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

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

is a closed proper convex function. Furthermore, the normal cone operator $N_C := \partial\delta_C$ is maximal monotone and $P_C = (I + \alpha N_C)^{-1}$, where $\alpha > 0$. See [16–18] for more details.

Below we state two basic properties of the forward–backward splitting, which were given by the first author in her Master's thesis, Zhengzhou University, March of 2011. Note that the following proof is a refined version.

Lemma 1. *Let $B : \mathcal{H} \to 2^{\mathcal{H}}$ be maximal monotone. For any given $x, z \in \mathcal{H}$, denote*

$$r(x,\alpha) = x - (I + \alpha B)^{-1}(x + \alpha z), \quad \alpha > 0.$$

Then when $\alpha \geqslant \alpha' > 0$, the following hold

$$\|r(x,\alpha)\| \geqslant \|r(x,\alpha')\|, \qquad \frac{\|r(x,\alpha)\|}{\alpha} \leqslant \frac{\|r(x,\alpha')\|}{\alpha'}.$$

Proof. For any given positive number $\alpha > 0$ and $u \in \mathcal{H}$, it is easy to check that

$$\alpha^{-1}\left(I - (I + \alpha B)^{-1}\right)(u) \in B\left((I + \alpha B)^{-1}(u)\right).$$

By using the monotonicity of B and the relation above with $u := x + \alpha z$, we can get

$$\left\langle \frac{x + \alpha' z - (I + \alpha' B)^{-1}(x + \alpha' z)}{\alpha'} - \frac{x + \alpha z - (I + \alpha B)^{-1}(x + \alpha z)}{\alpha}, \; (I + \alpha' B)^{-1}(x + \alpha' z) - (I + \alpha B)^{-1}(x + \alpha z) \right\rangle \geqslant 0.$$

This is equivalent to saying that

$$\left\langle \frac{r(x,\alpha')}{\alpha'} - \frac{r(x,\alpha)}{\alpha}, \; r(x,\alpha) - r(x,\alpha') \right\rangle \geqslant 0, \tag{6}$$

which, together with the Cauchy–Schwarz inequality, implies

$$\frac{\|r(x,\alpha)\|^2}{\alpha} + \frac{\|r(x,\alpha')\|^2}{\alpha'} \leqslant \left(\frac{1}{\alpha} + \frac{1}{\alpha'}\right)\|r(x,\alpha)\|\|r(x,\alpha')\|.$$

If $\|r(x,\alpha')\| = 0$ then so is $\|r(x,\alpha)\|$. In this case, the assertion holds naturally. Otherwise, we must have $\|r(x,\alpha')\| \neq 0$. At this moment, we can re-express the inequality above in the following way

$$\frac{\|r(x,\alpha)\|^2}{\|r(x,\alpha')\|^2} - \left(1 + \frac{\alpha}{\alpha'}\right)\frac{\|r(x,\alpha)\|}{\|r(x,\alpha')\|} + \frac{\alpha}{\alpha'} \leqslant 0.$$

Since the associated parabola $p(t) := t^2 - (1 + \alpha/\alpha')t + \alpha/\alpha'$ opens upward (here both α and α' are given positive numbers) and the equation $p(t) = 0$ has two roots $t = 1$ and $t = \alpha/\alpha' \geqslant 1$, it follows that $p(t) \leqslant 0$ if and only if

$$1 \leqslant \frac{\|r(x,\alpha)\|}{\|r(x,\alpha')\|} \leqslant \frac{\alpha}{\alpha'}.$$

The proof is complete. □

When B is the sub-differential of the indicator function for some closed convex set in the Euclidean space $\mathcal{R}^n$, the first inequality and the second inequality in Lemma 1 reduce to two well-known projection properties in [14,15], respectively. Note that the proof techniques are in the same spirit as those due to [19], where these two projection properties in [14,15] were reproved in such a unified way. Yet, the first part of the proof up to (6) seems new and is more general and simpler than the counterpart of [19] in the setting of monotone variational inequality problems.

Lemma 2. *Let $F : \mathcal{H} \to 2^{\mathcal{H}}$ be continuous monotone in $\mathcal{H}$, and $B : \mathcal{H} \to 2^{\mathcal{H}}$ be maximal monotone on domB. Then for any given $x \in$ domB and $\alpha \geqslant \alpha' > 0$, the following hold*

$$\|(I + \alpha B)^{-1}(I - \alpha F)(x) - x\| \geqslant \|(I + \alpha' B)^{-1}(I - \alpha' F)(x) - x\|, \tag{7}$$

$$\|(I + \alpha B)^{-1}(I - \alpha F)(x) - x\|/\alpha \leqslant \|(I + \alpha' B)^{-1}(I - \alpha' F)(x) - x\|/\alpha', \tag{8}$$

$$\liminf_{\alpha \to 0} \|(I + \alpha B)^{-1}(I - \alpha F)(x) - x\|/\alpha = \min\{\|w\| : w \in B(x) + F(x)\}. \tag{9}$$

Moreover, if $B + F$ is maximal monotone on domB, then the minimum on the right-hand side must be uniquely attainable.

Proof. The inequality (7) and the inequality (8) immediately follow from Lemma 1 whereas the rest of this lemma follows from [2, Lemma 3.3]. □

Now several interesting and important properties of the forward–backward splitting have been given. For further results in this direction, we refer to [20].

3. Algorithm

In this section, we are ready to detail, step by step, the algorithm roughly described by (4) and (5). Furthermore, a comparison to Tseng's splitting algorithm is made.

Algorithm 1. *A Proximal-Descent Algorithm*

Step 0. Choose $x^0 \in \mathcal{X},\ \alpha_{-1} = 1,\ \beta, \rho \in (0,1), \theta \in (0,2)$. Set $k := 0$.
Step 1. (*step length*) Choose $\tilde{\alpha}_{k-1}$ by either of the following two ways

$$\tilde{\alpha}_{k-1} = \alpha_{k-1}, \quad \text{if } F \text{ is strongly monotone; otherwise,} \tag{10}$$

$$\tilde{\alpha}_{k-1} = \alpha_{k-1}, \quad k \leqslant K,\ \tilde{\alpha}_{k-1} = 2\alpha_{K-1},\ k > K, \tag{11}$$

where K is a prescribed natural number, and determine the step length α_k to be the largest

$$\alpha \in \{\tilde{\alpha}_{k-1}\beta^{-1}, \alpha_{k-1}\beta^i,\ i = 0, 1, \ldots\}, \quad x^k(\alpha) = (I + \alpha B)^{-1}(I - \alpha F)(x^k),$$

such that

$$\alpha\langle x^k - x^k(\alpha),\ F(x^k) - F(x^k(\alpha))\rangle \leqslant (1-\rho)\|x^k - x^k(\alpha)\|^2. \tag{12}$$

Step 2. (*Proximal step*) Use the step length α_k just determined in Step 1, and compute

$$x^k(\alpha_k) = (I + \alpha_k B)^{-1}(I - \alpha_k F)(x^k). \tag{13}$$

If $x^k(\alpha_k) = x^k$, then stop. Otherwise go to Step 3.
Step 3. (*Descent step*) First compute the relaxation factor γ_k through the following formula

$$\gamma_k = \frac{\theta\rho\|x^k - x^k(\alpha_k)\|^2}{\|x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\|^2}. \tag{14}$$

Then the new iterate is given by

$$x^{k+1} = P_{\mathcal{X}}[x^k - \gamma_k(x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k)))]. \tag{15}$$

Set $k := k + 1$, and go to Step 1.

Note that the proximal-descent algorithm allows for scaling, see [13, Section 4] for more details. As to the proximal step, it is equivalent to solving the monotone inclusion

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

When B is the gradient of a continuously differentiable convex function in $\mathcal{R}^n$, a practical PR+ conjugate gradient method only using gradient [21] can serve as an appropriate choice of solving it. Of course, if such gradient is linear, the celebrated conjugate gradient method of Hestenes and Stiefel is applied directly.

Next, by following one referee's suggestions, we discuss the range of γ_k in the descent step. In fact, it follows from the inequality (12) that

$$\begin{aligned}\rho\|x^k - x^k(\alpha_k)\|^2 &\leqslant \langle x^k - x^k(\alpha_k), x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\rangle \\ &\leqslant \|x^k - x^k(\alpha_k)\|\|x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\|.\end{aligned}$$

Then

$$\rho\|x^k - x^k(\alpha_k)\|/\|x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\| \leqslant 1.$$

Combining it with the definition of γ_k in (14) obtains $0 < \gamma_k \leqslant \theta/\rho$.

Remark 1. Here we would like to remark two points. One is that, almost in the same time, Noor [22] independently suggested a conceptual method similar to our Algorithm 1 but without our self-adaptive choice of α_k, and he proved the method's convergence in the finite-dimensional space. The other is: to make a comparison, we give the details of the ways of determining step length in Tseng's splitting algorithm. Given $\sigma > 0,\ \beta \in (0,1)$ and $\rho > 0$. At kth iteration, find α_k to be the largest

$$\alpha \in \{\sigma\beta^i,\ i = 0, 1, \ldots\}, \quad x^k(\alpha) := (I + \alpha B)^{-1}(I - \alpha F)(x^k), \tag{16}$$

such that

$$\alpha\|F(x^k) - F(x^k(\alpha))\| \leqslant (1-\rho)\|x^k - x^k(\alpha)\|. \tag{17}$$

Moreover, an alternative to the format is the following self-adaptive strategy [2, p. 434]: At kth iteration, find α_k to be the largest

$$\alpha = \max\{\alpha_{k-1}\beta^i,\ i = 0,1,\ldots\}, \quad x^k(\alpha) := (I+\alpha B)^{-1}(I-\alpha F)(x^k) \tag{18}$$

such that the inequality (17) holds.

Obviously, our proposed step length rule has an appealing property. It combines their individual advantages of these two formats for α_k described in Tseng's splitting algorithm. And a sequence $\{\alpha_k\}$ ($k \geqslant 0$) generated by our proposed step length rule is not necessarily decreasing. Thus it is more practical.

Lemma 3. *Let α_k be the step length involved in Algorithm 1. Then such step length is well-defined and can be determined within finite trials. Moreover, there must exist constant numbers $\alpha > 0$ and $\hat{\alpha} > 0$ such that the following implication relations hold*

(i) *F's Lipschitz continuity implies $\alpha_k \geqslant \alpha > 0$;*
(ii) *F's strong monotonicity implies $\alpha_k \leqslant \hat{\alpha} < +\infty$.*

Proof. First of all, we prove that the associated step length is well-defined and can be determined within finite trials. Otherwise, there exists one iteration, and for all $i = 0,1,\ldots$, the inequality (12) fails to hold, i.e., by using the notation $\bar{\alpha}_k := \alpha_{k-1}\beta^i$, we have

$$\bar{\alpha}_k\langle x^k - x^k(\bar{\alpha}_k),\ F(x^k) - F(x^k(\bar{\alpha}_k))\rangle > (1-\rho)\|x^k - x^k(\bar{\alpha}_k)\|^2.$$

It can be rewritten as

$$\left\langle \frac{x^k - x^k(\bar{\alpha}_k)}{\|x^k - x^k(\bar{\alpha}_k)\|},\ F(x^k) - F(x^k(\bar{\alpha}_k))\right\rangle > (1-\rho)\|x^k - x^k(\bar{\alpha}_k)\|/\bar{\alpha}_k, \ \ i = 0,1,\ldots. \tag{19}$$

In views of Lemma 2, we can conclude that

$$\liminf_{i\to+\infty}\|x^k(\bar{\alpha}_k) - x^k\|/\bar{\alpha}_k = \min\left\{\|w\| : w \in F(x^k) + B(x^k)\right\}. \tag{20}$$

Since x^k is not a solution to the problem (1), the closed convex set $F(x^k) + B(x^k)$ will not include the origin. Thus, by Lemma 2, the minimum on the right-hand side is uniquely attainable, and it must be some positive number. Then

$$\liminf_{i\to+\infty}\|x^k(\bar{\alpha}_k) - x^k\|/\bar{\alpha}_k > 0.$$

On the other hand, it follows $x^k(\bar{\alpha}_k) = (I+\bar{\alpha}_k B)^{-1}(I-\bar{\alpha}_k F)(x^k)$ and $\bar{\alpha}_k = \alpha_{k-1}\beta^i$ that

$$\|x^k(\bar{\alpha}_k) - x^k\| \to 0, \quad \text{as } i \to +\infty,$$

which, together with (19), the continuity of F and the boundedness of $(x^k - x^k(\bar{\alpha}_k))/\|x^k - x^k(\bar{\alpha}_k)\|$, implies

$$\liminf_{i\to+\infty}\|x^k(\bar{\alpha}_k) - x^k\|/\bar{\alpha}_k = 0,$$

giving a contradiction. So, such a step length α_k is well-defined and can be determined within finite trials.

Next, we prove the implication relation i). In fact, in view of Step 1 in Algorithm 1, the chosen step length falls into two cases. One is that α_k takes $\tilde{\alpha}_{k-1}\beta^{-1}$ or α_{k-1}. In such case, it is not difficult (but somewhat complicated) to get its positive lower bound. The other is that α_k takes $\alpha_{k-1}\beta^i$ for some $i \in \{1,2,\ldots\}$. In such case, $\alpha_k\beta^{-1}$ does not satisfy (12), i.e.,

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

Combining this with the Cauchy–Schwarz inequality and F's Lipschitz continuity yields $\alpha_k > (1-\rho)\beta/L$, where L is Lipschitz constant. Thus, the implication relation i) must hold.

Finally, we prove the implication relation ii). From (12), we have

$$\alpha_k\langle x^k - x^k(\alpha_k),\ F(x^k) - F(x^k(\alpha_k))\rangle \leqslant (1-\rho)\|x^k - x^k(\alpha_k)\|^2,$$

which, together with F's strong monotonicity, implies $\alpha_k \leqslant (1-\rho)/\mu$, where μ is the strong monotonicity modulus. $\square$

4. Convergence

In this section, we analyze the global weak convergence of Algorithm 1 even if the forward operator is not assumed to be Lipschitz continuous.

First of all, we make the following three assumptions, which are standard in the literature [2].

Assumption 1. *(i) For any $x^k \in \mathcal{X} \subset \mathrm{dom}B$ and $y^k \in \mathrm{dom}B$, if the sequences $\{x^k\}$ and $\{y^k\}$ converge weakly, respectively, and $\|x^k - y^k\| \to 0$, then $\|F(x^k) - F(y^k)\| \to 0$; (ii) The distance of the origin to the set $F(x^k) + B(x^k)$ is uniformly bounded on $\mathcal{X}$ whenever the sequence $\{x^k\}$ converges weakly; (iii) The sum operator $F + B$ is maximal monotone on domB.*

Note that the first item corresponds to the locally uniform continuity of the forward operator, which is implied by Lipschitz continuity assumption. Also, be aware of that the first item is slightly different from the one in [2] because we add the assumption on weak convergence of the sequence $\{x^k\}$. In the setting of monotone variational inequality problems that corresponds to $B := N_{\mathcal{C}}$, where $\mathcal{C}$ is some nonempty closed convex subset in $\mathcal{H}$, the assumption (i) holds automatically. As for the second one, the reader may consult [2] for further discussions. About maximal monotonicity criteria for the sum operator, we refer to [23–25] and the references therein.

To simplify the proof of the main convergence theorem in this paper, the next lemma is introduced, which is well known (cf. [26, p.27] and [27, p.105]).

Lemma 4. *Consider any maximal monotone mapping $T : \mathcal{H} \to 2^{\mathcal{H}}$. Assume that the sequence $\{x^k\}$ in $\mathcal{H}$ converges weakly to some x, and the sequence $\{s^k\}$ on $\mathrm{dom}\, T$ converges strongly to some s. If $T(x^k) \ni s^k$ for all k, then the relation $T(x) \ni s$ must hold.*

Theorem 1. *If Assumption 1 holds, then the sequence $\{x^k\}$ generated by Algorithm 1 must converge weakly to an element of the solution set (if nonempty) of the problem (1).*

Proof. The notation $x^k(\alpha_k) = (I + \alpha_k B)^{-1}(I - \alpha_k F)(x^k)$ is equivalent to saying

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

which may be rewritten as

$$\alpha_k^{-1}(x^k - x^k(\alpha_k)) - F(x^k) \in B(x^k(\alpha_k)). \tag{21}$$

Take x^* from the intersection of $\mathcal{X}$ and the solution set of the problem (1). Then, we have $-F(x^*) \in B(x^*)$.So, it follows from the monotonicity of B that

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

which, together with $\alpha_k > 0$ and the monotonicity of F, implies

$$\begin{aligned} 0 &\leqslant \langle x^k(\alpha_k) - x^*, x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^*)\rangle \\ &= \langle x^k(\alpha_k) - x^*, x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\rangle - \alpha_k\langle x^k(\alpha_k) - x^*, F(x^k(\alpha_k)) - F(x^*)\rangle \\ &\leqslant \langle x^k(\alpha_k) - x^*, x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\rangle. \end{aligned} \tag{22}$$

In view of $x^k(\alpha_k) - x^* = x^k - x^* - (x^k - x^k(\alpha_k))$ and the step length rule (12), it is immediate that

$$\begin{aligned} \langle x^k - x^*, x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\rangle &\geqslant \langle x^k - x^k(\alpha_k), x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\rangle \\ &= \|x^k - x^k(\alpha_k)\|^2 - \alpha_k\langle x^k - x(\alpha_k), F(x^k) - F(x^k(\alpha_k))\rangle \\ &\geqslant \rho\|x^k - x^k(\alpha_k)\|^2. \end{aligned} \tag{23}$$

So, it follows from (15) that

$$\begin{aligned} \|x^{k+1} - x^*\|^2 &= \|P_{\mathcal{X}}[x^k - \gamma_k(x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k)))] - P_{\mathcal{X}}(x^*)\|^2 \\ &\leqslant \|x^k - x^* - \gamma_k(x^k - x^k(\alpha_k) - \lambda F(x^k) + \lambda F(x^k(\alpha_k))\|^2 \\ &= \|x^k - x^*\|^2 - 2\gamma_k\langle x^k - x^*, x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\rangle + \gamma_k^2\|x^k - x^k(\alpha_k) - \alpha_k F(x^k) \\ &\quad + \alpha_k F(x^k(\alpha_k))\|^2 \\ &\leqslant \|x^k - x^*\|^2 - 2\gamma_k\rho\|x^k - x^k(\alpha_k)\|^2 + \gamma_k^2\|x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\|^2 \\ &= \|x^k - x^*\|^2 - \theta(2-\theta)\rho^2\frac{\|x^k - x^k(\alpha_k)\|^4}{\|x^k - x^k(\alpha_k) - \alpha_k F(x^k) + \alpha_k F(x^k(\alpha_k))\|^2}. \end{aligned} \tag{24}$$

The relation $x^* = P_{\mathcal{X}}(x^*)$ comes from $x^* \in \mathcal{X}$, and the first inequality follows from the well-known non-expansive property of the projection (i.e, $\|P_{\mathcal{X}}(u) - P_{\mathcal{X}}(v)\| \leqslant \|u - v\|$ for all $u, v \in \mathcal{H}$), the second inequality from (23), and the last equality from the formula (14) for computing the relaxation factor.

The inequality (24) shows that $\{\|x^k - x^*\|\}$ is bounded. On the other hand, $\{\alpha_k\}$ is also bounded. Thus, $\{(x^k, \alpha_k)\}$ must be bounded and at least has one weak cluster point, say $(x^\infty, \alpha_\infty)$. Let $\{(x^{k_j}, \alpha_{k_j})\}$ be some subsequence converging weakly to $(x^\infty, \alpha_\infty)$. Then, $\{x^{k_j}(\alpha_{k_j})\}$ also converges weakly to x^∞ in that $\{\|x^k - x^k(\alpha_k)\|\}$ converges strongly to zero. Furthermore, it can be easily seen that $x^{k_j}(\alpha_{k_j}) \in \mathrm{dom}B$ is bounded in norm for all k_j.

It follows from (21) that

$$s^k := \alpha_k^{-1}(x^k - x^k(\alpha_k)) - (F(x^k) - F(x^k(\alpha_k))) \in (F + B)(x^k(\alpha_k)). \tag{25}$$

Case 1. $\alpha_\infty > 0$. In views of (25), the item (i) of Assumption 1 and the fact that $\{\|x^k - x^k(\alpha_k)\|\}$ converges strongly to zero, the desired result immediately follows from Lemma 4 and the item (iii) of Assumption 1.

Case 2. $\alpha_\infty = 0$. Note that our choice of α_k implies that the inequality (12) fails to hold for $\bar{\alpha}_k := \alpha_k\beta^{-1}$, if k is large enough. Thus, we have

$$\left\langle \frac{x^k - x^k(\bar{\alpha}_k)}{\|x^k - x^k(\bar{\alpha}_k)\|}, F(x^k) - F(x^k(\bar{\alpha}_k)) \right\rangle > (1-\rho)\|x^k - x^k(\bar{\alpha}_k)\|/\bar{\alpha}_k. \tag{26}$$

By using Lemma 2 and $\bar{\alpha}_k := \alpha_k\beta^{-1} > \alpha_k$, we have

$$\|x^k(\bar{\alpha}_k) - x^k\|/\bar{\alpha}_k \leqslant \|x^k(\alpha_k) - x^k\|/\alpha_k \quad \Rightarrow \quad \|x^k(\bar{\alpha}_k) - x^k\| \leqslant \beta^{-1}\|x^k(\alpha_k) - x^k\|.$$

So, we can conclude that $\lim_{k\to+\infty}\|x^k(\bar{\alpha}_k) - x^k\| = 0$. Combining this with (26) and the item i) of Assumption 1 yields

$$\lim_{k\to+\infty}\|x^k(\bar{\alpha}_k) - x^k\|/\bar{\alpha}_k = 0.$$

From $\bar{\alpha}_k := \alpha_k\beta^{-1} > \alpha_k$, (7) and the limit above, we further know that

$$\|x^k(\alpha_k) - x^k\|/\alpha_k = \beta^{-1}\|x^k(\alpha_k) - x^k\|/\bar{\alpha}_k \leqslant \beta^{-1}\|x^k(\bar{\alpha}_k) - x^k\|/\bar{\alpha}_k \to 0, \quad \text{as } k \to +\infty,$$

which, together with $\|x^k - x^k(\alpha_k)\| \to 0$ and the item i) of Assumption 1, implies that the sequence $\{s^k\}$ defined by (25) converges strongly to zero. Meanwhile, we also have known that the subsequence $\{x^{k_j}(\alpha_{k_j})\}$ converges weakly to x^∞. Therefore, it follows from Lemma 4 and the item iii) of Assumption 1 that $(F + B)(x^\infty) \ni 0$. The proof technique of the uniqueness of the weak cluster point is standard (cf. [1,2]), thus is omitted here. $\square$

Remark 2. The first part of this proof up to (25) follows that of [9, Theorem 4.2.3], see [13] for recent discussions. Furthermore, as pointed out in [9, Theorem 4.2.3], the corresponding weak convergence is valid in the real Hilbert space $\mathcal{H}$ when F is Lipschitz continuous monotone, and it can be at a locally linear rate if some additional assumptions hold as well. Although Tseng [2] previously analyzed weak convergence of his splitting algorithm without assuming Lipschitz continuity of the forward operator, it is more convenient and new to prove weak convergence of the proximal-descent algorithm by means of the two new properties of the forward–backward splitting.

5. Numerical experiments

In this section, we implemented Algorithm 1 to check its numerical performance. We also compared it with Tseng's splitting algorithm (TsengSA2 for short) with the self-adaptive strategy which outperformed the other one.

All numerical experiments were run in MATLAB R2013a on a desktop computer with an Intel (R) Xeon (R) 2.40 GHz CPU, 6.00 GB of RAM and Windows operating system.

Our first test problem is to solve the following complementarity problem

$$Ax - b + \partial\delta_{\mathcal{C}}(x) \ni 0,$$

where $\mathcal{C} = \{x : x_i \geqslant 0, i = 1, \ldots, n\}$ and $A = A_1 + A_2$ is an $n \times n$ block tridiagonal matrix with

$$A_1 = \begin{pmatrix} B & -I & & & \\ -I & B & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & B & -I \\ & & & -I & B \end{pmatrix}, \quad A_2 = \frac{hc}{2}\begin{pmatrix} O & I & & & \\ -I & O & I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & O & I \\ & & & -I & O \end{pmatrix},$$

and $n = m^2$, where $h = 1/(m+1)$, $c = 100$, I is an $m \times m$ identity matrix and B is an $m \times m$ matrix of the form $4I + Q + Q^T$ with $Q_{ij} = -1$ whenever $j = i + 1$, $i = 1, \ldots, m - 1$, otherwise $Q_{ij} = 0$. To know the solution of this problem in advance, we set

$b = Ae_1$, where e_1 is the first column of the corresponding identity matrix. Thus, $x^* = e_1$ is the unique solution of this complementarity problem. Obviously, the problem corresponds to the monotone inclusion (1) with

$$F(x) = Ax - b, \quad B(x) = \partial\delta_C(x).$$

For this problem, we took the parameters of Algorithm 1 to be $\beta = 0.9,\ \rho = 0.9,\ \theta = 1.9$ and the parameters of TsengSA2 to be $\beta = 0.5,\ \rho = 0.4$. To make our comparison as fair as possible, we took the maximum trial number of choosing step length to be 10. The starting point was $x^0 = (1, \ldots, 1)^T \in R^n$, and the stopping criterion was

$$\|x^k - x^*\|_\infty \leqslant 10^{-6}\|x^0 - x^*\|_\infty,$$

where $\|x\|_\infty = \max_{i=1,\ldots,n}\{|x_i|\}$. Moreover, we also terminated the algorithm if it failed to solve the problem after 20000 iterations and used "–" to stand for this situation.

We reported the corresponding numerical results in Table 1, where "Iter" and "CPUtime" mean the number of iterations and the elapsed CPU time (in seconds), respectively.

Our second test problem is to solve the following monotone inclusion

$$A^T(Ax - b) + \lambda\partial\|x\|_1 \ni 0,$$

which serves as optimality condition of the basis pursuit denoising [28] defined by

$$\min_{x\in R^n} \frac{1}{2}\|Ax - b\|_2^2 + \lambda\|x\|_1, \tag{27}$$

where $A \in R^{m\times n}$ $(m \ll n)$, $b \in R^m$, $\|x\|_1 = \sum_{i=1}^n |x_i|$ and $\lambda > 0$ is a constant. In our experiments, we solved a sparse signal recovery problem of the form (27) studied in [29]. The data are randomly generated with the default settings in Matlab. The matrix A is random, with $m = 1024$ and $n = 4096$. Its elements are chosen from Gaussian distribution with mean zero and variance $1/(2n)$. The observed vector $b \in R^{1024}$ is $b = Ax_{true} + v$, where v is sampled from Gaussian distribution with mean zero and variance 10^{-4}, and $x_{true} \in R^{4096}$ is a vector whose elements are zeros except 160 randomly placed ± 1 spikes. For the parameter λ, as in [29], we set $\lambda = 0.1\|A^Tb\|_\infty$.

In practical implementations, we took the parameters of Algorithm 1 to be $\beta = 0.5,\ \rho = 0.6,\ \theta = 1.9$ and the parameters of TsengSA2 to be $\beta = 0.9,\ \rho = 0.1$. To make our comparison as fair as possible, we took the maximum trial number of choosing step length to be 10. The starting point was $x^0 = A^Tb$, and the stopping criterion was

$$\|x^k - x^k(\alpha_k)\|_\infty \leqslant 10^{-4}.$$

Moreover, $(I + \alpha B)^{-1}(x)$ reduces to the so-called soft shrinkage or soft thresholding function given componentwise by

Table 1
Numerical results on the first test problem.

	α_{-1}	Algorithm 1			σ	TsengSA2		
		Iter	CPUtime	$\|x_k - x^*\|_\infty$		Iter	CPUtime	$\|x_k - x^*\|_\infty$
$n = 50^2$	10^{-2}	213	0.076	8.92e−7	10^{-2}	4696	1.386	9.98e−7
	10^{-1}	203	0.070	9.38e−7	10^{-1}	491	0.148	9.92e−7
	10^{0}	207	0.077	9.55e−7	10^{0}	399	0.118	9.12e−7
	10^{1}	220	0.082	8.49e−7	10^{1}	615	0.187	9.88e−7
	10^{2}	232	0.085	9.76e−7	10^{2}	501	0.147	9.45e−7
$n = 150^2$	10^{-2}	1912	4.080	9.79e−7	10^{-2}	–	–	–
	10^{-1}	1845	4.041	9.81e−7	10^{-1}	4252	6.383	9.97e−7
	10^{0}	1846	4.158	9.94e−7	10^{0}	3405	5.033	9.90e−7
	10^{1}	1883	4.296	9.89e−7	10^{1}	5429	8.055	9.93e−7
	10^{2}	1923	4.309	9.93e−7	10^{2}	4351	6.438	9.94e−7

Table 2
Numerical results on the second test problem.

α_{-1}	Algorithm 1			σ	TsengSA2		
	Iter	CPUtime	f_val		Iter	CPUtime	f_val
10^{-2}	159	2.33	3.21	10^{-2}	2720	26.09	3.45
10^{-1}	160	2.29	3.21	10^{-1}	1867	17.98	3.21
10^{0}	145	2.13	3.21	10^{0}	346	3.34	3.21
10^{1}	152	2.23	3.21	10^{1}	349	3.55	3.21
10^{2}	164	2.44	3.21	10^{2}	453	4.73	3.21

$$\left((I + \alpha\partial\|\cdot\|_1)^{-1}(x)\right)_i = \mathrm{sgn}(x_i)\max\{|x_i| - \alpha, 0\}$$

for all $i = 1, 2, \ldots, n$.

We reported the corresponding numerical results in Table 2, where "Iter", "CPUtime" and "f_val" mean the number of iterations, the elapsed CPU time (in seconds) and the objective function value, respectively.

From Tables 1 and 2, we observed that Algorithm 1 is not sensitive to the value of α_{-1}, it needs fewer iterations and less CPU time in achieving the same medium accuracy, when compared to Tseng's splitting algorithm. Moreover, the latter is sensitive to the σ's value.

6. Conclusions

In this paper, we discussed a proximal-descent algorithm for finding a zero of the sum of two maximal monotone operators in a real Hilbert space and analyzed its weak convergence without assuming Lipschitz continuity of the forward operator. Different from Tseng's, our convergence analysis relies on two new properties of the forward–backward splitting, introduced in this paper. Furthermore, based on theoretical analysis, we suggested a new strategy in the Armijo-like condition to locate the (not necessarily decreasing) step length self-adaptively, and it makes the proximal-descent algorithm practically effective.

Very recently, Dong [30, Theorem 3.1] put forth a new idea to discuss a fundamental property of the proximal point algorithm, and thus improved a classical result of Brézis and Lions [31]. How to apply his idea to our case will be an interesting research topic.

References

[1] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1976) 877–898.
[2] P. Tseng, A modified forward–backward splitting method for maximal monotone mappings, SIAM J. Control Optim. 38 (2) (2000) 431–446.
[3] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979) 964–979.
[4] G.B. Passty, Ergodic convergence to a zero of the sum of monotone operators in Hilbert space, J. Math. Anal. Appl. 72 (1979) 383–390.
[5] M. Sibony, Méthodes itératives pour les équations et inéquations aux dérivées partielles nonlinéares de type monotone, Calcolo 7 (1970). 65–18.
[6] P.H. Calamai, J.J. Moré, Projected gradient method for linearly constrained problems, Math. Program. 39 (1987) 93–116.
[7] H.G. Chen, R.T. Rockafellar, Convergence rates in forward–backward splitting, SIAM J. Optim. 7 (1997) 421–444.
[8] M.V. Solodov, B.F. Svaiter, A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator, Set-Valued Anal. 7 (1999) 323–345.
[9] Y.D. Dong, Splitting methods for monotone inclusions (Ph.D. dissertation), Nanjing University, 2003.
[10] B.S. He, A class of projection and contraction methods for monotone variational inequalities, Appl. Math. Optim. 35 (1997) 69–76.
[11] M.V. Solodov, P. Tseng, Modified projection-type methods for monotone variational inequalities, SIAM J. Control Optim. 34 (1996) 1814–1830.
[12] D.F. Sun, A class of iterative methods for solving nonlinear projection equations, J. Optim. Theory Appl. 91 (1) (1996) 123–140.
[13] Y.D. Dong, A variable metric proximal-descent algorithm for monotone operators, J. Appl. Math. Comput. 60 (2012) 563–571.
[14] E.M. Gafni, D.P. Bertsekas, Two-metric projection methods for constrained optimization, SIAM J. Control Optim. 6 (22) (1984) 936–964.
[15] PH.L. Toint, Global convergence of a class of trust region methods for nonconvex minimization in Hilbert space, IMA J. Numer. Anal. 8 (2) (1988) 231–252.
[16] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer-Verlag, New York, 2011.
[17] J.B. Hiriart-Urruty, C. Lemaréchal, Fundamentals of Convex Analysis, Grundlehren Text Editions, Springer-Verlag, 2001.
[18] R.T. Rockafellar, R. J-B. Wets, Variational Analysis, Springer, Berlin, 2009.
[19] T. Zhu, Z.G. Yu, A simple proof for some important properties of the projection mapping, Math. Inequal. Appl. 7 (2004) 453–456.
[20] Y.D. Dong, An extension of Luque's growth condition, Appl. Math. Lett. 22 (9) (2009) 1390–1393.
[21] Y.D. Dong, A practical PR+ conjugate gradient method only using gradient, Appl. Math. Comput. 219 (4) (2012) 2041–2052.
[22] M.A. Noor, Mixed quasi variational inequalities, Appl. Math. Comput. 146 (2003) 553–578.
[23] T. Pennanen, Dualization of generalized equations of maximal monotone type, SIAM J. Optim. 10 (2000) 809–835.
[24] R.T. Rockafellar, On the maximality of sums of nonlinear monotone operators, Trans. Am. Math. Soc. 149 (1970) 75–88.
[25] M.D. Voisei, C. Zălinescu, Maximal monotonicity criteria for the composition and the sum under weak interiority conditions, Math. Program. 123 (1, Ser. B) (2010) 265–283.
[26] H. Brézis, Opérateurs maximaux monotones, North-Holland, Amsterdam, 1973.
[27] D. Pascali, S. Sburlan, Nonlinear Mappings of Monotone Type, Editura Academiei, Bucharest, 1978.
[28] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1998) 33–61.
[29] S. Kim, K. Koh, M. Lusig, S. Boyd, D. Gorinevsky, An interior-point method for large-scale l_1-regularized least squares, IEEE J. Sel. Top. Signal Process. 1 (2007) 606–617.
[30] Y.D. Dong, The proximal point algorithm revisited, J. Optim. Theory Appl. http://dx.doi.org/10.1007/s10957-013-0351-3.
[31] H. Brézis, P.L. Lions, Produits infinis de resolvantes, Israel J. Math. 29 (1978) 329–345.