



Heriot-Watt University
Research Gateway

A regularized tri-linear approach for optical interferometric imaging

Citation for published version:

Birdi, J, Repetti, A & Wiaux, Y 2017, 'A regularized tri-linear approach for optical interferometric imaging', *Monthly Notices of the Royal Astronomical Society*, vol. 468, no. 1, pp. 1142–1155.
<https://doi.org/10.1093/mnras/stx415>

Digital Object Identifier (DOI):

[10.1093/mnras/stx415](https://doi.org/10.1093/mnras/stx415)

Link:

[Link to publication record in Heriot-Watt Research Portal](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Monthly Notices of the Royal Astronomical Society

Publisher Rights Statement:

This article has been accepted for publication in Monthly Notices of the Royal Astronomical Society, ©: 2017 Oxford University Press Published by Oxford University Press on behalf of the Royal Astronomical Society. All rights reserved.

General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact open.access@hw.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

A regularized tri-linear approach for optical interferometric imaging

Jasleen Birdi,[★] Audrey Repetti and Yves Wiaux

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

Accepted 2017 February 15. Received 2016 December 22; in original form 2016 September 1

ABSTRACT

In the context of optical interferometry, only undersampled power spectrum and bispectrum data are accessible. It poses an ill-posed inverse problem for image recovery. Recently, a tri-linear model was proposed for monochromatic imaging, leading to an alternated minimization problem. In that work, only a positivity constraint was considered, and the problem was solved by an approximated Gauss–Seidel method. In this paper, we propose to improve the approach on three fundamental aspects. First, we define the estimated image as a solution of a regularized minimization problem, promoting sparsity in a fixed dictionary using either an ℓ_1 or a (re)weighted- ℓ_1 regularization term. Secondly, we solve the resultant non-convex minimization problem using a block-coordinate forward–backward algorithm. This algorithm is able to deal both with smooth and non-smooth functions, and benefits from convergence guarantees even in a non-convex context. Finally, we generalize our model and algorithm to the hyperspectral case, promoting a joint sparsity prior through an $\ell_{2,1}$ regularization term. We present simulation results, both for monochromatic and hyperspectral cases, to validate the proposed approach.

Key words: techniques: image processing – techniques: interferometric.

1 INTRODUCTION

With the advent of astronomical interferometers, it has become possible to image the sky at very high angular resolution. An interferometer basically consists of an array of telescopes such that each pair of telescopes probes a spatial frequency in the Fourier plane (denoted by u – v plane) of the image of interest. Given the limited number of telescopes, incomplete sampling of the u – v plane is obtained. In particular, for radio interferometry, measurements consist of complex visibilities, related to Fourier coefficients of the intensity image of interest (Thompson, Moran & Swenson 2001). In this context, the incomplete Fourier sampling leads to a linear ill-posed inverse problem for image reconstruction, and iterative algorithms need to be designed to solve this problem. Classical reconstruction methods for radio interferometry are mainly based on iterative deconvolution (CLEAN; Högbom 1974), and on maximum entropy methods (MEM) to impose smoothness on the sought image by maximizing the entropy of the image (Cornwell & Evans 1985). More recently, imaging techniques within the framework of compressive sensing have been proposed (Wiaux et al. 2009). These methods rely on finding an image that is sparse in a given dictionary, using convex optimization algorithms (Boyd & Vandenberghe 2004; Combettes & Pesquet 2010).

As compared to the radio interferometers, optical interferometers involve a less number of telescopes, which in turn provides a

sparser u – v coverage. Moreover, atmospheric turbulence at optical wavelengths causes random phase fluctuations leading to cancellation of the visibility values. Indeed, the measurements consist of phase-insensitive observables: power spectrum and bispectrum, resulting into loss of partial phase information (Thiébaud & Giovannelli 2010). This induces non-linearity in the inverse problem for image reconstruction in optical interferometry. Thus, the image recovery methods used in radio interferometry cannot be directly applied, and new methods need to be developed.

Research in this direction has led to the development of various algorithms, based on different approaches. In Thiébaud (2008), the so-called MIRA method has been developed, using a maximum a posteriori (MAP) approach to recover the image, where different types of quadratic regularization can be considered. The author proposed to solve the resulting minimization problem using a limited variable metric algorithm that accounts for parameter bounds (namely, the VMLMB algorithm; Thiébaud 2002). Another technique, proposed by Meimon, Mugnier & Le Besnerais (2005), namely WISARD, makes use of a self-calibration approach to solve for missing phase information, using smooth regularizations. The so-called BSMEM method, proposed in Buscher (1994), consists of using MEM to impose smoothness on the estimated image. Recently, Hofmann, Weigelt & Schertl (2014) proposed the IRBis method (image reconstruction software using the bispectrum), which solves the minimization problem from a MAP approach, considering smooth regularization terms, and employing a non-linear optimization algorithm based on conjugate gradients (Hager & Zhang 2005, 2006). However, due to the non-linearity of the considered inverse problem,

[★] E-mail: jb36@hw.ac.uk

the minimization problems solved by the above methods perform only local optimization. For global minimum search, different approaches have been proposed these last years. In particular, techniques based on a Markov chain Monte Carlo (MCMC) method (Gammerman & Lopes 2006) have been adopted in MACIM (Ireland, Monnier & Thureau 2006) and SQUEEZE (Baron, Kloppenborg & Monnier 2012), while in Auria et al. (2014), a tensor approach has been proposed. In the latter, following the idea of phase-lift methods for phase retrieval problems (Candès, Strohmer & Voroninski 2011; Waldspurger, D’Aspremont & Mallat 2013), the data model is lifted from a vector to a supersymmetric rank-1 order-3 tensor formed by the tensor product of the vector representing the sought image with itself. This yields a linear inverse problem, and a convex minimization problem can be deduced from a MAP approach. In Auria et al. (2015), the tensor approach has been extended to account for the signal sparsity and thereby improving the reconstruction quality. However, solving for order-3 tensor instead of an image (i.e. a vector) increases the dimensionality of the problem drastically and makes this approach computationally very expensive. Thus, Auria et al. (2014) proposed another method that involves solving linear and convex sub-problems alternately and iteratively for three images. Although the global minimization problem remains non-convex and dependent on the initial guess, in practice it has been shown that it provides much better reconstruction quality and accelerates the convergence speed as compared to the tensor approach. Moreover, contrary to the state-of-the-art methods, it brings convexity to the sub-problems. However, Auria et al. (2014) proposed to solve the tri-linear problem using a Gauss–Seidel method (Zangwill 1969; Ortega & Rheinboldt 1970, chapter 7; Bertsekas 1999, chapter 2), which does not have any convergence guarantees in this context. Additionally, only positivity constraints have been considered, without imposing any other a priori information on the underlying image.

All of the above-mentioned methods are designed to reconstruct monochromatic images. However, electromagnetic radiations at different wavelengths can be emitted from an astrophysical source, corresponding to its spectrum. In order to exploit the spectrum of the source, modern optical interferometers are paving the way for multi-wavelength imaging. Instruments such as AMBER (Petrov et al. 2000), GRAVITY (Eisenhauer et al. 2007) and MATISSE (Lopez et al. 2008) can take measurements at multiple wavelength channels. This necessitates the progression of imaging techniques from monochromatic to hyperspectral case. Lately, initial work is done in the direction of hyperspectral imaging for optical interferometry. In particular, the method proposed by Kluska et al. (2014), namely SPARCO, is a semi-parametric approach for image reconstruction of chromatic objects, whereas the method proposed by Thiébaud, Soulez & Denis (2013) deals with a sparsity regularized approach considering the observed scene to be a collection of point-like sources. Recently, the use of differential phases for hyperspectral imaging has been proposed in PAINTER (Schutz et al. 2014). The methods proposed by Thiébaud et al. (2013) and Schutz et al. (2014) use the alternating direction method of multipliers (ADMM) algorithm (Boyd et al. 2010) to solve the considered minimization problem.

In this paper, we propose an image reconstruction algorithm that can be applied both for monochromatic and hyperspectral cases in optical interferometry. More precisely, in the monochromatic case, we propose to improve the method based on the tri-linear data model proposed by Auria et al. (2014). First, we propose to impose sparsity as a regularization term, by means of an ℓ_1 norm, either in the image domain or in a given basis (Wiaux et al. 2009; Car-

rillo, McEwen & Wiaux 2012), leveraging the recent compressive sensing theory (Donoho 2006). In addition, we have developed an algorithm, based on the block-coordinate forward–backward algorithm recently proposed, e.g., by Bolte, Sabach & Teboulle (2014), Frankel, Garrigos & Peyrouquet (2015) and Chouzenoux, Pesquet & Repetti (2016), which allows us to deal with non-necessarily smooth regularization terms such as the ℓ_1 norm. Moreover, this algorithm benefits from the convergence guarantees even for the non-convex global minimization problems. Finally, we generalize the proposed method to the hyperspectral case. It translates to a new approach for hyperspectral imaging in optical interferometry. We exploit the joint sparsity of the image cube through an $\ell_{2,1}$ norm (Thiébaud et al. 2013).

The rest of the paper is organized as follows. Section 2 describes the observation model, whereas the corresponding regularized minimization problem is detailed in Section 3. In Section 4, the proposed algorithm to solve the resultant minimization problem is presented along with the implementation details, incorporating various regularization terms. The simulations performed and the results obtained thereby for monochromatic case are discussed in Section 5. Section 6 is devoted to the hyperspectral case. Starting with the problem statement, the optimization details and the simulations performed are then presented with the results obtained. Finally, the conclusion is provided in Section 7.

2 OPTICAL INTERFEROMETRIC OBSERVATION MODEL

Consider the intensity image of interest be represented by real and positive vector $\bar{x} = (\bar{x}_n)_{1 \leq n \leq N} \in \mathbb{R}_+^N$. Its discrete Fourier transform is denoted by $\hat{x} = (\hat{x}_n)_{1 \leq n \leq N} \in \mathbb{C}^N$. An interferometer probes discrete spatial frequencies in the u – v plane of the image of interest. Each spatial frequency sampled by a pair of telescopes, separated by a distance d , is given by (d/λ) , with λ being the observation wavelength (Thiébaud & Giovannelli 2010). Note that the total flux is assumed to be measured independently and the zero-frequency Fourier coefficient, denoted by \hat{x}_c , is normalized to be equal to 1.

In optical interferometry, the measurements are composed by $M_{\mathcal{P}}$ power spectrum measurements, corresponding to the squared modulus of the complex visibilities, and by $M_{\mathcal{B}}$ bispectrum measurements, corresponding to a triple product of three different complex visibilities. Thus, each measurement can be represented by the triple product of Fourier coefficients of the image of interest, i.e. $\hat{x}_i \hat{x}_j \hat{x}_k$, where i, j and k belong to $\{1, \dots, N\}$. Considering the Hermitian symmetry, we denote by \hat{x}_{i^*} the Fourier coefficient at the opposite spatial frequency to that related with \hat{x}_i . Following this notation, the power spectrum measurements are obtained by choosing indices $j = i^*$ and $k = c$, thus giving triple product of the form $\hat{x}_i \hat{x}_{i^*} \hat{x}_c = |\hat{x}_i|^2$. Similarly, for the bispectrum measurements, phase closure should be satisfied so that the spatial frequencies corresponding to \hat{x}_i, \hat{x}_j and \hat{x}_k sum to zero (Monnier 2007). As a result, the bispectrum measurements are given by $\hat{x}_i \hat{x}_j \hat{x}_{(i+j)^*}$.

It is to be mentioned here that in general, for a fixed number A of telescopes in an interferometer, the independent spatial frequencies sampled, each probed by a pair of telescopes, are equal to $\binom{A}{2} = A(A-1)/2$, and the number of possible closing triangles (i.e. phase closures) is $\binom{A}{3} = A(A-1)(A-2)/(3 \times 2)$. However, out of these only $\binom{A-1}{2} = (A-1)(A-2)/2$ number of phase closures are independent (Monnier 2007). As a result, most of the Fourier phase information is missing. Combined with the sparseness

in the u - v coverage, this poses a highly underdetermined inverse problem.

In view of the description provided above, the inverse problem can be written as follows:

$$\mathbf{y} = [(\mathbf{T}_1 \bar{\mathbf{x}}) \cdot (\mathbf{T}_2 \bar{\mathbf{x}}) \cdot (\mathbf{T}_3 \bar{\mathbf{x}})] + \boldsymbol{\eta}, \quad (1)$$

where \cdot denotes the Hadamard product, $\mathbf{y} = (y_m)_{1 \leq m \leq M} \in \mathbb{C}^M$, with $M = M_{\mathcal{P}} + M_{\mathcal{B}}$, $\boldsymbol{\eta} \in \mathbb{C}^M$ is a realization of an additive i.i.d. Gaussian noise and $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$ are linear operators from \mathbb{R}^N to \mathbb{C}^M . More precisely, for every $p \in \{1, 2, 3\}$, \mathbf{T}_p performs a discrete 2D Fourier transform $\mathbf{F} \in \mathbb{C}^{N \times N}$, followed by selection operators, denoted by $\mathbf{S} \in \mathbb{R}^{M_{\mathcal{P}} \times N}$ and $\mathbf{L}_p \in \mathbb{R}^{M \times M_{\mathcal{P}}}$, i.e.

$$\mathbf{T}_p = \mathbf{L}_p \mathbf{S} \mathbf{F}. \quad (2)$$

First, the operator \mathbf{S} selects $M_{\mathcal{P}}$ Fourier coefficients corresponding to the spatial frequencies given by the telescopes' position. Note that due to Hermitian symmetry, only half of the Fourier plane is sampled. Then, the operators $\mathbf{L}_1, \mathbf{L}_2$ and \mathbf{L}_3 select the different coefficients from $\mathbf{S} \mathbf{F} \bar{\mathbf{x}}$, in order to construct the triple products corresponding to the power spectrum and bispectrum measurements. This makes these three operators different from each other.

3 PROPOSED REGULARIZED MINIMIZATION PROBLEM

3.1 Problem formulation

The data model in equation (1) being non-linear, applying directly a MAP approach would lead to a non-convex minimization problem. To bring linearity in equation (1), following the model proposed by Auria et al. (2014), we introduce $(\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \bar{\mathbf{u}}_3) \in (\mathbb{R}_+^N)^3$ such that

$$\bar{\mathbf{u}}_1 = \bar{\mathbf{u}}_2 = \bar{\mathbf{u}}_3 = \bar{\mathbf{x}}. \quad (3)$$

Then, the data model (1) is equivalent to

$$\mathbf{y} = [(\mathbf{T}_1 \bar{\mathbf{u}}_1) \cdot (\mathbf{T}_2 \bar{\mathbf{u}}_2) \cdot (\mathbf{T}_3 \bar{\mathbf{u}}_3)] + \boldsymbol{\eta}, \quad (4)$$

where $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ and $\bar{\mathbf{u}}_3$ correspond to the unknown image that is to be estimated. The new model described in equation (4) is tri-linear, i.e. it is linear in each of the variables $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ and $\bar{\mathbf{u}}_3$. Thus, the problem can be solved separately for each of these variables, keeping other two fixed.

We propose to use a MAP approach to find an estimation of the original image $\bar{\mathbf{x}}$. More precisely, we propose to define the estimation of $(\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2, \bar{\mathbf{u}}_3)$ as a solution to

$$\underset{(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \in (\mathbb{R}^N)^3}{\text{minimize}} \quad f(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) + \sum_{p=1}^3 r(\mathbf{u}_p), \quad (5)$$

where $f : \mathbb{R}^N \rightarrow]-\infty, +\infty[$ is the data fidelity term ensuring consistency of the solution with the measurements, and $r : \mathbb{R}^N \rightarrow]-\infty, +\infty[$ is a regularization term incorporating a priori information on the target image $\bar{\mathbf{x}}$. Here, due to equality (3), we propose to choose the same regularization for $\mathbf{u}_1, \mathbf{u}_2$ and \mathbf{u}_3 .

Since $\boldsymbol{\eta}$ in equation (4) is assumed to be a realization of an i.i.d. Gaussian noise, the usual least-squares criterion can be used for the data fidelity term:

$$f(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \frac{1}{2} \|\mathbf{y} - (\mathbf{T}_1 \mathbf{u}_1) \cdot (\mathbf{T}_2 \mathbf{u}_2) \cdot (\mathbf{T}_3 \mathbf{u}_3)\|_2^2. \quad (6)$$

Note that here we have assumed that the noise variance is the same for both the power spectrum and bispectrum measurements. However, in practice, the bispectrum measurements are degraded by

a noise with greater variance than that of the noise associated with the power spectrum (Pauls et al. 2005). In such scenario, one can use a weighted least-squares data fidelity term in order to incorporate information from the noise covariance matrix (Hofmann et al. 2014).

In order to ensure a good reconstruction quality, we propose to use a hybrid regularization term

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad r(\mathbf{x}) = \iota_{\mathbb{R}_+^N}(\mathbf{x}) + \mu g(\mathbf{x}), \quad (7)$$

where $\iota_{\mathbb{R}_+^N}(\mathbf{x})$ denotes the indicator function equal to 0 if $\mathbf{x} \in \mathbb{R}_+^N$, and $+\infty$ otherwise, $\mu \in]0, +\infty[$ is a regularization parameter, and $g : \mathbb{R}^N \rightarrow]-\infty, +\infty[$ is a convex non-necessarily smooth function. Thus, the proposed formulation can be seen as a generalization of the model proposed in Auria et al. (2014). Indeed, Auria et al. (2014) proposed to solve equation (5) using f defined in equation (6), and r given by equation (7) when $\mu \equiv 0$.

3.2 Symmetrized data fidelity term

Problem (5) can be solved by alternating sequentially between the estimation of each variable $\mathbf{u}_1, \mathbf{u}_2$ and \mathbf{u}_3 while keeping the other two fixed. Since the vectors are solved separately in each sub-problem, the three estimated vectors can converge to different estimations. One method to avoid this issue is to add the information (3) in the regularization term, e.g. to consider quadratic terms controlling the distance between the variables $\mathbf{u}_1, \mathbf{u}_2$ and \mathbf{u}_3 . However, introducing such regularization terms involves additional regularization parameters to be tuned. Thus, to ensure convergence of the three vectors to similar estimations, while avoiding to complicate the minimization problem with additional regularization parameters, we propose to consider a symmetric data fidelity term for $\mathbf{u}_1, \mathbf{u}_2$ and \mathbf{u}_3 , instead of considering the usual least-squares criterion (6). More precisely, in order to take into account the symmetry between $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ and $\bar{\mathbf{u}}_3$, we propose to consider the following data fidelity term:

$$\begin{aligned} \tilde{f}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) &= \frac{1}{6} (f(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) + f(\mathbf{u}_1, \mathbf{u}_3, \mathbf{u}_2) \\ &\quad + f(\mathbf{u}_2, \mathbf{u}_1, \mathbf{u}_3) + f(\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_1) \\ &\quad + f(\mathbf{u}_3, \mathbf{u}_1, \mathbf{u}_2) + f(\mathbf{u}_3, \mathbf{u}_2, \mathbf{u}_1)), \end{aligned} \quad (8)$$

where f is given by equation (6). In this case, it can be noticed that $\mathbf{u}_1, \mathbf{u}_2$ and \mathbf{u}_3 are commutative in equation (8), i.e. we have

$$\begin{aligned} \tilde{f}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) &= \tilde{f}(\mathbf{u}_1, \mathbf{u}_3, \mathbf{u}_2) = \tilde{f}(\mathbf{u}_2, \mathbf{u}_1, \mathbf{u}_3) \\ &= \tilde{f}(\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_1) = \tilde{f}(\mathbf{u}_3, \mathbf{u}_1, \mathbf{u}_2) \\ &= \tilde{f}(\mathbf{u}_3, \mathbf{u}_2, \mathbf{u}_1). \end{aligned} \quad (9)$$

The symmetrization of the data fidelity term can be explained as follows. Due to equality (3), images $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ and $\bar{\mathbf{u}}_3$ correspond to the sought image $\bar{\mathbf{x}}$. Let $\hat{\mathbf{u}}_p = (\hat{u}_{p,i})_{1 \leq i \leq N}$ denote the Fourier transform of $\bar{\mathbf{u}}_p$, for $p \in \{1, 2, 3\}$. Then, for a given frequency index i , we have $\hat{u}_{1,i} = \hat{u}_{2,i} = \hat{u}_{3,i}$. This implies that each measurement y_{ijk} , where (i, j, k) is a triplet of frequency indices, can be given by $\hat{u}_{p,i} \hat{u}_{q,j} \hat{u}_{s,k}$, for all the possible permutations of $(p, q, s) \in (\{1, 2, 3\})^3$, with $p \neq q \neq s$.

Thus, following this symmetrized approach, we propose to

$$\underset{(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \in (\mathbb{R}^N)^3}{\text{minimize}} \quad \left\{ h(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \tilde{f}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) + \sum_{p=1}^3 r(\mathbf{u}_p) \right\}, \quad (10)$$

where \tilde{f} is defined by equation (8), and r is given by equation (7). Note that, since the data fidelity term is symmetrized and the same regularization term is used for $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$, the global cost function h is symmetric as well with respect to $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$. Furthermore, the minimization problem is solved using identical initialization for the unknown vectors $\mathbf{u}_1, \mathbf{u}_2$ and \mathbf{u}_3 , and the final estimation \mathbf{x}^* of $\bar{\mathbf{x}}$ is taken to be the mean of the three estimated vectors.

We will demonstrate in Section 5, through simulation results, that the recovered estimations of $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ and $\bar{\mathbf{u}}_3$ are very close.

3.3 Alternated minimization

As discussed earlier, problem (10) can be solved sequentially, alternating between the estimations of $\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_2$ and $\bar{\mathbf{u}}_3$. To describe the three corresponding sub-problems, additional notations are introduced.

In particular, according to Section 3.2, let us rewrite the considered symmetrized data fidelity term (8) as follows:

$$\tilde{f}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) = \frac{1}{2} \left\| \tilde{\mathbf{y}} - (\tilde{\mathbf{T}}_1 \mathbf{u}_1) \cdot (\tilde{\mathbf{T}}_2 \mathbf{u}_2) \cdot (\tilde{\mathbf{T}}_3 \mathbf{u}_3) \right\|_2^2, \quad (11)$$

where $\tilde{\mathbf{T}}_1, \tilde{\mathbf{T}}_2$ and $\tilde{\mathbf{T}}_3$ are linear operators defined to be the concatenations of the permutations of the operators $(\mathbf{T}_p)_{1 \leq p \leq 3}$:

$$\tilde{\mathbf{T}}_1 = \frac{1}{6^{1/6}} \begin{bmatrix} \mathbf{T}_1 \\ \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_3 \end{bmatrix}, \quad \tilde{\mathbf{T}}_2 = \frac{1}{6^{1/6}} \begin{bmatrix} \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_1 \\ \mathbf{T}_3 \\ \mathbf{T}_1 \\ \mathbf{T}_2 \end{bmatrix} \quad \text{and} \quad \tilde{\mathbf{T}}_3 = \frac{1}{6^{1/6}} \begin{bmatrix} \mathbf{T}_3 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_1 \end{bmatrix}, \quad (12)$$

and $\tilde{\mathbf{y}} \in \mathbb{C}^{(6M)}$ is the concatenation of the corresponding six permutations of the observation vector \mathbf{y} , divided by $6^{1/2}$. Let $(p, q, s) \in \{1, 2, 3\}$. Fix $\mathbf{u}_q \in [0, +\infty[^N$ and $\mathbf{u}_s \in [0, +\infty[^N$ such that $p \neq q \neq s$, and consider the operator $\tilde{\mathbf{T}}_{(u_q, u_s)} : \mathbb{R}^N \rightarrow \mathbb{C}^M$ defined by

$$\tilde{\mathbf{T}}_{(u_q, u_s)} \mathbf{u}_p = \left[(\tilde{\mathbf{T}}_1 \mathbf{u}_q) \cdot (\tilde{\mathbf{T}}_2 \mathbf{u}_s) \cdot (\tilde{\mathbf{T}}_3 \mathbf{u}_p) \right]. \quad (13)$$

Then, the minimization of h with respect to \mathbf{u}_p (while \mathbf{u}_q and \mathbf{u}_s are fixed) can be rewritten as follows:

$$\underset{\mathbf{u}_p \in \mathbb{R}^N}{\text{minimize}} \quad \tilde{f}_p(\mathbf{u}_p | \mathbf{u}_q, \mathbf{u}_s) + r(\mathbf{u}_p), \quad (14)$$

where r is given by equation (7) and

$$\tilde{f}_p(\mathbf{u}_p | \mathbf{u}_q, \mathbf{u}_s) = \frac{1}{2} \left\| \tilde{\mathbf{y}} - \tilde{\mathbf{T}}_{(u_q, u_s)} \mathbf{u}_p \right\|_2^2. \quad (15)$$

Note that the data fidelity term $\tilde{f}_p(\cdot | \mathbf{u}_q, \mathbf{u}_s)$ defined by equation (15) is a convex differentiable function, with its gradient given by

$$\nabla \tilde{f}_p(\mathbf{u}_p | \mathbf{u}_q, \mathbf{u}_s) = \tilde{\mathbf{T}}_{(u_q, u_s)}^\dagger \left(\tilde{\mathbf{T}}_{(u_q, u_s)} \mathbf{u}_p - \tilde{\mathbf{y}} \right). \quad (16)$$

Moreover, $\nabla \tilde{f}_p$ is $\kappa(\mathbf{u}_q, \mathbf{u}_s)$ -Lipschitzian (Bauschke & Combettes 2011, definition 1.46) with

$$\kappa(\mathbf{u}_q, \mathbf{u}_s) = \left\| \tilde{\mathbf{T}}_{(u_q, u_s)}^\dagger \tilde{\mathbf{T}}_{(u_q, u_s)} \right\|_S, \quad (17)$$

$\|\cdot\|_S$ denoting the spectral norm.

3.4 Choice of the regularization term

Concerning the choice of g in equation (7), it is important to emphasize that astronomical images are usually sparse, otherwise they can have sparse representation (Starck, Murtagh & Fadili 2010). Mathematically, this means that the original image can be expressed as

$$\bar{\mathbf{x}} = \Psi \alpha, \quad (18)$$

where $\Psi \in \mathbb{R}^{N \times J}$ is a given dictionary such that $\bar{\mathbf{x}}$ is represented by a sparse vector $\alpha \in \mathbb{R}^J$ in this dictionary. For instance, in the particular case when the image $\bar{\mathbf{x}}$ itself is assumed to be sparse (such as the point source image), one can choose Ψ to be the identity matrix (i.e. the Dirac basis). More generally, for the sparse representation of continuous extended structures, Ψ can be considered as wavelet basis (Mallat 2009), a possibly redundant or a concatenation of non-redundant wavelet basis (Carrillo et al. 2012).

In this context, the theory of compressive sensing has proven its worth in numerous cases to obtain a unique solution from a highly underdetermined problem, relying on the sparsity of the underlying signal (Wiaux et al. 2009; Duarte & Eldar 2011). This drives us to use the regularization function g in equation (7) to promote sparsity in our minimization problem.

A natural way to find the sparsest solution is by considering regularization term of the form

$$g(x) = \|\Psi^\dagger \mathbf{x}\|_0, \quad (19)$$

where $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm counting the non-zero entries of its argument (Donoho 2006). Note that in practice this function is difficult to manage due to its non-convexity and non-differentiability. Thus, non-convexity is often relaxed by the use of the ℓ_1 norm (Chen, Donoho & Saunders 2001), so that the sparsity prior is taken to be

$$g(x) = \|\Psi^\dagger \mathbf{x}\|_1. \quad (20)$$

However, unlike the ℓ_0 pseudo-norm, the ℓ_1 norm is dependent on the magnitude of the coefficients of the signal. Thus, these last years, several approximations of the ℓ_0 pseudo-norm have been proposed (Candès, Wakin & Boyd 2008; Chouzenoux et al. 2013; Repetti et al. 2015). In particular, as proposed in Candès et al. (2008), ℓ_0 minimization behaviour can be nicely approximated by reweighted- ℓ_1 minimization. The authors have shown through several experiments that in many sparse signal recovery problems, reweighted- ℓ_1 minimization can outperform ℓ_1 minimization. In the context of radio interferometry, this has been demonstrated numerically by Carrillo et al. (2012). Thus, in our approach, we will consider both ℓ_1 and reweighted- ℓ_1 to promote sparsity. In the reweighted- ℓ_1 method, a sequence of weighted- ℓ_1 minimization problems is considered, i.e. problem (10) with

$$g(x) = \|\mathbf{W} \Psi^\dagger \mathbf{x}\|_1, \quad (21)$$

where the weights $\mathbf{W} = \text{Diag}(w_1, \dots, w_J)$, with $(w_j)_{1 \leq j \leq J} \in]0, +\infty[^J$, are computed from the current estimation of $\bar{\mathbf{x}}$. Note that in the case when \mathbf{W} is the identity matrix, the usual ℓ_1 regularization (20) is recovered. The calculation of weights will be discussed more in detail in Section 4.4.

4 PROPOSED ALGORITHM

4.1 Algorithm formulation

In this section, we will describe more in detail the proposed alternating minimization algorithm to solve problem (10). We exploit

the variable block structure described in Section 3.3 using a block-coordinate forward–backward algorithm (Bolte et al. 2014; Frankel et al. 2015; Chouzenoux et al. 2016). In this method, \mathbf{u}_1 , \mathbf{u}_2 and \mathbf{u}_3 are updated sequentially, by solving equation (14), as described in Algorithm 1. More precisely, this algorithm consists in computing, at each iteration $k \in \mathbb{N}$,

- (i) $\mathbf{u}_1^{(k+1)}$ while $(\mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})$ are fixed,
- (ii) $\mathbf{u}_2^{(k+1)}$ while $(\mathbf{u}_1^{(k+1)}, \mathbf{u}_3^{(k)})$ are fixed,
- (iii) $\mathbf{u}_3^{(k+1)}$ while $(\mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)})$ are fixed.

The update of each variable $(\mathbf{u}_p^{(k+1)})_{1 \leq p \leq 3}$ is computed with the forward–backward iterations described in steps 7–12 of Algorithm 1. Each iteration involves alternating between

- Step 9: gradient step (or forward step) on the corresponding differentiable function, i.e. $f_1(\cdot | \mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})$ for \mathbf{u}_1 , $\tilde{f}_2(\cdot | \mathbf{u}_1^{(k+1)}, \mathbf{u}_3^{(k)})$ for \mathbf{u}_2 and $\tilde{f}_3(\cdot | \mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)})$ for \mathbf{u}_3 ,
- Step 10: proximity step (or backward step) on the non-necessarily smooth function r .

The proximity operator of r at $\mathbf{x} \in \mathbb{R}^N$ is defined as

$$\text{prox}_r(\mathbf{x}) = \underset{\mathbf{u} \in \mathbb{R}^N}{\text{argmin}} \mathbf{u} + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2. \quad (22)$$

This operator has been introduced by Moreau (1965) and extensively used in signal and image processing to deal with non-smooth functions (Combettes, Duñg & Vũ 2011). An interesting fact concerning this operator is that it admits explicit formulae for a wide class of functions such as ℓ_p norms, for $p > 0$. Finally, it can be seen as a generalization of the projection operator on to a non-empty closed convex set \mathcal{C} when r is chosen to be the indicator function of \mathcal{C} :

$$\begin{aligned} P_{\mathcal{C}}(\mathbf{x}) &= \underset{\mathbf{u} \in \mathbb{R}^N}{\text{argmin}} \iota_{\mathcal{C}}(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2, \\ &= \underset{\mathbf{u} \in \mathcal{C}}{\text{argmin}} \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|_2^2, \end{aligned} \quad (23)$$

thus finding the closest point to \mathbf{x} belonging to the set \mathcal{C} .

Intuitively, the forward–backward iterations can be understood as follows. For each of the variables \mathbf{u}_p , consider the minimization problem (14). Here the objective function is a sum of smooth and non-smooth functions. First, a gradient step is performed on the differentiable function $\tilde{f}_p(\cdot | \mathbf{u}_q, \mathbf{u}_s)$, giving $\mathbf{z}^{(t)}$ (step 9 in Algorithm 1). Then a proximity step (step 10) is applied to the non-smooth function r . Here in the computation of the proximity operator of r , the quadratic term controls the distance between the solution of this step and $\mathbf{z}^{(t)}$. Finally, as a result of these forward–backward iterations, the solution obtained is basically the minimizer of the global objective function in equation (14).

Note that, in Algorithm 1, for every $k \in \mathbb{N}$, the gradient of $\tilde{f}_1(\cdot | \mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})$ (resp. $\tilde{f}_2(\cdot | \mathbf{u}_1^{(k+1)}, \mathbf{u}_3^{(k)})$ and $\tilde{f}_3(\cdot | \mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)})$) depends on the current iterates $(\mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})$ (resp. $(\mathbf{u}_1^{(k+1)}, \mathbf{u}_3^{(k)})$ and $(\mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)})$). Thus, the linear operator $\tilde{\mathbf{T}}_{(\mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})}$ (resp. $\tilde{\mathbf{T}}_{(\mathbf{u}_1^{(k+1)}, \mathbf{u}_3^{(k)})}$ and $\tilde{\mathbf{T}}_{(\mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)})}$) needs to be updated at each iteration $k \in \mathbb{N}$.

Algorithm 1 Block-coordinate forward–backward algorithm.

- 1: **Initialization:** Let $\mathbf{u}_1^{(0)} = \mathbf{u}_2^{(0)} = \mathbf{u}_3^{(0)} \in \mathbb{R}_+^N$, $t_{\max} \in \mathbb{N}^*$, and, for every $k \in \mathbb{N}$, let $(\delta_1^{(k,t)})_{0 \leq t \leq t_{\max}-1}$, $(\delta_2^{(k,t)})_{0 \leq t \leq t_{\max}-1}$ and $(\delta_3^{(k,t)})_{0 \leq t \leq t_{\max}-1}$ be positive sequences.
 - 2: **For** $k = 0, 1, \dots$
 - 3: **for** $p = 1, 2, 3$
 - 4: **if** $p = 1$; $\mathbf{T} = \tilde{\mathbf{T}}_{(\mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})}$; **end if**
 - 5: **if** $p = 2$; $\mathbf{T} = \tilde{\mathbf{T}}_{(\mathbf{u}_1^{(k+1)}, \mathbf{u}_3^{(k)})}$; **end if**
 - 6: **if** $p = 3$; $\mathbf{T} = \tilde{\mathbf{T}}_{(\mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)})}$; **end if**
 - 7: $\tilde{\mathbf{u}}^{(0)} = \mathbf{u}_p^{(k)}$
 - 8: **for** $t = 0, \dots, t_{\max} - 1$
 - 9: $\mathbf{z}^{(t)} = \tilde{\mathbf{u}}^{(t)} - \delta_p^{(k,t)} \mathbf{T}^\dagger (\mathbf{T} \tilde{\mathbf{u}}^{(t)} - \mathbf{y})$
 - 10: $\tilde{\mathbf{u}}^{(t+1)} = \text{prox}_{\delta_p^{(k,t)} r}(\mathbf{z}^{(t)})$
 - 11: **end for**
 - 12: $\mathbf{u}_p^{(k+1)} = \tilde{\mathbf{u}}^{(t_{\max})}$
 - 13: **end for**
 - 14: **end for**
 - 15: **Return:** $\mathbf{x}^* = (\mathbf{u}_1^* + \mathbf{u}_2^* + \mathbf{u}_3^*)/3$, where $\mathbf{u}_1^* = \lim_k \mathbf{u}_1^{(k)}$, $\mathbf{u}_2^* = \lim_k \mathbf{u}_2^{(k)}$, $\mathbf{u}_3^* = \lim_k \mathbf{u}_3^{(k)}$.
-

4.2 Convergence results

The key point of the proposed Algorithm 1 is that its convergence can be derived from Bolte et al. (2014) and Chouzenoux et al. (2016). We present the convergence results in the following theorem.

Theorem 1. Let $(\mathbf{u}_1^{(k)})_{k \in \mathbb{N}}$, $(\mathbf{u}_2^{(k)})_{k \in \mathbb{N}}$ and $(\mathbf{u}_3^{(k)})_{k \in \mathbb{N}}$ be sequences generated by Algorithm 1. Assume that, for every $k \in \mathbb{N}$ and $t \in \{0, \dots, t_{\max} - 1\}$,

$$\begin{cases} \delta_1^{(k,t)} \in]0, 2/\kappa(\mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})[, \\ \delta_2^{(k,t)} \in]0, 2/\kappa(\mathbf{u}_1^{(k+1)}, \mathbf{u}_3^{(k)})[, \\ \delta_3^{(k,t)} \in]0, 2/\kappa(\mathbf{u}_1^{(k+1)}, \mathbf{u}_2^{(k+1)})[, \end{cases} \quad (24)$$

where $\kappa(\cdot, \cdot)$ is defined by equation (17). If g is a semi-algebraic function,¹ then $(\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)})_{k \in \mathbb{N}}$ converges to a critical point $(\mathbf{u}_1^*, \mathbf{u}_2^*, \mathbf{u}_3^*)$ of h , and $(h(\mathbf{u}_1^{(k)}, \mathbf{u}_2^{(k)}, \mathbf{u}_3^{(k)}))_{k \in \mathbb{N}}$ is a non-increasing function converging to $h(\mathbf{u}_1^*, \mathbf{u}_2^*, \mathbf{u}_3^*)$.

Note that, according to Chouzenoux et al. (2016), to ensure the convergence of Algorithm 1, t_{\max} needs to be finite [and equal to 1 in Bolte et al. (2014)]. In the limit case that $t_{\max} \rightarrow +\infty$, Algorithm 1 can be viewed as an approximated Gauss–Seidel algorithm (Zangwill 1969; Ortega & Rheinboldt 1970, chapter 7; Bertsekas 1999, chapter 2). However, up to the best of our knowledge, the most general convergence results for the Gauss–Seidel method are presented in Tseng (2001), and require technical assumptions on $\tilde{f}_p + r$ that are not necessarily satisfied in our minimization problem, due to the selection operators involved in equation (4).² Thus, it is important to note that our method is in contrast with the

¹ A function is semi-algebraic if its graph is a finite union of sets defined by a finite number of polynomial inequalities. Semi-algebraicity property is satisfied by a wide class of functions. In particular, it is satisfied by the different functions g described in Section 3.4.

² In particular, convexity of sub-problems $\tilde{f}_p + r$, $p \in \{1, 2, 3\}$ is not enough to ensure the convergence of the Gauss–Seidel algorithm (Powell 1973).

algorithm proposed by Auria et al. (2014), where an approximated Gauss–Seidel method is adopted.

4.3 Implementation details

As mentioned in Section 4.1, each sub-problem (14) is solved using the forward–backward iterations. At each sub-iteration $t \in \{0, \dots, t_{\max} - 1\}$, for every $p \in \{1, 2, 3\}$, step 10 performs the proximity operator of r , computed as follows:

$$\begin{aligned} \tilde{\mathbf{u}}^{(t+1)} &= \text{prox}_{\delta_p^{(k,t)} r}(\mathbf{z}^{(t)}) \\ &= \underset{\mathbf{u} \in \mathbb{R}^N}{\text{argmin}} \, t_{\mathbb{R}_+^N}(\mathbf{u}) + \zeta_p^{(k,t)} g(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{z}^{(t)}\|_2^2, \end{aligned} \quad (25)$$

where $\zeta_p^{(k,t)} = \delta_p^{(k,t)} \mu$. The computation of the proximity operator in equation (25) depends on the choice of g . It can either have an explicit formulation or need to be computed using sub-iterations. In the following, we briefly describe the proximity steps obtained for the different regularization terms g discussed in Section 3.4.

4.3.1 Positivity and reality

In Auria et al. (2014), only positivity and reality constraints have been considered. Thus, the regularization term (7) corresponds to the case when $\mu = 0$. In this case, the proximity step 10 boils down to the projection of the current iterate on to the real positive set \mathbb{R}_+^N , and is given by

$$\tilde{\mathbf{u}}^{(t+1)} = \text{Proj}_{\mathbb{R}_+^N}(\mathbf{z}^{(t)}) = (\max\{\text{Re}(z_n^{(t)}), 0\})_{1 \leq n \leq N}, \quad (26)$$

where $\text{Re}(\cdot)$ denotes the real part operator.

4.3.2 Positivity, reality and sparsity in the image space

In the case when the original image is known to be sparse, function g can be used to promote sparsity directly in the image space. This corresponds to regularization (20) (resp. (21)) with Ψ chosen equal to the identity matrix. Then, according to Combettes & Pesquet (2010, table 10.2(ix)), the proximity step 10 has an explicit formulation, given by

$$\tilde{\mathbf{u}}^{(t+1)} = (p_n^{(t+1)})_{1 \leq n \leq N}, \quad (27)$$

where, for every $n \in \{1 \leq n \leq N\}$,

$$p_n^{(t+1)} = \begin{cases} \text{Re}(z_n^{(t)}) - \omega_n & \text{if } \text{Re}(z_n^{(t)}) \geq \omega_n, \\ 0 & \text{otherwise,} \end{cases} \quad (28)$$

with $\omega_n = \zeta_p^{(k,t)}$ (resp. $\omega_n = w_n \zeta_p^{(k,t)}$). This operator is called the *positive soft-thresholding operator*. It involves setting all the components of $z_n^{(t)}$ smaller than the soft-thresholding parameter ω_n to zero, while the other components are reduced by ω_n . Thus, this operator sparsifies the vector $z_n^{(t)}$, while imposing positivity and reality.

4.3.3 Positivity, reality and sparsity in a given dictionary

As discussed in Section 3.4, if an astronomical image is not sparse, it can have a sparse representation in a given dictionary Ψ . In this case, regularization (20) or (21) can be used, where Ψ^\dagger is a general dictionary. However, the proximity operator (25) does not have a closed form solution. Its computation in step 10 involves sub-iterations, which we propose to perform using the so-called

dual forward–backward algorithm (Combettes & Pesquet 2010; Combettes et al. 2011), described in Algorithm 2.

Algorithm 2 Dual forward–backward algorithm to compute equation (25).

1: **Initialization:** Let $\tilde{\mathbf{p}}^{(0)} \in \mathbb{R}^N, \epsilon \in]0, \min\{1, 1/\|\mathbf{W}\Psi^\dagger\|^2\}[$ and $\gamma \in [\epsilon, 2/\|\mathbf{W}\Psi^\dagger\|^2 - \epsilon]$.
 2: **For** $\ell = 0, 1, \dots$
 3: $\mathbf{v}^{(\ell)} = \text{Proj}_{\mathbb{R}_+^N}(\mathbf{z}^{(\ell)} - \Psi\mathbf{W}^\dagger\tilde{\mathbf{p}}^{(\ell)})$
 4: $\mathbf{s}^{(\ell)} = \tilde{\mathbf{p}}^{(\ell)} + \gamma\mathbf{W}\Psi^\dagger\mathbf{v}^{(\ell)}$
 5: $\tilde{\mathbf{p}}^{(\ell+1)} = \mathbf{s}^{(\ell)} - \gamma \text{prox}_{\gamma^{-1}\zeta_p^{(k,t)} g}(\gamma^{-1}\mathbf{s}^{(\ell)})$
 6: **end for**
 7: **Return:** $\tilde{\mathbf{u}}^{(t+1)} = \lim_\ell \mathbf{v}^{(\ell)}$.

In Algorithm 2, \mathbf{W} is the identity matrix if the ℓ_1 regularization (20) is used, or \mathbf{W} corresponds to a diagonal matrix with positive weights (w_1, \dots, w_J) if weighted- ℓ_1 regularization (21) is chosen. Moreover, step 3 is computed using definition (26) in the image space, while the proximity operator in step 5 corresponds to the *soft-thresholding operator* (Chaux et al. 2007) computed in the dictionary space. It is given by

$$\text{prox}_{\gamma^{-1}\zeta_p^{(k,t)} g}(\gamma^{-1}\mathbf{s}^{(\ell)}) = (p_j^{(\ell)})_{1 \leq j \leq J}, \quad (29)$$

where, for every $j \in \{1, \dots, J\}$, $p_j^{(\ell)}$ is defined by

$$p_j^{(\ell)} = \begin{cases} \gamma^{-1}(s_j^{(\ell)} + \omega_j) & \text{if } s_j^{(\ell)} < -\omega_j, \\ 0 & \text{if } -\omega_j \leq s_j^{(\ell)} \leq \omega_j, \\ \gamma^{-1}(s_j^{(\ell)} - \omega_j) & \text{otherwise,} \end{cases} \quad (30)$$

where $\omega_j = \zeta_p^{(k,t)}$ (resp. $\omega_j = w_j \zeta_p^{(k,t)}$) if regularization (20) (resp. (21)) is considered. The *soft-thresholding operator* sparsifies the vector $\gamma^{-1}\mathbf{s}^{(\ell)}$, by setting all its components satisfying $|\gamma^{-1}s_j^{(\ell)}| \leq \gamma^{-1}\omega_j$ to zero. Note that unlike the *positive soft-thresholding operator* described in Section 4.3.2, it does not impose positivity.

4.4 Reweighting approach

As discussed in Section 3.4, we propose to use a reweighted- ℓ_1 regularization term to promote sparsity. In particular, we propose to compute the weights \mathbf{W} in equation (21) according to the weighting procedure developed in Candès et al. (2008). More precisely, let \mathbf{x}^* be a critical point obtained from Algorithm 1, where the function r is defined by equation (7) with either $\mu = 0$ or g given by an ℓ_1 regularization function (20). Then, \mathbf{x}^* is used to compute the weights for the first weighting procedure, essentially computed from the inverse of the values of $\Psi^\dagger \mathbf{x}^*$:

$$(\forall j \in \{1, \dots, J\}) \quad w_j = \frac{1}{\epsilon + [|\Psi^\dagger \mathbf{x}^*|]_j}, \quad (31)$$

where $\epsilon > 0$, and $[\Psi^\dagger \mathbf{x}^*]_j$ denotes the j th component of $\Psi^\dagger \mathbf{x}^*$ (if Ψ is chosen to be identity, or if $\mu = 0$, then $J = N$ and $w_n = \frac{\epsilon}{\epsilon + |x_n^*|}$). Note that ϵ in equation (31) can be viewed as a stabilization parameter (see Candès et al. 2008, section 2). In particular, choosing $\epsilon \rightarrow 0$ leads to an approximation of the ℓ_0 pseudo-norm, limiting the dependence of the weighted- ℓ_1 regularization on the magnitude of the signal coefficients.

Finally, Algorithm 1 is used again to solve the new minimization problem, taking into account the weighted- ℓ_1 regularization (21)

with weights computed by equation (31). The new solution \mathbf{x}_1^* obtained from the weighted- ℓ_1 minimization problem can be used to compute new weights from equation (31), where \mathbf{x}^* is replaced by \mathbf{x}_1^* . The reweighted- ℓ_1 minimization problem obtained can be solved in turn using Algorithm 1. This reweighting procedure can be repeated until a stable solution is obtained.

5 SIMULATIONS AND RESULTS

In this section, to show the good behaviour of the proposed method, we will present simulation results, obtained by implementing the proposed algorithm in `MATLAB` [version R2015a].

5.1 Simulation setting

All the simulations are performed on the image `LkH α` shown in Fig. 1, taken from the 2004 Optical Interferometric Imaging Beauty Contest (Lawson et al. 2004), with $N = 64^2$. Two types of u - v coverages are considered.

(i) Fig. 2(a): synthetic u - v coverage, which consists of random variable-density sampling scheme in 2D discrete Fourier space. In this case, the u - v coverage is generated by random Gaussian sampling such that low frequencies are more likely to be sampled than high frequencies.

(ii) Fig. 2(b): realistic u - v coverage, corresponding to discretized version of 2016 Optical Interferometric Imaging Beauty Contest coverage plan (Sanchez-Bermudez et al. 2016). It corresponds to the measurements made by the GRAVITY instrument at the Very Large

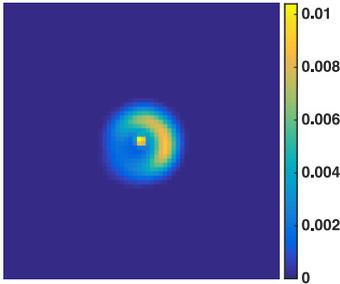


Figure 1. Original image `LkH α` , of size 64×64 , used for simulations, taken from the 2004 Imaging Beauty Contest (Lawson et al. 2004).

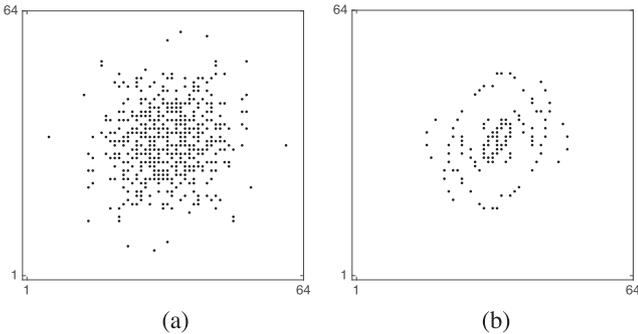


Figure 2. Discretized spatial frequencies coverage plans for the image of size 64×64 . (a) Synthetic u - v coverage for $M_{\mathcal{P}}/N = 0.05$: consists of random variable-density sampling scheme. (b) Realistic u - v coverage: taken from the 2016 Interferometric Imaging Beauty Contest (Sanchez-Bermudez et al. 2016).

Telescope Interferometer (VLTI).³ The observation wavelength is $1.95 \mu\text{m}$. It samples 72 points in the u - v plane resulting into 72 power spectrum measurements.

For both coverages, the bispectrum points are chosen at random, relaxing the phase closure constraint, mainly from the low-frequency region. It is taken care that no two bispectrum measurements correspond to the same triple product.

In both the cases, the simulated measurements in equation (4) are obtained by taking the input signal-to-noise ratio (iSNR) equal to 30 dB, where

$$\text{iSNR} = 20 \log_{10} \left(\frac{\|\mathbf{y}\|_2}{\sqrt{M}\sigma_{\eta}} \right), \quad (32)$$

σ_{η}^2 being the variance of the noise.

For quantitative comparison of the reconstructed images, signal-to-noise ratio (SNR) is considered. For a given estimated image \mathbf{x}^* of an original image $\bar{\mathbf{x}}$, the SNR is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\bar{\mathbf{x}}\|_2}{\|\mathbf{x}^* - \bar{\mathbf{x}}\|_2} \right). \quad (33)$$

In our simulations, results are presented considering a stopping criterion for Algorithm 1, given by $\max_{p \in \{1, 2, 3\}} (\|\mathbf{u}_p^{(k)} - \mathbf{u}_p^{(k-1)}\|_2 / \|\mathbf{u}_p^{(k)}\|_2) \leq 10^{-2}$.

Finally, let us define the power spectrum undersampling ratio as $u_{\mathcal{P}} = M_{\mathcal{P}}/N$, and the bispectrum undersampling ratio as $u_{\mathcal{B}} = M_{\mathcal{B}}/N$. Note that due to the Hermitian symmetry, $M_{\mathcal{P}}$ power spectrum measurements in fact correspond to $2M_{\mathcal{P}}$ sampled spatial frequencies in the Fourier plane. This implies that in the particular case when $u_{\mathcal{P}} = 0.5$, all the spatial frequencies in the Fourier plane are sampled.

As discussed in Section 2, the number of spatial frequencies probed $M_{\mathcal{P}}$ depends on the number of telescopes A . Thus, $u_{\mathcal{P}}$ will change, depending on A . Also for a given $u_{\mathcal{P}}$, there can be at most $\binom{A}{3}$ possible bispectrum measurements, i.e. $M_{\mathcal{B}} \leq \binom{A}{3}$. Keeping this in mind, for a fixed $u_{\mathcal{P}}$, we have performed simulations by varying the number of bispectrum measurements considered, which results into varying $u_{\mathcal{B}}$. Furthermore, for each pair $(u_{\mathcal{P}}, u_{\mathcal{B}})$, 10 simulations are performed, varying the noise realization, and, for the synthetic u - v coverage, the sampling pattern as well.

5.2 Synthetic u - v coverage

This section presents the simulations performed on the image `LkH α` considering the synthetic u - v coverage given in Fig. 2(a). Simulations corresponding to the different regularization terms are described below.

5.2.1 Positivity and reality constraints

We consider the simplest case, described by Auria et al. (2014), corresponding to the minimization problem (10) with only positivity and reality constraints taken into account. Details of the implementation of Algorithm 1 in this case are described in Section 4.3.1.

As mentioned in Section 4.2, given the non-convexity of the minimization problem (10), Algorithm 1 can only converge to a critical point of h . Thus, the reconstructed image depends on the initialization. To avoid local minima, we propose to run Algorithm 1

³ <https://www.eso.org/sci/facilities/paranal/telescopes/vlti.html>

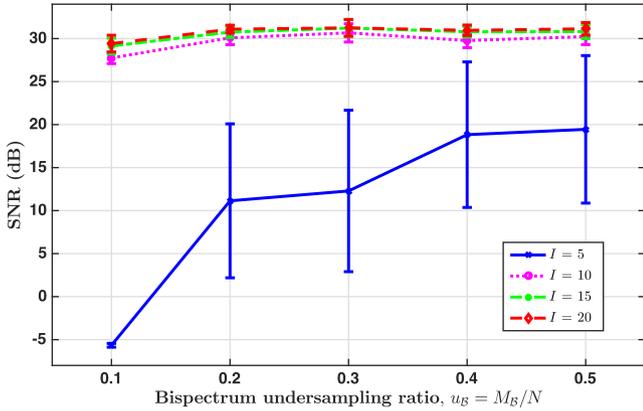


Figure 3. SNR graph obtained for positivity and reality constrained case with LkH α image and synthetic u - v coverage for $u_{\mathcal{P}} = 0.2$, considering iSNR = 30 dB and varying u_B . The graph shows the comparison of average SNR values (over 10 simulations), and corresponding 1-standard-deviation error bars, for different number of initializations I : $I = 5$ (solid blue), $I = 10$ (dotted pink), $I = 15$ (dot-dash green) and $I = 20$ (dashed red).

several times, for I random initializations $\mathbf{x}_i^{(0)} = \mathbf{u}_1^{(0)} = \mathbf{u}_2^{(0)} = \mathbf{u}_3^{(0)}$, with $i \in \{1, \dots, I\}$. Let \mathbf{x}_i^* be the estimation found with initialization $\mathbf{x}_i^{(0)}$. Then the best estimation \mathbf{x}^* is selected by taking $\mathbf{x}^* = \mathbf{x}_{i^*}^*$, where i^* corresponds to the initialization index with minimum value of the objective function, i.e. for every $i \in \{1, \dots, I\}$, $f(\mathbf{x}_{i^*}^*) + r(\mathbf{x}_{i^*}^*) \leq f(\mathbf{x}_i^*) + r(\mathbf{x}_i^*)$.

To choose the number I of random initializations, first tests for different I are performed and presented in Fig. 3. Four curves are depicted, corresponding to the different number of initializations considered, $I \in \{5, 10, 15, 20\}$. Each curve represents the average SNR values over 10 simulations, along with 1-standard-deviation error bars, as a function of the undersampling ratio u_B , for a fixed $u_{\mathcal{P}} = 0.2$. It can be seen from the graph that the SNR changes a lot as the number of random initializations increases from 5 to 20. It reflects the sensitivity of the minimization problem to the number of initializations. However, between 15 and 20 initializations, the SNR not only saturates, in fact it exhibits very small standard deviation error bars. Thus, in all the subsequent simulations, when only positivity and reality constraints are taken into account, we consider $I = 15$ random initializations for each pair $(u_{\mathcal{P}}, u_B)$.

5.2.2 ℓ_1 and weighted- ℓ_1 regularizations

In order to solve the minimization problem (10) promoting sparsity, we consider the regularization function given by equation (7), and we examine both ℓ_1 and weighted- ℓ_1 regularizations defined respectively by equations (20) and (21), using Ψ to be Daubechies 8 wavelet basis (Mallat 2009). In this case, we use Algorithm 1 with the implementation details given in Section 4.3.3, and the reweighting process described in Section 4.4.

Concerning the initialization, both for ℓ_1 and weighted- ℓ_1 minimization problems, two different cases have been tested. On the one hand, we considered the same initialization strategy as described in Section 5.2.1, with $I = 15$. On the other hand, we used the final estimation obtained from the positivity constrained problem, itself initialized with $I = 15$ (Section 5.2.1). Preliminary simulations indicated that the results obtained in the two cases have the similar reconstruction quality. However, the computation time was much longer considering several random initializations than using the solution obtained from the positivity constrained problem.

Thus, for computational efficiency, all further simulations for ℓ_1 and weighted- ℓ_1 regularization are performed using the final solution obtained when only positivity constraint is considered, as described in Section 5.2.1.

To inspect the quality of reconstruction, we consider two sub-cases for ℓ_1 minimization with different number of forward-backward sub-iterations (corresponding to steps 9 and 10 in Algorithm 1): $t_{\max} = 200$ and 400. In addition, for the weighting scheme, two sub-cases are considered for different number of weighting iterations: a weighted- ℓ_1 regularization (with only one weighting computation), and a second weighting iteration (i.e. reweighted- ℓ_1).⁴ As discussed in Section 4.4, the weights are computed using equation (31), where, for the weighted- ℓ_1 regularization, we take \mathbf{x}^* to be the solution obtained from the positivity constrained minimization problem, whereas for the reweighted- ℓ_1 regularization, \mathbf{x}^* is the solution obtained from the weighted- ℓ_1 minimization problem.

Note that during weighted and reweighted- ℓ_1 , t_{\max} is taken to be 200. In the simulations performed, regularization parameter μ in equation (7) is tuned to maximize the SNR: $\mu = 10^{-5}$ (resp. $\mu = 1.5 \times 10^{-5}$) for ℓ_1 (resp. weighted and reweighted- ℓ_1) minimization problem.

5.2.3 Simulation results

We have implemented several tests to analyse the performance of the proposed method with respect to the number of measurements made by the interferometer. More precisely, to take into account different undersampling ratios of the u - v plane, we have performed simulations by varying $u_{\mathcal{P}}$ and u_B . First, concerning the choice of $u_{\mathcal{P}}$, we have considered two cases: $u_{\mathcal{P}} = 0.05$ corresponding to highly undersampled u - v plane, and $u_{\mathcal{P}} = 0.2$ to simulate a less undersampled data set. Secondly, for each of the considered values of $u_{\mathcal{P}}$, we have varied number of bispectrum measurements, i.e. u_B . Taking these different values of $u_{\mathcal{P}}$ and u_B into account, Fig. 4 shows the SNR graphs corresponding to the reconstructed images, as a function of u_B for $u_{\mathcal{P}} = 0.05$ (left) and 0.2 (right), respectively. Typically, the range over which u_B is varied is chosen depending on the value of $u_{\mathcal{P}}$. As such, we have taken the values of u_B comparable to and greater than $u_{\mathcal{P}}$. Consequently, for the smaller value of $u_{\mathcal{P}} = 0.05$, we have considered less number of bispectrum measurements with $u_B \in \{0.04, 0.2\}$, whereas for the larger value of $u_{\mathcal{P}} = 0.2$, the number of bispectrum measurements considered is also increased, $u_B \in \{0.05, 0.5\}$.

In each graph of Fig. 4, comparisons are given for the results obtained using the different regularizations described in Sections 5.2.1 and 5.2.2. For visual assessment, reconstructed images corresponding to median SNR are shown in Figs 5 and 6. The reconstructed images for ℓ_1 regularization with different t_{\max} are visually very similar. Same is the case for reconstructed images with weighted- ℓ_1 and reweighted- ℓ_1 regularization. Hence, in Figs 5 and 6, we show the images corresponding to positivity constrained case, ℓ_1 regularization with $t_{\max} = 200$ and reweighted- ℓ_1 regularization. The respective error images are also displayed to show the absolute error $|\mathbf{x}^* - \bar{\mathbf{x}}|$ between the reconstructed image \mathbf{x}^* and the true image $\bar{\mathbf{x}}$.

⁴ Note that the simulations were performed with more than two weighting iterations. However, preliminary results indicated that after the second weighting iteration, a stable solution was achieved both in terms of the SNR and visual quality.

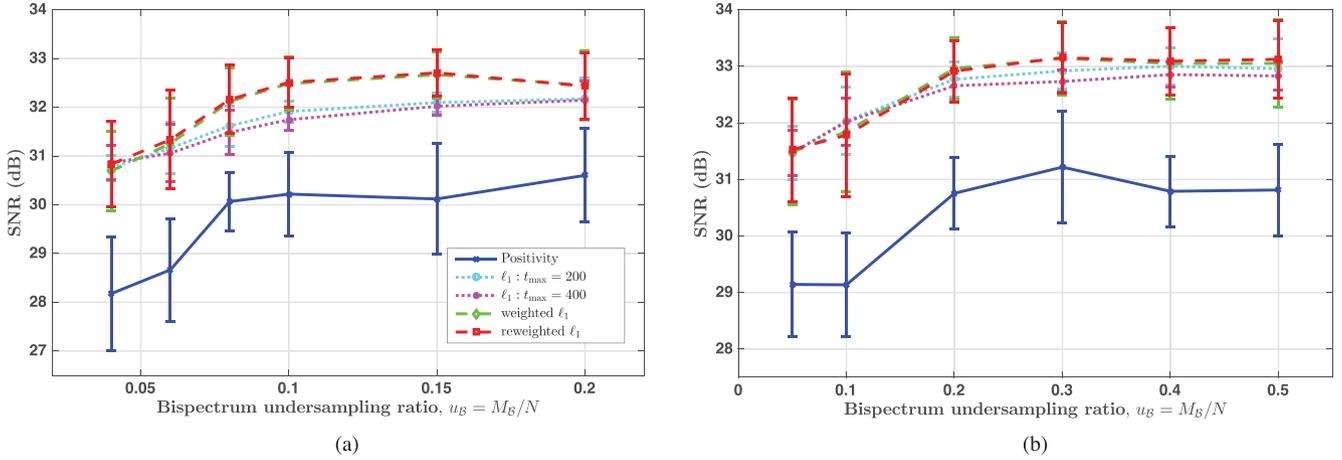


Figure 4. SNR graphs obtained with LkH α image and synthetic u - v coverage, considering iSNR = 30 dB, varying u_B for two different power spectrum undersampling ratios: (a) $u_P = 0.05$ and (b) $u_P = 0.2$. In each graph, comparison of average SNR values (over 10 simulations), and corresponding 1-standard-deviation error bars, for different regularization terms is shown: positivity constraints (solid blue), ℓ_1 regularization with $t_{\max} = 200$ (dotted cyan) and $t_{\max} = 400$ (dotted pink), weighted- ℓ_1 regularization (dashed green) and reweighted- ℓ_1 regularization (dashed red).

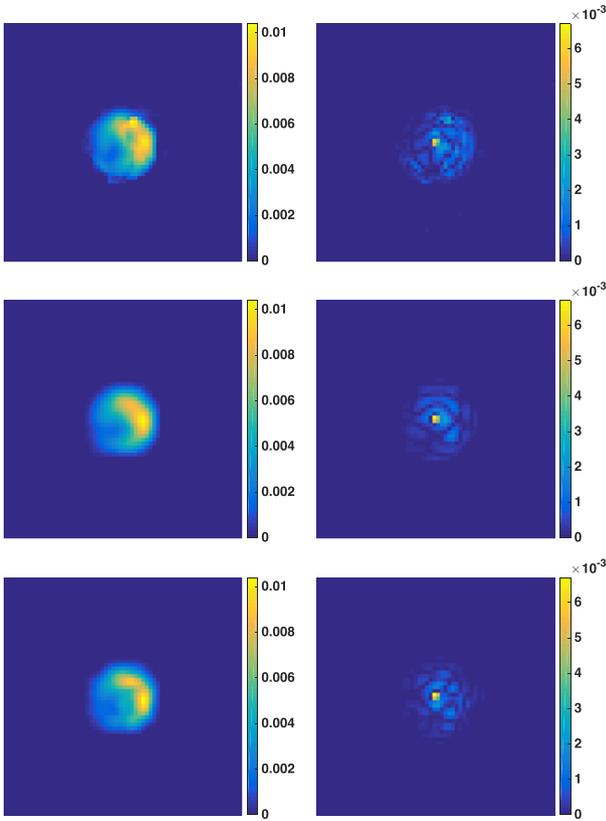


Figure 5. Reconstructed images (first column) and error images (second column) obtained by considering the true image LkH α , corresponding to median SNR (over 10 simulations), with synthetic u - v coverage for $(u_P, u_B) = (0.05, 0.1)$. For both the columns, in each row, images corresponding to different regularization terms are shown. First row: positivity constraint, second row: ℓ_1 regularization with $t_{\max} = 200$ and third row: reweighted- ℓ_1 regularization.

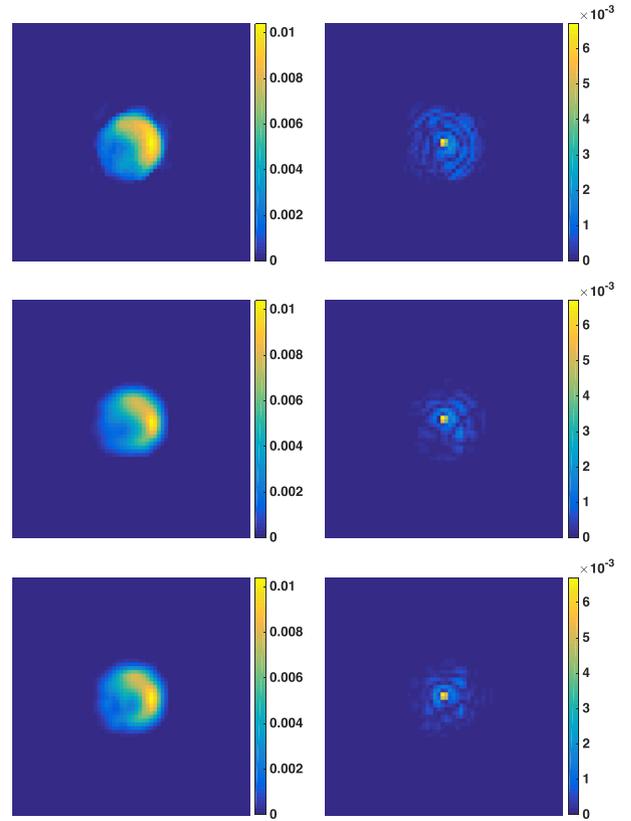


Figure 6. Reconstructed images (first column) and error images (second column) obtained by considering the true image LkH α , corresponding to median SNR (over 10 simulations), with synthetic u - v coverage for $(u_P, u_B) = (0.2, 0.3)$. For both the columns, in each row, images corresponding to different regularization terms are shown. First row: positivity constraint, second row: ℓ_1 regularization with $t_{\max} = 200$ and third row: reweighted- ℓ_1 regularization.

From Figs 4–6, we can observe that promoting sparsity, either by ℓ_1 , weighted- ℓ_1 or reweighted- ℓ_1 regularization term, gives better reconstruction quality, and hence lesser residual in the error images, than the positivity and reality constrained case

[SNR improves between 2 and 3 dB depending on the considered (u_P, u_B)].

Moreover, from the results given in Fig. 4, it can be seen that when $u_P = 0.2$ (Fig. 4b), the quality of reconstruction obtained

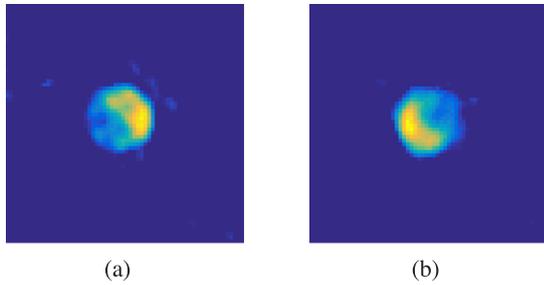


Figure 7. Reconstructed images obtained for the true image $LkH\alpha$, corresponding to two different initializations and the respective median SNR (over 10 simulations) for positivity and reality constrained case, with synthetic $u-v$ coverage for $(u_P, u_B) = (0.05, 0)$ (considering only power spectrum measurements). The figure illustrates the orientation uncertainty when no phase information is taken into account. (a) Reconstructed image with the correct orientation of the true image $LkH\alpha$, (b) Reconstructed image with the opposite orientation.

with the ℓ_1 regularization and the (re)weighted- ℓ_1 regularization is almost the same. In contrast, when $u_P = 0.05$ (Fig. 4a), as u_B is increased, the SNR values obtained with either of the weighted- ℓ_1 or reweighted- ℓ_1 regularization terms are greater than the SNR obtained using an ℓ_1 regularization. This implies that weighting scheme tends to be more beneficial for the case of highly undersampled $u-v$ plane.

Considering the importance of symmetrization, it is worth mentioning here that the reconstructed images for the final solution $\mathbf{x}^* = (1/3)(\mathbf{u}_1^* + \mathbf{u}_2^* + \mathbf{u}_3^*)$ as well as for the solutions of \mathbf{u}_1^* , \mathbf{u}_2^* , \mathbf{u}_3^* are visually indistinguishable. This observation is supported by the small values of the variations between the solutions: $\|\mathbf{u}_1^* - \mathbf{u}_2^*\|_2$, $\|\mathbf{u}_2^* - \mathbf{u}_3^*\|_2$ and $\|\mathbf{u}_3^* - \mathbf{u}_1^*\|_2$, which are of the order of 10^{-2} , 10^{-4} and 10^{-2} , respectively.

5.2.4 Image reconstruction without the bispectrum measurements

In order to emphasize the benefits of using phase information from bispectrum measurements, simulations have been performed considering only the power spectrum measurements, i.e. with $u_B = M_B = 0$. In this case, Algorithm 1 has been implemented by considering only positivity and reality constraints, as described in Section 5.2.1. Moreover, as explained in this section, owing to the non-convexity of the minimization problem (10), several simulations are performed with different random initializations.

Considering the synthetic $u-v$ coverage with $u_P = 0.05$ and $u_B = 0$ (no bispectrum measurements), the reconstructed images obtained from two different random initializations for positivity and reality constrained case are shown in Fig. 7. Since the power spectrum measurements do not contain any phase information, it can be observed that the reconstructed images suffer from phase ambiguity. This arises from the space-reversal property of the Fourier transform, i.e. if a signal is inverted in the spatial domain, then in the Fourier domain, this inversion only reverses the sign of the phase of the Fourier coefficients. It implies that with no phase information, the uncertainty related to signal inversion remains. While the image in Fig. 7(a) is recovered with correct orientation, i.e. the same orientation as that of the original image $LkH\alpha$ given in Fig. 1, the image in Fig. 7(b) is recovered with the opposite orientation.

On the one hand, this indicates that the proposed Algorithm 1 is still able to restore images with only power spectrum measurements,

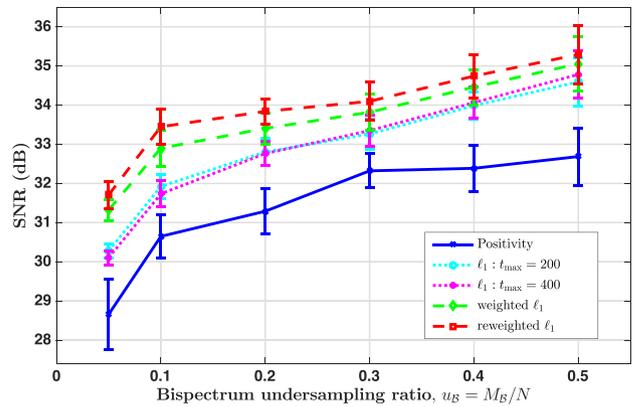


Figure 8. SNR graph obtained with $LkH\alpha$ image and realistic $u-v$ coverage, considering $iSNR = 30$ dB, varying u_B . In the graph, comparison of average SNR values (over 10 simulations), and corresponding 1-standard-deviation error bars, for different regularization terms is shown: positivity constraints (solid blue), ℓ_1 regularization with $t_{\max} = 200$ (dotted cyan) and $t_{\max} = 400$ (dotted pink), weighted- ℓ_1 regularization (dashed green) and reweighted- ℓ_1 regularization (dashed red).

i.e. without any phase information, though with the uncertainty in the orientation. On the other hand, the results obtained from the case when $u_B > 0$ highlight that the incorporation of phase information is essential to recover properly oriented images.

5.3 Realistic $u-v$ coverage

The performance of the proposed algorithm has been assessed for the realistic $u-v$ coverage given in Fig. 2(b). We have performed several simulations by varying the number of bispectrum measurements and thus in turn the bispectrum undersampling ratio u_B . Note that, as mentioned in Section 5.1, for the considered realistic $u-v$ coverage, $M_P = 72$. With $N = 64^2$, this implies that $u_P \simeq 0.018$.

Figs 8 and 9 illustrate the results obtained for different regularization terms, as discussed in Section 5.2. While Fig. 8 depicts the SNR graph for the reconstructed images as a function of $u_B \in \{0.05, 0.5\}$, the corresponding recovered images and the error images for $u_B = 0.2$, with median SNR, are shown in Fig. 9. Here again considering the visual similarity between the reconstructed images for ℓ_1 regularization with different t_{\max} , and that between images for weighted and reweighted- ℓ_1 regularization, we only show the images for positivity constraint, ℓ_1 with $t_{\max} = 200$ and reweighted- ℓ_1 .

It is to be remarked here that the results obtained are in coherence with the observations made for the synthetic $u-v$ coverage in Section 5.2.3. More precisely, the results indicate the superiority of promoting sparsity relative to just positivity and reality over the full undersampling range, leading to an improvement of the SNR between 3 and 4 dB, depending on the considered value of u_B . Moreover, given the small value of u_P , the SNR gets better not only with increasing u_B , but also by considering the (re)weighted- ℓ_1 regularization term.

6 HYPERSPECTRAL IMAGING

6.1 Problem statement

As described in Section 2, the sampled spatial frequencies depend on the observation wavelength. Thus, interferometric measurements made at different wavelengths correspond to probing different

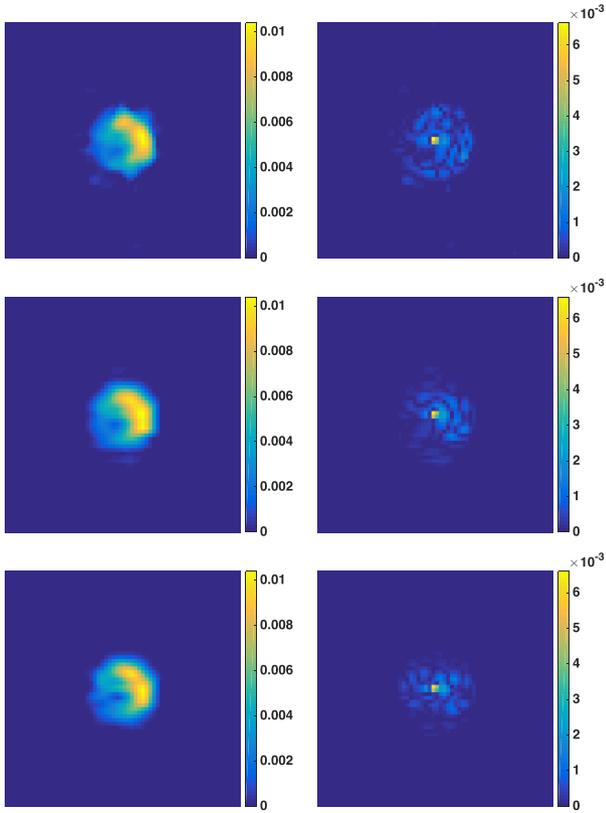


Figure 9. Reconstructed images (first column) and error images (second column) obtained by considering the true image $LkH\alpha$, corresponding to median SNR (over 10 simulations), with realistic $u-v$ coverage for $u_B = 0.2$. For both the columns, in each row, images corresponding to different regularization terms are shown. First row: positivity constraint, second row: ℓ_1 regularization with $t_{\max} = 200$ and third row: reweighted- ℓ_1 regularization.

spatial frequencies in the $u-v$ plane of the image of interest. Considering L spectral channels, in accordance with the data model proposed for the monochromatic case (1), the measurement equation at each spectral channel $l \in \{1, \dots, L\}$ can be written as

$$\mathbf{y}_l = [(\mathbf{T}_{1,l}\bar{\mathbf{x}}_l) \cdot (\mathbf{T}_{2,l}\bar{\mathbf{x}}_l) \cdot (\mathbf{T}_{3,l}\bar{\mathbf{x}}_l)] + \boldsymbol{\eta}_l, \quad (34)$$

where $\mathbf{y}_l \in \mathbb{C}^M$ denotes the measurement vector, $\bar{\mathbf{x}}_l \in \mathbb{R}_+^N$ is the intensity image, $\boldsymbol{\eta}_l \in \mathbb{C}^M$ is a realization of an additive Gaussian noise and, in analogy with equation (2), the l th measurement operators are given by $\mathbf{T}_{p,l} = \mathbf{L}_{p,l}\mathbf{S}_l\mathbf{F}$, for every $p \in \{1, 2, 3\}$. Following the approach adopted in the monochromatic case and considering $\bar{\mathbf{u}}_{1,l} = \bar{\mathbf{u}}_{2,l} = \bar{\mathbf{u}}_{3,l} = \bar{\mathbf{x}}_l$ for $1 \leq l \leq L$, the tri-linear counterpart of the inverse problem (34) becomes

$$\mathbf{y}_l = [(\mathbf{T}_{1,l}\bar{\mathbf{u}}_{1,l}) \cdot (\mathbf{T}_{2,l}\bar{\mathbf{u}}_{2,l}) \cdot (\mathbf{T}_{3,l}\bar{\mathbf{u}}_{3,l})] + \boldsymbol{\eta}_l. \quad (35)$$

Then, concatenating all the spectral channels, we define the ill-posed hyperspectral inverse problem as

$$\mathbf{Y} = [\mathbf{T}_1(\bar{\mathbf{U}}_1) \cdot \mathbf{T}_2(\bar{\mathbf{U}}_2) \cdot \mathbf{T}_3(\bar{\mathbf{U}}_3)] + \mathbf{H}, \quad (36)$$

where $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{C}^{M \times L}$ is the measurement matrix, for every $p \in \{1, 2, 3\}$, $\bar{\mathbf{U}}_p = [\bar{\mathbf{u}}_{p,1}, \dots, \bar{\mathbf{u}}_{p,L}] \in \mathbb{R}_+^{N \times L}$ is the image matrix, $\mathbf{H} = [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_L] \in \mathbb{C}^{M \times L}$ is the noise matrix and $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3$ are the concatenated measurement operators such that, for $p \in \{1, 2, 3\}$, $\mathbf{T}_p(\bar{\mathbf{U}}_p) = (\mathbf{T}_{p,l}\bar{\mathbf{u}}_{p,l})_{1 \leq l \leq L}$. More precisely, column $l \in \{1, \dots, L\}$ of $\bar{\mathbf{U}}_p$ represents the intensity image at wavelength λ_l , whereas row $n \in \{1, \dots, N\}$ represents the variation of pixel values along the spectral channels.

In analogy with the monochromatic case and the minimization problem described in equation (10) by symmetrizing the data fidelity term, we propose to define the estimate of $(\bar{\mathbf{U}}_1, \bar{\mathbf{U}}_2, \bar{\mathbf{U}}_3)$ as a solution to

$$\underset{(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) \in (\mathbb{R}^{N \times L})^3}{\text{minimize}} \left\{ h(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) = \tilde{f}(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) + \sum_{p=1}^3 r(\mathbf{U}_p) \right\}, \quad (37)$$

where the same regularization term

$$(\forall \mathbf{X} \in \mathbb{R}^{N \times L}) \quad r(\mathbf{X}) = t_{\mathbb{R}_+^{N \times L}}(\mathbf{X}) + \mu g(\mathbf{X}) \quad (38)$$

is chosen for $\mathbf{U}_1, \mathbf{U}_2$ and \mathbf{U}_3 , and \tilde{f} is the symmetrized data fidelity term given by

$$\begin{aligned} \tilde{f}(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) &= \frac{1}{2} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{T}}_1(\mathbf{U}_1) \cdot \tilde{\mathbf{T}}_2(\mathbf{U}_2) \cdot \tilde{\mathbf{T}}_3(\mathbf{U}_3)\|_2^2 \\ &= \sum_{l=1}^L \check{f}_l(\mathbf{u}_{1,l}, \mathbf{u}_{2,l}, \mathbf{u}_{3,l}), \end{aligned} \quad (39)$$

with

$$\begin{aligned} \check{f}_l(\mathbf{u}_{1,l}, \mathbf{u}_{2,l}, \mathbf{u}_{3,l}) &= \frac{1}{2} \|\tilde{\mathbf{y}}_l - (\tilde{\mathbf{T}}_{1,l}\mathbf{u}_{1,l}) \cdot (\tilde{\mathbf{T}}_{2,l}\mathbf{u}_{2,l}) \cdot (\tilde{\mathbf{T}}_{3,l}\mathbf{u}_{3,l})\|_2^2. \end{aligned} \quad (40)$$

$\tilde{\mathbf{Y}} = [\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_L] \in \mathbb{C}^{6M \times L}$ and $\tilde{\mathbf{T}}_p(\mathbf{U}_p) = (\tilde{\mathbf{T}}_{p,l}\mathbf{u}_{p,l})_{1 \leq l \leq L}$ are the symmetrized versions of the measurements matrix and the linear operators for $p \in \{1, 2, 3\}$, respectively, in accordance with Section 3.3.

As discussed in the earlier sections, given the voids in the $u-v$ coverage, ensuring data consistency is not sufficient to obtain a good estimation from the measurements, and imposing a priori information is essential. In the monochromatic case, we have considered promoting sparsity prior with a, possibly weighted, ℓ_1 regularization term (Section 3.4). In the context of hyperspectral imaging, joint sparsity gives an additional degree of possible regularization, in the spectral dimension, that should be leveraged to improve the overall image reconstruction quality compared to reconstructing each channel separately (Soulez et al. 2011; Thiébaud et al. 2013; Abdulaziz et al. 2016). Mathematically, joint sparsity is defined for a set of sparse signals such that the non-zero entries of each signal are located at the same spatial position. From physical point of view, if a source is absent, i.e. the corresponding pixel has a zero value in a spectral channel, then the pixels at the same spatial positions along all the spectral channels will be zero. Thus, the joint sparsity prior enforces spatial sparsity while imposing spectral continuity. We propose to promote the joint sparsity prior using an $\ell_{2,1}$ norm (Fornasier & Rauhut 2008; Thiébaud et al. 2013) for the regularization term, defined as follows:

$$g(\mathbf{X}) = \sum_{j=1}^J \left(\sum_{l=1}^L |[\boldsymbol{\Psi}^\dagger \mathbf{x}_l]_j|^2 \right)^{1/2}, \quad (41)$$

where $\boldsymbol{\Psi}$ can either be identity matrix or a given dictionary belonging to $\mathbb{R}^{J \times N}$. The $\ell_{2,1}$ norm is characterized by taking ℓ_2 norm along the columns and then ℓ_1 norm of the resultant vector.

In order to solve the minimization problem (37), we propose to adopt the same methodology as developed for monochromatic case.

6.2 Implementation details

The implementation of Algorithm 1 to solve equation (37) requires replacing the variables and the operators with the corresponding variables and operators for hyperspectral case, as defined in Section 6.1.

First, according to equations (39) and (40), for every $l \in \{1, \dots, L\}$, partial gradients of \check{f}_l are independent. Thus, the gradient descent step 9 of Algorithm 1 can be computed in parallel for each spectral channel. Secondly, the proximity operator of the non-smooth function r defined by equation (38), with g given by equation (41), does not have a closed form solution. In order to compute this, we propose to resort once more to Algorithm 2. In this case, step 3 in Algorithm 2 requires performing the proximity operator of equation (41), defined, for every $B \in \mathbb{R}^{J \times L}$ and $\nu > 0$, as

$$\text{prox}_{\nu \|\cdot\|_{2,1}}(B) = \begin{bmatrix} p_1 \\ \vdots \\ p_J \end{bmatrix}, \quad (42)$$

where, for every $j \in \{1, \dots, J\}$, p_j is a line vector given by

$$p_j = \begin{cases} b_j \frac{\|b_j\|_2 - \nu}{\|b_j\|_2} & \text{if } \|b_j\|_2 \geq \nu, \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad (43)$$

b_j denoting the j -th row of B . Thus, the proximity operator of the $\ell_{2,1}$ norm corresponds to a soft-thresholding operation row-wise.

6.3 Simulations and results

In this section, we will show the performance of the proposed Algorithm 1 for hyperspectral imaging by solving equation (37). Simulations are performed on two sets of images, with $N = 64^2$ for each image. More precisely, two original images are considered: LkH α , given in the top left of Fig. 11, and an image consisting of two simulated uniform discs, which we refer to as *synthetic image*, shown as the top-left image in Fig. 12. These images correspond to the observed image at the first spectral channel \bar{x}_1 . Then, the images corresponding to other spectral channels $l \in \{2, \dots, L\}$ are obtained by following power-law model. In this context, we have, for $\bar{x}_l = (\bar{x}_{l,n})_{1 \leq n \leq N}$,

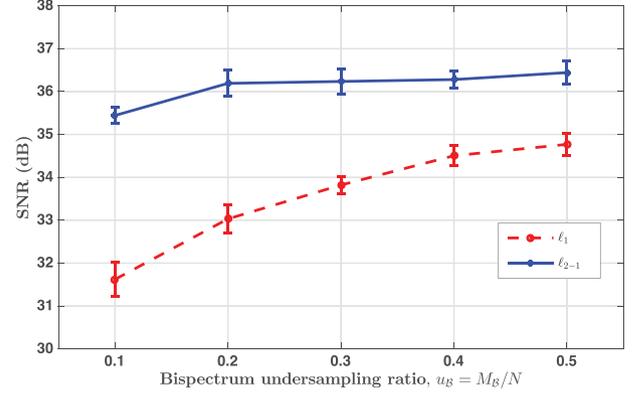
$$\bar{x}_{l,n} = \bar{x}_{1,n} \left(\frac{\lambda_1}{\lambda_l} \right)^{\alpha_n}, \quad (44)$$

where λ_l denotes the wavelength at spectral channel l , and $\alpha = (\alpha_n)_{1 \leq n \leq N}$ is the spectral indices' vector (Rau & Cornwell 2011). Spatial correlation is ensured by taking α to be a linear combination of a random Gaussian field and the reference image convolved with a Gaussian kernel of size 3×3 at full width at half-maximum (Junklewitz, Bell & Enßlin 2015).

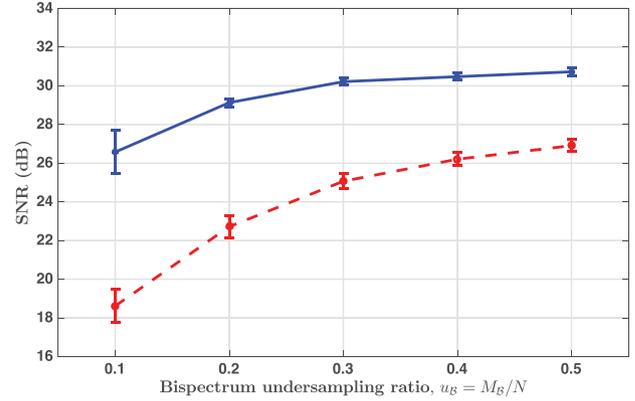
For both the images, $L = 8$ spectral channels in the wavelength range 1.95–1.97 μm are considered. The corresponding u - v coverage plan is given in Fig. 2(b) for observation wavelength 1.95 μm . The generated ground-truth images for $l = 8$ are shown as top-right images in Figs 11 and 12, respectively, for LkH α and *synthetic image*.

We compare the results obtained considering the $\ell_{2,1}$ norm with the case when each channel is treated separately, considering an ℓ_1 norm on each image produced by each spectral channel:

$$(\forall X \in \mathbb{R}^{N \times L}) \quad g(X) = \sum_{l=1}^L \|\Psi^\dagger x_l\|_1. \quad (45)$$



(a)



(b)

Figure 10. SNR graphs obtained for the reconstruction of two different hyperspectral image cubes with the realistic u - v coverage, considering iSNR = 30 dB for each spectral channel, varying u_B . For the two graphs, the ground-truth images at first spectral channel are given by (a) LkH α (top-left image in Fig. 11) and (b) *synthetic image* (top-left image in Fig. 12). Each graph depicts the comparison of the average SNR values (over 10 simulations) and corresponding 1-standard-deviation error bars, between single-channel reconstruction with ℓ_1 regularization (45) (red dashed) and reconstruction by considering joint sparsity with $\ell_{2,1}$ regularization (41) (blue solid).

While the case considering ℓ_1 norm is initialized with the solution of problem (37) solved with only positivity and reality constrained case [i.e. $\mu = 0$ in equation (38)], the solution obtained for each channel by ℓ_1 regularized case is in turn used to initialize $\ell_{2,1}$ regularized case. For both cases, the forward-backward iterations (steps 8–11 in Algorithm 1) are performed with $t_{\max} = 200$.

In the hyperspectral case, we observed that considering Ψ as the identity matrix gives better reconstruction results than using Daubechies wavelets. Moreover, the SNR of the reconstructed image matrix X^* is computed as the mean of the SNRs from the reconstructed images of each channel $(x_l^*)_{1 \leq l \leq L}$. The SNR comparisons between the regularizations (41) and (45) are provided in Fig. 10. For both cases, average SNR curves with 1-standard-deviation error bars are presented (performed over 10 simulations, varying both the noise realization and the measured bispectrum). From these plots, we can observe that using $\ell_{2,1}$ norm as a regularization term leads to better reconstruction than considering only ℓ_1 independently in each channel. The reconstructed and the corresponding error images for the first and the last spectral channels, considering Ψ to be the

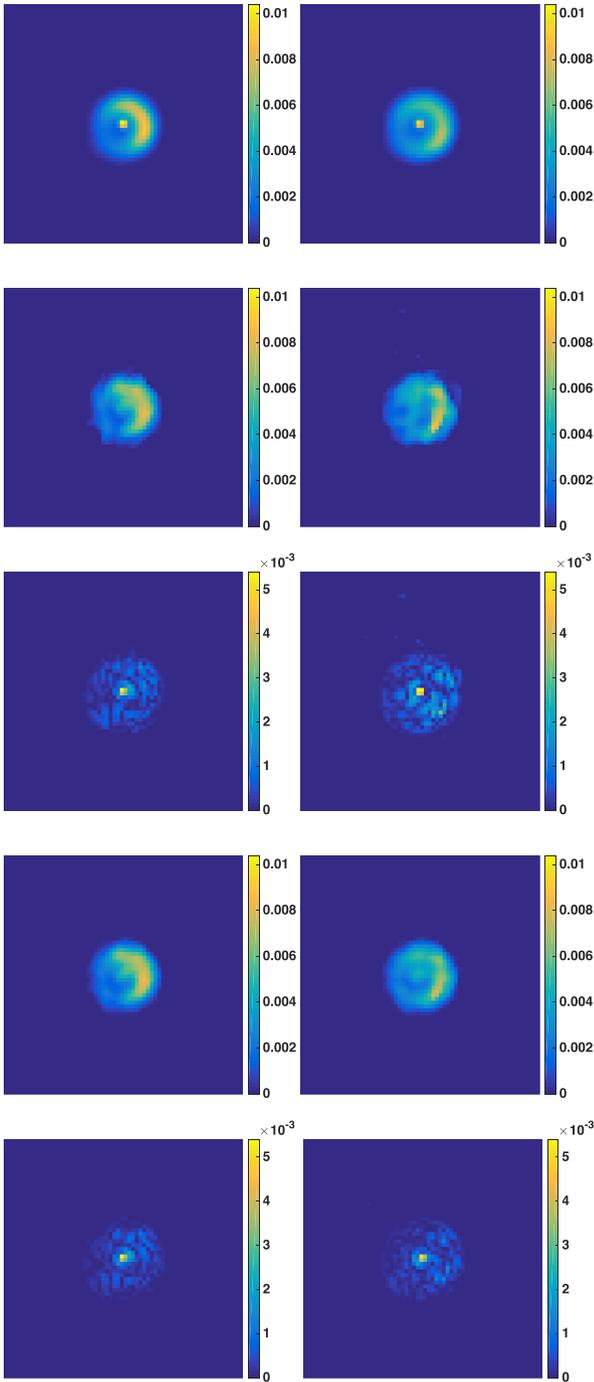


Figure 11. Results for hyperspectral imaging with realistic $u-v$ coverage for $L = 8$, $u_B = 0.1$ and $LkH\alpha$ as the original image (top left). Left column: images corresponding to first spectral channel, $l = 1$. Right column: images corresponding to last spectral channel, $l = 8$. In each column, the images shown are – first row: original image (top), second row: reconstructed image with ℓ_1 regularization (45), third row: error image with ℓ_1 regularization, fourth row: reconstructed image with $\ell_{2,1}$ regularization (41) and fifth row: error image with $\ell_{2,1}$ regularization (41).

identity matrix, are shown in Figs 11 and 12. For the two image examples, the figures demonstrate the superiority of solving globally for the hyperspectral channels over single-channel reconstruction, where no advantage of inter-channel information is taken.

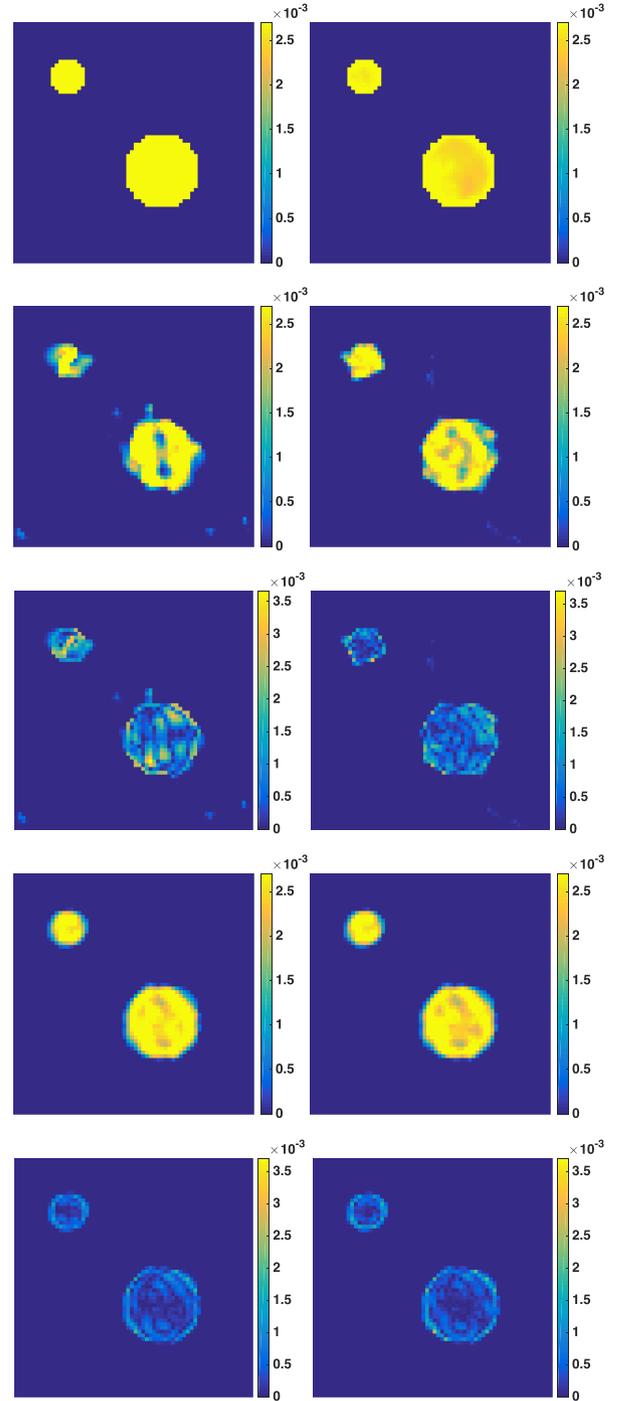


Figure 12. Results for hyperspectral imaging with realistic $u-v$ coverage for $L = 8$, $u_B = 0.1$ and synthetic image as the original image (top left). Left column: images corresponding to first spectral channel, $l = 1$. Right column: images corresponding to last spectral channel, $l = 8$. In each column, the images shown are – first row: original image (top), second row: reconstructed image with ℓ_1 regularization (45), third row: error image with ℓ_1 regularization, fourth row: reconstructed image with $\ell_{2,1}$ regularization (41) and fifth row: error image with $\ell_{2,1}$ regularization (41).

7 CONCLUSION

We have presented a new method for image reconstruction in optical interferometry, based on the tri-linear data model proposed in Auria et al. (2014). While only monochromatic imaging was considered

in the previous work, we have extended this model to hyperspectral imaging. Furthermore, to improve the reconstruction quality, since in Auria et al. (2014) only positivity constraints were considered, we have proposed additionally to promote sparsity using either an ℓ_1 or a weighted- ℓ_1 regularization term in the monochromatic case, and an $\ell_{2,1}$ regularization term in the hyperspectral case. Moreover, in order to solve the resultant minimization problem, we have developed an alternated minimization algorithm, based on a block-coordinate forward-backward algorithm. This algorithm presents convergence guarantees, and benefits from the fact that it can be designed to work with smooth functions, using gradient steps, and with non-necessarily smooth functions thanks to proximity steps. We have assessed the performance of the proposed method on several simulations both for synthetic and realistic $u-v$ coverages, in monochromatic and hyperspectral cases. On the one hand, for monochromatic imaging, adding a sparsity prior gives promising results. On the other hand, for hyperspectral imaging, we have shown numerically that exploiting joint sparsity, using an $\ell_{2,1}$ norm, improves drastically the quality of reconstruction as compared to single-channel reconstruction. To summarize, we have proposed a method that presents a general framework, where the regularization term can be non-smooth and adapted either for the monochromatic case or for the hyperspectral case. Future work includes testing the proposed algorithm on realistic data sets and comparing our method with the state-of-the-art methods in optical interferometry.

ACKNOWLEDGEMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC, grant EP/M008843/1). We would like to thank André Ferrari for insightful discussions.

REFERENCES

- Abdulaziz A., Dabbech A., Onose A., Wiaux Y., 2016, in Proc. EUSIPCO. IEEE, p. 388
- Auria A., Carrillo R., Thiran J.-P., Wiaux Y., 2014, MNRAS, 437, 2083
- Auria A., Carrillo R., Thiran J.-P., Wiaux Y., 2015, in 2014 IEEE International Conference on Image Processing (ICIP). IEEE, p. 6026
- Baron F., Kloppenborg B., Monnier J., 2012, Proc. SPIE, 8445, 84451D
- Bauschke H. H., Combettes P. L., 2011, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer-Verlag, New York
- Bertsekas D., 1999, Nonlinear Programming, 2nd edn. Athena Scientific, Belmont, MA
- Bolte J., Sabach S., Teboulle M., 2014, Math. Program., 146, 459
- Boyd S., Vandenberghe L., 2004, Convex Optimization. Cambridge Univ. Press, Cambridge
- Boyd S., Parikh N., Chu E., Peleato B., Eckstein J., 2010, Found. Trends Mach. Learn., 3, 1
- Buscher D. F., 1994, in Robertson J. G., Tango W. J., eds, Proc. IAU Symp. 158, Very High Angular Resolution Imaging. Kluwer, Dordrecht, p. 91
- Candès E., Wakin M., Boyd S., 2008, J. Fourier Anal. Appl., 14, 877
- Candès E., Strohmer T., Voroninski V., 2011, Commun. Pure Appl. Math., 66, 1241
- Carrillo R., McEwen J., Wiaux Y., 2012, MNRAS, 426, 1223
- Chaux C., Combettes P. L., Pesquet J.-C., Wajs V. R., 2007, Inverse Probl., 23, 1495
- Chen S. S., Donoho D. L., Saunders M. A., 2001, SIAM Rev., 43, 129
- Chouzenoux E., Jezierska A., Pesquet J.-C., Talbot H., 2013, SIAM J. Imaging Sci., 6, 563
- Chouzenoux E., Pesquet J.-C., Repetti A., 2016, J. Global Optim., 66, 457
- Combettes P. L., Pesquet J.-C., 2010, Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer, New York
- Combettes P. L., Duñg D., Vũ B. C., 2011, J. Math. Anal. Appl., 380, 680
- Cornwell T. J., Evans K. F., 1985, A&A, 143, 77
- Donoho D. L., 2006, IEEE Trans. Inf. Theory, 52, 1289
- Duarte M. F., Eldar Y. C., 2011, IEEE Trans. Signal Process., 59, 4053
- Eisenhauer F. et al., 2007, Proc. IAU Symp. 3, A Giant Step: from Milli- to Micro-arcsecond Astrometry. p. 100
- Fornasier M., Rauhut H., 2008, SIAM J. Numer. Anal., 46, 577
- Frankel P., Garrigos G., Peypouquet J., 2015, J. Optim. Theory Appl., 165, 874
- Gamerman A., Lopes H. F., 2006, Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. Chapman and Hall/CRC
- Hager W. W., Zhang H., 2005, SIAM J. Optim., 16, 170
- Hager W. W., Zhang H., 2006, SIAM J. Optim., 17, 526
- Hofmann K.-H., Weigelt G., Schertl D., 2014, A&A, 565, A48
- Högbom J. A., 1974, A&A, 15, 417
- Ireland M., Monnier J., Thureau N., 2006, Proc. SPIE, 6268, 62681T
- Junklewitz H., Bell M. R., Enßlin T., 2015, A&A, 581, A59
- Kluska J. et al., 2014, A&A, 564, A80
- Lawson P. R. et al., 2004, in Traub W. A., ed., Proc. SPIE Conf. Ser. Vol. 5491, New Frontiers in Stellar Interferometry. SPIE, Bellingham, p. 886
- Lopez B. et al., 2008, in Schöller M., Danchi W. C., Delplancke F., eds, Proc. SPIE Conf. Ser. Vol. 7013, Optical and Infrared Interferometry. SPIE, Bellingham, p. 70132B
- Mallat S. G., 2009, A Wavelet Tour of Signal Processing: The Sparse Way, 2nd edn. Academic Press, Burlington, MA
- Meimon S., Mugnier L., Le Besnerais G., 2005, Opt. Lett., 30, 1809
- Monnier J. D., 2007, New Astron. Rev., 51, 604
- Moreau J.-J., 1965, Bull. Soc. Math. France, 93, 273
- Ortega J. M., Rheinboldt W. C., 1970, Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York
- Pauls T. A., Young J. S., Cotton W. D., Monnier J. D., 2005, PASP, 117, 1255
- Petrov R. et al., 2000, in Léna P., Quirrenbach A., eds, Proc. SPIE Conf. Ser. Vol. 4006, Interferometry in Optical Astronomy. SPIE, Bellingham, p. 68
- Powell M. J. D., 1973, Math. Program., 4, 193
- Rau U., Cornwell T. J., 2011, A&A, 532, A71
- Repetti A., Pham M. Q., Duval L., Chouzenoux E., Pesquet J.-C., 2015, IEEE Signal Process. Lett., 22, 539
- Sanchez-Bermudez J. et al., 2016, in Malbet F., Creech-Eakman M. J., Tuthill P. G., eds, Proc. SPIE Conf. Ser. Vol. 9907, Optical and Infrared Interferometry and Imaging V. SPIE, Bellingham, p. 99071D
- Schutz A., Ferrari A., Mary D., Soulez F., Thiébaud E., Vannier M., 2014, J. Opt. Soc. Am. A, 31, 2334
- Soulez F., Bongard S., Thiébaud E., Bacon R., 2011, in 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). IEEE, p. 1
- Starck J.-L., Murtagh F., Fadili J., 2010, Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity. Cambridge Univ. Press, Cambridge
- Thiébaud E., 2002, in Starck J.-L., Murtagh F. D., eds, Proc. SPIE Conf. Ser. Vol. 4847, Astronomical Data Analysis II. SPIE, Bellingham, p. 174
- Thiébaud E., 2008, Proc. SPIE, 7013, 70131I
- Thiébaud E., Giovannelli J., 2010, IEEE Signal Process. Mag., 27, 97
- Thiébaud E., Soulez F., Denis L., 2013, J. Opt. Soc. Am. A, 30, 160
- Thompson A., Moran J., Swenson G., 2001, Interferometry and Synthesis in Radio Astronomy. Wiley-Interscience, New York
- Tseng P., 2001, J. Optim. Theory Appl., 109, 475
- Waldspurger I., D'Aspremont A., Mallat S., 2013, Math. Program., 149, 47
- Wiaux Y., Jacques L., Puy G., Scaife A., Vandergeheynst P., 2009, MNRAS, 395, 1733
- Zangwill W. I., 1969, Nonlinear Programming: A Unified Approach. Prentice-Hall, Englewood Cliffs, NJ

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