Heriot-Watt University
Research Gateway

Facet-Based Regularization for Scalable Radio-Interferometric Imaging

Citation for published version:
Naghibzadeh, S, Repetti, A, van der Veen, A-J & Wiaux, Y 2018, Facet-Based Regularization for Scalable Radio-Interferometric Imaging. in *EUSIPCO 2018.* IEEE, 26th European Signal Processing Conference 2018, Rome, Italy, 3/09/18.

Link:
Link to publication record in Heriot-Watt Research Portal

Document Version:
Peer reviewed version

Published In:
EUSIPCO 2018

Publisher Rights Statement:
© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

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.

Download date: 24. Jun. 2024

Facet-Based Regularization for Scalable Radio-Interferometric Imaging

Shahrzad Naghibzadeh
Faculty of EEMCS
Delft University of Technology
Delft, The Netherlands
s.naghibzadeh@tudelft.nl

Audrey Repetti
EPS School
Heriot-Watt University
Edinburgh, United Kingdom
a.repetti@hw.ac.uk

Alle-Jan van der Veen
Faculty of EEMCS
Delft University of Technology
Delft, The Netherlands
a.j.vanderveen@tudelft.nl

Yves Wiaux
EPS School
Heriot-Watt University
Edinburgh, United Kingdom
y.wiaux@hw.ac.uk

***Abstract*—Current and future radio telescopes deal with large volumes of data and are expected to generate high resolution gigapixel-size images. The imaging problem in radio interferometry is highly ill-posed and the choice of prior model of the sky is of utmost importance to guarantee a reliable reconstruction. Traditionally, one or more regularization terms (e.g. sparsity and positivity) are applied for the complete image. However, radio sky images can often contain individual source facets in a large empty background. More precisely, we propose to divide radio images into source occupancy regions (facets) and apply relevant regularizing assumptions for each facet. Leveraging a stochastic primal dual algorithm, we show the potential merits of applying facet-based regularization on the radio-interferometric images which results in both computation time and memory requirement savings.**

I. Introduction

In radio interferometry (RI), the radio emissions from the sky are observed via a telescope array with the objective of constructing radio sky images over the Field of View (FoV) of the radio telescope. Radio telescope arrays probe the sky through incomplete sampling of the sky image of interest in the Fourier domain, leading to an ill-posed inverse deconvolution problem for image recovery. Any attempt to image reconstruction requires the regularization of this problem by postulating an appropriate signal model. Point sources and extended emissions can coexist in a radio image and require different models to obtain the best reconstruction performance as studied in [1]. Among the most famous deconvolution algorithms, CLEAN [2] models the radio sky as a collection of point sources and is more adequate when the sky is composed of a number of distinct point sources, Multi-Scale CLEAN (MS-CLEAN) [3] improves the performance of CLEAN in retrieving extended emissions by modeling the sky brightness via a collection of components with different scales. More recently, convex optimization approaches have been proposed, leveraging the versatility of optimization theory in incorporating complex signal models into the image recovery. The "Sparsity Averaging Re-weighted Analysis" approach (SARA) has been extensively shown to provide significant improvements in imaging quality in comparison to CLEAN [4]–[7].

The design of the imaging algorithms for the next generation radio telescopes faces extreme challenges due to the unprecedented data volumes and the demand for recovery of sky images at a new range of resolutions and sensitivities. In this context, modern convex optimization algorithmic structures such as the primal dual algorithm [8]–[11] enable to split both data and image into an arbitrary number of blocks which can be handled in parallel. An additional randomization functionality of the primal-dual algorithm enables to visit a subset of the blocks at each iteration. A recent work has studied the benefits of the primal-dual algorithm for radio-interferometric imaging by splitting the data into blocks, and randomizing over the data blocks [6].

The present work is a preliminary attempt to investigate the computational benefit of splitting the image under scrutiny into blocks as well, here called facets, using optimization theory. Faceting is a common technique in RI. Traditionally, they have been introduced to handle the so-called direction-dependent effects (DDEs) [12]. Recently, Tasse et al. [13] proposed a faceting approach to parallelize the facet computation, and accelerate the global image reconstruction. This method is a CLEAN-based technique, assuming sparsity in the image domain (e.g. ℓ_1 regularization). The method we develop in this article is similar to [13] in the parallelization of facet computation, but leverages advanced stochastic optimization techniques rather than greedy approaches. Moreover, optimization theory allows to use versatile regularization terms and benefits from convergence guaranties [10]. In fact, the proposed method can be seen as a multi-facet generalization of the data-block SARA method developed in [6]. In particular, beyond the obvious benefit of parallelization of the facet computation at each iteration, we focus on the following two aspects. Firstly, our work stems from the realization that a large class of radio images of interest are mostly empty, with few structured sources appearing in distinct facets. In this context, it should be computationally more efficient to apply a very simple prior model (i.e. ℓ_1) on the background, and confine sophisticated regularization priors (i.e. sparsity averaging) to specific source regions (facets). We assume prior knowledge of the facet decomposition, which can for example be obtained from low-resolution reconstruction. Secondly, we investigate how the computation can be further lightened by relying on a stochastic functionality and propose a procedure to process only a fraction of the facets.

The remainder of the manuscript is organized as follows.

This work was supported by the NWO DRIFT project (contract 628.002.002) and the UK Engineering and Physical Sciences Research Council (EPSRC, grant EP/M008843/1 and EP/M019306/1).

In Section II we describe the RI imaging problem and recall the data splitting approach developed in [6]. We introduce our facet-based approach and our algorithm in Section III. We illustrate the performance of our approach on simulated data in Section IV. Finally, we draw our conclusions in Section V.

II. Radio-interferometric Imaging

A. Problem description

Starting from an array processing perspective [14], the telescope array contains distinct antennas that capture radio signals that are split into narrow frequency bands. The FoV of the radio telescope is decomposed into Q pixels that can be independently treated as the source signals impinging on the array. The received signals are contaminated by the receiver noise modeled as mutually independent zero mean Gaussian signals. The objective is to find an estimate $\mathbf{x}^\star$ of the original unknown sky brightness distribution $\overline{\mathbf{x}} \in \mathbb{R}^Q$ from the degraded measurement $\mathbf{y} \in \mathbb{C}^M$ (also called visibilities), corresponding to the pairwise correlation of the signals output from the telescope array. This results in a linear measurement model

$$\mathbf{y} = \mathbf{\Phi}\overline{\mathbf{x}} + \mathbf{e} \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{C}^{M\times Q}$ is the linear measurement operator and $\mathbf{e} \in \mathbb{C}^M$ is a realization of an additive random noise with bounded energy, i.e. there exists $\epsilon > 0$ such that $\|\mathbf{e}\|_2 \leq \epsilon$. We assume that the measurement operator $\mathbf{\Phi}$ is perfectly known, such that no calibration is needed [15]. It is modeled as the product between a matrix $\mathbf{G}$ and an oversampled Fourier operator $\mathbf{F}$ (implemented using the Fast Fourier Transform - FFT). The matrix $\mathbf{G}$ contains the compact support kernels enabling the computation of the continuous Fourier samples from the discrete Fourier coefficients provided by the FFT [16], [17].

B. Compressive sensing and data splitting

In [5], the authors have proposed to define the estimate of $\overline{\mathbf{x}}$ as a solution to a convex minimization problem, leveraging compressive sensing theory [18], [19]. In this context, the sky image is assumed to have a sparse representation into a given basis $\mathbf{\Psi} \in \mathbb{C}^{Q\times L}$, with $L \geq Q$. In other words, it is assumed that $\mathbf{\Psi}^\dagger\overline{\mathbf{x}}$ has only few non-zero coefficients, where $\mathbf{\Psi}^\dagger$ denotes the Hermitian transpose matrix of $\mathbf{\Psi}$. This sparsity-aware approach has been adopted in multiple works such as [5], [6], [20], [21] with different choices of sparsity basis Ψ. In [5], the authors proposed to promote average sparsity by choosing $\mathbf{\Psi}$ to be the concatenation of the Dirac basis with the first 8 Daubechies wavelet transforms [22]. More recently, to achieve a highly parallelizable algorithmic structure, it has been proposed in [6] to split the visibilities into $n_d > 0$ blocks. Formally, $\mathbf{y} = (\mathbf{y}_j)_{1\leq j\leq n_d}$ where $\mathbf{y}_j \in \mathbb{C}^{M_j}$ is the jth data block of dimension M_j. In this context, the linear system (1) can be rewritten as follows:

$$(\forall j \in \{1,\ldots,n_d\}) \quad \mathbf{y}_j = \mathbf{\Phi}_j\overline{\mathbf{x}} + \mathbf{e}_j \tag{2}$$

where $\mathbf{\Phi}_j = \mathbf{G}_j\mathbf{F} \in \mathbb{C}^{M_j\times Q}$ is the associated block of the measurement matrix and $\mathbf{e}_j \in \mathbb{C}^{M_j}$ is the jth block of the additive noise $\mathbf{e}$. Since the additive noise is assumed to have a bounded energy, for every $j \in \{1,\ldots,n_d\}$, there exists $\epsilon_j > 0$ such that $\|\mathbf{e}_j\|_2 \leq \epsilon_j$. Using this structure, the authors have proposed to define the estimate of the sky image as a solution to

$$\underset{\mathbf{x}\in\mathbb{R}_+^Q}{\text{minimize}} \; \| \mathbf{\Psi}^\dagger\mathbf{x} \|_1 + \sum_{j=1}^{n_d} \iota_{\mathcal{B}_2(\mathbf{y}_j,\epsilon_j)}(\mathbf{\Phi}_j\mathbf{x}), \tag{3}$$

where $\iota_{\mathcal{B}_2(\mathbf{y}_j,\epsilon_j)}$ denotes the indicator function of the ℓ_2 ball $\mathcal{B}_2(\mathbf{y}_j,\epsilon_j)$ centered in $\mathbf{y}_j$ with radius ϵ_j. The indicator function at a point $\mathbf{z}$ is equal to 0 if $\mathbf{z} \in \mathcal{B}_2(\mathbf{y}_j,\epsilon_j)$, and $+\infty$ otherwise.

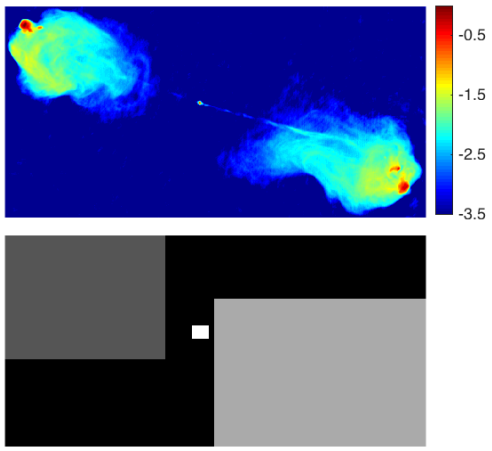


Fig. 1. Top: Image of Cygnus A ($Q = 512 \times 1024$) in log scale. Bottom: Image showing the 4 considered facets (3 facets with structures, and 1 facet for the background).

III. Proposed Approach

A. Facet-based Imaging

Giga-pixel images of the radio sky $\overline{\mathbf{x}}$ usually consist of a large black background which is mostly empty and separate extended structures which can be clustered together. An example of this type of radio image is shown in Fig. 1(top) which shows the radio emission from the Cygnus A radio galaxy image with separate emissions and a large emission-free background. We propose to take advantage of this particular sky image structure and split these images into non-overlapping n_c facets. Formally, we define $\overline{\mathbf{x}} = (\overline{\mathbf{x}}_k)_{1\leq k\leq n_c}$ where, for every $k \in \{1,\ldots,n_c\}$, $\overline{\mathbf{x}}_k \in \mathbb{R}^{Q_k}$ consists of a subpart of the image, and $Q = Q_1 + \ldots + Q_{n_c}$. In Fig. 1(bottom), we show an example of the image splitting of the Cygnus A image. We can see that 3 main facets are identified, corresponding to the 3 main structures of the image. In addition, the background (in black in Fig. 1(bottom)) is considered to be a 4-th facet. In this work, we assume that a pre-processing clustering step has been performed (e.g. using a low resolution estimate), and that we have access to the support of each facet.

In this context, problem (1) can be rewritten for the jth data block as

$$\mathbf{y}_j = \sum_{k=1}^{n_c} [\mathbf{\Phi}_j\overline{\mathbf{x}}]_k + \mathbf{e}_j, \tag{4}$$

where $[.]_k$ denotes the kth block of its argument. Since the considered facets can contain either complex structures, or almost only zero coefficients for the background, we propose to choose different sparsity regularization terms for the different

facets. In particular, the facet associated with the background, chosen to be the last facet $k = n_c$, is mostly empty in the image domain. To avoid losing any weak emission that might be present in the background, we propose to choose a simple regularization, the ℓ_1 norm, to regularize this particular facet. For all the other facets containing sophisticated structures, we use the average sparsity regularization introduced in [5]. The associated minimization problem can be formulated as

$$\underset{\mathbf{x}=(\mathbf{x}_1,\ldots,\mathbf{x}_{n_c})\in\mathbb{R}_+^Q}{\text{minimize}} \sum_{k=1}^{n_c} \gamma_k \|\boldsymbol{\Psi}_k^\dagger \mathbf{x}_k\|_1 + \sum_{j=1}^{n_d} \iota_{\mathcal{B}_2(\mathbf{y}_j,\epsilon_j)}\big(\sum_{k=1}^{n_c} \boldsymbol{\Phi}_{j,k}\mathbf{x}_k\big), \quad (5)$$

where, for every $k \in \{1,\ldots,n_c\}$, $\gamma_k > 0$ is a regularization parameter, associated with the kth facet. In this formulation, for every $k \in \{1,\ldots,n_c-1\}$, $\boldsymbol{\Psi}_k \in \mathbb{R}^{L_k \times Q}$ corresponds to the concatenation of the Dirac basis and the first 8 Daubechies wavelet transforms associated to the kth facet $\mathbf{x}_k$. The last facet $\mathbf{x}_{n_c}$ being dedicated to the background, $\boldsymbol{\Psi}_k$ is chosen to be the Dirac basis (i.e. identity matrix with $L_k = Q$).

B. Proposed algorithm

Problem (5) consists in minimizing a sum of composite non-differentiable convex functions. Primal-dual proximal algorithms are particularly efficient to solve such problems [11]. This class of iterative optimization algorithms have been already used during the last decade in the context of RI, when sparsity-based regularization terms are considered [4]–[6]. In particular, in [6], the authors proposed to leverage the stochastic primal-dual proximal algorithm developed in [10], to solve problem (3). Basically, in this work, the authors proposed to utilize the stochastic properties to reduce the computational complexity of the algorithm by activating, randomly, only a subset of the data blocks per iteration.

In the proposed facet-based imaging approach, not only the data are divided into blocks, but also the the image is divided into facets. We propose to use the block-coordinate structure of the stochastic primal-dual proximal algorithms developed in [10] to handle efficiently the block data terms and the facet-based regularization terms. The resulting *faceting primal dual algorithm*, to solve problem (5), is described in Algorithm 1.

In Algorithm 1, we can distinguish three main parallel loops. The first loop (step 6) is used to update the facets. The second loop (step 10) is used to handle the ℓ_1 facet-based regularization terms. Finally, the third loop (step 21) is used to handle the ℓ_2-ball constraints related to the data blocks.

To handle the positivity and the ℓ_2-ball constraints, in steps 7 and 24 respectively, projection steps are performed. The projection onto a convex, closed, non-empty subset $\mathcal{C}$ of $\mathbb{R}^Q$ of $\mathbf{c} \in \mathbb{R}^P$, is defined as $\Pi_{\mathcal{C}}(\mathbf{c}) = \underset{\boldsymbol{\vartheta}\in\mathbb{R}^P}{\operatorname{argmin}} \|\boldsymbol{\vartheta} - \mathbf{c}\|_2$. On the one hand, in the case when $\mathcal{C} = \mathbb{R}_+^Q$, this projection reduces to

$$(\forall \mathbf{x} \in \mathbb{R}^Q) \quad \Pi_{\mathbb{R}_+^Q}(\mathbf{x}) = \max\big\{\mathbf{0}, \mathbf{x}\big\}, \quad (6)$$

Algorithm 1: Faceting primal-dual algorithm

1 **Initialization:** $\mathbf{x}_0 \in \mathbb{R}^Q$, $\boldsymbol{\varsigma}_0 \in \mathbb{R}^Q$, $\boldsymbol{\upsilon}_0 \in \mathbb{R}^Q$, $\nu \in]0,+\infty[$, $\tau \in]0,+\infty[$, $(\forall k \in \{1,\ldots,n_c\})$ $\mathbf{v}_{k,0} \in \mathbb{R}^{L_k}$, and $\gamma_k \in]0,+\infty[$, and $(\forall j \in \{1,\ldots,n_d\})$ $\mathbf{z}_{j,0} \in \mathbb{R}^{M_j}$
2 **Iterations:**
3 **for** $n = 0, 1, \ldots$ **do**
4 Choose randomly $\mathbb{S}_n \subset \{1,\ldots,n_c\}$
5 Choose randomly $\mathbb{D}_n \subset \{1,\ldots,n_d\}$
6 **for** $k \in \{1,\ldots,n_c\}$ *(facet parallel update)* **do**
7 $\mathbf{x}_{k,n+1} = \Pi_{\mathbb{R}_+^Q}\big(\mathbf{x}_{k,n} - \tau(\boldsymbol{\varsigma}_{k,n} + \boldsymbol{\upsilon}_{k,n})\big)$
8 $\widetilde{\mathbf{x}}_{k,n} = 2\mathbf{x}_{k,n+1} - \mathbf{x}_{k,n}$
9 **end**
10 **for** $k \in \{1,\ldots,n_c\}$ *(regularization parallel step)* **do**
11 **if** $k \in \mathbb{S}_n$ **then**
12 $\mathbf{u}_{k,n} = \mathbf{v}_{k,n} + \gamma_k \boldsymbol{\Psi}_k^\dagger \widetilde{\mathbf{x}}_{k,n}$
13 $\mathbf{v}_{k,n+1} = \mathbf{u}_{k,n} - \gamma_k \mathcal{T}_{\gamma_k^{-1}}\big(\gamma_k^{-1}\mathbf{u}_{k,n}\big)$
14 $\boldsymbol{\upsilon}_{k,n} = \boldsymbol{\Psi}_k \mathbf{v}_{k,n}$
15 **else**
16 $\mathbf{v}_{k,n+1} = \mathbf{v}_{k,n}$
17 $\boldsymbol{\upsilon}_{k,n+1} = \boldsymbol{\upsilon}_{k,n}$
18 **end**
19 **end**
20 $\mathbf{b}_n = \mathbf{F}\widetilde{\mathbf{x}}_n$
21 **for** $j \in \{1,\ldots,n_d\}$ *(data parallel step)* **do**
22 **if** $j \in \mathbb{D}_n$ **then**
23 $\mathbf{s}_{j,n} = \mathbf{z}_{j,n} + \nu \mathbf{G}_j \mathbf{b}_{j,n}$
24 $\mathbf{z}_{j,n+1} = \mathbf{s}_{j,n} - \nu \Pi_{\mathcal{B}_2(\mathbf{y}_j,\epsilon_j)}(\nu^{-1}\mathbf{s}_{j,n})$
25 $\mathbf{g}_{j,n+1} = \mathbf{G}_j^\dagger \mathbf{z}_{j,n+1}$
26 **else**
27 $\mathbf{z}_{j,n+1} = \mathbf{z}_{j,n}$
28 $\mathbf{g}_{j,n+1} = \mathbf{g}_{j,n}$
29 **end**
30 **end**
31 $\boldsymbol{\varsigma}_{n+1} = \mathbf{F}^\dagger \mathbf{g}_{n+1}$
32 **end**

where the max operation is performed component-wise. On the other hand, in the case when $\mathcal{C} = \mathcal{B}_2(\mathbf{y}_j, \epsilon_j)$, for every $j \in \{1,\ldots,n_d\}$, we have, for every $\mathbf{s}_j \in \mathbb{C}^{M_j}$,

$$\Pi_{\mathcal{B}_2(\mathbf{y}_j,\epsilon_j)}(\mathbf{s}_j) = \begin{cases} \mathbf{s}_j & \text{if } \mathbf{s}_j \in \mathcal{B}_2(\mathbf{y}_j,\epsilon_j), \\ \mathbf{y}_j + \epsilon_j \dfrac{\mathbf{s}_j - \mathbf{y}_j}{\|\mathbf{s}_j - \mathbf{y}_j\|_2} & \text{otherwise.} \end{cases} \quad (7)$$

Similarly, to handle the ℓ_1 regularization terms, soft-thresholding operations are performed in step 13. For every facet $k \in \{1,\ldots,n_c\}$, for every $\gamma_k > 0$ and $\mathbf{u}_k \in \mathbb{R}^{L_k}$, this operator is defined as follows [23]:

$$\mathcal{T}_{\gamma_k^{-1}}(\mathbf{u}_k) = \begin{cases} -\mathbf{u}_k^{(i)} + \gamma_k^{-1} & \text{if } \mathbf{u}_k < -\gamma_k^{-1}, \\ 0 & \text{if } -\gamma_k^{-1} < \mathbf{u}_k < \gamma_k^{-1}, \\ \mathbf{u}_k - \gamma_k^{-1} & \text{otherwise.} \end{cases} \quad (8)$$

Algorithm 1 is a stochastic algorithm in the sense that, at each iteration $n \in \mathbb{N}$, a subset $\mathbb{S}_n$ of the n_c facet-based regularization terms and a subset $\mathbb{D}_n$ of the n_d data terms are selected randomly (steps 4 and 5, respectively), and only the related variables are updated. It is in particular interesting to notice that the sparsifying operator $\mathbf{\Psi}_k$ is only applied if the facet $k \in \{1, \dots, n_c\}$ is selected (see parallel loop in step 10). Similarly, as proposed in [6], the gridding matrix $\mathbf{G}_j$ is only applied if the data block $j \in \{1, \dots, n_d\}$ is selected (see parallel step 21). However, the oversampled FFT operator $\mathbf{F}$ and its adjoint must be performed at each iteration (steps 20 and 31, respectively). Note that when $n_c = 1$, then the algorithm proposed in [6] is recovered. In addition, if, at each iteration $n \in \mathbb{N}$, we choose $\mathbb{S}_n = \{1, \dots, n_c\}$ and $\mathbb{D}_n = \{1, \dots, n_d\}$, then the algorithm reduces to a deterministic primal dual algorithm [8], [9].

Finally, it is important to emphasize that the proposed algorithm benefits from the convergence properties of the general stochastic primal dual algorithms developed in [10]. Let $\nu > 0$, $\tau > 0$ and, for every $k \in \{1, \dots, n_c\}$, let $\gamma_k > 0$ satisfying $\tau^{-1} > \nu \|\mathbf{\Phi}\|_S^2 + \sum_{k=1}^{n_c} \gamma_k \|\mathbf{\Psi}_k\|_S^2$, where $\|.\|_S$ denotes the spectral norm. Then, the sequence $(\mathbf{x}_n)_{n \in \mathbb{N}}$ generated by Algorithm 1 converges almost surely to a random variable $\mathbf{x}^\star$ solution to Problem (5).

IV. Simulation Results

We evaluate the performance of Algorithm 1 on simulated radio-interferometric data. We choose as the test image the Cygnus A radio galaxy image of size $Q = 512 \times 1024$ shown in in Fig. 1(top), in log scale. The image is manually split into the $n_c = 4$ facets shown in Fig. 1(bottom). In this image, the black facet represents the background, indexed by $k = n_c$. The visibilities are generated according to model 1, using a random Gaussian undersampling $u-v$ coverage, with $M = Q$. The $u - v$ coverage, split into $n_d = 16$ blocks, is shown in Fig. 2. In model 1, the additive noise is generated following to a zero-mean complex Gaussian distribution, considering an input signal-to-noise ratio (iSNR) of 20 dB, where iSNR $= 20\log_{10}\left(\frac{\|\mathbf{y}\|_2}{\sqrt{M}\sigma}\right)$, σ^2 being the variance of the noise. For the details of data generation and the estimation of the bounds $(\epsilon_j)_{1 \leq j \leq n_d}$ the reader is referred to [6].

We compare the performances of the proposed faceting primal-dual algorithm, with and without randomization over the facet-based regularization terms, with the primal-dual algorithm developed in [6] to judge the benefit of faceting. Since the data blocks are handled exactly in the same manner in Algorithm 1 and [6], in our simulations, we focus on the interest of using a facet-based approach, with possible randomization. Consequently, in our simulations, we choose, for every $n \in \mathbb{N}$, $\mathbb{D}_n = \{1, \dots, n_d\}$ (i.e. no randomization over the data terms). The reader is referred to [6] for a complete investigation of randomizing the primal dual algorithm over the data blocks. Concerning $\mathbb{S}_n$, we investigate two cases. In the first case, at each iteration $n \in \mathbb{N}$, we choose $\mathbb{S}_n = \{1, \dots, n_c\}$. It corresponds to a deterministic version of Algorithm 1 where all the facet-based regularization terms are selected at each iteration. This first configuration is used to emphasize the advantage of the proposed facet-based approach, considering the simple ℓ_1 regularization term on the background. In the second case, we use the randomization property of Algorithm 1. In this case, we choose to activate, at each iteration $n \in \mathbb{N}$ the complete background (i.e. facet $k = n_c = 4$) and select randomly one of the three other facets. In this context, we choose $\mathbb{S}_n = \{k_n, 4\}$, where $k_n \in \{1, 2, 3\}$ is chosen following a uniform distribution.

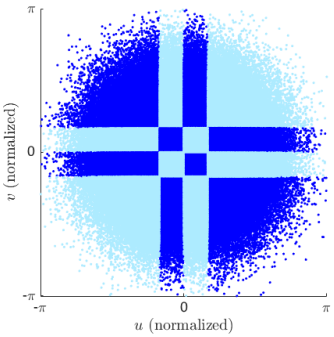


Fig. 2. Normalized random Gaussian $u - v$ coverage. The light blue and dark blue colors emphasize the different n_d data blocks

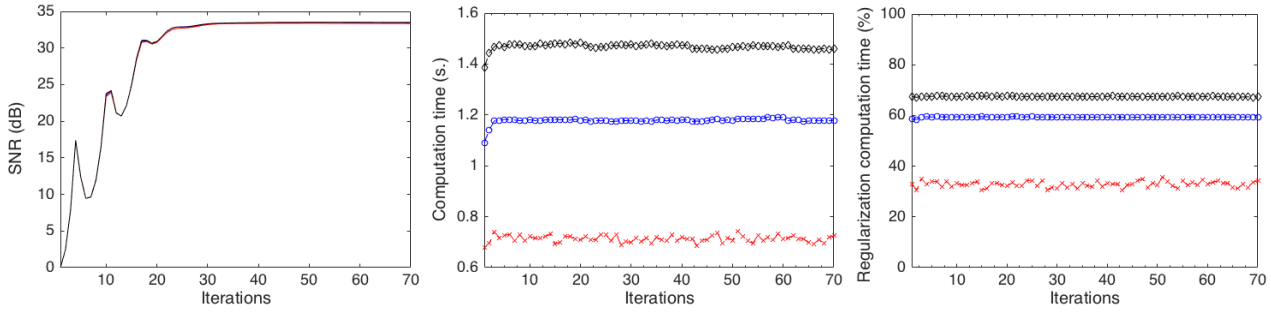
We present in Fig. 3 the results of our simulations. Fig. 3(a) shows the signal-to-noise ratio (SNR) as a function of the iterations $n \in \mathbb{N}$. The SNR (in dB), for the current iterate $\mathbf{x}_n$ is defined as

$$\text{SNR} = 20 \log_{10} \left(\frac{\|\overline{\mathbf{x}}\|_2}{\|\overline{\mathbf{x}} - \mathbf{x}_n\|_2} \right). \tag{9}$$

These curves emphasize that the proposed facet-based regularization term is not affecting the reconstruction quality. In Figs. 3(b) and (c), for the three considered algorithms, we give the time necessary to compute each global iteration, and the % of time spent to perform the regularization steps (i.e. step 10 in Algorithm 1) per global iteration, respectively. By comparing the black curve ([6]) and the blue curve (Algorithm 1 without randomization), we can conclude that the proposed facet-based regularization term reduces the total computational time. This is due to the fact that the wavelet transforms are only performed on the facets with complex structures. In particular, the theoretical complexity to perform the wavelet decomposition with Algorithm 1 (without randomization) is $\sum_{k=1}^{n_c-1} \mathcal{O}(L_k)$, which is smaller than the complexity of performing a wavelet decomposition on the global image (which is equal to $\mathcal{O}(L)$). The red curves are associated with the randomized version of Algorithm 1. It can be observed that leveraging the stochastic properties of this algorithm allows to divide by 2 the computation time per iteration (w.r.t. [6]). In addition, with the randomized version, performing the regularization steps only requires $\sim$ 30% of the iteration computation time, while it requires $\sim$ 60% without randomization and $\sim$ 70% for [6]. We note that in this comparison, the facets being identified manually, the computational contribution from the facet-selection procedure is not included in Fig. 3.

V. Conclusion and Future Works

We have reported our attempts to achieve a facet-based regularization for scalable interferometric imaging. Firstly, we



(a) (b) (c)

Fig. 3. Results obtained using the primal dual algorithm from [6] (black - diamond marks), Algorithm 1 without randomization (blue - circle marks), and Algorithm 1 with randomization on the facet-based regularization terms (red - cross marks). (a) SNR (dB) as a function of iterations. The SNR is an averaged SNR for 10 realizations of noise and under-sample distributions. (b) Computation time per global iteration. (c) Computation time to handle the regularization term (i.e. step 10 in Algorithm 1), as a % of the time necessary to perform a complete iteration. For (b)-(c), the computation time is an averaged computation time for 10 realizations of noise and under-sample distributions, with 10 runs on each setting.

have extended the primal-dual algorithm developed in [6] to incorporate image facets and enable randomization of the facet computation at each iteration. Secondly, we have shown that based on the nature of radio images where individual source facets exist in a sea of mostly-empty background, we can apply the regularization more wisely in order to save computations without sacrificing the reconstruction quality. Thirdly, we have shown that randomization over facets further saves computations.

For future works, we plan to implement an automated faceting schemes based on a low resolution initial image and integrate general faceting schemes which also include overlapping facets in the algorithm by introducing consensus steps as presented in [10]. In addition, similar to the faceting approach proposed in [13], we will perform splitting of the measurement operator over the facets. Finally, we plan to develop a facet-based DDE calibration version of the proposed method, by leveraging a block-coordinate approach [21], [24].

REFERENCES

[1] S. Yatawatta, "Fundamental limitations of pixel based image deconvolution in radio astronomy," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2010 IEEE*. IEEE, 2010, pp. 69–72.

[2] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy and Astrophysics Supplement Series*, vol. 15, p. 417, 1974.

[3] T. J. Cornwell, "Multiscale clean deconvolution of radio synthesis images," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 5, pp. 793–801, 2008.

[4] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Monthly Notices of the Royal Astronimical Society*, vol. 395, no. 3, pp. 1733–1742, 2009.

[5] R. Carrillo, J. McEwen, and Y. Wiaux, "Sparsity averaging reweighted analysis (SARA): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.

[6] 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," *Monthly Notices of the Royal Astronomical Society*, vol. 462, no. 4, pp. 4314–4335, 2016.

[7] A. Dabbech, A. Onose, A. Abdulaziz, R. A. Perley, O. M. Smirnov, and Y. Wiaux, "Cygnus a super-resolved via convex optimisation from vla data," *arXiv preprint arXiv:1710.08810*, 2017.

[8] L. Condat, "A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *Journal of Optimization Theory and Applications*, vol. 158, no. 2, pp. 460–479, 2013.

[9] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comp. Math.*, vol. 38, no. 3, pp. 667–681, 2013.

[10] J.-C. Pesquet and A. Repetti, "A class of randomized primal-dual algorithms for distributed optimization," *Journal of Nonlinear and Convex Analysis*, vol. 16, no. 12, pp. 2353–2490, 2015.

[11] N. Komodakis and J.-C. Pesquet, "Playing with duality: An overview of recent primal dual approaches for solving large-scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.

[12] L. Kogan and E. W. Greisen, "Faceted imaging in aips," *AIPS memo*, 2009.

[13] C. Tasse, B. Hugo, M. Mirmont, O. Smirnov, M. Atemkeng, L. Bester, M. Hardcastle, R. Lakhoo, S. Perkins, and T. Shimwell, "Faceting for direction-dependent spectral deconvolution," *Astronomy & Astrophysics*, vol. 611, p. A87, 2018.

[14] A.-J. van der Veen and S. J. Wijnholds, "Signal processing tools for radio astronomy," in *Handbook of Signal Processing Systems*. Springer, 2013, pp. 421–463.

[15] A. Repetti, J. Birdi, A. Dabbech, and Y. Wiaux, "Non-convex optimization for self-calibration of direction-dependent effects in radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 470, no. 4, pp. 3981–4006, 2017.

[16] L. Greengard and J.-Y. Lee, "Accelerating the nonuniform fast fourier transform," *SIAM review*, vol. 46, no. 3, pp. 443–454, 2004.

[17] J. Fessler and B. Sutton, "Nonuniform fast fourier transforms using min-max interpolation," *IEEE Tran. Sig. Proc.*, vol. 51, no. 2, pp. 560–574, Feb 2003.

[18] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[19] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[20] S. J. Hardy, "Direct deconvolution of radio synthesis images using l1 minimisation," *A&A*, vol. 557, p. A134, 2013. [Online]. Available: http://dx.doi.org/10.1051/0004-6361/201321833

[21] A. Repetti, J. Birdi, A. Dabbech, and Y. Wiaux, "Non-convex optimization for self-calibration of direction-dependent effects in radio interferometric imaging," *Monthly Notices of the Royal Astronimical Society*, vol. 470, no. 4, pp. 3981–4006, Oct. 2017.

[22] I. Daubechies, *Ten lectures on wavelets*. Siam, 1992, vol. 61.

[23] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*. Springer, 2011, pp. 185–212.

[24] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward–backward algorithm," *Journal of Global Optimization*, vol. 66, no. 3, pp. 457–485, 2016.