STOCHASTIC FORWARD-BACKWARD AND PRIMAL-DUAL APPROXIMATION ALGORITHMS WITH APPLICATION TO ONLINE IMAGE RESTORATION

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

[1]Sorbonne Universités – UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions,

Paris, France

plc@ljll.math.upmc.fr

[2]Université Paris-Est, Laboratoire d'Informatique Gaspard Monge – CNRS UMR 8049,

Champs-sur-Marne, France

jean-christophe.pesquet@univ-paris-est.fr

ABSTRACT

Stochastic approximation techniques have been used in various contexts in data science. We propose a stochastic version of the forward-backward algorithm for minimizing the sum of two convex functions, one of which is not necessarily smooth. Our framework can handle stochastic approximations of the gradient of the smooth function and allows for stochastic errors in the evaluation of the proximity operator of the nonsmooth function. The almost sure convergence of the iterates generated by the algorithm to a minimizer is established under relatively mild assumptions. We also propose a stochastic version of a popular primal-dual proximal splitting algorithm, establish its convergence, and apply it to an online image restoration problem.

Index Terms— convex optimization, nonsmooth optimization, primal-dual algorithm, stochastic algorithm, parallel algorithm, proximity operator, recovery, image restoration.

1. INTRODUCTION

A large array of optimization problems arising in signal processing involve functions belonging to $\Gamma_0(\mathsf{H})$, the class of proper lower semicontinuous convex function from H to $]-\infty, +\infty]$, where H is a finite-dimensional real Hilbert space with norm $\|\cdot\|$. In particular, the following formulation has proven quite flexible and far reaching [18].

Problem 1.1 Let $\mathsf{f} \in \Gamma_0(\mathsf{H})$, let $\vartheta \in]0, +\infty[$, and let $\mathsf{h}\colon \mathsf{H} \to \mathbb{R}$ be a differentiable convex function such that $\nabla\mathsf{h}$ is ϑ^{-1}-Lipschitz continuous on H. The goal is to

$$\underset{\mathsf{x}\in\mathsf{H}}{\text{minimize}}\ \mathsf{f}(\mathsf{x}) + \mathsf{h}(\mathsf{x}), \qquad (1)$$

under the assumption that the set F of minimizers of $\mathsf{f} + \mathsf{h}$ is nonempty.

A standard method to solve Problem 1.1 is the forward-backward algorithm [6, 9, 10, 16, 18], which constructs a sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H via the recursion

$$(\forall n \in \mathbb{N}) \quad \mathsf{x}_{n+1} = \text{prox}_{\gamma_n\mathsf{f}}\big(\mathsf{x}_n - \gamma_n\nabla\mathsf{h}(\mathsf{x}_n)\big), \qquad (2)$$

where $\gamma_n \in]0, 2\vartheta[$ and $\text{prox}_{\gamma_n\mathsf{f}}$ is the proximity operator of function $\gamma_n\mathsf{f}$, i.e., [3]

$$\text{prox}_{\gamma_n\mathsf{f}}\colon \mathsf{x} \to \underset{\mathsf{y}\in\mathsf{H}}{\text{argmin}}\ \left(\mathsf{f}(\mathsf{y}) + \frac{1}{2\gamma_n}\|\mathsf{x} - \mathsf{y}\|^2\right). \qquad (3)$$

In practice, it may happen that, at each iteration n, $\nabla\mathsf{h}(\mathsf{x}_n)$ is not known exactly and is available only through some stochastic approximation u_n, while only a deterministic approximation f_n to f is known; see, e.g., [29]. To solve (1) in such uncertain environments, we propose to investigate the following stochastic version of (2). In this algorithm, at iteration n, a_n stands for a stochastic error term modeling inexact implementations of the proximity operator of $\gamma_n\mathsf{f}_n$, $(\Omega, \mathcal{F}, \mathsf{P})$ is the underlying probability space, and $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ denotes the space of H-valued random variable x such that $\mathsf{E}\|x\|^2 < +\infty$. Our algorithmic model is the following.

Algorithm 1.2 Let x_0, $(u_n)_{n\in\mathbb{N}}$, and $(a_n)_{n\in\mathbb{N}}$ be random variables in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 1]$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2\vartheta[$, and let $(\mathsf{f}_n)_{n\in\mathbb{N}}$ be a sequence of functions in $\Gamma_0(\mathsf{H})$. For every $n \in \mathbb{N}$, set

$$x_{n+1} = x_n + \lambda_n\big(\text{prox}_{\gamma_n\mathsf{f}_n}(x_n - \gamma_n u_n) + a_n - x_n\big). \quad (4)$$

The first instances of the stochastic iteration (4) can be traced back to [31] in the context of the gradient descent method, i.e., when $\mathsf{f}_n \equiv \mathsf{f} = 0$. Stochastic approximations in the gradient method were then investigated in the Russian literature of the late 1960s and early 1970s [21,23,36]. Stochastic gradient methods have also been used extensively in adaptive signal processing, in control, and in machine learning,

(e.g., in [2, 26, 40]). More generally, proximal stochastic gradient methods have been applied to various problems; see for instance [1, 20, 27, 32, 35, 37, 38].

The first objective of the present work is to provide a thorough convergence analysis of the stochastic forward-backward algorithm described in Algorithm 1.2. In particular, our results do not require that the proximal parameter sequence $(\gamma_n)_{n\in\mathbb{N}}$ be vanishing. A second goal of our paper is to show that the extension of Algorithm 1.2 for solving monotone inclusion problems allows us to derive a stochastic version of a recent primal-dual algorithm [39] (see also [17, 19]). Note that our algorithm is different from the random block-coordinate approaches developed in [4, 30], and that it is more in the spirit of the adaptive method of [28].

The organization of the paper is as follows. Section 2 contains our main result on the convergence of the iterates of Algorithm 1.2. Section 3 presents a stochastic primal-dual approach for solving composite convex optimization problems. Section 4 illustrates the benefits of this algorithm in signal restoration problems with stochastic degradation operators. Concluding remarks appear in Section 5.

2. A STOCHASTIC FORWARD-BACKWARD ALGORITHM

Throughout, given a sequence $(x_n)_{n\in\mathbb{N}}$ of H-valued random variables, the smallest σ-algebra generated by $x_0, \dots, x_n$ is denoted by $\sigma(x_0, \dots, x_n)$, and we denote by $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ a sequence of sigma-algebras such that

$$(\forall n \in \mathbb{N}) \quad \mathcal{X}_n \subset \mathcal{F} \quad \text{and} \quad \sigma(x_0, \dots, x_n) \subset \mathcal{X}_n \subset \mathcal{X}_{n+1}. \tag{5}$$

Furthermore, $\ell_+(\mathscr{X})$ designates the set of sequences of $[0, +\infty[$-valued random variables $(\xi_n)_{n\in\mathbb{N}}$ such that, for every $n \in \mathbb{N}$, ξ_n is $\mathcal{X}_n$-measurable, and we define

$$\ell_+^{1/2}(\mathscr{X}) = \left\{ (\xi_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{X}) \ \Big| \ \sum_{n\in\mathbb{N}} \xi_n^{1/2} < +\infty \ \mathsf{P}\text{-a.s.} \right\}, \tag{6}$$

and

$$\ell_+^\infty(\mathscr{X}) = \left\{ (\xi_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{X}) \ \Big| \ \sup_{n\in\mathbb{N}} \xi_n < +\infty \ \mathsf{P}\text{-a.s.} \right\}. \tag{7}$$

We now state our main convergence result.

Theorem 2.1 *Consider the setting of Problem 1.1, let $(\tau_n)_{n\in\mathbb{N}}$ be a sequence in $[0, +\infty[$, let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 1.2, and let $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ be a sequence of sub-sigma-algebras satisfying (5). Suppose that the following are satisfied:*

(a) $\sum_{n\in\mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|a_n\|^2 \,|\, \mathcal{X}_n)} < +\infty.$

(b) $\sum_{n\in\mathbb{N}} \sqrt{\lambda_n} \|\mathsf{E}(u_n \,|\, \mathcal{X}_n) - \nabla\mathsf{h}(x_n)\| < +\infty.$

(c) *For every $\mathsf{z} \in \mathsf{F}$, there exists $(\zeta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$ such that $\left(\lambda_n \zeta_n(\mathsf{z})\right)_{n\in\mathbb{N}} \in \ell_+^{1/2}(\mathscr{X})$ and*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}(\|u_n - \mathsf{E}(u_n \,|\, \mathcal{X}_n)\|^2 \,|\, \mathcal{X}_n)$$
$$\leqslant \tau_n \|\nabla\mathsf{h}(x_n) - \nabla\mathsf{h}(\mathsf{z})\|^2 + \zeta_n(\mathsf{z}). \tag{8}$$

(d) *There exist sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ in $[0, +\infty[$ such that $\sum_{n\in\mathbb{N}} \sqrt{\lambda_n}\alpha_n < +\infty$, $\sum_{n\in\mathbb{N}} \lambda_n\beta_n < +\infty$, and*

$$(\forall n \in \mathbb{N})(\forall \mathsf{x} \in \mathsf{H})$$
$$\|\mathrm{prox}_{\gamma_n\mathsf{f}_n}\mathsf{x} - \mathrm{prox}_{\gamma_n\mathsf{f}}\mathsf{x}\| \leqslant \alpha_n\|\mathsf{x}\| + \beta_n. \tag{9}$$

(e) $\inf_{n\in\mathbb{N}} \gamma_n > 0$, $\sup_{n\in\mathbb{N}} \tau_n < +\infty$, *and* $\sup_{n\in\mathbb{N}}(1 + \tau_n)\gamma_n < 2\vartheta$.

(f) *Either $\inf_{n\in\mathbb{N}} \lambda_n > 0$ or $\big[\gamma_n \equiv \gamma, \sum_{n\in\mathbb{N}} \tau_n < +\infty$, and $\sum_{n\in\mathbb{N}} \lambda_n = +\infty\big]$.*

Then the following hold for every $\mathsf{z} \in \mathsf{F}$ and for some F-valued random variable x:

(i) $\sum_{n\in\mathbb{N}} \lambda_n \|\nabla\mathsf{h}(x_n) - \nabla\mathsf{h}(\mathsf{z})\|^2 < +\infty$ P-a.s.

(ii) $\sum_{n\in\mathbb{N}} \lambda_n \|x_n - \gamma_n\nabla\mathsf{h}(x_n) - \mathrm{prox}_{\gamma_n\mathsf{f}}\big(x_n - \gamma_n\nabla\mathsf{h}(x_n)\big) + \gamma_n\nabla\mathsf{h}(\mathsf{z})\|^2 < +\infty$ P-a.s.

(iii) $(x_n)_{n\in\mathbb{N}}$ *converges almost surely to x.*

In the deterministic case, Theorem 2.1(iii) can be found in [7, Corollary 6.5]. The proof the above stochastic version is based on the theoretical tools of [12] (see [13] for technical details and extensions to infinite-dimensional Hilbert spaces).

It should be noted that the existing works which are the most closely related to ours do not allow any approximation of the function f and make some additional restrictive assumptions. For example, in [1, Corollary 8] and [33], $(\gamma_n)_{n\in\mathbb{N}}$ is a decreasing sequence. In [1, Corollary 8], [33], and [34], no error term is allowed in the numerical evaluations of the proximity operators ($a_n \equiv 0$). In addition, in the former work, it is assumed that $(x_n)_{n\in\mathbb{N}}$ is bounded, whereas the two latter ones assume that the approximation of the gradient of h is unbiased, that is

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}(u_n \,|\, \mathcal{X}_n) = \nabla\mathsf{h}(x_n). \tag{10}$$

3. STOCHASTIC PRIMAL-DUAL SPLITTING

The subdifferential

$$\partial\mathsf{f}: \mathsf{x} \mapsto \big\{ \mathsf{u} \in \mathsf{H} \ \big| \ (\forall \mathsf{y} \in \mathsf{H}) \ \langle \mathsf{y} - \mathsf{x} \,|\, \mathsf{u} \rangle + \mathsf{f}(\mathsf{x}) \leqslant \mathsf{f}(\mathsf{y}) \big\} \tag{11}$$

of a function $\mathsf{f} \in \Gamma_0(\mathsf{H})$ is an example of a maximally monotone operator [3]. Forward-backward splitting has been developed in the more general framework of solving monotone inclusions [3, 7]. This powerful framework makes it possible to design efficient primal-dual strategies for optimization problems; see for instance [17, 25] and the references therein.

More precisely, we are interested in the following optimization problem [11, Section 4].

Problem 3.1 Let $f \in \Gamma_0(H)$, let $\vartheta \in {]}0, +\infty[$, let $h\colon H \to \mathbb{R}$ be convex and differentiable with a ϑ^{-1}-Lipschitz-continuous gradient, and let q be a strictly positive integer. For every $k \in \{1, \dots, q\}$, let G_k be a finite-dimensional Hilbert space, let $g_k \in \Gamma_0(G_k)$, and let $L_k\colon H \to G_k$ be linear. Let $G = G_1 \oplus \cdots \oplus G_q$ be the direct Hilbert sum of $G_1, \dots, G_q$, and suppose that there exists $\bar{x} \in H$ such that

$$0 \in \partial f(\bar{x}) + \sum_{k=1}^{q} L_k^* \partial g_k(L_k \bar{x}) + \nabla h(\bar{x}). \tag{12}$$

Let F be the set of solutions to the problem

$$\underset{x \in H}{\text{minimize}} \quad f(x) + \sum_{k=1}^{q} g_k(L_k x) + h(x) \tag{13}$$

and let F^* be the set of solutions to the dual problem

$$\underset{v \in G}{\text{minimize}} \quad (f^* \square h^*)\left(-\sum_{k=1}^{q} L_k^* v_k\right) + \sum_{k=1}^{q} g_k^*(v_k), \tag{14}$$

where $\square$ denotes the infimal convolution operation, φ^* is the Legendre conjugate of a function φ, and $v = (v_1, \dots, v_q)$ designates a generic point in G. The objective is to find a point in $F \times F^*$.

We are interested in the case when only stochastic approximations of the gradients of h and approximations of the function f are available to solve Problem 3.1. The following algorithm, which can be viewed as a stochastic extension of those of [5,8,17,19,22,24,39], will be the focus of our investigation.

Algorithm 3.2 Let $\rho \in {]}0, +\infty[$, let $(f_n)_{n \in \mathbb{N}}$ be a sequence of functions in $\Gamma_0(H)$, let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in ${]}0, 1]$ such that $\sum_{n \in \mathbb{N}} \lambda_n = +\infty$, and, for every $k \in \{1, \dots, q\}$, let $\sigma_k \in {]}0, +\infty[$. Let x_0, $(u_n)_{n \in \mathbb{N}}$, and $(b_n)_{n \in \mathbb{N}}$ be random variables in $L^2(\Omega, \mathcal{F}, P; H)$, and let v_0 and $(c_n)_{n \in \mathbb{N}}$ be random variables in $L^2(\Omega, \mathcal{F}, P; G)$. Iterate

$$
\begin{aligned}
&\text{for } n = 0, 1, \dots \\
&\left\lfloor
\begin{aligned}
&y_n = \text{prox}_{\rho f_n}\left(x_n - \rho\left(\sum_{k=1}^{q} L_k^* v_{k,n} + u_n\right)\right) + b_n \\
&x_{n+1} = x_n + \lambda_n(y_n - x_n) \\
&\text{for } k = 1, \dots, q \\
&\left\lfloor
\begin{aligned}
&w_{k,n} = \text{prox}_{\sigma_k g_k^*}\left(v_{k,n} + \sigma_k L_k(2y_n - x_n)\right) + c_{k,n} \\
&v_{k,n+1} = v_{k,n} + \lambda_n(w_{k,n} - v_{k,n}).
\end{aligned}
\right.
\end{aligned}
\right.
\end{aligned}
\tag{15}
$$

One of main benefits of the proposed algorithm is that it allows us to solve jointly the primal problem (13) and the dual one (14) in a fully decomposed fashion, where each function

and linear operator is activated individually. In particular, it does not require any inversion of some linear operator related to the operators $(L_k)_{1 \leqslant k \leqslant q}$ arising in the original problem. The convergence of the algorithm is guaranteed by the following result which follows from [13, Proposition 5.3].

Proposition 3.3 *Consider the setting of Problem 3.1, let $\mathscr{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sub-sigma-algebras of $\mathcal{F}$, and let $(x_n)_{n \in \mathbb{N}}$ and $(v_n)_{n \in \mathbb{N}}$ be sequences generated by Algorithm 3.2. Suppose that the following are satisfied:*

(a) $(\forall n \in \mathbb{N}) \; \sigma(x_{n'}, v_{n'})_{0 \leqslant n' \leqslant n} \subset \mathcal{X}_n \subset \mathcal{X}_{n+1}.$

(b) $\sum_{n \in \mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|b_n\|^2 \,|\, \mathcal{X}_n)} < +\infty$ *and* $\sum_{n \in \mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|c_n\|^2 \,|\, \mathcal{X}_n)} < +\infty.$

(c) $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n} \|\mathsf{E}(u_n \,|\, \mathcal{X}_n) - \nabla h(x_n)\| < +\infty.$

(d) *There exists a summable sequence $(\tau_n)_{n \in \mathbb{N}}$ in $[0, +\infty[$ such that, for every $x \in F$, there exists $(\zeta_n(x))_{n \in \mathbb{N}} \in \ell_+^\infty(\mathscr{X})$ such that $(\lambda_n \zeta_n(x))_{n \in \mathbb{N}} \in \ell_+^{1/2}(\mathscr{X})$ and*

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \mathsf{E}(\|u_n - \mathsf{E}(u_n \,|\, \mathcal{X}_n)\|^2 \,|\, \mathcal{X}_n) \\
\leqslant \tau_n \|\nabla h(x_n) - \nabla h(x)\|^2 + \zeta_n(x). \quad (16)
\end{aligned}
$$

(e) *There exist sequences $(\alpha_n)_{n \in \mathbb{N}}$ and $(\beta_n)_{n \in \mathbb{N}}$ in $[0, +\infty[$ such that $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n} \alpha_n < +\infty$, $\sum_{n \in \mathbb{N}} \lambda_n \beta_n < +\infty$, and*

$$
\begin{aligned}
(\forall n \in \mathbb{N})(\forall x \in H) \\
\|\text{prox}_{\rho f_n} x - \text{prox}_{\rho f} x\| \leqslant \alpha_n \|x\| + \beta_n. \quad (17)
\end{aligned}
$$

(f) $\left(\rho^{-1} - \sum_{k=1}^{q} \sigma_k \|L_k\|^2\right) \vartheta > 1/2.$

Then, for some F-valued random variable x and some F^*-valued random variable v, $(x_n)_{n \in \mathbb{N}}$ converges almost surely to x and $(v_n)_{n \in \mathbb{N}}$ converges almost surely to v.

4. APPLICATION TO ONLINE SIGNAL RECOVERY

We consider the recovery of a signal $\bar{x} \in H = \mathbb{R}^N$ from the observation model

$$(\forall n \in \mathbb{N}) \quad z_n = K_n \bar{x} + e_n, \tag{18}$$

where K_n is a $\mathbb{R}^{M \times N}$-valued random matrix and e_n is a $\mathbb{R}^M$-valued random noise vector. The objective is to recover $\bar{x}$ from $(K_n, z_n)_{n \in \mathbb{N}}$, which is assumed to be an identically distributed sequence. Such recovery problems have been addressed in [14]. In this context, we propose to solve the primal problem (13) with $q = 1$ and

$$(\forall x \in \mathbb{R}^N) \quad h(x) = \frac{1}{2} \mathsf{E}\|K_0 x - z_0\|^2, \tag{19}$$

while functions f and $g_1 \circ L_1$ are used to promote prior information on the target solution. Since the statistics of the sequence $(K_n, z_n)_{n \in \mathbb{N}}$ are not assumed to be known a priori

and have to be learnt online, at iteration $n \in \mathbb{N}$, we employ the empirical estimate

$$u_n = \frac{1}{m_{n+1}} \sum_{n'=0}^{m_{n+1}-1} K_{n'}^\top (K_{n'} x_n - z_{n'}) \qquad (20)$$

of $\nabla \mathsf{h}(x_n)$. The following statement, which can be deduced from [13, Section 5.2], illustrates the applicability of the results of Section 3.

Proposition 4.1 *Consider the setting of Problem 3.1 and Algorithm 3.2, where* $\mathsf{f}_n \equiv \mathsf{f}$, $b_n \equiv 0$, *and* $\boldsymbol{c}_n \equiv 0$. *Let* $(m_n)_{n \in \mathbb{N}}$ *be a strictly increasing sequence in* $\mathbb{N}$ *such that* $m_n = O(n^{1+\delta})$ *with* $\delta \in \ ]0, +\infty[$, *and let*

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{\mathcal{X}}_n = \sigma \big(x_0, \boldsymbol{v}_0, (K_{n'}, e_{n'})_{0 \leqslant n' < m_n} \big). \qquad (21)$$

Suppose that the following are satisfied:

(a) *The domain of* f *is bounded.*

(b) $(K_n, e_n)_{n \in \mathbb{N}}$, *is an independent and identically distributed (i.i.d.) sequence such that* $\mathsf{E}\|K_0\|^4 < +\infty$ *and* $\mathsf{E}\|e_0\|^4 < +\infty$.

(c) $\lambda_n = O(n^{-\kappa})$, *where* $\kappa \in \ ]1-\delta, 1] \cap [0, 1]$.

Then Assumptions (a)-(e) *in Proposition 3.3 hold.*

Based on this result, we apply Algorithm 3.2 to a practical scenario in which a grayscale image of size 256×256 with pixel values in $[0, 255]$ is degraded by a stochastic blur. The stochastic operator corresponds to a uniform i.i.d. subsampling of a uniform 5×5 blur, performed in the discrete Fourier domain. More precisely, the value of the frequency response at each frequency bin is kept with probability 0.3 or it is set to zero. In addition, the image is corrupted by an additive zero-mean white Gaussian noise with standard deviation equal to 5. The average signal-to-noise ratio (SNR) is initially equal to 3.4 dB.

In our restoration approach, the function f is the indicator function of the set $[0, 255]^N$, while $\mathsf{g}_1 \circ \mathsf{L}_1$ is a classical isotropic total variation regularizer, where L_1 is the concatenation of the horizontal and vertical discrete gradient operators. Fig. 1 displays the original image, the restored image, as well as two realizations of the degraded images. The SNR for the restored image is equal to 28.1 dB. Fig. 2 shows the convergence behavior of the algorithm. In these experiments, in accordance with Proposition 4.1, we have chosen

$$(\forall n \in \mathbb{N}) \quad \begin{cases} m_n = n^{1.1} \\ \lambda_n = (1 + (n/500)^{0.95})^{-1}. \end{cases} \qquad (22)$$

5. CONCLUSION

We have proposed two stochastic proximal splitting algorithms for solving nonsmooth convex optimization problems.
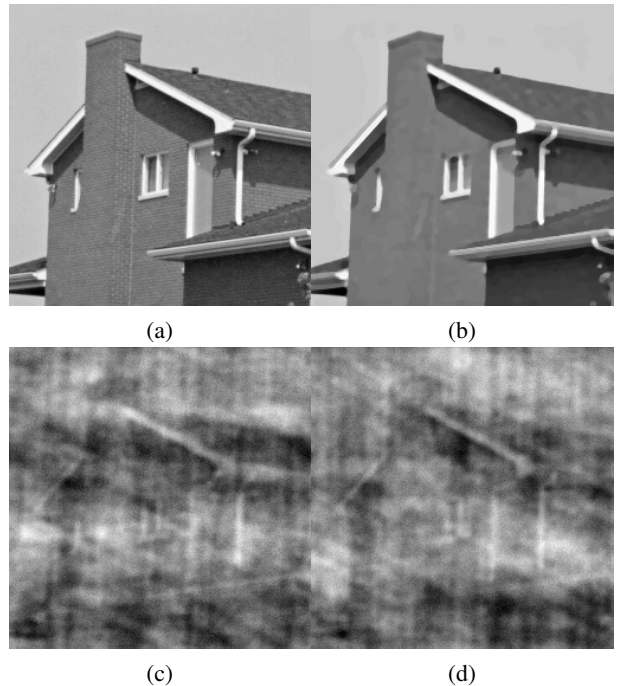

Fig. 1. Original image $\overline{x}$ (a), restored image (b), degraded image 1 (SNR = 0.14 dB) (c), and degraded image 2 (SNR = 12.0 dB) (d).

These methods require only approximations of the functions used in the formulation of the optimization problem, which is of the utmost importance for solving online signal processing problems. The almost sure convergence of these algorithms has been established. The stochastic version of the primal-dual algorithm that we have investigated has been evaluated in an online image restoration problem in which the data are blurred by a stochastic point spread function and corrupted with noise.

REFERENCES

[1] Y. F. Atchadé, G. Fort, and E. Moulines, "On stochastic proximal gradient algorithms," 2014. http://arxiv.org/abs/1402.2365
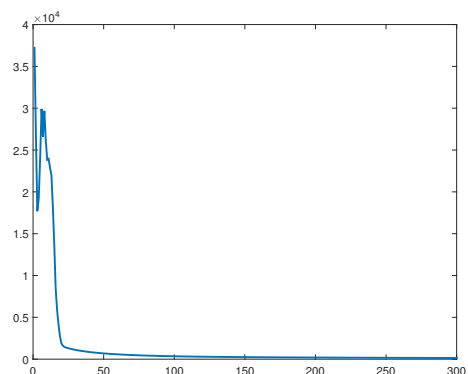
Fig. 2. $\|x_n - x_\infty\|$ versus the iteration number n.

[2] F. Bach and E. Moulines, "Non-asymptotic analysis of stochastic approximation algorithms for machine learning", in *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, Granada, Spain, Dec. 12-17, 2011, pp. 451–459.

[3] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, New York, 2011.

[4] P. Bianchi, W. Hachem, and F. Iutzeler, "A stochastic coordinate descent primal-dual algorithm and applications to large-scale composite optimization," 2014. http://arxiv.org/abs/1407.0898

[5] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision,* vol. 40, pp. 120–145, 2011.

[6] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems,* vol. 23, pp. 1495–1518, 2007.

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

[8] P. L Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ. "A forward-backward view of some primal-dual optimization methods in image recovery," *Proc. IEEE Int. Conf. Image Process.*, Paris, France, 27-30 Oct. 2014, pp. 4141–4145.

[9] P. L. Combettes and J.-C. Pesquet, "Proximal thresholding algorithm for minimization over orthonormal bases", *SIAM J. Optim.,* vol. 18, pp. 1351–1376, 2007.

[10] 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.

[11] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, vol. 20, pp. 307–330, 2012.

[12] 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.

[13] 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.

[14] P. L. Combettes and H. J. Trussell, "Methods for digital restoration of signals degraded by a stochastic impulse response," *IEEE Trans. Acoustics, Speech, Signal Process.,* vol. 37, pp. 393–401, 1989.

[15] P. L. Combettes and B. C. Vũ, "Variable metric forward-backward splitting with applications to monotone inclusions in duality," *Optimization,* vol. 63, pp. 1289–1318, 2014.

[16] P. L. Combettes and I. Yamada, "Compositions and convex combinations of averaged nonexpansive operators," *J. Math. Anal. Appl.,* vol. 425, pp. 55–70, 2015.

[17] P. L. Combettes and B. C. Vũ, "Variable metric forward-backward splitting with applications to monotone inclusions in duality," *Optimization,* vol. 63, pp. 1289–1318, 2014.

[18] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.,* vol. 4, pp. 1168–1200, 2005.

[19] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.,* vol. 158, pp. 460–479, 2013.

[20] J. Duchi and Y. Singer, "Efficient online and batch learning using forward backward splitting," *J. Mach. Learn. Res.,* vol. 10, pp. 2899–2934, 2009.

[21] Yu. M. Ermoliev and Z. V. Nekrylova, "The method of stochastic gradients and its application," in *Seminar: Theory of Optimal Solutions,* no. 1, *Akad. Nauk Ukrain. SSR, Kiev,* pp. 24–47, 1967.

[22] E. Esser, X. Zhang, and T. Chan, "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science," *SIAM J. Imaging Sci.*, vol. 3, pp. 1015–1046, 2010.

[23] O. V. Guseva, "The rate of convergence of the method of generalized stochastic gradients," *Kibernetika (Kiev),* vol. 1971, pp. 143–145, 1971.

[24] B. He and X. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective," *SIAM J. Imaging Sci.*, vol. 5, pp. 119–149, 2012.

[25] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, pp. 31–54, 2015.

[26] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms with Applications*, 2nd ed. Springer, New York, 2003.

[27] J. Konečný, J. Liu, P. Richtárik, and M. Takáč, "Minibatch semi-stochastic gradient descent in the proximal setting," *IEEE J. Selected Topics Signal Process.*, vol. 10, pp. 242–255, 2016.

[28] S. Ono, M. Yamagishi, and I. Yamada, "A sparse system identification by using adaptively-weighted total variation via a primal-dual splitting approach," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Vancouver, Canada, 26-31 May 2013, pp. 6029–6033.

[29] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. O. Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE J. Selected Topics Signal Process.*, vol. 10, pp. 224–241, 2016.

[30] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," *J. Nonlinear Convex Anal.*, vol. 16, pp. 2453–2490, 2015.

[31] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statistics*, vol. 22, pp. 400–407, 1951.

[32] L. Rosasco, S. Villa, and B. C. Vũ, "Convergence of stochastic proximal gradient algorithm," 2014. http://arxiv.org/abs/1403.5074

[33] L. Rosasco, S. Villa, and B. C. Vũ, "Stochastic forward-backward splitting for monotone inclusions," *J. Optim. Theory Appl.,* vol. 169, pp. 388–406, 2016.

[34] L. Rosasco, S. Villa, and B. C. Vũ, "A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions," *Optimization,* to appear.

[35] S. Shalev-Shwartz and T. Zhang, "Stochastic dual coordinate ascent methods for regularized loss minimization," *J. Mach. Learn. Res.*, vol. 14, pp. 567–599, 2013.

[36] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions.* Springer, New York, 1985.

[37] L. Xiao and T. Zhang, "A proximal stochastic gradient method with progressive variance reduction," *SIAM J. Optim.*, vol. 24, pp. 2057–2075, 2014.

[38] M. Yamagishi, M. Yukawa, and I. Yamada, "Acceleration of adaptive proximal forward-backward splitting method and its application to sparse system identification," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Prague, Czech Republic, May 22-27, 2011, pp. 4296–4299.

[39] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[40] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1985.