关于两个非线性算子的一个分裂方法 *

董云达
(郑州大学数学系, 郑州 450001)

A SPLITTING METHOD FOR TWO NONLINEAR OPERATORS

Dong Yunda
(Department of Mathematics, Zhengzhou University, Zhengzhou 450001)

Abstract This paper generalizes a class of projection type methods for monotone variational inequalities to general monotone inclusion. It is shown that when the normal cone operator in projection is replaced by any maximal monotone operator, the resulting method inherits all attractive convergence properties of projection type methods, and allows an adjusting step size rule. Weaker convergence assumption entails an extra projection at each iteration. Moreover, this paper also addresses applications of the resulting method to convex programs and monotone variational inequalities.

Key words monotone inclusion, splitting method, monotone variational inequalities, projection method, weak convergence.
AMS(2000) subject classifications 90C33, 49J40
中图法分类号 O221. 2

1 Introduction

This paper considers the problem of finding an x in a Hilbert space H such that

$$0 \in F(x) + B(x), \tag{1}$$

*收稿日期: 2006-04-12.

where F is a Lipschitz continuous monotone operator and B is a maximal monotone operator in H.

It is well known that x solves (1) if and only if it solves

$$x=(I+\alpha B)^{-1}(x-\alpha F(x))$$

where α is a positive scalar. When B is the normal cone operator to some nonempty closed convex set C in R^n, it just reduces to the projection equation

$$x=P_C(x-\alpha F(x))$$

where P_C denotes the orthogonal projection.

The goal of this paper is to generalize a class of efficient methods for solving the projection equation above [4,5] to the case of the problem (1). Specifically, this goal is to show that when the normal operator in projection is replaced by any maximal monotone operator, the resulting method has globally weak convergence provided that the solution set of the problem (1), say $T^{-1}(0)$, is nonempty, and F is additionally Lipschitz continuous in H. Yet, for possibly weaker convergence assumptions, the version presented in this paper eventually does a projection to arrive at the next iterate (see (4)). So, the Lipschitz continuity for convergence is assumed not in the whole space H but in the union of the effective domain of B and the projection set X, where X may be any closed convex set such that $X\cap T^{-1}(0)\neq\emptyset$.

Compared with known method [6] for solving the problem (1), the proposed method shares all its nice convergence properties. It entails neither the strong monotonicity required in the forward-backward splitting method [3] nor the inversion operation of $I+\alpha F$ in the Douglas-Rachford splitting method [3], which is in general computationally expensive. Especially, the proposed method includes a more practical step size rule than [6]. This rule may be partly viewed as a generalized version of the Armijo type rule [4], with an improvement: The resulting step size can be re-adjusted to better extent once it is chosen conservatively. In contrast, [4,5,6] typically require the corresponding step size sequences to decrease monotonically. For closely related discussions, we refer to [2,1].

Here we would like to point out that our analysis does not simply follow those in [5,4]. This is because that their analysis is based on a known projection property [7]

$$\langle x-P_C(x),\, P_C(x)-y\rangle\geq 0,\quad \forall x\in R^n,\ \forall y\in C,$$

whereas its generalized version

$$\langle x-(I+B)^{-1}(x),\, (I+B)^{-1}(x)-y\rangle\geq 0,\quad \forall x\in R^n,\ \forall y\in \mathrm{dom}B,$$

is, in general, not still in force. To see this, we consider the $H=R^1$ case. Let $B(x)=x$. Then $\mathrm{dom}B=R^1$ and $(I+B)^{-1}(x)=x/2$. Taking $x=2, y=2\in\mathrm{dom}B$ yields

$$\langle x-(I+B)^{-1}(x),\, (I+B)^{-1}(x)-y\rangle=(2-1)\times(1-2)=-1<0.$$

This, in turn, implies that our results are not attainable by simply following their line of analysis [4,5].

NOTATION. $\langle x,y\rangle$ and $\|x\| = \sqrt{\langle x,x\rangle}$ denote the usual inner product and induced norm for $x,y \in H$. A multi-valued mapping T is called monotone if

$$\langle x-x',\, y-y'\rangle \geq 0,\ \forall y \in T(x),\ \forall y' \in T(x'),$$

and is called maximal monotone if in addition its graph can not be properly contained in the graph of any other monotone mapping on H. A continuous mapping F is called Lipschitz continuous on D if there exists $l > 0$ such that

$$\|F(x) - F(x')\| \leq l\,\|x-x'\|,\ \forall x,x' \in D.$$

T^{-1} denotes the inverse of a multi-valued mapping T, and $\mathrm{dom}T := \{x \mid T(x) \neq \emptyset\}$. Furthermore, the following notation is used frequently

$$J(x,\alpha) = (I+\alpha B)^{-1}(x - \alpha F(x)).$$

2 Method and its convergence

This section describes the proposed method and proves its globally weak convergence.

Now the proposed method is formally described as follows.

Algorithm 1

Step 0. Choose: $\alpha_{-1} \in (0,+\infty)$, a non-increasing nonnegative summable sequence $\{\delta_k\}$, $\rho \in (0,1)$, $\beta \in (0,1)$, $\theta \in (0,2)$. Choose a closed convex set such that $X \cap T^{-1}(0) \neq \emptyset$. And choose any starting point $x^0 \in H$. Set $k := 0$.

Step 1 Let α_k be the maximum of the values $\alpha_{k-1}(1+\delta_k)$, α_{k-1}, $\alpha_{k-1}\beta$, $\alpha_{k-1}\beta^2, \cdots$ such that

$$\alpha_k\langle x^k - J(x^k,\alpha_k),\, F(x^k) - F(J(x^k,\alpha_k))\rangle \leq (1-\rho)\|x^k - J(x^k,\alpha_k)\|^2. \tag{2}$$

If $x^k = J(x^k,\alpha_k)$, then stop. Otherwise

Step 2 Compute

$$\gamma_k = \frac{\rho\,\theta\,\|x^k - J(x^k,\alpha_k)\|^2}{\|x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\|^2}. \tag{3}$$

Set

$$x^{k+1} := P_X[x^k - \gamma_k(x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k)))], \tag{4}$$

and $k := k+1$. Go to Step 1.

For δ_k, a practical way is to take $\delta_k = \sigma^k$, $k = 0,1,2,\cdots$, where $\sigma \in (0,1)$ is some constant close to 1. So, the sequence $\{\delta_k\}$ is summable and decreases *slowly* so to keep better adjusting property as much as possible.

For θ, a good choice is to take $\theta \in (1,2)$. It has already been experimentally confirmed [4] that in general such choice is beneficial to practical performance.

For X, a useful way is to choose it to be bounded, especially if dom$B$ is bounded then X should be taken to be a bounded box or ball containing dom$B$ (if possible) so to do a projection trivially. The advantage of doing so is to weaken convergence assumption. This is because that, in this context, it is practically mild to assume F to be Lipschitz continuous on the compact convex subset X (i. e. , dom$B \cup X$). Other choices can be found in [6].

It looks interesting to compare the proposed method with the modified forward-backward splitting method [6]. The iterative formulae of the latter are

$$\begin{aligned} J(x^k,\alpha_k) &= (I+\alpha_k B)^{-1}(x^k-\alpha_k F(x^k)) \\ x^{k+1} &= P_X[J(x^k,\alpha_k)-\alpha_k F(x^k)+\alpha_k F(J(x^k,\alpha_k))]. \end{aligned}$$

This two methods have close relations to each other. For example, if F is Lipschitz continuous in H and $X = H$, then the two methods coincide whenever $\gamma_k = 1$. Therefore, there is no surprise that both have similar convergence properties. On the other hand, there are some improvements in this paper. For example, the step size rules presented here have good adjusting property absent in [6], and allows for possibly less conservative α_k (see Example 1 below) and an adaptive parameter γ_k. Therefore, these can be viewed as an improved versions of [6].

Below we give the convergence results on the proposed method.

Theorem 2.1 Suppose that, in Algorithm 1, F is additionally Lipschitz continuous on dom$B \cup X$. Then the following hold

(a) α_k is well defined. Furthermore, there exist two positive scalars $\bar{\alpha}, \tilde{\alpha}$ such that $\{\alpha_k\} \subset [\bar{\alpha}, \tilde{\alpha}]$.

(b) any resulting sequence $\{x^k\}$ is weakly convergent.

Proof. The proof of the part (a). In fact, possible values of α_k fall into three cases. (i) if $\alpha_k = \alpha_{k-1}(1+\delta_k)$, then $\alpha_k \geq \alpha_{k-1}$. (ii) if $\alpha_k = \alpha_{k-1}$, then $\alpha_{k-1}(1+\delta_k)$ fails to hold in (2). The Lipschitz continuity of F implies $\alpha_{k-1}(1+\delta_k)l > (1-\rho)$. So, it is immediate that $\alpha_{k-1} > (1-\rho)/(l+\delta_k l) \geq (1-\rho)/(l+\delta_0 l)$, thus, $\alpha_k > (1-\rho)/(l+\delta_0 l)$. (iii) if $\alpha_k = \alpha_{k-1}\beta^i\ (i = 1,2,\cdots)$, then $\alpha_{k-1}\beta^{i-1}$ (i. e. , α_k/β) fails to hold in (2), so it follows from the Lipschitz continuity of F that $l\alpha_k/\beta > (1-\rho)$. Thus $\alpha_k > (1-\rho)\beta/l$. The proceeding analysis, together with an induction argument, implies that α_k is well defined and

$$\alpha_k \geq \min\{\alpha_{-1}, (1-\rho)/(l+\delta_0 l), (1-\rho)\beta/l\} > 0.$$

On the other hand, α_k is bounded above. Denote $\delta = \prod_{i=0}^{\infty}(1+\delta_i)$. Then, δ is a finite value for δ_i is summable. So, it follows from step size rule that $\alpha_k \leq \alpha_{-1}\prod_{i=0}^{k}(1+\delta_i) \leq \alpha_{-1}\delta$. Take $\bar{\alpha} = \min\{\alpha_{-1}, (1-\rho)/(l+\delta_0 l), (1-\rho)\beta/l\}$, $\tilde{\alpha} = \alpha_{-1}\delta$. Then the claim holds.

The proof of the part (b). Clearly, $J(x^k,\alpha_k) = (I+\alpha_k B)^{-1}(x^k-\alpha_k F(x^k))$ is equivalent to saying

$$J(x^k,\alpha_k)+\alpha_k B(J(x^k,\alpha_k)) \ni x^k-\alpha_k F(x^k),$$

which may be rewritten as

$$\alpha_k^{-1}(x^k - J(x^k,\alpha_k)) - F(x^k) \in B(J(x^k,\alpha_k)). \tag{5}$$

Take $x^* \in X \cap T^{-1}(0)$. Then

$$-F(x^*) \in B(x^*).$$

So, it follows from monotonicity of B that

$$0 \le \langle J(x^k,\alpha_k) - x^*, \alpha_k^{-1}(x^k - J(x^k,\alpha_k)) - F(x^k) + F(x^*)\rangle,$$

which, together with $\alpha_k \ge \bar{\alpha} > 0$ and monotonicity of F, implies

$$\begin{aligned} 0 &\le \langle J(x^k,\alpha_k) - x^*, x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^*)\rangle \\ &= \langle J(x^k,\alpha_k) - x^*, x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\rangle \\ &\quad -\alpha_k \langle J(x^k,\alpha_k) - x^*, F(J(x^k,\alpha_k)) - F(x^*)\rangle \\ &\le \langle J(x^k,\alpha_k) - x^*, x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\rangle \end{aligned} \tag{6}$$

In view of $J(x^k,\alpha_k) - x^* = x^k - x^* - (x^k - J(x^k,\alpha_k))$ and the step size rule (2), it is immediate that

$$\begin{aligned} & \langle x^k - x^*, x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\rangle \\ \ge\; & \langle x^k - J(x^k,\alpha_k), x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\rangle \\ =\; & \|x^k - J(x^k,\alpha_k)\|^2 - \alpha_k \langle x^k - J(x^k,\alpha_k), F(x^k) - F(J(x^k,\alpha_k))\rangle \\ \ge\; & \rho \|x^k - J(x^k,\alpha_k)\|^2. \end{aligned} \tag{7}$$

So, it follows from (4) that

$$\begin{aligned} & \|x^{k+1} - x^*\|^2 \\ =\; & \|P_X(x^k - \gamma_k(x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k)))) - P_X(x^*)\|^2 \\ \le\; & \|x^k - x^* - \gamma_k(x^k - J(x^k,\alpha_k) - \lambda F(x^k) + \lambda F(J(x^k,\alpha_k))\|^2 \\ =\; & \|x^k - x^*\|^2 - 2\gamma_k \langle x^k - x^*, x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\rangle \\ & +\gamma_k^2 \|x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\|^2 \\ \le\; & \|x^k - x^*\|^2 - 2\gamma_k \rho \|x^k - J(x^k,\alpha_k)\|^2 \\ & +\gamma_k^2 \|x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\|^2 \\ =\; & \|x^k - x^*\|^2 - \theta(2-\theta)\rho^2 \|x^k - J(x^k,\alpha_k)\|^4 \\ & \cdot \|x^k - J(x^k,\alpha_k) - \alpha_k F(x^k) + \alpha_k F(J(x^k,\alpha_k))\|^{-2}. \end{aligned} \tag{8}$$

The relation that $x^* = P_X(x^*)$ comes from $x^* \in X$, and the first inequality follows from the non-expansive property [7] of the projection, the second inequality from (7), and the last equality from (3).

The inequality (8) shows that $\{\|x^k - x^*\|\}$ is bounded. In addition, $\{\alpha_k\}$ is also bounded. Thus, $\{(x^k,\alpha_k)\}$ must be bounded and at least has one weak cluster point, say (x^∞,α_∞). Let $\{(x^{k(j)},\alpha_{k(j)})\}$

be some subsequence converging weakly to $(x^\infty, \alpha_\infty)$. Then, $\{J(x^{k(j)}, \alpha_{k(j)})\}$ converges weakly to x^∞ in that $\{x^k - J(x^k, \alpha_k)\}$ converges strongly to zero. Furthermore, it can be seen that $J(x^{k(j)}) \in \mathrm{dom}B$ is bounded in norm for all $k(j)$.

On the other hand, it follows from (5) and the monotonicity of B that for all $x \in \mathrm{dom}B$, $\omega \in B(x)$ the following holds

$$\langle \alpha_k^{-1}(x^k - J(x^k, \alpha_k)) - F(x^k) - \omega, J(x^k, \alpha_k) - x\rangle \geq 0.$$

So, taking the limit along $k(j)$ yields

$$\langle -F(x^\infty) - \omega, x^\infty - x\rangle \geq 0.$$

This is equivalent to saying $-F(x^\infty) \in B(x^\infty)$ since B is maximal. Thus, x^∞ is a solution of the problem (1).

The proof of the uniqueness of the weak cluster point is standard (cf. [6]) and thus is omitted.

3 Applications

The section addresses applications of the proposed method to convex programs and monotone variational inequalities.

Example 1 Consider the convex program problem:

$$\min\{f(x) + g(y) \mid Ax - y = 0\}$$

where f, g are proper closed convex functions on R^n and R^m, respectively. A is any given $m \times n$ matrix. Denote by A^T the transpose of A. Under suitable condition, it is equivalent to solving

$$0 \in F(x,y,\lambda) + B(x,y,\lambda),$$

where

$$F(x,y,\lambda) = (A^T\lambda, -\lambda, -Ax + y), \quad B(x,y,\lambda) = \partial f(x) \times \partial g(y) \times \{0\}.$$

Clearly, F is Lipschitz continuous in R^{n+2m} with the constant

$$l = \sqrt{\|A^T\|^2 + \|-A\|^2 + \|-I\|^2 + \|I\|^2},$$

where $\|A^T\|$ and $\|A\|$ denote the usual matrix norms of A^T and A, respectively. So, in this setting, the proposed method may be expressed as

$$\begin{aligned}
\bar{x}^k &= \arg\min\{\alpha_k f(x) + \frac{1}{2}\|x - x^k + \alpha_k A^T\lambda^k\|^2 \mid x \in R^n\} \\
\bar{y}^k &= \arg\min\{\alpha_k g(y) + \frac{1}{2}\|y - y^k - \alpha_k\lambda^k\|^2 \mid y \in R^m\} \\
\bar{\lambda}^k &= \lambda^k - \alpha_k(Ax^k - y^k) \\
x^{k+1} &= x^k - \gamma_k(x^k - \bar{x}^k - \alpha_k A^T(\lambda^k - \bar{\lambda}^k)) \\
y^{k+1} &= y^k - \gamma_k(y^k - \bar{y}^k + \alpha_k(\lambda^k - \bar{\lambda}^k)) \\
\lambda^{k+1} &= \lambda^k - \gamma_k(\lambda^k - \bar{\lambda}^k - \alpha_k(A\bar{x}^k - Ax^k + y^k - \bar{y}^k)).
\end{aligned}$$

It can be easily seen that α_k may be properly chosen in $(0,+\infty)$. In contrast, the step size in [6] is possibly forced to be close to $1/l$. So, it may be conservative whenever $\|A^T\|$ and/or $\|A\|$ are large.

Example 2 Consider the projection equation corresponding to monotone variational inequalities

$$x = P_C(x-\alpha F(x)),$$

where C is a nonempty closed convex subset of R^n. In this setting, the proposed method may be expressed as

$$\begin{aligned} \bar{x}^k &= P_C(x^k-\alpha_k F(x^k)) \\ x^{k+1} &= P_X[x^k-\gamma_k(x^k-\bar{x}^k-\alpha_k F(x^k)+\alpha_k F(\bar{x}^k))]. \end{aligned}$$

If F is Lipschitz continuous on $C\cup X$ then the resulting sequence is convergent. If F is Lipschitz continuous on the whole space R^n then taking $X=R^n$ yields

$$\begin{aligned} \bar{x}^k &= P_C(x^k-\alpha_k F(x^k)) \\ x^{k+1} &= x^k-\gamma_k(x^k-\bar{x}^k-\alpha_k F(x^k)+\alpha_k F(\bar{x}^k)). \end{aligned}$$

In this context, it coincides with the projection type·method in [4]. So, it is convergent.

参 考 文 献

1 Dong Y D. An LS-free splitting method for composite mappings. Appl. Math. Letters, 2005, 18: 843-848

2 He B S. Solving a class of linear projection equations. Numer. Math., 1994, 68: 71-80

3 Lions P and Mercier B. Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal., 1979, 16: 964-979

4 Solodov M. and Tseng P. Modified projection-type methods for monotone variational inequalities. SIAM J. Control Optim., 1996, 34: 1814-1830

5 Sun D F. A class of iterative methods for solving nonlinear projection equations. J. Optim. Theory Appl., 1996, 91: 123-140

6 Tseng P. A modified forward-backward splitting method for maximal monotone mapping. SIAM J. Control Optim., 2000, 38: 431-446

7 Zarantonello E. Projections on Convex Sets in Hilbert Space and Spectral Theory. In Zarantonello, E. H., Contributions to Nonlinear Functional Analysis, 237-424. Academic Press, New York, 1971