A Low-Rank and Joint-Sparsity Model for Hyper-Spectral Radio-Interferometric Imaging

Abdullah Abdulaziz, Arwa Dabbech, Alexandru Onose and Yves Wiaux

Institute of Sensors, Signals and Systems, Heriot-Watt University,
Edinburgh EH14 4AS, United Kingdom

***Abstract*—With the advent of the next-generation radio-interferometric telescopes, like the Square Kilometre Array, novel signal processing methods are needed to provide the expected imaging resolution and sensitivity from extreme amounts of hyper-spectral data. In this context, we propose a generic non-parametric low-rank and joint-sparsity image model for the regularisation of the associated wide-band inverse problem. We pose a convex optimisation problem and propose the use of an efficient algorithmic solver. The proposed optimisation task requires only one tuning parameter, namely the relative weight between the low-rank and joint-sparsity constraints. Our preliminary simulations suggest superior performance of the model with respect to separate single band imaging, as well as to other recently promoted non-parametric wide-band models leveraging convex optimisation.**

***Index Terms*—hyper-spectral image processing, radio-interferometry**

I. Introduction

The next-generation of radio-interferometric (RI) telescopes, like the Square Kilometre Array, will be able to survey the radio sky at an unprecedented wide range of frequencies. New regimes of radio emission will be probed improving our knowledge in cosmology and astrophysics. To take advantage of such powerful instruments, revolutionary developments in wide-band imaging techniques are needed both in terms of imaging quality and scalability of the underlying algorithmic structures to handle large data.

Radio wide-band synthesis has been exploited for decades to achieve better Fourier sampling of the sky brightness on narrow bandwidths. This assumes a flat spectral behaviour and overcomes the need of increasing the number of antennas. When this assumption does not hold, particularly for wide-band data, each band is imaged separately without using any inter-band information.

In the context of wide-band imaging, two types of approaches have been proposed recently for the joint recovery of the hyper-spectral image cube. The first family is a parametric approach, where the imaging problem reduces to the joint estimation of the model parameters and the sky surface brightness at a reference frequency. The spectral behaviour is modelled as a low-order polynomial expansion in the logarithm of the frequency, the coefficients of which have to be mapped. The first order coefficient is known as spectral index [1]. The standard first order expansion in the field simply corresponds to a power law model for the spectral signature. The multi-frequency version of the standard RI imaging algorithm CLEAN builds on this parametric approach. Bayesian inference techniques have also been proposed in [2] to map the surface brightness and the spectral index. These methods have shown to be relevant for radio emissions with continuous spectra such as the synchrotron emission. They remain non-generic, especially given the new regimes to be unravelled by the new telescopes, with expected complex spectral behaviour. The second type of methods exploit the full hyper-spectral data cube with no explicit parametric model of the spectral behaviour. The hyper-spectral imaging problem is defined as a convex optimisation problem including different regularisation priors. First applications of compressive sensing (CS) techniques imposing spatial and spectral sparsity priors on the hyper-spectral cube in wavelet dictionaries and taking into account the smooth spectral variation of the radio emission have shown promising results in recovering wide-band RI data [3], [4]. These non-parametric approaches offer the important advantage of being more generic.

The work herein fits into the second category. We investigate a low-rank and joint-sparsity priors for wide-band RI imaging in the context of CS. The approach was shown to produce good reconstruction performance on both simulated data and satellite observation of the Earth [5], [6] which motivates the present study for RI imaging. We also propose the use of an efficient solver for the resulting convex optimisation minimisation task.

The contents of the paper are presented as follows. In Section II we present the wide-band RI problem and review some of the existing image reconstruction methods. Section III details the low-rank and joint-sparsity convex optimisation problem that we propose for wide-band RI. A sketch of the algorithmic solver is presented in Section IV. Simulations and results are presented in Section V.

II. wide-band imaging in radio-interferometry

Aperture synthesis in radio-interferometry consists of combining the electromagnetic signal coming from distant sources in the sky and sensed by a collection of antennas,

This work was supported by EPSRC, grants EP/M011089/1 and EP/M008843/1

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

hence simulating a single dish telescope with a diameter corresponding to the largest distance separating a pair of antennas. Under the assumptions of a small field of view and a planar array with respect to the direction of observation, for a monochromatic radio emission at a frequency ν, the radio measurements $y(\cdot,\nu)$ relate to the sky surface brightness $x(\cdot,\nu)$ as follows [7]:

$$\mathrm{y}(\mathbf{u},\nu) = S(\mathbf{u},\nu) \int D(\mathbf{l},\mathbf{u},\nu)\; x(\mathbf{l},\nu)\; e^{-2i\pi\mathbf{u}\cdot\mathbf{l}}\; \mathrm{d}\mathbf{l}^2, \quad (1)$$

where $\mathbf{u} = (u,v)$ are the coordinates of the vectorial distance separating each antenna pair in units of the wavelength in the plane perpendicular to the direction of the observation and $\mathbf{l} = (l,m)$ are the coordinates in a parallel plane of the observed region of the radio sky. $S(\mathbf{u},\nu)$ is the sampling function that is equal to 1 for a measured $\mathbf{u}$ point and 0 otherwise, and is defined by the array configuration and the observation time. The collection of the measured (u,v) points is usually termed by the uv-coverage. $D(\mathbf{l},\mathbf{u},\nu)$ are the so-called direction dependent effects which can be of geometrical nature representing the diffraction of the radio emission due to the non-coplanar array or perturbations introduced by the receivers and the propagation medium. These effects are usually corrected for during calibration and will be ignored in the present paper. The radio measurements $y(\cdot,\nu)$ at a given frequency ν reduce to Fourier samples of the sky surface brightness $x(\cdot,\nu)$.

Considering L frequency bands and sketching the intensity images at each band $\nu_{l=1,..L}$ as vectors, the discrete version of the measurement model associated with the ill-posed inverse problem reads:

$$\mathbf{y}_l = \mathbf{\Phi}_l \mathbf{x}_l + \mathbf{n}_l, \quad (2)$$

where $\mathbf{x}_l \in \mathbb{R}_+^N$ are the intensity model images, $\mathbf{y}_l \in \mathbb{C}^M$ are the complex measurements, and $\mathbf{n}_l \in \mathbb{C}^M$ are additive white Gaussian noise on each band ν_l. $\mathbf{\Phi}_l$ is the measurement operator at the frequency ν_l and is given by $\mathbf{\Phi}_l = \mathbf{G}_l\mathbf{F}\mathbf{Z}$, where $\mathbf{G}_l \in \mathbb{C}^{M\times N}$ contains on the rows convolutional kernels to model the non-uniform sampling in the Fourier domain of the measurements, and $\mathbf{Z}$ is an oversampling operator (a factor of 2 is used herein), also taking into account the possible imperfections due to the non-uniform re-sampling from the discrete Fourier coefficients provided by $\mathbf{F}$. The hyper-spectral data cube is defined in a matrix form as $\mathbf{Y} = (\mathbf{y}_1,..,\mathbf{y}_L) \in \mathbb{C}^{M\times L}$, the hyper-spectral image cube as $\mathbf{X} = (\mathbf{x}_1,..,\mathbf{x}_L) \in \mathbb{R}_+^{N\times L}$, and the additive noise as $\mathbf{N} = (\mathbf{n}_1,..,\mathbf{n}_L) \in \mathbb{C}^{M\times L}$. The linear operator $\mathbf{\Phi}$, for all the bands, is defined such that $\mathbf{\Phi}(\mathbf{X}) = ([\mathbf{\Phi}_l\mathbf{x}_l]_{l=1:L})$. Under these notations, the hyper-spectral data model reduces to:

$$\mathbf{Y} = \mathbf{\Phi}(\mathbf{X}) + \mathbf{N}. \quad (3)$$

In the context of convex optimisation for hyper-spectral imaging for RI, to the best of our knowledge two methods promoting sparsity have been proposed [3], [4]. In the literature, sparsity is promoted within two frameworks. Synthesis problems solve explicitly for the sparsity coefficients, while analysis problems solve for the signal itself, imposing sparsity in the adequate transform domain. The approaches are equivalent for orthonormal sparsity bases. The analysis problem is usually superior for redundant sparsity dictionaries [8], [9].

In [3], authors propose an unconstrained inverse problem promoting sparsity by synthesis. The method allows for non-smooth features in the spectral domain with ℓ_1 penalisation on such components, a second prior is also adopted promoting joint-sparsity by synthesis of the hyper-spectral cube through an $\ell_{\infty,1}$ norm. Similarly, [10] presents an unconstrained inverse problem promoting sparsity by analysis on both the spectral information using the DCT dictionary $\mathbf{W}$ and the spatial information using the wavelet dictionary $\mathbf{\Psi}$ introduced in the SARA algorithm [9]. This dictionary promotes average sparsity over multiple orthonormal transforms. Finally, in addition to positivity, a prior promoting smoothness of the hyper-spectral cube is adopted through the Frobenius norm of $\mathbf{X}$. The minimisation problem is as follows:

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|^2 + \frac{\mu_1}{2}\|\mathbf{X}\|_F^2 + \mu_2\|\mathbf{\Psi}^\dagger\mathbf{X}\|_1 + \mu_3\|\mathbf{X}\mathbf{W}\|_1 + \iota_{\mathbb{R}_+^{N\times L}}(\mathbf{X}), \quad (4)$$

where $\iota_{\mathbb{R}_+^{N\times L}}$ is the indicator function[1] of the convex set $\mathbb{R}_+^{N\times L}$ enforcing positivity of the hyper-spectral cube. The appearance of multiple arbitrary parameters (μ_1, μ_2, μ_3) representing here the trade off between the different priors, is usually problematic for the final reconstruction quality.

III. Sparsity and low-rank model

We propose a minimisation problem promoting simultaneous a low-rank solution and joint-sparsity on the hyper-spectral cube. This is based on the assumption that hyper-spectral images in RI can be decomposed into few sources ρ, each with a distinct spectral signature. We adopt the linear mixture model $\mathbf{X} = \mathbf{S}\mathbf{H}^\dagger$ proposed by [5], where the columns of the matrix $\mathbf{S} \in \mathbb{C}^{N\times\rho}$ represent the sources present in the sky images and the columns of the matrix $\mathbf{H} \in \mathbb{C}^{\rho\times L}$ are their corresponding spectral signatures. This model is motivated in RI by the understanding that the image results from a limited number of independent physical sources. The model actually implies the low-rankness of $\mathbf{X}$, as the rank is given by ρ. It also implies joint-sparsity over the spectral bands: if none of the sources is active at a given spatial point, a full spectral line of $\mathbf{X}$ will be automatically equal to zero. Solving for $\mathbf{S}$ and $\mathbf{H}$ would explicitly imply a non-linear non-convex problem. Instead, one can easily impose low-rankness by resorting to a convex nuclear norm relaxation prior. Joint-sparsity in some adequate sparsity basis $\mathbf{\Psi}$ can be encapsulated in

[1] The indicator function of a convex set $\mathcal{C}$ is: $\iota_{\mathcal{C}}(\mathbf{Z}) \triangleq \begin{cases} 0 & \mathbf{Z} \in \mathcal{C} \\ +\infty & \mathbf{Z} \notin \mathcal{C}. \end{cases}$

an analysis $\ell_{2,1}$ prior. A similar hyper-spectral imaging problem was proposed in [5]. We therefore define the following analysis convex minimisation problem:

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* + \mu\|\mathbf{\Psi}^\dagger\mathbf{X}\|_{\ell_{2,1}} \text{ s.t. } \begin{cases} \|\mathbf{Y} - \mathbf{\Phi}(\mathbf{X})\|_F \leq \varepsilon, \\ \mathbf{X} \in \mathbb{R}_+^{N\times L}. \end{cases} \quad (5)$$

The notation $\|\mathbf{\Psi}^\dagger\mathbf{X}\|_{\ell_{2,1}}$ stands for the component-wise $\ell_{2,1}$ norm. It promotes not only joint-sparsity in $\mathbf{\Psi}$, but also smoothness of the spectral lines. The dictionary $\mathbf{\Psi}$ is a concatenation of 9 orthogonal basis, the identity basis and the eight first Daubechies wavelet dictionaries. It was proven to be very efficient in promoting sparsity for the narrow-band radio-interferometric image reconstruction problem [9], [11], [12]. The notation $\|\mathbf{X}\|_*$ stands for the nuclear norm and represents the ℓ_1 norm of the vector of its singular values. Note that the problem defined in (5) is a constrained minimisation problem, where the data fidelity term is defined as a bound on the Frobenius norm of the residual $\mathbf{Y} - \mathbf{\Phi}(\mathbf{X})$. The bound ε is derived from the noise statistics which are usually known [9]. We will refer to this as the ℓ_2 ball constraint. This constrained problem approach enables to reduce the number of arbitrary parameters. One such parameter remains in the form of μ, which sets the trade off between the two priors.

The minimisation problem (5) can be redefined as

$$\min_{\mathbf{X}} f(\mathbf{X}) + \mu g_1(\mathbf{\Psi}^\dagger\mathbf{X}) + g_2(\mathbf{X}) + g_3\big(\mathbf{\Phi}(\mathbf{X})\big) \quad (6)$$

with the functions involved, defined as

$$\begin{aligned} &f(\mathbf{Z}) = \iota_{\mathcal{D}}(\mathbf{Z}), \;\; \mathcal{D} = \mathbb{R}_+^{N\times L}, \\ &g_1(\mathbf{Z}) = \|\mathbf{Z}\|_{\ell_{2,1}}, \;\; g_2(\mathbf{Z}) = \|\mathbf{Z}\|_*, \\ &g_3(\mathbf{Z}) = \iota_{\mathcal{B}}(\mathbf{Z}), \;\; \mathcal{B} = \{\mathbf{Z} \in \mathbb{C}^{M\times L} : \|\mathbf{Z} - \mathbf{Y}\|_F \leq \varepsilon\}. \end{aligned} \quad (7)$$

We use the indicator function $\iota_{\mathcal{C}}$ of a convex set $\mathcal{C}$ to impose the same constraints on the solution as in (5). This approach is completely equivalent and allows for the use of specialised convex optimisation solvers.

IV. The primal-dual forward-backward solver

We propose the use of a primal-dual (PD) algorithm using forward-backward iterations [13], [14] as a solver for the minimisation problem (6). The same algorithmic structure was recently proposed in the context of single-band RI imaging [12]. It is able to achieve the full splitting of the terms involved in the minimisation problem and has a highly parallelisable structure. The details are presented in Algorithm 1. We use a rescaled version of PD where, by using the Moreau decomposition, we express the algorithmic steps directly with respect to the proximity operator $\text{prox}_h(\mathbf{Z}) \triangleq \text{argmin}_{\bar{\mathbf{Z}}} h(\bar{\mathbf{Z}}) + \frac{1}{2}\|\mathbf{Z} - \bar{\mathbf{Z}}\|_F^2$ of the different terms f and g_i, $i \in \{1, 2, 3\}$, from (6).

The algorithm performs in parallel forward-backward iterations to update all the dual variables $\mathbf{V}_1$, $\mathbf{V}_2$ and $\mathbf{V}_3$ in steps 4–6. Firstly, a forward gradient-like step is performed. It is followed by the incorporation into the algorithmic structure of the non-smooth functions g_i through the application of their proximity operator. This produces a backward, implicit sub-gradient-like step. The proximity operator of the sparsity prior introduced by g_1 resolves to a soft-thresholding operation, defined for each row k as

$$\Big(\boldsymbol{\mathcal{S}}_\alpha^{\ell_{2,1}}(\mathbf{Z})\Big)_{k,:} \triangleq \begin{cases} \bar{\boldsymbol{z}}\dfrac{\|\bar{\boldsymbol{z}}\|_{\ell_2} - \alpha}{\|\bar{\boldsymbol{z}}\|_{\ell_2}} & \|\bar{\boldsymbol{z}}\|_{\ell_2} > \alpha \\ 0 & \|\bar{\boldsymbol{z}}\|_{\ell_2} \leq \alpha \end{cases} \forall k, \quad (8)$$

where we denote by $\bar{\boldsymbol{z}} = \boldsymbol{z}_{k,:}$ the row k of the input matrix $\mathbf{Z}$. The proximity operator to the nuclear norm from g_2 requires the computation of the singular value decomposition $\mathbf{Z} = \mathbf{U}_1\mathbf{\Sigma}\mathbf{U}_2^\dagger$ and is defined as

$$\boldsymbol{\mathcal{S}}_\alpha^*(\mathbf{Z}) \triangleq \mathbf{U}_1\, \boldsymbol{\mathcal{S}}_\alpha^{\ell_1}(\mathbf{\Sigma})\mathbf{U}_2^\dagger. \quad (9)$$

Since $\mathbf{\Sigma}$ is diagonal, this results in the soft-thresholding of the eigenvalues of the matrix $\mathbf{Z}$ via an operator $\boldsymbol{\mathcal{S}}_\alpha^{\ell_1}$ defined similarly to (8) [9]. Data fidelity is enforced by performing the projection on the ℓ_2 ball of size ε defined by the concatenation of the residuals from all frequency bands. This is expressed in an equivalent formulation as

$$\boldsymbol{\mathcal{P}}_{\mathcal{B}}(\mathbf{Z}) \triangleq \begin{cases} \epsilon\dfrac{\mathbf{Z} - \mathbf{Y}}{\|\mathbf{Z} - \mathbf{Y}\|_F} + \mathbf{Y} & \|\mathbf{Z} - \mathbf{Y}\|_F > \varepsilon \\ \mathbf{Z} & \|\mathbf{Z} - \mathbf{Y}\|_F \leq \varepsilon. \end{cases} \quad (10)$$

After their update, all the dual variables are used in step 8 to update the primal variable, the hyper-spectral images of interest. This is performed through a similar forward-backward step. The updated dual variables are incorporated through the use of the update steps τ and σ_i, $i \in \{1, 2, 3\}$. The proximity operator to the indicator of the positive orthant is defined as the projection

$$\Big(\boldsymbol{\mathcal{P}}_{\mathcal{D}}(\mathbf{Z})\Big)_{k,j} \triangleq \begin{cases} \Re(z_{k,j}) & \Re(z_{k,j}) > 0 \\ 0 & \Re(z_{k,j}) \leq 0 \end{cases} \forall k, j. \quad (11)$$

The convergence of the algorithm is achieved if the update parameters satisfy $\tau\left(\sigma_1\|\mathbf{\Psi}^\dagger\|_{\text{S}}^2 + \sigma_2 + \sigma_3\|\mathbf{\Phi}\|_{\text{S}}^2\right) < 1$ [14]. The configuration parameter μ needs to be positive.

Compared to the other convex optimisation solvers proposed for RI imaging, such as the alternating direction

Algorithm 1 PD with forward-backward iterations.

1: **given** $\mathbf{X}^{(0)}, \tilde{\mathbf{X}}^{(0)}, \mathbf{V}_1^{(0)}, \mathbf{V}_2^{(0)}, \mathbf{V}_3^{(0)}, \mu, \tau, \sigma_1, \sigma_2, \sigma_3$
2: **repeat for** $t = 1, \ldots$
3: **do in parallel**
4: $\mathbf{V}_1^{(t)} = \boldsymbol{V}_1^{(t-1)} + \mathbf{\Psi}^\dagger\tilde{\mathbf{X}}^{(t-1)} - \boldsymbol{\mathcal{S}}_{\mu/\sigma_1}^{\ell_{2,1}}\Big(\mathbf{V}_1^{(t-1)} + \mathbf{\Psi}^\dagger\tilde{\mathbf{X}}^{(t-1)}\Big)$
5: $\mathbf{V}_2^{(t)} = \mathbf{V}_2^{(t-1)} + \tilde{\mathbf{X}}^{(t-1)} - \boldsymbol{\mathcal{S}}_{1/\sigma_2}^*\Big(\mathbf{V}_2^{(t-1)} + \tilde{\mathbf{X}}^{(t-1)}\Big)$
6: $\mathbf{V}_3^{(t)} = \mathbf{V}_3^{(t-1)} + \mathbf{\Phi}\big(\tilde{\mathbf{X}}^{(t-1)}\big) - \boldsymbol{\mathcal{P}}_{\mathcal{B}}\Big(\mathbf{V}_3^{(t-1)} + \mathbf{\Phi}\big(\tilde{\mathbf{X}}^{(t-1)}\big)\Big)$
7: **end**
8: $\mathbf{X}^{(t)} = \boldsymbol{\mathcal{P}}_{\mathcal{C}}\Big(\mathbf{X}^{(t-1)} - \tau\Big(\sigma_1\mathbf{\Psi}\mathbf{V}_1^{(t)} + \sigma_2\mathbf{V}_2^{(t)} + \sigma_3\mathbf{\Phi}^\dagger\big(\mathbf{V}_3^{(t)}\big)\Big)\Big)$
9: $\tilde{\mathbf{X}}^{(t)} = 2\mathbf{X}^{(t)} - \mathbf{X}^{(t-1)}$
10: **until convergence**

method of multipliers (ADMM) [12] and the simultaneous direction method of multipliers (SDMM) [11], the PD method is more flexible and has great parallelisation capabilities with limited overhead. ADMM is limited to only two functions in the minimisation problem. For our minimisation task, it requires sub-iterations and, given the computational cost of the priors, especially of the nuclear norm, such approach is unfeasible. SDMM suffers from the need to compute the solution to a linear system of equations of the same size as $\mathbf{X}$ which, for hyper-spectral imaging, can easily become an important bottleneck.

The PD structure, with the full splitting of the operators and functions, does not have these drawbacks. It can scale to any number of functions and more parallelism may be exploited. Redefining the minimisation problem (6) with one ℓ_2 ball per band will take full advantage of this functionality. Randomised updates are also supported such that only some of the terms defining the minimisation problem are activated per iteration. This lowers the infrastructure requirements and makes the algorithmic structure extremely scalable, at the cost of increasing the number of iterations for convergence. The heaviest prior, the nuclear norm, is a heavy computational burden for any solver. It involves the estimation of the SVD through an iterative QR factorisation. The use of randomisation, which the PD algorithmic framework fully supports, can partially alleviate this. Performing an inexact SVD could also be feasible. We leave all these approaches for a further study.

V. Simulations and results

In this section, results are given for simulations using a radio emission map from an HII region in the M31 galaxy. The image is of size $N = 256 \times 256$ pixels and is considered as the original sky image $\mathbf{x}_0$ at the reference frequency $\nu_0 = 1.4$GHz (Fig. 2, second row; left panel). The hyper-spectral cube is simulated using the basic power-law model $\mathbf{x}_\nu = \mathbf{x}_0 \, (\frac{\nu}{\nu_0})^{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha}$ is the vector of spectral indices of size N [1]. In order to ensure a spatial correlation in the spectral index map, the latter is generated in an ad hoc manner, similarly to [2], [4], that is a linear combination of the reference sky image smoothed with a Gaussian kernel of size 3×3 at full width half maximum, and a random Gaussian field. The hyper-spectral data cube is simulated using a non-uniform random sampling in Fourier with a Gaussian density profile at the reference frequency ν_0. To mimic radio-interferometric uv-coverages, holes are introduced in the sampling function through an inverse Gaussian profile, placing the missing Fourier content predominantly in the high spatial frequencies (Fig. 2, top row; left panel). For each band ν, its corresponding uv-coverage is obtained by scaling the reference uv-coverage with ν. This is intrinsic to radio-interferometric wide-band acquisition. The coordinates $\mathbf{u}$ defining the uv-coverages correspond to the vectorial distance in units of the wavelength. Wide-band data cubes are generated within the range $[1.4, 2.8]$ GHz, with uniformly sampled

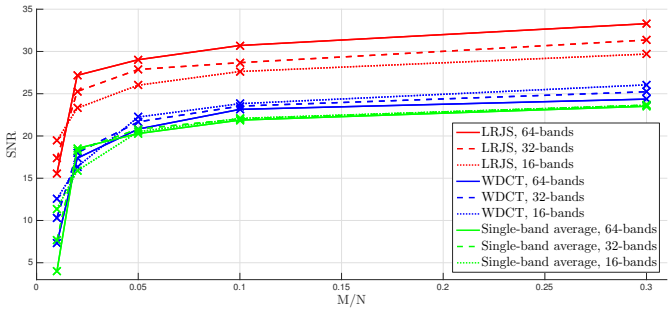

Fig. 1: Reconstruction results for the different methods on the data cubes with $L \in \{16, 32, 64\}$: LRJS (red), WDCT (blue) and single-band recovery (green). The average SNR of the estimated hyper-spectral cubes (y-axis) is plotted as a function of the sampling rate (x-axis). Each point corresponds to the mean value of 10 noise realisations on the data cube.

bands. Tests are carried out on three data cubes with a total number of bands $L = \{16, 32, 64\}$, and varying the sampling rate, that is the ratio between the total number of measurements and the size of the cube, ranging from 0.01 to 1.

The adopted metric to assess the reconstruction quality of the different methods is the signal to noise ratio. For a single band ν_l, it is defined by $\text{SNR}_l = 20 \log_{10} \left(\|\mathbf{X}_l\|_2 / \|\mathbf{X}_l - \widehat{\mathbf{X}}_l\|_2 \right)$, and for the hyper-spectral cube, the average SNR is adopted; $\text{SNR} = (1/L) \sum_l \text{SNR}_l$. The performance of the proposed algorithm, denoted by LRJS, is compared with that of the approach proposed in [4], denoted by WDCT, and to the single reconstruction of each band. The two benchmark methods are also implemented using the PD algorithm [13]. For the reconstruction, the optimal value of the regularisation parameter in LRJS is found to be $\mu = 10^{-4}$. This means more weight is given to the nuclear norm prior with respect to the $\ell_{2,1}$. The optimal values for the three parameters of WDCT are found to be $\mu_1 = 10^{-3}$, $\mu_2 = 10^2$ and $\mu_3 = 2 \cdot 10^2$, giving more emphasis to the ℓ_1 components in coherence with [4].

Reconstruction results of the hyper-spectral cubes are shown in Fig. 1. They clearly demonstrate a significantly higher performance of the proposed approach compared to the benchmark methods, for different sampling rates and number of bands. This suggests that the LRJS regularisation model is significantly superior to both the WDCT regularisation model and the purely spatial model of the single-band approach. Although only sampling rates below 0.3 are plotted, the same behaviour is obtained with larger rates. It is also noticeable that for regimes with sampling rates above 0.01, increasing the number of bands enhances significantly the recovery of the LRJS approach. In fact, a gain of almost 2dB on the average SNR of the cube is achieved when doubling its size, which shows the power of the nuclear norm prior in recovering the low-rank hyper-spectral cube. This is not the case for the benchmark methods. While it is expected for single

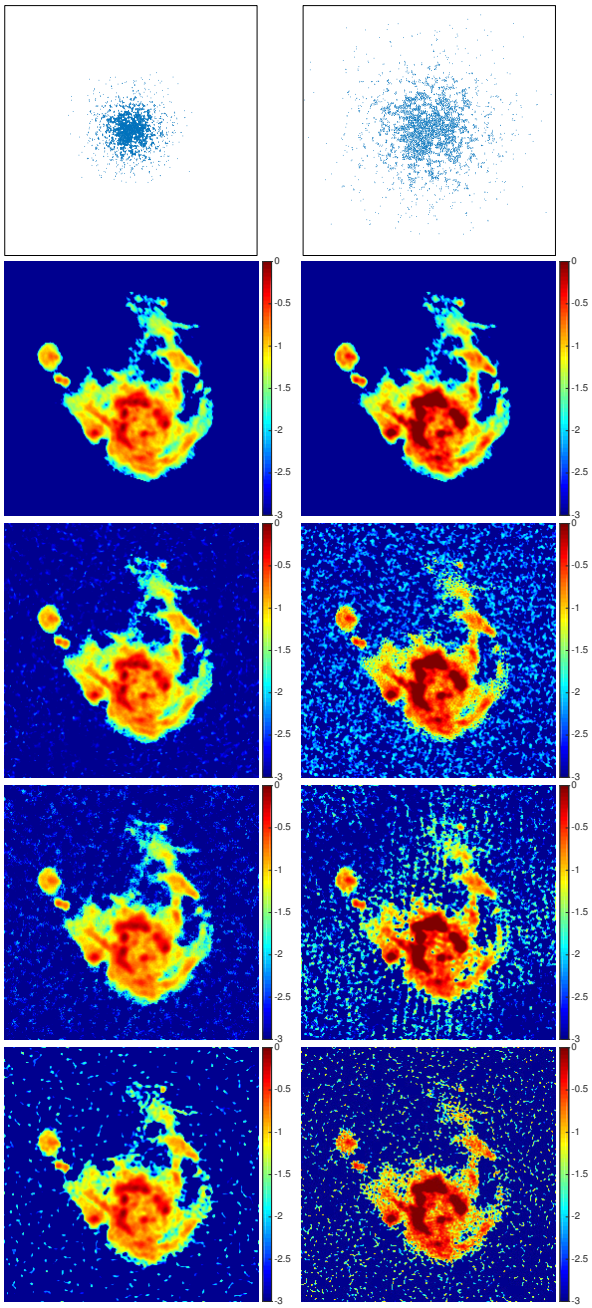

Fig. 2: Simulations and results for the data cube $\mathbf{Y} \in \mathbb{C}^{0.1N \times 64}$, left: first band $\nu_1 = 1.4$GHz, right: last band $\nu_{64} = 2.8$GHz. From top to bottom: uv-coverages, original sky images, reconstructed model images for the LRJS ($\text{SNR}_{\nu_1} = 28.5$dB, $\text{SNR}_{\nu_{64}} = 28.9$dB), WDCT ($\text{SNR}_{\nu_1} = 21.3$dB, $\text{SNR}_{\nu_{64}} = 21.9$dB) and single-band ($\text{SNR}_{\nu_1} = 20.4$dB, $\text{SNR}_{\nu_{64}} = 21.3$dB).

band reconstruction, since no inter-band information is exploited, the WDCT approach presents a small loss in the average SNR when doubling the number of bands.

The visual inspection of the reconstructed images in Fig. 2, for a data cube with 64 bands and a sampling rate equal to $0.1N$, confirms the higher efficiency of the proposed method. It is worth mentioning that for regimes with sampling rate below 0.01, the increase of the number of bands results in worse recovery of the hyper-spectral cube for the three methods, yet surprisingly the proposed approach reaches an average SNR that is almost 7dB higher than the benchmark methods.

VI. Conclusions

We studied a generic non-parametric low-rank and joint-sparsity image model for the regularisation of the ill-posed inverse problem occurring in wide-band RI. Unlike the unconstrained WDCT approach, our LRJS minimisation problem requires only one tuning hyper-parameter. Expressed as a convex optimisation task, the problem was solved using a PD algorithm that shows promise to scale to the large data sizes expected from future radio telescopes. The preliminary simulations suggest superior performance of the LRJS model with respect to separate single band imaging, as well as to WDCT.

References

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

[2] H. Junklewitz, M. R. Bell, and T. Enßlin, "A new approach to multifrequency synthesis in radio interferometry," *AA*, vol. 581, p. A59, Sep. 2015.

[3] S. Wenger and M. Magnor, "A sparse reconstruction algorithm for multi-frequency radio images," 2014.

[4] 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," *ArXiv e-prints*, Apr. 2015.

[5] M. Golbabaee and P. Vandergheynst, "Compressed Sensing of Simultaneous Low-Rank and Joint-Sparse Matrices," *IEEE Tran. Info. Th.*, p. 32, 2012.

[6] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Tran. Geosc. Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, Aug 2014.

[7] A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Synthesis in Radio Astronomy*. Wiley-Interscience, New York, 2001.

[8] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, Jun. 2007.

[9] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging," *MNRAS*, vol. 426, no. 2, pp. 1223–1234, 2012.

[10] A. Ferrari, D. Mary, R. Flamary, and C. Richard, "Distributed image reconstruction for very large arrays in radio astronomy," *Sensor Array Multich. Sig. Proc. Workshop*, vol. 1507.00501, 2014.

[11] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," *MNRAS*, vol. 439, no. 4, pp. 3591–3604, 2014.

[12] 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 e-prints*, Jan. 2016.

[13] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *Sig. Proc. Mag.*, vol. 32, no. 6, pp. 31–54, Nov 2015.

[14] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," *J. Nonlinear Convex Anal.*, vol. 16, no. 12, 2015.