Linear convergence of stochastic block-coordinate fixed point algorithms

Patrick L Combettes, Jean-Christophe Pesquet

LINEAR CONVERGENCE OF STOCHASTIC BLOCK-COORDINATE FIXED POINT ALGORITHMS

Patrick L. Combettes[1] *and Jean-Christophe Pesquet*[2]

[1] Department of Mathematics, North Carolina State University, Raleigh, NC, USA
plc@math.ncsu.edu
[2] Center for Visual Computing, INRIA, CentraleSupélec, Université Paris-Saclay, Gif sur Yvette, France
jean-christophe@pesquet.eu

ABSTRACT

Recent random block-coordinate fixed point algorithms are particularly well suited to large-scale optimization in signal and image processing. These algorithms feature random sweeping rules to select arbitrarily the blocks of variables that are activated over the course of the iterations and they allow for stochastic errors in the evaluation of the operators. The present paper provides new linear convergence results. These convergence rates are compared to those of standard deterministic algorithms both theoretically and experimentally in an image recovery problem.

Index Terms— Block-coordinate algorithm, fixed-point algorithm, linear convergence, stochastic algorithm

1. INTRODUCTION

Many algorithms used in applied mathematics and in signal processing rely on fixed point methods. Such a method typically generates a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ in some underlying real Hilbert space $\mathbf{H}$ via the iterative scheme

$$\begin{aligned}&\text{for } n = 0, 1, \dots\\ &\lfloor\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n\big(\mathbf{T}_n\boldsymbol{x}_n - \boldsymbol{x}_n\big),\end{aligned} \tag{1}$$

where $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence of relaxation parameters in $[0, +\infty[$ and $(\mathbf{T}_n)_{n\in\mathbb{N}}$ is a sequence of operators from $\mathbf{H}$ to $\mathbf{H}$. Under suitable assumptions on the relaxation parameters and the operators, the sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges to a point in the intersection $\mathbf{F}$ of the fixed point sets $(\operatorname{Fix}\mathbf{T}_n)_{n\in\mathbb{N}}$ [2,6–8].

In recent years, an increasing challenge in data analysis has been to process massive data sets, especially in the field of inverse problems. In high dimensional applications, the implementation of (1) may raise serious computational issues. In particular, it may be too demanding in terms of memory requirements. An efficient strategy for overcoming this limitation consists of splitting the variables in (1) into m blocks and to update only some of them at each iteration, while leaving the others unchanged. More specifically, let us assume that $\mathbf{H}$ is decomposed into the direct Hilbert sum

$$\mathbf{H} = \mathsf{H}_1 \oplus \cdots \oplus \mathsf{H}_m \tag{2}$$

where, for every $i \in \{1, \dots, m\}$, H_i is a real Hilbert space. Then, for every $n \in \mathbb{N}$ and every $i \in \{1, \dots, m\}$, the i-th block of $\boldsymbol{x}_n$, denoted by $x_{i,n} \in \mathsf{H}_i$ can be updated or remain unchanged. In this context, a block-coordinate approach aims at devising efficient block update rules while guaranteeing convergence. The extension of the convergence results existing for general fixed point algorithms to block-coordinate forms is however quite delicate. In [9], a probabilistic framework was developed for designing block-coordinate fixed point algorithms that generate provably convergent sequences and in which the blocks are activated in a random manner. Other block-coordinate methods focused on specialized minimization problems, and featuring possibly weaker types of convergence, include [1, 5, 12–14].

In this paper, a block-coordinate extension of (1) is investigated, in which the blocks are activated randomly. Conditions ensuring the almost sure convergence of the iterates are provided. We also show that, under a strict quasinonexpansiveness assumption, mean square convergence is obtained. The same assumption allows us to derive linear convergence results. One important practical question in this context is to assess the impact of the block decomposition on the speed of convergence. Our work provides theoretical elements to answer this question.

In Section 2, we present our stochastic block-coordinate fixed point algorithm. In Section 3, we provide general convergence results for this algorithm. In Section 4, we further investigate the mean square behavior of the algorithm over a finite number of iterations. This allows us to compare block-coordinate approaches and non-coordinate ones in terms of linear convergence rate. We also examine the impact of a stochastic error in the computation of the involved operators. In Section 5, we illustrate our theoretical results with experiments in multicomponent image recovery [3].

[0]The work of P. L. Combettes was supported by the National Science Foundation under grant CCF-1715671.

2. BLOCK-COORDINATE FIXED POINT ALGORITHM

To avoid technicalities, the underlying space $\mathbf{H}$ is assumed to be finite dimensional throughout this paper. Our results can however be extended to the infinite dimensional case [11].

For every $n \in \mathbb{N}$, the operator $\mathbf{T}_n \colon \mathbf{H} \to \mathbf{H}$ is block-decomposed as

$$(\forall \mathbf{x} \in \mathbf{H}) \quad \mathbf{T}_n \mathbf{x} = (\mathsf{T}_{i,n}\, \mathbf{x})_{1 \leqslant i \leqslant m} \tag{3}$$

where, for every $i \in \{1, \ldots, m\}$, $\mathsf{T}_{i,n} \colon \mathbf{H} \to \mathsf{H}_i$ is measurable. The proposed block-coordinate algorithm is as follows.

Algorithm 2.1 Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ and set $\mathsf{D} = \{0,1\}^m \smallsetminus \{\mathbf{0}\}$. Let $\boldsymbol{x}_0$ and $(\boldsymbol{a}_n)_{n\in\mathbb{N}} = (a_{i,n})_{1\leqslant i\leqslant n}$ be $\mathbf{H}$-valued random variables, and let $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ be identically distributed D-valued random variables. Iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \quad \text{for } i = 1, \ldots, m \\ \quad\quad x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\lambda_n\big(\mathsf{T}_{i,n}(x_{1,n}, \ldots, x_{m,n}) \\ \quad\quad\qquad\qquad + a_{i,n} - x_{i,n}\big). \end{array} \tag{4}$$

For every $n \in \mathbb{N}$ and $i \in \{1, \ldots, m\}$, $\varepsilon_{i,n}$ is a binary random variable that signals the activation of the i-th block $\mathsf{T}_{i,n}$ of the operator $\mathbf{T}_n$ and $a_{i,n}$ is an H_i-valued random variable modeling some possible stochastic error. Such error may arise because of finite precision computations or approximations to the operator $\mathsf{T}_{i,n}$ [10]. For every $n \in \mathbb{N}$, let $\boldsymbol{\varepsilon}_n = (\varepsilon_{i,n})_{1\leqslant i\leqslant n}$. Note that, as a special case of Algorithm 2.1, (1) is recovered if $\boldsymbol{a}_n \equiv 0$ and $\boldsymbol{\varepsilon}_n \equiv (1, \ldots, 1)$, almost surely.

3. ASYMPTOTIC ANALYSIS

3.1. Notation and assumptions

Given a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ of $\mathbf{H}$-valued random variables in a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, the smallest σ-algebra generated by $(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n)$ is denoted by $\sigma(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n)$, and we denote by $(\mathcal{F}_n)_{n\in\mathbb{N}}$ a sequence of sigma-algebras such that

$$(\forall n \in \mathbb{N}) \quad \mathcal{F}_n \subset \mathcal{F} \quad \text{and} \quad \sigma(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_n) \subset \mathcal{F}_n \subset \mathcal{F}_{n+1}. \tag{5}$$

Let $\|\cdot\|$ denote the norm of $\mathbf{H}$ (the same notation will be used also for other Hilbert spaces). $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbf{H})$ denotes the space of $\mathbf{H}$-valued random variable $\boldsymbol{x}$ such that $\mathsf{E}\|\boldsymbol{x}\|^2 < +\infty$.

In order to study the convergence of Algorithm 2.1, we make the following assumptions.

Assumption 3.1

(i) $\mathbf{F} \neq \varnothing$.

(ii) $\inf_{n\in\mathbb{N}} \lambda_n > 0$.

(iii) There exists a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $[0, +\infty[$ such that $\sum_{n\in\mathbb{N}} \sqrt{\alpha_n} < +\infty$ and

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}(\|\boldsymbol{a}_n\|^2 \,|\, \mathcal{F}_n) \leqslant \alpha_n. \tag{6}$$

(iv) For every $n \in \mathbb{N}$, $\mathcal{E}_n = \sigma(\boldsymbol{\varepsilon}_n)$ and $\mathcal{F}_n$ are independent.

(v) For every $i \in \{1, \ldots, m\}$, $\mathsf{p}_i = \mathsf{P}[\varepsilon_{i,0} = 1] > 0$.

3.2. Convergence results

We first show that, under quasinonexpansiveness properties for the operators $(\mathbf{T}_n)_{n\in\mathbb{N}}$, almost sure convergence results are obtained. Recall that an operator $\mathbf{T} \colon \mathbf{H} \to \mathbf{H}$ with fixed point set $\operatorname{Fix} \mathbf{T}$ is quasinonexpansive if

$$(\forall \mathbf{z} \in \operatorname{Fix} \mathbf{T})(\forall \mathbf{x} \in \mathbf{H}) \quad \|\mathbf{T}\mathbf{x} - \mathbf{z}\| \leqslant \|\mathbf{x} - \mathbf{z}\|. \tag{7}$$

Theorem 3.2 [9] *Let $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 2.1. Suppose that $\sup_{n\in\mathbb{N}} \lambda_n < 1$ and that, for every $n \in \mathbb{N}$, $\mathbf{T}_n$ is quasinonexpansive. Then, under Assumption 3.1, the following hold:*

(i) *$(\mathbf{T}_n \boldsymbol{x}_n - \boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges to $\mathbf{0}$ almost surely.*

(ii) *Suppose that, almost surely, every sequential cluster point of $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ belongs to $\mathbf{F}$. Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges almost surely to an $\mathbf{F}$-valued random variable.*

Remark 3.3

(i) The condition required in (ii) is actually met for many fixed point algorithms for solving monotone inclusion problems, e.g., the forward-backward algorithm or the Douglas-Rachford algorithm [9, Section 5].

(ii) This convergence result can be extended to the more general case when the operators $(\mathbf{T}_n)_{n\in\mathbb{N}}$ are averaged operators [9, Corollary 3.8].

In order to obtain more accurate convergence results, we make the strict contraction assumption

$$\begin{cases} \mathbf{F} = \{\overline{\mathbf{x}}\} = \{(\overline{\mathsf{x}}_i)_{1\leqslant i\leqslant m}\} \\ (\forall n \in \mathbb{N})(\forall \mathbf{x} = (\mathsf{x}_i)_{1\leqslant i\leqslant m} \in \mathbf{H}) \\ \qquad\qquad \|\mathbf{T}_n \mathbf{x} - \overline{\mathbf{x}}\|^2 \leqslant \displaystyle\sum_{i=1}^{m} \tau_{i,n} \|\mathsf{x}_i - \overline{\mathsf{x}}_i\|^2, \end{cases} \tag{8}$$

where, for every $n \in \mathbb{N}$, $(\tau_{i,n})_{1\leqslant i\leqslant m}$ are strictly positive parameters such that $(\forall i \in \{1, \ldots, m\})\ \sup_{n\in\mathbb{N}} \tau_{i,n} < 1$.

Theorem 3.4 [11] *Let $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 2.1. Suppose that $\boldsymbol{x}_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbf{H})$ and $(\mathbf{T}_n)_{n\in\mathbb{N}}$ satisfy* (8)*. Then, under Assumption 3.1, $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges to $\overline{\mathbf{x}}$ both in the mean square and almost sure senses.*

4. MEAN SQUARE BEHAVIOR

4.1. Mean square error bound

Let us assume in this section that Assumption 3.1 holds and that the sequence $(\mathsf{T}_n)_{n\in\mathbb{N}}$ satisfies (8). In order to provide non asymptotic bounds on the mean square estimation error, introduce a few parameters, namely

$$(\forall n\in\mathbb{N})\quad \begin{cases} \xi_n = \dfrac{\alpha_n}{\min\limits_{1\leqslant i\leqslant m} \mathsf{p}_i} \\ \mu_n = 1-\min\limits_{1\leqslant i\leqslant m} \mathsf{p}_i\big(1-\tau_{i,n}\big) \\ \chi_n = 1-\lambda_n(1-\mu_n)+\sqrt{\xi_n}\lambda_n(1-\lambda_n+\lambda_n\sqrt{\mu_n}) \\ \overline{\eta}_n = \displaystyle\sum_{k=0}^{n}\Bigg[\prod_{\ell=k+1}^{n}\chi_\ell\Bigg]\lambda_k\big((1-\lambda_k+\lambda_k\sqrt{\mu_k})\sqrt{\xi_k}+\lambda_k\xi_k\big). \end{cases}$$

We are now ready to state our main result:

Theorem 4.1 *Under the same assumptions as in Theorem 3.4,*

$$(\forall n\in\mathbb{N})\quad \mathsf{E}\|\boldsymbol{x}_{n+1}-\overline{\mathbf{x}}\|^2 \leqslant \frac{\max\limits_{1\leqslant i\leqslant m}\mathsf{p}_i}{\min\limits_{1\leqslant i\leqslant m}\mathsf{p}_i}\Bigg(\prod_{k=0}^{n}\chi_k\Bigg)\mathsf{E}\|\boldsymbol{x}_0-\overline{\mathbf{x}}\|^2+\overline{\eta}_n. \quad (9)$$

The proof of this result is given in [11].

4.2. Behavior in the absence of stochastic errors

Let us consider the case when there are no errors, i.e., when $\alpha_n\equiv 0$. Set

$$(\forall n\in\mathbb{N})\quad \overline{\chi}=1-\inf_{n\in\mathbb{N}}\Big(\lambda_n\min_{1\leqslant i\leqslant m}\mathsf{p}_i(1-\tau_{i,n})\Big)\in[0,1[. \quad (10)$$

Then we derive from Theorem 4.1 that

$$(\forall n\in\mathbb{N})\quad \mathsf{E}\|\boldsymbol{x}_n-\overline{\mathbf{x}}\|^2\leqslant C\,\overline{\chi}^n, \quad (11)$$

where

$$C=\frac{\max\limits_{1\leqslant i\leqslant m}\mathsf{p}_i}{\min\limits_{1\leqslant i\leqslant m}\mathsf{p}_i}\mathsf{E}\|\boldsymbol{x}_0-\overline{\mathbf{x}}\|^2. \quad (12)$$

This shows that a linear convergence rate is obtained.

Let us now assess the impact of the activation probabilities of the blocks. For simplicity, let us further assume that the blocks are processed uniformly in the sense that $(\forall i\in\{1,\dots,m\})\ \mathsf{p}_i=\mathsf{p}$. Set

$$\chi=1-\inf_{n\in\mathbb{N}}\Big(\lambda_n\Big(1-\max_{1\leqslant i\leqslant m}\tau_{i,n}\Big)\Big)\in[0,1[. \quad (13)$$

Then

$$\overline{\chi}=1-(1-\chi)\mathsf{p}. \quad (14)$$

When $\mathsf{p}=1$, the upper bound in (14) on the convergence rate is minimal and equal to χ. This is consistent with the intuition that frequently activating the coordinates should favor the convergence speed as a function of the iteration number. On the other hand, activating the blocks less frequently induces a reduction of the computational load per iteration. In Algorithm 2.1, the cost of computing $\mathsf{T}_{i,n}(x_{1,n},\dots,x_{m,n})$, here assumed to be independent of i and the iteration number n, is on the average p times lower than in the standard non block-coordinate approach. Let us introduce the quantity

$$\varrho(\mathsf{p})=-\frac{\ln\big(1-(1-\chi)\mathsf{p}\big)}{\mathsf{p}} \quad (15)$$

to evaluate the convergence rate normalized by the probability p accounting for computational cost. Then (14) yields

$$\overline{\chi}^n=\exp\big(-\varrho(\mathsf{p})\mathsf{p}n\big). \quad (16)$$

As n iterations of the block-coordinate algorithm have the same computational cost as $\mathsf{p}n$ iterations of a non block-coordinate approach, $\varrho(\mathsf{p})$ appears to be a relevant quantity to evaluate the convergence rate normalized by the computational cost. The ratio $\rho(\mathsf{p})/\varrho(1)$ can thus be used to provide a fair comparison of a block-coordinate approach versus a non block-coordinate one.

Fig. 1 shows that, for values of χ not too small, the decrease in the normalized convergence rate remains limited with respect to a deterministic approach in which all the blocks are activated. For example, if $\chi>0.2$, then $\varrho(\mathsf{p})/\varrho(1)\in[0.49,1]$.

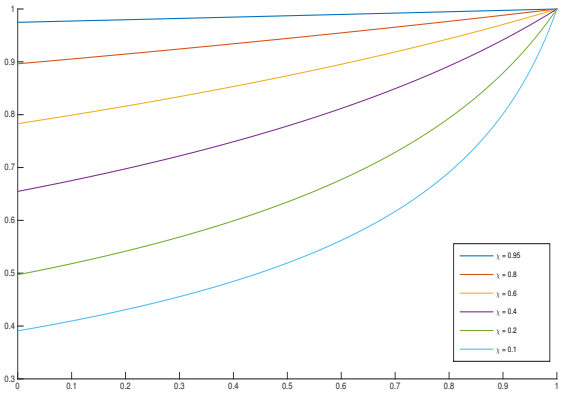

Fig. 1. Variations of $\varrho(\mathsf{p})/\varrho(1)$ as a function of p for various values of χ.

4.3. Influence of stochastic errors

Since $\alpha_n\to 0$, there exists $n_0\in\mathbb{N}$ such that $\widetilde{\chi}=\sup_{n\geqslant n_0}\chi_n<1$. Without loss of generality, we assume that

$n_0 = 0$. Using standard majorizations, we obtain

$$(\forall n \in \mathbb{N}) \quad \overline{\eta}_n \leqslant \left(1 + \sup_{k\in\mathbb{N}} \sqrt{\xi_k}\right) \sum_{k=0}^{n} \sqrt{\xi_k}\, \widetilde{\chi}^{n-k}. \quad (17)$$

Let us now assume that $\alpha_n = O(n^{-\theta})$ with $\theta \in]2, +\infty[$, which is a choice compatible with Assumption 3.1(iii) and consistent with standard stochastic approximation techniques. It then follows from (17) that $\overline{\eta}_n = O(n^{-\theta/2})$ and, consequently, Theorem 4.1, yields $\mathsf{E}\|\boldsymbol{x}_n - \overline{\mathbf{x}}\|^2 = O(n^{-\theta/2})$. We thus lose the linear convergence property. This illustrates the fact that care should be taken in order to control the stochastic error term in Algorithm 2.1.

5. SIMULATION EXAMPLE

We consider the problem of denoising a sequence of $m = 4$ color images of size 512×512 acquired in burst mode. Each color component of these images has been corrupted with a zero-mean white Gaussian noise with standard deviation 60, leading to an initial signal-to-noise ratio equal to 8.00 dB (see top row of Fig. 2). We denote by $\widetilde{\mathbf{x}} = (\widetilde{\mathsf{x}}_i)_{1\leqslant i\leqslant m}$ the original image and by $\mathbf{y} = (\mathsf{y}_i)_{1\leqslant i\leqslant m}$ the noisy one. In this application, for every $i \in \{1, \ldots, m\}$, $\mathsf{H}_i = \mathbb{R}^K$ with $K = 512^2$. Given $\mathbf{y}$, we want to generate an estimate $\overline{\mathbf{x}}$ of $\widetilde{\mathbf{x}}$ by solving the variational problem

$$\underset{\mathbf{x}\in\mathbf{H}}{\text{minimize}} \quad \sum_{i=1}^{m} \mathsf{f}_i(\mathsf{x}_i) + \sum_{i=1}^{m-1} \mathsf{g}(\mathsf{x}_{i+1} - \mathsf{x}_i) \quad (18)$$

where, for every $i \in \{1, \ldots, m\}$, f_i is the proper lower-semicontinuous convex function defined by

$$(\forall \mathsf{x}_i \in \mathsf{H}_i) \qquad \mathsf{f}_i(\mathsf{x}_i) = \frac{1}{2}\|\mathsf{x}_i - \mathsf{y}_i\|^2 + \kappa \mathsf{h}(\mathsf{W}\mathsf{x}_i). \quad (19)$$

$\kappa \in]0, +\infty[$ is an intra-image regularization parameter, $\mathsf{W} \in \mathbb{R}^{N\times N}$ is an orthogonal wavelet decomposition performed over 4 resolution levels using Symlet-4 Daubechies wavelets, and h is an ℓ_1-norm applied on the wavelet detail coefficients. On the other hand, g is an $1/\delta$-Lipschitz differentiable convex function serving to perform an inter-image regularization, which is defined as

$$(\forall \mathsf{u} = (\upsilon_k)_{1\leqslant k\leqslant K} \in \mathbb{R}^K)\ \mathsf{g}(\mathsf{u}) = \zeta \sum_{k=1}^{K} \sqrt{|\upsilon_k|^2 + \delta^2} \quad (20)$$

where $(\zeta, \delta) \in]0, +\infty[^2$. Problem (18) is solved by a block-coordinate forward-backward algorithm [9, Section 5.2], which is a special case of Algorithm (4) with, for every $i \in \{1, \ldots, m\}$, $n \in \mathbb{N}$, and $\mathbf{x} \in \mathbf{H}$,

$$\mathsf{T}_{i,n}(\mathsf{x}_1, \ldots, \mathsf{x}_m) = \mathrm{prox}_{\gamma \mathsf{f}_i}\left(\mathsf{x}_i - \gamma \sum_{\substack{j=1 \\ |j-i|=1}}^{m} \nabla \mathsf{g}(\mathsf{x}_i - \mathsf{x}_j)\right), \quad (21)$$

where $\gamma \in]0, +\infty[$. Since $(\mathsf{f}_i)_{1\leqslant i\leqslant m}$ are 1-strongly convex, (8) is satisfied with $(\forall i \in \{1, \ldots, m\})\ \tau_{i,n} \equiv 1/(1+\gamma)^2$. In our experiments, we set $\kappa = 84$, $\zeta = 5$, $\delta = 0.5$, $\gamma = 5.83 \times 10^{-2}$, and $\lambda_n \equiv 1$. The denoised images $\overline{\mathbf{x}}$ obtained after running the algorithm for a large number of iterations with all the blocks activated (i.e. $(\forall i \in \{1, \ldots, m\})\ \mathsf{p}_i = 1$) is displayed on the bottom row of Fig. 2, resulting in an improved signal-to-noise ratio equal to 17.54 dB.

We also show in Fig. 3 the variations of $\mathsf{E}\|\boldsymbol{x}_n - \overline{\mathbf{x}}\|^2$ as a function of the iteration number n, in three cases. The first corresponds to the non block-coordinate case, whereas the second (resp. third) corresponds to the stochastic case when $(\forall i \in \{1, \ldots, m\})\ \mathsf{p}_i = 0.8$ (resp. 0.46). In the latter two cases, the estimation of the mean square estimation error is performed over 10 realizations. In all the cases, the algorithm is initialized with the noisy images. The experimental plots are consistent with the upper bound expressions derived in (9). In particular, we observe that these ones provide good estimates of the asymptotic convergence rate. As expected, this rate value is lower when the probability of activation decays. In Fig. 4, we compare the same simulation scenarios, by plotting now the mean square estimation error as a function of the computation time. We observe then that the algorithm has a similar convergence behavior in all cases. This is in agreement with our discussion in Section 4.2 since, in our experimental setting, the parameter χ is close to 1.

REFERENCES

[1] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli, Dual block-coordinate forward-backward algorithm with application to deconvolution and deinterlacing of video sequences, *J. Math. Imaging Vision*, vol. 59, p. 415–431, 2017.

[2] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed. Springer, New York, 2017.

[3] L. M. Briceño-Arias and P. L. Combettes, Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery, *Numer. Math. Theory Methods Appl.,* vol. 2, pp. 485–508, 2009.

[4] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces,* Lecture Notes in Mathematics, vol. 2057. Springer, Heidelberg, 2012.

[5] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, A block-coordinate variable metric forward-backward algorithm, *J. Global Optim.,* vol. 66, pp. 457–485, 2016.

[6] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization,* vol. 53, pp. 475–504, 2004.

[7] P. L. Combettes and L. E. Glaudin, Quasinonexpansive iterations on the affine hull of orbits: From Mann's mean value algorithm to inertial methods, *SIAM J. Optim.,* vol. 27, pp. 2356–2380, 2017.

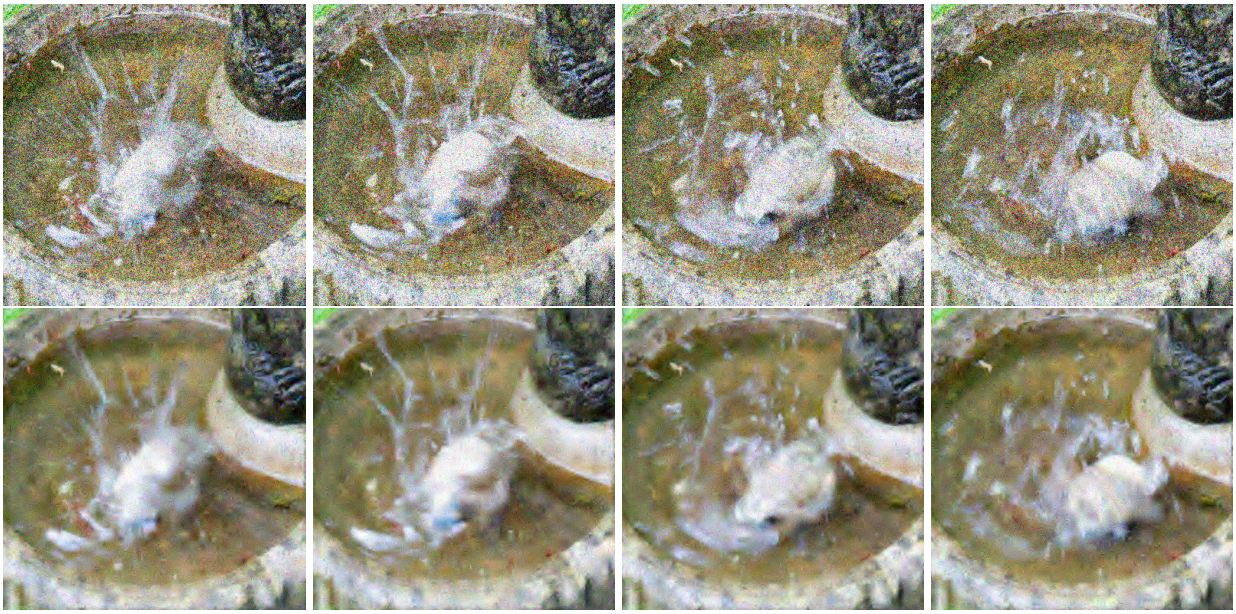

Fig. 2. Top: noisy images y_1, y_2, y_3, y_4. Bottom: denoised images $\overline{\mathsf{x}}_1$, $\overline{\mathsf{x}}_2$, $\overline{\mathsf{x}}_3$, $\overline{\mathsf{x}}_4$.

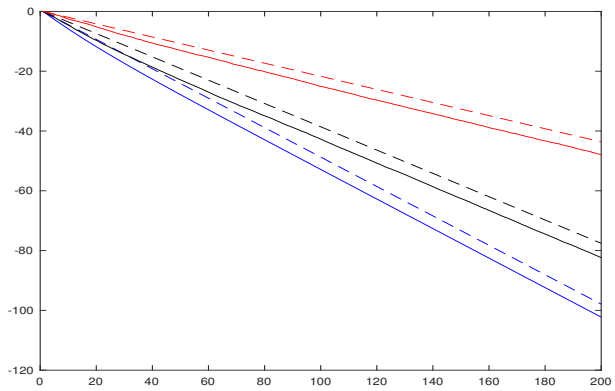

Fig. 3. $\mathsf{E}\|\boldsymbol{x}_n - \overline{\mathbf{x}}\|^2 / \mathsf{E}\|\boldsymbol{x}_0 - \overline{\mathbf{x}}\|^2$ (in dB) versus iteration number n: in blue when the blocks are always activated, in black (resp. red) when their probability of activation is 0.8 (resp. 0.46). The upper bound in (9) is shown in dashed lines for each case.

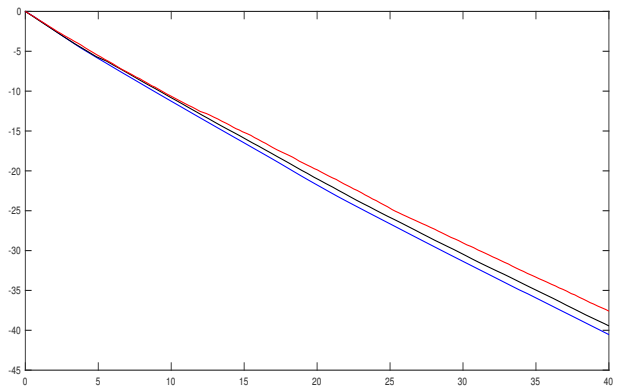

Fig. 4. $\mathsf{E}\|\boldsymbol{x}_n - \overline{\mathbf{x}}\|^2 / \mathsf{E}\|\boldsymbol{x}_0 - \overline{\mathbf{x}}\|^2$ (in dB) versus computation time in seconds (using Matlab R2018 with a 3.3 GHz Intel i7 processor): in blue when the blocks are always activated, in black (resp. red) when their probability of activation is 0.8 (resp. 0.46).

[8] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, (H. H. Bauschke et al., eds), pp. 185–212. Springer, New York, 2011.

[9] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, *SIAM J. Optim.,* vol. 25, pp. 1221–1248, 2015.

[10] P. L. Combettes and J.-C. Pesquet, Stochastic approximations and perturbations in forward-backward splitting for monotone operators, *Pure Appl. Funct. Anal.,* vol. 1, pp. 13–37, 2016.

[11] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping II: Mean-square and linear convergence, *Math. Program.,* published on line 2018-05-26.

[12] Z. Lu and L. Xiao, On the complexity analysis of randomized block-coordinate descent methods, *Math. Program.,* vol. A152, pp. 615–642, 2015.

[13] I. Necoara and A. Patrascu, A random coordinate descent algorithm for optimization problems with composite objective function and linear coupled constraints, *Comput. Optim. Appl.,* vol. 57, pp 307–337, 2014.

[14] P. Richtárik and M. Takáč, Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function, *Math. Program.,* vol. A144, pp. 1–38, Apr. 2014.

[15] E. Zeidler, *Nonlinear Functional Analysis and Its Applications I: Fixed-Point Theorems,* Springer-Verlag, New York, 1986.