TWO DISTRIBUTED ALGORITHMS FOR THE DECONVOLUTION OF LARGE RADIO-INTERFEROMETRIC MULTISPECTRAL IMAGES

Céline Meillier, Pascal Bianchi, Walid Hachem

LTCI, CNRS, Telecom Paristech, Université Paris-Saclay, 75013, Paris, France

ABSTRACT

We address in this paper the deconvolution issue for radio-interferometric multispectral images. Whereas this problem has been widely explored in the recent literature for single images, a few algorithms are able to reconstruct multispectral images (three-dimensional images) [1], [2]. We propose in this paper two new distributed algorithms based on the optimization methods ADMM and projected gradient (PG) for the reconstruction of radio-interferometric multispectral images. We present an original distributed architecture and a comparison of their performance on a quasi-real data cube.

***Index Terms*—** ADMM, deconvolution, distributed optimization, projected gradient, radio-interferometry, multispectral images.

1. INTRODUCTION

With the advent of new generations of radio interferometers such as the Low Frequency Array (LOFAR) and the Square Kilometer Array (SKA), a large amount of multispectral images will be produced in the next few years. These new interferometers have a very large field of view (millions of pixels) with a high spectral resolution (hundreds of frequency bands in the radio wave domain). These massive observed data are corrupted by the noise and by the instrument response. A challenging point is the design of deconvolution algorithms that are able to deal with the large size of the observations. A good deal of the recent research is focused on distributed optimization algorithms aiming to solve the deconvolution issue in two cases:

- Large two-dimensional monochromatic (*i.e.*, with only one spectral band) image deconvolution with different approaches: let us cite the so-called PURIFY algorithm [3] based on the Simultaneous Direction Method of Multipliers (SDMM), [4] where the authors propose in particular two new scalable splitting algorithms for image reconstruction, or [5] where a compressive sensing approach is proposed.
- The extension to the multispectral (3D) images bearing two spatial dimensions and a spectral one. The third dimension obviously increases the size of the deconvolution problem. The authors of [2] propose a method based on the Alternating Direction Method of Multipliers (ADMM) that is known amongst other things for its ability to be distributed [6].

This work was partly supported by the Agence Nationale pour la Recherche, France, under the project ANR-14-CE23-0004-01 MAGELLAN.

This work starts with the regularized optimization problem described in [2] and recalled in Section 2. In Section 3, a simpler version of ADMM than in [2] is proposed by removing one Lagrange multiplier vector. Concurrently, we also propose to solve the optimization problem by resorting to the projected gradient method (PG) in the dual space (Section 4). Distributed implementations for both these algorithms are proposed in Section 5. This distributed architecture differs from the classical master/slave one by the fact that none of the nodes stores the entire multispectral image (in classical architecture, the global image is stored on the master node). Thanks to a minimal amount of data exchange between the nodes, the required memory and the computational cost supported by a single node are decreased. The performances of the two algorithms are finally compared in Section 6.

2. PROBLEM FORMULATION

Assuming that the observations are made on L frequency bands, the 3D data cube can be seen as a collection of L monochromatic images of N pixels each. Similar to [2], the image observed at the frequency $\nu_l \in \{\nu_1, \cdots, \nu_L\}$, also called the "dirty image" at ν_l, is given by the equation

$$\mathbf{y}_l = \mathbf{H}_l \mathbf{x}_l + \mathbf{n}_l \in \mathbb{R}^N \tag{1}$$

where $\mathbf{x}_l \in \mathbb{R}^N$ is the "true" image vector, $\mathbf{n}_l$ is the noise vector and $\mathbf{H}_l$ is a convolution matrix representing the so-called Point Spread Function (PSF) of the radio-interferometer. Stacking the L dirty images in the vector $\mathbf{y} = [\mathbf{y}_1^T, \cdots, \mathbf{y}_L^T]^T \in \mathbb{R}^M$ where $M = N \times L$, and denoting respectively as $\|\cdot\|_2$ and $\|\cdot\|_1$ the Euclidean and the ℓ_1 norms, the optimization problem is written

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\mu_\epsilon}{2}\|\mathbf{x}\|_2^2 + \imath_{\mathbb{R}^+}(\mathbf{x}) + \|\mathbf{W}\mathbf{x}\|_1, \tag{2}$$

where the first term is the objective function, the second is a Tikhonov regularization term controlled by the parameter

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

$\mu_\epsilon > 0$, the third is a positivity constraint (since we are recovering sky brightnesses) where $\imath_{\mathbb{R}^+}(\mathbf{x}) = \infty$ if one of the elements of $\mathbf{x}$ at least is negative, and 0 otherwise, and the last term is a sparsity term. Here, the sparsity is induced in the multiresolution (wavelet) domain for each monochromatic image, and in the Discrete Cosine Transform (DCT) domain at each multifrequency pixel. Specifically,

$$\mathbf{W} = \begin{pmatrix} \mu_s \tilde{\mathbf{W}}_s \\ --- \\ \mu_\nu \tilde{\mathbf{W}}_\nu \end{pmatrix}$$

where $\mu_s, \mu_\nu > 0$ are regularization parameters, and

$$\tilde{\mathbf{W}}_s = \begin{pmatrix} \mathbf{W}_s & & 0 \\ & \ddots & \\ 0 & & \mathbf{W}_s \end{pmatrix} \in \mathbb{R}^{m_s M \times M}$$

is a block-diagonal matrix where each block $\mathbf{W}_s \in \mathbb{R}^{m_s N \times N}$, acting on a monochromatic image, is the concatenation of m_s orthogonal wavelet bases. Similar to [2] and [3], we identify $\mathbf{W}_s$ with a dictionary consisting in the concatenation of the first eight Daubechies wavelet bases ($m_s = 8$). Finally,

$$\tilde{\mathbf{W}}_\nu = \begin{pmatrix} \mathbf{W}_\nu & & 0 \\ & \ddots & \\ 0 & & \mathbf{W}_\nu \end{pmatrix} \mathbb{P} \in \mathbb{R}^{M \times M}$$

where $\mathbb{P}$ is the permutation matrix that rearranges the elements of the vector $\mathbf{x}$ pixel by pixel, each of these pixels being represented by a vector of L frequencies, and where each block $\mathbf{W}_\nu$ is a $L \times L$ matrix representing the DCT. Note that $\tilde{\mathbf{W}}_s^T \tilde{\mathbf{W}}_s = m_s \mathbf{I}_M$ and $\tilde{\mathbf{W}}_\nu^T \tilde{\mathbf{W}}_\nu = \mathbf{I}_M$. It is moreover obvious that $\|\mathbf{W}\mathbf{x}\|_1 = \mu_s \|\tilde{\mathbf{W}}_s \mathbf{x}\|_1 + \mu_\nu \|\tilde{\mathbf{W}}_\nu \mathbf{x}\|_1$.

3. ADMM DESCRIPTION

In order to solve Problem (2), we reformulate it as follows:

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{z}) \quad \text{subject to: } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{0}$$

where

$$\begin{aligned} f(\mathbf{x}) &= (1/2)\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + (\mu_\epsilon/2)\|\mathbf{x}\|_2^2, \\ g(\mathbf{z}) &= \imath_{\mathbb{R}^+}(\mathbf{p}) + \mu_s\|\mathbf{t}\|_1 + \mu_\nu\|\mathbf{v}\|_1 \quad \text{where} \\ \mathbf{z}^T &= (\mathbf{p}^T, \mathbf{t}^T, \mathbf{v}^T) \in \mathbb{R}^M \times \mathbb{R}^{m_s M} \times \mathbb{R}^M, \\ \mathbf{A} &= \begin{pmatrix} \mathbf{I}_M \\ \tilde{\mathbf{W}}_s \\ \tilde{\mathbf{W}}_\nu \end{pmatrix}, \text{ and } \mathbf{B} = \begin{pmatrix} -\mathbf{I}_M & 0 & 0 \\ 0 & -\mathbf{I}_{m_s M} & 0 \\ 0 & 0 & -\mathbf{I}_M \end{pmatrix}. \end{aligned}$$

The associated augmented Lagrangian for $\rho > 0$ is

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\gamma}) &= f(\mathbf{x}) + g(\mathbf{z}) \\ &+ \boldsymbol{\gamma}_{\mathbf{p}}^T(\mathbf{x} - \mathbf{p}) + \boldsymbol{\gamma}_{\mathbf{t}}^T(\tilde{\mathbf{W}}_s\mathbf{x} - \mathbf{t}) + \boldsymbol{\gamma}_{\mathbf{v}}^T(\tilde{\mathbf{W}}_\nu\mathbf{x} - \mathbf{v}) \\ &+ \frac{\rho}{2}\|\mathbf{x} - \mathbf{p}\|_2^2 + \frac{\rho}{2}\|\tilde{\mathbf{W}}_s\mathbf{x} - \mathbf{t}\|_2^2 + \frac{\rho}{2}\|\tilde{\mathbf{W}}_\nu\mathbf{x} - \mathbf{v}\|_2^2 \quad (3) \end{aligned}$$

where $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_{\mathbf{p}}^T, \boldsymbol{\gamma}_{\mathbf{t}}^T, \boldsymbol{\gamma}_{\mathbf{v}}^T]^T \in \mathbb{R}^{(2+m_s)M}$ is the vector of Lagrange multipliers, decomposed in accordance with the right hand side of (3). Note that the dimension of this vector is smaller than in [2], where four sets of Lagrange mutlipliers were used instead of three here.

As it is well known, ADMM consists of the following iterations:

$$\mathbf{x}^{k+1} = \operatorname*{argmin}_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^k, \boldsymbol{\gamma}^k) \quad (4)$$

$$\begin{aligned} \mathbf{z}^{k+1} &= [{\mathbf{p}^{k+1}}^T, {\mathbf{t}^{k+1}}^T, {\mathbf{v}^{k+1}}^T]^T \\ &= \operatorname*{argmin}_{\mathbf{z}} \mathcal{L}_\rho(\mathbf{x}^{k+1}, \mathbf{z}, \boldsymbol{\gamma}^k) \quad (5) \end{aligned}$$

$$\begin{aligned} \boldsymbol{\gamma}^{k+1} &= [{\boldsymbol{\gamma}_{\mathbf{p}}^{k+1}}^T, {\boldsymbol{\gamma}_{\mathbf{t}}^{k+1}}^T, {\boldsymbol{\gamma}_{\mathbf{v}}^{k+1}}^T]^T \\ &= \boldsymbol{\gamma}^k + \rho(\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1}). \quad (6) \end{aligned}$$

We now write $\mathbf{x}^k = [{\mathbf{x}_1^k}^T, \ldots, {\mathbf{x}_L^k}^T]^T$ where each block is of size N and thus corresponds to a monochromatic image. We do the same decomposition for $\mathbf{p}^k$ and $\boldsymbol{\gamma}_{\mathbf{p}}^k$. Similar decompositions are also done for $\mathbf{t}^k$ and $\boldsymbol{\gamma}_{\mathbf{t}}^k$ where the dimensions of the blocks are this time equal to $m_s N$.

Let us consider the minimization (4). Solving the equation shows that the minimization w.r.t. $\mathbf{x}$ is separable with respect to the frequencies thanks to the block diagonal structure of $\tilde{\mathbf{W}}_s$ and to the orthogonality of $\tilde{\mathbf{W}}_\nu$. After a straightforward computation, we obtain that $\mathbf{x}_l^{k+1} = \mathbf{Q}_l^{-1}\mathbf{b}_l^k$ for each $l \in \{1, \ldots, L\}$ where

$$\begin{aligned} \mathbf{Q}_l &= \mathbf{H}_l^T\mathbf{H}_l + (\mu_\epsilon + (2 + m_s)\rho)\mathbf{I}_N \quad \text{and} \\ \mathbf{b}_l^k &= \mathbf{H}_l^T\mathbf{y}_l - \boldsymbol{\gamma}_{\mathbf{p},l}^k - \mathbf{W}_s^T\left(\boldsymbol{\gamma}_{\mathbf{t},l}^k - \rho\mathbf{t}_l^k\right) \\ &\quad + \rho\left(\mathbf{p}_l^k + \left(\mathbb{P}^T(\mathbf{v}^k)\right)_l\right) - \left(\mathbb{P}^T(\tilde{\mathbf{W}}_\nu^T\boldsymbol{\gamma}_{\mathbf{v}}^k)\right)_l \end{aligned} \quad . \quad (7)$$

Here $(\cdot)_l$ denotes the l^{th} size-N block of a vector. The computations leading to the expression of $\mathbf{Q}_l$ make use of the isometric nature of $\tilde{\mathbf{W}}_s$. Observe that the $\mathbf{Q}_l$ can be computed once at the beginning of the algorithm while the vectors $\mathbf{b}_l^k$ need to be computed at every iteration. Finally, since $\mathbf{H}_l$ is a convolution operator, the equation $\mathbf{Q}_l^{-1}\mathbf{b}_l^k$ can be practically approximated by using the Fast Fourier Transform operator.

Minimizations w.r.t. the vectors $\mathbf{p}$ and $\mathbf{t}$ in (5) are structurally separable with respect to the frequencies. Writing $\tilde{\mathbf{p}}_l^{k+1} = \rho^{-1}\boldsymbol{\gamma}_{\mathbf{p},l}^k + \mathbf{x}_l^{k+1}$, we get

$$\begin{aligned} \mathbf{p}_l^{k+1} &= \arg\min_{\mathbf{u} \in \mathbb{R}^M} \imath_{\mathbb{R}^+}(\mathbf{u}) + {\boldsymbol{\gamma}_{\mathbf{p},l}^k}^T(\mathbf{x}_l^{k+1} - \mathbf{u}) \\ &\quad + \frac{\rho}{2}\|\mathbf{x}_l^{k+1} - \mathbf{u}\|_2^2 \\ &= \max(0, \tilde{\mathbf{p}}_l^{k+1}) \end{aligned}$$

where $\max$ is taken elementwise.

Writing $\tilde{\mathbf{t}}_l^{k+1} = \mathbf{W}_s \mathbf{x}_l^{k+1} + \rho^{-1}\boldsymbol{\gamma}_{\mathbf{t},l}^k$, we also have

$$\begin{aligned}\mathbf{t}_l^{k+1} &= \arg\min_{\mathbf{u}\in\mathbb{R}^{m_s M}} \mu_s \|\mathbf{u}\|_1 + \frac{\rho}{2}\|\mathbf{u} - \tilde{\mathbf{t}}_l^{k+1}\|_2^2 \\ &= \tilde{\mathbf{t}}_l^{k+1} \bullet \max\Big(0, 1 - \frac{\rho^{-1}\mu_s}{|\tilde{\mathbf{t}}_l^{k+1}|}\Big)\end{aligned}$$

where $\bullet$ is the elementwise product. We recognize here the usual soft thresholding operator.

The update of the vector $\mathbf{v}^k$ is done at each multifrequency pixel. Write this time $\mathbf{v}^k = [\mathbf{v}_1^{k^T}, \dots, \mathbf{v}_N^{k^T}]^T$ and $\boldsymbol{\gamma}_{\mathbf{v}}^k = [\boldsymbol{\gamma}_{\mathbf{v},1}^{k^T}, \dots, \boldsymbol{\gamma}_{\mathbf{v},1}^{k^T}]^T$ where the blocks within these vectors have the size L. Let $\tilde{\mathbf{v}}_i^{k+1} = (\tilde{\mathbf{W}}_\nu \mathbf{x}^{k+1})_i + \rho^{-1}\boldsymbol{\gamma}_{\mathbf{v},i}^k$ where $(\cdot)_i$ is a size-L block, we get

$$\begin{aligned}\mathbf{v}_i^{k+1} &= \arg\min_{\mathbf{u}\in\mathbb{R}^{L}} \mu_\nu \|\mathbf{u}\|_1 + \frac{\rho}{2}\|\mathbf{u} - \tilde{\mathbf{v}}_i^{k+1}\|_2^2 \\ &= \tilde{\mathbf{v}}_i^{k+1} \bullet \max\Big(0, 1 - \frac{\rho^{-1}\mu_\nu}{|\tilde{\mathbf{v}}_i^{k+1}|}\Big).\end{aligned}$$

Finally, the inspection of Equation (6) shows that $\boldsymbol{\gamma}_{\mathbf{p}}^k$ and $\boldsymbol{\gamma}_{\mathbf{t}}^k$ are updated at the level of the monochormatic images while $\boldsymbol{\gamma}_{\mathbf{v}}^k$ is updated at the multifrequency pixel level.

4. PROJECTED GRADIENT ON THE DUAL PROBLEM

4.1. Primal and dual problems

In this section, we replace Problem (2) with the problem

$$\min_{\mathbf{x}\in\mathbb{R}^M} f(\mathbf{x}) + h(\mathbf{W}\mathbf{x}), \tag{8}$$

where we recall that $f(\mathbf{x}) = (1/2)\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + (\mu_\epsilon/2)\|\mathbf{x}\|_2^2$, and where we set $h : \mathbb{R}^{(m_s+1)M} \to \mathbb{R}_+,\ \mathbf{u} \mapsto \|\mathbf{u}\|_1$. Note that the positivity assumption is now absent. In order to solve this problem, we start by writing its dual problem

$$\min_{\boldsymbol{\lambda}\in\mathbb{R}^{(m_s+1)M}} f^*(-\mathbf{W}^T\boldsymbol{\lambda}) + h^*(\boldsymbol{\lambda}) \tag{9}$$

where f^* is the Legendre-Fenchel transform of f, defined as $f^*(\boldsymbol{\phi}) = \sup_{\mathbf{x}}\langle\boldsymbol{\phi}, \mathbf{x}\rangle - f(\mathbf{x})$. After a standard calculation, we get

$$f^*(\boldsymbol{\phi}) = \frac{1}{2}\boldsymbol{\phi}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\phi} + \boldsymbol{\phi}^T\boldsymbol{\Delta}^{-1}\mathbf{H}^T\mathbf{y} + \frac{1}{2}\mathbf{y}^T(\mathbf{H}\boldsymbol{\Delta}^{-1}\mathbf{H}^T - \mathbf{I})\mathbf{y}$$

where $\boldsymbol{\Delta} = (\mathbf{H}^T\mathbf{H} + \mu_\epsilon\mathbf{I})$ is a block diagonal matrix of size $M \times M$ with Toeplitz blocks, and

$$h^*(\boldsymbol{\lambda}) = \imath_{\mathcal{B}_\infty}(\boldsymbol{\lambda})$$

where $\mathcal{B}_\infty = \{\boldsymbol{\lambda} \in \mathbb{R}^{(m_s+1)M} : \|\boldsymbol{\lambda}\|_\infty \leqslant 1\}$ is the unit ball for the $\|\cdot\|_\infty$ norm.

The inspection of f and h shows that the qualification conditions for the duality gap to be zero hold. Moreover, the saddle point $(\mathbf{x}^\star, \boldsymbol{\lambda}^\star)$ satisfies $\mathbf{x}^\star = \nabla f^*(-\mathbf{W}^T\boldsymbol{\lambda}^\star)$ where ∇f^* is the gradient of f^*, given by

$$\nabla f^*(\boldsymbol{\phi}) = \boldsymbol{\Delta}^{-1}(\boldsymbol{\phi} + \mathbf{H}\mathbf{y}). \tag{10}$$

4.2. Solving the dual problem using PG

The dual problem (9) can be reformulated as:

$$\min_{\boldsymbol{\lambda}:\|\boldsymbol{\lambda}\|_\infty\leqslant 1} f^*(-\mathbf{W}^T\boldsymbol{\lambda}).$$

Since f^* is smooth, this problem can be solved with the help of PG (see e.g. [7]). In our context, this algorithm reads:

$$\begin{aligned}\boldsymbol{\lambda}^{k+1} &= \mathcal{P}_\infty\Big(\boldsymbol{\lambda}^k + \rho\mathbf{W}\nabla f^*(-\mathbf{W}^T\boldsymbol{\lambda}^k)\Big) \\ &= \mathcal{P}_\infty\Big(\boldsymbol{\lambda}^k - \rho\mathbf{W}\boldsymbol{\Delta}^{-1}(\mathbf{W}^T\boldsymbol{\lambda}^k - \mathbf{H}\mathbf{y})\Big)\end{aligned}$$

where $\rho > 0$ and where $\mathcal{P}_\infty$ is the projection operator on $\mathcal{B}_\infty$, being the proximity operator of h^*. At the last iteration k_{max}, the 3D cube $\mathbf{x}^{k_{max}}$ is recovered according to the equation $\mathbf{x}^{k_{max}} = \nabla f^*(-\mathbf{W}^T\boldsymbol{\lambda}^{k_{max}}) = \boldsymbol{\Delta}^{-1}(\mathbf{H}^T\mathbf{y} - \mathbf{W}^T\boldsymbol{\lambda}^{k_{max}})$.

In order to provide a distributed implementation of this algorithm, we write $\boldsymbol{\lambda}^k = [\boldsymbol{\lambda}_s^{k^T}, \boldsymbol{\lambda}_\nu^{k^T}]^T$ where $\boldsymbol{\lambda}_s^k \in \mathbb{R}^{m_s M}$ is processed at the level of the monochromatic images and $\boldsymbol{\lambda}_\nu^k \in \mathbb{R}^M$ is processed at the level of the multifrequency pixels. Details are provided in the next section.

5. DISTRIBUTED ARCHITECTURE

Implementation of deconvolution algorithms on a distributed architecture is needed for memory charge reasons; SKA multispectral data cubes are expected to be 80 tera bytes.

5.1. Structure of the cluster

From the 2D + 1D structure of the multispectral data and the spatial and spectral sparsity constraints of the minimization problems (2) and (8), the variables $\mathbf{x}, \mathbf{p}, \mathbf{t}, \boldsymbol{\gamma}_{\mathbf{p}}, \boldsymbol{\gamma}_{\mathbf{t}}, \boldsymbol{\lambda}_s$ (resp. $\mathbf{v}, \boldsymbol{\gamma}_{\mathbf{v}}, \boldsymbol{\lambda}_\nu$) can be evaluated only on monochromatic images (resp. on pixels). All of these calculations can be done with parallel programming w.r.t. the frequencies (resp. the pixels). We use a cluster of machines that we divide into two groups: one for the calculations w.r.t. the frequencies (group A), the other for the calculations w.r.t. the pixels (group B). Figure 1 illustrates the two groups of nodes architecture and exchanges between nodes. For the sake of simplicity, in figure 1, we assume there are as many nodes in group A (resp. in group B) as frequency bands (resp. pixels) in the multispectral image. Note that in real implementation, each node is in charge of several frequencies or pixels, depending on the capacity of the cluster and the image size.

5.2. Distributed implementation of ADMM and PG

Algorithm 1 summarizes the distribution of the alternated updating steps of the primal and dual variables described in section 3 for the ADMM algorithm. All the calculations are distributed on the two groups of nodes architecture introduced in the previous paragraph w.r.t. the frequencies (resp. w.r.t.

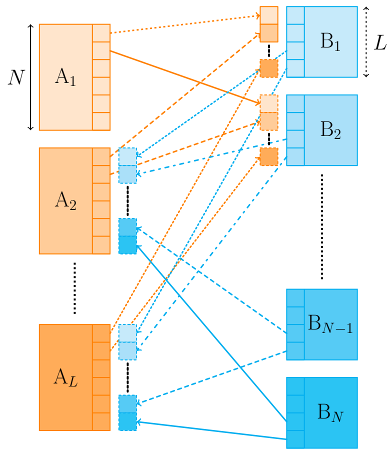

Fig. 1: Illustration of the distributed architecture. Nodes of group A (orange blocks) do parallel calculation w.r.t. the frequencies, while nodes of group B (blue blocks) do parallel calculation w.r.t. the pixels.

Algorithm 1 Distributed ADMM algorithm

Initialize $\mathbf{x}$, $\mathbf{p}$, $\mathbf{t}$, $\mathbf{v}$, $\gamma_{\mathbf{p}}$, $\gamma_{\mathbf{t}}$, $\gamma_{\mathbf{v}}$
while $\delta_{\mathbf{x}}^k \geqslant 10^{-5}$ **do**
  for nodes in group A **do**
    Evaluate $\mathbf{b}_l^{k+1}$
    Solve $\mathbf{Q}_l \mathbf{x}_l^{k+1} = \mathbf{b}_l^{k+1}$
    Send $\mathbf{x}^{k+1}$ to nodes of group B
  end for
  MPI synchronization barrier
  for nodes in group A **do**
    $\mathbf{p}_l^{k+1} \leftarrow \max(0, \tilde{\mathbf{p}}_l^{k+1})$
    $\mathbf{t}_l^{k+1} \leftarrow \tilde{\mathbf{t}}_l^{k+1} \bullet \max\left(0, 1 - \frac{\rho^{-1}\mu_s}{|\tilde{\mathbf{t}}_l^{k+1}|}\right)$
    $\gamma_{\mathbf{p},l}^{k+1} \leftarrow \gamma_{\mathbf{p},l}^{k} + \rho(\mathbf{x}_l^{k+1} - \mathbf{p}_l^{k+1})$
    $\gamma_{\mathbf{t},l}^{k+1} = \gamma_{\mathbf{t},l}^{k} + \rho(\mathbf{W}_s \mathbf{x}_l^{k+1} - \mathbf{t}_l^{k+1})$
  end for
  for nodes in group B **do**
    $\mathbf{v}_i^{k+1} \leftarrow \tilde{\mathbf{v}}_i^{k+1} \bullet \max\left(0, 1 - \frac{\rho^{-1}\mu_\nu}{|\tilde{\mathbf{v}}_i^{k+1}|}\right)$
    $\gamma_{\mathbf{v},i}^{k+1} \leftarrow \gamma_{\mathbf{v},i}^{k} + \rho((\tilde{\mathbf{W}}_\nu \mathbf{x}^{k+1})_i - \mathbf{v}_i^{k+1})$
    Send $\mathbf{v}_i^{k+1}$ and $\gamma_{\mathbf{v}}^{k+1}$ to nodes of group A
  end for
  MPI synchronization barrier
  k = k+1
end while
return $\mathbf{x}^{k-1}$

the pixels). Algorithm 2 summarizes the PG approach where the calculation of the gradient defined in equation (10) is distributed on the two groups of nodes architecture. The parallel programming of both the algorithms on the distributed architecture (figure 1) is done using message passing interface (MPI) that allows the nodes of the cluster to exchange data.

6. RESULTS AND CONCLUSION

The two algorithms are tested on a quasi-real multispectral image of size $256 \times 256 \times 100$ pixels. We use the radio emission image from an HII region of the M31 galaxy (figure 2 on the left) as the reference image. We build a multispectral image by applying a sine wave spectrum to each pixel of the image. The dirty image (figure 2 on the right) is obtained by convolution with a 2D Gaussian PSF and corrupted by a white Gaussian noise whose signal-to-noise ratio (SNR) is from 15dB to 25dB according to the frequencies.

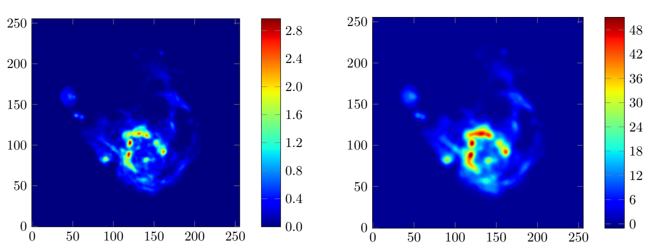

Fig. 2: Left: reference image of the M31 galaxy. Right: Dirty image at a given frequency (SNR = 25dB).

We present in figure 3 the deconvolution results produced by ADMM and PG approaches. The regularization parameters for ADMM and PG on the dual are set to the following values: $\mu_\epsilon = 1$, $\mu_s = 0.1$ and $\mu_\nu = 0.5$ that provide the best reconstruction on synthetic multispectral images. The ADMM parameter ρ is set to 1, while $\rho = 1/K$ for PG where $K = \frac{\mu_s^2 m_s + \mu_\nu^2}{\mu_\epsilon}$ is the Lipschitz constant of the gradient $\nabla[f^* \circ -\mathbf{W}^T]$. The cluster contains 40 nodes (20 for group A and 20 for group B). Since the two problems (2) and (8) are not strictly equivalent, we propose to use the relative variation $\delta_{\mathbf{x}}$ for the reconstructed image for the convergence criterion and the mean square error

$$\delta_{\mathbf{x}} = \frac{\|\mathbf{x}^{k+1} - \mathbf{x}^k\|_2}{\|\mathbf{x}^{k+1}\|_2} \quad \text{and} \quad \mathrm{MSE} = \mathrm{E}\left[(\mathbf{x}^k - \mathbf{x})^2\right], \quad (11)$$

where $\mathbf{x}$ is the theoretical deconvolved image. As expected, using PG on the dual requires less iterations to converge to a solution than using ADMM algorithm: the variation of the reconstructed image is lower than for ADMM case. Evaluation of the reconstruction quality is not the purpose of this paper, but MSE curves are coherent according to images reconstructed with ADMM and PG. In both case, the solution is slightly biased (it can be seen on reconstructed spectrum) due to the Tikhonov regularization.

In this work we have proposed a preliminary study of two distributed optimization algorithm to solve the regularized optimization problem described in [2]. Details of both the algorithms will be reported in an extended paper where the distributed architecture will be more precisely developed. In this paper we applied the algorithms on a $256 \times 256 \times 100$ image.

Algorithm 2 Distributed PG for dual problem

Initialization
for nodes in group A **do**
  Eval. $\mathbf{\Delta}^{-1}\mathbf{H}^T\mathbf{y}$
  Send $\mathbf{\Delta}^{-1}\mathbf{H}^T\mathbf{y}$ to B
  Evaluate and save $\mu_s \mathbf{W}_s \mathbf{\Delta}^{-1}\mathbf{H}^T\mathbf{y}_l$
end for
for nodes in group B **do**
  Receive $\mathbf{\Delta}^{-1}\mathbf{H}^T\mathbf{y}$ from A
  Evaluate and save $\mu_\nu \tilde{\mathbf{W}}_\nu \mathbf{\Delta}^{-1}\mathbf{H}^T\mathbf{y}$
end for
while $\delta_{\boldsymbol{\lambda}}^k \geqslant 10^{-5}$ **do**
  for nodes in group A **do**
    Eval. $\mu_s \tilde{\mathbf{W}}_s^T \boldsymbol{\lambda}_s^k$
  end for
  for nodes in group B **do**
    Eval. $\mu_\nu \tilde{\mathbf{W}}_\nu^T \boldsymbol{\lambda}_\nu^k$
    Send $\mu_\nu \tilde{\mathbf{W}}_\nu^T (\boldsymbol{\lambda}_\nu^k)$ to A
  end for
  MPI synchronization barrier
  for nodes in group A **do**
    Receive $\mu_\nu \tilde{\mathbf{W}}_\nu^T \boldsymbol{\lambda}_\nu^k$ from B
    Eval. $\mathbf{\Delta}^{-1}\left(\mu_s \tilde{\mathbf{W}}_s^T \boldsymbol{\lambda}_s^k + \mu_\nu \tilde{\mathbf{W}}_\nu^T \boldsymbol{\lambda}_\nu^k\right)$
    Send $\mathbf{\Delta}^{-1}\left(\mu_s \tilde{\mathbf{W}}_s^T \boldsymbol{\lambda}_s^k + \mu_\nu \tilde{\mathbf{W}}_\nu^T \boldsymbol{\lambda}_\nu^k\right)$ to B
    Eval. $\boldsymbol{\theta}_s = \mu_s \tilde{\mathbf{W}}_s \mathbf{\Delta}^{-1}\left(\mu_s \tilde{\mathbf{W}}_s^T \boldsymbol{\lambda}_s^k + \mu_\nu \tilde{\mathbf{W}}_\nu^T \boldsymbol{\lambda}_\nu^k\right)$
    $\boldsymbol{\lambda}_s^{k+1} \leftarrow \mathcal{P}_\infty\left(\boldsymbol{\lambda}_s^k - \rho\left(\boldsymbol{\theta}_s - \mu_s \mathbf{W}_s \mathbf{\Delta}^{-1}\mathbf{H}^T\mathbf{y}\right)\right)$
  end for
  for nodes in group B **do**
    Receive $\mathbf{\Delta}^{-1}\left(\mu_s \tilde{\mathbf{W}}_s^T \boldsymbol{\lambda}_s^k + \mu_\nu \tilde{\mathbf{W}}_\nu^T \boldsymbol{\lambda}_\nu^k\right)$ from A
    Eval. $\boldsymbol{\theta}_\nu = \mu_\nu \mathbf{W}_\nu \mathbf{\Delta}^{-1}\left(\mu_s \tilde{\mathbf{W}}_s^T \boldsymbol{\lambda}_s^k + \mu_\nu \tilde{\mathbf{W}}_\nu^T \boldsymbol{\lambda}_\nu^k\right)$
    $\boldsymbol{\lambda}_\nu^{k+1} \leftarrow \mathcal{P}_\infty\left(\boldsymbol{\lambda}_\nu^k - \rho\left(\boldsymbol{\theta}_\nu - \mu_\nu \mathbf{W}_\nu \mathbf{\Delta}^{-1}\mathbf{H}^T\mathbf{y}\right)\right)$
  end for
  k = k+1
end while
return $\boldsymbol{\lambda}^{k-1}$

This helped to test the algorithms before applying them to real data of size $(2048 \times 2048 \times 256)$ that should be soon available. We also would like to modify the minimization problem formulation to try smoother way for imposing the positivity on the reconstructed image.

7. ACKNOWLEDGEMENT

We thank Maxime Flauder for useful suggestions regarding the distributed implementation.

REFERENCES

[1] U. Rau and T. J. Cornwell, "A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry," *Astronomy & Astrophysics*, vol. 532, pp. A71, 2011.

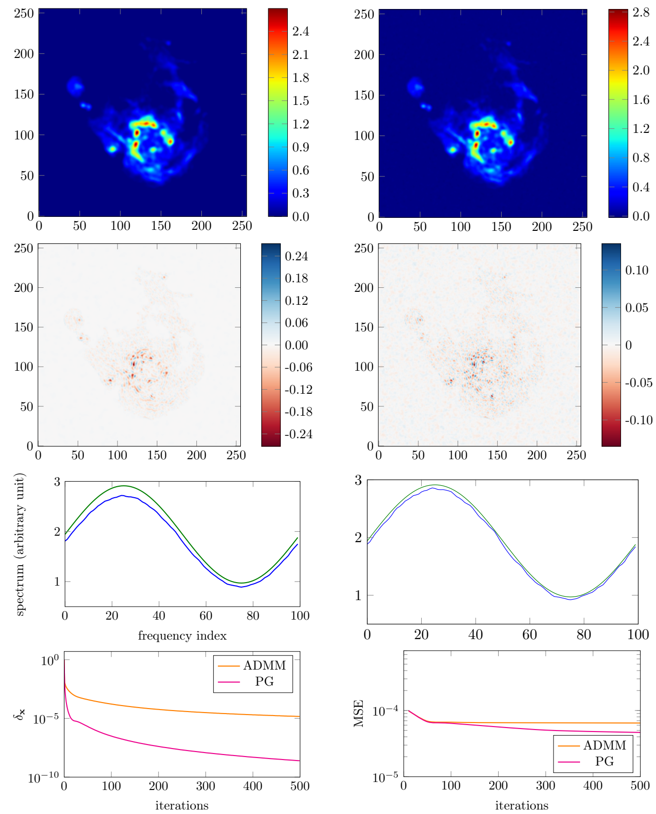

Fig. 3: Left column: results from ADMM algorithm. Right column: results from PG on the dual problem. From top to bottom: deconvolution result at a given frequency, error between the reference image and the reconstructed one, estimated spectrum (blue) vs theoretical spectrum (green), last line: relative variation of reconstructed image (left) and mean square error (right) for both ADMM (orange lines) and PG (magenta lines).

[2] A. Ferrari, J. Deguignet, C. Ferrari, D. Mary, A. Schutz, and O. Smirnov, "Multi-frequency image reconstruction for radio interferometry. A regularized inverse problem approach," in *SKA Pathfindes Radio Continuum Surveys (SPARCS)*, 2015.

[3] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Purify: a new approach to radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 439, no. 4, pp. 3591–3604, 2014.

[4] A. Onose, R. E. Carrillo, A. Repetti, J. D. McEwen, J.-P. Thiran, J.-C. Pesquet, and Y. Wiaux, "Scalable splitting algorithms for big-data interferometric imaging in the SKA era," *arXiv preprint arXiv:1601.04026*, 2016.

[5] H. Garsden, J.N. Girard, J.-L. Starck, S. Corbel, C. Tasse, A. Woiselle, J.-P. Mckean, A. S. Van Amesfoort, J. Anderson, I. M. Avruch, et al., "Lofar sparse image reconstruction," *Astronomy & astrophysics*, vol. 575, pp. A90, 2015.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[7] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.