New Advances in the Study of the Proximal Point Algorithm

Gheorghe Moroşanu

Department of Mathematics, Central European University, Nador u. 9, 1051 Budapest, Hungary

Abstract. Consider in a real Hilbert space H the inexact, Halpern-type, proximal point algorithm

$$x_{n+1} = \alpha_n u + (1-\alpha_n) J_{\beta_n} x_n + e_n, \ \ n = 0, 1, \dots, \qquad (H-PPA)$$

where $u, x_0 \in H$ are given points, $J_{\beta_n} = (I + \beta_n A)^{-1}$ for a given maximal monotone operator A, and (e_n) is the error sequence, under new assumptions on $\alpha_n \in (0,1)$ and $\beta_n \in (0, +\infty)$. Several strong convergence results for the $H-PPA$ are presented under the general condition that the error sequence converges strongly to zero, thus improving the classical Rockafellar's summability condition on $(\|e_n\|)$ that has been extensively used so far for different versions of the proximal point algorithm. Our results extend and improve some recent ones. These results can be applied to approximate minimizers of convex functionals. Convergence rate estimates are established for a sequence approximating the minimum value of such a functional.

Keywords: Halpern-type proximal point algorithm, monotone operator, strong convergence, convex function, minimizer, convergence rate.
PACS: 47J05, 47J25, 47H05, 47N10

INTRODUCTION

Let H be a real Hilbert space with scalar product $(\cdot,\cdot)$ and the Hilbertian norm $\|\cdot\|$. Let $A : D(A) \to 2^H$ be a maximal monotone operator (for details on such operators see, e.g., [10], Chapter I). An important topic in nonlinear analysis and optimization theory concerns iterative methods for solving nonlinear problems of the form:

$$\text{Find } x \in D(A), \ 0 \in Ax. \qquad (1)$$

One of the most popular iterative methods for solving such nonlinear equations is the so-called proximal point algorithm (PPA), which was introduced in 1970 by Martinet [8] and then developed systematically by R.T. Rockafellar. His investigations have stimulated much further work on the topic. The initial form of the PPA, as formulated by Rockafellar [11], was

$$x_{n+1} = J_{\beta_n} x_n + e_n, \ \ n = 0, 1, \dots, \qquad (R-PPA)$$

where $x_0 \in H$ is a given starting point, $J_{\beta_n} = (I + \beta_n A)^{-1}$ (the proximal mapping), with $\beta_n \in (0, +\infty)$ $(n = 0, 1, \dots)$, and (e_n) is the sequence of computational errors. In other words, x_{n+1} is an approximate solution of the equation

$$0 \in Ax + \beta_n^{-1}(x - x_n). \qquad (2)$$

Note that for a given x_n there is a unique solution x_{n+1} of equation $(R-PPA)$, so the sequence (x_n) is well defined. Rockafellar proved the weak convergence of algorithm $R-PPA$ for a general maximal monotone operator A assuming that:

the sequence (β_n) is bounded below away from zero, i.e.

$$\liminf \beta_n > 0; \qquad (3)$$

equation (1) has at least a solution, i.e.

$$\emptyset \neq F := A^{-1}0; \qquad (4)$$

and the sequence $(\|e_n\|)$ is summable, i.e.

$$\sum_{n=0}^{\infty} \|e_n\| < \infty. \qquad (5)$$

CP1281, *ICNAAM, Numerical Analysis and Applied Mathematics, International Conference 2010*
edited by T. E. Simos, G. Psihoyios, and Ch. Tsitouras

The weak limit of (x_n) is a point of F, i.e., a solution of problem (1).

As far as assumption (3) is concerned, it is acceptable. In fact, Rockafellar's result is valid under the weaker condition

$$\sum_{n=0}^{\infty} \beta_n^2 = \infty,$$

as proved by Brézis and Lions [5] for the exact $R-PPA$ (i.e., when all $e_n = 0$), and then by Moroşanu for $R-PPA$ with errors satisfying (5) (see [10], Chapter II, Section 3).

Condition (4) is a natural one, since our goal is to approximate solutions of equation (1). In fact, it was proved in [1] that if all $e_n = 0$, $\sum_{n=0}^{\infty} \beta_n = \infty$, and the sequence (x_n) generated by $R-PPA$ is bounded for some $x_0 \in H$, then F is necessarily nonempty. Note that (4) guarantees boundedness for all sequences (x_n) generated by $R-PPA$ provided that (5) holds.

Now, if A is the subdifferential of a proper, convex, lower semicontinuous (LSC) function $f: H \to (-\infty, +\infty]$, i.e. $A = \partial f$, then the set $F = A^{-1}0$ coincides with the set of all minimizers of f and so problem (1) can be interpreted as the minimization problem

$$\min_{x \in H} f(x). \tag{6}$$

Recall that many convex programming problems with or without constraints can be written in this form. In this case the $R-PPA$ is connected to the Moreau proximal mapping J_λ defined by (see [9])

$$J_\lambda x = \arg\min_{v \in H} \{f(v) + \frac{1}{2\lambda}\|v - x\|^2\} \ \forall x \in H, \lambda > 0, \tag{7}$$

and obviously $J_\lambda = (I + \lambda A)^{-1}$, thus justifying the name of the PPA.

Note that for the case $A = \partial f$ better results on the $R-PPA$ have been obtained by different authors. For example, the above result of Brézis and Lions holds under the weaker condition

$$\sum_{n=0}^{\infty} \beta_n = \infty,$$

as proved by the same authors.

Although the $R-PPA$ has been extensively used since its birth with good numerical results, though two main questions have arisen:

1. Rockafellar's open question: Is the $R-PPA$ strongly convergent?

2. Can we replace the summability condition (5) on errors by a weaker one, more acceptable from the computational point of view?

As far as the former question is concerned, Güler [6] constructed an example of A in $H = l^2$, even a subdifferential, showing that the $R-PPA$ fails to converge strongly in general. Fortunately, there have been proposed some modifications of the $R-PPA$ which generate sequences strongly convergent to solutions of problem (1) (see Xu [13], Solodov and Svaiter [12]). Among them, Xu's version is the simplest one. It has been recently extended slightly by Boikanyo and Moroşanu [1], namely

$$x_{n+1} = \alpha_n u + (1 - \alpha_n) J_{\beta_n} x_n + e_n, \ \ n = 0, 1, ..., \tag{$H-PPA$}$$

where $u, x_0 \in H$ are given points (Xu's version corresponds to $u = x_0$). We call it a Halpern-type algorithm (and denote it $H-PPA$) since it is similar to the old Halpern's iterative process associated with a map T that is nonexpanding on the unit ball [7].

Concerning Question 2, Boikanyo and Moroşanu [2] extended recently the strong convergence of the $H-PPA$ to general errors e_n satisfying the much weaker condition $\|e_n\| \to 0$, at the expense of choosing suitable control coefficients α_n depending on errors. Thus **Question 2 has a satisfactory positive answer.** In addition, several new results on the $H-PPA$ have been reported in [3] and [4]. Our assumptions on the coefficients α_n, β_n are weaker than the previous known conditions, and our results extend and improve some recent results such as those of Xu. These results can be used to approximate minimizers of convex functionals. Convergence rate estimates are established for a sequence approximating the minimum value of such a functional. Some extensions of $H-PPA$ are also investigated in [4]. We do not discuss these extensions here.

SELECTED RESULTS

We first introduce some assumptions which are frequently used in this section:

$$(A1)\ \ \alpha_n \to 0, \quad (A2)\ \ \sum_{n=0}^{\infty} \alpha_n = \infty; \quad (E1)\ \ \sum_{n=0}^{\infty} \|e_n\| < \infty, \quad (E2)\ \ \lim \frac{\|e_n\|}{\alpha_n} = 0;$$

Theorem 1 [2]. *Assume that A is maximal monotone, with* (4); $\alpha_n \in (0,1)$, $(A1)$, $(A2)$; *either* $(E1)$ *or* $(E2)$ *holds;* $\beta_n \to \infty$. *Then, for any fixed* $u, x_0 \in H$, *the sequence* (x_n) *generated by algorithm* $H-PPA$ *converges strongly to* $P_F u$, *the projection of* u *on* F.

Comments. The most important advance here as compared to the previous papers is the extension to errors satisfying $(E2)$. This allows one to also resolve the case $\sum_{n=0}^{\infty} \|e_n\| = +\infty$, with $\|e_n\| \to 0$, which is more adequate from a numerical point of view. We can do so by choosing suitable coefficients α_n, depending on errors, so that strong convergence of the $H-PPA$ is preserved. Indeed, one can choose, e.g., $\alpha_n = \|e_n\|^{1/2}$ for n large and $e_n \neq 0$, and $\|e_n\| = 1/(n+2)$ otherwise, thus Theorem 1 above works. So, we can answer positively Question 2 above if $\beta_n \to \infty$, or under some other conditions involving both α_n and β_n (see our results below), preserving at the same time strong convergence of the $H-PPA$.

Note that conditions $(E1)$, $(E2)$ are distinct, as simple examples show: $\|e_n\| = 1/n,\ \alpha_n = 1/\sqrt{n}$, and $\|e_n\| = 1/n^2,\ \alpha_n = 1/n + (-1)^n/(n+1)$.

One of the main steps in the proof of Theorem 1 (and of our next theorems as well) is the boundedness of (x_n) under the new condition $(E2)$. In fact, this conclusion holds under the weaker condition: $\limsup \|e_n\|/\alpha_n < \infty$, where $\alpha_n \in (0,1)$ for all n (no other assumption is needed) (see [2]).

Theorem 2 [3]. *Assume that A is maximal monotone, with* (4); $\alpha_n \in (0,1)$, $(A1)$, $(A2)$ *and*

$$\lim \frac{\alpha_{n+1} - \alpha_n}{\alpha_n^2} = 0;$$

either $(E1)$ *or* $(E2)$ *holds;* β_n *converges to some* $\beta \in (0,\infty)$. *Then, for any fixed* $u, x_0 \in H$, *the sequence* (x_n) *generated by algorithm* $H-PPA$ *converges strongly to* $P_F u$.

Remark. Our condition on (β_n) is weaker than Xu's conditions ([14], Thm. 3.3): $\liminf \beta_n > 0$, $\sum_{n=0}^{\infty} |\beta_{n+1} - \beta_n| < \infty$. As an example, for any $\beta > 0$, the sequence $\beta_n = \beta + (-1)^n/n$ for n large, satisfies our condition, but not Xu's summability condition.

Theorem 3 [4]. *Assume that A is maximal monotone, with* (4); $\alpha_n \in (0,1)$, $(A1)$, $(A2)$; (β_n) *is an increasing sequence of positive numbers;*

$$\sum_{n=1}^{\infty} \frac{|\alpha_n - \alpha_{n-1}|}{\beta_n} < \infty;$$

either $(E1)$ *or* $(E2)$ *holds.Then, for any fixed* $u, x_0 \in H$, *the sequence* (x_n) *generated by algorithm* $H-PPA$ *converges strongly to* $P_F u$.

Theorem 4 [4]. *Assume that A is maximal monotone, with* (4); $\alpha_n \in (0,1)$, $(A1)$, $(A2)$; $\liminf \beta_n > 0$, *and*

$$\text{either } \sum_{n=1}^{\infty} \left|\frac{\alpha_n}{\beta_{n+1}} - \frac{\alpha_{n-1}}{\beta_n}\right| < \infty \quad \text{or} \quad \lim \frac{1}{\beta_{n+1}}\left(\frac{\alpha_n}{\alpha_{n-1}} - \frac{\beta_{n+1}}{\beta_n}\right) = 0.$$

Then, for any fixed $u, x_0 \in H$, the sequence (x_n) generated by algorithm $H-PPA$ converges strongly to $P_F u$.

Setting $v_n = (1-\alpha_{n-1})^{-1}(x_n - \alpha_{n-1}u - e_{n-1})$, the $H-PPA$ becomes

$$v_{n+1} = J_{\beta_n}(\alpha_{n-1}u + (1-\alpha_{n-1})v_n + e_{n-1}),\ \ n = 1,2,... \qquad (X-PPA)$$

This algorithm has been investigated by Xu [14] (that is why we call it $X-PPA$). In fact it is "equivalent" to $H-PPA$ provided that $\alpha_n \to 0$ and $\|e_n\| \to 0$, as proved in [2]. Sometimes it is convenient to use $X-PPA$ or to combine $H-PPA$ and $X-PPA$ to derive convergence results.

If $A = \partial f$, where $f : H \longrightarrow (-\infty, +\infty]$ is proper, convex and LSC, then the above results can be used to approximate minimizers of f. In addition, new results can be obtained in this particular case under weaker assumptions, such as for example

Theorem 5 [3]. *Let f be as above and $A = \partial f$. Assume that (4) holds, i.e., f has at least a minimizer. For any fixed $u, v_1 \in H$, let (v_n) be the sequence generated by algorithm $X-PPA$ and let*

$$w_n := \frac{1}{\sigma_n}\sum_{k=1}^{n}\beta_k v_{k+1},\ \ where\ \ \sigma_n = \sum_{k=1}^{n}\beta_k.$$

If (E_1) holds, then we have the estimate

$$f(w_n) - f(z) \le \frac{\|v_1 - z\|^2 + K\sum_{k=1}^{n}(\alpha_{k-1} + \|e_{k-1}\|)}{2\sigma_n}\ \ \forall z \in H,$$

for some constant $K > 0$. If (E_2) holds, we have

$$f(w_n) - f(z) \le \frac{\|v_1 - z\|^2 + M\sum_{k=1}^{n}\alpha_{k-1}}{2\sigma_n}\ \ \forall z \in H,$$

for some constant $M > 0$. If, in addition,

$$\sigma_n \to \infty \quad and \quad \frac{\sum_{k=1}^{n}\alpha_{k-1}}{\sigma_n} \to 0\ \ as\ \ n \to \infty,$$

then

$$f(w_n) \to inf_H f.$$

REFERENCES

1. O. A. Boikanyo, and G. Moroşanu, Modified Rockafellar's algorithms, *Math. Sci. Res. J.* **13** (5) (2009), 101–122.
2. O. A. Boikanyo, and G. Moroşanu, A proximal point algorithm converging strongly for general errors, *Optim. Letters*, in press.
3. O. A. Boikanyo, and G. Moroşanu, Inexact Halpern-type proximal point algorithm, submitted.
4. O. A. Boikanyyo and G. Moroşanu, paper in preparation.
5. H. Brézis, and P. L. Lions, Produits infinis de resolvantes, *Israel J. Math.* **29** (1978), 329–345.
6. O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.* **29** (1991), 403–419.
7. B. Halpern, Fixed points of nonexpanding mappings, *Bull. Amer. Math. Soc.* **73** (1967), 957–961.
8. B. Martinet, Régularisation d′inéquations variationnelles par approximation successives, *Rev. Française Informat. Recherche Opérationnelle* **4** (1970), Ser. R-3, 154–158.
9. J. J. Moreau, Proximité et dualité dans un espace Hilbertien, *Bull. Soc. Math. France* **93** (1965), 273–299.
10. G. Moroşanu, *Nonlinear Evolution Equations and Applications*, Reidel, Dordrecht, 1988.
11. R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* **14** (1976), 877–898.
12. M. V. Solodov, and B. F. Svaiter, Forcing strong convergence of proximal point interations in a Hilbert space, *Math. Program.*, Ser. **A 87** (2000), 189–202.
13. H. K. Xu, Iterative algorithms for nonlinear operators, *J. London Math. Soc.* **66** (2) (2002), 240–256.
14. H. K. Xu, A regularization method for the proximal point algorithm, *J. Global. Optim.* **36** (2006), 115–125.