) - \beta F(u^*)\}^T \{u - P_\Omega[u^* - \beta F(u^*)]\} \leq 0, \quad \forall u \in \Omega. \tag{1.13}$$

Since $e(u^*, \beta) = 0$, $u^* = P_\Omega(\cdot) \in \Omega$ and $P_\Omega[u^* - \beta F(u^*)] = u^*$. Substituting them in (1.13), we get

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

Thus u^* is a solution of VI(Ω, F). The proof is complete. $\square$

The following theorem tells us, for any fixed u, $\|e(u, \beta)\|$ is a non-decreasing function of β, and $\{\|e(u, \beta)\|/\beta\}$ is a non-increasing function of β. Our proof is taken from [14], where the tool inequality (1.5) plays a key role.

Theorem 1.2 *For any given* $u \in \Re^n$ *and* $\tilde{\beta} \geq \beta > 0$*, we have*

$$\|e(u,\tilde{\beta})\| \geq \|e(u,\beta)\| \tag{1.14}$$

and

$$\frac{\|e(u,\tilde{\beta})\|}{\tilde{\beta}} \leq \frac{\|e(u,\beta)\|}{\beta}. \tag{1.15}$$

Proof. Let $t = \|e(x,\tilde{\beta})\|/\|e(x,\beta)\|$, the assertions of this theorem is equivalent to

$$1 \leq t \leq \tilde{\beta}/\beta.$$

In other words, t is the solution of the quadratic inequality

$$(t-1)(t-\tilde{\beta}/\beta) \leq 0. \tag{1.16}$$

First, since $P_\Omega(w) \in \Omega$, according to the tool inequality (1.5), we have

$$(v - P_\Omega(v))^T (P_\Omega(v) - P_\Omega(w)) \geq 0, \quad \forall\, v \in \Re^n. \tag{1.17}$$

Setting $v := u - \beta F(u)$ and $w := P_\Omega[u - \tilde{\beta}F(u)]$ in (1.17), using the definition of $e(u,\beta)$, it follows that $(v - P_\Omega(v)) = e(u,\beta) - \beta F(u)$ and

$$P_\Omega(v) - P_\Omega(w) = P_\Omega[u - \beta F(u)] - P_\Omega[u - \tilde{\beta}F(u)] = e(u,\tilde{\beta}) - e(u,\beta).$$

Substituting it in (1.17), we get

$$\{e(u,\beta)-\beta F(u)\}^T\{e(u,\tilde\beta)-e(u,\beta)\}\geq 0. \tag{1.18}$$

Change the position of β and $\tilde\beta$ in (1.18), it follows that

$$\{e(u,\tilde\beta)-\tilde\beta F(u)\}^T\{e(u,\beta)-e(u,\tilde\beta)\}\geq 0. \tag{1.19}$$

Multiplying (1.18) and (1.19) by $\tilde\beta$ and β, respectively, and then adding them up, we have

$$\{\tilde\beta e(u,\beta)-\beta e(u,\tilde\beta)\}^T\{e(u,\tilde\beta)-e(u,\beta)\}\geq 0,$$

and thus

$$\beta\|e(x,\tilde\beta)\|^2-(\beta+\tilde\beta)e(x,\beta)^Te(x,\tilde\beta)+\tilde\beta\|e(x,\beta)\|^2\leq 0. \tag{1.20}$$

Using Cauchy-Schwarz inequality to (1.20),

$$\beta\|e(x,\tilde\beta)\|^2-(\beta+\tilde\beta)\|e(x,\beta)\|\cdot\|e(x,\tilde\beta)\|+\tilde\beta\|e(x,\beta)\|^2\leq 0. \tag{1.21}$$

Dividing (1.21) by $\beta\|e(x,\beta)\|^2$, and using $t=\|e(x,\tilde\beta)\|/\|e(x,\beta)\|$,

$$t^2-\Big(1+\frac{\tilde\beta}{\beta}\Big)t+\frac{\tilde\beta}{\beta}\leq 0.$$

Thus, the inequality (1.16) is true, and the proof of this theorem is complete. □

Although Theorem 1.1 indicates that, for any $\beta > 0$, u is a solution of (1.1) if and only if $e(u, \beta) = 0$. Theorem 1.2 tell us, if we use $\|e(u, \beta)\|$ as the error measure, the constant parameter $\beta > 0$ should not be too large or too small. Generally, it should be considered in combination with the physical significance of the considered problem.

2 Three fundamental inequalities and Projection and Contraction Methods

Three fundamental inequalities Let u^* be any given solution of the monotone variational inequality (1.1). Because $\tilde{u} = P_\Omega[u - \beta F(u)] \in \Omega$, according to the definition of VI, we have

$$\text{(FI1)} \quad (\tilde{u} - u^*)^T \beta F(u^*) \geq 0.$$

Setting $v = u - \beta F(u)$ in (1.5), since $\tilde{u} = P_\Omega[u - \beta F(u)] = P_\Omega[v]$ and $u^* \in \Omega$, according to the tool inequality (1.5), we have

$$\text{(FI2)} \quad (\tilde{u} - u^*)^T \big([u - \beta F(u)] - \tilde{u} \big) \geq 0.$$

The variational inequalities considered in this Lecture Series is monotone, according to the monotonicity of the operator F, we have

$$\text{(FI3)} \quad (\tilde{u} - u^*)^T (\beta F(\tilde{u}) - \beta F(u^*)) \geq 0.$$

The search directions of some projection contraction algorithms are derived from these basic (but fundamental) inequalities.

The basic framework of the projection and contraction methods

Projection and Contraction Methods is a kind of contraction algorithms based on projection. For given $\beta > 0$ and he current point u^k, we get a predictor $\tilde{u}^k$ by making the projection $\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)]$. According to Theorem 1.1,

u^k is a solution of (1.1) if and only if $u^k = \tilde{u}^k$.

Error measure function A nonnegative function $\varphi(u^k, \tilde{u}^k)$ is called the error measure function of VI(Ω, F) (1.1), if there is a $\delta > 0$, such that

$$\varphi(u^k, \tilde{u}^k) \geq \delta\|u^k - \tilde{u}^k\|^2 \quad \text{and} \quad \varphi(u^k, \tilde{u}^k) = 0 \Leftrightarrow u^k = \tilde{u}^k. \tag{2.1a}$$

Profitable direction A vector $d(u^k, \tilde{u}^k)$ $\big(\|d(u^k, \tilde{u}^k)\| = \mathcal{O}\|u^k - \tilde{u}^k\|\big)$ is called a profitable direction associated with $\varphi(u^k, \tilde{u}^k)$, if

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \varphi(u^k, \tilde{u}^k), \ \forall\, u^* \in \Omega^*. \tag{2.1b}$$

- Although u^* is unknown, for any fixed u^*, $(u^k - u^*)$ is the gradient of the distance function $\frac{1}{2}\|u - u^*\|^2$ at u^k. Thus, we call $\varphi(u, \tilde{u})$ and $d(u, \tilde{u})$ "the error measure function" and "the profitable direction", respectively.
- The basic idea of the projection and contraction algorithm is to construct a direction $d(u^k, \tilde{u}^k)$, such that (2.1b) is satisfied for any $u^* \in \Omega^*$. The definitions (2.1) indicate that $-d(u^k, \tilde{u}^k)$ is a descent direction of the distance function $\frac{1}{2}\|u - u^*\|^2$ at the current point u^k.
- In the contraction methods, it is required that the sequence $\{\|u^k - u^*\|^2\}$ is strictly monotone decreasing, where $\{u^k\}$ is the sequence generated by the algorithm and u^* is any fixed solution point.

Using the conditions (2.1a) and (2.1b), we give the following general algorithm.

The general algorithm When the condition (2.1b) is satisfied, let

$$u^{k+1}(\alpha) = u^k - \alpha d(u^k, \tilde{u}^k) \tag{2.2}$$

be the step-size α dependent new iterate. Now, we consider how to maximize the "profit" of the square of the distance, namely

$$\vartheta_k(\alpha) = \|u^k - u^*\|^2 - \|u^{k+1}(\alpha) - u^*\|^2. \tag{2.3}$$

Notice that

$$\begin{aligned} \vartheta_k(\alpha) &= \|u^k - u^*\|^2 - \|u^k - u^* - \alpha d(u^k, \tilde{u}^k)\|^2 \\ &= 2\alpha(u^k - u^*)^T d(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \end{aligned}$$

For any given u^*, the last equation indicates that $\vartheta_k(\alpha)$ is a quadratic function of α. Because u^* is unknown, we can not directly maximize $\vartheta_k(\alpha)$. However, using (2.1b), we have

$$\vartheta_k(\alpha) \geq 2\alpha\varphi(u^k, \tilde{u}^k) - \alpha^2 \|d(u^k, \tilde{u}^k)\|^2. \tag{2.4}$$

We define the right hand side of the last inequality as $q_k(\alpha)$, so, we get

$$q_k(\alpha) = 2\alpha\varphi_k(u^k, \tilde{u}^k) - \alpha^2\|d(u^k, \tilde{u}^k)\|^2, \tag{2.5}$$

which is a lower bound function of $\vartheta_k(\alpha)$. The quadratic function $q_k(\alpha)$ reaches its maximum at α_k^*,

$$\alpha_k^* = \frac{\varphi(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \tag{2.6}$$

If we take

$$u^{k+1} = u^k - \alpha_k^* d(u^k, \tilde{u}^k), \tag{2.7}$$

it follows from (2.5) and (2.6) that

$$q_k(\alpha_k^*) = \alpha_k^*\varphi(u^k, \tilde{u}^k).$$

The new iterate generates by (2.7) is not necessarily in Ω, however, it satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \alpha_k^*\varphi(u^k, \tilde{u}^k).$$

The original intention of the contraction algorithms is to maximize the quadratic function $\vartheta_k(\alpha)$ (see (2.3)) in each iteration, because it contains the unknown u^*, we have to maximize its lower bound function $q_k(\alpha)$.

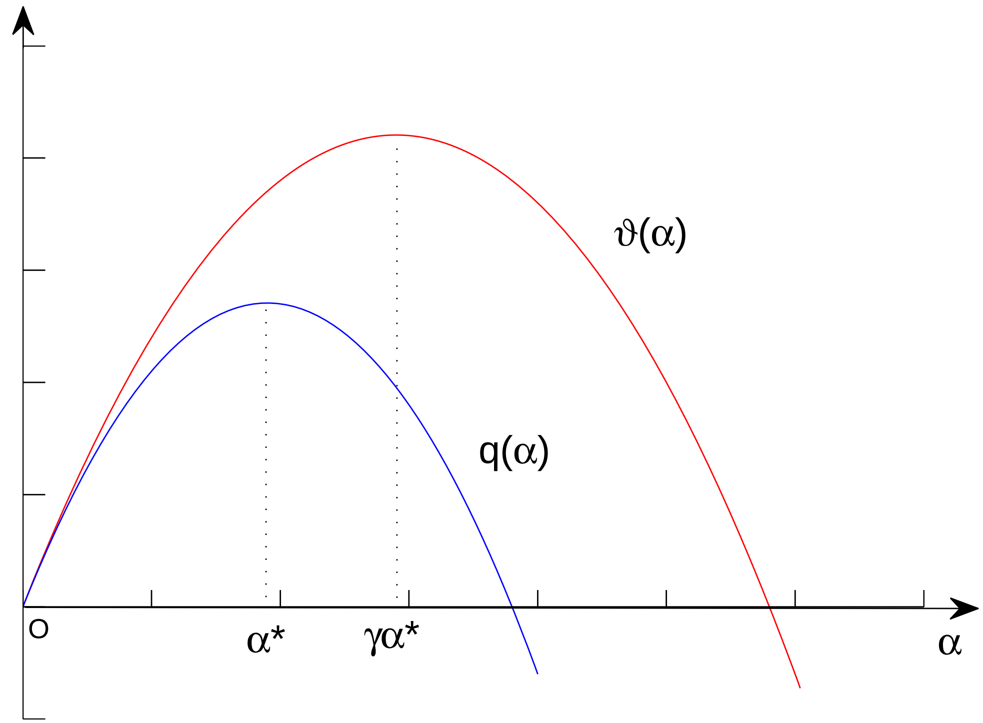

Fig.5 Interpretation of $\gamma \in [1, 2)$

Thus, in the practical computation, we take a relaxed factor $\gamma \in [1, 2)$, and set

$$u^{k+1} = u^k - \gamma \alpha_k^* d(u^k, \tilde{u}^k). \tag{2.8}$$

The reason for taking $\gamma \in [1,2)$ in the step size is depicted in Fig. 5.

From (2.5) and (2.6), we get

$$\begin{aligned} q_k(\gamma\alpha_k^*) &= 2\gamma\alpha_k^*\varphi(u^k,\tilde{u}^k) - \gamma^2(\alpha_k^*)^2\|d(u^k,\tilde{u}^k)\|^2 \\ &= \gamma(2-\gamma)\alpha_k^*\varphi(u^k,\tilde{u}^k). \end{aligned}$$

Thus, the new iterate u^{k+1} updated by (2.8) satisfies

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \gamma(2-\gamma)\alpha_k^*\varphi(u^k,\tilde{u}^k). \qquad (2.9)$$

By using (2.6) and (2.8), we have

$$\alpha_k^*\varphi(u^k,\tilde{u}^k) = \|\alpha_k^* d(u^k,\tilde{u}^k)\|^2 = \frac{1}{\gamma^2}\|u^k - u^{k+1}\|^2.$$

Substituting it in (2.9), we get the following inequality:

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{2-\gamma}{\gamma}\|u^k - u^{k+1}\|^2.$$

Primary algorithm Replacing $\varphi(u^k, \tilde{u}^k) \geq \delta\|u^k - \tilde{u}^k\|^2$ (see (2.1a)) by

$$\varphi(u^k, \tilde{u}^k) \geq \frac{1}{2}\big(\|d(u^k, \tilde{u}^k)\|^2 + \tau\|u^k - \tilde{u}^k\|^2\big), \quad (\tau > 0). \tag{2.10}$$

When the conditions (2.1b) and (2.10) are satisfied, we use the simple form

$$u^{k+1} = u^k - d(u^k, \tilde{u}^k), \tag{2.11}$$

to generate the new iterate. Since it uses the unit step size, we call (2.11) as **Primary Algorithm**. By a manipulation, we get

$$\begin{array}{rcl} \|u^{k+1} - u^*\|^2 & = & \|(u^k - u^*) - d(u^k, \tilde{u}^k)\|^2 \\ & = & \|u^k - u^*\|^2 - 2(u^k - u^*)^T d(u^k, \tilde{u}^k) + \|d(u^k, \tilde{u}^k)\|^2 \\ \text{(use (2.1b))} & \leq & \|u^k - u^*\|^2 - (2\varphi(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2). \end{array}$$

Since (2.10) is satisfied, the new iterate u^{k+1} satisfies

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \tau\|u^k - \tilde{u}^k\|^2. \tag{2.12}$$

The inequalities (2.9) and (2.12) tell us that the sequence $\{u^k\}$ is bounded, they are key inequalities for the convergence. By using (2.1a) and Theorem 1.1, it is easy to prove the following theorem.

Theorem 2.1 *Assume that* Ω^**, the solution set of the problem VI*(Ω, F)*, is nonempty, then the sequence generated by the projection and contraction methods converges to some* $u^* \in \Omega^*$*.*

It should be noticed that usually we advocate to use the general algorithm to determine the new iterate by calculating the step size. Although the primary algorithm does not need to calculate the step size, according to our experience, the general algorithm converges faster than the primal one.

- The projection and contraction methods can also be regarded as a prediction-correction methods.
- The vector $\tilde{u}^k$, which is obtained by the projection, can be viewed as a predictor. It provides us the error measure function and a profitable direction.
- The updating procedure, (2.8) or (2.11), which offered u^{k+1}, can be viewed as the correction pcocess.
- Whether the projection and contraction method or prediction-correction method, their first letter is P and C, so it is called PC Methods for short.

The differentiable convex optimization problem $\min\{f(x)\mid x\in\Omega\}$ is equivalent to the variational inequality

$$x\in\Omega, \qquad (x'-x)^T\nabla f(x)\geq 0, \;\; \forall x'\in\Omega. \tag{2.13}$$

If $\nabla f(x)$ is differentiable, its Hessian matrix, $\nabla^2 f(x)$ is symmetric. Especially, when $f(x)$ is a quadratic convex function, its Hessian matrix, is symmetric and positive semidefinite.

When we consider the general nonlinear monotone variational inequality VI(Ω, F), it is only required that

$$(u-v)^T(F(u)-F(v))\geq 0.$$

The Jacobian of F, $\nabla F(u)$, if it exist, it is not necessarily to be symmetric.

In linear variational inequality, the mapping $F(u)=Mu+q$ is affine. A linear variational inequality means that $M+M^T$ is positive semidefinite, but M is not necessarily symmetric.

Differentiable convex optimization problem is a kind of variational inequality (2.13) with special properties. We will give some better contraction algorithms for such convex optimization problems in Lecture 10.

3 PC Methods for LVI based on FI1+ FI2

In the linear variational inequality VI(Ω, F), $F(u) = Mu + q$ is an affine operator. Notice that $\tilde{u} = P_\Omega[u - \beta F(u)]$. Adding FI1 and FI2,

$$\begin{cases} (\tilde{u} - u^*)^T \beta F(u^*) \geq 0. & \text{(FI1)} \\ (\tilde{u} - u^*)^T \big([u - \beta F(u)] - \tilde{u}\big) \geq 0. & \text{(FI2)} \end{cases}$$

and using $F(u) = Mu + q$, we get

$$\{(u - u^*) - (u - \tilde{u})\}^T \{(u - \tilde{u}) - \beta M(u - u^*)\} \geq 0, \quad \forall u \in \Re^n, \; u^* \in \Omega^*.$$

Consequently, it follows that

$$(u - u^*)^T (I + \beta M^T)(u - \tilde{u}) \geq \|u - \tilde{u}\|^2 + \beta (u - u^*)^T M (u - u^*).$$

Since $(u-u^*)^TM(u-u^*) = \frac{1}{2}(u-u^*)^T(M^T+M)(u-u^*) \geq 0$, we have

$$(u-u^*)^T(I+\beta M^T)(u-\tilde{u}) \geq \|u-\tilde{u}\|^2, \quad \forall u \in \Re^n,\ u^* \in \Omega^*. \qquad (3.1)$$

Let

$$\varphi(u,\tilde{u}) = \|u-\tilde{u}\|^2 \qquad (3.2)$$

and

$$d(u,\tilde{u}) = (I+\beta M^T)(u-\tilde{u}). \qquad (3.3)$$

The $\varphi(u,\tilde{u})$ defined by (3.2) with the $d(u,\tilde{u})$ defined by (3.3), satisfy the conditions (2.1a) and (2.1b). Especially, in (2.1a), the positive parameter $\delta = 1$.

If we use the general algorithm, the step size α_k^* is determined (2.6) and the new iterate is given by (2.8). In details, for given u^k and $\beta > 0$, let

$$\tilde{u}^k = P_\Omega[u^k - \beta(Mu^k+q)]$$

to produce a predictor. The new iterate is given by

$$u^{k+1} = u^k - \gamma \alpha_k^* (I + \beta M^T)(u^k - \tilde{u}^k)$$

where

$$\alpha_k^* = \frac{\|u^k - \tilde{u}^k\|^2}{\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2}.$$

The generated sequence $\{u^k\}$ satisfies the following contractive property

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \gamma(2-\gamma)\alpha_k^* \|u^k - \tilde{u}^k\|^2.$$

In such method, usually the parameter β is sensitive for the convergence. Thus, we suggest to adjust the parameter β_k dynamically (in every 5-10 iterations), such that

$$\beta_k \|M^T(u^k - \tilde{u}^k)\| = \mathcal{O}(\|u^k - \tilde{u}^k\|).$$

♣ If the parameter β is selected to satisfy

$$\|(I + \beta M^T)(u^k - \tilde{u}^k)\|^2 \le (2-\tau)\|u^k - \tilde{u}^k\|^2, \quad \tau \in (0,1), \tag{3.4}$$

the $\varphi(u,\tilde{u})$ defined by (3.2) with the $d(u,\tilde{u})$ defined by (3.3) satisfy

$$2\varphi(u,\tilde{u}) \geq \|d(u,\tilde{u})\|^2 + \tau\|u-\tilde{u}\|^2.$$

This inequality tells us that the condition (2.10) is satisfied, and thus we can use the primary algorithm (2.11) (with step size 1) to update the new iterate.

According to (2.12), the generated sequence $\{u^k\}$ by the primary algorithm satisfies

$$\|u^{k+1}-u^*\|^2 \leq \|u^k-u^*\|^2 - \tau\|u^k-\tilde{u}^k\|^2.$$

More projection and contraction methods for LVI based on FI1+ FI2 can be found in [4] and [5]. Besides the contraction methods in Euclidean-norm, we can also build the contraction methods in G-norm, where G is a symmetric positive definite matrix. Especially, if we take $G=(I+\beta M^T)(I+\beta M)$ and consider to reduce $\|u-u^*\|_G^2$, the update form is

$$u^{k+1} = u^k - \gamma(I+\beta M)^{-1}(u^k-\tilde{u}^k), \quad \gamma \in (0,2). \tag{3.5}$$

According to this formula, we have

$$\begin{aligned}\|u^{k+1} - u^*\|_G^2 &= \|(u^k - u^*) - \gamma(I + \beta M)^{-1}(u^k - \tilde{u}^k)\|_G^2 \\ &= \|u^k - u^*\|_G^2 - 2\gamma(u^k - u^*)^T(I + \beta M^T)(u^k - \tilde{u}^k) \\ &\quad + \gamma^2\|(I + \beta M)^{-1}(u^k - \tilde{u}^k)\|_G^2.\end{aligned}$$

Using (3.1) and $G = (I + \beta M^T)(I + \beta M)$, from the last inequality follows that

$$\begin{aligned}\|u^{k+1} - u^*\|_G^2 &\le \|u^k - u^*\|_G^2 - 2\gamma\|u^k - \tilde{u}^k\|^2 + \gamma^2\|u^k - \tilde{u}^k\|^2 \\ &= \|u^k - u^*\|_G^2 - \gamma(2 - \gamma)\|u^k - \tilde{u}^k\|^2.\end{aligned}$$

By using this contraction method, the update form (3.5) is equivalent to solving the following system of equations:

$$(I + \beta M)(u^{k+1} - u^k) = \gamma(\tilde{u}^k - u^k).$$

Since the matrix in each iteration is in variant, we just need to do one LU decomposition for the matrix $(I + \beta M)$ in the whole iteration process.

M is a symmetric positive definite matrix

When M is symmetric positive definite, we replacing M by H, (3.1) becomes

$$(u-u^*)^T(I+\beta H)(u-\tilde{u}) \geq \|u-\tilde{u}\|^2, \quad \forall u \in \Re^n. \tag{3.6}$$

In this case, $G = I + \beta H$ is symmetric and positive definite. We consider the contraction method in G-norm. In this case, we use

$$u^{k+1} = u^k - \gamma\alpha_k^*(u^k-\tilde{u}^k), \quad \alpha_k^* = \frac{\|u^k-\tilde{u}^k\|^2}{\|u^k-\tilde{u}^k\|_G^2}, \quad \gamma \in (0,2) \tag{3.7}$$

to produce the new iterate. By using (3.6) and (3.7), we get

$$\begin{aligned}
\|u^{k+1}-u^*\|_G^2 \;&=\; \|(u^k-u^*) - \gamma\alpha_k^*(u^k-\tilde{u}^k)\|_G^2 \\
&= \|u^k-u^*\|_G^2 - 2\gamma\alpha_k^*(u^k-u^*)^T G(u^k-\tilde{u}^k) + \gamma^2(\alpha_k^*)^2\|u^k-\tilde{u}^k\|_G^2 \\
&\leq \|u^k-u^*\|_G^2 - 2\gamma\alpha_k^*\|u^k-\tilde{u}^k\|^2 + \gamma^2(\alpha_k^*)^2\|u^k-\tilde{u}^k\|_G^2.
\end{aligned}$$

Since $\alpha_k^*\|u^k-\tilde{u}^k\|_G^2 \overset{(3.7)}{=} \|u^k-\tilde{u}^k\|^2$, it follows from the above inequality

$$\|u^{k+1}-u^*\|_G^2 \leq \|u^k-u^*\|_G^2 - \gamma(2-\gamma)\alpha_k^*\|u^k-\tilde{u}^k\|^2, \quad \forall u^* \in \Omega^*$$

which is the key inequality for the proof of the convergence.

4 PC Methods for NVI based on FI1+FI2+FI3

We consider the contraction method for the general nonlinear variational inequality (1.1). Again, for given u, $\tilde{u} = P_\Omega[u - \beta F(u)]$ is obtained by the projection and it is a predictor. Adding FI1, FI2 and FI3,

$$\begin{cases} (\tilde{u} - u^*)^T \beta F(u^*) \ge 0 & \text{(FI1)} \\ (\tilde{u} - u^*)^T \big([u - \beta F(u)] - \tilde{u}\big) \ge 0 & \text{(FI2)} \\ (\tilde{u} - u^*)^T \big(\beta F(\tilde{u}) - \beta F(u^*)\big) \ge 0 & \text{(FI3)} \end{cases}$$

we get

$$\{(u^k - u^*) - (u^k - \tilde{u})\}^T \{(u - \tilde{u}) - \beta[F(u) - F(\tilde{u})]\} \ge 0. \tag{4.1}$$

By defining

$$d(u, \tilde{u}) = (u - \tilde{u}) - \beta\big(F(u) - F(\tilde{u})\big), \tag{4.2}$$

and

$$\varphi(u, \tilde{u}) = (u - \tilde{u})^T d(u, \tilde{u}), \tag{4.3}$$

it follows from (4.1) that

$$(u-u^*)^T d(u,\tilde{u}) \ge \varphi(u,\tilde{u}), \quad \forall\, u \in \Re^n.$$

The condition (2.1b) is satisfied. Whether $\varphi(u,\tilde{u})$ is an error measure function ? Under the assumption that F is Lipschitz continuous, for a given $\nu \in (0,1)$, we can use **Armijo** line-search strategy to get a β such that

$$(u-\tilde{u})^T \big(\beta F(u) - \beta F(\tilde{u})\big) \le \nu \|u-\tilde{u}\|^2, \;\; \nu \in (0,1). \tag{4.4}$$

Usually, such β also fits the request $\beta\|F(u)-F(\tilde{u})\| = \mathcal{O}(\|u-\tilde{u}\|)$. According to (4.2) and (4.4), we have

$$\begin{aligned} \varphi(u,\tilde{u}) &= (u-\tilde{u})^T d(u,\tilde{u}) \\ &= \|u-\tilde{u}\|^2 - (u-\tilde{u})^T \beta\big(F(u)-F(\tilde{u})\big) \\ &\ge (1-\nu)\|u-\tilde{u}\|^2. \end{aligned}$$

Indeed, $\varphi(u,\tilde{u})$ is an error measure function, when (4.4) is satisfied.

The $\varphi(u,\tilde{u})$ defined by (4.3) with the $d(u,\tilde{u})$ defined by (4.2), satisfy the conditions (2.1a) and (2.1b). In (2.1a), the positive parameter $\delta = 1-\nu$.

By using the update form (2.8) with α_k^* given by (2.6), the generated sequence $\{u^k\}$ satisfies

$$\|u^{k+1}-u^*\|^2 \le \|u^k-u^*\|^2 - \gamma(2-\gamma)\alpha_k^*(1-\nu)\|u^k-\tilde{u}^k\|^2. \quad (4.5)$$

♣ Especially, when

$$\beta\|F(u)-F(\tilde{u})\| \le \nu\|u-\tilde{u}\|, \quad (4.6)$$

the direction $d(u,\tilde{u})$ defined by (4.2) and the $\varphi(u,\tilde{u})$ defined by (4.3) satisfy

$$\begin{aligned} & 2\varphi(u,\tilde{u}) - \|d(u,\tilde{u})\|^2 \\ = \;& 2(u-\tilde{u})^T d(u,\tilde{u}) - \|d(u,\tilde{u})\|^2 \\ = \;& d(u,\tilde{u})^T\big\{2(u-\tilde{u}) - d(u,\tilde{u})\big\} \\ = \;& \big\{(u-\tilde{u}) - \beta\big(F(u)-F(\tilde{u})\big)\big\}^T\big\{(u-\tilde{u}) + \beta\big(F(u)-F(\tilde{u})\big)\big\} \\ = \;& \|u-\tilde{u}\|^2 - \beta^2\|F(u)-F(\tilde{u})\|^2 \;\ge\; (1-\nu^2)\|u-\tilde{u}\|^2. \end{aligned} \quad (4.7)$$

This tell us that the condition (2.10) is satisfied with $\tau = 1 - \nu^2$.

Thus, under the condition (4.6), we can use the primary algorithm to update the new iterate, namely,

$$u^{k+1} = u^k - d(u^k, \tilde{u}^k).$$

According to (2.12), we have

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

The earliest PC Algorithms for LVI appears in [4, 5] which is based on adding FI1 and FI2. The late PC Methods for nonlinear variational inequality [6, 8, 12] are based on FI1+FI2+FI3 .

For nonlinear variational inequality, even though the ‘strict’ condition (4.6) is satisfied, we still advocate using the computational step size for determining the next iteration point

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k),$$

where

$$\alpha_k^* = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}.$$

Since the (4.6) is satisfied, the left hand side of (4.7) is strictly greater than 0. Thus, for any $k > 0$, we have $\alpha_k^* > 1/2$, and

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{\gamma(2-\gamma)(1-\nu)}{2}\|u^k - \tilde{u}^k\|^2. \qquad (4.8)$$

The formulation of three fundamental inequalities and related methods can be seen in [6] (the first version is the preprint 94-11, Institute of mathematics, Nanjing University) and [15]. After graduating from Nanjing University, Defeng Sun also independently found the direction (4.2) and constructed the projection contraction algorithm [12]. Thank him for citing [6] in the footnote and the references of his paper [12].

The greatest truths are the simplest ! The construction of projection contraction algorithms is based on three fundamental inequalities. Its principle is simple and unified, it gives us beautiful enjoyment !

5 The extra-gradient method

The extra-gradient method [9] (abbreviated to EG method) can be applied to solve the nonlinear monotone variational inequality. Some young scholars in the universities north America use the EG method to solve the problems arising from the area of information science, such as speech recognition, optical fiber network, machine learning and other-s. In their doctoral dissertation, they have mentioned the relationship between the EG method and PC Algorithms. For preparing to compare their numerical efficiency in the next section, here, we will introduce the extra-gradient method. In fact, the extra gradient algorithm is a prediction-correction form of the proximal point algorithm.

The proximal point algorithm Let us first briefly review the Proximal Point Algorithm (abbreviated to PPA) VI(Ω, F) (see (1.1)). PPA is an iterative method. For given u^k and $r > 0$, the new iterate u^{k+1} is the solution of the following variational inequality:

$$u^{k+1} \in \Omega, \quad (u - u^{k+1})^T \{F(u^{k+1}) + r(u^{k+1} - u^k)\} \geq 0, \ \forall \ u \in \Omega. \tag{5.1}$$

It is clear that u^{k+1} is a solution of (1.1) if and only if $u^{k+1} = u^k$. In the case of $u^{k+1} \neq u^k$, by setting $u = u^*$ in (5.1), we obtain

$$(u^{k+1} - u^*)^T r(u^k - u^{k+1}) \geq (u^{k+1} - u^*)^T F(u^{k+1}). \tag{5.2}$$

Because F is monotone, we have

$$(u^{k+1}-u^*)^T F(u^{k+1}) = (u^{k+1}-u^*)^T F(u^*) \geq 0$$

and consequently from (5.4), we obtain

$$(u^{k+1}-u^*)^T (u^k - u^{k+1}) \geq 0,$$

and thus

$$(u^k-u^*)^T (u^k - u^{k+1}) \geq \|u^k - u^{k+1}\|^2.$$

By using the last inequality, we obtain

$$\|u^{k+1}-u^*\|^2 \leq \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2. \tag{5.3}$$

The sequence $\{u^k\}$ generated by PPA has the nice convergence property, however, the subproblem (5.1) is almost difficult as the original problem (1.1). Thus, the classical PPA is not widely used in the practical application.

By using $\beta = 1/r$ in (5.1), u^{k+1} can be viewed as

$$u^{k+1} \in \Omega, \quad (u-u^{k+1})^T\{(u^{k+1}-u^k) + \beta F(u^{k+1})\} \geq 0, \ \forall\, u \in \Omega. \tag{5.4}$$

By using the equivalent representation of VI (see Theorem 1.1), it can be written as

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

In other words,

$$u^{k+1} = P_\Omega[u^k - \beta F(u^{k+1})]. \tag{5.5}$$

The extra-gradient method It is difficult to directly get the solution of (5.5), because the both sides of this equation include the unknown u^{k+1}. Replacing u^{k+1} in the right hand side of (5.5) by u^k, we denote the output by $\tilde{u}^k$ which is produced by the projection

$$\tilde{u}^k = P_\Omega[u^k - \beta F(u^k)], \tag{5.6a}$$

We call $\tilde{u}^k$ the predictor. Then, replacing u^{k+1} in the right hand side of (5.5) by the predictor $\tilde{u}^k$, we obtain the (corrector) new iterate

$$u^{k+1} = P_\Omega[u^k - \beta F(\tilde{u}^k)] \tag{5.6b}$$

The method (5.6) is called the extra-gradient method (EG-method). Each iteration of the EG method includes two projections on Ω. In the prediction step, the parameter β should be chosen to satisfy the following condition:

$$\beta\|F(u^k) - F(\tilde{u}^k)\| \le \nu\|u^k - \tilde{u}^k\|, \quad \nu \in (0,1), \tag{5.7}$$

which is same as in (4.6).

Convergence analysis of the extra-gradient algorithm The analysis is based on the basic property of the projection (1.5) and its consequent inequality (1.9).

♣ First, since u^{k+1} is the projection of $[u^k - \beta F(\tilde{u}^k)]$ on Ω, according to (1.9), we have

$$\|u^{k+1} - u^*\|^2 \le \|(u^k - \beta F(\tilde{u}^k)) - u^*\|^2 - \|(u^k - \beta F(\tilde{u}^k)) - u^{k+1}\|^2. \tag{5.8}$$

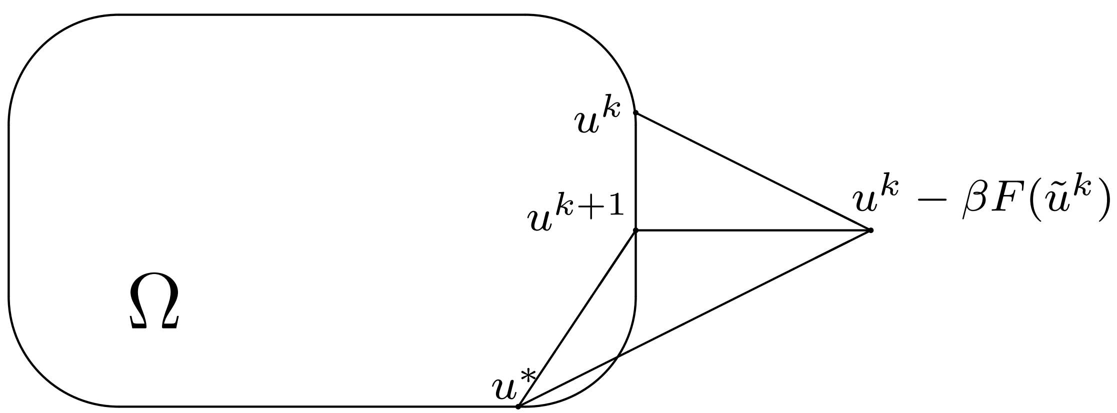

Fig.6 Geometric interpretation of inequality (5.8)

By a manipulation, it follows from (5.8) that

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2 - 2(u^{k+1} - u^*)^T \beta F(\tilde{u}^k). \quad (5.9)$$

Since $(\tilde{u}^k - u^*)^T F(\tilde{u}^k) \ge (\tilde{u}^k - u^*)^T F(u^*) \ge 0$, adding the nonnegative term $2(\tilde{u}^k - u^*)^T F(\tilde{u}^k)$ to the right hand side of (5.9), we get

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \|u^k - u^{k+1}\|^2 - 2(u^{k+1} - \tilde{u}^k)^T \beta F(\tilde{u}^k). \quad (5.10)$$

The quadratic term $\|u^k - u^{k+1}\|^2$ in the right hand side of (5.10) can be written in form

$$\|u^k - u^{k+1}\|^2 = \|u^k - \tilde{u}^k\|^2 + \|\tilde{u}^k - u^{k+1}\|^2 + 2(\tilde{u}^k - u^{k+1})^T (u^k - \tilde{u}^k).$$

Substituting it in (5.10), we get

$$\begin{aligned} \|u^{k+1} - u^*\|^2 \quad \le \quad & \|u^k - u^*\|^2 - \|u^k - \tilde{u}^k\|^2 - \|\tilde{u}^k - u^{k+1}\|^2 \\ & +2(u^{k+1} - \tilde{u}^k)^T [u^k - \beta F(\tilde{u}^k) - \tilde{u}^k]. \end{aligned} \quad (5.11)$$

♣ Now, we use the tool inequality (1.5). Setting $v = u^k - \beta F(u^k)$, then $\tilde{u}^k = P_\Omega(v)$, because $u^{k+1} \in \Omega$, according to (1.5), we have

$$2(\tilde{u}^k - u^{k+1})^T \{[u^k - \beta F(u^k)] - \tilde{u}^k\} \ge 0.$$

Adding the left term to the right hand side of (5.11), we get

$$\begin{aligned} \|u^{k+1} - u^*\|^2 \quad &\leq \quad \|u^k - u^*\|^2 - \|u^k - \tilde{u}^k\|^2 - \|\tilde{u}^k - u^{k+1}\|^2 \\ &\quad +2(u^{k+1} - \tilde{u}^k)^T \beta[F(u^k) - F(\tilde{u}^k)]. \end{aligned} \tag{5.12}$$

♣ Applying the Cauchy–Schwarz inequality to the cross term of the RHS of (5.12),

$$2(u^{k+1} - \tilde{u}^k)^T \beta[F(u^k) - F(\tilde{u}^k)] \leq \|u^{k+1} - \tilde{u}^k\|^2 + \beta^2\|F(u^k) - \beta F(\tilde{u}^k)\|^2.$$

Substituting it in (5.12), it follows that

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \|u^k - \tilde{u}^k\|^2 + \beta^2\|F(u^k) - F(\tilde{u}^k)\|^2. \tag{5.13}$$

The contractive property of the extra-gradient method According (5.13), when the condition (5.7) is satisfied, the sequence generated by the extra-gradient method (5.6) has the following contractive property:

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - (1 - \nu^2)\|u^k - \tilde{u}^k\|^2. \tag{5.14}$$

The inequality (5.14) is the key for the proof of the convergence of the extra-gradient method. This short proof is based on the strategy in [8]. For the convergence proof of the extra-gradient algorithm, readers can also refer to [1] (Vol. II pp. 1115-1118).

6 Numerical experiments

For comparing the efficiency of the PC Algorithms and the EG Method, we test the nonlinear complementarity problem (a class of VI(Ω, F) with $\Omega = \Re^n_+$)

$$u \geq 0, \qquad F(u) \geq 0, \qquad u^T F(u) = 0.$$

In the test examples, we take

$$F(u) = D(u) + Mu + q,$$

where $Mu + q$ and $D(u)$ are the linear part and nonlinear part of $F(u)$, respectively. The linear part $Mu + q$ is generated as in [2][a], using Matlab, it produced by

```
A=(rand(n,n)-0.5)*10;  B=(rand(n,n)-0.5)*10;   B=B-B';  M=A'*A+B;
q=(rand(n,1)-0.5)*1000;        or          q=(rand(n,1)-1.0)*500;
```

In the nonlinear part $D(u)$, each element is given by $D_j(u) = d_j * \arctan(u_j)$, where d_j is a random variable between $(0, 1)$, similarly as in [13][b].

[a] In the paper by Harker and Pang [2], the matrix $M = A^T A + B + D$, where A and B are the same matrices as here, and D is a diagonal matrix with uniformly distributed random variable $d_{jj} \in (0.0, 0.3)$. In our test examples $d_{jj} \equiv 0$.

[b] In [13], the components of nonlinear mapping $D(u)$ are $D_j(u) =$ constat $* \arctan(u_j)$. Thus, $D_j(u)$ in our test example is more general.

As in [8], we have refined the EG method, and thus use the following procedure.

Refined extra-gradient method:

Step 0. Set $\beta_0 = 1$, $\nu \in (0,1)$, $u^0 \in \Omega$ 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 $r_k > \nu$, $\beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\}$,

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

$$r_k := \frac{\beta_k \|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

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

If $r_k \le \mu$ then $\beta_k := \beta_k * 1.5$, **end(if)**

Step 2. $\beta_{k+1} = \beta_k$ and $k = k+1$, go to Step 1.

When the EG method applied to solve NCP, the other people use the above program but omit the sentence **If** $r_k \le \mu$ then $\beta_k := \beta_k * 1.5$ **end(if)**. Our calculation experiments shows that if this sentence is omitted, the number of iteration steps will be greatly increased, sometimes even leading to calculation failure.

The PC Algorithm (the extra computation than EG Method is indicated in a small box):

Projection and Contraction Method:

Step 0. Set $\beta_0 = 1$, $\nu \in (0,1)$, $u^0 \in \Omega$ 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 $r_k > \nu$, $\beta_k := \frac{2}{3}\beta_k * \min\{1, \frac{1}{r_k}\}$,

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

$$r_k := \frac{\beta_k \|F(u^k) - F(\tilde{u}^k)\|}{\|u^k - \tilde{u}^k\|},$$

end(while)

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

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

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

If $r_k \le \mu$ then $\beta_k := \beta_k * 1.5$, **end(if)**

Step 2. $\beta_{k+1} = \beta_k$ and $k = k + 1$, go to Step 1.

Both the EG Method and the PC Algorithm compared here can be regarded as prediction correction methods. They use the same formula

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

to produce the predictor $\tilde{u}^k$. In order to compare the efficiency, we all require that the predictor points meet (see (5.7))

$$\beta\|F(u^k) - F(\tilde{u}^k)\| \le \nu\|u^k - \tilde{u}^k\|, \quad \nu \in (0,1).$$

The only difference is that EG Method uses (see (5.6b))

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

to update the new iterate u^{k+1}. while the PC Algorithm gives the next iterate u^{k+1} by

$$u^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k), \quad \gamma \in [1,2)$$

where (see (4.2))

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta\big(F(u^k) - F(\tilde{u}^k)\big),$$

and $\alpha_k^* = (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)/\|d(u^k, \tilde{u}^k)\|^2$. The extra work for calculating α_k^* in P-C Algorithm is small, and the correction does not need to doing projection.

In the following, we give Matlab codes for the two different algorithms

The Matlab Code of the Refined Extra-Gradient Method

```
function REG(n,M,q,d,xstart,tol,pfq)                                      %(1)
fprintf('Extragragient Method by Korpelevich   n=%4d \n',n);             %(2)
x=xstart;  Fx= d.*atan(x) + M*x + q;  stopc=norm(x-max(x-Fx,0),inf);    %(3)
beta=1;        k=0;         l=0;         tic;                             %(4)
while (stopc>tol && k<=2000)                                              %(5)
  if mod(k,pfq)==0 fprintf(' k=%4d    epsm=%9.3e  \n',k,stopc); end;     %(6)
    x0=x;   Fx0=Fx;    k=k+1;                                             %(7)
    x=max(x0-Fx0*beta,0);        Fx=d.*atan(x) + M*x + q;   l=l+1;       %(8)
       dx=x0-x;   df=(Fx0-Fx)*beta;                                       %(9)
       r=norm(df)/norm(dx);                                              %(10)
       while r>0.9     beta=0.7*beta*min(1,1/r);    l=l+1;               %(11)
         x=max(x0-Fx0*beta,0);    Fx=d.*atan(x) + M*x + q;               %(12)
         dx=x0-x;   df=(Fx0-Fx)*beta;     r=norm(df)/norm(dx);           %(13)
       end;                                                              %(14)
       x=max(x0-Fx*beta,0);                                              %(15)
       Fx= d.*atan(x) + M*x + q;     l=l+1;                              %(16)
       ex=x-max(x-Fx,0);     stopc=norm(ex,inf);                         %(17)
       if r <0.4  beta=beta*1.5;  end;                                   %(18)
end;  toc;   fprintf(' k=%4d   epsm=%9.3e     l=%4d  \n',k,stopc,l); %%%%
```

Replacing (15) in the EG code by (15a)-(15b) in the next code, we get the code for PC Algorithm.

The Matlab Code of The Projection and Contraction Method

```
function PC_G(n,M,q,d,xstart,tol,pfq)                                        %(1)
fprintf('PC Method use Direction D1 with gamma a*  n=%4d \n',n);            %(2)
x=xstart;  Fx= d.*atan(x) + M*x + q;  stopc=norm(x-max(x-Fx,0),inf);         %(3)
beta=1;        k=0;        l=0;        tic;                                  %(4)
while (stopc>tol && k<=2000)                                                 %(5)
  if mod(k,pfq)==0 fprintf(' k=%4d    epsm=%9.3e  \n',k,stopc); end;          %(6)
    x0=x;   Fx0=Fx;    k=k+1;                                                %(7)
    x=max(x0-Fx0*beta,0);        Fx=d.*atan(x) + M*x + q;   l=l+1;            %(8)
       dx=x0-x;   df=(Fx0-Fx)*beta;                                          %(9)
       r=norm(df)/norm(dx);                                                 %(10)
       while r>0.9     beta=0.7*beta*min(1,1/r);   l=l+1;                    %(11)
         x=max(x0-Fx0*beta,0);    Fx=d.*atan(x) + M*x + q;                   %(12)
         dx=x0-x;   df=(Fx0-Fx)*beta;     r=norm(df)/norm(dx);               %(13)
       end;                                                                 %(14)
       dxf=dx-df;    r1=dx'*dxf;    r2=dxf'*dxf;     alpha=r1/r2;           %(15a)
       x=x0- dxf*alpha*1.9;                                                 %(15b)
       Fx= d.*atan(x) + M*x + q;      l=l+1;                                %(16)
       ex=x-max(x-Fx,0);      stopc=norm(ex,inf);                           %(17)
       if r <0.4  beta=beta*1.5;  end;                                      %(18)
end;  toc;   fprintf(' k=%4d   epsm=%9.3e     l=%4d  \n',k,stopc,l); %%%%
```

Instead of (15) in the EG method code, the Matlab code of PC Algorithm is (15a) and (15b).

Table 1. Numerical results for Easy Problems $q \in (-500, 500)$

	Extra-gradient Method			General PC-Method		
$n =$	No. It	No. F	CPU	No. It	No. F	CPU
500	724	1485	0.26	468	977	0.17
1000	804	1650	2.85	514	1079	1.86
2000	776	1593	10.33	407	864	5.63

Table 2. Numerical results for Hard Problems $q \in (-500, 0)$

	Extra-gradient Method			General PC-Method		
$n =$	No. It	No. F	CPU	No. It	No. F	CPU
500	1453	2983	0.53	865	1824	0.33
1000	2026	4159	7.12	1199	2553	4.38
2000	1702	3494	22.45	1025	2177	14.00

The PC method converges faster than the refined extra-gradient method.

$$\frac{\text{It. No. of Projection and Contraction Method}}{\text{It. No. of The refined extra-gradient Method}} \approx 60\%.$$

Usually, for the same problems, the PC Algorithms can save about 40% CPU time than the EG Method.

In both of the PC Algorithm and the EG Method, the computational load of each iteration is $\mathcal{O}(n^2)$. In each iteration of PC Algorithm, the computational cost for determining the step length is $\mathcal{O}(n)$, which is a small proportion in the whole calculations.

Some Ph.D Dissertations which use the EG Method to solve their problems

- Fei Sha, Large Margin Training of Acoustic Models for Speech Recognition, PhD Thesis, Computer and Information Science, University of Pennsylvania, 2007. 语音识别
- Yan Pan, A game theoretical approach to constrained OSNR optimization problems in optical network, PhD Thesis, Electrical & Computer Engineering. University of Toronto, 2009. 光纤网络
- Simon Lacoste-Julien, Discriminative Machine Learning with Structure, PhD Thesis, Computer Science. University of California, Berkeley, 2009. 机器学习
- A. G. Howard, Large Margin, Transformation Learning, PhD Thesis, Graduate School of Arts and Science. Columbia University, 2009. 机器学习

All these theses have mentioned the PC Algorithm[6]. If PC Algorithm is used instead of EG Method, the convergence speed will be greatly improved.

References

[1] F. Facchinei and J.S. Pang, Finite-dimensional variational inequality and complementarity problems, Volume I and II, Springer Series in Operations Research, 2003.

[2] Harker, P.T. Harker and J.S. Pang *A damped-Newton method for the linear complementarity problem,* Lectures in Applied Mathematics **26**, 265–284, 1990.

[3] B.S. He, A projection and contraction method for a class of linear complementarity problems and its application in convex quadratic programming, Applied Mathematics and Optimization, **25**, 247–262, 1992.

[4] B.S. He, A new method for a class of linear variational inequalities, Math. Progr., **66**, 137-144, 1994.

[5] B.S. He, Solving a class of linear projection equations, Numerische Mathematik, **68**, 71-80, 1994.

[6] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Applied Mathematics and Optimization, **35**, 69-76, 1997.

[7] B. S. He, Solving trust region problem in large scale optimization, J. Compu. Math. **18**, 1-12, 2000

[8] B.S He and L-Z Liao, Improvements of some projection methods for monotone nonlinear variational inequalities, Journal of Optimization Theory and Applications, **112**, 111-128, 2002

[9] G. M. Korpelevich. The extragradient method for finding saddle points and other problems, Ekonomika i Matematicheskie Metody, **12**, 747-756, 1976.

[10] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimension I*. Springer-Verlag, 1970.

[11] D.F. Sun, A new step-size skill for solving a class of nonlinear projection equations, JCM, **13**, 357-368, 1995.

[12] D.F. Sun, A class of iterative methods for solving nonlinear projection equations, JOTA **91**, 123-140, 1996.

[13] K. Taji, M. Fukushima, and T. Ibaraki, A globally convergent Newton method for solving strongly monotone variational inequalities, Math. Progr. **58**, 369-383, 1993.

[14] T. Zhu and Z. G. Yu, A simple proof for some important properties of the projection mapping. Math. Inequal. Appl. **7**, 453–456, 2004.

[15] 何炳生, 论求解单调变分不等式的一些投影收缩算法,〖计算数学〗, **18**, 54-60, 1996.