Stochastic Proximal Algorithms with Applications to Online Image Recovery

Patrick Louis Combettes[1] and Jean-Christophe Pesquet[2]

[1] Mathematics Department, North Carolina State University, Raleigh, USA
[2] Center for Visual Computing, CentraleSupelec, University Paris-Saclay, Grande Voie des Vignes, 92295 Châtenay-Malabry, France

S^3 Seminar - 24 March 2017

Outline

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*)
  $\rightsquigarrow$ large number of variables (inverse problems)
  $\rightsquigarrow$ large number of observations (machine learning)
- Use of more sophisticated data structures
  (*graph signal processing*)

Variational formulation

GOAL:

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

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
- $h\colon H \to \mathbb{R}$: differentiable convex function such that ∇h is ϑ^{-1}-Lipschitz continuous with $\vartheta \in]0, +\infty[$
- $F = \operatorname{Argmin}(f + h)$ assumed to be nonempty.

Algorithm

CLASSICAL SOLUTION [Combettes and Wajs - 2005]

FORWARD-BACKWARD ALGORITHM

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

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

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

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

Algorithm

CLASSICAL SOLUTION [Combettes and Wajs - 2005]

FORWARD-BACKWARD ALGORITHM

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

In the context of online processing and machine learning,
what to do if $\nabla \mathsf{h}$ and f are not known exactly ?

Proposed Solution

Stochastic FB Algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\mathrm{prox}_{\gamma_n \mathsf{f}_n}(x_n - \gamma_n u_n) + a_n - x_n\big),$$

where

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

Proposed Solution

STOCHASTIC FB ALGORITHM

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

where

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

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\mathrm{prox}_{\gamma_n \mathsf{f}_n}(x_n - \gamma_n u_n) + a_n - x_n\big),$$

where

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

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\mathrm{prox}_{\gamma_n \mathsf{f}_n}(x_n - \gamma_n u_n) + a_n - x_n\big),$$

where

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

Assumptions

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

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

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

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

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

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

Assumptions

Let $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ 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:

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

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

Assumptions

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

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

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

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

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

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

Assumptions

Let $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ 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 $\left[\gamma_n \equiv \gamma, \sum_{n\in\mathbb{N}} \tau_n < +\infty, \text{ and } \sum_{n\in\mathbb{N}} \lambda_n = +\infty\right]$.

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

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

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

Stochastic Quasi-Fejér Sequences

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

$$(\forall n\in\mathbb{N})\quad \phi(\|\mathsf{x}_{n+1}-\mathsf{z}\|)\leqslant\phi(\|\mathsf{x}_n-\mathsf{z}\|)$$

Stochastic Quasi-Fejér Sequences

- Let $\phi\colon\ [0,+\infty[\,\to[0,+\infty[$, $\phi(t)\uparrow+\infty$ as $t\to+\infty$
- **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 $\mathsf{z}\in$ F,

$$(\forall n\in\mathbb{N})\ \mathsf{E}(\phi(\|x_{n+1}-\mathsf{z}\|)\,|\,\mathcal{X}_n)\leqslant\phi(\|x_n-\mathsf{z}\|)$$

Stochastic Quasi-Fejér Sequences

- Let $\phi\colon\ [0,+\infty[\ \to [0,+\infty[$, $\phi(t)\uparrow +\infty$ as $t\to+\infty$
- **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 $\mathsf{z}\in\mathsf{F}$, there exist $(\chi_n(\mathsf{z}))_{n\in\mathbb{N}}\in\ell_+^1(\mathscr{X})$, $(\vartheta_n(\mathsf{z}))_{n\in\mathbb{N}}\in\ell_+(\mathscr{X})$, and $(\eta_n(\mathsf{z}))_{n\in\mathbb{N}}\in\ell_+^1(\mathscr{X})$ such that

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

Stochastic Quasi-Fejér Sequences

- Let $\phi\colon\ [0,+\infty[\ \to [0,+\infty[$, $\phi(t) \uparrow +\infty$ as $t \to +\infty$
- **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 $\mathsf{z} \in \mathsf{F}$, there exist $(\chi_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, $(\vartheta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, and $(\eta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$ such that

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

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

- $(\forall \mathsf{z} \in \mathsf{F})\ \left[\ \sum_{n\in\mathbb{N}} \vartheta_n(\mathsf{z}) < +\infty\ \text{P-a.s.}\right]$
- $[\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{F}\ \ \text{P-a.s.}] \Leftrightarrow [(x_n)_{n\in\mathbb{N}}$ 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:

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

where

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

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

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

More General Problem

GOAL:

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

where

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

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

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

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

- $F = \operatorname{zer}(A + B)$ 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

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \Big(\mathsf{J}_{\gamma_n \mathsf{A}_n}(x_n - \gamma_n u_n) + a_n - x_n \Big),$$

where

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

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\mathsf{J}_{\gamma_n \mathsf{A}_n}(x_n - \gamma_n u_n) + a_n - x_n \big),$$

where

- $\lambda_n \in \,]0,1]$ and $\gamma_n \in \,]0, 2\vartheta[$
- $\mathsf{J}_{\gamma_n \mathsf{A}_n}$: resolvent of a maximally monotone operator $\gamma_n \mathsf{A}_n \colon \mathsf{H} \to 2^{\mathsf{H}}$ approximating $\gamma_n \mathsf{A}$

Proposed Solution

Stochastic FB Algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\mathsf{J}_{\gamma_n \mathsf{A}_n}(x_n - \gamma_n u_n) + a_n - x_n \big),$$

where

- $\lambda_n \in \,]0,1]$ and $\gamma_n \in \,]0, 2\vartheta[$
- $\mathsf{J}_{\gamma_n \mathsf{A}_n}$: resolvent of a maximally monotone operator $\gamma_n \mathsf{A}_n \colon \mathsf{H} \to 2^{\mathsf{H}}$ approximating $\gamma_n \mathsf{A}$
- u_n second-order random variable: approximation to $\mathsf{B} x_n$

Proposed Solution

STOCHASTIC FB ALGORITHM

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\mathsf{J}_{\gamma_n \mathsf{A}_n}(x_n - \gamma_n u_n) + a_n - x_n \big),$$

where

- $\lambda_n \in]0,1]$ and $\gamma_n \in]0, 2\vartheta[$
- $\mathsf{J}_{\gamma_n \mathsf{A}_n}$: resolvent of a maximally monotone operator $\gamma_n \mathsf{A}_n \colon \mathsf{H} \to 2^{\mathsf{H}}$ approximating $\gamma_n \mathsf{A}$
- u_n second-order random variable: approximation to $\mathsf{B} x_n$
- a_n second-order random variable: possible additional error term

Convergence Conditions

Let $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ 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:

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

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

Convergence Conditions

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

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

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

<u>Assumptions on the resolvent approximation:</u>

- 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

$$(\forall n \in \mathbb{N})(\forall \mathsf{x} \in \mathsf{H}) \quad \|\mathsf{J}_{\gamma_n \mathsf{A}_n}\mathsf{x} - \mathsf{J}_{\gamma_n \mathsf{A}}\mathsf{x}\| \leqslant \alpha_n\|\mathsf{x}\| + \beta_n.$$

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

Convergence Conditions

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

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

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

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 $\left[\gamma_n \equiv \gamma, \sum_{n\in\mathbb{N}} \tau_n < +\infty\right.$, and $\left.\sum_{n\in\mathbb{N}} \lambda_n = +\infty\right]$.

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 $x \in \operatorname{dom} A$ if, for every sequence $(x_n, u_n)_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ and every $u \in Ax$ such that $x_n \rightharpoonup x$ and $u_n \to u$, we have $x_n \to x$.
Example: A strongly monotone, i.e. there exists $\alpha \in]0, +\infty[$ such that $A - \alpha \mathrm{Id}$ is monotone.

Primal-Dual Splitting

GOAL:

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

where

- H: real Hilbert space
- $f \in \Gamma_0(H)$
- $h\colon H \to \mathbb{R}$: 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
- $\exists \overline{x} \in H$ such that $0 \in \partial f(\overline{x}) + \sum_{k=1}^{q} L_k^* \partial g_k(L_k \overline{x}) + \nabla h(\overline{x})$.

Reformulation

Let

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

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 $n = 0, 1, \dots$

$$
\left\lfloor
\begin{array}{l}
y_n = \mathrm{prox}_{\rho \mathsf{f}_n} \left(x_n - \rho \left(\sum_{k=1}^{q} \mathsf{L}_k^* v_{k,n} + u_n \right) \right) + b_n \\
x_{n+1} = x_n + \lambda_n (y_n - x_n) \\
\text{for } k = 1, \dots, q \\
\left\lfloor
\begin{array}{l}
w_{k,n} = \mathrm{prox}_{\sigma_k \mathsf{g}_k^*} \left(v_{k,n} + \sigma_k \mathsf{L}_k (2y_n - x_n) \right) + c_{k,n} \\
v_{k,n+1} = v_{k,n} + \lambda_n (w_{k,n} - v_{k,n}).
\end{array}
\right.
\end{array}
\right.
$$

where

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

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \dots$

$$y_n = \text{prox}_{\rho \mathsf{f}_n}\left(x_n - \rho\left(\sum_{k=1}^{q} \mathsf{L}_k^* v_{k,n} + u_n\right)\right) + b_n$$

$$x_{n+1} = x_n + \lambda_n (y_n - x_n)$$

for $k = 1, \dots, q$

$$w_{k,n} = \text{prox}_{\sigma_k \mathsf{g}_k^*}\left(v_{k,n} + \sigma_k \mathsf{L}_k (2y_n - x_n)\right) + c_{k,n}$$

$$v_{k,n+1} = v_{k,n} + \lambda_n (w_{k,n} - v_{k,n}).$$

where

- $\mathsf{f}_n \in \Gamma_0(\mathsf{H})$: approximation to f

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \ldots$

$$y_n = \operatorname{prox}_{\rho \mathsf{f}_n} \left(x_n - \rho \Big(\sum_{k=1}^{q} \mathsf{L}_k^* v_{k,n} + u_n \Big) \right) + b_n$$

$$x_{n+1} = x_n + \lambda_n (y_n - x_n)$$

for $k = 1, \ldots, q$

$$w_{k,n} = \operatorname{prox}_{\sigma_k \mathsf{g}_k^*} \big(v_{k,n} + \sigma_k \mathsf{L}_k (2y_n - x_n) \big) + c_{k,n}$$

$$v_{k,n+1} = v_{k,n} + \lambda_n (w_{k,n} - v_{k,n}).$$

where

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

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \ldots$

$$y_n = \text{prox}_{\rho \mathsf{f}_n}\left(x_n - \rho\left(\sum_{k=1}^{q} \mathsf{L}_k^* v_{k,n} + u_n\right)\right) + b_n$$

$$x_{n+1} = x_n + \lambda_n(y_n - x_n)$$

for $k = 1, \ldots, q$

$$w_{k,n} = \text{prox}_{\sigma_k \mathsf{g}_k^*}\big(v_{k,n} + \sigma_k \mathsf{L}_k(2y_n - x_n)\big) + c_{k,n}$$

$$v_{k,n+1} = v_{k,n} + \lambda_n(w_{k,n} - v_{k,n}).$$

where

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

Resulting Algorithm

STOCHASTIC PRIMAL-DUAL ALGORITHM

for $n = 0, 1, \ldots$

$$y_n = \mathrm{prox}_{\rho \mathsf{f}_n}\left(x_n - \rho\left(\sum_{k=1}^{q} \mathsf{L}_k^* v_{k,n} + u_n\right)\right) + b_n$$

$$x_{n+1} = x_n + \lambda_n (y_n - x_n)$$

for $k = 1, \ldots, q$

$$w_{k,n} = \mathrm{prox}_{\sigma_k \mathsf{g}_k^*}\big(v_{k,n} + \sigma_k \mathsf{L}_k (2y_n - x_n)\big) + c_{k,n}$$

$$v_{k,n+1} = v_{k,n} + \lambda_n (w_{k,n} - v_{k,n}).$$

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 $(\mathsf{L}_k)_{1 \leqslant k \leqslant q}$ required.

Assumptions

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

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

Assumptions on the gradient approximation:

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

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

Assumptions

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

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

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

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

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

Convergence Result

- F: set of solutions to the primal problem
- $\mathbf{F}^*$: 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 $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ converges weakly a.s. to an $\mathbf{F}^*$-valued random variable.

Online Image Recovery

Observation model

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

where

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

Objective

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

Application of Primal-Dual Algorithm

Formulation

- Mean square error criterion

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

  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 $\nabla\mathsf{h}(x_n)$:

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

  where $(m_n)_{n\in\mathbb{N}}$ is strictly increasing sequence in $\mathbb{N}$
- f and $\mathsf{g}_1 \circ \mathsf{L}_1$ $(q = 1)$: regularization terms

Application of Primal-Dual Algorithm

- Mean square error criterion

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

  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 $\nabla\mathsf{h}(x_n)$:

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

  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 $\mathsf{E}\|K_0\|^4 < +\infty$ and $\mathsf{E}\|e_0\|^4 < +\infty$.
- Approximation to $\nabla \mathsf{h}(x_n)$:

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

  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 $\kappa \in]1-\delta, 1] \cap [0, 1]$.
- $\mathsf{f}_n \equiv \mathsf{f}$ and the domain of f is bounded.
- $b_n \equiv 0$ and $\boldsymbol{c}_n \equiv 0$.

Simulation example

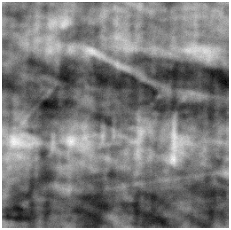
- Grayscale image of size 256×256 with pixel values in $[0, 255]$
- Stochastic blur (uniform i.i.d. subsampling of a uniform 5×5 blur performed in the discrete Fourier domain with 70% frequency bins set to zero).
- Additive white $\mathcal{N}(0, 5^2)$ noise.
- $\mathsf{f} = \iota_{[0,255]^N}$ and $\mathsf{g}_1 \circ \mathsf{L}_1 =$ isotropic total variation.
- Parameter choice:

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

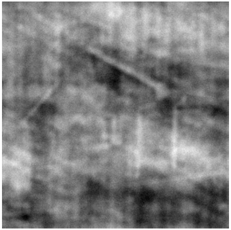
Simulation example

Original image $\overline{x}$

Restored image (SNR = 28.1 dB)

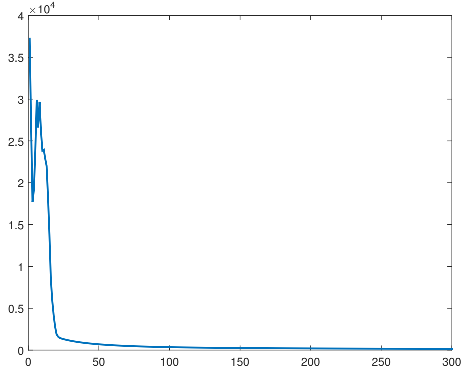


Degraded image 1 (SNR = 0.14 dB)



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.