The Proximal Point Algorithm Is $\mathcal{O}(1/\epsilon)$

Yunda Dong

Department of Mathematics, Zhengzhou University, Zhengzhou, 450001, PR China

Abstract In this paper, we give two new results on the proximal point algorithm for finding a zero point of any given maximal monotone operator. First, we prove that if the sequence of regularization parameters is non-increasing then so is that of residual norms. Then, we use it to show that the proximal point algorithm can find an approximate zero point (assume the existence of such zero point) in at most $\mathcal{O}(1/\epsilon)$ outer iterations, where $\epsilon > 0$ is the accuracy parameter described in the stopping criterion.

Keywords Monotone operator · Proximal point algorithm · Residual norm

1 Introduction

In this paper, we consider the problem of finding an x in a Hilbert space $\mathcal{H}$ such that

$$T(x) \ni 0, \tag{1}$$

where $T : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximal monotone operator. From now on, we call such x a zero point of the operator T, and always assume that at least one such zero point exists.

The problem above can be solved by the proximal point algorithm (PPA for short) [1, 2]. For any given starting point $x^0 \in \mathcal{H}$, it generates a sequence of iterates $\{x^k\}$ by the relation

$$0 \in T(x^{k+1}) + \lambda_k^{-1}(x^{k+1} - x^k), \quad k \geq 0, \tag{2}$$

where $\{\lambda_k\}$ is a sequence of regularization parameters with $\lambda_k \geq \lambda > 0$. When specialized to the minimization of a given proper closed convex function f on R^n, it reads

$$0 \in \partial f(x^{k+1}) + \lambda_k^{-1}(x^{k+1} - x^k), \quad k \geq 0,$$

where ∂f is the sub-differential of f. In this case, the uniquely determined x^{k+1} is called the proximal point [3] of f at x^k.

Based on a result due to Minty [4], who showed that for any given $x \in \mathcal{H}$ and $\lambda > 0$, there must exist a unique $\tilde{x}$ such that $(I + \lambda T)(\tilde{x}) \ni x$, we can rewrite (2) as

$$x^{k+1} = (I + \lambda_k T)^{-1}(x^k), \quad k \geq 0,$$

where I stands for the identity operator.

To date, the majority of the PPA's fundamental properties have been discovered and analyzed. (i) Rockafellar [2] proved its global convergence in the presence of summable computational errors, and quite recently, Zaslavski [5] analyzed it in the non-summable form. (ii) Under certain assumptions, Rockafellar [2] showed its local convergence at a linear rate, see [6–8] for further discussions. (iii) Güler [9] demonstrated via an example that the PPA may fail to converge strongly.

In this paper, we give two new properties of the PPA. First, as for the sequence of residual norms generated by the PPA, we prove that such a sequence is non-increasing if that of regularization parameters is non-increasing, i.e.,

$$\lambda_{k+1} \leq \lambda_k \quad \Rightarrow \quad \|x^{k+1} - (I + \lambda_{k+1} T)^{-1}(x^{k+1})\| \leq \|x^k - (I + \lambda_k T)^{-1}(x^k)\|.$$

Then, we use it to show that the PPA can find an approximate zero point in at most $\mathcal{O}(1/\epsilon)$ outer iterations whenever the corresponding stopping criterion of the proximal point algorithm is

$$\|x^k - (I + \lambda_k T)^{-1}(x^k)\|^2 \leq \epsilon, \tag{3}$$

where $\epsilon > 0$ is a prescribed accuracy parameter. To the best of the author's knowledge, the latter is the first upper bound of the number of the PPA's outer iterations in the setting of maximal monotone operators.

2 Preliminary Results

We first review some basic definitions and then provide some auxiliary results for later use.

Throughout this paper, we denote by $\langle x, y \rangle$ for the usual inner product for any $x, y \in \mathcal{H}$, and by $\|x\| = \sqrt{\langle x, x \rangle}$ for the induced norm. T^{-1} denotes the inverse of T. Clearly, $T^{-1}T = I$. In what follows, the notation $(x, w) \in T$ and the notation $x \in \mathcal{H}$, $w \in T(x)$ have the same meaning.

Definition 1. Let $T : \mathcal{H} \rightrightarrows \mathcal{H}$ be a possibly multi-valued operator. It is called monotone if

$$\langle x - x',\ w - w' \rangle \geq 0, \qquad \forall (x,\ w) \in T,\ \ (x',\ w') \in T;$$

maximal monotone if it is monotone and its graph $\{(x,\ w) : x \in \mathcal{H}, w \in T(x)\}$ can not be enlarged without loss of monotonicity.

Lemma 1. *Let $r^k := x^k - (I + \lambda_k T)^{-1}(x^k)$, where $\{(x^k, \lambda_k)\}$ is the sequence generated by the PPA. Then*

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

In particular, if $\lambda_{k+1} \leq \lambda_k$ then it reduces to

$$\|r^{k+1}\|^2 \leq \|r^k\|^2 - \|r^{k+1} - r^k\|^2 \qquad \Rightarrow \qquad \|r^{k+1}\| \leq \|r^k\|.$$

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

$$(x^k - r^k) + \lambda_k T(x^k - r^k) \ni x^k \quad \Leftrightarrow \quad T(x^k - r^k) \ni \lambda_k^{-1} r^k.$$

Meanwhile, we also have

$$T(x^{k+1} - r^{k+1}) \ni \lambda_{k+1}^{-1} r^{k+1}.$$

Thus, it follows from the T's monotonicity that

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

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

$$\langle -r^{k+1},\ \lambda_{k+1}^{-1}\, r^{k+1} - \lambda_k^{-1}\, r^k \rangle \geq 0 \quad \Leftrightarrow \quad \lambda_k \lambda_{k+1}^{-1} \|r^{k+1}\|^2 \leq \langle r^{k+1},\ r^k \rangle$$

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 results. □

3 Main Result

In this section, by using Lemma 1, we give an upper bound of the number of the PPA's outer iterations.

Theorem 1. *Let $\{(x^k, \lambda_k)\}$ be the sequence generated by the PPA with the stopping criterion (3). If T has a zero point $x^* \in \mathcal{H}$ and $\lambda_{k+1} \leq \lambda_k$, then*

$$\|x^k - (I + \lambda_k T)^{-1}(x^k)\|^2 \leq \mathcal{O}(1/k).$$

Moreover, the PPA can find an approximate zero point in at most $\mathcal{O}(1/\epsilon)$ outer iterations.

Proof. In view of the iterative formula (2), we have

$$T(x^{k+1}) \ni \lambda_k^{-1}(x^k - x^{k+1}).$$

Thus, by the assumption $T(x^*) \ni 0$ and the T's monotonicity, we can further get

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

which, together with $\lambda_k > 0$, implies that

$$\langle x^{k+1} - x^*,\ x^k - x^{k+1} \rangle \geq 0.$$

This shows that

$$\begin{aligned}
\|x^k - x^*\|^2 &= \|x^k - x^{k+1} + x^{k+1} - x^*\|^2 \\
&= \|x^k - x^{k+1}\|^2 + 2\langle x^k - x^{k+1},\ x^{k+1} - x^* \rangle + \|x^{k+1} - x^*\|^2 \\
&\geq \|x^k - x^{k+1}\|^2 + \|x^{k+1} - x^*\|^2
\end{aligned}$$

By the equivalent iterative relation $x^{k+1} := x^k - r^k$, we can further get

$$\|r^i\|^2 \leq \|x^i - x^*\|^2 - \|x^{i+1} - x^*\|^2,\ \ i \geq 0.$$

Summing up for $i = 0, 1, ..., k$ on both sides yields

$$\sum_{i=0}^{k} \|r^i\|^2 \leq \|x^0 - x^*\|^2 - \|x^{k+1} - x^*\|^2. \tag{4}$$

On the other side, it easily follows from Lemma 2.1 that the assumption $\lambda_{k+1} \leq \lambda_k$ implies

$$\|r^{k+1}\| \leq \|r^k\|.$$

Combining this with (4) yields

$$(k+1)\|r^k\|^2 \leq \sum_{i=0}^{k} \|r^i\|^2 \leq \|x^0 - x^*\|^2 \quad \Rightarrow \quad \|r^k\|^2 \leq \mathcal{O}(1/k).$$

So, the rest of the assertions directly follows from the stopping criterion (3) and this inequality above. □

Here we would like to emphasize that if the sequence $\{\|r^i\|\}$ were not monotonically decreasing then we would get from (4) that

$$\min_{i=0,1,...,k} \|r^i\|^2 \leq \mathcal{O}(1/k),$$

which is less interesting.

References

[1] Martinet B.: Regularisation d'inéquations variationelles par approximations successives. Revue Francaise d'Informatique et de Recherche Operationelle. **4**, 154-158 (1970)

[2] Rockafellar R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877-898 (1976)

[3] Moreau J.J.: Proximité et dualité dans un espace Hilbertien. Bulletin de la Société Mathématique de France. **93**, 273-299 (1965)

[4] Minty G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**, 341-346 (1962)

[5] 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)

[6] Luque F.J.: Asymptotic convergence analysis of the proximal point algorithm. SIAM J. Control Optim. **22**, 277-293, (1984)

[7] Pennanen T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**, 193-202 (2002)

[8] Dong Y.D.: A new relative error criterion for the proximal point algorithm. Nonlinear Anal. **64**, 2143-2148 (2006)

[9] Güler O.: On the convergence of the proximal point algorithm for convex minimization. SIAM J. Control Optim. **29**, 403-419 (1991)