A PARALLEL BLOCK-COORDINATE APPROACH FOR PRIMAL-DUAL SPLITTING WITH ARBITRARY RANDOM BLOCK SELECTION

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

The solution to many applied problems relies on finding the minimizer of a sum of smooth and/or nonsmooth convex functions possibly involving linear operators. In the last years, primal-dual methods have shown their efficiency to solve such minimization problems, their main advantage being their ability to deal with linear operators with no need to invert them. However, when the problem size becomes increasingly large, the implementation of these algorithms can be complicated, due to memory limitation issues. A simple way to overcome this difficulty consists of splitting the original numerous variables into blocks of reduced dimension, corresponding to the available memory, and to process separately each of them. In this paper we propose a random block-coordinate primal-dual algorithm, converging almost surely to a solution to the considered minimization problem. Moreover, an application to large-size 3D mesh denoising is provided to show the numerical efficiency of our method.

Index Terms— convex optimization, nonsmooth optimization, primal-dual algorithm, stochastic algorithm, parallel algorithm, random block-coordinate approach, proximity operator, mesh denoising.

1. INTRODUCTION

In the last years, the curse of dimensionality has become a prevailing concern in a number of application areas. One can mention large-scale graph processing applications, as well as 3D imaging, or machine learning using huge databases. Not only the number of data is becoming increasingly large, but there is also a need to process them with fast methods. Many of the problems of interest are actually related to optimization tasks involving cost functions which are often nonsmooth so as to leverage sparsity assumptions on the sought variables. In this context, two basic ideas can be exploited in order to lower the memory requirements and computational cost in high-dimensional optimization methods.

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

The first one consists of splitting the data into different blocks of reduced dimension which can be processed separately at a given time. By doing so, data lengths of the order of the blocksize can be kept in RAM. Early techniques for processing large-size images, which are actually still used by practitioners, partition the target image into non-overlapping blocks, each of them being dealt with in an independent manner. The drawback of these approaches is that they introduce artifacts around the borders of the blocks. These undesirable effects can be alleviated by considering overlapping blocks and/or by making use of a post-processing smoothing step, but the optimality of the resulting solution is no longer guaranteed. Note that block-coordinate strategies can also lead to an improvement in terms of convergence speed as they may, for instance, yield better adapted values of the algorithm parameters (e.g. larger stepsize values in the case of a gradient algorithm) [1, 2].

The second idea for efficiently solving large-scale optimization problems is to make use of massively parallel computing architectures. Due to the current technological limitations in terms of CPU speed, designing parallel algorithms allowing to benefit from multicore systems or GPGPU boards often is of paramount importance to obtain fast implementations. For this reason, in the last decade, a great deal of effort has been devoted to proximal splitting algorithms where the computation of the proximity operators and the gradient steps can be performed in a parallel manner [3–6]. These algorithms can be viewed as extensions of the parallel algorithms developed earlier for solving convex admissibility problems [7, 8]. Although they are able to solve more general convex optimization problems, they are less flexible in the sense that they require the proximity or gradient operators to be all activated at each iteration, whereas parallel projection-based methods make it possible to activate only a subset of the required projections. If the number of available processors is less than the number of operations to be performed in parallel, most of the existing parallel proximal algorithms become then of limited interest.

In this work, we overcome the latter limitation by proposing a block-coordinate algorithm applicable to a wide array of convex optimization problems. The blocks are activated

randomly, and the computations can be performed in a parallel manner by adjusting the computation load to the number of available processors. Our approach is grounded on recent developments concerning primal-dual convex optimization methods [9–13] as well as on novel convergence results on stochastic fixed point algorithms [14].

This paper is organized as follows. Section 2 gives the formulation of the considered optimization problem and introduces our notation. Section 3 describes the proposed random block-coordinate primal-dual proximal splitting algorithm, discusses its merits with respect to existing methods, and states its convergence properties. Section 4 illustrates the good performance of our method on a 3D mesh denoising problem. Finally, some conclusions are drawn in Section 5.

2. PROBLEM FORMULATION

In this paper, we consider a quite general form of convex optimization problems which can be expressed as

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

where for every $j\in\{1,\ldots,p\}$, H_j is a separable real Hilbert space, $\mathsf{h}_j\colon\mathsf{H}_j\to\,]-\infty,+\infty]$ is a convex differentiable function with a Lipschitzian gradient, for every $k\in\{1,\ldots,q\}$, G_k is a separable real Hilbert space, $\mathsf{g}_k\in\Gamma_0(\mathsf{G}_k)$,[1] $\mathsf{l}_k\in\Gamma_0(\mathsf{G}_k)$ is assumed to be strongly convex, and $\mathsf{L}_{k,j}$ is a linear bounded operator from H_j to G_k satisfying

$$\mathbb{L}_k=\big\{j'\in\{1,\ldots,p\}\;\big|\;\mathsf{L}_{k,j'}\neq 0\big\}\neq\varnothing, \tag{2}$$

$$\mathbb{L}_j^*=\big\{k'\in\{1,\ldots,q\}\;\big|\;\mathsf{L}_{k',j}\neq 0\big\}\neq\varnothing. \tag{3}$$

We recall that the *inf-convolution of two functions* $\mathsf{g}\colon\mathsf{H}\to\,]-\infty,+\infty]$ and $\mathsf{l}\colon\mathsf{H}\to\,]-\infty,+\infty]$ is defined as $\mathsf{g}\,\square\,\mathsf{l}\colon\mathsf{y}\mapsto\inf_{\mathsf{z}\in\mathsf{H}}\big(\mathsf{g}(\mathsf{z})+\mathsf{l}(\mathsf{y}-\mathsf{z})\big)$. The identity element of the inf-convolution is the indicator function of $\{0\}$, denoted by $\iota_{\{0\}}$. Then, if $\mathsf{l}=\iota_{\{0\}}$, $\mathsf{g}\,\square\,\mathsf{l}$ reduces to g.

In the remaining of the paper, it is assumed that Problem (1) has at least one solution and that one of the following two technical conditions holds:

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

The dual problem associated to Problem (1) is then given by

$$\underset{\mathsf{v}_1\in\mathsf{G}_1,\ldots,\mathsf{v}_q\in\mathsf{G}_q}{\text{minimize}} \quad \sum_{j=1}^{p}\mathsf{h}_j^*\Big(-\sum_{k=1}^{q}\mathsf{L}_{k,j}^*\mathsf{v}_k\Big)+\sum_{k=1}^{q}\big(\mathsf{g}_k^*(\mathsf{v}_k)+\mathsf{l}_k^*(\mathsf{v}_k)\big). \tag{4}$$

[1] $\Gamma_0(\mathsf{G}_k)$ denotes the set of proper lower-semicontinuous convex functions from G_k to $]-\infty,+\infty]$.

This dual problem involves the *Fenchel-Legendre conjugate functions* of $(\mathsf{h}_j)_{1\leqslant j\leqslant p}$, $(\mathsf{g}_k)_{1\leqslant k\leqslant q}$ and $(\mathsf{l}_k)_{1\leqslant k\leqslant q}$, here denoted by a star symbol. We recall that the conjugate f^* of a proper lower-semicontinuous convex function f is defined as $(\forall\mathsf{x}\in\mathsf{H})\ \mathsf{f}^*(\mathsf{x})=\sup_{\mathsf{y}\in\mathsf{H}}\big(\langle\mathsf{y}\mid\mathsf{x}\rangle-\mathsf{f}(\mathsf{y})\big)$.

Subsequently, we denote by $\mathbf{F}$ (resp. $\mathbf{F}^*$) the set of solutions to Problem (1) (resp. (4)).

3. BLOCK-COORDINATE PRIMAL-DUAL ALGORITHM

3.1. Proposed algorithm

We propose a primal-dual proximal splitting approach for solving Problem (1). Our method is based on a randomized block-coordinate strategy, described in Algorithm 1. For conciseness, in the remainder of this work, $\mathbf{H}=\mathsf{H}_1\oplus\cdots\oplus\mathsf{H}_p$ 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, for every $j\in\{1,\ldots,p\}$, $\mathsf{x}_j\in\mathsf{H}_j$. Similarly, we define $\mathbf{G}=\mathsf{G}_1\oplus\cdots\oplus\mathsf{G}_q$ with generic element $\mathbf{v}=(\mathsf{v}_k)_{1\leqslant k\leqslant q}$ where, for every $k\in\{1,\ldots,q\}$ $\mathsf{v}_k\in\mathsf{G}_k$. Moreover, $\mathbb{D}=\{0,1\}^{p+q}\smallsetminus\{\mathbf{0}\}$ is the set of nonzero binary vectors of length $p+q$. Finally, the *proximity operator* of a proper lower-semicontinuous function f defined on a Hilbert space H, relative to the metric induced by some strongly positive self-adjoint bounded linear operator $\mathsf{U}\colon\mathsf{H}\to\mathsf{H}$, is defined as

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

where $\langle\cdot\mid\cdot\rangle$ is the scalar product endowing H. We refer the reader to [15] for further mathematical details.

We now discuss the main advantages offered by the proposed algorithm:

(i) In order to accelerate the convergence speed of the method, we modify the underlying space metrics in a manner similar to [16, 17], by introducing preconditioning linear operators $(\mathsf{W}_j)_{1\leqslant j\leqslant p}$ and $(\mathsf{U}_k)_{1\leqslant k\leqslant q}$. These operators are assumed to be bounded, self-adjoint and strongly positive.

(ii) A random block alternating strategy over indices $j\in\{1,\ldots,p\}$ and $k\in\{1,\ldots,q\}$ is adopted in order to take advantage of the block structure of the problem. More precisely, at each iteration n of Algorithm 1, $(\varepsilon_{j,n})_{1\leqslant j\leqslant p}$ (resp. $(\varepsilon_{p+k,n})_{1\leqslant k\leqslant q}$) are Boolean variables signaling the primal variables $(x_{j,n})_{1\leqslant j\leqslant p}$ (resp. dual variables $(v_{k,n})_{1\leqslant k\leqslant q}$) that are activated. In addition, at iteration n, for every $j\in\{1,\ldots,p\}$, the Boolean $\eta_{j,n}$ is chosen such that $s_{j,n}$ and $y_{j,n}$ are updated only if there exists some $k\in\{1,\ldots,q\}$ such that both $\varepsilon_{p+k,n}$ and $\mathsf{L}_{k,j}$ are nonzero.

(iii) For every $j\in\{1,\ldots,p\}$, and $k\in\{1,\ldots,q\}$, the random variables $(a_{j,n})_{n\in\mathbb{N}}$, $(b_{k,n})_{n\in\mathbb{N}}$ and $(c_{k,n})_{n\in\mathbb{N}}$ model

some possible errors arising in the computation of $\nabla \mathsf{h}_j$, $\mathrm{prox}_{\mathsf{g}_k^*}^{\mathsf{U}_k^{-1}}$ and $\nabla \mathsf{l}_k^*$, respectively. The algorithm is robust to such stochastic errors as soon as they fulfilled, almost surely (a.s.),

$$\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{a}_n\|^2 \,|\, \mathcal{X}_n)} < +\infty, \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{b}_n\|^2 \,|\, \mathcal{X}_n)} < +\infty,$$
$$\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{c}_n\|^2 \,|\, \mathcal{X}_n)} < +\infty, \text{with } \mathcal{X}_n = (\boldsymbol{x}_{n'}, \boldsymbol{v}_{n'})_{0\leqslant n'\leqslant n}.$$

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 $\boldsymbol{x}_0$, $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ be $\mathbf{H}$-valued random variables, and let $\boldsymbol{v}_0$, $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$, and $(\boldsymbol{c}_n)_{n\in\mathbb{N}}$ be $\mathbf{G}$-valued random variables.
Iterations:
for $n = 0, 1, \ldots$
  Select randomly a vector $\boldsymbol{\varepsilon}_n = (\varepsilon_{i,n})_{1\leqslant i\leqslant p+q} \in \mathbb{D}$
  for $j = 1, \ldots, p$
    $\eta_{j,n} = \max\big\{\varepsilon_{p+k,n} \;\big|\; k \in \mathbb{L}_j^*\big\}$
    $s_{j,n} = \eta_{j,n}\Big(x_{j,n} - \mathsf{W}_j\big(\nabla \mathsf{h}_j(x_{j,n}) + a_{j,n}\big)\Big)$
    $y_{j,n} = \eta_{j,n}\big(s_{j,n} - \mathsf{W}_j \sum_{k\in\mathbb{L}_j^*} \mathsf{L}_{k,j}^* v_{k,n}\big)$
  for $k = 1, \ldots, q$
    $u_{k,n} = \varepsilon_{p+k,n}\Big(\mathrm{prox}_{\mathsf{g}_k^*}^{\mathsf{U}_k^{-1}}\Big(v_{k,n} + \mathsf{U}_k \sum_{j\in\mathbb{L}_k} \mathsf{L}_{k,j} y_{j,n} - \mathsf{U}_k\big(\nabla \mathsf{l}_k^*(v_{k,n}) + c_{k,n}\big)\Big) + b_{k,n}\Big)$
    $v_{k,n+1} = v_{k,n} + \lambda_n \varepsilon_{p+k,n}(u_{k,n} - v_{k,n})$
  for $j = 1, \ldots, p$
    $p_{j,n} = \varepsilon_{j,n}\Big(s_{j,n} - \mathsf{W}_j \sum_{k\in\mathbb{L}_j^*} \mathsf{L}_{k,j}^* u_{k,n}\Big)$
    $x_{j,n+1} = x_{j,n} + \lambda_n \varepsilon_{j,n}(p_{j,n} - x_{j,n})$.

3.2. Link with existing works

It can be noticed that, when $p = 1$ and $(\forall n \in \mathbb{N})\ \mathsf{P}[\varepsilon_{1,n} = 1] = 1$, Algorithm 1 is related to a number of existing primal-dual algorithms :

(i) The deterministic algorithms proposed in [18, 19] are recovered when $q = 1$, $(\forall n \in \mathbb{N})\ \mathsf{P}[\varepsilon_{2,n} = 1] = 1$, H_1 and G_1 are finite dimensional spaces, $\mathsf{l}_1 = \iota_{\{0\}}$, $\mathsf{W}_1 = \tau\mathsf{I}$ and $\mathsf{U} = \rho\mathsf{I}$, with $(\tau, \rho) \in]0, +\infty[^2$, no relaxation ($\lambda_n \equiv 1$) or a constant one ($\lambda_n \equiv \lambda < 1$) is performed, and in the absence of errors.

(ii) The deterministic fully parallel algorithm proposed in [9, Prop. 4.3] is recovered when $(\forall n \in \mathbb{N})\ (\forall k \in \{1, \ldots, q\})$ $\mathsf{P}[\varepsilon_{p+k,n} = 1] = 1$, and the errors $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$, $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{c}_n)_{n\in\mathbb{N}}$ are deterministic and summable.

(iii) A different stochastic block primal-dual algorithm was recently proposed in [20] (see also [21] for some particular case) but this algorithm requires more restrictive convergence conditions than those provided in the next section.

3.3. Convergence result

The following theorem, deduced from [22], guarantees that the sequence $(\boldsymbol{x}_n, \boldsymbol{v}_n)_{n\in\mathbb{N}}$ asymptotically provides a pair of primal-dual solutions to the problems (1)-(4).

Theorem 3.1. *Suppose that the following assumptions hold:*

(i) *$(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ are identically distributed $\mathbb{D}$-valued random variables and, for every $n \in \mathbb{N}$, $\boldsymbol{\varepsilon}_n$ and $\mathcal{X}_n$ are independent. In addition, $(\forall j \in \{1, \ldots, p\})\ \mathsf{P}[\varepsilon_{j,0} = 1] > 0$, and $(\forall k \in \{1, \ldots, q\})\ \varepsilon_{p+k,n} = \max\big\{\varepsilon_{j,n} \;\big|\; j \in \mathbb{L}_k\big\}$.*

(ii) *For every $j \in \{1, \ldots, p\}$ and for every $k \in \{1, \ldots, q\}$, $\mathsf{h}_j \circ \mathsf{W}_j^{1/2}$ and $\mathsf{l}_k^* \circ \mathsf{U}_k^{1/2}$ have Lipschitzian gradients with positive constants μ_j^{-1} and ν_k^{-1}, respectively. Moreover,*

$$\min\Big\{\mu, \nu\Big(1 - \|\mathbf{U}^{1/2}\mathbf{L}\mathbf{W}^{1/2}\|^2\Big)\Big\} > 1/2, \qquad (5)$$

where $\mathbf{L}\colon (\mathsf{x}_j)_{1\leqslant j\leqslant p} \mapsto \big(\sum_{j=1}^p \mathsf{L}_{k,j}\mathsf{x}_j\big)_{1\leqslant k\leqslant q}$, $\mathbf{U}\colon (\mathsf{v}_k)_{1\leqslant k\leqslant q} \mapsto (\mathsf{U}_1\mathsf{v}_1, \ldots, \mathsf{U}_q\mathsf{v}_q)$, $\mathbf{W}\colon (\mathsf{x}_j)_{1\leqslant j\leqslant p} \mapsto (\mathsf{W}_1\mathsf{x}_1, \ldots, \mathsf{W}_p\mathsf{x}_p)$, $\mu = \min\{\mu_1, \ldots, \mu_p\}$, and $\nu = \min\{\nu_1, \ldots, \nu_q\}$.

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.*

Note that in the case when, for every $k \in \{1, \ldots, q\}$, $\mathsf{l}_k = \iota_{\{0\}}$, the condition (5) simplifies into

$$\|\mathbf{U}^{1/2}\mathbf{L}\mathbf{W}^{1/2}\| < 1 \quad \text{and} \quad \mu > 1/2. \qquad (6)$$

Finally, let us emphasize that the assumption we make on $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ is very weak, so that our method benefits from a very high flexibility on the way the blocks can be selected. Basically, almost any arbitrary random sampling strategy fits into our framework.

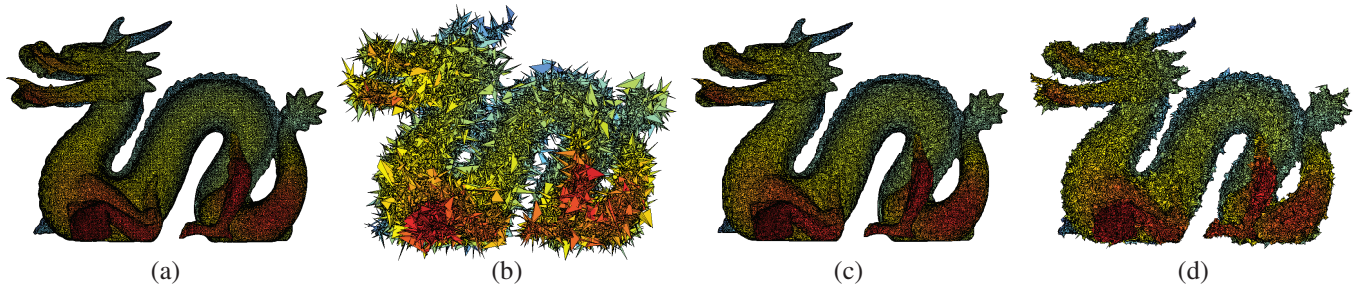
4. APPLICATION TO 3D MESH DENOISING

4.1. Problem statement

We focus on the inverse problem of restoring the spatial positions $\overline{\mathbf{x}} = (\overline{\mathbf{x}}_j^X, \overline{\mathbf{x}}_j^Y, \overline{\mathbf{x}}_j^Z)_{1\leqslant j\leqslant p} \in \mathbb{R}^{p\times 3}$ of the nodes of a large 3D mesh from measurements $\mathbf{z} = (\mathbf{z}^X, \mathbf{z}^Y, \mathbf{z}^Z)$ corrupted with an i.i.d. noise. The topology of the observed mesh $\mathbf{z}$ is assumed to be similar to the one of $\overline{\mathbf{x}}$ (i.e. both meshes have the same adjacency matrix).

We propose to define the estimate $\widehat{\mathbf{x}} \in \mathbb{R}^{p\times 3}$ of $\overline{\mathbf{x}}$ as a solution to Problem (1) [23], where:

- The sum over $j \in \{1, \ldots, p\}$ of functions $(\mathsf{h}_j)_{1\leqslant j\leqslant p}$ represents the data fidelity term incorporating some informations regarding the observation model. For every



(a) (b) (c) (d)

Fig. 1. (a) Original mesh $\overline{\mathsf{x}}$, (b) noisy mesh z with MSE = 2.89×10^{-6}, and reconstructed meshes $\widehat{\mathsf{x}}$ using (c) Algorithm 1 with MSE = 8.09×10^{-8} and (d) Laplacian smoothing with MSE = 5.23×10^{-7}.

$\mathsf{x}_j = (\mathsf{x}_j^D)_{D\in\{X,Y,Z\}} \in \mathsf{H}_j = \mathbb{R}^3$, we set $\mathsf{h}_j(\mathsf{x}_j) = \sum_{D\in\{X,Y,Z\}} \Psi(\mathsf{x}_j^D - \mathsf{z}_j^D)$, where Ψ is the Huber function with threshold $\delta > 0$, defined as

$$(\forall \xi \in \mathbb{R}) \quad \Psi(\xi) = \begin{cases} \frac{1}{2}\xi^2 & \text{if } |\xi| \leqslant \delta, \\ \delta|\xi| - \frac{1}{2}\delta^2 & \text{otherwise.} \end{cases} \tag{7}$$

Note that Ψ is 1-Lipschitz differentiable.

- The sum over $k \in \{1,\ldots,q\}$, with $q = 2p$, of functions $(\mathsf{g}_k)_{1\leqslant k\leqslant q}$ models a hybrid regularization term introducing some a priori knowledge on the sought node positions (we take $\mathsf{l}_k \equiv \iota_{\{0\}}$). First, following [24, 25], we promote the smoothness of the restored mesh by defining, for all $\mathsf{x} \in \mathbb{R}^{p\times 3}$, for all $k \in \{1,\ldots,p\}$, $\mathsf{g}_k(\sum_{j=1}^p \mathsf{L}_{k,j}\mathsf{x}_j) = \beta \sum_{D\in\{X,Y,Z\}} \|(\mathsf{x}_k^D - \mathsf{x}_i^D)_{i\in\mathcal{V}_k}\|_2$,with $\beta > 0$, and $\mathcal{V}_k$ the set of neighbors of node k. Moreover, we constrain the spatial positions of the estimated nodes to belong to the box with bounds $(\mathsf{x}_{\min}^D, \mathsf{x}_{\max}^D)_{D\in\{X,Y,Z\}} \in \mathbb{R}^{3\times 2}$ by choosing, for all $\mathsf{x} \in \mathbb{R}^{p\times 3}$, for all $k \in \{1,\ldots,p\}$ $\mathsf{g}_{p+k}(\sum_{j=1}^p \mathsf{L}_{k,j}\mathsf{x}_j) = \sum_{D\in\{X,Y,Z\}} \iota_{[\mathsf{x}_{\min}^D, \mathsf{x}_{\max}^D]}(\mathsf{x}_j^D)$.

4.2. Implementation

In our simulations, we simply take $(\forall j \in \{1,\ldots,p\})$ $\mathsf{W}_j \equiv \tau\mathsf{I}$, $\mathsf{U}_j \equiv \rho_1\mathsf{I}$ and $\mathsf{U}_{p+j} \equiv \rho_2\mathsf{I}$, where (τ, ρ_1, ρ_2) are positive constants satisfying Condition (6).

At each iteration $n \in \mathbb{N}$, for every $k \in \{1,\ldots,q\}$, we need to compute the proximity operator of g_k^* for updating variable $u_{k,n}$. Since, for every $j \in \{1,\ldots,p\}$, g_j identifies with a separable sum of ℓ_2 norms, $\mathrm{prox}_{\mathsf{g}_j}^{\mathsf{U}_j}$ has an explicit form [4], from which the expression of $\mathrm{prox}_{\mathsf{g}_k^*}^{\mathsf{U}_k^{-1}}$ can be deduced by applying the Moreau decomposition theorem [26]. Moreover the proximity operators of functions g_{p+j} relative to the metric induced by U_{p+j} correspond to the projection onto $[\mathsf{x}_{\min}^X, \mathsf{x}_{\max}^X] \times [\mathsf{x}_{\min}^Y, \mathsf{x}_{\max}^Y] \times [\mathsf{x}_{\min}^Z, \mathsf{x}_{\max}^Z]$.

Note that the inner loops over $j \in \{1,\ldots,p\}$ and $k \in \{1,\ldots,q\}$ in Algorithm 1 can be computed in a parallel manner.

Finally, when one deals with a very large amount of data, memory requirement becomes a more limiting factor than the computational time. This implies that only a limited number of variables can be handled at each iteration of the algorithm. At each iteration $n \in \mathbb{N}$, the activation variables $(\varepsilon_{j,n})_{1\leqslant j\leqslant p}$ must thus be chosen so as to satisfy some budget constraint, which can be expressed as $\sum_{j=1}^p \varepsilon_{j,n} = r$, with $r \leqslant p$. To this aim, we propose to divide the considered mesh into p/r non-overlapping blocks, and to set $(\varepsilon_{j,n})_{1\leqslant j\leqslant p, n\in\mathbb{N}}$ so as to activate a single block chosen randomly at every iteration.

4.3. Numerical results

To show the performance of the proposed method, we consider a large 3D mesh, namely the standard `Dragon` available at **http://graphics.stanford.edu/data/3Dscanrep/**, with $p = 100250$ nodes and 301207 edges, displayed in Figure 1(a). The positions of the original mesh are corrupted through a Gaussian mixture noise model. The resulting observed mesh z is displayed in Figure 1(b). The restored mesh obtained using Algorithm 1, with $\lambda_n \equiv 1$, is represented in Figure 1(c). We choose the regularization parameters (β, δ) so as to minimize the Mean Square Error (MSE). For comparison, Figure 1(d) displays the estimated mesh obtained using Laplacian smoothing [27].

Figure 2 illustrates the time necessary to reach the stopping criterion $\|\boldsymbol{x}_n - \widehat{\boldsymbol{x}}\| \leqslant 10^{-3}\|\widehat{\boldsymbol{x}}\|$, where $\widehat{\boldsymbol{x}}$ has been precomputed using a large number of iterations, as a function of the number of blocks p/r. The memory space required for running the algorithm in Matlab R2013a is also reported. As one could expect, the best reconstruction time (equals to 5 s.) is obtained for $r = p$, i.e. with a deterministic strategy. The choice of smaller blocks leads to a decrease of the occupied memory, at the expense of a decrease of the convergence speed. A major advantage of our method is that it allows the user to select the random sampling strategy leading to a minimal reconstruction time, without exceeding the available memory on its own computer architecture.

5. CONCLUSION

In this paper, we proposed a random block-coordinate version of a primal-dual algorithm for solving minimization problems involving convex (possibly non differentiable) functions and

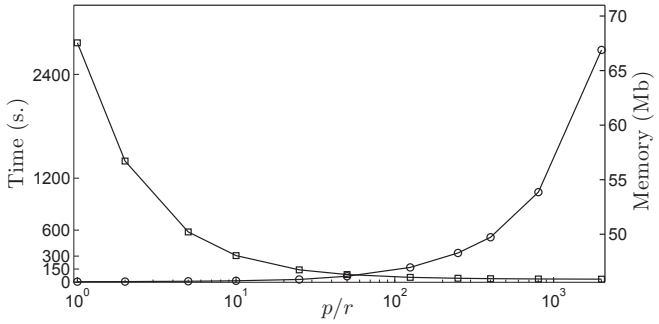


Fig. 2. Reconstruction time (circles) and required memory (squares) for several block numbers p/r.

linear operators, whose convergence is established. Our approach generalizes existing primal-dual methods, by allowing to update randomly, at each iteration, only a subset of the components of the iterates. This is particularly useful for the resolution of large scale problems with limited allocated memory. The good performance of our method has been illustrated through an application on a large 3D mesh denoising problem.

REFERENCES

[1] 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.

[2] 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.

[3] J.-C. Pesquet and N. Pustelnik, "A parallel inertial proximal optimization method," *Pac. J. Optim.*, vol. 8, no. 2, pp. 273–305, Apr. 2012.

[4] 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.

[5] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Restoration of Poissonian images using alternating direction optimization," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3133–3145, Dec. 2010.

[6] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Visual Communication and Image Representation*, vol. 21, no. 3, pp. 193–199, Apr. 2010.

[7] P. L. Combettes, "The foundations of set theoretic estimation," in *Proc. of the IEEE*, Feb. 1993, vol. 81, pp. 182–208.

[8] P. L. Combettes, "A block-iterative surrogate constraint splitting method for quadratic signal recovery," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 596–636, June 2003.

[9] 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," in *Proc. IEEE Int. Conf. Image Process.*, Paris, France, 27-30 Oct. 2014, pp. 4141–4145.

[10] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, vol. 20, no. 2, pp. 307–330, June 2012.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

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

[16] 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.

[17] 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.

[18] P. Chen, J. Huang, and X. Zhang, "A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration," *Inverse Problems*, vol. 29, no. 2, pp. 025011, 2013.

[19] I. Loris and C. Verhoeven, "On a generalization of the iterative soft-thresholding algorithm for the case of non-separable penalty," *Inverse Problems*, vol. 27, no. 12, pp. 125007, 2011.

[20] A. Repetti, E. Chouzenoux, and J.-C. Pesquet, "A random block-coordinate primal-dual proximal algorithm with application to 3D mesh denoising," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brisbane, Australia, 21-25 Apr. 2015.

[21] 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.

[22] 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.

[23] 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.

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

[25] 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.

[26] P. L. Combettes and B. C. Vũ, "Variable metric forward-backward splitting with applications to monotone inclusions in duality," *Optimization*, vol. 63, no. 9, pp. 1289–1318, Sept. 2014.

[27] G. Taubin, T. Zhang, and G. Golub, "Optimal surface smoothing as filter design," in *Computer Vision ECCV '96*, Bernard Buxton and Roberto Cipolla, Eds., vol. 1064 of *Lecture Notes in Computer Science*, pp. 283–292. Springer Berlin Heidelberg, 1996.