J Optim Theory Appl
DOI 10.1007/s10957-013-0351-3

The Proximal Point Algorithm Revisited

Yunda Dong

Received: 28 March 2012 / Accepted: 29 May 2013

Abstract In this paper, we consider the proximal point algorithm for the problem of finding zeros of any given maximal monotone operator in an infinite-dimensional Hilbert space. For the usual distance between the origin and the operator's value at each iterate, we put forth a new idea to achieve a new result on the speed at which the distance sequence tends to zero globally, provided that the problem's solution set is nonempty and the sequence of squares of the regularization parameters is non-summable. We show that it is comparable to a classical result of Brézis and Lions in general and becomes better whenever the proximal point algorithm does converge strongly. Furthermore, we also reveal its similarity to Güler's classical results in the context of convex minimization in the sense of strictly convex quadratic functions, and we discuss an application to an ϵ-approximation solution of the problem above.

Keywords Monotone operator · Convex minimization · Proximal point algorithm · Rate of convergence

1 Introduction

In this paper, we consider the problem of finding zeros of any given maximal monotone operator in an infinite-dimensional Hilbert space. When the operator is taken to be the sub-differential of a closed proper convex function, it corresponds to the optimality condition of minimizing such a function in this space.

An iterative procedure for solving the monotone inclusion above is the proximal point algorithm, first introduced by Martinet [1] for convex minimization and further generalized by Rockafellar [2] to get today's version. Rockafellar proved its global

Communicated by Qamrul Hasan Ansari.

Y. Dong (✉)
School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, P.R. China
e-mail: ydong@zzu.edu.cn

Published online: 18 June 2013

weak convergence by assuming that the solution set is nonempty and the sequence of the regularization parameters has a positive lower bound. Shortly after, Brézis and Lions [3, Proposition 8] weakened the latter assumption merely require the sequence of squares of the regularization parameters to be nonsummable. Furthermore, for the usual distance between the origin and the operator's value (which corresponds to a set!) at each iterate, they gave a speed at which the distance sequence tends to zero globally. Later on, in the context of convex minimization, Güler [4] well studied convergence behaviors of the proximal point algorithm and achieved some fundamental results. (i) Güler [4, Sect. 5] gave an example for which the proximal point algorithm fails to converge strongly. (ii) Assuming that the sequence of the regularization parameters is nonsummable, he deduced a convergence rate of the proximal point algorithm in terms of the difference of the function's value at each iterate and its value at a minimizer (see (8) below). Furthermore, if the sequence of the iterates does converge strongly to a minimizer of the function, the corresponding convergence rate can be improved (see (9) below).

Then, for the proximal point algorithm in the context of monotone inclusions, an important issue is to ask whether or not the speed proposed by Brézis and Lions can be improved whenever the proximal point algorithm does converge strongly.

In this paper, we will put forth a new idea to give a lemma (see Lemma 2.2 below) and then make use of it to resolve the aforementioned issue. As a result, under the same assumption, we achieve a new estimate (see (18) below) of the speed at which the distance sequence tends to zero globally. As shown below, this new one is comparable to a classical result of Brézis and Lions [3, Proposition 8] in general and becomes better whenever the proximal point algorithm does converge strongly.

It is generally accepted that Güler's estimates are classical results on convergence rates of the proximal point algorithm for convex minimization. We believe that our proposed one will play a similar role in analyzing convergence behaviors of the proximal point algorithm for monotone inclusion. In fact, as demonstrated at the end of Sect. 3, in the sense of strictly convex quadratic functions, our new estimate has been as good as Güler's for convex minimization.

The rest of this paper is organized as follows. In Sect. 2, we give some useful concepts and preliminary results. In particular, we put forth a new idea so as to achieve an interesting result (see Lemma 2.2 below), which plays a key role in the proof of the main theorem. In Sect. 3, we derive the new estimate and others. Furthermore, we reveal a remarkable similarity between our new estimate for monotone inclusion and Güler's for convex minimization. Finally, we discuss an application to an ϵ-approximation solution of monotone inclusion, a notion introduced and analyzed by Zaslavski (cf. [5]) very recently. In Sect. 4, we close this paper by some concluding remarks.

2 Preliminary Results

In this section, we first give some basic definitions and then provide some auxiliary results for later use.

Let H be an infinite-dimensional Hilbert space, in which $\langle x, y\rangle$ stands for the usual inner product and $\|x\| := \sqrt{\langle x, x\rangle}$ for the induced norm for any $x, y \in H$. Let U be a nonempty closed convex set in H. We use

$$|\cdot - U| := \min\big\{\|\cdot - u\| : u \in U\big\}$$

to stand for the usual distance between a point and this set U.

Definition 2.1 Let $f : H \to]-\infty, +\infty]$ be a closed proper convex function. Then for any given $x \in H$ the sub-differential of f at x is defined by

$$\partial f(x) := \big\{s \in H : f(y) - f(x) \geq \langle s, y - x\rangle,\ \forall y \in H\big\}.$$

Each element s is called a sub-gradient of f at x. 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.

To concisely give the following definition, we agree on that the notation $(x, w) \in A$ and $x \in H$, $w \in A(x)$ have the same meaning.

Definition 2.2 Let $A : H \rightrightarrows H$ be an operator. It is called monotone iff

$$\big\langle x - x', w - w'\big\rangle \geq 0, \quad \forall (x, w) \in A,\ \forall \big(x', w'\big) \in A;$$

maximal monotone iff it is monotone and for given $\hat{x} \in H$ and $\hat{w} \in H$ the following implication relation holds:

$$\langle x - \hat{x}, w - \hat{w}\rangle \geq 0, \quad \forall (x, w) \in A \quad \Rightarrow \quad (\hat{x}, \hat{w}) \in A.$$

As is well known, the sub-differential of any closed proper convex function in an infinite-dimensional Hilbert space is maximal monotone as well. Furthermore, for any given maximal monotone operator $A : H \rightrightarrows H$, it is Minty [6] who proved that there must exist a unique $y \in H$ such that $(I + \lambda A)(y) \ni x$ for all $x \in H$ and $\lambda > 0$, where I stands for the identity operator, i.e., $I(x) = x$ for all $x \in H$. This implies that the corresponding operator $(I + \lambda A)^{-1}$ is single-valued.

For any given maximal monotone operator $A : H \rightrightarrows H$, there are other related properties. (i) For all $x \in H$, the set $A(x)$ must be either empty or nonempty closed convex; see [7]. (ii) The set $A^{-1}(0) := \{x : 0 \in A(x)\}$ is either empty or nonempty closed convex. Therefore, for all $x \in H$, the distance between the set $A(x)$ and the origin is uniquely determined, and so is the distance between the set $A^{-1}(0)$ and the origin.

Based on these facts, we state the problem under consideration in this paper as follows. It is to find an $x \in H$ such that

$$0 \in A(x), \tag{1}$$

where $A : H \rightrightarrows H$ is a maximal monotone operator. Henceforth, for this monotone inclusion, the solution set $A^{-1}(0)$ is always assumed to be nonempty.

An important instance of the problem above is the optimality condition of minimizing a closed proper convex function $f : H \to]-\infty, +\infty]$, which reads

$$0 \in \partial f(x). \tag{2}$$

An iterative procedure for solving (2) is the proximal point algorithm [1]: For any given starting point $x^0 \in H$, it recursively generates a sequence of iterates $\{x^k\}$ by

$$0 \in \partial f\big(x^{k+1}\big) + \lambda^{-1}\big(x^{k+1} - x^k\big), \quad k = 0, 1, \ldots, \tag{3}$$

where $\lambda > 0$ is the regularization parameter. Note that such x^{k+1} is also the unique solution to the corresponding convex minimization $\min_{x\in H}\{f(x) + \frac{1}{2\lambda}\|x - x^k\|^2\}$ and called the proximal point (see Moreau [8]) of f at x^k.

Replacing the sub-differential operator by a general maximal monotone operator $A : H \rightrightarrows H$, Rockafellar [2] got today's form of the proximal point algorithm:

$$0 \in A\big(x^{k+1}\big) + \lambda_k^{-1}\big(x^{k+1} - x^k\big), \quad k = 0, 1, \ldots, \tag{4}$$

where $\lambda_k > 0$ can vary from iteration to iteration. From the result of Minty, we can rewrite the proximal point algorithm (4) as

$$x^{k+1} = (I + \lambda_k A)^{-1}\big(x^k\big), \quad k = 0, 1, \ldots. \tag{5}$$

In an infinite-dimensional Hilbert space, we say that the proximal point algorithm does converge strongly if and only if the sequence of the iterates does converge strongly to an element, say x^∞, of the solution set in the sense of

$$\lim_{k\to+\infty} \big\|x^k - x^\infty\big\| = 0. \tag{6}$$

Of course, this relation (6) will always hold in the finite-dimensional inner product space $\mathbb{R}^n$, in which there is no distinction between strong convergence and weak convergence so that we simply say one word "convergence".

Proposition 2.1 *Let* $\{(x^k, \lambda_k)\}$ *be the sequence in the proximal point algorithm. Then we have*

$$\big|A\big(x^k\big) - 0\big|^2 \sum_{i=0}^{k-1} \lambda_i^2 \le \big|x^0 - A^{-1}(0)\big|^2. \tag{7}$$

In the context of convex minimization, we have

$$\big(f\big(x^k\big) - \min f(x)\big) \sum_{i=0}^{k-1} \lambda_i \le \frac{1}{2}\big|x^0 - A^{-1}(0)\big|^2. \tag{8}$$

Furthermore, if the proximal point algorithm does converge strongly, then the associated convergence rate can be improved above to

$$\lim_{k\to+\infty} \big(f\big(x^k\big) - \min f(x)\big) \sum_{i=0}^{k-1} \lambda_i \to 0, \quad \textit{whenever} \quad \lim_{k\to+\infty} \sum_{i=0}^{k-1} \lambda_i = +\infty. \tag{9}$$

Note that this proposition characterizes some of the fundamental properties of the proximal point algorithm. The inequality (7) corresponds to a classical result of Brézis and Lions [3, Proposition 8] and the others are attributed to Güler [4].

For the proximal point algorithm (5), we refer the reader to Rockafellar [2] and Zaslavski [9] for two different approximate versions. Furthermore, the reader may consult [10–12] for other related discussions of convergence rates given in terms of the closeness of the iterate to the solution set (but with some restrictive assumptions).

Very recently, Zaslavski (cf. [5]) introduced and analyzed a notion of an ϵ-approximation solution of the monotone inclusion (1). It corresponds to the problem of finding an $x \in H$ such that

$$A(x) \ni w, \quad \text{with } \|w\| \le \epsilon.$$

This notion's importance can be understood through a simple but demonstrative example. Let us consider the problem of minimizing $|x|$ in the space $\mathbb{R}$, and the associated optimality condition is

$$0 \in \partial|x| = \begin{cases} \{s : |s| \le 1\}, & \text{if } x = 0, \\ \{x/|x|\}, & \text{if } x \ne 0. \end{cases}$$

Since the operator $A := \partial|\cdot|$ is maximal monotone in the space $\mathbb{R}$, this optimality condition becomes a monotone inclusion. For the latter, once we let $\epsilon < 1$ (say $\epsilon = 0.99$), its ϵ-approximation solution must be unique and coincides with its exact solution $x^* = 0$. Since ϵ is not necessarily required to be small, this fact outstands the value of the notion of an ϵ-approximation solution in some cases.

Now we state a lemma, which will be used in the proof of the main result of this paper.

Lemma 2.1 *Let $r^k := x^k - (I + \lambda_k A)^{-1}(x^k) = x^k - x^{k+1}$, where $\{(x^k, \lambda_k)\}$ is the sequence in the proximal point algorithm. Then the sequence $\{\lambda_k^{-1}\|r^k\|\}$ is decreasing and the sequence $\{\|r^k\|\}$ satisfies*

$$\left(2\lambda_k\lambda_{k+1}^{-1} - 1\right)\left\|r^{k+1}\right\|^2 \le \left\|r^k\right\|^2 - \left\|r^{k+1} - r^k\right\|^2. \tag{10}$$

Proof In view of the notation $r^k = x^k - (I + \lambda_k A)^{-1}(x^k)$, we have

$$x^k \in \left(x^k - r^k\right) + \lambda_k A\left(x^k - r^k\right) \quad \Leftrightarrow \quad \lambda_k^{-1} r^k \in A\left(x^k - r^k\right).$$

Similarly, $\lambda_{k+1}^{-1} r^{k+1} \in A(x^{k+1} - r^{k+1})$. Thus, it follows from A's monotonicity that

$$\left\langle x^{k+1} - r^{k+1} - \left(x^k - r^k\right), \lambda_{k+1}^{-1} r^{k+1} - \lambda_k^{-1} r^k \right\rangle \ge 0,$$

which, together with the equivalent iterative relation $x^{k+1} = x^k - r^k$, implies

$$\left\langle -r^{k+1}, \lambda_{k+1}^{-1} r^{k+1} - \lambda_k^{-1} r^k \right\rangle \ge 0 \quad \Leftrightarrow \quad \lambda_k\lambda_{k+1}^{-1}\left\|r^{k+1}\right\|^2 \le \left\langle r^{k+1}, r^k \right\rangle. \tag{11}$$

By using the identity

$$2\langle r^{k+1}, r^k\rangle = \|r^{k+1}\|^2 + \|r^k\|^2 - \|r^{k+1} - r^k\|^2,$$

we can further get the desired result. Meanwhile, applying the Cauchy–Schwartz inequality to (11) yields the result that the decreasing property of $\{\lambda_k^{-1}\|r^k\|\}$, equivalently $\{\|\lambda_k^{-1}(x^k - x^{k+1})\|\}$. □

Note that, in the case of A being single-valued, the associated proof easily follows from A's monotonicity:

$$\langle A(x^{k+1}) - A(x^k), x^{k+1} - x^k\rangle \geq 0$$

and the iterative formula $x^{k+1} - x^k = -\lambda_k A(x^{k+1})$, $\lambda_k > 0$. So, combining these with the Cauchy–Schwartz inequality yields the decreasing property of $\{\|A(x^{k+1})\|\}$, as was done in [3].

The decreasing property of $\{\|\lambda_k^{-1}(x^k - x^{k+1})\|\}$ can be viewed as an extension of [4, Lemma 2.1] to monotone operators, and a closely related terminology is the Yosida approximation, denoted by $\lambda^{-1}(x - (I + \lambda A)^{-1}(x))$.

Obviously, from Lemma 2.1 above, we can get an abstract inequality: If $A : H \rightrightarrows H$ is maximal monotone, then for any given $x \in H$ and $\tilde{x} = (I + \lambda A)^{-1}(x)$, the relation

$$\begin{aligned}
&(2\lambda\tilde{\lambda}^{-1} - 1)\|\tilde{x} - (I + \tilde{\lambda}A)^{-1}(\tilde{x})\|^2 \\
&\quad \leq \|x - (I + \lambda A)^{-1}(x)\|^2 - \|\tilde{x} - (I + \tilde{\lambda}A)^{-1}(\tilde{x}) - (x - (I + \lambda A)^{-1}(x))\|^2
\end{aligned}$$

holds for all $\tilde{\lambda} > 0$ and for all $\lambda > 0$ Yet, we will not further discuss it in this paper.

Next, we give a fundamental result of characterizing interrelations of three sequences of positive numbers frequently arise in analysis of convergence behaviors of some optimization methods, and it is crucial in proving the main result of this paper. Note that, to the best of the author's knowledge, the idea of the proof is new.

Lemma 2.2 *Let* $\{\alpha_k\}, \{\beta_k\}, \{\gamma_k\}$ *be sequences of positive numbers. Assume that they satisfy*

$$\alpha_{k+1}^2 \leq \alpha_k^2 - \beta_k\gamma_k, \quad k = 0, 1, \ldots, \tag{12}$$

the sequence $\{\beta_k\}$ *is nonsummable and the sequence* $\{\gamma_k\}$ *is decreasing. Then there exists* ε_k *such that*

$$\gamma_k \sum_{i=0}^{k} \beta_i \leq 2\alpha_0\varepsilon_k, \tag{13}$$

$$\alpha_k \leq \varepsilon_k \leq \alpha_0, \qquad \lim_{k\to+\infty} \varepsilon_k = \lim_{k\to+\infty} \alpha_k. \tag{14}$$

Proof It follows from (12) that

$$\alpha_{i+1} \leq \sqrt{\alpha_i^2 - \beta_i \gamma_i} \leq \alpha_i - \beta_i \gamma_i (2\alpha_i)^{-1}, \quad i = 0, 1, \ldots.$$

This is the desired recursive inequality. Summing it for $i = 0, 1, \ldots, k$ and rearranging terms yield

$$\sum_{i=0}^{k} \beta_i \gamma_i (2\alpha_i)^{-1} \leq \alpha_0 - \alpha_{k+1} \leq \alpha_0,$$

which, together with the decreasing property of the sequence $\{\gamma_i\}$, implies

$$\gamma_k \sum_{i=0}^{k} \beta_i (2\alpha_i)^{-1} \leq \alpha_0.$$

Hence, we can further get

$$\gamma_k \sum_{i=0}^{k} \beta_i \leq \frac{\alpha_0 \sum_{i=0}^{k} \beta_i}{\sum_{i=0}^{k} \beta_i (2\alpha_i)^{-1}} = 2\alpha_0 \frac{\sum_{i=0}^{k} \beta_i \alpha_i^{-1} \alpha_i}{\sum_{i=0}^{k} \beta_i \alpha_i^{-1}}. \tag{15}$$

Denote

$$\varepsilon_k := \frac{\sum_{i=0}^{k} \beta_i \alpha_i^{-1} \alpha_i}{\sum_{i=0}^{k} \beta_i \alpha_i^{-1}} = \sum_{i=0}^{k} \frac{\beta_i \alpha_i^{-1}}{\sum_{i=0}^{k} \beta_i \alpha_i^{-1}} \alpha_i.$$

Form the term on the right-hand side, we can see that ε_k must lie in the interval $[\alpha_k, \alpha_0]$ because the sequence $\{\alpha_k\}$ is clearly decreasing and $\beta_i \alpha_i^{-1} \geq \beta_i \alpha_0^{-1} > 0$ for $i = 0, 1, \ldots, k$. Meanwhile, for the nonnegative and decreasing sequence $\{\alpha_k\}$, its limit must exist, and for the sequence $\{\beta_k \alpha_k^{-1}\}$, it is also nonsummable. So, we can use the Silverman–Toeplitz theorem [13, p. 43] to conclude that $\{\varepsilon_k\}$ and $\{\alpha_k\}$ have the same limit. The proof is complete. □

3 Main Results

In this section, we derive our new result on the proximal point algorithm for monotone inclusion. Moreover, we reveal its similarity to Güler's for convex minimization, and we discuss an application to an ϵ-approximation solution of monotone inclusion in the sense of Zaslavski.

Theorem 3.1 *Let* $\{(x^k, \lambda_k)\}$ *be the corresponding iterate-parameter sequence in the proximal point algorithm. Assume that* $\lim_{k\to+\infty} \sum_{i=0}^{k-1} \lambda_i^2 = +\infty$. *Then*

$$\left\|\lambda_k^{-1}\left(x^k - x^{k+1}\right)\right\|^2 \sum_{i=0}^{k} \lambda_i^2 \leq 2\left|x^0 - A^{-1}(0)\right| \varepsilon_k, \tag{16}$$

$$\left|x^k - A^{-1}(0)\right| \le \varepsilon_k \le \left|x^0 - A^{-1}(0)\right|, \qquad \lim_{k\to+\infty} \varepsilon_k = \lim_{k\to+\infty} \left|x^k - A^{-1}(0)\right|. \tag{17}$$

As a consequence, the following relation:

$$\lim_{k\to+\infty} \sup \left|A\left(x^k\right) - 0\right|^2 \sum_{i=0}^{k-1} \lambda_i^2 \le 2\left|x^0 - A^{-1}(0)\right| \lim_{k\to+\infty} \left|x^k - A^{-1}(0)\right| \tag{18}$$

holds as well.

Proof In view of the iterative formula (4), we have

$$\lambda_k^{-1}\left(x^k - x^{k+1}\right) \in A\left(x^{k+1}\right).$$

Thus, by the assumption $0 \in A(x^*)$ (because the existence of a solution has been assumed throughout this paper) and A's monotonicity, we can further get

$$\left\langle x^{k+1} - x^*, \lambda_k^{-1}\left(x^k - x^{k+1}\right) - 0\right\rangle \ge 0 \quad \overset{\lambda_k > 0}{\Longrightarrow} \quad \left\langle x^{k+1} - x^*, x^k - x^{k+1}\right\rangle \ge 0.$$

Therefore

$$\begin{aligned} \left\|x^k - x^*\right\|^2 &= \left\|x^k - x^{k+1} + x^{k+1} - x^*\right\|^2 \\ &= \left\|x^k - x^{k+1}\right\|^2 + 2\left\langle x^k - x^{k+1}, x^{k+1} - x^*\right\rangle + \left\|x^{k+1} - x^*\right\|^2 \\ &\ge \left\|x^k - x^{k+1}\right\|^2 + \left\|x^{k+1} - x^*\right\|^2. \end{aligned} \tag{19}$$

This implies

$$\left|x^{k+1} - A^{-1}(0)\right|^2 \le \left|x^k - A^{-1}(0)\right|^2 - \lambda_k^2 \left\|\lambda_k^{-1}\left(x^k - x^{k+1}\right)\right\|^2. \tag{20}$$

Denote

$$\alpha_k := \left|x^k - A^{-1}(0)\right|, \qquad \beta_k := \lambda_k^2, \qquad \gamma_k := \left\|\lambda_k^{-1}\left(x^k - x^{k+1}\right)\right\|^2.$$

Then it follows from Lemma 2.2 that

$$\left\|\lambda_k^{-1}\left(x^k - x^{k+1}\right)\right\|^2 \sum_{i=0}^{k} \lambda_i^2 \le 2\left|x^0 - A^{-1}(0)\right| \varepsilon_k,$$

$$\left|x^k - A^{-1}(0)\right| \le \varepsilon_k \le \left|x^0 - A^{-1}(0)\right|, \qquad \lim_{k\to+\infty} \varepsilon_k = \lim_{k\to+\infty} \left|x^k - A^{-1}(0)\right|$$

whenever $\lim_{k\to+\infty} \sum_{i=0}^{k-1} \lambda_i^2 = +\infty$. Thus, the relation (16) is proved.

Next, we need to prove (18). In fact, since $\lambda_k^{-1}(x^k - x^{k+1}) \in A(x^{k+1})$, we can use the relation (16) to further obtain

$$\left|A\left(x^{k+1}\right) - 0\right|^2 \sum_{i=0}^{k} \lambda_i^2 \le 2\left|x^0 - A^{-1}(0)\right| \varepsilon_k. \tag{21}$$

On the other hand, it follows from the existence of the limit of $\{|x^k - A^{-1}(0)|\}$ and (17) that

$$\lim_{k\to+\infty} \varepsilon_k = \lim_{k\to+\infty} \left|x^k - A^{-1}(0)\right| = \lim_{k\to+\infty} \left|x^{k+1} - A^{-1}(0)\right|.$$

Hence, taking the limits on both sides of (21) and then using $k \leftarrow k+1$ yield the desired result.

It remains to prove global weak convergence. Denote $w^k := \lambda_{k-1}^{-1}(x^{k-1} - x^k)$. From (16) and the assumption $\lim_{k\to+\infty} \sum_{i=0}^{k-1} \lambda_i^2 = +\infty$, we know that the sequence $\{\|w^k\|\}$ converges to zero. Meanwhile, it follows from (19) that the sequence $\{\|x^k - x^*\|\}$ has the limit, thus the sequence $\{x^k\}$ is bounded and there must exist a weak accumulation point. From these two facts, A's maximality and the relation $w^k \in A(x^k)$, we can follow [2] to prove global weak convergence of the proximal point algorithm in an infinite-dimensional Hilbert space. Therefore, our new estimate makes sense whenever $\lim_{k\to+\infty} \sum_{i=0}^{k-1} \lambda_i^2 = +\infty$. □

From this theorem, it can be easily seen that our new result (18) is comparable to the classical result (7) of Brézis and Lions in general. In fact, under the assumption above, which is equivalent to saying that

$$\delta_k := \left(\sum_{i=0}^{k-1} \lambda_i^2\right)^{1/2} \to +\infty, \quad \text{as } k \to +\infty, \tag{22}$$

either result tells us that $|A(x^k) - 0| \le O(1/\delta_k)$. More importantly, only our new result (18) can be improved above to $|A(x^k) - 0| = o(1/\delta_k)$ whenever the proximal point algorithm does converge strongly. This is because, if it is so, then there must exist an element x^∞ of the solution set such that (6) holds. Thus, we further have

$$\left|x^k - A^{-1}(0)\right| \le \left\|x^k - x^\infty\right\| \to 0, \quad \text{as } k \to +\infty,$$

which, together with (18), yields the desired conclusion.

Remark 3.1 In Theorem 3.1, we follow [3] to make the assumption equivalent to (22). This assumption is weaker than the one in [2], where the sequence of the regularization parameters was assumed to have a positive lower bound. The assumption here will be still meaningful to consider the continuous version of the proximal point algorithm:

$$0 \in \frac{dx(t)}{dt} + \partial f\big(x(t)\big), \quad x(0) = x,$$

where x belongs to the closure of the f's effective domain $\{x : f(x) < +\infty\}$.

The corresponding backward Euler (implicit) approximation of this differential inclusion is

$$0 \in \frac{x^{k+1} - x^k}{t_{k+1} - t_k} + \partial f\big(x(t_{k+1})\big), \quad x^0 \in H,\ k = 0, 1, \ldots.$$

If $\lambda_k := t_{k+1} - t_k > 0$ is large, then the approximation is poor. To prevent it, we may require λ_k be small. Furthermore, the sequence $\{\lambda_k\}$ needs to satisfy $\lambda_k \downarrow 0$ and $\delta_k \uparrow +\infty$ as $k \to +\infty$. When the ∂f above is replaced by some maximal monotone operator, a related discussion is similar.

Interestingly, there is a remarkable similarity between our new estimate (18) for monotone inclusion and Güler's for convex minimization. First, let us show that the sequence $\{|\partial f(x^k) - 0|^2\}$ is of the same order as $\{f(x^k) - \min f(x)\}$ in the sense of strictly convex quadratic functions. To this end, now we set

$$f(x) = \frac{1}{2}\langle x, Mx\rangle + \langle q, x\rangle + c,$$

where M is an $n \times n$ symmetric positive definite matrix, q is an n-dimensional vector and c is a constant. Clearly, for the function's sub-differential, we have $\partial f(x) = \{\nabla f(x)\} = \{Mx + q\}$. Thus, the optimality condition of minimizing $f(x)$ is $\nabla f(x^*) = 0$, where x^* is the unique solution. Below we consider the left-hand side of the inequality (8)

$$\begin{aligned} f(x^k) - f(x^*) &= \langle \nabla f(x^*), x^k - x^* \rangle + \frac{1}{2}\langle x^k - x^*, M(x^k - x^*)\rangle \\ &= \frac{1}{2}\langle x^k - x^*, M(x^k - x^*)\rangle. \end{aligned}$$

Meanwhile, for the left-hand side of the inequality (18), in views of the optimality condition above and the iterative formula, we can get

$$\begin{aligned} |\partial f(x^k) - 0|^2 &= \|\nabla f(x^k)\|^2 \\ &= \|\nabla f(x^k) - \nabla f(x^*)\|^2 \\ &= \|M(x^k - x^*)\|^2. \end{aligned}$$

Thus, the assertion follows easily. Obviously, in the case of λ_k being a constant, our new estimate (18) becomes $|\partial f(x^k) - 0|^2 = o(1/k)$ whereas Güler's estimate (9) becomes $f(x^k) - \min f(x) = o(1/k)$. This implies that, in the sense of strictly convex quadratic functions, our new estimate (18) for monotone inclusion has been as good as Güler's (9) for convex minimization.

Remark 3.2 The aforementioned are accuracy measures in the cases of convex minimization and monotone inclusion. For other related discussions in a finite-dimensional space, we refer to a recent paper by Nemirovski et al. [14] and the references cited therein.

Below we would like to list the corresponding estimates of convergence rates of the proximal point algorithm when the problem setting is finite-dimensional and the sequence of the regularization parameters has a positive lower bound. They are direct consequences of (18) and Güler's estimate.

Corollary 3.1 *In the space $\mathbb{R}^n$, if the sequence of the regularization parameters has a positive lower bound, then the proximal point algorithm has the following estimate of convergence rate in the context of convex minimization*:

$$f\big(x^k\big) - \min f(x) = o(1/k).$$

And in the context of the monotone inclusion (1), *we have*

$$\big|A\big(x^k\big) - 0\big|^2 = o(1/k).$$

At the end of this section, we would like to discuss the following question: How many iterations in the worst case are needed for the proximal point algorithm with constant regularization parameters to find an ϵ-approximation solution of the monotone inclusion $0 \in A(x)$ in the space $\mathbb{R}^n$? According to Theorem 3.1, the answer is satisfactory. It follows from the iterative formula

$$w^k := \lambda_{k-1}^{-1}\big(x^{k-1} - x^k\big), \quad w^k \in A\big(x^k\big)$$

and

$$o(1/k) = \big\|\lambda_{k-1}^{-1}\big(x^{k-1} - x^k\big)\big\|^2 \leq \epsilon^2$$

that the proximal point algorithm used will find an ϵ-approximation solution in at most K iterations, where K is much fewer than $O(1/\epsilon^2)$. In contrast, if we make use of the classical result of Brézis and Lions [3, Proposition 8], then the associated K becomes $O(1/\epsilon^2)$ in the worst case.

4 Conclusions

In this paper, we have studied the proximal point algorithm for monotone inclusion in an infinite-dimensional Hilbert space. For the usual distance between the origin and the operator's value at each iterate, under the weakest possible assumptions, we gave a new result on the speed at which the distance sequence tends to zero globally. In contrast to the classical result of Brézis and Lions, our new result is comparable in general and becomes better whenever the proximal point algorithm does converge strongly.

Here it should be specially stressed that, to the best of the author's knowledge, the idea embodied in the proof of Lemma 2.2 is new in the optimization literature. How to further apply it to analyze the Douglas–Rachford operator splitting method of Lions and Mercier [15] (cf. [16–19]) for monotone inclusion is one of our on-going research topics.

Acknowledgements The author is very grateful to the associated editor and the referees for their valuable suggestions and helpful comments which improve the presentation of this paper.

References

1. Martinet, B.: Regularisation d'inéquations variationelles par approximations successives. Rev. Fr. Inform. Rech. Oper. **4**, 154–158 (1970)
2. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
3. Brézis, H., Lions, P.L.: Produits infinis de resolvantes. Isr. J. Math. **29**, 329–345 (1978)
4. Güler, O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403–419 (1991)
5. Zaslavski, A.J.: Proximal point algorithm for finding a common zero of a finite family of maximal monotone operators in the presence of computational errors. Nonlinear Anal. **75**(16), 6071–6087 (2012)
6. Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**, 341–346 (1962)
7. Aubin, J.P., Ekeland, I.: Applied Nonlinear Analysis, Interscience Publications. Wiley, New York (1984)
8. Moreau, J.J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. Fr. **93**, 273–299 (1965)
9. Zaslavski, A.J.: Maximal monotone operators and the proximal point algorithm in the presence of computational errors. J. Optim. Theory Appl. **150**, 20–32 (2011)
10. Luque, F.J.: Asymptotic convergence analysis of the proximal point algorithm. SIAM J. Control Optim. **22**, 277–293 (1984)
11. Pennanen, T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**, 193–202 (2002)
12. Dong, Y.D.: A new relative error criterion for the proximal point algorithm. Nonlinear Anal. **64**, 2143–2148 (2006)
13. Hardy, G.H.: Divergent Series. Clarendon Press, Oxford (1949)
14. Nemirovski, A., Onn, S., Rothblum, U.: Accuracy certificates for computational problems with convex structure. Math. Oper. Res. **35**(11), 52–78 (2010)
15. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
16. Lawrence, J., Spingarn, J.E.: On fixed points of non-expansive piecewise isometric mappings. Proc. Lond. Math. Soc. **55**, 605–624 (1987)
17. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**(3), 293–318 (1992)
18. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**, 475–504 (2004)
19. Dong, Y.D., Fischer, A.: A family of operator splitting methods revisited. Nonlinear Anal. **72**, 4307–4315 (2010)