



**HAL**  
open science

## A parallel and automatically tuned algorithm for multispectral image deconvolution

R Ammanouil, André Ferrari, D Mary, C Ferrari, F Loi

► **To cite this version:**

R Ammanouil, André Ferrari, D Mary, C Ferrari, F Loi. A parallel and automatically tuned algorithm for multispectral image deconvolution. *Monthly Notices of the Royal Astronomical Society*, 2019, 490 (1), pp.37-49. 10.1093/mnras/stz2193 . hal-03434392

**HAL Id: hal-03434392**

**<https://hal.science/hal-03434392>**

Submitted on 1 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A parallel and automatically tuned algorithm for multispectral image deconvolution

R. Ammanouil,<sup>1,2★</sup> A. Ferrari,<sup>2★</sup> D. Mary,<sup>2★</sup> C. Ferrari<sup>2</sup> and F. Loi<sup>3</sup>

<sup>1</sup>College of Engineering and Technology, American University of the Middle East, Kuwait

<sup>2</sup>Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Lab. J.-L. Lagrange, Nice 06108, France

<sup>3</sup>Dipartimento di Fisica e Astronomia, Università degli Studi di Bologna, Bologna 40126, Italy

Accepted 2019 August 5. Received 2019 August 5; in original form 2019 May 21

## ABSTRACT

In the era of big data, radio astronomical image reconstruction algorithms are challenged to estimate clean images given limited computing resources and time. This article is driven by the need for large-scale image reconstruction for the future Square Kilometre Array (SKA), which will become in the next decades the largest low and intermediate frequency radio telescope in the world. This work proposes a scalable wide-band deconvolution algorithm called MUFFIN, which stands for ‘MUlti Frequency image reconstruction For radio INterferometry’. MUFFIN estimates the sky images in various frequency bands, given the corresponding dirty images and point spread functions. The reconstruction is achieved by minimizing a data fidelity term and joint spatial and spectral sparse analysis regularization terms. It is consequently non-parametric w.r.t. the spectral behaviour of radio sources. MUFFIN algorithm is endowed with a parallel implementation and an automatic tuning of the regularization parameters, making it scalable and well suited for big data applications such as SKA. Comparisons between MUFFIN and the state-of-the-art wide-band reconstruction algorithm are provided.

**Key words:** methods: data analysis – methods: numerical – techniques: image processing – techniques: interferometric.

## 1 INTRODUCTION

The new generation of radio interferometers is characterized by very wide fields of view, high-resolution, huge sensitivity and very large bandwidth. The future Square Kilometre Array (SKA) is emblematic of these extreme instruments. In its final phase of construction, SKA will consist of up to one million dipole antennas and two thousands parabolic antennas (Dewdney et al. 2009, 2013). These antennas, comprising a collecting area of approximately 1 km<sup>2</sup>, will be located in Australia and South Africa. The SKA radio telescope is expected to provide images of the sky in the giga pixel range with (sub-)arcsec angular resolution, fields of view up to 200 deg<sup>2</sup> and more sensitivity than any other present-day instrument with a dynamic range covering up to seven orders of magnitude (see e.g. table 1 in Dewdney et al. 2013).

The work presented here addresses the large-scale and high-precision radio-interferometric image reconstruction problem. Radio-interferometric data, the visibilities, are the cross-correlations between voltages from pairs of antennas. In the ideal case, two receivers with baseline  $\mathbf{b}$  observing in a narrow frequency

band  $c/\lambda$  measure a complex visibility equal to the Fourier spectrum of the intensity distribution of interest at spatial frequency  $\mathbf{b}/\lambda$ . The partial measurement of the Fourier plane is what makes the image reconstruction strongly ill posed. In practice, the measurement model corresponds to a linear map from the image domain to the visibility space, but includes multiple distortions such as (a) direction-dependent effects (DDE) related to wide-field effects, non-coplanarity of the baselines, the ionospheric perturbations, and associated Faraday rotation; (b) antenna level effects such as the primary beam patterns, element-based gains; and (c) computational approximation of operators such as the degriding required to use Fast Fourier Transform algorithms (see e.g. Bhatnagar et al. 2008; Rau et al. 2009; Wijnholds et al. 2010). Compensations of the perturbations (a) and (b) rely on alternating a (self-)calibration step and an image reconstruction step. This paper concentrates on the image reconstruction step.

The imaging task is realized by solving either directly the inverse problem related to the calibrated visibility measurements, or the inverse problem in the image domain after applying the adjoint measurement model to the visibilities. The latter approach relies on the construction of a ‘dirty image’ from calibrated visibilities. The dirty image is related to the sky image by a convolution if the DDEs are properly compensated. This is the approach followed in the classical major/minor loop (Tasse et al. 2013; Jongerius et al. 2014) and in the

\* E-mail: [ritaammanouil@gmail.com](mailto:ritaammanouil@gmail.com) (RA); [andre.ferrari@unice.fr](mailto:andre.ferrari@unice.fr) (AF); [david.mary@unice.fr](mailto:david.mary@unice.fr) (DM)

facet frameworks (Kogan & Greisen 2009; van Weeren et al. 2016) – see Tasse et al. (2018) for a detailed comparison between these two techniques. The main advantage of the image domain processing is not only the computing efficiency related to the fact that the direct and adjoint operators are simple convolutions but also the amount of data to process: the dirty image size equals the reconstructed sky image size, while the number of raw visibilities can be several order of magnitudes the number of pixels in the sky image.

Interestingly, sparsity was early recognized as a powerful tool for radio-interferometric image reconstruction and has led to the most populated family of imaging algorithms. Their patriarch is the CLEAN algorithm (Högbom 1974), which expresses and exploits the sparsity of the sky intensity distribution in the canonical basis. Variants of CLEAN as Cornwell (2008) generalize this approach to extended objects. More recently, various algorithms based on the theory of compressed sensing (Donoho 2006; Candès & Wakin 2008) and convex optimization emerged for the reconstruction of radio-interferometric images from the calibrated visibilities. These algorithms exploit the fact that the image has a sparse representation in some appropriate transform domain through the use of sparsifying regularizations (see e.g. Wiaux et al. 2009; McEwen & Wiaux 2011; Kartik et al. 2017). A greedy alternative to the minimization of a regularized cost function was proposed by Dabbech et al. (2015). Sparse image representations using wavelet dictionaries has indisputably shown substantial improvement with respect to classical CLEAN-based approaches, as demonstrated by Garsden et al. (2015) using LOFAR data and Dabbech et al. (2018) using VLA data. The price to pay is an increased computational cost. As a result, a lot of effort has gone into designing scalable methods able to cope with large data sets. For example, the authors of Onose et al. (2016a, b) achieve parallel computations by splitting the visibilities into multiple blocks and exploiting randomization over data blocks at each iteration.

The SKA will observe the sky in wide sub-bands (covering all together from 50 MHz to 14 GHz) and the reconstruction of both spatial and spectral behaviour of radio sources is essential to characterize the astrophysical origin of the detected radiation (see e.g. Kraus 1986). As a result, multi-frequency reconstruction algorithms are required in order to fully exploit the spectral and spatial resolution of the SKA telescope. Taking the path of CLEAN, the multifrequency synthesis (MFS; Conway, Cornwell & Wilkinson 1990) and the multiscale multifrequency synthesis (MS-MFS; Rau & Cornwell 2011), deconvolution algorithms were the first wide-band (WB) radio-interferometric reconstruction algorithms. At the time of writing, still relatively few such WB reconstruction algorithms exist (Bajkova & Pushkarev 2011; Wenger & Magnor 2014; Junklewitz, Bell & Enßlin 2015; Abdulaziz et al. 2016).

In Rau & Cornwell (2011), Bajkova & Pushkarev (2011), and Junklewitz et al. (2015), the spectrum of each pixel is modelled as a (generalized) power law, or its Taylor expansion. These ‘semiparametric’ methods rely on spectral models and thus clearly offer many advantages such as estimation accuracy when the model is appropriate. However, across the broad-frequency coverage of current radio facilities, radio sources exhibiting complex spectral shapes (not simple power laws) are expected. For instance, Kellermann (1974) evidences that some sources may exhibit one or more relative minima, breaks, and turnovers. For the new generation of low-frequency telescopes such as LOFAR, recent studies have also shown that second-order broad-band spectral models are often insufficient (Scaife & Heald 2012). In addition to that, not assuming any spectral shape for radio sources in the deconvolution step can be useful for mutliwavelength polarization studies, as well as for simultane-

ously identifying spectral lines within WB continuum observations. However, few WB reconstruction algorithms consider the general case of reconstructing the spatio-spectral ‘image cube’ without constraining the spectral variation to a power law. In particular, few WB reconstruction algorithms have opted instead for a sparse regularization of the spectra. Among these, Wenger & Magnor (2014) formulate the problem as an inverse problem with a smooth spectral regularization allowing for local deviations; Abdulaziz et al. (2016) constrain the sky image cube to be low rank in order to enforce that only few sources, each with a distinct spectral signature, can exist in the field of view. This approach requires a singular value decomposition of the whole sky image cube at each iteration, so computational costs can become an issue. Finally, WB image reconstruction has also recently been tackled as a joint blind source separation and reconstruction problem by Jiang, Bobin & Starck (2017).

The present article focuses on total intensity WB and HI imaging task within the minor loop. Section 2 derives the proposed algorithm that is non-parametric in the sense that it does not rely on a particular assumption of the spectral structure, but rather on a sparse representation of the spectra in a dictionary. The spectral dimension critically blows up the size of the inverse problem and a particular effort is dedicated to the scalability of the algorithm. This results in an algorithm where most computationally intensive steps, such as the spatial wavelets decompositions and the associated adjoint operators, can be computed in parallel for each wavelength. Given the execution time for a large-scale reconstruction, the selection of the regularization parameters is particularly critical. Section 3 proposes an automatic tuning strategy based on the Stein Unbiased Risk Estimator. Finally, Section 4 presents simulation results that demonstrate the effectiveness of the algorithm.

It is worthy to note that, whereas simulations in Section 4 are realized in the deconvolution framework, the proposed algorithm can be used in the same way to reconstruct the sky image cube from dirty images or visibilities, depending on the direct operator.

## 2 INVERSE PROBLEM

In this section, we first formulate the multifrequency image deconvolution task as a regularized and constrained least-squares optimization problem. We then propose and discuss a primal-dual algorithm for solving the aforementioned optimization problem.

### 2.1 Optimization problem

Let  $\mathbf{x}_\ell^*$  and  $\mathbf{y}_\ell$  be  $N \times 1$  column vectors collecting the unknown (and sought-after) sky intensity image and the (measured) dirty intensity image, respectively. The subscript  $\ell$  in the notations  $\mathbf{x}_\ell^*$ ,  $\mathbf{y}_\ell$ , and in all following matrix and vector notations refers to the wavelength band  $\lambda_\ell$ , with  $\ell = 1, \dots, L$  and  $L$  the number of spectral channels. At each wavelength band, the dirty image  $\mathbf{y}_\ell$  is related to the sky image  $\mathbf{x}_\ell^*$  by a linear mapping and contaminated with noise. This takes the following form:

$$\mathbf{y}_\ell = \mathbf{H}_\ell \mathbf{x}_\ell^* + \mathbf{w}_\ell, \quad \ell = 1, \dots, L \quad (1)$$

where  $\mathbf{H}_\ell$  represents a convolution by the point spread function (PSF) and  $\mathbf{w}_\ell$  is a perturbation vector accounting for noise and modelling error. The perturbation vector  $\mathbf{w}_\ell$  is assumed to be Gaussian, independent and identically distributed (i.i.d) with zero mean. Without loss of generality we will assume that the variance of the noise does not depend on the wavelength and is equal to  $\sigma_w^2$ . More sophisticated correlation structures can be taken into account provided that the covariance matrix is known. Given

the dirty images and the PSF, equation (1) defines an ill-posed deconvolution problem. The problem is ill-posed owing to the partial coverage of the Fourier plane, i.e. due to the fact that  $\mathbf{H}_\ell$  is rank deficient. A meaningful solution can be obtained in a cost minimization framework by incorporating regularizations and constraints reflecting our prior knowledge on the unknown variables  $\mathbf{x}_1^*, \dots, \mathbf{x}_L^*$ , the true sky intensity images. Let  $\mathbf{x}_1, \dots, \mathbf{x}_L$  denote the optimization variables corresponding to  $\mathbf{x}_1^*, \dots, \mathbf{x}_L^*$ , and let  $\mathbf{X}$  denote the  $N \times L$  matrix  $\mathbf{X} := [\mathbf{x}_1, \dots, \mathbf{x}_L]$  obtained by stacking the image vectors next to each other. Let  $\mathbf{x}^1, \dots, \mathbf{x}^N$  denote the rows of  $\mathbf{X}$  that represent the pixels spectra. Hence,  $\mathbf{X}$  can be alternatively obtained by stacking the pixels spectra on top of each other, i.e.  $\mathbf{X} := [\mathbf{x}^{1\top}, \dots, \mathbf{x}^{N\top}]^\top$ . With these notations, we propose to minimize the following cost function:

$$\operatorname{argmin}_{\mathbf{X}} \sum_{\ell=1}^L \frac{1}{2} \|\mathbf{y}_\ell - \mathbf{H}_\ell \mathbf{x}_\ell\|^2 + \mathbf{1}_{\mathbb{R}^+}(\mathbf{X}) + h(\mathbf{X}), \quad (2)$$

with

$$h(\mathbf{X}) := \sum_{\ell=1}^L \mu_\ell \|\mathbf{W}_s \mathbf{x}_\ell\|_1 + \sum_{n=1}^N \gamma_n \|\mathbf{W}_\lambda \mathbf{x}^n\|_1. \quad (3)$$

The first term in the cost function in equation (2) is a data fidelity term penalizing the discrepancy between the measurements and the estimated model. The second term in equation (2) is a positivity constraint enforcing each element in  $\mathbf{X}$  to be positive, more precisely  $\mathbf{1}_{\mathbb{R}^+}(\mathbf{X}) = 0$  if  $\mathbf{X} \geq 0$  (where the inequality is element-wise) and  $\mathbf{1}_{\mathbb{R}^+}(\mathbf{X}) = \infty$  otherwise. The third term in equation (2), developed in equation (3), is a spatial-spectral regularization. More precisely,  $h(\mathbf{X})$  is the sum of a spatial regularization imposed on the images at all the wavelength bands and a spectral regularization imposed on the spectra at all the pixels. The matrices  $\mathbf{W}_s$  and  $\mathbf{W}_\lambda$  represent the spatial and spectral decomposition operators, respectively. The scalars  $\mu_\ell \in \mathbb{R}^+$  for  $\ell = 1, \dots, L$  and  $\gamma_n \in \mathbb{R}^+$  for  $n = 1, \dots, N$  are the spatial and the spectral regularization weights, respectively.

The spatial and spectral regularizations in equation (3) are sparse analysis regularizations (Elad, Milanfar & Rubinstein 2007). The sparse analysis approach consists in penalizing the  $\ell_1$ -norm of the correlations between the atoms in a dictionary, i.e.  $\mathbf{W}_s$  and  $\mathbf{W}_\lambda$  in the case of equation (3), and the images, i.e.  $\mathbf{x}_\ell$  for  $\ell = 1, \dots, L$  and the spectra, i.e.  $\mathbf{x}^n$  for  $n = 1, \dots, N$ . Penalizing the  $\ell_1$ -norm of these decompositions promotes a sparse correlation between the atoms in the dictionary and the variables. Hence, it encourages finding a solution that is not correlated to some atoms of the dictionary and conversely more correlated to others. The dictionaries can be orthonormal transforms or more generally redundant transforms also referred to as overcomplete dictionaries. Possible choices for  $\mathbf{W}_s$  are (concatenations of) Daubechies wavelets (Carrillo, McEwen & Wiaux 2014; Onose et al. 2016b) or the isotropic undecimated wavelet transform (IUWT; Starck & Bobin 2010; Dabbech et al. 2015; Garsden et al. 2015) that have proved their efficiency for various image processing tasks in astronomy. One possible choice for  $\mathbf{W}_\lambda$  that will be illustrated in the experiments is the discrete cosine transform (DCT), a Fourier-related transform. The DCT spectral regularization assumes that the DCT is compact, an assumption widely used in the sparse reconstruction literature.

Other types of regularizations based on sparse representations, such as synthesis priors or more generally representations based on 3D spatio-spectral dictionaries, can be of course considered. See Lanusse et al. (2014) for a general discussion on 3D wavelets and Paris et al. (2013) where a spatio-spectral dictionary is derived for a specific detection task. The drawback of these regularizations is

computational: contrarily to equation (3), they couple space and frequency, i.e. the lines and columns of  $\mathbf{X}$ , and prevent an efficient implementation of equation (2), e.g. with parallel steps. This point will be clarified in the next section. Finally, it is essential to note that if  $\forall \ell, \gamma_\ell = 0$ , the minimization problem in equation (2) becomes separable w.r.t.  $\mathbf{x}_\ell$  and reduces to  $L$  deconvolution problems that can be solved independently at each wavelength  $\lambda_\ell$ .

The regularization parameters  $\mu_\ell$  for  $\ell = 1, \dots, L$  and  $\gamma_n$  control the strength of the regularization term w.r.t. the data fidelity term. A large value of  $\mu_\ell$  tends to oversmooth structures at wavelength  $\ell$ , while a small value of  $\mu_\ell$  leads to underregularization. Similarly, a large value of  $\gamma_n$  tends to oversmooth spectra at pixel  $n$ . See e.g. Renard, Thiébaud & Malbet (2011) in the case of a single regularization parameter for monochromatic optical interferometry. Although the use of a different regularization parameter per pixel and band could improve the estimation performance compared to the use of one regularization parameter per regularization type, the tuning of  $(L + N)$  regularization parameters is in practice very challenging. For this reason, we propose to use two regularization parameters and assign a different (fixed) weight per pixel and per band. By doing this, we reduce the number of regularization parameters from  $(L + N)$  to 2 and we take into account the spatial and spectral variability of the images and spectra, respectively. We adopt the following model for the regularization parameters:

$$\begin{cases} \gamma_n &= \gamma \alpha_n, & n = 1, \dots, N, \\ \mu_\ell &= \mu \beta_\ell, & \ell = 1, \dots, L, \end{cases} \quad (4)$$

with

$$\begin{cases} \alpha_n &= \frac{1}{\|\mathbf{W}_\lambda(\mathbf{x}^n)_0\|_{2+\varepsilon}}, \\ \beta_\ell &= \frac{1}{\|\mathbf{W}_s(\mathbf{x}_\ell)_0\|_{2+\varepsilon}}, \end{cases} \quad (5)$$

where  $(\mathbf{x}^n)_0$  and  $(\mathbf{x}_\ell)_0$  correspond to the spectra and images of the sky image used for initializing the algorithm. The intuition for the weights setting is similar to the one in re-weighted  $\ell_1$  minimization schemes (Candès, Wakin & Boyd 2008), where the strength of the regularization parameter is inversely proportional to the norm of the corresponding regularized norm. The intuition for this setting is that the value of the regularization parameter should be high whenever the decomposition coefficient is close to zero in order to encourage sparsity, and it should be small otherwise. Nevertheless, in re-weighted  $\ell_1$  minimization schemes, the weights may be updated at each iteration or after a certain number of iterations using the current estimates for  $\mathbf{x}^n$  and  $\mathbf{x}_\ell$ . In this work, and in the simulations, we consider fixed weights computed from a coarse initial estimation of  $\mathbf{X}$ .

Finally, note that the proposed optimization problem in equation (2) is convex but does not admit a closed form solution. In order to find the minimizer of equation (2), we need to resort to an iterative procedure as described in the next section.

## 2.2 Optimization algorithm

The proposed optimization problem described in equations (2) and (3) is the sum of a smooth and differentiable function (the LS term), and three non-smooth and non-differentiable functions (the indicator function, the sparse analysis spectral regularization, and the sparse analysis spatial regularization). Nevertheless, the non-differentiable functions, the indicator function and the two sparse analysis regularizations, are simple functions in the sense that their proximity operators (Combettes & Pesquet 2011) have closed form representations. In order to solve the proposed optimization prob-

lem, we resort to the primal dual splitting algorithm proposed by Vü (2011); Condat (2014) also known as the Vü–Condat algorithm. The Vü–Condat algorithm is well suited for optimization problems similar to the one proposed in this work, i.e. problems that consist of a sum between a smooth function and several proximable functions involving linear operators. Note that other splitting algorithms could have been used such as the alternating direction method of multipliers (ADMM; Boyd et al. 2011), simultaneous direction method of multipliers (SDMM; Setzer, Steidl & Teuber 2010; Combettes & Pesquet 2011), or the primal dual splitting algorithm proposed by Combettes & Pesquet (2012). These algorithms operate by splitting the objective function to be minimized into smaller and simpler functions that are dealt with individually. However, the Vü–Condat algorithm is more appealing compared to ADMM and SDMM because it does not involve any implicit inversion of the operators  $\mathbf{I} + \alpha \mathbf{W}_\lambda^\top \mathbf{W}_\lambda$  and  $\mathbf{I} + \beta \mathbf{W}_s^\top \mathbf{W}_s$ , which can be very large and cumbersome to invert (see for example the work by Meillier et al. 2018). It can be considered as a generalization of Chambolle–Pock algorithm (Chambolle & Pock 2010), which includes a differentiable term and multiple analysis regularizations.

Vü–Condat algorithm is not symmetric with respect to the primal and dual variables. Switching the roles of these variables gives rise to two different versions. A straightforward application of the first Vü–Condat algorithm (algorithm 3.1 of Condat 2014) to the optimization problem described in equations (2) and (3) with no relaxation ( $\rho_n = 1$ ) leads to Algorithm 1. These choices are discussed at the end of this sub-section. We refer to Algorithm 1 as MUFFIN for Multi Frequency image reconstruction For radio INterferometry. Algorithm 1 requires using the proximity operator of the positivity indicator function, namely the projection on the positive orthant defined as

$$(u)_+ := \begin{cases} u & \text{if } u > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

The proximity operator of the  $\ell_1$ -norm is the well-known soft thresholding operator. Algorithm 1 requires using the proximity operator of the Fenchel conjugate of the  $\ell_1$ -norm function which can be easily computed from the soft thresholding operator using Moreau’s identity. This leads to the saturation function defined as

$$\text{sat}(u) := \begin{cases} -1 & \text{if } u < -1 \\ 1 & \text{if } u > 1 \\ u & \text{if } |u| \leq 1. \end{cases} \quad (7)$$

$$\tau \left( \frac{\beta}{2} + \sigma \left( \sum_{\ell=1}^L \mu_\ell^2 \|\mathbf{W}_s\|^2 + \sum_{n=1}^N \gamma_n^2 \|\mathbf{W}_\lambda\|^2 \right) \right) < 1, \quad (8)$$

where  $\beta$  is the Lipschitz constant of the gradient of the LS term in equation (2).

Hereafter, we briefly describe each step in Algorithm 1. In equation (9), the gradient of the LS term is computed. This consists of applying the linear operator  $\mathbf{H}_\ell^\dagger$  on the residue of the model for  $\ell = 1, \dots, L$ . In equation (10),  $\tilde{\mathbf{X}}$  is computed. The adjoint operators  $\mathbf{W}_s^\dagger$  and  $\mathbf{W}_\lambda^\dagger$  are applied on the dual variables  $\mathbf{U}$  and  $\mathbf{V}^\top$ . More precisely, the third term is where the operator  $\mathbf{W}_s^\dagger$  is applied on  $\mathbf{U}$ , and the fourth term is where the operator  $\mathbf{W}_\lambda^\dagger$  is applied on  $\mathbf{V}^\top$  ( $\mathbf{V}^\top \mathbf{W}_\lambda = (\mathbf{W}_\lambda^\dagger \mathbf{V}^\top)^\top$ ). The primal variable  $\mathbf{X}$  is updated using the gradient and the aforementioned inverse transforms, and a projection on the positive orthant is then applied. The result is then assigned to the intermediate primal variable  $\tilde{\mathbf{X}}$ . In equation (11), the

first dual variable  $\mathbf{U}$  is computed. The linear operator  $\mathbf{W}_s$  is applied on  $(2\tilde{\mathbf{X}} - \mathbf{X})$ . The dual variable  $\mathbf{U}$  is updated using previously computed transform, and the saturation function is then applied. The result is assigned to the dual variable  $\mathbf{U}$ . In equation (12), similarly to the previous step, the second dual variable  $\mathbf{V}$  is computed. The linear operator  $\mathbf{W}_\lambda$  is applied on  $(2\tilde{\mathbf{X}} - \mathbf{X})^\top$ . The dual variable  $\mathbf{V}$  is updated using previously computed transform, and the saturation function is then applied. The result is assigned to the dual variable  $\mathbf{V}$ . In equation (13),  $\mathbf{X}$  is updated by simply assigning it to  $\tilde{\mathbf{X}}$ . Finally, note how the proposed iterative Algorithm 1 activates separately: the gradient of the LS term, i.e. the linear operator  $\mathbf{H}_\ell^\dagger$  for  $\ell = 1, \dots, L$ , the linear operators  $\mathbf{W}_s$ ,  $\mathbf{W}_\lambda$ , and their adjoint operators  $\mathbf{W}_s^\dagger$ , and  $\mathbf{W}_\lambda^\dagger$ . For example, the spatial and spectral operators are applied on images and spectra, respectively, in separate steps (see for example  $\mathbf{W}_s \mathbf{X}$  in equation (11) and  $\mathbf{W}_\lambda \mathbf{X}^\top$  in equation 12). Finally, for a better understanding of the derivation of Algorithm 1, the reader can refer to Algorithm 3 (section 4 in Condat 2014).

From (11), the size of  $\mathbf{U}$  is the size of  $\mathbf{W}_s \mathbf{X}$  and consequently several times the size of the cube of images  $\mathbf{X}$  when  $\mathbf{W}_s$  is a redundant transform such as a union of orthogonal wavelets or an IUWT decomposition. Symmetrically to equation (11), the second version of the Vü–Condat algorithm (algorithm 3.2 Condat 2014) requires, for the update of  $\mathbf{X}$ , the update of  $\mathbf{U}$  in the current iteration and in the previous iteration. Keeping these two variables in memory can be too cumbersome when  $\mathbf{X}$  is large and when the spatial decomposition is redundant. For this reason, we privileged the first version of the algorithm (algorithm 3.1 Condat 2014). Similarly, in order to reduce memory requirements, the relaxed version of the algorithm is not used ( $\rho_n = 1$ ).

### 2.3 Parallel implementation

The parallel implementation consists in parallelizing the steps that are carried on the images at different wavelength bands separately. The advantage of this parallelization is twofold. First, the memory load can be distributed among several slave nodes, a slave node only stores the images at certain wavelength bands and not at all wavelength bands. Secondly, most of the computations and most interestingly the most expensive ones (such as the application of  $\mathbf{W}_s$  and  $\mathbf{W}_s^\dagger$ ) can be also carried out in parallel. More precisely, a group of  $L$  slave nodes, where each slave node is assigned a band, handles the gradient computation, the application of the spatial operator  $\mathbf{W}_s$ , and its adjoint operator  $\mathbf{W}_s^\dagger$ .

The parallel implementation of Algorithm 1 is described in Algorithm 2. The computations in equations (9)–(11), and equation (13), except for the last term in equation (10) are computed in a parallel manner for each band. The parallel computations of equations (9)–(11), and equation (13) in Algorithm 1, are described in equations (14)–(20) in Algorithm 2. After the slave nodes have performed their computations in parallel, each node sends the new value of  $\mathbf{x}_\ell$  and  $\tilde{\mathbf{x}}_\ell$  to the master node. The master node centralizes the computation of the last term in equations (10) and (12), since these computations operate on spectra rather than images. It is worthy to note that if  $\forall n \gamma_n = 0$ , i.e. the spectral regularization is omitted, computation steps on the master (21)–(23) can be removed: each slave reconstructs the image corresponding to its wavelength independently to the others. From a memory point of view, the master node stores the whole dirty image cube and psf, whereas each slave node as mentioned previously only stores the dirty image and PSF at the corresponding band. Finally, without any loss of generality, a node can be assigned a group of bands, i.e. it can

handle a sub-cube of the image rather than just one band. In this case, the communication overhead between the master node and the slave node can be reduced compared to the case where each band is assigned to a node.

$$\nabla = \left[ \mathbf{H}_1^\dagger (\mathbf{H}_1 \mathbf{x}_1 - \mathbf{y}_1) \mid \dots \mid \mathbf{H}_L^\dagger (\mathbf{H}_L \mathbf{x}_L - \mathbf{y}_L) \right] \quad (9)$$

$$\tilde{\mathbf{X}} = (\mathbf{X} - \tau (\nabla + \mathbf{W}_s^\dagger \mathbf{U} \mathbf{D}_\mu + \mathbf{D}_\gamma \mathbf{V} \mathbf{W}_\lambda))_+ \quad (10)$$

$$\mathbf{U} = \text{sat} (\mathbf{U} + \sigma \mathbf{W}_s (2\tilde{\mathbf{X}} - \mathbf{X}) \mathbf{D}_\mu), \quad (11)$$

$$\mathbf{V} = \text{sat} (\mathbf{V} + \sigma \mathbf{D}_\gamma (2\tilde{\mathbf{X}} - \mathbf{X}) \mathbf{W}_\lambda^\dagger) \quad (12)$$

$$\mathbf{X} = \tilde{\mathbf{X}} \quad (13)$$

**Algorithm 1:** MUFFIN update

Each node  $\ell = 1 \dots L$  computes sequentially:

$$\mathbf{x}_\ell = \tilde{\mathbf{x}}_\ell, \quad (14)$$

$$\nabla_\ell = \mathbf{H}_\ell^\dagger (\mathbf{H}_\ell \mathbf{x}_\ell - \mathbf{y}_\ell) \quad (15)$$

$$\mathbf{s}_\ell = \mu_\ell \mathbf{W}_s^\dagger \mathbf{u}_\ell \quad (16)$$

$$\mathbf{m}_\ell = \mathbf{x}_\ell - \tau (\nabla_\ell + \mathbf{s}_\ell + \mathbf{t}_\ell) \quad (17)$$

$$\tilde{\mathbf{x}}_\ell = (\mathbf{m}_\ell)_+ \quad (18)$$

$$\mathbf{p}_\ell = \mathbf{u}_\ell + \sigma \mu_\ell \mathbf{W}_s (2\tilde{\mathbf{x}}_\ell - \mathbf{x}_\ell) \quad (19)$$

$$\mathbf{u}_\ell = \text{sat} (\mathbf{p}_\ell), \quad (20)$$

and sends  $\mathbf{x}_\ell$  and  $\tilde{\mathbf{x}}_\ell$  to the master node.

The master computes sequentially, for  $n = 1 \dots N$ :

$$\tilde{\mathbf{v}}^n = \mathbf{v}^n + \sigma \gamma_n \mathbf{W}_\lambda (2\tilde{\mathbf{x}}^n - \mathbf{x}^n) \quad (21)$$

$$\mathbf{v}^n = \text{sat} (\tilde{\mathbf{v}}^n) \quad (22)$$

$$\mathbf{t}^n = \gamma_n \mathbf{W}_\lambda^\dagger \mathbf{v}^n, \quad (23)$$

and sends the col.  $\ell$  of  $\mathbf{T}$ , denoted as  $\mathbf{t}_\ell$ , to node  $\ell$ .

**Algorithm 2:** Parallel MUFFIN update

### 3 RISK ESTIMATION FOR OPTIMAL PARAMETER SELECTION

The regularization parameters values  $\mu$  and  $\gamma$  defined in equation (4), control the strength of the spatial and spectral regularizations w.r.t. the strength of the data fidelity term. Setting the regularization parameters values appropriately is crucial to ensure a good performance of the deconvolution algorithm, and to obtain a good estimated image  $\mathbf{X}$  of the true sky  $\mathbf{X}^*$ . However, judging the estimated image quality and its closeness to the true and sought after image is a non-trivial and complicated task. This is due to the lack of ground truth, i.e. due to the fact that the true image is unknown. In addition to this, a ‘trial-and-error’ technique consisting in the inspection of multiple reconstructions is extremely time-consuming when dealing with large-scale problems and multiple parameters.

The literature on optimal parameter selection, within the frequentist statistical paradigm, provides various quantitative measures for evaluating the quality of an estimated image without prior knowledge of the true image. These measures can be broadly classified as those based on the discrepancy principle (Morozov 1966; Karl 2005), the  $L$ -curve method (Hansen & O’Leary 1993; Regińska 1996), the generalized cross-validation (GCV; Golub, Heath & Wahba 1979), and the (generalized) Stein Unbiased Risk Estimator

(SURE; Stein 1981; Eldar 2009). Among these measures, the SURE is one of the most attractive: unlike the GCV, the  $L$ -curve, and the discrepancy principle, the SURE benefits from optimal properties for non-linear estimators and it is not restricted for linear ones. It provides an unbiased estimate of the risk, i.e. the mean square error (MSE), a quantitative measure often used for quality assessment in image processing. The SURE measure was originally applicable in the case of image denoising, i.e. when all the  $\mathbf{H}_\ell$  are equal to the identity matrix in equation (1). Nevertheless, the concepts underlying the SURE have been generalized to the case of image reconstruction. Weighted versions of the SURE allow the unbiased estimation of a weighted MSE in the case, where  $\mathbf{H}_\ell$  are not equal to the identity matrix as it is the case in equation (1). In this work, we consider a weighted version of the SURE known as the predicted SURE (PSURE). The PSURE and other weighted SURE metrics have been successfully adopted for automatic tuning in image deconvolution (see e.g. Giryes, Elad & Eldar 2011; Ramani et al. 2012; Deledalle et al. 2014; Ammanouil et al. 2017; and references therein).

A straightforward strategy for optimal parameter selection based on the SURE criteria or any other metric is by exhaustive search. More precisely, image reconstructions using all combinations of the regularization parameters are tested to achieve the minimum SURE estimate (see e.g. the work of Ramani, Blu & Unser 2008 and Eldar 2009). Compared to the grid search, a golden section allows to narrow the range of values inside which the extremum exists (Ramani et al. 2012). However, the golden section search is most suited for the case of one parameter value. In the case of multiple regularization parameters, it can be either used to successively find the optimal regularization parameters values as in Ammanouil et al. (2017) or more sophisticated method such as the simplex method can be used. In general, exhaustive search, or bisection strategies for multiple parameters, can become rapidly computationally prohibitive.

More recently, Deledalle et al. (2014) proposed to minimize the SURE (or PSURE) using a gradient descent approach. This approach is based on an approximation of the SURE that is weakly differentiable w.r.t. the regularization parameters. This approximation, denoted as SURE-FDMC, relies on Finite Differences and Monte Carlo approximations (FDMC). The weak gradient of the SURE-FDMC, called the Stein Unbiased Gradient estimator of the Risk (SUGAR), is used in a gradient descent scheme to find the optimal parameters. Among other advantages, the gradient descent scheme allows setting multiple parameters more easily than grid search and golden section approaches.

The exhaustive search using SURE and the gradient descent using SUGAR all require running the reconstruction algorithm until convergence various times, with different values for the regularization parameters, which can be prohibitive when the algorithm is computationally complex. Giryes et al. (2011) proposed to update the regularization parameter in parallel to the reconstruction algorithm: at each iteration of the reconstruction algorithm, the regularization parameter is updated by a minimization of the SURE computed a number of iterations ahead using a golden section algorithm. As noticed by Giryes et al. (2011), for a given number of iterations of the reconstruction algorithm and golden section, the computational cost of this greedy approach is larger than the global approach. In this work, we propose to estimate the optimal regularization parameters in a more computationally efficient way. Towards this goal, they are updated in parallel to the reconstruction algorithm using a gradient descent strategy based on SUGAR. More precisely, the regularization parameters are updated every  $\Delta$  iterations of the reconstruction algorithm such that the SURE-FDMC is reduced.

We refer to this method as ADAptive Parameter Tuning (ADA-PT). To the best of our knowledge, Garsden et al. (2015) and Jiang et al. (2017) are the only works in image reconstruction to propose an ‘online’ self-tuning of the parameters. Their approach relies on the fact that in many sparse reconstruction algorithms based on primal optimization methods, the regularization parameters act at each iteration as thresholds levels on reconstructed images. This is the case in Garsden et al. (2015) where Beck & Teboulle (2009) FISTA is used with a sparse synthesis prior. The regularization parameters can then be set as a function of a noise level estimation associated with the image update. In contrast with this approach, the application of the ADA-PT approach is not limited to MUFFIN and can be extended to most iterative reconstruction algorithms.

In what follows, we briefly review the essential concepts behind the estimation of the Predicted Stein Unbiased Risk Estimator based on Finite Differences and Monte Carlo simulations (PSURE-FDMC) and the Stein Unbiased Gradient estimator of the Risk introduced by Deledalle et al. (2014). We derive the iterative steps necessary for the computation of the SUGAR in the case of MUFFIN. We then show how SUGAR, in ADA-PT, is used to automatically and adaptively select the optimal regularization parameters. We also show that the steps required for self-tuning can be performed in parallel and in a memory efficient way.

### 3.1 Risk estimation: PSURE-FDMC

The SURE provides an unbiased estimate of the MSE. However, for ill-posed inverse problems, it is generally not possible to estimate the MSE since the measurements may only contain partial information about the true image. The concepts underlying the SURE have been extended to the case of image reconstruction using weighted versions of the SURE measure (Giryes et al. 2011; Ramani et al. 2012): the PSURE. The predicted risk, or predicted mean square error (PMSE) for model (1) refers to the norm of the discrepancy between the true sky and its estimate in the measurement domain, in other terms it refers to the reconstruction error measured in the  $\ell_2$ -norm sense:

$$R = \sum_{\ell=1}^L R_{\ell}, \quad (24)$$

where

$$R_{\ell} = \mathbb{E} \|\mathbf{H}_{\ell}(\mathbf{x}_{\ell}^* - \mathbf{x}_{\ell})\|^2. \quad (25)$$

The PSURE is an unbiased estimator of the PMSE, but unfortunately, it is not generally differentiable, e.g. when considering non-smooth regularizers as in equation (3). For this reason, we focus from now on the PSURE-FDMC. The PSURE-FDMC is an asymptotically unbiased estimator for the PMSE in the case of a linear model with additive white Gaussian noise as in equation (1). Let  $\hat{R}$  be the overall PSURE-FDMC. According to Deledalle et al. (2014), the PSURE-FDMC is given by

$$\hat{R} = \sum_{\ell=1}^L \hat{R}_{\ell}, \quad (26)$$

where

$$\hat{R}_{\ell} = \|\mathbf{y}_{\ell} - \mathbf{d}_{\ell}\|^2 - N\sigma_w^2 + \sigma_w^2 \frac{2}{\epsilon} \langle \bar{\mathbf{d}}_{\ell} - \mathbf{d}_{\ell}, \boldsymbol{\delta}_{\ell} \rangle, \quad (27)$$

where  $\boldsymbol{\delta}_{\ell}$  are independent random Gaussian vector direction such that  $\boldsymbol{\delta}_{\ell} \sim \mathcal{N}(0, \mathbf{I})$ ,  $\mathbf{d}_{\ell} = \mathbf{H}_{\ell}\mathbf{x}_{\ell}$  and  $\bar{\mathbf{d}}_{\ell} = \mathbf{H}_{\ell}\bar{\mathbf{x}}_{\ell}$  with  $\bar{\mathbf{x}}_{\ell}$  the output of the deconvolution algorithm when  $\bar{\mathbf{y}}_{\ell} := \mathbf{y}_{\ell} + \epsilon\boldsymbol{\delta}_{\ell}$  is used instead of

$\mathbf{y}_{\ell}$  for  $\ell = 1, \dots, L$  i.e. when the observations are slightly perturbed,  $\epsilon > 0$  is a small constant that should be chosen as small as possible as long as it does not rise to numerical instabilities. Note that  $\boldsymbol{\delta}_{\ell}$  does not have to be necessary Gaussian as explained by Avron & Toledo (2011). As mentioned in the beginning of this section, the PSURE-FDMC provides an unbiased estimate of the PMSE:

$$\lim_{\epsilon \rightarrow 0} \mathbb{E} [\hat{R}] = R. \quad (28)$$

The PSURE-FDMC in equation (27) only depends on the measurements, the outputs of the deconvolution algorithm ( $\mathbf{x}_{\ell}$  and  $\bar{\mathbf{x}}_{\ell}$ ,  $\ell = 1, \dots, L$ ), and requires knowing the noise variances  $\sigma_w^2$ . Knowledge of the noise variance is generally assumed in interferometric imaging, as in (Carrillo et al. 2014; Onose et al. 2016b), or it can be estimated in the image domain using as in Garsden et al. (2015) robust statistics such as the median absolute deviation. Most importantly, note how the PSURE-FDMC in equation (27) does not require the knowledge of  $\mathbf{x}_{\ell}^*$ , unlike its PMSE counterpart in equation (25). As a result, it measures the quality of the estimation for given values of the regularization parameters, and the regularization parameters that gives the best estimate (i.e. the lowest PSURE-FDMC) are the best one among the tested values.

### 3.2 Gradient estimation: SUGAR-FDMC

Contrarily to the SURE, the PSURE-FDMC defined in (27) is differentiable in the weak sense w.r.t. the regularization parameters. Noting that  $\mathbf{x}_{\ell}$  and  $\bar{\mathbf{x}}_{\ell}$  are the only terms that depend on the regularization parameters, a straightforward derivation of PSURE-FDMC in equation (27) w.r.t. one of the regularization parameters yields the SUGAR-FDMC:

$$\hat{R}'_{\ell} = 2(\mathbf{H}_{\ell}\mathbf{x}_{\ell} - \mathbf{y}_{\ell})^{\top} \mathbf{H}_{\ell}\mathbf{x}'_{\ell} + \frac{2\sigma_w^2}{\epsilon} \langle \bar{\mathbf{d}}'_{\ell} - \mathbf{d}'_{\ell}, \boldsymbol{\delta}_{\ell} \rangle, \quad (29)$$

where  $\mathbf{d}'_{\ell} = \mathbf{H}_{\ell}\mathbf{x}'_{\ell}$  and  $\bar{\mathbf{d}}'_{\ell} = \mathbf{H}_{\ell}\bar{\mathbf{x}}'_{\ell}$ , and the notation  $(\cdot)'$  refers to the weak derivative w.r.t. one of the regularization parameters, in other terms  $(\cdot)'$  corresponds to either  $\frac{\partial(\cdot)}{\partial\mu}$  or  $\frac{\partial(\cdot)}{\partial\gamma}$ . Interestingly, it was shown by Deledalle et al. (2014) that SUGAR-FDMC defined in equation (29) as the weak gradient of PSURE-FDMC is an unbiased estimate of the weak gradient of the PMSE in equation (25) when  $\epsilon$  tends to zero:

$$\lim_{\epsilon \rightarrow 0} \mathbb{E} [\hat{R}'] = \mathbb{E} [R'], \quad (30)$$

which means that equation (29) gives access to an unbiased estimate of the risk's gradient without prior knowledge of the true sky.

The computation of SUGAR-FDMC requires  $\mathbf{x}'_{\ell}$  and  $\bar{\mathbf{x}}'_{\ell}$ , the weak derivatives of  $\mathbf{x}_{\ell}$  and  $\bar{\mathbf{x}}_{\ell}$ , respectively, for  $\ell = 1, \dots, L$ . Given that  $\mathbf{x}_{\ell}$  and  $\bar{\mathbf{x}}_{\ell}$  are estimated iteratively,  $\mathbf{x}'_{\ell}$  and  $\bar{\mathbf{x}}'_{\ell}$  are also estimated iteratively. This is done by simply deriving each equation in Algorithm 2 w.r.t. the regularization parameter under investigation. The iterative algorithms necessary for estimating  $\frac{\partial\mathbf{x}_{\ell}}{\partial\mu}$  and  $\frac{\partial\bar{\mathbf{x}}_{\ell}}{\partial\mu}$  are described in Algorithms 3 and 4, respectively. The weak derivatives  $\frac{\partial\mathbf{x}_{\ell}}{\partial\gamma}$  and  $\frac{\partial\bar{\mathbf{x}}_{\ell}}{\partial\mu}$  are estimated by replacing  $\mathbf{y}_{\ell}$  by  $\bar{\mathbf{y}}_{\ell}$  in Algorithms 3 and 4. Each iteration in Algorithm 3 is obtained by deriving the corresponding iteration in Algorithm 2 w.r.t.  $\mu$ . When a variable implicitly depends on  $\mu$ , which is the case for  $\mathbf{x}_{\ell}$  and all other intermediate variables, this consists in simply replacing the variable by its derivative (see e.g. equation 36). When the variable depends on the product between  $\mu$  and another variable (that depends on  $\mu$ ), this requires a simple application of the chain rule (see e.g. equation 38). The functions  $\mathcal{U}(\cdot)$  and  $\Pi(\cdot)$ , in equations (40) and (42), respectively, correspond to the weak gradients of

the projection on the positive orthant function and the saturation function defined in equations (6) and (7), respectively. They are defined as follows:

$$\mathcal{U}(u) := \begin{cases} 0 & \text{if } u \leq 0, \\ 1 & \text{if } u > 0, \end{cases} \quad (31)$$

and

$$\Pi(u) := \begin{cases} 1 & \text{if } -1 \leq u \leq 1, \\ 0 & \text{if otherwise,} \end{cases} \quad (32)$$

When the input is a matrix, they are applied component-wise. The same concepts are applied for the derivation of Algorithm 4.

Running  $m$  iterations of Algorithms 3 and 4 with Algorithm 2 starting from iteration  $k$ , provides an iterative procedure for estimating the weak derivative of the reconstructed image  $m$  steps ahead  $k$  w.r.t.  $\mu$  and  $\gamma$ , respectively, and similarly for  $\bar{x}_\ell$  when  $\mathbf{y}_\ell$  is replaced by  $\bar{\mathbf{y}}_\ell$ . Plugging these derivatives in equation (29) provides an asymptotically unbiased estimate of the gradient of the risk  $m$  iteration ahead  $k$ . We will denote the corresponding gradient as  $\nabla \hat{R}_{k:k+m}(\boldsymbol{\theta})$  where  $\boldsymbol{\theta}$  represents either  $\gamma$ ,  $\mu$ , or  $(\gamma, \mu)^\top$ .

### 3.3 Adaptive parameter selection

Ideally, one would like to find the optimal parameters  $\mu^*$  and  $\gamma^*$  such that the PMSE at the convergence of the MUFFIN algorithm is minimized. An alternative solution for estimating the regularization parameters is to use the PSURE-FDMC instead of the PMSE:

$$\boldsymbol{\theta}^* = \operatorname{argmin}_{\boldsymbol{\theta}} \hat{R}. \quad (33)$$

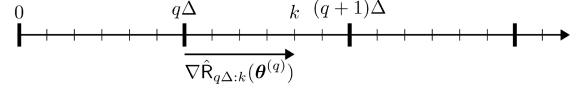
In order to solve problem (33) avoiding grid search techniques and for multiple parameters, the authors of Deledalle et al. (2014) propose to take advantage of the gradient estimate of the PSURE-FDMC defined in the previous section in order to find the optimal parameters using a gradient descent scheme at convergence of the reconstruction algorithm. The gradient descent approach finds the optimal regularization parameters according to the following updates:

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \nabla \hat{R}_{0:n_c}(\boldsymbol{\theta}^{(i)}), \quad (34)$$

where  $\eta$  is a sufficiently small step size,  $\nabla \hat{R}_{0:n_c}$  is computed as explained in the previous section and  $n_c$  corresponds to the number of iterations required to achieve convergence of MUFFIN. Note that  $n_c$  may depend on  $\boldsymbol{\theta}^{(i)}$  but can be set without any additional cost monitoring the successive values of  $k \mapsto \|\nabla \hat{R}_{0:k}(\boldsymbol{\theta}^{(i)})\|$ . This gradient descent approach requires at iteration  $i$  running until convergence MUFFIN with the regularization parameter value  $\boldsymbol{\theta}^{(i)}$  and the derivatives  $\nabla \hat{R}$ . Afterwards, the parameters  $\boldsymbol{\theta}^{(i)}$  are updated to  $\boldsymbol{\theta}^{(i+1)}$ , MUFFIN and the derivatives estimation algorithms should be run again and so on.

In order to reduce the computational cost of this algorithm, we propose an adaptive gradient descent approach that simultaneously updates the estimates of  $\mathbf{x}_\ell^*$  and the regularization parameters. This approach, in the spirit of Giryes et al. (2011), was introduced by Ammanouil, Ferrari & Richard (2018) for general iterative algorithms. More precisely, the regularization parameters are updated every  $\Delta$  iterations of MUFFIN according to a gradient descent direction such that the PSURE-FDMC at the current iteration is reduced (see Fig. 1). At iteration  $k$  of MUFFIN, if  $k = (q+1)\Delta$ ,  $\boldsymbol{\theta}^{(q)}$  is updated using

$$\boldsymbol{\theta}^{(q+1)} = \boldsymbol{\theta}^{(q)} - \eta \nabla \hat{R}_{q\Delta:k}(\boldsymbol{\theta}^{(q)}). \quad (35)$$



**Figure 1.** ADA-PT update rule. Thin vertical lines indicate MUFFIN iterations and thick vertical lines the updates of the regularization parameters.

Hence, rather than running the MUFFIN Algorithm 2 until convergence and then updating the regularization parameters using equation (34), it requires running Algorithms 2–4 only once. We refer to this method as ADA-PT. More precisely, the ADA-PT strategy consists in the iterations described in Algorithm 5. The parameter  $\Delta$  sets the trade-off between a greedy algorithm,  $\Delta = 1$ , which will update the parameters at each iteration and the global algorithm (34),  $\Delta = n_c$ . As pointed by Giryes et al. (2011), one of the problems of the greedy strategy ( $\Delta = 1$ ) is that it will update the parameters in order to reduce the estimated PMSE of the current iteration but can harm the overall PMSE. Thus, instead of choosing to reduce the estimated PMSE at each iteration, ADA-PT reduces the estimated PMSE every  $\Delta > 1$  iterations, typically few tens. It is worthy to note that the value of  $\Delta$  does not affect the complexity of the algorithm.

Finally, it is worth pointing out that although estimating the risk gradient entails the additional operations described in Algorithms 3 and 4, the global memory requirement and computational complexity of ADA-PT’ed MUFFIN are multiplied only by three w.r.t. MUFFIN. Moreover, these additional computations benefit from the same parallel implementation as MUFFIN described in Section 2.3.

Each node  $\ell = 1 \dots L$  computes sequentially:

$$\mathbf{x}'_\ell = \tilde{\mathbf{x}}'_\ell, \quad (36)$$

$$\nabla'_\ell = \mathbf{H}_\ell^\dagger \mathbf{H}_\ell \mathbf{x}'_\ell \quad (37)$$

$$\mathbf{s}'_\ell = \mu \beta_\ell \mathbf{W}_s^\dagger \mathbf{u}'_\ell + \beta_\ell \mathbf{W}_s^\dagger \mathbf{u}_\ell \quad (38)$$

$$\mathbf{m}'_\ell = \mathbf{x}'_\ell - \tau(\nabla'_\ell + \mathbf{s}'_\ell + \mathbf{t}'_\ell) \quad (39)$$

$$\tilde{\mathbf{x}}'_\ell = \mathcal{U}(\mathbf{m}'_\ell) \mathbf{m}'_\ell \quad (40)$$

$$\mathbf{p}'_\ell = \mathbf{u}'_\ell + \sigma \beta_\ell \mathbf{W}_s (2\tilde{\mathbf{x}}_\ell - \mathbf{x}_\ell + \mu(2\tilde{\mathbf{x}}'_\ell - \mathbf{x}'_\ell)) \quad (41)$$

$$\mathbf{u}'_\ell = \Pi(\mathbf{p}'_\ell) \mathbf{p}'_\ell, \quad (42)$$

and sends  $\mathbf{x}'_\ell$  and  $\tilde{\mathbf{x}}'_\ell$  to the master node.

The master computes sequentially, for  $n = 1 \dots N$ :

$$(\tilde{\mathbf{v}}^n)' = (\mathbf{v}^n)' + \sigma \gamma \alpha_n \mathbf{W}_\lambda (2\tilde{\mathbf{x}}^n' - (\mathbf{x}^n)') \quad (43)$$

$$(\mathbf{v}^n)' = \Pi(\tilde{\mathbf{v}}^n)' (\tilde{\mathbf{v}}^n)' \quad (44)$$

$$(\mathbf{t}^n)' = \gamma \alpha_n \mathbf{W}_\lambda^\dagger (\mathbf{v}^n)', \quad (45)$$

and sends the col.  $\ell$  of  $\mathbf{T}'$ , denoted as  $\mathbf{t}'_\ell$ , to node  $\ell$ .

**Algorithm 3:** Parallel update of  $\mathbf{x}'_\ell = \frac{\partial \mathbf{x}_\ell}{\partial \mu}$

## 4 TESTS ON SIMULATED DATA

The reconstruction results reported in this section were obtained with the proposed distributed implementation of MUFFIN. The aim of these simulations is twofold: (1) to demonstrate the efficiency of the distributed implementation and (2) to illustrate the relevance of the ADA-PT strategy for automatically finding the optimal regularization parameters.



Each node  $\ell = 1 \dots L$  computes sequentially:

$$\mathbf{x}'_\ell = \bar{\mathbf{x}}'_\ell, \quad (46)$$

$$\nabla'_\ell = \mathbf{H}_\ell^\dagger \mathbf{H}_\ell \mathbf{x}'_\ell \quad (47)$$

$$\mathbf{s}'_\ell = \mu \beta_\ell \mathbf{W}_s^\dagger \mathbf{u}'_\ell \quad (48)$$

$$\mathbf{m}'_\ell = \mathbf{x}'_\ell - \tau(\nabla'_\ell + \mathbf{s}'_\ell + \mathbf{t}'_\ell) \quad (49)$$

$$\bar{\mathbf{x}}'_\ell = \mathcal{U}(\mathbf{m}'_\ell) \mathbf{m}'_\ell \quad (50)$$

$$\mathbf{p}'_\ell = \mathbf{u}'_\ell + \sigma \mu \beta_\ell \mathbf{W}_s (2\bar{\mathbf{x}}'_\ell - \mathbf{x}'_\ell) \quad (51)$$

$$\mathbf{u}'_\ell = \Pi(\mathbf{p}'_\ell) \mathbf{p}'_\ell, \quad (52)$$

and sends  $\mathbf{x}'_\ell$  and  $\bar{\mathbf{x}}'_\ell$  to the master node.

The master computes sequentially, for  $n = 1 \dots N$ :

$$(\tilde{\mathbf{v}}^n)' = (\mathbf{v}^n)' + \sigma \alpha_n \mathbf{W}_\lambda (2\bar{\mathbf{x}}^n - \mathbf{x}^n + \gamma(2(\tilde{\mathbf{x}}^n)' - (\mathbf{x}^n)')) \quad (53)$$

$$(\mathbf{v}^n)' = \Pi(\tilde{\mathbf{v}}^n)' (\tilde{\mathbf{v}}^n)' \quad (54)$$

$$(\mathbf{t}^n)' = \gamma \alpha_n \mathbf{W}_\lambda^\dagger (\mathbf{v}^n)' + \alpha_n \mathbf{W}_\lambda^\dagger (\mathbf{v}^n), \quad (55)$$

and sends the col.  $\ell$  of  $\mathbf{T}'$ , denoted as  $\mathbf{t}'_\ell$ , to node  $\ell$ .

**Algorithm 4:** Parallel update of  $\mathbf{x}'_\ell = \frac{\partial \mathbf{x}_\ell}{\partial \gamma}$

- 1: Set  $k = 0$  and  $q = 0$ . Initialize  $\theta^{(0)}$ ,  $\mathbf{X}^{(0)}$  and  $\bar{\mathbf{Y}} = \mathbf{Y} + \epsilon \delta$ .
- 2: 1 iteration of alg.~2 using  $\theta^{(q)}$  and  $\mathbf{Y}$ :  $\mathbf{X}^{(k+1)}$
- 3: 1 iteration of alg.~2 using  $\theta^{(q)}$  and  $\bar{\mathbf{Y}}$ :  $\bar{\mathbf{X}}^{(k+1)}$
- 4: 1 iteration of algs.~3 and 4 for  $\mathbf{X}^{(k+1)}$
- 5: 1 iteration of algs.~3 and 4 for  $\bar{\mathbf{X}}^{(k+1)}$
- 6: **if**  $k = (q + 1)\Delta$  **then**
- 7: Compute  $\nabla \hat{R}_{q\Delta:k}(\theta^{(k)})$  using (29).
- 8: Update  $\theta^q$  to  $\theta^{q+1}$  using (35).
- 9:  $q \leftarrow q + 1$
- 10: **end if**
- 11:  $k \leftarrow k + 1$ . Go to 1:

**Algorithm 5:** ADA-PT'ed MUFFIN algorithm

#### 4.1 Data generation description

The data sets used in the following are composed by two simulated, multiwavelength radio sky emissions in which we have introduced complex spatial and spectral features in order to challenge the reconstruction algorithms. The raw data are obtained by convolving these sky emissions with wavelength-dependent PSFs and adding noise. We describe in more details the considered sky emissions and the PSFs below. Note that in all figures the flux scale is arbitrary as we are mostly interested in comparing the relative intensities of reference and reconstructed images/cubes. Point estimation of the signal-to-noise ratio (SNR) and of the PMSE denoted as Weighted MSE (WMSE):

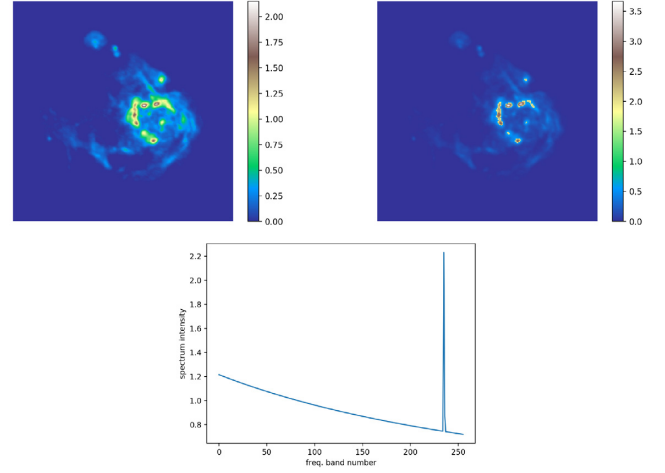
$$\text{SNR}(\mathbf{X}) = 10 \log_{10} \left( \frac{\sum_{\ell=1}^L \|\mathbf{x}_\ell^*\|_2^2}{\sum_{\ell=1}^L \|\mathbf{x}_\ell^* - \mathbf{x}_\ell\|_2^2} \right) \quad (56)$$

$$\text{WMSE}(\mathbf{X}) = \frac{1}{NL} \sum_{\ell=1}^L \|\mathbf{H}_\ell(\mathbf{x}_\ell^* - \mathbf{x}_\ell)\|_2^2 \quad (57)$$

will however be indicated.

##### 4.1.1 Simulated galaxy with H I emission line ('Galaxy-H I')

The first data set is a RA-Dec- $\nu$  cube generated from a well-known image representing a star-forming region in the M31 galaxy (see the top left-hand panel of Fig. 2). For each of the  $256 \times 256$



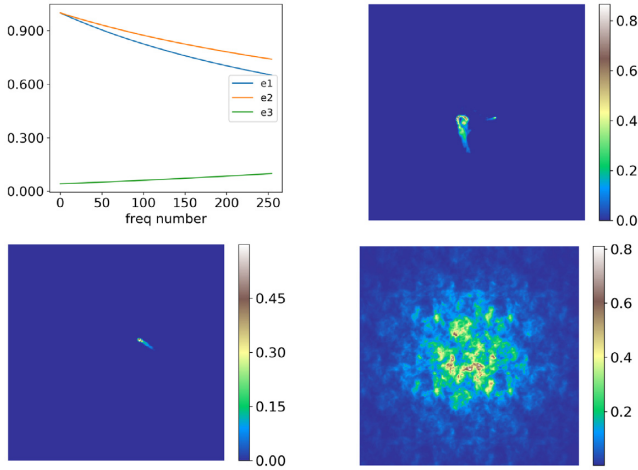
**Figure 2.** ‘Galaxy-H I’ data set. Top left-hand panel: Considered galaxy data image in the first wavelength channel. The H I emission is added to the brightest pixels (in red, above 1.6 in this colour scale). Top right-hand panel: Image in the H I emission line (236th wavelength band). Bottom panel: The spectrum of a bright pixel. The intensity peak caused by the H I emission is visible in the 235th frequency band. The shape of the continuum is a power law.

pixels in this image, we computed a spectrum in 256 adjacent frequency bands of constant bandwidth (2 MHz), starting from 950 MHz. Each spectrum follows a first-order power-law spectrum model depending on a spectral index. The  $256 \times 256$  map of spectral indices was constructed following the procedure detailed by Junklewitz et al. (2015): for each pixel, the spectral index is taken as a linear combination of a homogeneous Gaussian field and the reference sky image.

As an additional spectral feature, we inserted in the spectra of the brightest pixels (see the legend of Fig. 2) a H I line simulating the localized presence of neutral hydrogen gas. The line is centred at 1.42 GHz, corresponding to our 235th channel. We assume here a simple model where the galaxy has a gas consumption time-scale of typically 1 Gyr, a H mass of  $10^9$  solar masses and thus a star formation rate of one solar mass per year. Assuming further a velocity dispersion of  $200 \text{ km s}^{-1}$ , we obtain lines whose flux is about twice the flux of the continuum at 1.42 GHz, a value that may typically fluctuate within one order of magnitude (Serra private communication). Translated into frequency, the velocity dispersion leads to a line width of 1 MHz, with tails that slightly smear out on adjacent channels. In view of these elements, the spectral H I line profile was computed as a continuous Gaussian emission-line profile with a maximum corresponding to twice the local continuum and a full width at half-maximum (FWHM) of 1 MHz. The profile was further convolved by a boxcar function of width 2 MHz, sampled at the locations of the considered 256 frequency bands and added to the continuum. The top right-hand panel of Fig. 2 shows the simulated galaxy in the channel where the H I line has most energy (235th wavelength band). Fig. 2 bottom panel shows the spectrum of a bright pixel from the Galaxy data set where the H I emission was added. We refer to this simulated sky data cube as ‘Galaxy-H I’ for short.

##### 4.1.2 Simulated galaxy cluster with diffuse halo ('Halo')

The second data set is a galaxy cluster radio model based on a refined version of 2D radio simulations considered in our previous



**Figure 3.** ‘Halo’ data set. Top left-hand panel: The end member spectra  $e_k$ ,  $k = 1, 2, 3$ , corresponding to the three sources. Top right-hand, bottom left-hand, and bottom right-hand panels: Abundance maps for the two compact sources and the Halo, respectively.

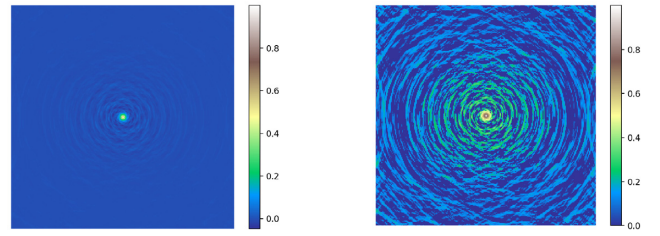
works (Ferrari et al. 2015). This model includes various kinds of radio sources that can be found in ‘radio-loud’ clusters, in particular unresolved point-like objects, bright elongated radio galaxies with complex morphologies and a giant low-surface-brightness source of synchrotron radiation (or ‘radio halo’; see e.g. Ferrari et al. 2008; Feretti et al. 2012) and references therein. In addition to these features, this simulation also properly models the number counts of both the cluster radio source and the foreground/background population. The simulated radio halo has been produced following the approach described in Murgia et al. (2004), while discrete radio sources have been modelled according to what described in Loi et al. (2018)

To generate the 3D data cube, we segmented the simulated 2D radio map described above into three contributions: two giant radio galaxies (with the typical head–tail and narrow-angle-tail morphology of FRI in clusters) and a low-surface-brightness extended halo. Each of the three contribution was given a specific power-law spectrum  $e_k$  ( $k = 1, 2, \text{ or } 3$ ) that remains spatially constant up to a multiplicative factor  $\alpha_n(k)$ . Formally, spectrum  $n$  of the blended data cube is thus generated as:

$$\mathbf{x}^*{}^n = \sum_{k=1}^3 \alpha_n(k) e_k, \quad \forall n = 1, \dots, N.$$

The three considered spectra  $\{e_k\}$ , also called end members, are shown in the left panel of Fig. 3. The three considered segmentation maps (also called abundance maps), which show in each pixel the fraction  $\alpha_n(k)$  of the  $k$ th contribution, are displayed in the three other panels of Fig. 3. Note that the halo is normalized such as its power equals approximately 11 per cent of the power of the final cube.

The purpose of this data set is twofold. First, to evaluate the ability of the proposed method to reconstruct scenes with very high dynamic range and complex structures (the halo in particular), which is one of SKA first objectives. Secondly, to show that, thanks to MUFFIN’s non-parametric approach, the reconstructed cube can be further processed to achieve a specific task such as the deblending of the sources using their spectral properties. For this spectral unmixing case, a two-step procedure (reconstruction and calibration within a major/minor loop followed by the deblending), compared to a joint procedure such as in (Abdulaziz et al. 2016; Jiang et al.



**Figure 4.** PSF in the first channel in linear scale (left) and power scale (0.3, right). The PSF has a central lobe and ringed side-lobes extending at large angular distances.

2017), can be advocated since (i) it avoids the resolution of global non-convex optimization problems and (ii) it gives some space to a human expertise during the unmixing task, e.g. for identifying the end member when complex objects are superposed to a halo.

The final cube, which will be referred to as ‘Halo’ for short, spans a frequency range from 950 MHz to 1.458 GHz, with 256 spectral channels of equal bandwidth.

#### 4.1.3 Point spread function

We simulated the PSFs of SKA1-MID with its 197 dishes using the HI-inator package<sup>1</sup> based on the MEQTTREES software by Noordam & Smirnov (2010). The simulated PSFs correspond to a Fourier coverage produced by a total observation time of 8 h. Fig. 4 shows the PSF at the first wavelength band. The PSF in each channel is characterized by a central lobe and ringed side-lobes at larger angular distances. This cube of PSFs was used to convolve channel by channel the Galaxy-H I and the Halo data sets. Gaussian noise was added to the result to produce the dirty data cubes with an SNR of 20 dB for the dirty Galaxy-H I data set and an SNR of 40 dB for the dirty Halo data set.

## 4.2 Deconvolution examples

### 4.2.1 Implementation details

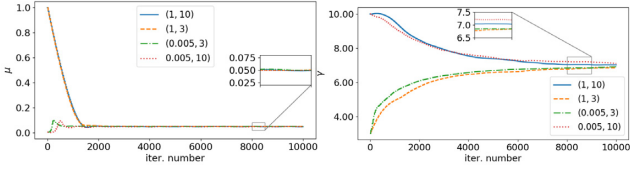
We tested the proposed deconvolution approach, namely ADAPT<sup>ed</sup> MUFFIN, on the two data sets described above. In all the simulations,  $\mathbf{W}_s$  and  $\mathbf{W}_\lambda$  were, respectively, set to the union of the first 8 Daubechies wavelet bases and to the DCT. The parameter of the deconvolution algorithm  $\sigma$  was set to  $\sigma = 10$  and the corresponding value of  $\tau$  was calculated as explained in Section 2.2. The parameter  $\Delta$  was set to 10.

We compare the results of the proposed approach with the MS-MFS deconvolution algorithm. We used the implementation developed within the SKA Science Data Processor Consortium available online<sup>2</sup> with the default parameters. In particular, the number of iterations is equal to  $10^3$ , the scales are set to (0, 3, 10, 30) and order of three polynomials are used to fit the spectra. We tested MS-MFS using the default parameters and using other parameters, the default parameters gave the best results.

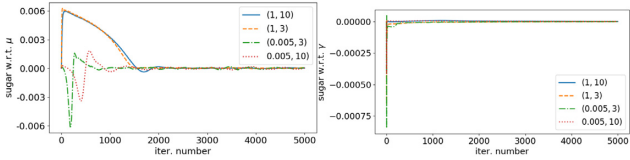
This work was granted access to HPC resources. In particular, the simulations were performed using 80 compute nodes, 1 master node and 79 slave nodes, with 64 Go of RAM per node. This

<sup>1</sup><https://github.com/SpheMakh/HI-Inator>

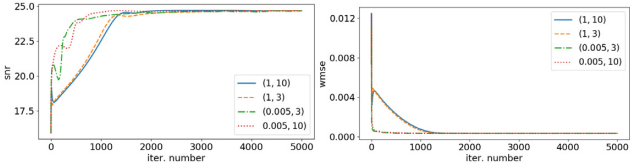
<sup>2</sup><https://github.com/SKA-ScienceDataProcessor/algorithm-reference-library>



**Figure 5.** Evolution of  $\mu$  (left) and  $\gamma$  (right) as a function of the iteration number obtained with four different initialization settings for the regularization parameters  $(\mu, \gamma)$ : (1,10), (1,3), (0,005,3), and (0.005,10).



**Figure 6.** Variation of SUGAR w.r.t.  $\mu$  and of SUGAR w.r.t.  $\gamma$  as a function of the iteration number for the four considered initialization settings.



**Figure 7.** Variation of SNR and WMSE as a function of the iteration number obtained for the four considered initialization settings.

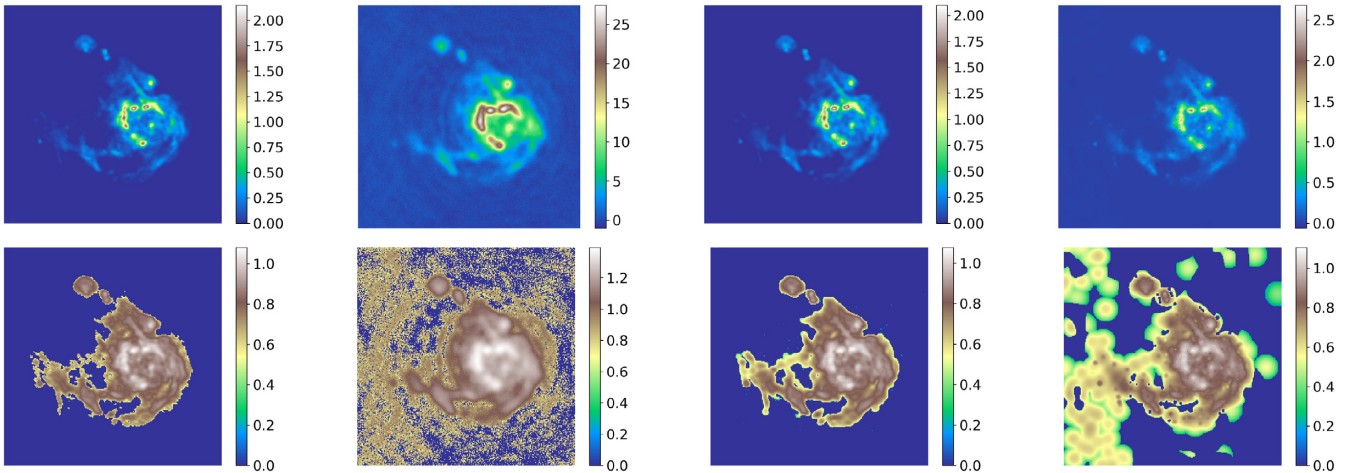
setting corresponds to 4 spectral bands at most per slave node. The algorithm is written in PYTHON, and the Message Passing Interface (MPI) is used to deal with the parallel computing architecture. The computing time to process a  $256 \times 256 \times 256$  cube is 3 s per iteration for MUFFIN and 15 s per iteration for the full ADP-PT’ed code. The code is available at <https://github.com/andferrari/MUFFIN>.

The most computationally intensive steps of MUFFIN are the  $L = 256$  spatial wavelets decompositions and recompositions that are performed at each iteration. Nevertheless, these steps are performed in parallel (see equations 16 and 19 in Algorithm 2). On the other hand, MS-MFS requires 12 scale decompositions (four scales and order of three polynomials) that are performed at the beginning of the algorithm, i.e. before the first iteration. Memory wise, the amount of memory required by MUFFIN is to store the matrices  $U$ ,  $X$ ,  $\tilde{X}$ , and  $V$ . As a result, the global memory footprint of MUFFIN is mainly proportional to the total size of these matrices that is equal to  $11LN$ . Note that the memory load is spread across the  $L$  computing nodes. On the other hand, the MS-MFS memory usage is  $12N$ . The memory requirements for MUFFIN are doubled for ADA-PT’ed MUFFIN.

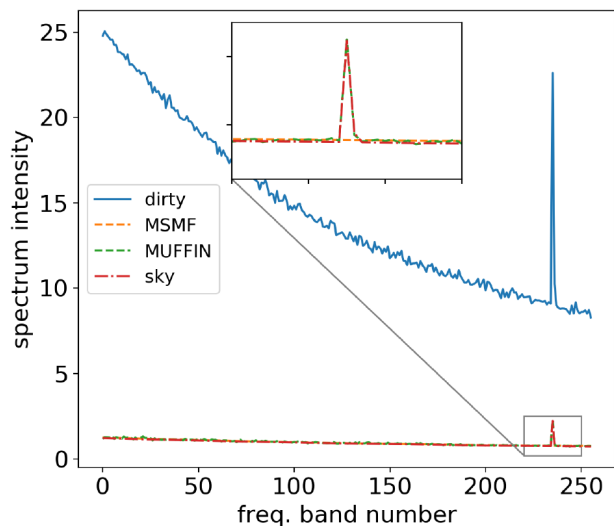
4.2.2 *Galaxy- $H_1$  data set*

The spatial and spectral regularization parameters  $(\mu, \gamma)$  were initialized to four different couples of values: (1,10), (1,3), (0,005,3), and (0.005,10). The ADA-PT’ed MUFFIN algorithm was ran for  $10^4$  iterations and Fig. 5 shows the evolution of each regularization parameters for the four different initialization settings as a function of the iterations. In all cases, the regularization parameter  $\mu$  converges to 0.05 and the regularization parameter  $\gamma$  converges to 7 (on average). Note that the four initialization settings cover the four distinct scenarios where both parameter values are greater than the optimal ones, less than the optimal ones, and where one is greater and one is less than the optimal value.

Fig. 6 shows the evolution of  $k \rightarrow \nabla \hat{R}_{q,\Delta;k}(\theta^{(k)})$ , i.e. of SUGAR w.r.t.  $\mu$  and of SUGAR w.r.t.  $\gamma$  as a function of the iteration number for the four considered initializations values. Note how both SUGAR curves converge to zero, which justifies the regularization parameters convergence. Fig. 7 shows the variation of SNR and WMSE as a function of the iteration number for the four considered initialization cases. The SNR converges to a value of 24 dB and the WMSE converges to  $3.4 \times 10^{-4}$ . Note that the SNR and WMSE curves are only shown for the first 5000 iterations, as they remain stable after the first 2000 iterations.

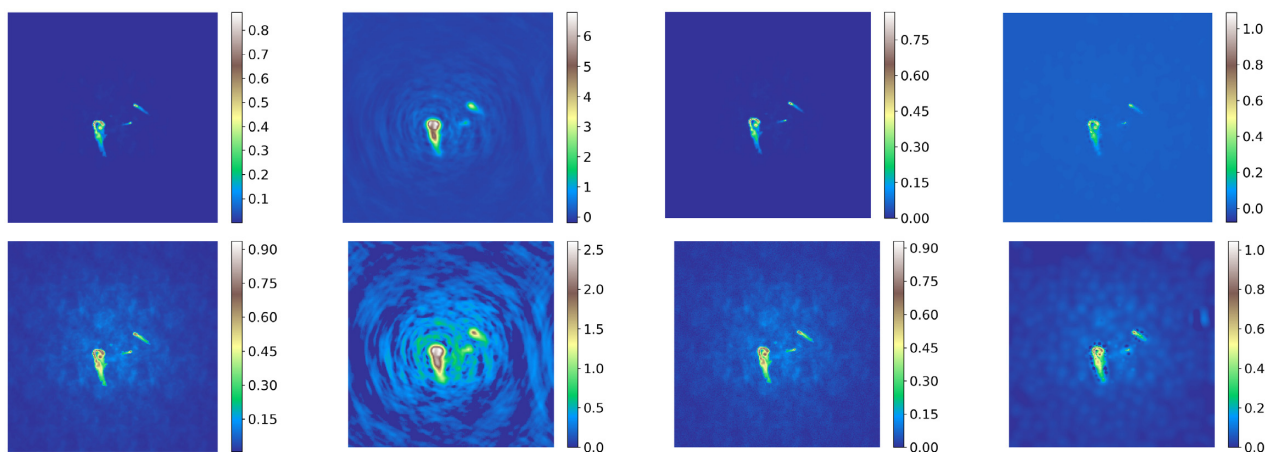


**Figure 8.** First column: Galaxy- $H_1$  reference images. Second column: Dirty images. Third column: Reconstructed images using MUFFIN. Fourth column: Reconstructed images using MS-MFS. All images are shown at the first frequency band. First row shows the images intensities at linear scale, while the second row shows the images intensities to the power 0.1 in order to better highlight the small intensities.



**Figure 9.** Reference, dirty and reconstructed spectra for the bright pixel characterized by an H I emission shown in Fig. 2.

We compare in Fig. 8 the proposed approach with Rau & Cornwell (2011) MS-MFS. The third and fourth columns in the first row of Fig. 8 show that the deconvolved images obtained using MUFFIN and MS-MFS are both very close to the sky image and ‘clean’ when compared to the dirty images shown in the second column. The second row of Fig. 8 which highlights the low intensities, reveals that MUFFIN was able to better reconstruct the image while some smooth artefacts are seen in the MS-MFS image. Note that the SNR of the reconstructed image cube is 24.6 dB for MUFFIN and 18.8 for MS-MFS. Finally, Fig. 9 shows the reconstruction results for the spectrum of the bright pixel in the image characterized by an H I emission and shown in Fig. 2. It can be seen that both MUFFIN and MS-MFS are able to estimate and denoise the continuum part of the spectrum. Turning to the inset panel (zoom on the H I emission line), we see that the line is logically absent in the MS-MFS spectrum, because this algorithm is aimed at reconstructing a power-law spectrum. In contrast, MUFFIN is able to preserve the H I emission in the reconstructed spectra.



**Figure 10.** First column: Halo sky images. Second column: Dirty images. Third column: Halo-reconstructed image using MUFFIN. Fourth column: Halo-reconstructed images using MS-MFS. All images are shown at the first frequency band. First row shows the images intensities at linear scale, while the second row shows the images intensities to the power 0.5 in order to better highlight the small intensities.

#### 4.2.3 Halo data set

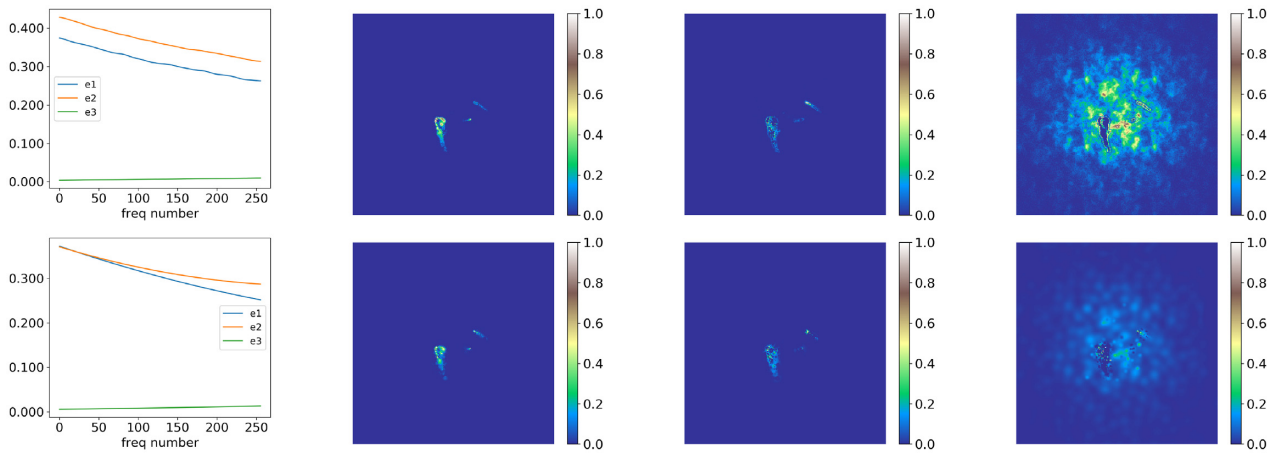
The proposed ADA-PT strategy was run on the Halo data set, for different initialization settings and for  $10^4$  iterations. The regularization parameter  $\mu$  and  $\gamma$  converge, respectively, to  $7 \times 10^{-6}$  and to  $2.6 \times 10^{-2}$ . The SNR converges to a value of 31 dB and the WMSE converges to  $1.6 \times 10^{-7}$ .

Fig. 10 compares the performance of the proposed approach with MS-MFS. The third (MUFFIN) and fourth (MS-MFS) columns lead to the same conclusion as for the Galaxy-H I data set. Note that the SNR of the reconstructed image cube is 31.44 dB for MUFFIN and 14.06 for MS-MFS.

Finally, Fig. 11 shows the source separation result. We chose three spectra belonging to brightest pixels: one from the faint halo region, one from the largest galaxy, and one from the smallest galaxy. The spectra of these pixels are shown in the first column of Fig. 11. We then solved a positively constrained least-squares problem in order to find the contribution of each spectrum for each pixel of the reconstructed data cubes. Columns 2–4 show the recovered abundance maps (to be compared to the maps shown in Fig. 3): the first row corresponds to MUFFIN and the second row to MS-MFS. It can be seen that in the case of MUFFIN, the two galaxies are better separated than in the MS-MFS case. As for the Halo, it is well separated in both cases, but for MS-MFS it presents larger artefacts and differences with respect to the reference.

## 5 CONCLUSIONS

In this work, we have developed a spatial-spectral deconvolution method for large multispectral data named MUFFIN. For this algorithm, we designed a specific parallel implementation and an original procedure for automatically self-tuning the regularization parameters. We demonstrated the efficiency of the proposed deconvolution method on simulated but realistic astronomical multispectral images. The parallel strategy and the SURE-based automatic parameter selection were specifically adapted to the spatial-spectral reconstruction strategy of the MUFFIN algorithm. However, the tools used in this work should provide a foundation for developing analogous parallel implementations and automated parameter selection procedures in the case of different deconvolution problems.



**Figure 11.** Estimated abundance maps. First column shows the selected spectra from the reconstructed image; the second, third, and fourth columns show the abundance maps obtained by solving the positively constrained LS problem. The first and second rows correspond to the results obtained with MUFFIN and MS-MFS, respectively.

## ACKNOWLEDGEMENTS

This work was supported by the Agence Nationale pour la Recherche, France (MAGELLAN project, ANR-14-CE23-0004-01). The numerical simulations in this work were performed at the Centre for Intensive Computing ‘Mésocentre sigamm’ hosted by the Observatoire de la Côte d’Azur. The authors would like to thank Alain Miniussi for his help on the cluster and P. Serra from Osservatorio Astronomico di Cagliari for his help on the model of the H I line. The authors are very grateful to the referee, Stephen Ord, for his insightful suggestions and helpful comments to improve the quality of the original manuscript.

## REFERENCES

- Abdulaziz A., Dabbech A., Onose A., Wiaux Y., 2016, Signal Processing Conference (EUSIPCO), 2016 24th European. IEEE, Budapest, Hungary, p. 388
- Ammanouil R., Ferrari A., Flamary R., Ferrari C., Mary D., 2017, European Signal Processing Conference (EUSIPCO). IEEE, Kos, Greece
- Ammanouil R., Ferrari A., Richard C., 2018, IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Calgary, Canada
- Avron H., Toledo S., 2011, JACM, 58, 8
- Bajkova A. T., Pushkarev A. B., 2011, MNRAS, 417, 434
- Beck A., Teboulle M., 2009, IEEE Transactions on Image Processing, 18, IEEE
- Bhatnagar S., Cornwell T. J., Golap K., Uson J. M., 2008, A&A, 487, 419
- Boyd S., Parikh N., Chu E., Peleato B., Eckstein J., 2011, Found. Trends Mach. Learn., 3, 1
- Candes E. J., Wakin M. B., 2008, IEEE Signal Process. Mag., 25, 21
- Candès E. J., Wakin M. B., Boyd S. P., 2008, J. Fourier Anal. Appl., 14, 877
- Carrillo R. E., McEwen J. D., Wiaux Y., 2014, MNRAS, 439, 3591
- Chambolle A., Pock T., 2010, J. Math. Imaging Vision, 40, 120
- Combettes P., Pesquet J.-C. P., 2012, Set-Valued and Variational Analysis, 20, 307
- Combettes P. L., Pesquet J.-C., 2011, Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer, New York, NY, p. 185
- Condat L., 2014, IEEE Signal Process. Lett., 21, 985
- Conway J. E., Cornwell T. J., Wilkinson P. N., 1990, MNRAS, 246, 490
- Cornwell T. J., 2008, ISTSP, 2, 793
- Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, A&A, 576, A7
- Dabbech A., Onose A., Abdulaziz A., Perley R. A., Smirnov O. M., Wiaux Y., 2018, MNRAS, 476, 2853
- Deledalle C.-A., Vaiter S., Fadili J., Peyré G., 2014, SIAM J. Imaging Sci., 7, 2448
- Dewdney P., Hall P., Schilizzi R., Lazio T., Joseph L. W., 2009, Proc. IEEE, 97, 1482
- Dewdney P., Turner W., Millenaar R., McCool R., Lazio J., Cornwell T., 2013, Technical Report 1, SKA1 System Baseline Design. SKA Office
- Donoho D. L., 2006, IEEE Trans. Inf. Theory, 52, 1289
- Elad M., Milanfar P., Rubinstein R., 2007, Inverse Probl., 23, 947
- Eldar Y. C., 2009, IEEE Trans. Signal Process., 57, 471
- Feretti L., Giovannini G., Govoni F., Murgia M., 2012, A&AR, 20, 54
- Ferrari C., Govoni F., Schindler S., Bykov A. M., Rephaeli Y., 2008, Space Sci. Rev., 134, 93
- Ferrari C. et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14). Proceedings of Science, Giardini Naxos, Italy, p. 75
- Garsden H. et al., 2015, A&A, 575, A90
- Giryres R., Elad M., Eldar Y. C., 2011, Appl. Comput. Harmon. Anal., 30, 407
- Golub G. H., Heath M., Wahba G., 1979, Technometrics, 21, 215
- Hansen P. C., O’Leary D. P., 1993, SIAM J. Sci. Comput., 14, 1487
- Högbom J. A., 1974, A&AS, 15, 417
- Jiang M., Bobin J., Starck J.-L., 2017, SIAM J. Imaging Sci., 10, 1997
- Jongering R., Wijnholds S., Nijboer R., Corporaal H., 2014, Computer, 47, 48
- Junklewitz H., Bell M. R., Enßlin T., 2015, A&A, 581, A59
- Karl W. C., 2005, Handbook of Image Video Processing. Academic Press, p. 141
- Kartik S. V., Carrillo R., Thiran J.-P., Wiaux Y., 2017, MNRAS, 468, 2382
- Kellermann K., 1974, in Verschuur G., Kellermann K., eds, Galactic and Extragalactic Radio Astronomy, Chapt. Radio Galaxies and Quasars. Springer-Verlag
- Kogan L., Greisen E. W., 2009, Technical Report, Faceted imaging in AIPS. AIPS Memos
- Kraus J. D., 1986, Radio Astronomy. Cygnus-Quasar, Powell, OH
- Lanusse F., Starck J.-L., Woiselle A., Fadili M., 2014, Advances in Imaging and Electron Physics. Elsevier, p. 99
- Loi F., Murgia M., Govoni F., Vacca V., Prandoni I., Li H., Feretti L., Giovannini G., 2018, Galaxies, 6, 133
- McEwen J., Wiaux Y., 2011, 2011 18th IEEE International Conference on Image Processing (ICIP). IEEE, Brussels, Belgium, p. 1313

- Meillier C., Ammanouil R., Ferrari A., Bianchi P., 2018, *Signal Process. Image Commun.*, 67, 149
- Morozov V. A., 1966, in *Soviet Math. Dokl.* p. 414
- Murgia M., Govoni F., Feretti L., Giovannini G., Dallacasa D., Fanti R., Taylor G. B., Dolag K., 2004, *A&A*, 424, 429
- Noordam J. E., Smirnov O. M., 2010, *A&A*, 524, A61
- Onose A., Carrillo R. E., McEwen J. D., Wiaux Y., 2016a, *Signal Processing Conference (EUSIPCO), 2016 24th European. IEEE, Budapest, Hungary*, p. 1448
- Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016b, *MNRAS*, 462, 4314
- Paris S., Suleiman R., Mary D., Ferrari A., 2013, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Vancouver, Canada
- Ramani S., Blu T., Unser M., 2008, IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Las Vegas, USA, p. 905
- Ramani S., Liu Z., Rosen J., Nielsen J.-F., Fessler J. A., 2012, *IEEE Trans. Image Process.*, 21, 3659
- Rau U., Cornwell T. J., 2011, *A&A*, 532, A71
- Rau U., Bhatnagar S., Voronkov M. A., Cornwell T. J., 2009, *Proc. IEEE*, 97, 1472
- Regińska T., 1996, *SIAM J. Sci. Comput.*, 17, 740
- Renard S., Thiébaud É., Malbet F., 2011, *A&A*, 533, A64
- Scaife A. M. M., Heald G. H., 2012, *MNRAS*, 423, L30
- Setzer S., Steidl G., Teuber T., 2010, *J. Visual Commun. Image Represent.*, 21, 193
- Starck J.-L., Bobin J., 2010, *Proc. IEEE*, 98, 1021
- Stein C. M., 1981, *The Annals of Statistics. Institute of Mathematical Statistics*, p. 1135
- Tasse C., van der Tol S., van Zwieten J., van Diepen G., Bhatnagar S., 2013, *A&A*, 553, A105
- Tasse C. et al., 2018, *A&A*, 611, A87
- van Weeren R. J. et al., 2016, *ApJS*, 223, 2
- Vũ B. C., 2011, *Adv. Comput. Math.*, 38, 667
- Wenger S., Magnor M., 2014, Technical Report, A Sparse Reconstruction Algorithm for Multi-Frequency Radio Images. Computer Graphics Lab, TU Braunschweig
- Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vanderghenst P., 2009, *MNRAS*, 395, 1733
- Wijnholds S., van der Tol S., Nijboer R., van der Veen A.-J., 2010, *IEEE Signal Process. Mag.*, 27, 30

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.