Stochastic splitting methods for solving large-scale convex optimization problems

Jean-Christophe PESQUET

Laboratoire d'Informatique Gaspard Monge - CNRS
Univ. Paris-Est, France

SESO 2015

In collaboration with

E. Chouzenoux

Univ. Paris-Est

P. Combettes

Univ. PM Curie

A. Repetti

Univ. Paris-Est

Context

Fast development of optimization methods over the last decade

Why?

Context

Fast development of optimization methods over the last decade

Why?

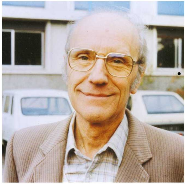
- Interest in nonsmooth cost functions (*sparsity*)
- Need for optimal processing of massive datasets (*big data*)
  $\rightsquigarrow$ large number of variables
  (inverse problems, energy management)
  $\rightsquigarrow$ large number of observations (machine learning)
- Use of more sophisticated data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

Proximity operator

Let H be a real Hilbert space.
$\Gamma_0(\mathsf{H})$ is the class of lower-semicontinuous convex functions from H to $]-\infty,+\infty]$ with a nonempty domain.

Let $f \in \Gamma_0(\mathsf{H})$. Let $\mathsf{U} : \mathsf{H} \to \mathsf{H}$ be a strongly positive self-adjoint linear operator.
The proximity operator $\mathrm{prox}_f^{\mathsf{U}}(x)$ of f at $x \in \mathsf{H}$ relative to the metric induced by U is the unique vector $\widehat{y} \in \mathsf{H}$ such that

$$f(\widehat{y}) + \frac{1}{2}\|\widehat{y} - x\|_{\mathsf{U}}^2 = \inf_{y \in \mathsf{H}} f(y) + \frac{1}{2}\langle y - x \mid \mathsf{U}(y - x)\rangle.$$



Jean-Jacques Moreau
(1923–2014)

Proximity operator

Let $f \in \Gamma_0(\mathsf{H})$. Let $\mathsf{U} : \mathsf{H} \to \mathsf{H}$ be a strongly positive self-adjoint linear operator.
The proximity operator $\mathrm{prox}_f^{\mathsf{U}}(x)$ of f at $x \in \mathsf{H}$ relative to the metric induced by U is the unique vector $\widehat{y} \in \mathsf{H}$ such that

$$f(\widehat{y}) + \frac{1}{2}\|\widehat{y} - x\|_{\mathsf{U}}^2 = \inf_{y \in \mathsf{H}} f(y) + \frac{1}{2}\langle y - x \mid \mathsf{U}(y - x)\rangle.$$

Remark:

- If C is a nonempty closed convex subset of H. Let

$$(\forall x \in \mathsf{H}) \quad \iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

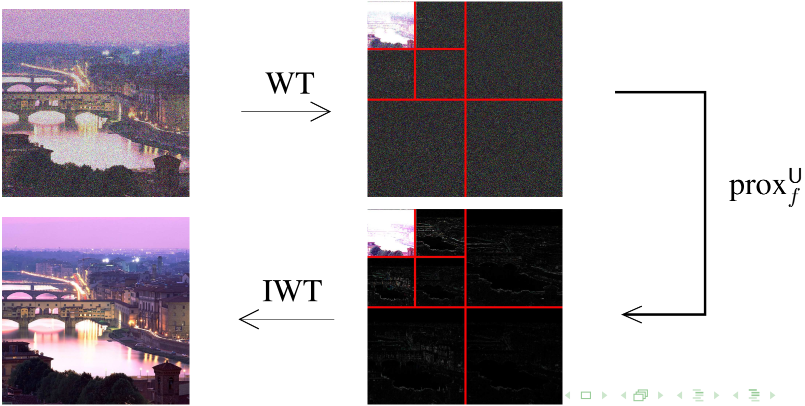
Then, $\mathrm{prox}_{\iota_C}^{\mathrm{Id}}$ reduces to the projection operator onto C.

Proximity operator: Bayesian interpretation

- If $\mathsf{H} = \mathbb{R}^N$ and

$$x = \overline{y} + w$$

where $\overline{y}$ is a realization of a random vector with probability density function $\exp(-f)$ and w is a realization of a $\mathcal{N}(0, \mathsf{U}^{-1})$ noise, then $\mathrm{prox}_f^{\mathsf{U}}(x)$ is a Maximum A Posteriori estimate of $\overline{y}$.



Examples of proximity operators

- Explicit form for objective functions associated to the usual log-concave probability densities when $\mathrm{U} = \mathrm{Id}$. [Chaux *et al.* - 2007]
    - Laplace
    - Generalized Gaussian
    - maximum entropy
    - gamma
    - uniform
    - Weibull
    - Generalized inverse Gaussian
    - Gaussian
    - Huber
    - Smoothed Laplace
    - chi
    - triangular
    - Pearson type I
    - ...

- And many other functions ! [Combettes and Pesquet - 2011]

Minimization problems

- Minimization of one function
  Let H be a Hilbert space. Let $f \in \Gamma_0(\mathsf{H})$.
  We want to

$$\underset{x\in\mathsf{H}}{\text{minimize}} \;\; f(x)$$

Minimization problems

- Minimization of one function
  Let H be a Hilbert space. Let $f \in \Gamma_0(\mathsf{H})$.
  We want to

$$\underset{x\in\mathsf{H}}{\text{minimize}} \;\; f(x)$$

- Standard first-order algorithmic solutions
  - gradient algorithm

$$\boxed{(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla f(x_n)}$$

$\rightsquigarrow$ f β- Lipschitz differentiable with $\beta \in]0,+\infty[$
$0 < \inf_{n\in\mathbb{N}} \gamma_n$ and $\sup_{n\in\mathbb{N}} \gamma_n < 2\beta^{-1}$

Minimization problems

- Minimization of one function
  Let H be a Hilbert space. Let $f \in \Gamma_0(\mathsf{H})$.
  We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \quad f(x)$$

- Standard first-order algorithmic solutions
  - gradient algorithm

$$\boxed{(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla f(x_n)}$$

$\rightsquigarrow f$ β- Lipschitz differentiable with $\beta \in]0, +\infty[$
$0 < \inf_{n \in \mathbb{N}} \gamma_n$ and $\sup_{n \in \mathbb{N}} \gamma_n < 2\beta^{-1}$

  - proximal point algorithm

$$\boxed{(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}^{\text{Id}}_{\gamma_n f} \, x_n}$$

$(\gamma_n)_{n \in \mathbb{N}}$ sequence in $]0, +\infty[$ such that $\sum_{n=1}^{+\infty} \gamma_n = +\infty$
$\rightsquigarrow f$ proximable

Minimization problems

- Splitting in a sum of 2 functions
  Let H be a Hilbert space. Let $g_1 \in \Gamma_0(\mathsf{H})$ and $g_2 \in \Gamma_0(\mathsf{H})$.
  We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \;\; g_1(x) + g_2(x)$$

Minimization problems

- Splitting in a sum of 2 functions
  Let H be a Hilbert space. Let $g_1 \in \Gamma_0(\mathsf{H})$ and $g_2 \in \Gamma_0(\mathsf{H})$.
  We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \;\; g_1(x) + g_2(x)$$

- Standard algorithmic solutions
  - forward-backward algorithm

$$\boxed{(\forall n \in \mathbb{N}) \quad x_{n+1} = \text{prox}^{\text{Id}}_{\gamma_n g_1}\left(x_n - \gamma_n \nabla g_2(x_n)\right)}$$

g_2 β-Lipschitz differentiable with $\beta \in]0, +\infty[$
$0 < \inf_{n \in \mathbb{N}} \gamma_n$ and $\sup_{n \in \mathbb{N}} \gamma_n < 2\beta^{-1}$
Particular case: projected gradient algorithm when $g_1 = \iota_C$ where C nonempty closed convex subset of H

Minimization problems

- Splitting in a sum of 2 functions
  Let H be a Hilbert space. Let $g_1 \in \Gamma_0(\mathsf{H})$ and $g_2 \in \Gamma_0(\mathsf{H})$.
  We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \;\; g_1(x) + g_2(x)$$

- Standard algorithmic solutions
  - forward-backward algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \Big(\mathrm{prox}^{\mathsf{U}^{-1}}_{\gamma_n g_1} \big(x_n - \gamma_n \mathsf{U} \nabla g_2(x_n) \big) - x_n \Big)$$

g_2 β-Lipschitz differentiable with $\beta \in]0, +\infty[$
$0 < \inf_{n \in \mathbb{N}} \gamma_n$ and $\sup_{n \in \mathbb{N}} \gamma_n < 2\beta^{-1}$
$\mathsf{U}\colon \mathsf{H} \to \mathsf{H}$ strongly positive self-adjoint linear operator
$(\lambda_n)_{n \in \mathbb{N}}$ sequence in $]0, 1]$ such that $\inf_{n \in \mathbb{N}} \lambda_n > 0$

Minimization problems

- Splitting in a sum of 2 functions
  Let H be a Hilbert space. Let $g_1 \in \Gamma_0(\mathsf{H})$ and $g_2 \in \Gamma_0(\mathsf{H})$.
  We want to

$$\underset{x\in\mathsf{H}}{\text{minimize}} \;\; g_1(x) + g_2(x)$$

- Standard algorithmic solutions
  - forward-backward algorithm

$$\boxed{(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\mathrm{prox}^{\mathsf{U}^{-1}}_{\gamma_n g_1}\big(x_n - \gamma_n \mathsf{U}\nabla g_2(x_n)\big) - x_n\big)}$$

  - Douglas-Rachford algorithm

$$\boxed{(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \frac{\lambda_n}{2}\Big((2\mathrm{prox}^{\mathrm{Id}}_{\gamma g_1} - \mathrm{Id}\,) \circ (2\mathrm{prox}^{\mathrm{Id}}_{\gamma g_2} - \mathrm{Id}\,)x_n - x_n\Big)}$$

  $\gamma \in \,]0,+\infty[$, $(\lambda_n)_{n\in\mathbb{N}}$ sequence in $[0,2]$ such that
  $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$
  Particular case: **ADMM amounts to applying DR to the dual problem**.

Minimization problems

- Splitting in a sum of $q \geqslant 2$ functions
  Let H be a Hilbert space.
  For every $k \in \{1, \dots, q\}$, let $g_k \in \Gamma_0(\mathsf{H})$. We want to

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

Minimization problems

- Splitting in a sum of $q \geqslant 2$ functions
  Let H be a Hilbert space.
  For every $k \in \{1, \ldots, q\}$, let G_k be a Hilbert space, let $g_k \in \Gamma_0(\mathsf{G}_k)$, let $\mathsf{L}_k \colon \mathsf{H} \to \mathsf{G}_k$ be a linear and bounded operator. We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \quad \sum_{k=1}^{q} g_k(\mathsf{L}_k x)$$

Minimization problems

- Splitting in a sum of $q \geqslant 2$ functions
  Let H be a Hilbert space.
  For every $k \in \{1, \ldots, q\}$, let G_k be a Hilbert space, let $g_k \in \Gamma_0(\mathsf{G}_k)$, let $\mathsf{L}_k \colon \mathsf{H} \to \mathsf{G}_k$ be a linear and bounded operator. We want to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \quad \sum_{k=1}^{q} g_k(\mathsf{L}_k x)$$

- Algorithmic solutions
  - Simultaneous direction method of multipliers-like algorithms (PPXA, PPXA+,...)
  - Proximal primal-dual algorithms

Proximal primal-dual algorithm

Advantages:

- No linear operator inversion.
- Use of proximable or/and differentiable functions.
- Amenable to parallel implementations.
- Special cases: Forward-Backward and Douglas-Rachford algorithms.

Proximal primal-dual algorithm

Advantages:

- No linear operator inversion.
- Use of proximable or/and differentiable functions.
- Amenable to parallel implementations.
- Special cases: Forward-Backward and Douglas-Rachford algorithms.

Bibliographical remarks:

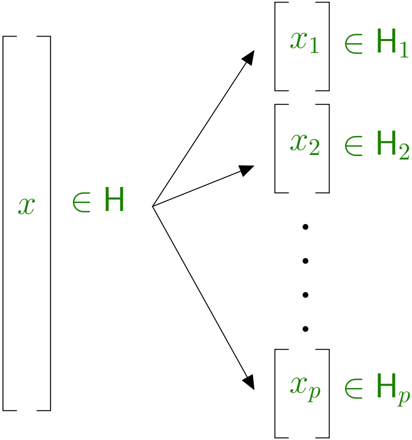
- Methods based on Forward-Backward iteration
  - type I: [Vu - 2013][Condat - 2013]
    (extensions of [Esser *et al.* - 2010][Chambolle and Pock - 2011])
  - type II: [Combettes *et al.* - 2014]
    (extensions of [Loris and Verhoeven - 2011][Chen *et al.* - 2014])
- Methods based on Forward-Backward-Forward iteration
  [Combettes and Pesquet - 2012] [Boţ and Hendrich,2014]
- Projection based methods
  [Alotaibi *et al.* - 2013]
- ...

Acceleration via block alternation

- Idea: variable splitting.

Acceleration via block alternation

- Idea: variable splitting.



$$\times_{j=1}^{p} \mathsf{H}_j = \mathsf{H}$$

where $\mathsf{H}_1, \ldots, \mathsf{H}_p$ separable real Hilbert spaces.

Block-coordinate strategy

$\Rightarrow$ At each iteration n, update only a subset of components ($\sim$ Gauss-Seidel).

Block-coordinate strategy

$\Rightarrow$ At each iteration n, update only a subset of components ($\sim$ Gauss-Seidel).

Advantages:

- Reduced complexity
- Less memory requirements per iteration
- More flexibility

$\Rightarrow$ Useful for large-scale optimization (see e.g. [Richtárik andTakáč, 2014])

Parallel proximal primal-dual algorithm

Optimization problem: Minimization of

$$(\forall x \in \mathsf{H}) \qquad \Phi(x) = \sum_{j=1}^{p} \big(f_j(x_j) + h_j(x_j)\big) + \sum_{k=1}^{q} (g_k \,\square\, l_k)\big(\mathsf{L}_k x\big)$$

Parallel proximal primal-dual algorithm

Optimization problem: Minimization of

$$(\forall x \in \mathsf{H}) \qquad \Phi(x) = \sum_{j=1}^{p} \big(f_j(x_j) + h_j(x_j)\big) + \sum_{k=1}^{q} (g_k \,\square\, l_k)\big(\mathsf{L}_k x\big)$$

where

- $(\forall j \in \{1, \ldots, p\})\ f_j \in \Gamma_0(\mathsf{H}_j)$
- $h_j \in \Gamma_0(\mathsf{H}_j)$ μ_j Lipschitz-differentiable with $\mu_j \in \,]0, +\infty[$
- $(\forall k \in \{1, \ldots, q\})\ g_k \in \Gamma_0(\mathsf{G}_k)$, G_k separable real Hilbert space
- $l_k \in \Gamma_0(\mathsf{G}_k)$ ν_k-strongly convex with $\nu_k \in \,]0, +\infty[$

Parallel proximal primal-dual algorithm

Optimization problem: Minimization of

$$(\forall x \in \mathsf{H}) \qquad \Phi(x) = \sum_{j=1}^{p} \big(f_j(x_j) + h_j(x_j)\big) + \sum_{k=1}^{q} (g_k \,\square\, l_k)\big(\mathsf{L}_k x\big)$$

where

- $(\forall j \in \{1,\ldots,p\})$ $f_j \in \Gamma_0(\mathsf{H}_j)$
- $h_j \in \Gamma_0(\mathsf{H}_j)$ μ_j Lipschitz-differentiable with $\mu_j \in \,]0,+\infty[$
- $(\forall k \in \{1,\ldots,q\})$ $g_k \in \Gamma_0(\mathsf{G}_k)$, G_k separable real Hilbert space
- $l_k \in \Gamma_0(\mathsf{G}_k)$ ν_k-strongly convex with $\nu_k \in \,]0,+\infty[$
- $g_k \,\square\, l_k$ inf-convolution of g_k and l_k:

$$(\forall v_k \in \mathsf{G}_k) \qquad (g_k \,\square\, l_k)(v_k) = \inf_{v_k' \in \mathsf{G}_k} g_k(v_k') + h_k(v_k - v_k')$$

$$g_k \,\square\, \iota_{\{0\}} = g_k$$

Parallel proximal primal-dual algorithm

Optimization problem: Minimization of

$$(\forall x \in \mathsf{H}) \qquad \Phi(x) = \sum_{j=1}^{p} \big(f_j(x_j) + h_j(x_j)\big) + \sum_{k=1}^{q} (g_k \,\Box\, l_k)\big(\mathsf{L}_k x\big)$$

where

- $(\forall j \in \{1,\ldots,p\})\ f_j \in \Gamma_0(\mathsf{H}_j)$
- $h_j \in \Gamma_0(\mathsf{H}_j)$ μ_j Lipschitz-differentiable with $\mu_j \in]0,+\infty[$
- $(\forall k \in \{1,\ldots,q\})\ g_k \in \Gamma_0(\mathsf{G}_k)$, G_k separable real Hilbert space
- $l_k \in \Gamma_0(\mathsf{G}_k)$ ν_k-strongly convex with $\nu_k \in]0,+\infty[$
- $g_k \,\Box\, l_k$ inf-convolution of g_k and l_k
- $(\forall x \in \mathsf{H})\ \mathsf{L}_k x = \sum_{j=1}^{p} \mathsf{L}_{k,j} x_j$, where
  $(\forall j \in \{1,\ldots,p\})\ \mathsf{L}_{k,j} \colon \mathsf{H}_j \to \mathsf{G}_k$ bounded linear operator with

$$\mathbb{L}_k = \big\{ j \in \{1,\ldots,p\} \;\big|\; \mathsf{L}_{k,j} \neq 0 \big\} \neq \varnothing$$
$$\mathbb{L}_j^* = \big\{ k \in \{1,\ldots,q\} \;\big|\; \mathsf{L}_{k,j} \neq 0 \big\} \neq \varnothing.$$

Block parallel proximal primal-dual problem

Primal problem: Find an element of the set F of solutions to

$$\underset{x_1 \in \mathsf{H}_1,\dots,x_p \in \mathsf{H}_p}{\text{minimize}} \quad \sum_{j=1}^{p} \big(f_j(x_j) + h_j(x_j)\big) + \sum_{k=1}^{q} (g_k \,\square\, l_k)\left(\sum_{j=1}^{p} \mathsf{L}_{k,j} x_j\right)$$

Dual problem: Find an element of the set F^* of solutions to

$$\underset{v_1 \in \mathsf{G}_1,\dots,v_q \in \mathsf{G}_q}{\text{minimize}} \quad \sum_{j=1}^{p} (f_j^* \,\square\, h_j^*)\left(-\sum_{k=1}^{q} \mathsf{L}_{k,j}^* v_k\right) + \sum_{k=1}^{q} \big(g_k^*(v_k) + l_k^*(v_k)\big)$$

For every $j \in \{1,\dots,p\}$, $f_j^* \in \Gamma_0(\mathsf{H}_j)$ is the conjugate of f_j such that

$$(\forall y_j \in \mathsf{H}_j) \qquad f_j^*(y_j) = \sup_{x_j \in \mathsf{H}_j} \langle x_j \mid y_j \rangle - f_j(x_j).$$

Block parallel proximal primal-dual problem

Primal problem : Find an element of the set F of solutions to

$$\underset{x_1\in\mathsf{H}_1,\ldots,x_p\in\mathsf{H}_p}{\text{minimize}} \quad \sum_{j=1}^{p}\big(f_j(x_j)+h_j(x_j)\big)+\sum_{k=1}^{q}(g_k\,\square\, l_k)\Big(\sum_{j=1}^{p}\mathsf{L}_{k,j}x_j\Big)$$

Dual problem : Find an element of the set F^* of solutions to

$$\underset{v_1\in\mathsf{G}_1,\ldots,v_q\in\mathsf{G}_q}{\text{minimize}} \quad \sum_{j=1}^{p}(f_j^*\,\square\, h_j^*)\Big(-\sum_{k=1}^{q}\mathsf{L}_{k,j}^*v_k\Big)+\sum_{k=1}^{q}\big(g_k^*(v_k)+l_k^*(v_k)\big)$$

We assume that there exists $(\overline{x}_1,\ldots,\overline{x}_p)\in\mathsf{H}_1\times\cdots\times\mathsf{H}_p$ such that

$$(\forall j\in\{1,\ldots,p\})\quad 0\in\partial f_j(\overline{x}_j)+\nabla h_j(\overline{x}_j) + \sum_{k=1}^{q}\mathsf{L}_{k,j}^*\partial(g_k\,\square\, l_k)\Big(\sum_{j'=1}^{p}\mathsf{L}_{k,j'}\overline{x}_{j'}\Big).$$

Random block-coordinate primal-dual algorithm

for $n = 0, 1, \ldots$
  for $k = 1, \ldots, q$
    $u_{k,n} = \varepsilon_{p+k,n}\Big(\mathrm{prox}_{g_k^*}^{\mathsf{U}_k^{-1}}\big(v_{k,n} + \mathsf{U}_k(\sum_{j\in\mathbb{L}_k} \mathsf{L}_{k,j} x_{j,n} - \nabla l_k^*(v_{k,n}) + d_{k,n})\big) + b_{k,n}\Big)$
    $v_{k,n+1} = v_{k,n} + \lambda_n \varepsilon_{p+k,n}(u_{k,n} - v_{k,n}),$
  for $j = 1, \ldots, p$
    $y_{j,n} = \varepsilon_{j,n}\Big(\mathrm{prox}_{f_j}^{\mathsf{W}_j^{-1}}\big(x_{j,n} - \mathsf{W}_j(\sum_{k\in\mathbb{L}_j^*} \mathsf{L}_{k,j}^*(2u_{k,n} - v_{k,n}) + \nabla h_j(x_{j,n}) + c_{j,n}\big) + a_{j,n}\Big)$
    $x_{j,n+1} = x_{j,n} + \lambda_n \varepsilon_{j,n}(y_{j,n} - x_{j,n})$

Random block-coordinate primal-dual algorithm

for $n = 0, 1, \ldots$
  for $k = 1, \ldots, q$
    $u_{k,n} = \varepsilon_{p+k,n}\Big(\mathrm{prox}^{\mathsf{U}_k^{-1}}_{g_k^*}\big(v_{k,n} + \mathsf{U}_k(\sum_{j\in\mathbb{L}_k} \mathsf{L}_{k,j} x_{j,n} - \nabla l_k^*(v_{k,n}) + d_{k,n})\big) + b_{k,n}\Big)$
    $v_{k,n+1} = v_{k,n} + \lambda_n \varepsilon_{p+k,n}(u_{k,n} - v_{k,n}),$
  for $j = 1, \ldots, p$
    $y_{j,n} = \varepsilon_{j,n}\Big(\mathrm{prox}^{\mathsf{W}_j^{-1}}_{f_j}\big(x_{j,n} - \mathsf{W}_j(\sum_{k\in\mathbb{L}_j^*} \mathsf{L}_{k,j}^*(2u_{k,n} - v_{k,n}) + \nabla h_j(x_{j,n}) + c_{j,n}\big) + a_{j,n}\Big)$
    $x_{j,n+1} = x_{j,n} + \lambda_n \varepsilon_{j,n}(y_{j,n} - x_{j,n})$

where

- $(\varepsilon_n)_{n\in\mathbb{N}}$ identically distributed $\mathbb{D}$-valued random variables with $\mathbb{D} = \{0,1\}^{p+q} \smallsetminus \{\mathbf{0}\}$
  $\rightsquigarrow$ binary variables signaling the blocks to be activated

Random block-coordinate primal-dual algorithm

for $n = 0, 1, \ldots$
  for $k = 1, \ldots, q$
    $u_{k,n} = \varepsilon_{p+k,n}\Big(\mathrm{prox}_{g_k^*}^{\mathsf{U}_k^{-1}}\big(v_{k,n} + \mathsf{U}_k(\sum_{j \in \mathbb{L}_k} \mathsf{L}_{k,j} x_{j,n} - \nabla l_k^*(v_{k,n}) + d_{k,n})\big) + b_{k,n}\Big)$
    $v_{k,n+1} = v_{k,n} + \lambda_n \varepsilon_{p+k,n}(u_{k,n} - v_{k,n}),$
  for $j = 1, \ldots, p$
    $y_{j,n} = \varepsilon_{j,n}\Big(\mathrm{prox}_{f_j}^{\mathsf{W}_j^{-1}}\big(x_{j,n} - \mathsf{W}_j(\sum_{k \in \mathbb{L}_j^*} \mathsf{L}_{k,j}^*(2u_{k,n} - v_{k,n}) + \nabla h_j(x_{j,n}) + c_{j,n}\big) + a_{j,n}\Big)$
    $x_{j,n+1} = x_{j,n} + \lambda_n \varepsilon_{j,n}(y_{j,n} - x_{j,n})$

where

- $(\varepsilon_n)_{n\in\mathbb{N}} \rightsquigarrow$ binary variables signaling the blocks to be activated
- x_0, $(a_n)_{n\in\mathbb{N}}$, and $(c_n)_{n\in\mathbb{N}}$ H-valued random variables, v_0, $(b_n)_{n\in\mathbb{N}}$, and $(d_n)_{n\in\mathbb{N}}$ G-valued random variables with $\mathsf{G} = \mathsf{G}_1 \times \cdots \times \mathsf{G}_q$
  $\rightsquigarrow (a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$, and $(d_n)_{n\in\mathbb{N}}$: error terms

Random block-coordinate primal-dual algorithm

for $n = 0, 1, \ldots$
  for $k = 1, \ldots, q$
    $u_{k,n} = \varepsilon_{p+k,n}\Big(\text{prox}_{g_k^*}^{\mathsf{U}_k^{-1}}\big(v_{k,n} + \mathsf{U}_k(\sum_{j\in\mathbb{L}_k} \mathsf{L}_{k,j} x_{j,n} - \nabla l_k^*(v_{k,n}) + d_{k,n})\big) + b_{k,n}\Big)$
    $v_{k,n+1} = v_{k,n} + \lambda_n \varepsilon_{p+k,n}(u_{k,n} - v_{k,n}),$
  for $j = 1, \ldots, p$
    $y_{j,n} = \varepsilon_{j,n}\Big(\text{prox}_{f_j}^{\mathsf{W}_j^{-1}}\big(x_{j,n} - \mathsf{W}_j(\sum_{k\in\mathbb{L}_j^*} \mathsf{L}_{k,j}^*(2u_{k,n} - v_{k,n}) + \nabla h_j(x_{j,n}) + c_{j,n}\big) + a_{j,n}\Big)$
    $x_{j,n+1} = x_{j,n} + \lambda_n \varepsilon_{j,n}(y_{j,n} - x_{j,n})$

where

- $(\varepsilon_n)_{n\in\mathbb{N}} \rightsquigarrow$ binary variables signaling the blocks to be activated
- $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$, and $(d_n)_{n\in\mathbb{N}}$: error terms
- $(\forall j \in \{1, \ldots, p\})\ \mathsf{W}_j : \mathsf{H}_j \to \mathsf{H}_j$ and $(\forall k \in \{1, \ldots, q\})\ \mathsf{U}_k : \mathsf{G}_k \to \mathsf{G}_k$ strongly positive self-adjoint preconditioning linear operators such that

$$1 - \Big(\sum_{j=1}^{p}\sum_{k=1}^{q} \|\mathsf{U}_k^{1/2}\mathsf{L}_{k,j}\mathsf{W}_j^{1/2}\|^2\Big)^{1/2} > \frac{1}{2}\max\{(\|\mathsf{W}_j\|\mu_j)_{1\leqslant j\leqslant p}, (\|\mathsf{U}_k\|\nu_k)_{1\leqslant k\leqslant q}\}.$$

Random block-coordinate primal-dual algorithm

for $n = 0, 1, \ldots$
  for $k = 1, \ldots, q$

$$u_{k,n} = \varepsilon_{p+k,n}\Big(\mathrm{prox}^{\mathsf{U}_k^{-1}}_{g_k^*}\big(v_{k,n} + \mathsf{U}_k\big(\sum_{j\in\mathbb{L}_k} \mathsf{L}_{k,j} x_{j,n} - \nabla l_k^*(v_{k,n}) + d_{k,n}\big)\big) + b_{k,n}\Big)$$

$$v_{k,n+1} = v_{k,n} + \lambda_n \varepsilon_{p+k,n}(u_{k,n} - v_{k,n}),$$

  for $j = 1, \ldots, p$

$$y_{j,n} = \varepsilon_{j,n}\Big(\mathrm{prox}^{\mathsf{W}_j^{-1}}_{f_j}\big(x_{j,n} - \mathsf{W}_j\big(\sum_{k\in\mathbb{L}_j^*} \mathsf{L}_{k,j}^*(2u_{k,n} - v_{k,n}) + \nabla h_j(x_{j,n}) + c_{j,n}\big) + a_{j,n}\Big)$$

$$x_{j,n+1} = x_{j,n} + \lambda_n \varepsilon_{j,n}(y_{j,n} - x_{j,n})$$

where

- $(\varepsilon_n)_{n\in\mathbb{N}} \rightsquigarrow$ binary variables signaling the blocks to be activated
- $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$, and $(d_n)_{n\in\mathbb{N}}$: error terms
- $(\forall j \in \{1,\ldots,p\})$ $\mathsf{W}_j : \mathsf{H}_j \to \mathsf{H}_j$ and $(\forall k \in \{1,\ldots,q\})$ $\mathsf{U}_k : \mathsf{G}_k \to \mathsf{G}_k$ strongly positive self-adjoint preconditioning linear operators
- $(\forall n \in \mathbb{N})$ $\lambda_n \in]0,1]$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$.

Random block-coordinate primal-dual algorithm

Theorem

Let $(\Omega, \mathcal{F}, \mathsf{P})$ *be the underlying probability space.*
Set $(\forall n \in \mathbb{N})\ \mathfrak{X}_n = \sigma(x_{n'}, v_{n'})_{0 \leqslant n' \leqslant n}$. *Assume that*

1. $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|a_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|b_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$, *and* $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|d_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$ P-a.s.
2. *The variables* $(\varepsilon_n)_{n\in\mathbb{N}}$ *are identically distributed such that* $(\forall j \in \{1,\ldots,p\})\ \mathsf{P}[\varepsilon_{j,0} = 1] > 0$.
3. *For every* $n \in \mathbb{N}$, ε_n *and* $\mathfrak{X}_n$ *are independent.*
4. *For every* $k \in \{1,\ldots,q\}$ *and* $n \in \mathbb{N}$,
$$\bigcup_{j\in\mathbb{L}_k} \big\{\omega \in \Omega \,\big|\, \varepsilon_{j,n}(\omega) = 1\big\} \subset \big\{\omega \in \Omega \,\big|\, \varepsilon_{p+k,n}(\omega) = 1\big\}.$$

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

Proof: based on properties of quasi-Fejér stochastic sequences [Combettes and Pesquet, 2014].

Random block-coordinate primal-dual algorithm

Theorem

Let $(\Omega, \mathcal{F}, \mathsf{P})$ *be the underlying probability space.*
Set $(\forall n \in \mathbb{N})$ $\mathfrak{X}_n = \sigma(x_{n'}, v_{n'})_{0 \leqslant n' \leqslant n}$. *Assume that*

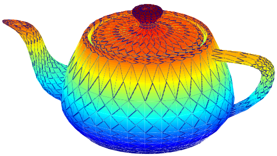
1. $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|a_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|b_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$, *and* $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|d_n\|^2 \,|\, \mathfrak{X}_n)} < +\infty$ P-a.s.
2. *The variables* $(\varepsilon_n)_{n\in\mathbb{N}}$ *are identically distributed such that* $(\forall j \in \{1, \ldots, p\})$ $\mathsf{P}[\varepsilon_{j,0} = 1] > 0$.
3. *For every* $n \in \mathbb{N}$, ε_n *and* $\mathfrak{X}_n$ *are independent.*
4. *For every* $k \in \{1, \ldots, q\}$ *and* $n \in \mathbb{N}$,

$$\varepsilon_{p+k,n} = \max_{1\leqslant j\leqslant p} \left\{ \varepsilon_{j,n} \;\middle|\; j \in \mathbb{L}_k \right\}.$$

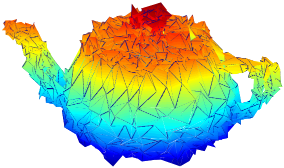
Then, $(x_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to an* F*-valued random variable, and* $(v_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to an* F^**-valued random variable.*

Proof: based on properties of quasi-Fejér stochastic sequences [Combettes and Pesquet, 2014].

Mesh denoising problem



Initial mesh $\overline{x}$



Observed mesh z

Undirected nonreflexive graph

OBJECTIVE:
Estimate $\overline{x} = (\overline{x}^{(i)})_{1 \leqslant i \leqslant M}$ from noisy observations $z = (z^{(i)})_{1 \leqslant i \leqslant M}$ where, for every $i \in \{1, \ldots, M\}$, $\overline{x}^{(i)} \in \mathbb{R}^3$ is the vector of 3D coordinates of the i-th vertex of a mesh

$\rightsquigarrow$ $\mathsf{H} = (\mathbb{R}^3)^M$

Mesh denoising problem

Objective:

Estimate $\overline{x} = (\overline{x}^{(i)})_{1 \leqslant i \leqslant M}$ from noisy observations $z = (z^{(i)})_{1 \leqslant i \leqslant M}$ where, for every $i \in \{1, \ldots, M\}$, $\overline{x}^{(i)} \in \mathbb{R}^3$ is the vector of 3D coordinates of the i-th vertex of a mesh

Cost function:

$$\Phi(x) = \sum_{j=1}^{M} \psi_j(x^{(j)} - z^{(j)}) + \iota_{C_j}(x^{(j)}) + \eta_j \|(x^{(j)} - x^{(i)})_{i \in \mathcal{N}_j}\|_{1,2}$$

where, for every $j \in \{1, \ldots, M\}$,

- $\psi_j : \mathbb{R}^3 \to \mathbb{R}$: $\ell_2 - \ell_1$ Huber function
  - robust data fidelity measure
- C_j: nonempty convex subset of $\mathbb{R}^3$
  - box constraint
- $\mathcal{N}_j$: neighborhood of j-th vertex
- $(\eta_j)_{1 \leqslant j \leqslant M}$: nonnegative regularization constants.

Mesh denoising problem

OBJECTIVE:

Estimate $\overline{x} = (\overline{x}^{(i)})_{1\leqslant i\leqslant M}$ from noisy observations $z = (z^{(i)})_{1\leqslant i\leqslant M}$ where, for every $i \in \{1,\dots,M\}$, $\overline{x}^{(i)} \in \mathbb{R}^3$ is the vector of 3D coordinates of the i-th vertex of a mesh

COST FUNCTION:

$$\Phi(x) = \sum_{j=1}^{M} \psi_j(x^{(j)} - z^{(j)}) + \iota_{C_j}(x^{(j)}) + \eta_j \|(x^{(j)} - x^{(i)})_{i\in\mathcal{N}_j}\|_{1,2}$$

IMPLEMENTATION DETAILS:

a block $\equiv$ a vertex $\Rightarrow$ $p = M$

Algorithm type I:

$(\forall j \in \{1,\dots,M\})$

- $h_j = \psi_j(\cdot - z^{(j)})$
- $f_j = \iota_{C_j}$

$q = M$

$(\forall k \in \{1,\dots,M\})(\forall x \in \mathsf{H})$

- $g_k(\mathsf{L}_k x) = \|(x^{(k)} - x^{(i)})_{i\in\mathcal{N}_k}\|_{1,2}$
- $l_k = \iota_{\{0\}}$

Algorithm type II:

$(\forall j \in \{1,\dots,M\})$

- $h_j = \psi_j(\cdot - z^{(j)})$

$q = 2M$

$(\forall k \in \{1,\dots,M\})(\forall x \in \mathsf{H})$

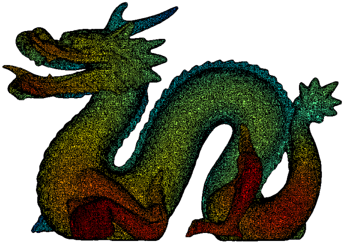
- $g_k(\mathsf{L}_k x) = \|(x^{(k)} - x^{(i)})_{i\in\mathcal{N}_k}\|_{1,2}$
- $g_{M+k}(\mathsf{L}_{M+k} x) = \iota_{C_k}(x^{(k)})$
- $l_k = \iota_{\{0\}}$

Simulation results (algorithm type II)

- positions of the original mesh are corrupted through an i.i.d zero-mean Gaussian mixture noise model.
- a limited number r of variables can be handled at each iteration, where

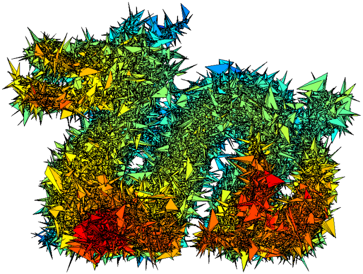
$$\sum_{j=1}^{p} \varepsilon_{j,n} = r \leqslant p.$$

- mesh decomposed into p/r non-overlapping sets.



Original mesh,

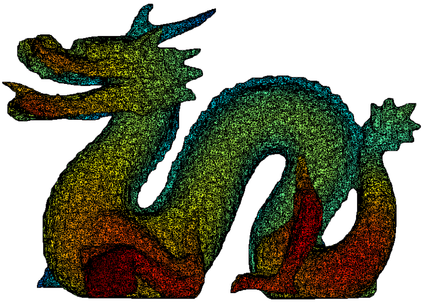
$M = 100250.$



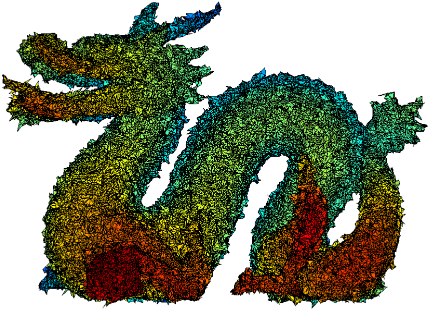
Noisy mesh,

MSE = 2.89×10^{-6}.

Simulation results (algorithm type II)

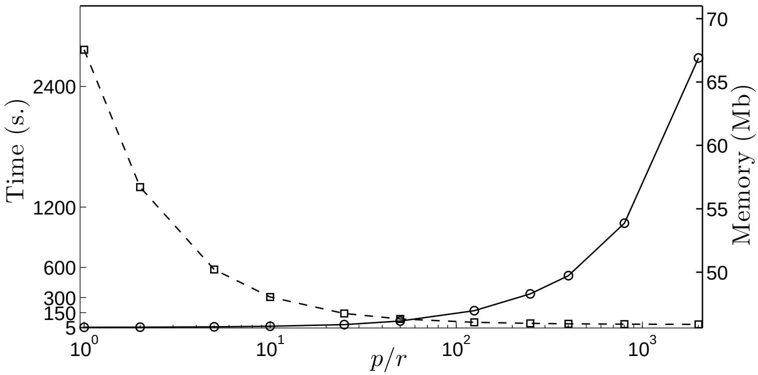


Proposed reconstruction,
MSE = 8.09×10^{-8}.



Laplacian smoothing,
MSE = 5.23×10^{-7}.

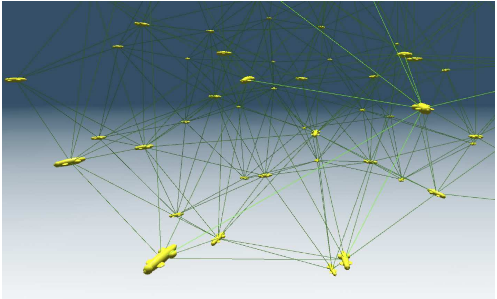
Complexity (algorithm type II)



- dashed line: required memory
- continuous line: reconstruction time

Available extensions

Asynchronous distributed algorithms
(stochastic, primal-dual, proximal, defined on a hypergraph)
[Pesquet and Repetti - 2014][Bianchi *et al.* - 2014]



Some references

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.

C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot
Dual constrained TV-based regularization on graphs
SIAM Journal on Imaging Sciences, vol. 6, no 3, pp. 1246-1273, 2013.

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
IEEE International Conference on Image Processing (ICIP 2014), 4141 - 4145, Paris, France, Oct. 27-30, 2014.

P. Combettes and J.-C Pesquet
Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping
to appear in *SIAM Journal on Optimization*, 2015.

N. Komodakis and J.-C. Pesquet
Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems
to appear in *IEEE Signal Processing Magazine*, 2015.

J.-C. Pesquet and A. Repetti
A class of randomized primal-dual algorithms for distributed optimization
to appear in *Journal of Nonlinear and Convex Analysis*, 2015.

A. Repetti, E. Chouzenoux and J.-C. Pesquet
A random block-coordinate primal-dual proximal algorithm with application to 3D mesh denoising
IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015). 5 p., Brisbane, Australia, April 19-24, 2015.