A RANDOM BLOCK-COORDINATE PRIMAL-DUAL PROXIMAL ALGORITHM WITH APPLICATION TO 3D MESH DENOISING

Audrey Repetti, Emilie Chouzenoux, and Jean-Christophe Pesquet

Université Paris-Est, Lab. d'Informatique Gaspard Monge, UMR CNRS 8049,
Champs-sur-Marne, 77454 Marne-la-Vallée, France
{first.last}@univ-paris-est.fr

ABSTRACT

Primal-dual proximal optimization methods have recently gained much interest for dealing with very large-scale data sets encoutered in many application fields such as machine learning, computer vision and inverse problems [1–3]. In this work, we propose a novel random block-coordinate version of such algorithms allowing us to solve a wide array of convex variational problems. One of the main advantages of the proposed algorithm is its ability to solve composite problems involving large-size matrices without requiring any inversion. In addition, the almost sure convergence to an optimal solution to the problem is guaranteed. We illustrate the good performance of our method on a mesh denoising application.

Index Terms— convex optimization, nonsmooth optimization, primal-dual algorithm, stochastic algorithm, block-coordinate algorithm, proximity operator, mesh processing, denoising, inverse problems.

1. INTRODUCTION

A quite general formulation of optimization problems arising in many application areas such as machine learning, computer vision or inverse problems [1–3] is as follows:

$$\underset{\mathsf{x}_1\in\mathsf{H}_1,\ldots,\mathsf{x}_p\in\mathsf{H}_p}{\text{minimize}} \quad \sum_{j=1}^{p}\big(\mathsf{f}_j(\mathsf{x}_j)+\mathsf{h}_j(\mathsf{x}_j)\big) + \sum_{k=1}^{q}(\mathsf{g}_k\,\square\,\mathsf{l}_k)\Big(\sum_{j=1}^{p}\mathsf{L}_{k,j}\mathsf{x}_j\Big) \tag{1}$$

where

- for every $j\in\{1,\ldots,p\}$, H_j is a separable real Hilbert space, f_j and h_j are proper lower-semicontinuous convex functions from H_j to $]-\infty,+\infty]$, h_j being assumed to be Lipschitz differentiable,[1]
- for every $k\in\{1,\ldots,q\}$, G_k is a separable real Hilbert space, g_k and l_k are proper lower-semicontinuous convex functions from G_k to $]-\infty,+\infty]$, l_k being assumed to be strongly convex,
- $\mathsf{L}_{k,j}$ is a linear bounded operator from H_j to G_k.

In (1), $\square$ denotes the inf-convolution operation defined by $\mathsf{g}_k\,\square\,\mathsf{l}_k\colon \mathsf{y}_k\mapsto \inf_{\mathsf{z}_k\in\mathsf{G}_k}\big(\mathsf{g}_k(\mathsf{z}_k)+\mathsf{l}_k(\mathsf{y}_k-\mathsf{z}_k)\big)$. In particular, if l_k is the indicator function of $\{0\}$ then $\mathsf{g}_k\,\square\,\mathsf{l}_k$ reduces to g_k. This special case in our framework probably corresponds to the most useful scenarios encountered in practice (e.g. total variation prior). Note that $(\mathsf{f}_j)_{1\leqslant j\leqslant p}$ and $(\mathsf{g}_k)_{1\leqslant k\leqslant q}$ can be chosen equal to the indicator functions of some closed convex sets so as to incorporate constraints on the solution to the optimization problem. See [4] for further details.

In the case when a single block of variables ($p=1$) is considered, Problem (1) can be efficiently solved using primal-dual splitting methods (see [5] and references therein) by performing, at each iteration, parallel computations of proximity operators and gradient steps. The main advantage of such algorithms relies on their ability to solve the primal problem (1) and its dual formulation with no need to invert linear operators. This ability turns out to be of primary importance in terms of computational complexity in the context of some large-scale problems (see e.g. [1,6–11]). Block-coordinate versions of various proximal algorithms have been recently proposed, such as the forward-backward algorithm [12–19], the Alternating Direction Method of Mutipliers [20, 21], and the Douglas-Rachford algorithm [12], where some of the block indices in $\{1,\ldots,p\}$ are activated at each iteration either in a random or deterministic manner. Although these methods have demonstrated a good performance, they are often limited either by the fact that they do not allow to fully split Problem (1) (in the sense that each function can be handled individually), or they require the inversion of some linear operators.

In this work we propose to combine the primal-dual strategy from [22–25] with the stochastic optimization tools recently developed in [12], giving rise to a new block-coordinate proximal primal-dual algorithm to solve Problem (1). Our algorithm makes it possible to activate randomly a subset of the $2(p+q)$ functions involved in the criterion. Such a strategy may be useful when p and q are large. In the line of [12, 15], the almost sure (a.s.) convergence to a solution to Problem (1) is proved even when some stochastic errors arise in the computation of the operations associated with the involved functions.

The remainder of this paper is organized as follows. First, we introduce our notation and better formulate the problem in Section 2. Then we describe in Section 3 the proposed random block-coordinate primal-dual proximal splitting algorithm and investigate its convergence properties. An application to 3D mesh denoising is provided in Section 4 to illustrate the good performance of our method. Finally, some conclusions are drawn in Section 5.

THIS WORK WAS SUPPORTED BY THE AGENCE NATIONALE DE LA RECHERCHE UNDER GRANT ANR GRAPHSIP.

[1] A differentiable function $\mathsf{f}\colon\mathsf{H}\to\mathbb{R}$ is said to be μ-Lipschitz differentiable if its gradient $\nabla\mathsf{f}$ is such that $(\forall(\mathsf{x},\mathsf{x}')\in\mathsf{H}^2)\ \|\nabla\mathsf{f}(\mathsf{x})-\nabla\mathsf{f}(\mathsf{x}')\|\leqslant\mu\|\mathsf{x}-\mathsf{x}'\|$.

2. PROBLEM FORMULATION

2.1. Optimization background

Let H be a real Hilbert space and let $\Gamma_0(\mathsf{H})$ denote the set of lower-semicontinuous convex functions from H to $]-\infty,+\infty]$ which are proper (i.e. with a nonempty domain). The proximity operator of a function $\mathsf{f} \in \Gamma_0(\mathsf{H})$ relative to the metric induced by some strongly positive bounded linear operator $\mathsf{U}\colon \mathsf{H} \to \mathsf{H}$ is defined as

$$\operatorname{prox}_{\mathsf{f}}^{\mathsf{U}}\colon \mathsf{H} \to \mathsf{H}\colon \mathsf{x} \to \operatorname*{argmin}_{\mathsf{y}\in\mathsf{H}} \ \mathsf{f}(\mathsf{y}) + \frac{1}{2}\langle \mathsf{x}-\mathsf{y} \mid \mathsf{U}(\mathsf{x}-\mathsf{y})\rangle,$$

where $\langle\cdot \mid \cdot\rangle$ is the scalar product endowing H.

The Fenchel-Legendre conjugate function $\mathsf{f}^* \in \Gamma_0(\mathsf{H})$ of f is defined by

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \mathsf{f}^*(\mathsf{x}) = \sup_{\mathsf{y}\in\mathsf{H}} \left(\langle \mathsf{y} \mid \mathsf{x}\rangle - \mathsf{f}(\mathsf{y})\right). \tag{2}$$

The Moreau decomposition theorem [26] ensures that $\operatorname{prox}_{\mathsf{f}^*}^{\mathsf{U}}(\cdot) = \mathsf{I} - \mathsf{U}^{-1}\operatorname{prox}_{\mathsf{f}}^{\mathsf{U}^{-1}}(\mathsf{U}\cdot)$.

2.2. Minimization problem

In this paper, we focus on problems of the form (1) satisfying the following technical assumption:

Assumption 2.1. *Problem* (1) *has at least one solution and one of the following two statements holds:*

- $(\forall j \in \{1,\dots,p\})$ f_j *is real-valued and, for every* $k \in \{1,\dots,q\}$, $(\mathsf{x}_j)_{1\leqslant j\leqslant p} \mapsto \sum_{j=1}^{p} \mathsf{L}_{k,j}\mathsf{x}_j$ *is surjective.*
- $(\forall k \in \{1,\dots,q\})$ *either* g_k *or* l_k *is real-valued.*

Moreover, the operators $(\mathsf{L}_{k,j})_{1\leqslant j\leqslant p,\, 1\leqslant k\leqslant q}$ *are non-trivial in the sense that, for every* $k \in \{1,\dots,q\}$ *and for every* $j \in \{1,\dots,p\}$,

$$\mathbb{L}_k = \big\{j' \in \{1,\dots,p\} \;\big|\; \mathsf{L}_{k,j'} \neq 0\big\} \neq \varnothing, \tag{3}$$

$$\mathbb{L}_j^* = \big\{k' \in \{1,\dots,q\} \;\big|\; \mathsf{L}_{k',j} \neq 0\big\} \neq \varnothing. \tag{4}$$

The dual problem of Problem (1) is then given by

$$\begin{aligned}\operatorname*{minimize}_{\mathsf{v}_1\in\mathsf{G}_1,\dots,\mathsf{v}_q\in\mathsf{G}_q} \ & \sum_{j=1}^{p}(\mathsf{f}_j^* \,\square\, \mathsf{h}_j^*)\Big(-\sum_{k=1}^{q}\mathsf{L}_{k,j}^*\mathsf{v}_k\Big) \\ & + \sum_{k=1}^{q}\big(\mathsf{g}_k^*(\mathsf{v}_k) + \mathsf{l}_k^*(\mathsf{v}_k)\big).\end{aligned} \tag{5}$$

We denote by $\mathbf{F}$ (resp. $\mathbf{F}^*$) the set of solutions to Problem (1) (resp. (5)). Our objective now is to generate a couple of random variables $(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{v}})$ taking its values in $\mathbf{F}\times\mathbf{F}^*$.

3. BLOCK-COORDINATE PRIMAL-DUAL ALGORITHM

3.1. Proposed algorithm

We propose to solve Problem (1) by adopting a primal-dual proximal splitting approach. Our algorithm is characterized by two main features. First, in order to take advantage of the block structure, a random block alternating strategy over indices $j \in \{1,\dots,p\}$ and $k \in \{1,\dots,q\}$ is implemented. Second, in accordance with recent works which have demonstrated the advantage in terms of convergence speed of adapting the underlying space metrics [10, 27], preconditioning linear operators $(\mathsf{W}_j)_{1\leqslant j\leqslant p}$ and $(\mathsf{U}_k)_{1\leqslant k\leqslant q}$ are introduced. We assume that these operators satisfy the following assumption:

Assumption 3.1.

(i) $(\forall j \in \{1,\dots,p\})$ $\mathsf{W}_j\colon \mathsf{H}_j \to \mathsf{H}_j$ *is a strongly positive self-adjoint bounded linear operator such that* $\mathsf{h}_j \circ \mathsf{W}_j^{1/2}$ *has a* μ_j^{-1}*-Lipschitzian gradient with* $\mu_j \in]0,+\infty[$.

(ii) $(\forall k \in \{1,\dots,q\})$ $\mathsf{U}_k\colon \mathsf{G}_k \to \mathsf{G}_k$ *is a strongly positive self-adjoint bounded linear operator such that* $\mathsf{l}_k^* \circ \mathsf{U}_k^{1/2}$ *has a* ν_k^{-1}*-Lipschitzian gradient with* $\nu_k \in]0,+\infty[$.

In the following, $\mathbf{H} = \mathsf{H}_1 \oplus \cdots \oplus \mathsf{H}_p$ denotes the Hilbert direct sum of $(\mathsf{H}_j)_{1\leqslant j\leqslant p}$. A generic element of $\mathbf{H}$ is denoted by $\mathbf{x} = (\mathsf{x}_j)_{1\leqslant j\leqslant p}$ with $\mathsf{x}_j \in \mathsf{H}_j$. Similarly, we define $\mathbf{G} = \mathsf{G}_1 \oplus \cdots \oplus \mathsf{G}_q$ and $\mathbf{v} = (\mathsf{v}_k)_{1\leqslant k\leqslant q}$ where $\mathsf{v}_k \in \mathsf{G}_k$ denotes a generic element of $\mathbf{G}$. Moreover, $\mathbb{D} = \{0,1\}^p \smallsetminus \{\mathbf{0}\}$ is the set of nonzero binary words of length p. We will keep on using such kind of notation throughout the paper. The proposed Randomized Block Primal-Dual Proximal algorithm is given below.

Algorithm 1 Randomized Block Primal-Dual Algorithm.

Initialization: Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ such that $\inf_{n\in\mathbb{N}}\lambda_n > 0$.
Let $\boldsymbol{x}_0$, $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$, and $(\boldsymbol{c}_n)_{n\in\mathbb{N}}$ be $\mathbf{H}$-valued random variables, and let $\boldsymbol{v}_0$, $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$, and $(\boldsymbol{d}_n)_{n\in\mathbb{N}}$ be $\mathbf{G}$-valued random variables.
Iterations:
for $n = 0, 1, \dots$
  Select randomly a vector of binary variables $\boldsymbol{\varepsilon}_n = (\varepsilon_{j,n})_{1\leqslant j\leqslant p}$.
  for $k = 1,\dots,q$
    $\eta_{k,n} = \max_{1\leqslant j\leqslant p}\big\{\varepsilon_{j,n} \;\big|\; k \in \mathbb{L}_j^*\big\}$
    $u_{k,n} = \eta_{k,n}\Big(\operatorname{prox}_{\mathsf{g}_k^*}^{\mathsf{U}_k^{-1}}\big(v_{k,n} - \mathsf{U}_k(\nabla \mathsf{l}_k^*(v_{k,n}) + d_{k,n})\big) + \mathsf{U}_k\sum_{j\in\mathbb{L}_k}\mathsf{L}_{k,j}x_{j,n}\Big) + b_{k,n}\Big)$
    $v_{k,n+1} = v_{k,n} + \lambda_n\eta_{k,n}(u_{k,n} - v_{k,n})$
  for $j = 1,\dots,p$
    $y_{j,n} = \varepsilon_{j,n}\Big(\operatorname{prox}_{\mathsf{f}_j}^{\mathsf{W}_j^{-1}}\big(x_{j,n} - \mathsf{W}_j(\nabla \mathsf{h}_j(x_{j,n}) + c_{j,n})\big) - \mathsf{W}_j\sum_{k\in\mathbb{L}_j^*}\mathsf{L}_{k,j}^*(2u_{k,n} - v_{k,n})\big) + a_{j,n}\Big)$
    $x_{j,n+1} = x_{j,n} + \lambda_n\varepsilon_{j,n}(y_{j,n} - x_{j,n})$.

The random Boolean variables $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ must satisfy the following assumption:

Assumption 3.2. $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ *are identically distributed* $\mathbb{D}$*-valued random variables and, for every* $n \in \mathbb{N}$, $\boldsymbol{\varepsilon}_n$ *and* $\boldsymbol{\mathcal{X}}_n = (\boldsymbol{x}_{n'}, \boldsymbol{v}_{n'})_{0\leqslant n'\leqslant n}$ *are independent. In addition,* $(\forall j \in \{1,\dots,p\})$ $\mathsf{P}[\varepsilon_{j,0} = 1] > 0$.

At each iteration n of Algorithm 1, $(\varepsilon_{j,n})_{1\leqslant j\leqslant p}$ signals the primal variables $(x_{j,n})_{1\leqslant j\leqslant p}$ that are activated. From a computational viewpoint, this means that if, for a given $j \in \{1,\dots,p\}$, $\varepsilon_{j,n} = 0$, the corresponding variable $x_{j,n}$ does not need to be updated. Moreover, for every $k \in \{1,\dots,q\}$, $\eta_{k,n}$ indicates the dual variables $u_{k,n}$ and $v_{k,n}$ that are activated. These dual variables need to be modified only when there exists $j \in \{1,\dots,p\}$ such that $\varepsilon_{j,n}$ and $\mathsf{L}_{k,j}$ are nonzero.

For every $j \in \{1,\dots,p\}$, and $k \in \{1,\dots,q\}$, the random variables $(a_{j,n})_{n\in\mathbb{N}}$ and $(b_{k,n})_{n\in\mathbb{N}}$ model stochastic errors possibly arising when applying $\operatorname{prox}_{\mathsf{f}_j}^{\mathsf{W}_j^{-1}}$ and $\operatorname{prox}_{\mathsf{g}_k^*}^{\mathsf{U}_k^{-1}}$, while $(c_{j,n})_{n\in\mathbb{N}}$ and $(d_{k,n})_{n\in\mathbb{N}}$ model some possible errors when computing the gradients of h_j and l_k^*. These random variables are required to fulfill the following assumption:

Assumption 3.3. *We have almost surely*

$$\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{a}_n\|^2\,|\,\mathcal{X}_n)} < +\infty, \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{b}_n\|^2\,|\,\mathcal{X}_n)} < +\infty,$$
$$\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{c}_n\|^2\,|\,\mathcal{X}_n)} < +\infty, \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{d}_n\|^2\,|\,\mathcal{X}_n)} < +\infty,$$

where for every $n \in \mathbb{N}$, $\mathcal{X}_n$ is defined as in Assumption 3.2.

It can be noticed that, when $p = 1$, Algorithm 1 allows us to generalize existing deterministic primal-dual algorithms. In particular, when $\mathsf{P}[\varepsilon_{1,0} = 1] = 1$, $\mathsf{W}_1 = \tau\mathsf{I}$ and, for every $k \in \{1,\dots,q\}$, $\mathsf{U}_k = \rho_k\mathsf{I}$, with $(\tau, \rho_1, \dots \rho_q) \in]0, +\infty[^{q+1}$, the algorithm from [23] is recovered, while the algorithm from [22] is obtained if it is further assumed that, for every $k \in \{1,\dots,q\}$, $\mathsf{l}_k = \iota_{\{0\}}$ where $\iota_{\{0\}}$ denotes the indicator function of $\{0\}$.

3.2. Convergence result

The following theorem, deduced from [15], guarantees that the sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ asymptotically provides a solution to the primal problem, while the sequence $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ asymptotically yields a solution to the dual one.

Theorem 3.1. *Let $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ be generated by Algorithm 1. Suppose that Assumptions 2.1-3.3 are satisfied. Moreover, suppose that*

$$2\left(1 - \|\mathbf{W}^{1/2}\mathbf{L}\mathbf{U}^{1/2}\|\right)\min\{(\mu_j)_{1\leqslant j\leqslant p}, (\nu_k)_{1\leqslant k\leqslant q}\} > 1, \quad (6)$$

with $\mathbf{L}\colon (\mathsf{x}_j)_{1\leqslant j\leqslant p} \mapsto \left(\sum_{j=1}^p \mathsf{L}_{k,j}\mathsf{x}_j\right)_{1\leqslant k\leqslant q}$, $\mathbf{U}\colon (\mathsf{v}_k)_{1\leqslant k\leqslant q} \mapsto (\mathsf{U}_1\mathsf{v}_1, \dots, \mathsf{U}_q\mathsf{v}_q)$, *and* $\mathbf{W}\colon (\mathsf{x}_j)_{1\leqslant j\leqslant p} \mapsto (\mathsf{W}_1\mathsf{x}_1, \dots, \mathsf{W}_p\mathsf{x}_p)$. *Then, $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly* a.s. *to an $\mathbf{F}$-valued random variable, and $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ converges weakly* a.s. *to an $\mathbf{F}^*$-valued random variable.*

Let us remark that, in the case when, for every $k \in \{1,\dots,q\}$, $\mathsf{l}_k = \iota_{\{0\}}$, the following condition, which is less restrictive than (6), is only required:

$$2\min\{(\mu_j)_{1\leqslant j\leqslant p}\}\left(1 - \|\mathbf{W}^{1/2}\mathbf{L}\mathbf{U}^{1/2}\|^2\right) > 1. \quad (7)$$

4. APPLICATION TO 3D MESH DENOISING

4.1. Problem statement

Let us consider an original 3D mesh described by its p nodes with spatial coordinates $\overline{\mathbf{x}} = (\overline{\mathsf{x}}_j^X, \overline{\mathsf{x}}_j^Y, \overline{\mathsf{x}}_j^Z)_{1\leqslant j\leqslant p} \in \mathbb{R}^{p\times 3}$ and its adjacency matrix $\mathsf{A} \in \mathbb{R}^{p\times p}$. We are interested in the problem of recovering an estimate of the original mesh from an observation of it, with similar topology (i.e. adjacency matrix A), but whose node positions, known in an imprecise manner, are given by

$$\mathbf{z} = (\mathbf{z}^X, \mathbf{z}^Y, \mathbf{z}^Z) = \overline{\mathbf{x}} + \mathbf{b}, \quad (8)$$

where $\mathbf{b} = (\mathsf{b}_j)_{1\leqslant j\leqslant p} \in \mathbb{R}^{p\times 3}$ models random uncertainties on the spatial positions of the mesh nodes. A method to produce an estimate $\widehat{\mathbf{x}} \in \mathbb{R}^{p\times 3}$ of $\overline{\mathbf{x}}$, is to define it as a solution to Problem (1), where functions $(\mathsf{f}_j)_{1\leqslant j\leqslant p}$, $(\mathsf{h}_j)_{1\leqslant j\leqslant p}$, and $(\mathsf{g}_k)_{1\leqslant k\leqslant q}$ incorporate some information regarding the observation model, and some a priori knowledge on the sought node positions [1].

Here, we will focus on the case when the intensity of the elements of $\mathbf{b}$ varies spatially. More precisely, two sets of nodes $\mathcal{N}_1$ and $\mathcal{N}_2$ are defined, with respective cardinality N_1 and N_2, such that $N_1 + N_2 = p$. We assume that the elements of $(\mathsf{b}_j)_{j\in\mathcal{N}_1}$ follow a white Gaussian distribution with zero mean and standard deviation σ_1, while the elements of $(\mathsf{b}_j)_{j\in\mathcal{N}_2}$ follow an i.i.d. Gaussian mixture distribution, with two mixture components parametrized by mixture weights $(\pi, 1-\pi)$ ($\pi \in [0,1]$), zero means and respective standard deviations σ_2 and σ_2'.

We define a data fidelity term allowing us to perform a robust estimation, which corresponds to the sum over $j \in \{1,\dots,p\}$ of functions:

$$(\forall \mathsf{x}_j = (\mathsf{x}_j^D)_{D\in\{X,Y,Z\}} \in \mathbb{R}^3) \quad \mathsf{h}_j(\mathsf{x}_j) = \sum_{D\in\{X,Y,Z\}} \Psi_j(\mathsf{x}_j^D - \mathsf{z}_j^D), \quad (9)$$

where, for every $j \in \{1,\dots,p\}$, Ψ_j is the Huber function parametrized by $\delta_j \in]0, +\infty]$, defined as

$$(\forall \xi \in \mathbb{R}) \quad \Psi_j(\xi) = \begin{cases} \frac{1}{2}\xi^2 & \text{if } |\xi| \leqslant \delta_j, \\ \delta_j|\xi| - \frac{1}{2}\delta_j^2 & \text{otherwise.} \end{cases} \quad (10)$$

Note that each function Ψ_j is 1-Lipschitz differentiable.

Moreover, in order to favor the smoothness of the restored mesh, we propose to use an isotropic total variation regularization function [28, 29] defined, for every $\mathbf{x} \in \mathbb{R}^{p\times 3}$, by

$$\sum_{k=1}^p \mathsf{g}_k(\mathsf{L}_k\mathbf{x}) = \sum_{k=1}^p \beta_k \sum_{D\in\{X,Y,Z\}} \|(\mathsf{x}_k^D - \mathsf{x}_i^D)_{i\in\mathcal{V}_k}\|_2, \quad (11)$$

where, for every $k \in \{1,\dots,p\}$, $\beta_k \in]0, +\infty[$ and $\mathcal{V}_k$ is the set of neighbors of node k, that is the set of indices $i \in \{1,\dots,p\}$ such that $\mathsf{A}_{ik} = 1$. Finally, we constrain the estimated nodes to belong to a box parametrized by its minimal and maximal sought spatial positions $(\mathsf{x}_{\min}^D, \mathsf{x}_{\max}^D)_{D\in\{X,Y,Z\}} \in \mathbb{R}^{3\times 2}$. To this end, we define

$$(\forall j \in \{1,\dots,p\}) \quad \mathsf{f}_j(\mathsf{x}_j) = \sum_{D\in\{X,Y,Z\}} \iota_{[\mathsf{x}_{\min}^D, \mathsf{x}_{\max}^D]}(\mathsf{x}_j^D).$$

Algorithm 1 requires the setting of matrices $(\mathsf{W}_j)_{1\leqslant j\leqslant p}$ and $(\mathsf{U}_k)_{1\leqslant k\leqslant p}$. Here, we simply take $\mathsf{W}_j \equiv \tau\mathsf{I}$ and $\mathsf{U}_k \equiv \rho\mathsf{I}$, where (τ, ρ) are positive constants such that Condition (7) is fulfilled. Since, for every $k \in \{1,\dots,p\}$, g_k corresponds to a separable sum of ℓ_2 norms, $\mathrm{prox}_{\mathsf{g}_k^*}^{\mathsf{U}_k^{-1}}$ has an explicit form [30]. Moreover the proximity operator of function f_j relative to the metric induced by W_j^{-1} corresponds to the projection onto $[\mathsf{x}_{\min}^X, \mathsf{x}_{\max}^X] \times [\mathsf{x}_{\min}^Y, \mathsf{x}_{\max}^Y] \times [\mathsf{x}_{\min}^Z, \mathsf{x}_{\max}^Z]$.

4.2. Numerical results

In our experiments, we use the standard `Bunny` 3D mesh, available at **http://graphics.stanford.edu/data/3Dscanrep/**, with $p = 8171$ nodes, displayed in Figure 1(top left). The noisy mesh $\mathbf{z}$, represented in Figure 1(top right), results from the observation model described in Section 4.1. We define $\mathcal{N}_2$, with cardinality $N_2 = 1327$, corresponding to the nodes located in the bunny ears and tail. The positions of these nodes are affected by outliers following a Gaussian mixture model with $(\sigma_2, \sigma_2') = (5\times 10^{-3}, 1.5\times 10^{-2})$ and $\pi = 0.98$, while a smaller noise standard deviation $\sigma_1 = 10^{-3}$ is used for the set $\mathcal{N}_1$. Figure 1(bottom left) displays the restored mesh obtained by applying Algorithm 1 with $\lambda_n \equiv 1$. The regularization parameters are set to $\beta_k = 5.5\times 10^{-4}$ for every $k \in \mathcal{N}_1$,

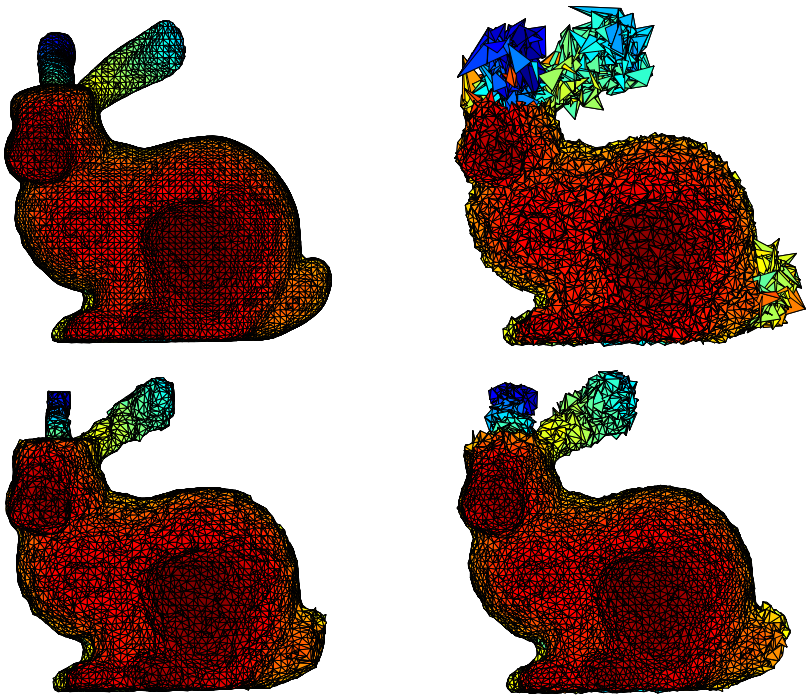

Fig. 1. Original mesh $\overline{\mathbf{x}}$ (top left), noisy mesh $\mathbf{z}$ with MSE = 5.45×10^{-6} (top right), and reconstructed mesh $\widehat{\mathbf{x}}$ using Algorithm 1 with MSE = 8.89×10^{-7} (bottom left) and using Laplacian smoothing with MSE = 1.29×10^{-6} (bottom right).

$\beta_k = 2.8 \times 10^{-3}$ otherwise, $\delta_j = +\infty$ for every $j \in \mathcal{N}_1$, $\delta_j = 10^{-2}$ otherwise, so as to minimize the Mean Square Error (MSE) of the restored mesh. For comparison, the result of Laplacian smoothing is given in Figure 1(bottom right).

Algorithm 1 allows a flexibility for the selection, at each iteration, of the updated nodes. To account for the spatial variability of the noise, we propose to choose, for every $n \in \mathbb{N}$ and $j \in \{1, \ldots, p\}$, $\varepsilon_{j,n}$ as a random variable following a Bernoulli distribution with probability p_j such that $\mathsf{p}_j = 1$ for $j \in \mathcal{N}_2$, and $\mathsf{p}_j = \mathsf{p} \in]0, 1]$ otherwise. Note that, for $\mathsf{p} = 1$, Algorithm 1 reduces to the deterministic primal-dual algorithm from [22]. We evaluate, for several values of probability p, the performance indicator:

$$\mathsf{C}(\mathsf{p}) = \overline{n}\, \frac{\mathsf{p} N_1 + N_2}{N_1 + N_2}, \tag{12}$$

where $\overline{n} \in \mathbb{N}^*$ is the first iteration of Algorithm 1 for which

$$\|\boldsymbol{x}_{\overline{n}} - \boldsymbol{x}_{\overline{n}-1}\| \leqslant 10^{-6} \sqrt{3p}.$$

For a given p, $\mathsf{C}(\mathsf{p})$ can be interpreted as the number of iterations necessary to reach the convergence, weighted by the mean number of updated nodes. Figure 2 illustrates the evolution of the indicator $\mathsf{C}(\mathsf{p})$, resulting from an averaging over ten runs of Algorithm 1 (i.e. ten different realizations of $(\varepsilon_n)_{n \in \mathbb{N}}$). One can observe that $\mathsf{C}(\mathsf{p})$ is minimized for $\mathsf{p} = 0.33$, which shows that the optimal strategy of Algorithm 1 in terms of convergence profile is obtained when only a fraction of the nodes in set $\mathcal{N}_1$, affected with the smallest noise, is updated at each iteration.

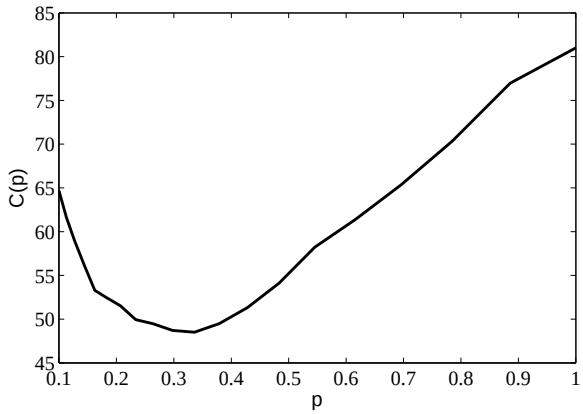

Fig. 2. Indicator $\mathsf{C}(\mathsf{p})$ for different values of probability p (average over ten runs of Algorithm 1).

5. CONCLUSION

In this paper, we have proposed an original approach for combining primal-dual proximal methods [22, 23] with randomized block-coordinate strategies [12]. We have shown that this gives rise to an efficient, provably convergent, method for solving large-scale convex optimization problems involving smooth and/or nonsmooth functions and linear operators, with no need to invert them. The good performance of the proposed method has been evaluated on a 3D mesh robust estimation problem. It can be expected that the proposed block-coordinate primal-dual algorithm will also be useful in other signal and image processing applications.

6. REFERENCES

[1] C. Couprie, L. Grady, L. Najman, J.-C. Pesquet, and H. Talbot, "Dual constrained TV-based regularization on graphs," *SIAM J. Imaging Sci.*, vol. 6, pp. 1246–1273, 2013.

[2] G. Chierchia, N. Pustelnik, B. Pesquet-Popescu, and J.-C. Pes-

quet, "A non-local structure tensor based approach for multicomponent image recovery problems," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5233–5248, Oct. 2014.

[3] I. Bayram and O. D. Akyildiz, "Primal-dual algorithms for audio decomposition using mixed norms," *Sig., Image and Video Proc.*, vol. 8, no. 1, pp. 95–110, Jan. 2014.

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

[5] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *Accepted for publication in IEEE Signal Process. Mag.*, 2014, http://www.optimization-online.org/DB_HTML/2014/06/4398.html.

[6] S. R. Becker and P. L. Combettes, "An algorithm for splitting parallel sums of linearly composed monotone operators, with applications to signal recovery," *J. Nonlinear Convex Anal.*, vol. 15, no. 1, pp. 137–159, Jan. 2014.

[7] S. Harizanov, J.-C. Pesquet, and G. Steidl, "Epigraphical projection for solving least squares Anscombe transformed constrained optimization problems," in *4th International Conference on Scale-Space and Variational Methods in Computer Vision*, A. Kuijper et al., Ed., Schloss Seggau, Leibnitz, Austria, 2-6 June 2013, vol. 7893 of *Lecture Notes in Computer Science*, pp. 125–136, Springer-Verlag, Berlin.

[8] A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot, "A primal-dual proximal splitting approach for restoring data corrupted with Poisson-Gaussian noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Kyoto, Japan, 25-30 Mar. 2012, pp. 1085–1088.

[9] N. Pustelnik, P. Borgnat, and P. Flandrin, "Empirical Mode Decomposition revisited by multicomponent nonsmooth convex optimization," *Signal Process.*, vol. 102, pp. 313–331, Sept. 2014.

[10] A. Repetti, E. Chouzenoux, and J.-C. Pesquet, "A penalized weighted least squares approach for restoring data corrupted with signal-dependent noise," in *Proc. Eur. Sig. and Image Proc. Conference*, Bucharest, Romania, 27-31 Aug. 2012, pp. 1553–1557.

[11] T. Teuber, G. Steidl, and R.-H. Chan, "Minimization and parameter estimation for seminorm regularization models with I-divergence constraints," *Inverse Problems*, vol. 29, pp. 035007, Mar. 2013.

[12] P. L. Combettes and J.-C. Pesquet, "Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping," 2014, http://www.optimization-online.org/DB_HTML/2014/04/4333.html.

[13] I. Necoara and A. Patrascu, "A random coordinate descent algorithm for optimization problems with composite objective function and linear coupled constraints," *Comput. Optim. Appl.*, vol. 57, no. 2, pp. 307–337, Mar. 2014.

[14] P. Richtárik and M. Takáč, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function," *Math. Program.*, vol. 144, no. 1–2, pp. 1–38, Apr. 2014.

[15] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," *Accepted for publication in J. Nonlinear Convex Anal.*, 2014, http://arxiv.org/abs/1406.6404.

[16] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Math. Program.*, vol. 1–2, pp. 459–494, Aug. 2014.

[17] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward-backward algorithm," 2014, http://www.optimization-online.org/DB_HTML/2013/12/4178.html.

[18] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[19] K. Slavakis and G. B. Giannakis, "Online dictionary learning from big data using accelerated stochastic approximation algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, 4-9 May 2014, pp. 16–20.

[20] 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.

[21] M. Hong, T.-H. Chang, X. Wang, M. Razaviyayn, S. Ma, and Z.-Q. Luo, "A block coordinate descent method of multipliers: Convergence analysis and applications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Florence, Italy, 4-9 May 2014, pp. 7739–7743.

[22] L. Condat, "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, no. 2, pp. 460–479, Aug. 2013.

[23] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, no. 3, pp. 667–681, Apr. 2013.

[24] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[25] 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, no. 4, pp. 1015–1046, 2010.

[26] P. L. Combettes and B. C. Vũ, "Variable metric forward-backward splitting with applications to monotone inclusions in duality," *Optimization*, , no. 9, pp. 1289–1318, Sept. 2014.

[27] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 6-13 2011, pp. 1762–1769.

[28] L. J. Grady and J. R. Polimeni, *Discrete Calculus: Applied Analysis on Graphs for Computational Science*, Springer-Verlag, New York, 2010.

[29] A. Elmoataz, O. Lezoray, and S. Bougleux, "Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing," *IEEE Trans. Image Process.*, vol. 17, pp. 1047–1060, 2008.

[30] 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. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2011.