

Stochastic Proximal Algorithms with Applications to Online Image Recovery

Patrick Louis Combettes¹ and Jean-Christophe Pesquet²

¹ Mathematics Department, North Carolina State University,
Raleigh, USA

² Center for Visual Computing, CentraleSupélec, University Paris-Saclay,
Grande Voie des Vignes, 92295 Châtenay-Malabry, France

Taras Shevchenko National University of Kyiv

Outline

1. Introduction
2. Stochastic Forward-Backward
3. Monotone Inclusion Problems
4. Primal-Dual Extension
5. Application
6. Conclusion

Context

Need for fast optimization methods over the last decade

Why?

Context

Need for fast optimization methods over the last decade

Why?

- ▶ Interest in nonsmooth cost functions (*sparsity*)
- ▶ Need for optimal processing of massive datasets (*big data*)
 - ↪ large number of variables (inverse problems)
 - ↪ large number of observations (machine learning)
- ▶ Use of more sophisticated data structures (*graph signal processing*)

Variational formulation

GOAL:

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

where

- H : signal space (real Hilbert space)
- $f \in \Gamma_0(H)$: class of convex lower-semicontinuous functions from H to $] -\infty, +\infty]$ with a nonempty domain
- $h: H \rightarrow \mathbb{R}$: differentiable convex function such that ∇h is ϑ^{-1} -Lipschitz continuous with $\vartheta \in]0, +\infty[$
- $F = \text{Argmin}(f + h)$ assumed to be nonempty.

Algorithm

CLASSICAL SOLUTION [Combettes and Wajs - 2005]

FORWARD-BACKWARD ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma_n f}(x_n - \gamma_n \nabla h(x_n)) - x_n),$$

where $\lambda_n \in]0, 1]$, $\gamma_n \in]0, 2\vartheta[$, and $\text{prox}_{\gamma_n f}$ is the proximity operator of $\gamma_n f$ [Moreau - 1965]:

$$\text{prox}_{\gamma_n f} : x \mapsto \underset{y \in H}{\text{argmin}} \quad f(y) + \frac{1}{2\gamma_n} \|x - y\|^2.$$

SPECIAL CASES: projected gradient method, iterative soft thresholding, Landweber algorithm,...

Algorithm

CLASSICAL SOLUTION [Combettes and Wajs - 2005]

FORWARD-BACKWARD ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma_n f}(x_n - \gamma_n \nabla h(x_n)) - x_n),$$

In the context of **online processing** and **machine learning**,
what to do if ∇h and f **are not known exactly** ?

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma_n f_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma_n f_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$
- $f_n \in \Gamma_0(H)$: approximation to f

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma_n f_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$
- $f_n \in \Gamma_0(H)$: approximation to f
- u_n second-order random variable: approximation to $\nabla h(x_n)$

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (\text{prox}_{\gamma_n f_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$
- $f_n \in \Gamma_0(H)$: approximation to f
- u_n second-order random variable: approximation to $\nabla h(x_n)$
- a_n second-order random variable: possible additional error term.

Assumptions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

where $\sigma(x_0, \dots, x_n)$ is the smallest σ -algebra generated by x_0, \dots, x_n .

$\ell_+(\mathcal{X})$: set of sequences of $[0, +\infty[$ -valued random variables $(\xi_n)_{n \in \mathbb{N}}$ such that $(\forall n \in \mathbb{N}) \xi_n$ is \mathcal{X}_n -measurable and

$$\ell_+^1(\mathcal{X}) = \left\{ (\xi_n)_{n \in \mathbb{N}} \in \ell_+(\mathcal{X}) \mid \sum_{n \in \mathbb{N}} \xi_n < +\infty \text{ P-a.s.} \right\}$$
$$\ell_+^\infty(\mathcal{X}) = \left\{ (\xi_n)_{n \in \mathbb{N}} \in \ell_+(\mathcal{X}) \mid \sup_{n \in \mathbb{N}} \xi_n < +\infty \text{ P-a.s.} \right\}.$$

Assumptions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

where $\sigma(x_0, \dots, x_n)$ is the smallest σ -algebra generated by x_0, \dots, x_n .

Assumptions on the gradient approximation:

- ▶ $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n} \|E(u_n | \mathcal{X}_n) - \nabla h(x_n)\| < +\infty$.
- ▶ For every $z \in F$, there exist sequences $(\tau_n)_{n \in \mathbb{N}} \in \ell_+$, $(\zeta_n(z))_{n \in \mathbb{N}} \in \ell_+^\infty(\mathcal{X})$ such that $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n \zeta_n(z)} < +\infty$ and

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

Assumptions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

where $\sigma(x_0, \dots, x_n)$ is the smallest σ -algebra generated by x_0, \dots, x_n .

Assumptions on the prox approximation:

- ▶ 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

$$(\forall n \in \mathbb{N})(\forall x \in H) \quad \|\text{prox}_{\gamma_n f_n} x - \text{prox}_{\gamma_n f} x\| \leq \alpha_n \|x\| + \beta_n.$$

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

Assumptions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

where $\sigma(x_0, \dots, x_n)$ is the smallest σ -algebra generated by x_0, \dots, x_n .

Assumptions on the algorithm parameters:

- ▶ $\inf_{n \in \mathbb{N}} \gamma_n > 0$, $\sup_{n \in \mathbb{N}} \tau_n < +\infty$, and $\sup_{n \in \mathbb{N}} (1 + \tau_n) \gamma_n < 2\vartheta$.
- ▶ **Either** $\inf_{n \in \mathbb{N}} \lambda_n > 0$ **or** $[\gamma_n \equiv \gamma, \sum_{n \in \mathbb{N}} \tau_n < +\infty, \text{ and } \sum_{n \in \mathbb{N}} \lambda_n = +\infty]$.

Convergence Result

Under the previous assumptions, the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the algorithm **converges weakly a.s.** to an F-valued random variable.

REMARKS:

- ★ Related works: [Rosasco et al. - 2014, Atchadé et al. - 2016]
- ★ Result valid for non vanishing step sizes $(\gamma_n)_{n \in \mathbb{N}}$.
- ★ We do not need to assume that

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

- ★ Proof based on properties of stochastic quasi-Fejér sequences [Combettes and Pesquet – 2015, 2016].

Stochastic Quasi-Fejér Sequences

- ▶ Let $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- ▶ **Deterministic definition:** A sequence $(x_n)_{n \in \mathbb{N}}$ in H is Fejér monotone with respect to F if for every $z \in F$,

$$(\forall n \in \mathbb{N}) \quad \phi(\|x_{n+1} - z\|) \leq \phi(\|x_n - z\|)$$

Stochastic Quasi-Fejér Sequences

- ▶ Let $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- ▶ **Stochastic definition 1:** A sequence $(x_n)_{n \in \mathbb{N}}$ of H-valued random variables is *stochastically Fejér monotone* with respect to F if, for every $z \in F$,

$$(\forall n \in \mathbb{N}) \mathbf{E}(\phi(\|x_{n+1} - z\|) \mid \mathcal{X}_n) \leq \phi(\|x_n - z\|)$$

Stochastic Quasi-Fejér Sequences

- ▶ Let $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- ▶ **Stochastic definition 2:** A sequence $(x_n)_{n \in \mathbb{N}}$ of H-valued random variables is *stochastically quasi-Fejér monotone* with respect to F if, for every $z \in F$, there exist $(\chi_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$, $(\vartheta_n(z))_{n \in \mathbb{N}} \in \ell_+(\mathcal{X})$, and $(\eta_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$ such that
$$(\forall n \in \mathbb{N}) \mathbf{E}(\phi(\|x_{n+1} - z\|) \mid \mathcal{X}_n) + \vartheta_n(z) \leq (1 + \chi_n(z))\phi(\|x_n - z\|) + \eta_n(z)$$

Stochastic Quasi-Fejér Sequences

- ▶ Let $\phi: [0, +\infty[\rightarrow [0, +\infty[$, $\phi(t) \uparrow +\infty$ as $t \rightarrow +\infty$
- ▶ **Stochastic definition 2:** A sequence $(x_n)_{n \in \mathbb{N}}$ of H-valued random variables is *stochastically quasi-Fejér monotone* with respect to F if, for every $z \in F$, there exist $(\chi_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$, $(\vartheta_n(z))_{n \in \mathbb{N}} \in \ell_+(\mathcal{X})$, and $(\eta_n(z))_{n \in \mathbb{N}} \in \ell_+^1(\mathcal{X})$ such that

$$(\forall n \in \mathbb{N}) \mathbb{E}(\phi(\|x_{n+1} - z\|) | \mathcal{X}_n) + \vartheta_n(z) \leq (1 + \chi_n(z))\phi(\|x_n - z\|) + \eta_n(z)$$

Suppose $(x_n)_{n \in \mathbb{N}}$ is stochastically quasi-Fejér monotone w.r.t. F. Then

- ▶ $(\forall z \in F) \left[\sum_{n \in \mathbb{N}} \vartheta_n(z) < +\infty \text{ P-a.s.} \right]$
- ▶ $[\mathfrak{W}(x_n)_{n \in \mathbb{N}} \subset F \text{ P-a.s.}] \Leftrightarrow [(x_n)_{n \in \mathbb{N}} \text{ converges weakly P-a.s. to an F-valued random variable}]$.

$\mathfrak{W}(x_n)_{n \in \mathbb{N}}$: set of weak sequential cluster points of $(x_n)_{n \in \mathbb{N}}$.

More General Problem

GOAL:

Find $x \in H$ such that $0 \in Ax + Bx$,

where

- $A: H \rightarrow 2^H$: maximally monotone operator, i.e.

$$(x, u) \in \text{gra } A \iff (\forall (y, v) \in \text{gra } A) \langle x - y \mid u - v \rangle \geq 0.$$

- If A is maximally monotone, then its resolvent $J_A = (\text{Id} + A)^{-1}$ is a firmly nonexpansive operator from H to H .

More General Problem

GOAL:

Find $x \in H$ such that $0 \in Ax + Bx$,

where

- $A: H \rightarrow 2^H$: maximally monotone operator, i.e.

$$(x, u) \in \text{gra } A \quad \Leftrightarrow \quad (\forall (y, v) \in \text{gra } A) \quad \langle x - y \mid u - v \rangle \geq 0.$$

- $B: H \rightarrow H$: ϑ -cocoercive operator, with $\vartheta \in]0, +\infty[$, i.e.

$$(\forall x \in H)(\forall y \in H) \quad \langle x - y \mid Bx - By \rangle \geq \vartheta \|Bx - By\|^2,$$

- $F = \text{zer}(A + B)$ assumed to be nonempty.

EXAMPLE: $A = \partial f$ with $f \in \Gamma_0(H)$ and $B = \nabla h$ with h convex with a ϑ^{-1} -Lipschitzian gradient.

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (J_{\gamma_n A_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (J_{\gamma_n A_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$
- $J_{\gamma_n A_n}$: resolvent of a maximally monotone operator
 $\gamma_n A_n: H \rightarrow 2^H$ approximating $\gamma_n A$

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (J_{\gamma_n A_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$
- $J_{\gamma_n A_n}$: resolvent of a maximally monotone operator
 $\gamma_n A_n: H \rightarrow 2^H$ approximating $\gamma_n A$
- u_n second-order random variable: approximation to Bx_n

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (J_{\gamma_n A_n}(x_n - \gamma_n u_n) + a_n - x_n),$$

where

- $\lambda_n \in]0, 1]$ and $\gamma_n \in]0, 2\vartheta[$
- $J_{\gamma_n A_n}$: resolvent of a maximally monotone operator
 $\gamma_n A_n : H \rightarrow 2^H$ approximating $\gamma_n A$
- u_n second-order random variable: approximation to Bx_n
- a_n second-order random variable: possible additional error term

Convergence Conditions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

where $\sigma(x_0, \dots, x_n)$ is the smallest σ -algebra generated by x_0, \dots, x_n .

Assumptions on the approximation to the cocoercive operator:

- ▶ $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n} \|E(u_n | \mathcal{X}_n) - Bx_n\| < +\infty$.
- ▶ For every $z \in F$, there exist sequences $(\tau_n)_{n \in \mathbb{N}} \in \ell_+$, $(\zeta_n(z))_{n \in \mathbb{N}} \in \ell_+^\infty(\mathcal{X})$ such that $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n} \zeta_n(z) < +\infty$ and

$$(\forall n \in \mathbb{N}) \quad E(\|u_n - E(u_n | \mathcal{X}_n)\|^2 | \mathcal{X}_n) \leq \tau_n \|Bx_n - Bz\|^2 + \zeta_n(z).$$

Convergence Conditions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

where $\sigma(x_0, \dots, x_n)$ is the smallest σ -algebra generated by x_0, \dots, x_n .

Assumptions on the resolvent approximation:

- ▶ 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

$$(\forall n \in \mathbb{N})(\forall x \in \mathbf{H}) \quad \|J_{\gamma_n A_n} x - J_{\gamma_n A} x\| \leq \alpha_n \|x\| + \beta_n.$$

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

Convergence Conditions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

where $\sigma(x_0, \dots, x_n)$ is the smallest σ -algebra generated by x_0, \dots, x_n .

Assumptions on the algorithm parameters:

- ▶ $\inf_{n \in \mathbb{N}} \gamma_n > 0$, $\sup_{n \in \mathbb{N}} \tau_n < +\infty$, and $\sup_{n \in \mathbb{N}} (1 + \tau_n) \gamma_n < 2\vartheta$.
- ▶ **Either** $\inf_{n \in \mathbb{N}} \lambda_n > 0$ **or** $[\gamma_n \equiv \gamma, \sum_{n \in \mathbb{N}} \tau_n < +\infty, \text{ and } \sum_{n \in \mathbb{N}} \lambda_n = +\infty]$.

Convergence Result

Under the previous assumptions, the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the algorithm **converges weakly a.s.** to an F -valued random variable.

In addition if A or B is demiregular at every $z \in F$, then the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the algorithm **converges strongly a.s.** to an F -valued random variable.

A is *demiregular* at $x \in \text{dom } A$ if, for every sequence $(x_n, u_n)_{n \in \mathbb{N}}$ in $\text{gra } A$ and every $u \in Ax$ such that $x_n \rightarrow x$ and $u_n \rightarrow u$, we have $x_n \rightarrow x$.

Example: A strongly monotone, i.e. there exists $\alpha \in]0, +\infty[$ such that $A - \alpha \text{Id}$ is monotone.

Primal-Dual Splitting

GOAL:

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

where

- H : real Hilbert space
- $f \in \Gamma_0(H)$
- $h: H \rightarrow \mathbb{R}$: differentiable convex function with ϑ^{-1} -Lipschitz continuous gradient
- $g_k \in \Gamma_0(G_k)$ with G_k real Hilbert space
- L_k : bounded linear operator from H to G_k
- $\exists \bar{x} \in H$ such that $0 \in \partial f(\bar{x}) + \sum_{k=1}^q L_k^* \partial g_k(L_k \bar{x}) + \nabla h(\bar{x})$.

Reformulation

Let

- ▶ $\mathbf{K} = \mathbf{H} \oplus \mathbf{G}$ with $\mathbf{G} = \mathbf{G}_1 \oplus \cdots \oplus \mathbf{G}_q$
- ▶ $\mathbf{g}: \mathbf{G} \rightarrow]-\infty, +\infty] : \mathbf{v} \mapsto \sum_{k=1}^q \mathbf{g}_k(\mathbf{v}_k)$
- ▶ $\mathbf{L}: \mathbf{H} \rightarrow \mathbf{G}: \mathbf{x} \mapsto (\mathbf{L}_k \mathbf{x})_{1 \leq k \leq q}$
- ▶ $\mathbf{A}: \mathbf{K} \rightarrow 2^{\mathbf{K}}: (\mathbf{x}, \mathbf{v}) \mapsto (\partial f(\mathbf{x}) + \mathbf{L}^* \mathbf{v}) \times (-\mathbf{L}\mathbf{x} + \partial \mathbf{g}^*(\mathbf{v}))$
- ▶ $\mathbf{B}: \mathbf{K} \rightarrow \mathbf{K}: (\mathbf{x}, \mathbf{v}) \mapsto (\nabla h(\mathbf{x}), \mathbf{0})$
- ▶ $\mathbf{V}: \mathbf{K} \rightarrow \mathbf{K}: (\mathbf{x}, \mathbf{v}) \mapsto (\rho^{-1}\mathbf{x} - \mathbf{L}^* \mathbf{v}, -\mathbf{L}\mathbf{x} + \mathbf{U}^{-1}\mathbf{v})$ with
 $\mathbf{U} = \text{Diag}(\sigma_1 \text{Id}, \dots, \sigma_q \text{Id})$ with $(\rho, \sigma_1, \dots, \sigma_q) \in]0, +\infty[^{q+1}$
 and $\rho \sum_{k=1}^q \sigma_k \|\mathbf{L}_k\|^2 < 1$.

In the renormed space $(\mathbf{K}, \|\cdot\|_{\mathbf{V}})$, $\mathbf{V}^{-1}\mathbf{A}$ is maximally monotone and $\mathbf{V}^{-1}\mathbf{B}$ is cocoercive. In addition, finding a zero of the sum of these operators is equivalent to finding a pair of primal-dual solutions.

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \dots$

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

where

- $\lambda_n \in]0, 1]$ with $\sum_{n \in \mathbb{N}} \lambda_n = +\infty$ and $(\rho^{-1} - \sum_{k=1}^q \sigma_k \|L_k\|^2) \vartheta > 1/2$

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \dots$

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

where

- $f_n \in \Gamma_0(H)$: approximation to f

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \dots$

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

where

- u_n second-order random variable: approximation to $\nabla h(x_n)$

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \dots$

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

where

- b_n and c_n second-order random variables: possible additional error terms

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \dots$

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

REMARKS:

- ★ Extension of the deterministic algorithms in
[Esser et al – 2010] [Chambolle and Pock – 2011] [Vũ – 2013] [Condat – 2013]
- ★ Parallel structure
- ★ No inversion of operators related to $(L_k)_{1 \leq k \leq q}$ required.

Assumptions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

Assumptions on the gradient approximation:

- ▶ $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n} \|E(u_n | \mathcal{X}_n) - \nabla h(x_n)\| < +\infty$.
- ▶ For every $z \in F$, there exists $(\zeta_n(z))_{n \in \mathbb{N}} \in \ell_+^\infty(\mathcal{X})$ such that $\sum_{n \in \mathbb{N}} \sqrt{\lambda_n \zeta_n(z)} < +\infty$ and

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

Assumptions

Let $\mathcal{X} = (\mathcal{X}_n)_{n \in \mathbb{N}}$ be a sequence of sigma-algebras such that

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

Assumptions on the prox approximations:

- ▶ 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

$$(\forall n \in \mathbb{N})(\forall x \in H) \quad \|\text{prox}_{\gamma_n f_n} x - \text{prox}_{\gamma_n f} x\| \leq \alpha_n \|x\| + \beta_n.$$

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

Convergence Result

- ▶ F : set of solutions to the primal problem
- ▶ F^* : set of solutions to the dual problem

Under the previous assumptions, the sequence $(x_n)_{n \in \mathbb{N}}$ converges weakly a.s. to an F -valued random variable and the sequence $(v_n)_{n \in \mathbb{N}}$ converges weakly a.s. to an F^* -valued random variable.

Online Image Recovery

OBSERVATION MODEL

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

where

- $\bar{x} \in \mathsf{H} = \mathbb{R}^N$: unknown image
- K_n : $\mathbb{R}^{M \times N}$ -valued random matrix
- e_n : \mathbb{R}^M -valued random noise vector.

OBJECTIVE

recover \bar{x} from $(K_n, z_n)_{n \in \mathbb{N}}$.

Application of Primal-Dual Algorithm

FORMULATION

- ▶ Mean square error criterion

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

assuming that $(K_n, z_n)_{n \in \mathbb{N}}$ are identically distributed

- ▶ Statistics of $(K_n, z_n)_{n \in \mathbb{N}}$ learnt online \Rightarrow Approximation to $\nabla \mathbf{h}(x_n)$:

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

where $(m_n)_{n \in \mathbb{N}}$ is strictly increasing sequence in \mathbb{N}

- ▶ f and $g_1 \circ L_1$ ($q = 1$): regularization terms

Application of Primal-Dual Algorithm

- Mean square error criterion

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

assuming that $(K_n, z_n)_{n \in \mathbb{N}}$ are identically distributed

- Statistics of $(K_n, z_n)_{n \in \mathbb{N}}$ learnt online \Rightarrow Approximation to $\nabla h(x_n)$:

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

where $(m_n)_{n \in \mathbb{N}}$ is strictly increasing sequence in \mathbb{N}

\Rightarrow recursive computation: $u_n = R_n x_n - c_n$ with

$$R_n = \frac{1}{m_{n+1}} \sum_{n'=0}^{m_{n+1}-1} K_{n'}^\top K_{n'} = \frac{m_n}{m_{n+1}} R_{n-1} + \frac{1}{m_{n+1}} \sum_{n'=m_n}^{m_{n+1}-1} K_{n'}^\top K_{n'}.$$

Application of Primal-Dual Algorithm

CONDITIONS FOR CONVERGENCE

- ▶ $(K_n, e_n)_{n \in \mathbb{N}}$ is an i.i.d. sequence such that $E\|K_0\|^4 < +\infty$ and $E\|e_0\|^4 < +\infty$.
- ▶ Approximation to $\nabla h(x_n)$:

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

where $(m_n)_{n \in \mathbb{N}}$ is strictly increasing sequence in \mathbb{N} such that $m_n = O(n^{1+\delta})$ with $\delta \in]0, +\infty[$.

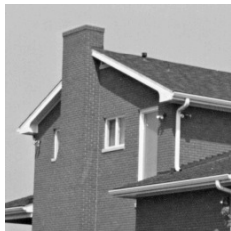
- ▶ $\lambda_n = O(n^{-\kappa})$, where $\kappa \in]1 - \delta, 1] \cap [0, 1]$.
- ▶ $f_n \equiv f$ and the domain of f is bounded.
- ▶ $b_n \equiv 0$ and $c_n \equiv 0$.

Simulation example

- ▶ Grayscale image of size 256×256 with pixel values in $[0, 255]$
- ▶ Stochastic blur (uniform i.i.d. subsampling of a uniform 5×5 blur performed in the discrete Fourier domain with 70% frequency bins set to zero).
- ▶ Additive white $\mathcal{N}(0, 5^2)$ noise.
- ▶ $f = \iota_{[0,255]^N}$ and $g_1 \circ L_1 =$ isotropic total variation.
- ▶ Parameter choice:

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

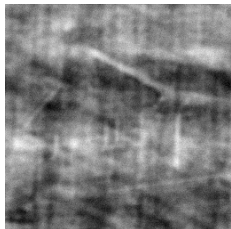
Simulation example



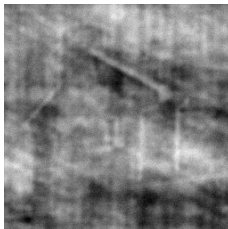
Original image \bar{x}



Restored image (SNR = 28.1 dB)

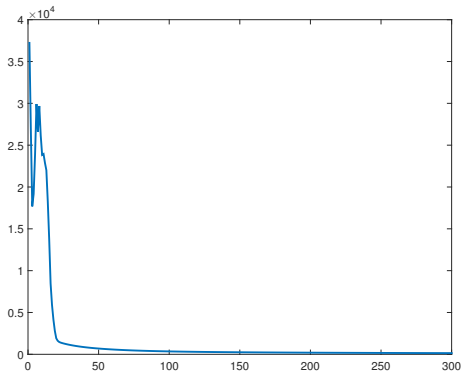


Degraded image 1 (SNR = 0.14 dB)



Degraded image 2 (SNR = 12.0 dB)

Simulation example



$\|x_n - x_\infty\|$ versus the iteration number n .

Conclusion

- Investigation of stochastic variants of Forward-Backward and Primal-Dual proximal algorithms.
- Stochastic approximations to both smooth and non smooth convex functions.
- Extension to monotone inclusion problems.
- Theoretical guaranties of convergence.
- Novel application to online image recovery.

Some references



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



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, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H.
Wolkowicz editors. Springer-Verlag, New York, pp. 185-212, 2011.



P. Combettes and J.-C. Pesquet
Stochastic quasi-Fejér block-coordinate fixed point iterations with random
sweeping
SIAM Journal on Optimization, vol. 25, no. 2, pp. 1221-1248, July 2015.



N. Komodakis and J.-C. Pesquet
Playing with duality: An overview of recent primal-dual approaches for solving
large-scale optimization problems
IEEE Signal Processing Magazine, vol. 32, no 6, pp. 31-54, Nov. 2015.



P. L. Combettes and J.-C. Pesquet
Stochastic approximations and perturbations in forward-backward splitting for
monotone operators
Pure and Applied Functional Analysis, vol. 1, no 1, pp. 13-37, Jan. 2016.