Convergence of the stochastic cyclic proximal algorithm
Sketch of the proof

Adil Salim, Pascal Bianchi, Walid Hachem and Jérémie Jakubowicz

May 3, 2016

Given a family of maximal monotone operators $\{A_i\}_{i=1}^L$ on a real Hilbert space, Passty [5] studied among other things the iterative algorithm

$$x_{n+1} = (I + \gamma_{n+1} A_L)^{-1} \cdots (I + \gamma_{n+1} A_2)^{-1} (I + \gamma_{n+1} A_1)^{-1} x_n$$

where (γ_n) is a sequence of step sizes belonging to $\ell^2 \setminus \ell^1$. When the operator $\sum_1^L A_i$ is maximal and when $Z(\sum_1^L A_i) = (\sum_1^L A_i)^{-1}(\{0\})$ is non empty, he showed that the sequence of weighted averages

$$\bar{x}_n = \frac{\sum_1^n \gamma_i x_i}{\sum_1^n \gamma_i}$$

converges towards an element of $Z(\sum_1^L A_i)$.

In these notes, we replace the operators A_i with random maximal monotone operators $A_i(\xi, \cdot)$ on the Euclidean space $\mathbb{R}^N$, where ξ is a random variable on some probability space $(\Xi, \mathscr{T}, \mu)$ (a rigorous definition of these operators is provided in [3]), and we consider the algorithm

$$x_{n+1} = (I + \gamma_{n+1} A_L(\xi_{n+1}, \cdot))^{-1} \cdots (I + \gamma_{n+1} A_2(\xi_{n+1}, \cdot))^{-1} (I + \gamma_{n+1} A_1(\xi_{n+1}, \cdot))^{-1} x_n$$

where (ξ_n) is a sequence of independent copies of ξ. In the manuscript [6], this algorithm is referred to as the *Stochastic Passty's Algorithm*. For any of the operators $A_i(s, \cdot)$, let

$$\mathcal{A}_i(\cdot) = \int A_i(s, \cdot)\, \mu(ds)$$

and assume that the monotone operator $\sum \mathcal{A}_i$ is maximal[1]. Still considering that $(\gamma_n) \in \ell^2 \setminus \ell^1$ and assuming that $\mathcal{Z} = Z(\sum_1^L \mathcal{A}_i)$ is nonempty, the purpose of these notes is to show that under some mild conditions, the sequence (x_n) converges almost surely to a random variable U supported by $\mathcal{Z}$.

In these notes, we shall confine ourselves to the case $L = 2$. For $L > 2$, the proofs are tedious without being essentially different from the case $L = 2$. Let us define the sets $\mathcal{S}_i$ for $i = 1, 2$ similarly as in [3] (in general, we shall reuse the notations of that paper, just adding the index i to specify which operator we are dealing with).

The following proposition is the analogue of [2, Prop. 1] or [3, Prop. 2].

[1]The issue of the maximality of the $\mathcal{A}_i$ was studied in [3]. The maximality of a finite sum of maximal monotone operators is a classical question in monotone operator theory, see [4] or [1].

Proposition 0.1. *Assume that $\mathcal{Z} \neq \emptyset$. Assume moreover that there exists $x_\star \in \mathcal{Z}$ such that the set*

$$\mathcal{R}_2(x_\star) = \Big\{ (\varphi_1, \varphi_2) \in \mathcal{S}^2_{A_1(\cdot, x_\star)} \times \mathcal{S}^2_{A_2(\cdot, x_\star)} \ : \ \int (\varphi_1 + \varphi_2) \, d\mu = 0 \Big\}$$

is not empty. Then

1. *The sequence (x_n) is bounded almost surely and in $\mathcal{L}^2$.*
2. *It holds that*

$$\mathbb{E}\Big[\sum_n \gamma_n^2 \int \Big(\| A_{1,\gamma_{n+1}}(s, x_n) \|^2 + \| A_{2,\gamma_{n+1}}(s, x_n - \gamma_{n+1} A_{1,\gamma_{n+1}}(s, x_n)) \|^2 \Big) \, \mu(ds) \Big] < \infty.$$

3. *The sequence $(\|x_n - x_\star\|)_n$ converges almost surely.*

Proof. Observing that

$$\begin{aligned} x_{n+1} &= J_{2,\gamma_{n+1}}(\xi_{n+1}, J_{1,\gamma_{n+1}}(\xi_{n+1}, x_n)) \\ &= J_{2,\gamma_{n+1}}(\xi_{n+1}, x_n - \gamma_{n+1} A_{1,\gamma_{n+1}}(\xi_{n+1}, x_n)) \\ &= x_n - \gamma_{n+1} A_{1,\gamma_{n+1}}(\xi_{n+1}, x_n) - \gamma_{n+1} A_{2,\gamma_{n+1}}(\xi_{n+1}, x_n - \gamma_{n+1} A_{1,\gamma_{n+1}}(\xi_{n+1}, x_n)), \end{aligned}$$

writing

$$\begin{aligned} \xi = \xi_{n+1}, \quad \gamma = \gamma_{n+1}, \quad A_{1,\gamma} = A_{1,\gamma_{n+1}}(\xi_{n+1}, x_n) \\ \text{and} \quad A_{2,\gamma} = A_{2,\gamma_{n+1}}(\xi_{n+1}, x_n - \gamma_{n+1} A_{1,\gamma_{n+1}}(\xi_{n+1}, x_n)) \end{aligned}$$

and expanding

$$\|x_{n+1} - x_\star\|^2 = \|x_n - x_\star\|^2 + 2\langle x_{n+1} - x_n, x_n - x_\star \rangle + \|x_{n+1} - x_n\|^2 ,$$

we obtain

$$\begin{aligned} \|x_{n+1} - x_\star\|^2 &= \|x_n - x_\star\|^2 - 2\gamma \langle A_{1,\gamma}, x_n - x_\star \rangle - 2\gamma \langle A_{2,\gamma}, x_n - x_\star \rangle + \gamma^2 \|A_{1,\gamma} + A_{2,\gamma}\|^2 \\ &= \|x_n - x_\star\|^2 - 2\gamma X_1 - 2\gamma X_2 + \gamma^2 X_3. \end{aligned}$$

Considering the functions φ_1 and φ_2 specified in the statement, and writing $J_{1,\gamma} = J_{1,\gamma_{n+1}}(\xi_{n+1}, x_n))$ and $J_{2,\gamma} = J_{2,\gamma_{n+1}}(\xi_{n+1}, x_n - \gamma_{n+1} A_{1,\gamma_{n+1}}(\xi_{n+1}, x_n))$, we have

$$\begin{aligned} X_1 &= \langle A_{1,\gamma} - \varphi_1(\xi), J_{1,\gamma} - x_\star \rangle + \gamma \langle A_{1,\gamma} - \varphi_1(\xi), A_{1,\gamma} \rangle + \langle \varphi_1(\xi), x_n - x_\star \rangle \\ &\geq \gamma \|A_{1,\gamma}\|^2 - \gamma \langle \varphi_1(\xi), A_{1,\gamma} \rangle + \langle \varphi_1(\xi), x_n - x_\star \rangle \end{aligned}$$

by the monotonicity of $A_1(s, \cdot)$, and

$$\begin{aligned} X_2 &= \langle A_{2,\gamma} - \varphi_2(\xi), J_{2,\gamma} - x_\star \rangle + \gamma \langle A_{2,\gamma} - \varphi_2(\xi), A_{2,\gamma} \rangle + \gamma \langle A_{2,\gamma} - \varphi_2(\xi), A_{1,\gamma} \rangle + \langle \varphi_2(\xi), x_n - x_\star \rangle \\ &\geq \gamma \langle A_{2,\gamma} - \varphi_2(\xi), A_{2,\gamma} \rangle + \gamma \langle A_{2,\gamma} - \varphi_2(\xi), A_{1,\gamma} \rangle + \langle \varphi_2(\xi), x_n - x_\star \rangle \\ &= \gamma \|A_{2,\gamma}\|^2 - \gamma \langle \varphi_2(\xi), A_{2,\gamma} \rangle + \gamma \langle A_{2,\gamma}, A_{1,\gamma} \rangle - \gamma \langle \varphi_2(\xi), A_{1,\gamma} \rangle + \langle \varphi_2(\xi), x_n - x_\star \rangle \end{aligned}$$

by the monotonicity of $A_2(s,\cdot)$. By expanding the term X_3, we obtain altogether

$$\begin{aligned}\|x_{n+1}-x_\star\|^2 \le \|x_n-x_\star\|^2 &- \gamma^2(\|A_{1,\gamma}\|^2+\|A_{2,\gamma}\|^2) + 2\gamma^2\langle\varphi_1(\xi)+\varphi_2(\xi),A_{1,\gamma}\rangle + 2\gamma^2\langle\varphi_2(\xi),A_{2,\gamma}\rangle \\ &- 2\gamma\langle\varphi_1(\xi)+\varphi_2(\xi),x_n-x_\star\rangle \\ \le \|x_n-x_\star\|^2 &- \gamma^2(1-\beta^{-1})(\|A_{1,\gamma}\|^2+\|A_{2,\gamma}\|^2) + \gamma^2\beta(\|\varphi_1(\xi)\|^2+\|\varphi_1(\xi)+\varphi_2(\xi)\|^2) \\ &- 2\gamma\langle\varphi_1(\xi)+\varphi_2(\xi),x_n-x_\star\rangle\end{aligned}$$

where we used the inequality $|\langle a,b\rangle| \le (\beta/2)\|a\|^2+\|b\|^2/(2\beta)$, with $\beta>0$ being otherwise arbitrary.

By assumption,

$$\int (\|\varphi_1(s)\|^2+\|\varphi_1(s)+\varphi_2(s)\|^2)\,\mu(ds)<\infty.$$

Moreover $\mathbb{E}_n\langle\varphi_1(\xi_{n+1})+\varphi_2(\xi_{n+1}),x_n-x_\star\rangle=0$. Thus,

$$\begin{aligned}\mathbb{E}_n\|x_{n+1}-x_\star\|^2 \le \|x_n-x_\star\|^2 \\ &- \gamma_{n+1}^2(1-\beta^{-1})\int\Big(\|A_{1,\gamma_{n+1}}(s,x_n)\|^2+\|A_{2,\gamma_{n+1}}(s,x_n-\gamma_{n+1}A_{1,\gamma_{n+1}}(s,x_n))\|^2\Big)\,\mu(ds) \\ &+ C\gamma_{n+1}^2.\end{aligned}$$

Choose $\beta>1$. By the supermartingale convergence theorem along with the assumptions $(\gamma_n)\in\ell^2$, the conclusions follow. □

References

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011.

[2] P. Bianchi. Ergodic convergence of a stochastic proximal point algorithm. *ArXiv e-prints, 1504.05400*, April 2015.

[3] P. Bianchi and W. Hachem. Dynamical behavior of a stochastic forward-backward algorithm using random monotone operators. *arXiv preprint arXiv:1508.02845*, 2015.

[4] H. Brézis. *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert.* North-Holland mathematics studies. Elsevier Science, Burlington, MA, 1973.

[5] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.*, 72(2):383–390, 1979.

[6] A. Salim, P. Bianchi, W. Hachem, and J. Jakubowicz. A stochastic proximal point algorithm for total variation regularization over large scale graphs. *submitted to the IEEE Conference on Decision and Control*, 2016.