Applied Mathematics and Computation 237 (2014) 60–68

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

New properties of forward–backward splitting and a practical proximal-descent algorithm

Yuanyuan Huang [a], Yunda Dong [b,*]

[a] School of Mathematics and Statistics, Xidian University, Xi'an 710126, PR China
[b] School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, PR China

ARTICLE INFO

Keywords:
Monotone operator
Lipschitz continuity
Forward–backward splitting
Proximal-descent algorithm

ABSTRACT

In this paper, we discuss a proximal-descent algorithm for finding a zero of the sum of two maximal monotone operators in a real Hilbert space. Some new properties of forward–backward splitting are given, which extend the well-known properties of the usual projection. Then, they are used to analyze the weak convergence of the proximal-descent algorithm without assuming Lipschitz continuity of the forward operator. We also give a new technique of choosing trial values of the step length involved in an Armijo-like condition, which returns the (not necessarily decreasing) step length self-adaptively. Rudimentary numerical experiments show that it is effective in practical implementations.

1. Introduction

Let $\mathcal{H}$ be a real infinite-dimensional Hilbert space with usual inner product $\langle x, y\rangle$ and induced norm $\|x\| = \sqrt{\langle x, x\rangle}$ for $x, y \in \mathcal{H}$. We consider the problem of finding an $x \in \mathcal{H}$ such that

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

where $F : \mathcal{H} \to 2^{\mathcal{H}}$ is a continuous monotone operator in the whole Hilbert space $\mathcal{H}$, and $B : \mathcal{H} \to 2^{\mathcal{H}}$ is a maximal monotone operator, with the effective domain $\mathrm{dom}B := \{x \in \mathcal{H} : B(x) \neq \emptyset\}$. This problem model covers the minimization of convex functions, computation of saddle points of convex-concave functions, solution of monotone complementarity and variational inequality problems and so on [1,2].

For the problem above, a very simple iterative procedure is the following forward–backward splitting method [3,4]:

$$x^{k+1} = (I + \alpha_k B)^{-1}(I - \alpha_k F)(x^k),$$

where I stands for the identity mapping, and $\alpha_k > 0$ is a step length. In this setting, F and B are usually called the forward operator and the backward operator, respectively. When we take the backward operator to be the normal cone operator of some nonempty closed convex set $\mathcal{C}$ in the Euclidean space, it reduces to a projection method for monotone variational inequalities [5]: $x^{k+1} = P_{\mathcal{C}}[x^k - \alpha_k F(x^k)]$, where $P_{\mathcal{C}}$ is a usual projection onto the set $\mathcal{C}$. This projection method is a direct generalization of a gradient projection method of Goldstein and of Levitin and Polyak, and see [6] for further discussions.

However, for global weak convergence, the forward–backward splitting method requires either the inverse of the forward operator be strongly monotone in $\mathcal{H}$ (cf. [7]) or the forward operator be Lipschitz continuous monotone in $\mathcal{H}$ and the sum operator $B + F$ be strongly monotone on $\mathrm{dom}B$ [7].

* Corresponding author.
E-mail addresses: yyuanhuang@126.com (Y. Huang), ydong@zzu.edu.cn (Y. Dong).

http://dx.doi.org/10.1016/j.amc.2014.03.062
0096-3003/

To weaken these restrictive convergence assumptions, Tseng [2] modified the forward–backward splitting method by adding an extra step at each iteration. More specifically, let $\mathcal{X}$ be some closed convex set intersecting the solution set of the problem (1), choose the starting point $x^0 \in \mathcal{X}$. At kth iteration, for known iterate x^k, choose $\alpha_k > 0$, then the new iterate x^{k+1} is given by

$$x^k(\alpha_k) := (I + \alpha_k B)^{-1}(I - \alpha_k F)(x^k), \tag{2}$$

$$x^{k+1} := P_{\mathcal{X}}[x^k(\alpha_k) - \alpha_k F(x^k(\alpha_k)) + \alpha_k F(x^k)]. \tag{3}$$

This iterative scheme described by (2) and (3) is sometimes called Tseng's splitting algorithm. When dom$B = \mathcal{H} = \mathcal{X}$, it can be viewed as an instance of the HPE algorithm proposed in [8].

As is well-known, Tseng's splitting algorithm has nice convergence properties. Its global weak convergence only requires the forward operator be (Lipschitz) continuous in $\mathcal{H}$, and the associated strong monotonicity is no longer assumed.

Subsequently, a relaxed form of Tseng's splitting algorithm was discussed in the second author's Ph.D. dissertation [9]: Choose the step length $\alpha_k > 0$ through an Armijo-like condition. Compute

$$x^k(\alpha_k) = (I + \alpha_k B)^{-1}(I - \alpha_k F)(x^k). \tag{4}$$

And compute a relaxation factor $\gamma_k > 0$. Then the new iterate is given by

$$x^{k+1} := P_{\mathcal{X}}[x^k - \gamma_k(x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k)))], \tag{5}$$

where $\mathcal{X}$ is the same set as defined in Tseng's splitting algorithm. When specialized to monotone variational inequality problems, such an iterative scheme just reduces to a projection-type method independently proposed in [10–12]. As shown in [9], it has the same nice convergence properties as Tseng's splitting algorithm. From now on, as in [13], we call it a proximal-descent algorithm for maximal monotone operators.

In this paper, our main goal is to further study the proximal-descent algorithm described by (4) and (5), and our contributions are threefold.

- Firstly, we give two new properties of forward–backward splitting (see Lemma 2 below), which are extensions of two well-known projection properties in [14,15]. Here we provide a simple and unified proof.
- Secondly, we make use of these new properties to analyze convergence behaviors of the proximal-descent algorithm for monotone operators, and prove its weak convergence beyond Lipschitz continuity of the forward operator F, with the same additional assumptions as those in [2]. The proof techniques take full advantage of these new properties of the forward–backward splitting and are obviously different from Tseng's.
- Thirdly, for the proximal-descent algorithm above, we give a new practical technique of choosing trial values of the step length involved in the Armijo-like condition, which returns the (not necessarily decreasing) step length self-adaptively. As a result, for our test problems, it needs fewer iterations and less CPU time in achieving the same medium accuracy compared to Tseng's splitting algorithm.

2. Preliminaries

In this section, we give and prove two new properties of forward–backward splitting, and they have interest in their own right.

First of all, let us review some useful definitions and concepts. Recall that $T : \mathcal{H} \to 2^{\mathcal{H}}$ is called monotone if

$$\langle s - s', x - x' \rangle \geqslant 0, \qquad \text{for all } x, x' \in \text{ dom } T,\ s \in T(x),\ s' \in T(x');$$

maximal monotone if it is monotone and its graph $\{(x, s) : x \in \mathcal{H}, s \in T(x)\}$ can not be enlarged without loss of the monotonicity. In addition, if there exists $\mu > 0$ such that $\langle s - s', x - x' \rangle \geqslant \mu\|x - x'\|^2$, for all $x, x' \in$ dom T, $s \in T(x)$, $s' \in T(x')$, then T is usually called strongly monotone. For a mapping $F : \mathcal{H} \to \mathcal{H}$, if there exists $L > 0$ such that

$$\|F(x) - F(y)\| \leqslant L\|x - y\|, \qquad \text{for all } x, y \in \mathcal{H},$$

then F is called Lipschitz continuous in $\mathcal{H}$. More on the Euclidean space $\mathcal{R}^n$. Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a closed proper convex function, then its sub-differential is defined by

$$\partial f(x) = \{s : f(y) - f(x) \geqslant \langle s, y - x \rangle, \text{ for all } y \in \mathcal{R}^n\}$$

for any given x in $\mathcal{R}^n$. Moreover, if f is further continuously differentiable, then $\partial f(x) = \{\nabla f(x)\}$, where $\nabla f(x)$ is the gradient of f at $x \in \mathcal{R}^n$. Let $\mathcal{C}$ be some nonempty closed convex set in $\mathcal{R}^n$, the usual projection is defined by $P_{\mathcal{C}}(u) = \text{argmin}\ \{\|u - x\| : x \in \mathcal{C}\}$. The associated indicator function defined by

$$\delta_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ \infty & \text{if } x \notin \mathcal{C}. \end{cases}$$