Proximal Point Methods Revisited

Oganeditse A. Boikanyo and Gheorghe Moroşanu

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

Abstract. The proximal point methods have been widely used in the last decades to approximate the solutions of nonlinear equations associated with monotone operators. Inspired by the iterative procedure defined by B. Martinet (1970), R.T. Rockafellar introduced in 1976 the so-called proximal point algorithm (PPA) for a general maximal monotone operator. The sequence generated by this iterative method is weakly convergent under appropriate conditions, but not necessarily strongly convergent, as proved by O. Güler (1991). This fact explains the introduction of different modified versions of the PPA which generate strongly convergent sequences under appropriate conditions, including the contraction-PPA defined by H.K. Xu in 2002. Here we discuss Xu's modified PPA as well as some of its generalizations. Special attention is paid to the computational errors, in particular the original Rockafellar summability assumption is replaced by the condition that the error sequence converges to zero strongly.

Keywords: proximal point algorithm, monotone operator, strong convergence, the method of alternating resolvents.
PACS: 47J05, 47J25, 47H05, 47N10

INTRODUCTION

Let H be a real Hilbert space equipped with scalar product $(\cdot,\cdot)$ and the associated Hilbertian norm $\|\cdot\|$. Let $A : D(A) \subset H \to 2^H$ be a maximal monotone operator (see, e.g., [13] for details on monotone operators). This condition on A will be assumed throughout this Note without explicit mention. We are interested in solving iteratively nonlinear operator equations (inclusions) of the form:

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

The study of iterative methods is an important topic in nonlinear analysis and optimization theory. Note that if A is the subdifferential of a proper, convex and lower semicontinuous function $f : H \to (-\infty, +\infty]$, then x is a solution of equation (1) if and only if x is a minimizer of f. Therefore equation (1) is intimately connected to convex programming problems. Special attention has been paid in the last decades to the so-called proximal point algorithm (PPA). The PPA was introduced in 1970 by B. Martinet [11] and then generalized and developed systematically by R.T. Rockafellar. The initial form of the PPA for a general maximal monotone operator A, as formulated by Rockafellar [14], was

$$x_{n+1} = J_{\beta_n} x_n + e_n,\ \ n = 0, 1, \ldots, \tag{$R-PPA$}$$

where $x_0 \in H$ is a given starting point (initial guess), $\beta_n \in (0, +\infty)$ $(n = 0, 1, \ldots)$, $J_{\beta_n} = (I + \beta_n A)^{-1}$ (the resolvent of A, which coincides with the Moreau proximal mapping if A is the subdifferential of a convex functional f), and (e_n) is the sequence of computational errors. Obviously, the (difference equation) R-PPA has a unique solution (x_n) for each $x_0 \in H$ since J_β is everywhere defined for all $\beta > 0$. The idea to formulate the $R-PPA$ was most probably suggested by the following simple remark: x is a solution of equation (1) if and only if x is a fixed point of J_β for every (equivalently, for a given) $\beta > 0$. Indeed, the $R-PPA$ reminds of the classical method of successive approximations. The seminal Rockafellar's work on the PPA has generated many other contributions on the topic. Some of them are discussed below.

Rockafellar proved in [14] that for every (initial guess) $x_0 \in H$ the sequence (x_n) generated by algorithm $R-PPA$ is weakly convergent provided that:

$$\liminf \beta_n > 0,\ \ A^{-1}(0) =: F \neq \emptyset,\ \text{ and } \sum_{n=0}^{\infty} \|e_n\| < \infty.$$

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

Numerical Analysis and Applied Mathematics ICNAAM 2011
AIP Conf. Proc. 1389, 893-896 (2011); doi: 10.1063/1.3636878

Rockafellar asked himself whether (x_n) is strongly convergent. This question remained open until 1991 when Güler [7] constructed an example of A in $H = l^2$, even a subdifferential, showing that the sequence (x_n) generated by the $R-PPA$ does not converge strongly in general. This counterexample generated much work on modifying the $R-PPA$ in order to achieve strong convergence of the corresponding sequence (see [9], [15], [16] and the references therein). The simplest modification of the R-PPA was introduced by Xu in 2002. In the next section we discuss this modified PPA as well as some of its generalizations.

RECENT RELATED RESULTS

In [16] Xu introduced the following modified PPA, called the contraction-PPA,

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

where $u, x_0 \in H$ are given points, $\alpha_n \in (0,1)$, $\beta_n > 0$ (in fact, in the original Xu's version $u = x_0$, but the above extension to an arbitrary u works similarly).

The following sufficient conditions have been extensively used

$$(A1) \;\; \alpha_n \to 0, \qquad (A2) \;\; \sum_{n=0}^{\infty} \alpha_n = \infty;$$

$$(B) \;\; \liminf \beta_n > 0;$$

$$\text{either} \;\; (E1) \;\; \sum_{n=0}^{\infty} \|e_n\| < \infty \;\; \text{or} \;\; (E2) \;\; \lim \frac{\|e_n\|}{\alpha_n} = 0,$$

plus various additional assumptions on (α_n) and (β_n), to derive the strong convergence of (x_n) generated by the $X-PPA$ (see Boikanyo and Morosanu [1], Marino and Xu [10], Xu [16] and the references therein). The strong limit of x_n is $P_F u$, the projection of u on F. (Note that F is a closed and convex subset of H since A (equivalently, A^{-1}) is a maximal monotone operator, thus $P_F u$ is well defined.) So in the case of strong convergence we have a precise characterization of the limit point. Similar results have been obtained for the following more general algorithm introduced by Yao and Noor [18]

$$x_{n+1} = \alpha_n u + \lambda_n x_n + \gamma_n J_{\beta_n} x_n + e_n, \;\; n = 0, 1, ..., \qquad (YN-PPA)$$

where $\alpha_n, \lambda_n, \gamma_n \in (0,1)$, with $\alpha_n + \lambda_n + \gamma_n = 1$. See Yao and Noor [18], Boikanyo and Moroşanu [2].

Recently, Wang and Cui [19] proved the following result which is better than all the previous related results:

Theorem [19]. *Assume that $A^{-1}0 =: F \neq \emptyset$; $\alpha_n \in (0,1)$, $(A1)$, $(A2)$; either $(E1)$ or $(E2)$ holds; both (B) and a similar condition for (γ_n) hold. Then, for any fixed $u, x_0 \in H$, the sequence (x_n) generated by algorithm $YN-PPA$ converges strongly to $P_F u$.*

Remark 1. Condition $(E2)$ was introduced by the authors in [1] in order to cover the case when $(E1)$ is not satisfied. In fact, it is enough to assume that $\|e_n\| \to 0$. Indeed, if $\sum_{n=0}^{\infty} \|e_n\| = \infty$, then one can choose appropriate α_n's such that condition $(E2)$ is fulfilled. Obviously, the new condition on errors, $\|e_n\| \to 0$, is more convenient from the computational point of view than Rockafellar's summability condition $(E1)$.

Remark 2. The above theorem includes the case $\lambda_n = 0$, so we have a nice result for the $X-PPA$, which is in fact equivalent with the following iterative scheme introduced in 2006 by Xu [17] (motivated by Lehdili and Moudafi [9]),

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

with an initial guess $x_0 \in H$. Thus the recent result obtained by Wang and Cui includes many previous contributions.

On the other hand, in [4] we pursue a different direction to derive a strong convergence result for the sequence generated by the algorithm

$$x_{n+1} = J_{\beta_n}(\alpha_n u + (1-\alpha_n)(x_n + e_n)), \;\; n = 0, 1, ..., \qquad (BM)$$

which is equivalent with $LMX - PPA$ if $\alpha_n \to 0$. Here we use an alternative condition: $\alpha_n \to 1$. More precisely, we have

Theorem [4]. *Assume that $A^{-1}0 =: F \neq \emptyset$; $\alpha_n \in (0,1)$, $\alpha_n \to 1$; $0 < \beta_n \to \infty$, and (e_n) is a bounded sequence. Then, for any $u, x_0 \in H$, the sequence (x_n) generated by algorithm BM converges strongly to $P_F u$.*

It is worth pointing out that the result holds for bounded errors, $\|e_n\| \leq \varepsilon$ for some $\varepsilon > 0$, which is very convenient from the computational point of view. However, it remains to check whether the above iterative process *BM* works well numerically.

Now, we are going to discuss the method of alternating resolvents which is a generalization of the R-PPA. More precisely, if A and B are two maximal monotone operators, we consider the algorithm

$$\begin{aligned} x_{2n+1} &= J^A_{\beta_n}(x_{2n} + e_n) \quad \text{for } n = 0, 1, \ldots, \\ x_{2n} &= J^B_{\mu_n}(x_{2n-1} + e'_n) \quad \text{for } n = 1, 2, \ldots, \end{aligned}$$

where x_0 is a given starting point, β_n and μ_n are positive numbers, while e_n, e'_n denote the computational errors. A particular case is the Neumann method of alternating projections,

$$H \ni x_0 \mapsto x_1 = P_{K_1}x_0 \mapsto x_2 = P_{K_2}x_1 \mapsto x_3 = P_{K_1}x_2 \mapsto x_4 = P_{K_2}x_3 \mapsto \cdots,$$

where K_1, K_2 are vector subspaces of H, which converges strongly to the point of $K_1 \cap K_2$ that is closest to the starting point x_0. In the case when K_1 and K_2 are arbitrary, closed and convex sets with nonempty intersection, Bregman [6] proved that the sequence (x_n) generated by the method of alternating projections converges weakly to a point in $K_1 \cap K_2$ which is nearest to x_0. Recently, Hundal [8] constructed an example in ℓ^2 showing that for any starting point $x_0 \in \ell^2$, there exist a hyperplane K_1 and a cone K_2, with $K_1 \cap K_2 = \{0\}$, such that the sequence of alternated projections (x_n) converges weakly to zero, but not strongly; see also Matouśková and Reich [12]. Therefore, we have the same problem for the method of alternating resolvents: (x_n) is not necessarily strongly convergent. In order to enforce strong convergence, one can use the same idea in modifying the algorithm of alernating resolvents. Specifically, in [5] we obtain the strong convergence of the sequence (x_n) generated by

$$x_{2n+1} = J^A_{\beta_n}(\alpha_n u + (1-\alpha_n)x_{2n} + e_n) \quad \text{for } n = 0, 1, \ldots, \tag{2}$$

$$x_{2n} = J^B_{\mu_n}(\lambda_n u + (1-\lambda_n)x_{2n-1} + e'_n) \quad \text{for } n = 1, 2, \ldots, \tag{3}$$

where $\alpha_n, \lambda_n \in (0,1)$, $\alpha_n \to 0$, $\lambda_n \to 0$; either $\sum_{n=0}^{\infty} \alpha_n = \infty$ or $\sum_{n=0}^{\infty} \lambda_n = \infty$; both (β_n) and (μ_n) are bounded from below away from zero, with

$$\lim_{n\to\infty}\left(1 - \frac{\beta_{n+1}}{\beta_n}\right) = 0 \quad \text{and} \quad \lim_{n\to\infty}\left(1 - \frac{\mu_{n+1}}{\mu_n}\right) = 0; \tag{$*$}$$

$A^{-1}0 \cap B^{-1}0 =: F \neq \emptyset$, and the error sequences satisfy appropriate conditions (basically it is enough to assume that both of them converge strongly to zero, if α_n's and λ_n's are appropriately adjusted; for example, if $\sum \|e_n\| = \infty$ and $\sum \|e'_n\| = \infty$, the algorithm works if $\|e_n\|/\alpha_n \to 0$ and $\|e'_n\|/\lambda_n \to 0$ and these conditions are satisfied for suitable α_n's and λ_n's; and there are many good sufficient conditions on errors and control parameters, at least fourteen in total, guaranteeing the functionality of the algorithm). The strong limit of (x_n) is precisely the point of F that is nearest to u.

We suspect that our result above remains valid without condition $(*)$ (see above), as in the case of a single operator (cf. [19]).

REFERENCES

1. O. A. Boikanyo and G. Moroşanu, A proximal point algorithm converging strongly for general errors, *Optim. Letters* **4** (2010), 635–641.
2. O. A. Boikanyo and G. Moroşanu, Four parameter proximal point algorithms, *Nonlinear Anal.* **74** (2011), 544–555.
3. O. A. Boikanyo and G. Moroşanu, Inexact Halpern-type proximal point algorithm, *J. Global Optimization*, to appear.

4. O. A. Boikanyo and G. Moroşanu, Strong convergence of a proximal point algorithm with bounded error sequence, submitted.
5. O. A. Boikanyo and G. Moroşanu, Strong convergence of the method of alternating resolvents, in preparation.
6. L. M. Bregman, The method of successive projection for finding a common point of convex sets, *Sov. Math. Dokl.* **6** (1965), 688–692.
7. O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.* **29** (1991), 403–419.
8. H. Hundal, An alternating projection that does not converge in norm, *Nonlinear Anal.* **57** (2004), no. 1, 35–61.
9. N. Lehdili and A. Moudafi, Combining the proximal point algorithm and Tikhonov method, *Optimization* **37** (1996), 239–252.
10. G. Marino and H.K. Xu, Convergence of generalized proximal point algorithm, *Comm. Pure Appl. Anal.* **3** (2004), 791–808.
11. B. Martinet, Régularisation d′inéquations variationnelles par approximation successives, *Rev. Française Informat. Recherche Opérationnelle* **4** (1970), Ser. R-3, 154–158.
12. E. Matoušková and S. Reich, The Hundal example revisited, *J. Nonlinear Convex Anal.* **4** (2003), 411–427.
13. G. Moroşanu, *Nonlinear Evolution Equations and Applications*, Reidel, Dordrecht, 1988.
14. R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.* **14** (1976), 877–898.
15. 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.
16. H. K. Xu, Iterative algorithms for nonlinear operators, *J. London Math. Soc.* **66** (2) (2002), 240–256.
17. H. K. Xu, A regularization method for the proximal point algorithm, *J. Global. Optim.* **36** (2006), 115–125.
18. Y. Yao and M.A. Noor, On convergence criteria of generalized proximal point algorithms, *J. Comput. Appl. Math.* **217** (2008), 46–55.
19. F. Wang and H. Cui, On the contraction-proximal point algorithms with multi-parameters, *J. Global Optimization*, to appear.