凸优化和单调变分不等式的收缩算法

第五讲: Lipschitz 连续的单调变分不等式投影收缩算法的收敛速率

Convergence rate of the PC methods for Lipschitz continuous monotone VIs

南京大学数学系　　何炳生

hebma@nju.edu.cn

The context of this lecture is based on the publication [2]

In 2005, Nemirovski's analysis 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, we have developed a class of Fejér monotone projection and contraction methods. 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.

1 Introduction

Let Ω be a nonempty closed convex subset of $\Re^n$, F be a continuous mapping from $\Re^n$ to itself. The variational inequality problem, denoted by VI(Ω, F), is to find a vector

$u^* \in \Omega$ such that

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

Notice that VI(Ω, F) is invariant when F is multiplied by some positive scalar $\beta > 0$. It is well known 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 Ω 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.*,

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

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

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

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

Among the algorithms for monotone variational inequalities, the extragradient (EG) method proposed by Korpelevich [9] is one of the 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

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

Since $\tilde{u}^k$ is not accepted as the new iterate, for designation convenience, we call it as a *predictor* and β_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

$$u^{k+1} = P_\Omega[u^k - \beta_k F(\tilde{u}^k)], \tag{1.4}$$

where β_k is called the *correction step size*. The sequence $\{u^k\}$ generated by the extragradient method is Fejér monotone with respect to the solution set, namely,

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - (1 - \nu^2)\|u^k - \tilde{u}^k\|^2. \tag{1.5}$$

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 (1.4) are equal. Thus the 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 [15], optical network [11] and speech recognition [12], etc. In addition, Nemirovski [10] and Tseng [16] 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, we devoted our effort to develop a class of projection and contraction (PC) methods for monotone variational inequalities [5, 6, 8, 13]. 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 (1.3) just as in the extragradient method. The PC methods exploit a pair of geminate directions [7, 8] offered by the predictor, namely, they are

$$d(u^k,\tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k(F(u^k) - F(\tilde{u}^k)) \quad \text{and} \quad \beta_k F(\tilde{u}^k). \tag{1.6}$$

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 more closed to the solution set. It leads to choosing the 'optimal' step length

$$\varrho_k = \frac{(u^k - \tilde{u}^k)^T d(u^k,\tilde{u}^k)}{\|d(u^k,\tilde{u}^k)\|^2}, \tag{1.7}$$

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

$$u^{k+1} = u^k - \gamma\varrho_k d(u^k,\tilde{u}^k), \tag{1.8}$$

or

$$u^{k+1} = P_\Omega[u^k - \gamma\varrho_k\beta_k F(\tilde{u}^k)]. \tag{1.9}$$

The PC methods (without line search) make one (or two) projection(s) on Ω at each iteration, and the distance of the iterates to the solution set monotonically converges to zero. According to the terminology in [1], 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 (1.4) and (1.9)). 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 Ω^* can be characterized as (see (2.3.2) in pp. 159 of [3])

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

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

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

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

$$\tilde{u} \in \Omega \quad \text{and} \quad \sup_{u\in\mathcal{D}} \big\{(\tilde{u}-u)^T F(u)\big\} \leq \epsilon. \tag{1.10}$$

As a byproduct of the complexity analysis, we find why taking a suitable relaxation factor

$\gamma \in (1,2)$ in the correction steps (1.8) and (1.9) 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 Section 3, we investigate the geminate descent directions of the distance function. Section 4 shows the contraction property of the PC methods. In Section 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. The solution methods study for variational inequality is helpful for investigating the splitting contraction methods for separable convex optimization, 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 Ω^*. 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 Ω in the Euclidean-norm has a closed form and it is easy to be carried out. Since

$$P_\Omega(v) = \text{argmin}\{\frac{1}{2}\|u-v\|^2 \mid u \in \Omega\},$$

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

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

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

$$\begin{aligned}
\|u-v\|^2 &= \|(u - P_\Omega(v)) - (v - P_\Omega(v))\|^2 \\
&= \|u - P_\Omega(v)\|^2 - 2(v - P_\Omega(v))^T(u - P_\Omega(v)) + \|v - P_\Omega(v)\|^2 \\
&\ge \|u - P_\Omega(v)\|^2 + \|v - P_\Omega(v)\|^2.
\end{aligned}$$

Therefore, we have

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

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 VI(Ω, F) because

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

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

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

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

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

Under the assumption that F is monotone we have

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

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.

3 Predictor and the ascent directions

For given u^k, the predictor $\tilde{u}^k$ in the projection and contraction methods [5, 6, 8, 13] is produced by (1.3). 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 make an initial guesses of $\beta = \nu/L$ and decrease β by a constant factor and repeat the procedure whenever (1.3b) is violated.

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[a] at the point u. A direction d is called an ascent direction of $\frac{1}{2}\|u - u^*\|^2$ at u if and only if the inner-product $(u - u^*)^T d > 0$.

[a]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.

3.1 Ascent directions by adding the fundamental inequalities

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

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

where

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k \big(F(u^k) - F(\tilde{u}^k)\big), \tag{3.2}$$

which is the same $d(u^k, \tilde{u}^k)$ defined in (1.6). It follows from (3.1) that

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k). \tag{3.3}$$

Note that, under the condition (1.3b), we have

$$\begin{aligned} & 2(u^k - \tilde{u}^k) d(u^k, \tilde{u}^k) - \|d(u^k, \tilde{u}^k)\|^2 \\ & \quad = \quad d(u^k, \tilde{u}^k)^T \{2(u^k - \tilde{u}^k) - d(u^k, \tilde{u}^k)\} \\ & \quad = \quad \|u^k - \tilde{u}^k\|^2 - \beta_k^2 \|F(u^k) - F(\tilde{u}^k)\|^2 \\ & \quad \geq \quad (1 - \nu^2) \|u^k - \tilde{u}^k\|^2. \end{aligned} \tag{3.4}$$

Consequently, from (3.3) and (3.4) we have

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \frac{1}{2}\big(\|d(u^k, \tilde{u}^k)\|^2 + (1-\nu^2)\|u^k - \tilde{u}^k\|^2\big).$$

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.

3.2 Geminate ascent directions

To the direction $d(u^k, \tilde{u}^k)$ defined in (3.2), there is a correlative ascent direction $\beta_k F(\tilde{u}^k)$. Use the notation of $d(u^k, \tilde{u}^k)$, the projection equation (1.3a) can be written as

$$\tilde{u}^k = P_\Omega\{\tilde{u}^k - [\beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k)]\}. \tag{3.5a}$$

It follows that $\tilde{u}^k$ is a solution of VI(Ω, F) if and only if $d(u^k, \tilde{u}^k) = 0$. Assume that there is a constant $c > 0$ such that

$$\varrho_k = \frac{(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2} \geq c, \quad \forall k \geq 0. \tag{3.5b}$$

In this paper, we call (3.5) with $c > 0$ *the general conditions* and the forthcoming analysis is based of these conditions. For given u^k, there are different ways to construct $\tilde{u}^k$ and $d(u^k, \tilde{u}^k)$ which satisfy the conditions (3.5) (see [8] for an example). If β_k satisfies (1.3b) and $d(u^k, \tilde{u}^k)$ is given by (3.2), the general conditions (3.5) are satisfied with $c \geq \frac{1}{2}$ (see (3.4)). Note that an equivalent expression of (3.5a) is

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^T \{\beta_k F(\tilde{u}^k) - d(u^k, \tilde{u}^k)\} \geq 0, \quad \forall\, u \in \Omega, \tag{3.6a}$$

and from (3.5b) we have

$$(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) = \varrho_k \|d(u^k, \tilde{u}^k)\|^2. \tag{3.6b}$$

In fact, $d(u^k, \tilde{u}^k)$ and $\beta_k F(\tilde{u}^k)$ in (3.5a) are a pair of geminate directions and usually 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 β_k. If $d(u^k, \tilde{u}^k) = u^k - \tilde{u}^k$, then $\tilde{u}^k$ in (3.6a) is the solution of the subproblem in the k-th iteration when PPA applied to solve VI(Ω, F). Hence, the projection and contraction methods considered in this paper belong to the *prox-like contraction methods*.

The following lemmas tell us that both the direction $d(u^k, \tilde{u}^k)$ (for $u^k \in \Re^n$) 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 the general conditions* (3.5) *be satisfied. Then we have*

$$(u^k - u^*)^T d(u^k, \tilde{u}^k) \geq \varrho_k \|d(u^k, \tilde{u}^k)\|^2, \quad \forall\, u^k \in \Re^n,\, u^* \in \Omega^*. \tag{3.7}$$

Proof. Note that $u^* \in \Omega$. By setting $u = u^*$ in (3.6a) (the equivalent expression of (3.5a)), we get

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

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

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

The assertion (3.7) is followed from the above inequality and (3.6b) directly. $\square$

Lemma 3.2 *Let the general conditions* (3.5) *be satisfied. If* $u^k \in \Omega$*, then we have*

$$(u^k - u^*)^T \beta_k F(\tilde{u}^k) \ge \varrho_k \|d(u^k, \tilde{u}^k)\|^2, \quad \forall\, u^* \in \Omega^*. \tag{3.8}$$

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

$$(u^k - u^*)^T \beta_k F(\tilde{u}^k) \ge (u^k - \tilde{u}^k)^T \beta_k F(\tilde{u}^k), \;\; \forall\, u^* \in \Omega^*.$$

Note that because $u^k \in \Omega$, by setting $u = u^k$ in (3.6a), we get

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

From the above two inequalities follows that

$$(u^k - u^*)^T \beta_k F(\tilde{u}^k) \ge (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k), \;\; \forall u^* \in \Omega^*.$$

The assertion (3.8) is followed from the above inequality and (3.6b) directly. $\Box$

Note that (3.7) holds for $u^k \in \Re^n$ while (3.8) is hold only for $u^k \in \Omega$.

4 Corrector in the contraction sense

Based on the pair of geminate ascent directions in (3.5), namely, $d(u^k, \tilde{u}^k)$ and $\beta_k F(\tilde{u}^k)$, we use the one of the following corrector forms to update the new iterate u^{k+1}:

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

or

$$\text{(Correction of PC Method-II)} \qquad u_{II}^{k+1} = P_\Omega[u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)], \tag{4.1b}$$

where $\gamma \in (0, 2)$ and ϱ_k is defined in (3.5b). Note that the same step size length is used in (4.1a) and (4.1b) even if the search directions are different. Recall that $\tilde{u}^k$ is obtained via a projection, by using the correction form (4.1b), we have to make an additional projection on Ω in the PC methods. Replacing $\gamma \varrho_k$ in (4.1b) by 1, it reduces to the update form of the extragradient method (see (1.4)).

For any solution point $u^* \in \Omega^*$, we define

$$\vartheta_I(\gamma) = \|u^k - u^*\|^2 - \|u_I^{k+1} - u^*\|^2 \tag{4.2a}$$

and

$$\vartheta_{II}(\gamma) = \|u^k - u^*\|^2 - \|u_{II}^{k+1} - u^*\|^2, \tag{4.2b}$$

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, 7, 8].

Theorem 4.1 *For given u^k, let the general conditions* (3.5) *be satisfied. If the corrector is updated by* (4.1a) *or* (4.1b)*, then for any $u^* \in \Omega^*$ and $\gamma > 0$, we have*

$$\vartheta_I(\gamma) \geq q(\gamma), \tag{4.3}$$

and

$$\vartheta_{II}(\gamma) \geq q(\gamma) + \|u_I^{k+1} - u_{II}^{k+1}\|^2, \tag{4.4}$$

respectively, where

$$q(\gamma) = \gamma(2-\gamma)\varrho_k^2\|d(u^k, \tilde{u}^k)\|^2. \tag{4.5}$$

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

$$\begin{aligned} \vartheta_I(\gamma) &= \|u^k - u^*\|^2 - \|u^k - u^* - \gamma\varrho_k d(u^k, \tilde{u}^k)\|^2 \\ &= 2\gamma\varrho_k(u^k - u^*)^T d(u^k, \tilde{u}^k) - \gamma^2\varrho_k^2\|d(u^k, \tilde{u}^k)\|^2. \end{aligned} \tag{4.6}$$

Recalling (3.7), we obtain

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

Substituting it in (4.6) 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

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

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

$$\begin{aligned} \|u^* - u_{II}^{k+1}\|^2 \quad &\leq \quad \|u^* - (u^k - \gamma\varrho_k\beta_k F(\tilde{u}^k))\|^2 \\ &\quad -\|u^k - \gamma\varrho_k\beta_k F(\tilde{u}^k) - u_{II}^{k+1}\|^2. \end{aligned} \tag{4.7}$$

Thus,

$$\begin{aligned} \vartheta_{II}(\gamma) \quad &= \quad \|u^k - u^*\|^2 - \|u_{II}^{k+1} - u^*\|^2 \\ &\geq \quad \|u^k - u^*\|^2 - \|(u^k - u^*) - \gamma\varrho_k\beta_k F(\tilde{u}^k)\|^2 \\ &\qquad +\|(u^k - u_{II}^{k+1}) - \gamma\varrho_k\beta_k F(\tilde{u}^k)\|^2 \\ &= \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k\beta_k(u_{II}^{k+1} - u^*)^T F(\tilde{u}^k) \\ &\geq \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k\beta_k(u_{II}^{k+1} - \tilde{u}^k)^T F(\tilde{u}^k). \end{aligned} \tag{4.8}$$

The last inequality in (4.8) 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 (3.6a), we get

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

and consequently, substituting it in the right hand side of (4.8), we obtain

$$\begin{aligned} \vartheta_{II}(\gamma) \quad &\geq \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k(u_{II}^{k+1} - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &= \quad \|u^k - u_{II}^{k+1}\|^2 + 2\gamma\varrho_k(u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ &\quad\quad -2\gamma\varrho_k(u^k - u_{II}^{k+1})^T d(u^k, \tilde{u}^k). \end{aligned} \tag{4.9}$$

To the two crossed term in the right hand side of (4.9), we have (by using (3.6b))

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

and

$$\begin{aligned} &-2\gamma\varrho_k(u^k - u_{II}^{k+1})^T d(u^k, \tilde{u}^k) \\ &\quad = \quad \|(u^k - u_{II}^{k+1}) - \gamma\varrho_k d(u^k, \tilde{u}^k)\|^2 \\ &\quad\quad\quad -\|u^k - u_{II}^{k+1}\|^2 - \gamma^2\varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2, \end{aligned}$$

respectively. Substituting them in the right hand side of (4.9) and using

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

we obtain

$$\begin{aligned} \vartheta_{II}(\gamma) &\geq \gamma(2-\gamma)\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2 + \|u_I^{k+1} - u_{II}^{k+1}\|^2 \\ &= q(\gamma) + \|u_I^{k+1} - u_{II}^{k+1}\|^2, \end{aligned} \quad (4.10)$$

and the proof is complete. □

Note that $q(\gamma)$ is a quadratic function of γ, it reaches its maximum at $\gamma^* = 1$. In practice, ϱ_k is the 'optimal' step size in (4.1) and γ 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) with
$\gamma \in [1, 2)$.

From Theorem 4.1 we obtain

$$\|u^{k+1} - u^*\|^2 \leq \|u^k - u^*\|^2 - \gamma(2-\gamma)\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2. \quad (4.11)$$

Convergence result follows from (4.11) directly. Due to the property (4.11) we call the methods which use different update forms in (4.1) PC Method-I and PC Method II, respectively. Note that the assertion (4.11) is derived from the general conditions (3.5). For the PC methods using correction form (1.8) or (1.9), because $\varrho_k > \frac{1}{2}$, by using (3.6b) and (1.3b), it follows from (4.11) that

$$\|u^{k+1} - u^*\|^2 \le \|u^k - u^*\|^2 - \frac{1}{2}\gamma(2-\gamma)(1-\nu)\|u^k - \tilde{u}^k\|^2. \tag{4.12}$$

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 pp. 159 of [3])

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

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

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

Although the PC Method I uses the update form (4.1a) and it does not guarantee that

$\{u^k\}$ belongs to Ω, 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 $u^k \in \Re^n$, let the general conditions* (3.5) *be satisfied. If the new iterate u^{k+1} is updated by* (4.1a) *with any $\gamma > 0$, then we have*

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

where $q(\gamma)$ is defined in (4.5).

Proof. Because (due to (3.6a))

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

and (see (4.1a))

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

we need only to show that

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

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

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

we obtain

$$\begin{array}{rcl}(u-\tilde{u}^k)^T(u^k-u^{k+1}) & = & \frac{1}{2}\big(\|u-u^{k+1}\|^2 - \|u-u^k\|^2\big) \\ & & +\frac{1}{2}\big(\|u^k-\tilde{u}^k\|^2 - \|u^{k+1}-\tilde{u}^k\|^2\big). \end{array} \tag{5.5}$$

By using $u^{k+1} = u^k - \gamma\varrho_k d(u^k,\tilde{u}^k)$ and (3.6b), we get

$$\begin{array}{rl} & \|u^k-\tilde{u}^k\|^2 - \|u^{k+1}-\tilde{u}^k\|^2 \\ = & \|u^k-\tilde{u}^k\|^2 - \|(u^k-\tilde{u}^k) - \gamma\varrho_k d(u^k,\tilde{u}^k)\|^2 \\ = & 2\gamma\varrho_k(u^k-\tilde{u}^k)^T d(u^k,\tilde{u}^k) - \gamma^2\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2 \\ = & \gamma(2-\gamma)\varrho_k^2\|d(u^k,\tilde{u}^k)\|^2. \end{array}$$

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 Ω. 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 the general conditions* (3.5) *be satisfied. If the new iterate* u^{k+1} *is updated by* (4.1b) *with any* $\gamma > 0$*, then we have*

$$(u - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) + \frac{1}{2}\big(\|u - u^k\|^2 - \|u - u^{k+1}\|^2\big) \geq \frac{1}{2} q(\gamma), \quad \forall\, u \in \Omega, \tag{5.6}$$

where $q(\gamma)$ *is defined in* (4.5)*.*

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

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

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 (3.6a) we get

$$\begin{aligned} & (u^{k+1} - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) \\ \geq \quad & \gamma \varrho_k (u^{k+1} - \tilde{u}^k)^T d(u^k, \tilde{u}^k) \\ = \quad & \gamma \varrho_k (u^k - \tilde{u}^k)^T d(u^k, \tilde{u}^k) - \gamma \varrho_k (u^k - u^{k+1})^T d(u^k, \tilde{u}^k). \end{aligned} \tag{5.7}$$

To the first crossed term of the right hand side of (5.7), using (3.6b), we have

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

To the second crossed term of the right hand side of (5.7), using the Cauchy-Schwarz Inequality, we get

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

Substituting them in the right hand side of (5.7), we obtain

$$(u^{k+1} - \tilde{u}^k)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) \geq \frac{1}{2}\gamma(2-\gamma)\varrho_k^2 \|d(u^k, \tilde{u}^k)\|^2 - \frac{1}{2}\|u^k - u^{k+1}\|^2. \tag{5.8}$$

Now, we turn to treat of the term $(u - u^{k+1})^T \gamma \varrho_k \beta_k F(\tilde{u}^k)$. Since u^{k+1} is updated by (4.1b), u^{k+1} is the projection of $\big(u^k - \gamma \varrho_k \beta_k F(\tilde{u}^k)\big)$ on Ω, it follows from (2.1) that

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

and consequently

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

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

$$\big(u - u^{k+1}\big)^T \gamma \varrho_k \beta_k F(\tilde{u}^k) \geq \frac{1}{2}\big(\|u - u^{k+1}\|^2 - \|u - u^k\|^2\big) + \frac{1}{2}\|u^k - u^{k+1}\|^2. \tag{5.9}$$

Adding (5.8) and (5.9) 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 Lemma 5.1 and Lemma 5.2, respectively. By setting $u = u^*$ in (5.3) and (5.6), we get

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

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

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

This is just the form (4.11) in Section 4. In other words, the contraction property (4.11) of PC methods is the consequent result of Lemma 5.1 and Lemma 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

$$(u - \tilde{u}^k)^T \varrho_k \beta_k F(u) + \frac{1}{2\gamma}\|u - u^k\|^2 \geq \frac{1}{2\gamma}\|u - u^{k+1}\|^2, \quad \forall\, u \in \Omega. \tag{5.10}$$

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*

$$(\tilde{u}_t - u)^T F(u) \leq \frac{1}{2\gamma \Upsilon_t}\|u - u^0\|^2, \quad \forall u \in \Omega, \tag{5.11}$$

where

$$\tilde{u}_t = \frac{1}{\Upsilon_t}\sum_{k=0}^{t} \varrho_k \beta_k \tilde{u}^k \qquad \textit{and} \qquad \Upsilon_t = \sum_{k=0}^{t} \varrho_k \beta_k. \tag{5.12}$$

Proof. Summing the inequality (5.10) over $k = 0, \ldots, t$, we obtain

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

Using the notations of Υ_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. $\square$

For given u^k, the predictor $\tilde{u}^k$ is given by (1.3a) and the prediction step size β_k satisfies the condition (1.3b). Thus, the general conditions (3.5) are satisfied with $\varrho_k \geq c = \frac{1}{2}$. We choose (4.1a) (for the case that u^k is not necessary in Ω) or (4.1b) (for the case that $u^k \in \Omega$) to generate the new iterate u^{k+1}. Because $\varrho_k \geq \frac{1}{2}$, $\inf_{k\geq 0}\{\beta_k\} \geq \beta_L$ and $\beta_L = O(1/L)$, it follows from (5.12) that

$$\Upsilon_t \geq \frac{t+1}{2}\beta_L,$$

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

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

iterations, where $\tilde{u}_t$ is defined in (5.12) and $D = \sup\{\|u - u^0\| \,|\, u \in \mathcal{D}\}$. This convergence rate is in the ergodic sense, the statement (5.11) suggests us to take a larger parameter $\gamma \in (0, 2]$ in the correction steps of the PC methods.

6 Conclusions and Remarks

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 (3.5) 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.

Besides its own applications, the solution methods study for variational inequality is helpful for investigating the splitting contraction methods for separable convex optimization. Following we give a sketch for such developments.

6.1 变分不等式的投影收缩算法

设 $\Omega \subset \Re^n$ 是一个非空闭凸集, F 是 $\Re^n \to \Re^n$ 的一个映射. 考虑单调变分不等式:

$$u^* \in \Omega, \quad (u-u^*)^\top F(u^*) \ge 0, \quad \forall\, u \in \Omega. \tag{6.1}$$

我们说变分不等式单调, 是指其中的算子 F 满足 $(u-v)^\top (F(u)-F(v)) \ge 0$.

引理 1 设 $\mathcal{X} \subset \Re^n$ 是闭凸集, $\theta(x)$ 和 $f(x)$ 都是凸函数, 其中 $f(x)$ 可微. 记 x^* 是凸优化问题 $\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\}$ 的解. 我们有

$$x^* = \arg\min\{\theta(x)+f(x)\,|\,x\in\mathcal{X}\} \tag{6.2a}$$

的充分必要条件是

$$x^*\in\mathcal{X}, \quad \theta(x)-\theta(x^*)+(x-x^*)^\top\nabla f(x^*) \ge 0, \quad \forall\, x\in\mathcal{X}. \tag{6.2b}$$

在求解 (6.1) 的投影收缩算法中, 对给定的当前点 u^k 和 $\beta_k > 0$, 我们利用投影

$$\tilde{u}^k = P_\Omega[u^k - \beta_k F(u^k)] = \text{Arg}\min\{\tfrac{1}{2}\|u-[u^k-\beta_k F(u^k)]\|^2 \,|\, u\in\Omega\} \tag{6.3}$$

生成一个预测点 $\tilde{u}^k$. 在投影 (6.3) 中, 我们假设选取的 β_k 满足

$$\beta_k\|F(u^k)-F(\tilde{u}^k)\| \le \nu\|u^k-\tilde{u}^k\|, \quad \nu\in(0,1). \tag{6.4}$$

由于 $\tilde{u}^k$ 是极小化问题 $\min\{\frac{1}{2}\|u - [u^k - \beta_k F(u^k)]\|^2 \mid u \in \Omega\}$ 的解, 根据引理 6.1 有

$$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^\top \{\tilde{u}^k - [u^k - \beta_k F(u^k)]\} \geq 0, \quad \forall u \in \Omega.$$

上式两边都加上 $(u - \tilde{u}^k)^\top d(u^k, \tilde{u}^k)$, 其中

$$d(u^k, \tilde{u}^k) = (u^k - \tilde{u}^k) - \beta_k [F(u^k) - F(\tilde{u}^k)], \tag{6.5}$$

就有我们需要的预测公式

[预测] $$\tilde{u}^k \in \Omega, \quad (u - \tilde{u}^k)^\top \beta_k F(\tilde{u}^k) \geq (u - \tilde{u}^k)^\top d(u^k, \tilde{u}^k), \quad \forall u \in \Omega. \tag{6.6}$$

将 (6.6) 中任意的 $u \in \Omega$ 选成 u^*, 就得到

$$(\tilde{u}^k - u^*)^\top d(u^k, \tilde{u}^k) \geq \beta_k (\tilde{u}^k - u^*)^\top F(\tilde{u}^k). \tag{6.7}$$

由单调性, $(\tilde{u}^k - u^*)^\top F(\tilde{u}^k) \geq (\tilde{u}^k - u^*)^\top F(u^*) \geq 0$, (6.7) 的右端非负, 随后得到

$$(u^k - u^*)^\top d(u^k, \tilde{u}^k) \geq (u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k). \tag{6.8}$$

由 $d(u^k, \tilde{u}^k)$ 的表达式 (6.5) 和假设 (6.4), 利用 Cauchy-Schwarz 不等式便可得到

$$(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \geq (1 - \nu)\|u^k - \tilde{u}^k\|^2. \tag{6.9}$$

因此，不等式 (6.8) 的右端为正. 这表示, 对任何正定矩阵 $H \in \Re^{n \times n}$, $H^{-1} d(u^k, \tilde{u}^k)$ 是未知函数 $\frac{1}{2}\|u - u^*\|_H^2$ 在 u^k 处 H-模意义下的一个上升方向. 我们用

[校正] $$u^{k+1} = u^k - \alpha H^{-1} d(u^k, \tilde{u}^k) \tag{6.10}$$

产生新的迭代点, 其中 $d(u^k, \tilde{u}^k)$ 由 (6.5) 给出. 考察与 α 相关的距离平方缩短量,

$$\vartheta_k(\alpha) := \|u^k - u^*\|_H^2 - \|u^{k+1} - u^*\|_H^2. \tag{6.11}$$

根据定义

$$\begin{aligned}
\vartheta_k(\alpha) &= \|u^k - u^*\|_H^2 - \|u^k - u^* - \alpha H^{-1} d(u^k, \tilde{u}^k)\|_H^2 \\
&= 2\alpha(u^k - u^*)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|H^{-1} d(u^k, \tilde{u}^k)\|_H^2 \\
\text{由(6.8)} \quad &\geq 2\alpha(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) - \alpha^2 \|H^{-1} d(u^k, \tilde{u}^k)\|_H^2 := q_k(\alpha)
\end{aligned} \tag{6.12}$$

上式说明 $q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 的一个下界函数. 在投影收缩算法中, 我们往往考虑在欧氏模下收缩, 即 H 为单位矩阵. 在此情况下,

$$\alpha_k^* = \text{argmax}\{q_k(\alpha)\} = \frac{(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k)}{\|d(u^k, \tilde{u}^k)\|^2}. \tag{6.13}$$

由假设 (6.4), 可得 $2(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) > \|d(u^k, \tilde{u}^k)\|^2$, 因而 $\alpha_k^* > \frac{1}{2}$. 接着就有

$$\begin{aligned}
&\|u^k - u^*\|^2 - \|u^{k+1} - u^*\|^2 \\
\geq \quad & q(\alpha_k^*) = \alpha_k^*(u^k - \tilde{u}^k)^\top d(u^k, \tilde{u}^k) \geq \tfrac{1}{2}(1-\nu)\|u^k - \tilde{u}^k\|^2.
\end{aligned} \tag{6.14}$$

这是证明算法收敛的关键不等式.

✠ 在投影收缩算法中, 分处不等式 (6.6) 两端的

$$\beta_k F(\tilde{u}^k) \quad 和 \quad d(u^k, \tilde{u}^k), \qquad 我们称之为一对孪生方向.$$

奇妙的是, 采用孪生方向, 相同步长的两种不同校正方法, 具有相同的收敛性质.

在实际计算中, 我们采用校正公式

$$(收缩算法\text{–}1) \quad u_I^{k+1} = u^k - \gamma\alpha_k^* d(u^k, \tilde{u}^k) \tag{6.15}$$

或者

$$(收缩算法\text{–}2) \quad u_{II}^{k+1} = P_\Omega[u^k - \gamma\alpha_k^* \beta_k F(\tilde{u}^k)] \tag{6.16}$$

产生新的迭代点 u^{k+1}, 其中 $\gamma \in (1.5, 1.8) \subset (0, 2)$, α_k^* 都由 (6.13) 给出.

证明在《高等学校计算数学学报》38卷 1, (2016) 74-96 或我主页的报告1中可以找到.

6.2 凸优化的分裂收缩算法

我们以两个可分离算子的线性约束凸优化问题

$$\min\{\theta_1(x)+\theta_2(y)\mid Ax+By=b,\ x\in\mathcal{X},y\in\mathcal{Y}\}. \tag{6.17}$$

为例. 这个问题的拉格朗日函数是定义在 $\mathcal{X}\times\mathcal{Y}\times\Re^m$ 上的

$$L(x,y,\lambda)=\theta_1(x)+\theta_2(y)-\lambda^\top(Ax+By-b). \tag{6.18}$$

一对 $\big((x^*,y^*),\lambda^*\big)$ 满足

$$L_{\lambda\in\Re^m}(x^*,y^*,\lambda)\le L(x^*,y^*,\lambda^*)\le L_{x\in\mathcal{X},y\in\mathcal{Y}}(x,y,\lambda^*)$$

就被称为拉格朗日函数的鞍点. 换句话说, 鞍点就是同时满足

$$\left\{\begin{array}{l} x^*=\arg\min\{L(x,y^*,\lambda^*)\,|\,x\in\mathcal{X}\},\\ y^*=\arg\min\{L(x^*,y,\lambda^*)\,|\,y\in\mathcal{Y}\},\\ \lambda^*=\arg\max\{L(x^*,y^*,\lambda)\,|\,\lambda\in\Re^m\}.\end{array}\right.$$

根据 (6.2), 上述优化子问题的最优性条件是

$$\left\{\begin{array}{lrl} x^*\in\mathcal{X}, & \theta_1(x)-\theta_1(x^*)+(x-x^*)^\top(-A^\top\lambda^*)\ge 0, & \forall\, x\in\mathcal{X},\\ y^*\in\mathcal{Y}, & \theta_2(y)-\theta_2(y^*)+(y-y^*)^\top(-B^\top\lambda^*)\ge 0, & \forall\, y\in\mathcal{Y},\\ \lambda^*\in\Re^m, & (\lambda-\lambda^*)^\top(Ax^*+By^*-b)\ge 0, & \forall\,\lambda\in\Re^m.\end{array}\right. \tag{6.19}$$

换句话说, 拉格朗日函数的鞍点 $((x^*, y^*), \lambda^*)$ 可以表示成以下变分不等式的解:

$$w^* \in \Omega, \quad \theta(u) - \theta(u^*) + (w - w^*)^\top F(w^*) \geq 0, \quad \forall\, w \in \Omega. \tag{6.20}$$

其中 $\theta(u) = \theta_1(x) + \theta_2(y)$,

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \quad w = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \quad F(w) = \begin{pmatrix} -A^\top \lambda \\ -B^\top \lambda \\ Ax + By - b \end{pmatrix}. \tag{6.21}$$

$\Omega = \mathcal{X} \times \mathcal{Y} \times \Re^m$. 注意到, (6.21) 中的 $F(w)$, 恰有 $(w - \tilde{w})^\top (F(w) - F(\tilde{w})) = 0$, 所以也是单调的. 求解 (6.20) 这样的问题, 许多方法都属于一个预测-校正方法框架.

[预测]. 对给定的 v^k (v 是向量 w 的部分分量), 求得预测点 $\tilde{w}^k \in \Omega$, 使其满足

$$\theta(u) - \theta(\tilde{u}^k) + (w - \tilde{w}^k)^\top F(\tilde{w}^k) \geq (v - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k), \; \forall w \in \Omega, \tag{6.22}$$

其中矩阵 Q 不一定对称, 但是 $Q^\top + Q$ 正定.

将 (6.22) 中任意的 $w \in \Omega$ 选成 w^*, 就得到

$$(\tilde{v}^k - v^*)^\top Q(v^k - \tilde{u}^k) \geq \theta(\tilde{u}^k) - \theta(u^*) + (\tilde{w}^k - w^*)^\top F(\tilde{w}^k). \tag{6.23}$$

由 $\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^\top F(\tilde{w}^k)=\theta(\tilde{u}^k)-\theta(u^*)+(\tilde{w}^k-w^*)^\top F(w^*)\ge 0$. 因而 (6.23) 的右端非负, $(\tilde{v}^k-v^*)^\top Q(v^k-\tilde{u}^k)\ge 0$. 随后得到

$$(v^k-v^*)^\top Q(v^k-\tilde{v}^k)\ge (v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k). \tag{6.24}$$

不等式 (6.24) 和 Q 正定告诉我们, $H^{-1}Q(v^k-\tilde{v}^k)$ 是未知函数 $\frac{1}{2}\|v-v^*\|_H^2$ 在 v^k 处 H-模意义下的一个上升方向. 我们用

$$v^{k+1}=v^k-\alpha H^{-1}Q(v^k-\tilde{v}^k) \tag{6.25}$$

产生新的迭代点. 记 $M=H^{-1}Q$, 校正公式就可以写成

[校正]. 新的核心变量 v^{k+1} 由

$$v^{k+1}=v^k-\alpha M(v^k-\tilde{v}^k). \tag{6.26}$$

考察与 α 相关的距离平方缩短量,

$$\vartheta_k(\alpha):=\|v^k-v^*\|_H^2-\|v^{k+1}-v^*\|_H^2. \tag{6.27}$$

根据定义

$$\begin{aligned}\vartheta_k(\alpha) &= \|v^k-v^*\|_H^2-\|v^k-v^*-\alpha M(v^k-\tilde{v}^k)\|_H^2\\ &= 2\alpha(v^k-v^*)^\top Q(v^k-\tilde{v}^k)-\alpha^2\|M(v^k-\tilde{v}^k)\|_H^2\\ &\ge 2\alpha(v^k-\tilde{v}^k)^\top Q(v^k-\tilde{v}^k)-\alpha^2\|M(v^k-\tilde{v}^k)\|_H^2:=q_k(\alpha)\end{aligned} \tag{6.28}$$

$q_k(\alpha)$ 是 $\vartheta_k(\alpha)$ 的一个下界函数. 使 $q_k(\alpha)$ 达到极大的 α_k^* 是

$$\alpha_k^* = \text{argmax}\{q_k(\alpha)\} = \frac{(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k)}{\|M(v^k - \tilde{v}^k)\|_H^2}. \tag{6.29}$$

更进一步, 如果矩阵

$$G = Q^\top + Q - M^\top HM \succ 0. \tag{6.30}$$

容易证明 $2(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k) > \|M(v^k - \tilde{v}^k)\|_H^2$, 因而 $\alpha_k^* > \frac{1}{2}$. 接着就有

$$\begin{aligned} &\|v^k - v^*\|_H^2 - \|v^{k+1} - v^*\|_H^2 \\ &\geq \quad q(\alpha_k^*) \;=\; \alpha_k^*(v^k - \tilde{v}^k)^\top Q(v^k - \tilde{v}^k) \;\geq\; \frac{1}{4}\|v^k - \tilde{v}^k\|_{(Q^T+Q)}^2. \end{aligned} \tag{6.31}$$

另外, 在条件 (6.30) 满足的情况下, 可以在 (6.26) 中取步长 $\alpha = 1$, 由校正公式

$$v^{k+1} = v^k - M(v^k - \tilde{v}^k)$$

生成序列 $\{v^k\}$. 由 (6.28), $q(1) = \|v^k - \tilde{v}^k\|_G^2$. 因此, $\{v^k\}$ 有收缩性质

$$\|v^{k+1} - v^*\|_H^2 \leq \|v^k - v^*\|_H^2 - \|v^k - \tilde{v}^k\|_G^2.$$

✠ 从 变分不等式的投影收缩算法 到 凸优化的分裂收缩算法, 一条主线, 一个模式！ ✠

References

[1] E. Blum and W. Oettli, *Mathematische Optimierung: Grundlagen und Verfahren. Ökonometrie und Unternehmensforschung*. Berlin-Heidelberg-New York, Springer-Verlag, 1975.

[2] X.J. Cai, G.Y. Gu and B. S. He, On the O(1/t) convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators, *Comput. Optim. Appl.,* **57**, pp. 339-363, 2013.

[3] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Vol. I and II*. Springer Series in Operations Research. Springer Verlag, New York, 2003.

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

[5] B.S. He, A class of projection and contraction methods for monotone variational inequalities, *Applied Mathematics and optimization* **35**, pp. 69–76, 1997.

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

[7] B.S. He and M.-H. Xu, A general framework of contraction methods for monotone variational inequalities, *Pacific J. Optimization* **4**, pp. 195-212, 2008.

[8] B.S. He, X.M. Yuan and J.J.Z. Zhang, Comparison of two kinds of prediction-correction methods for monotone variational inequalities, *Computational Optimization and Applications* **27**, pp. 247-267, 2004.

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

[10] A. Nemirovski, 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. Optimization* **15**, pp. 229-251, 2005.

[11] Y. Pan and L. Pavel, Games with coupled propagated constraints in optical networks with multi-link topologies, *Automatica* **45**, pp. 871-880, 2009.

[12] F. Sha, Large Margin Training of Acoustic Models for Speech Recognition, PhD Thesis, Computer and Information Science, University of Pennsylvania, 2007.

[13] M.V. Solodov and P. Tseng, Modified projection-type methods for monotone variational inequalities, *SIAM J. Control and Optimization* **34**, pp. 1814–1830, 1996.

[14] K. Taji, M. Fukushima and I. Ibaraki, A globally convergent Newton mmethod for solving strongly monotone variational inequalities. *Mathematical Programming* **58**, pp. 369-383, 1993.

[15] B. Taskar, S. Lacoste-Julien and M. I. Jordan. Structured prediction via Extragradient method, In Y. Weiss and B. Schoelkopf and J. Platt (Eds.), *Advances in Neural Information Processing Systems (NIPS)* **18**, 2006.

[16] P. Tseng, On accelerated proximal gradient methods for convex-concave optimization, Department of Mathematics, University of Washington, Seattle, WA 98195, USA, 2008.