Uncertainty Quantification in Imaging: When Convex Optimization Meets Bayesian Analysis

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

Marcelo Pereyra
MACS School
Heriot-Watt University
Edinburgh, United Kingdom
m.pereyra@hw.ac.uk

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

***Abstract*—We propose to perform Bayesian uncertainty quantification via convex optimization tools (BUQO), in the context of high dimensional inverse problems. We quantify the uncertainty associated with particular structures appearing in the maximum *a posteriori* estimate, obtained from a log-concave Bayesian model. A hypothesis test is defined, where the null hypothesis represents the non-existence of the structure of interest in the true image. To determine if this null hypothesis is rejected, we use the data and prior knowledge. Computing such test in the context of imaging problem is often intractable due to the high dimensionality involved. In this work, we propose to leverage probability concentration phenomena and the underlying convex geometry to formulate the Bayesian hypothesis test as a convex minimization problem. This problem is subsequently solved using a proximal primal-dual algorithm. The proposed method is applied to astronomical radio-interferometric imaging.**

***Index Terms*—Bayesian uncertainty quantification, hypothesis testing, astronomical imaging, inverse problem, proximal primal-dual algorithm**

I. Introduction

Many imaging problems can be formulated as an inverse problem. Here we focus on linear imaging inverse problems, where the objective is to estimate an unknown image $\overline{\boldsymbol{x}} \in \mathbb{R}^N$ from a degraded observation $\boldsymbol{y} \in \mathbb{C}^M$, related to $\bar{\boldsymbol{x}}$ by a model

$$\boldsymbol{y} = \boldsymbol{\Phi}\overline{\boldsymbol{x}} + \boldsymbol{w}, \tag{1}$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is the measurement matrix and $\boldsymbol{w} \in \mathbb{C}^M$ is a realization of an additive random noise. A common Bayesian method to solve (1) consists in modelling $\overline{\boldsymbol{x}}$ as a random vector with a prior distribution $p(\boldsymbol{x})$ that is log-concave, followed by *maximum a posteriori* (MAP) estimation given by

$$\boldsymbol{x}^\dagger \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}}\ f(\boldsymbol{x},\boldsymbol{y}) + g(\boldsymbol{x}) \tag{2}$$

where $f(\boldsymbol{x},\boldsymbol{y}) = -\log p(\boldsymbol{y}|\boldsymbol{x})$ associated with (1), $g(\boldsymbol{x}) = -\log p(\boldsymbol{x})$, and where we note that $f, g \in \Gamma_0(\mathbb{R}^N)$[1]. Problem (2) can generally be efficiently solved by using modern convex optimization techniques [1]–[10]. Using a MAP approach has the advantage of summarising $\boldsymbol{x}|\boldsymbol{y}$ as a single point estimate that can be easily displayed and visually analysed. However, in many applications it is necessary to go beyond point estimation and also analyze the uncertainty in the solutions delivered. This is the case for example in applications related to quantitative imaging, scientific inquiry, and image-driven decision-making, particularly in fields such as medicine or astronomy. Also note that uncertainty quantification analyses are particularly important in problems that are ill-posed or ill-conditioned because of their high intrinsic uncertainty [11].

The Bayesian framework provides a rich theory to analyse the uncertainty in the value of $\bar{\boldsymbol{x}}$. A first step in this analysis is to identify the regions of space where $\boldsymbol{x}|\boldsymbol{y}$ takes values with high probability [11], and to devise strategies to visually explore this uncertainty [12]. Following on from this, the next step is to enable quantitative uncertainty analyses, namely Bayesian hypothesis tests, to properly assess the degree of confidence in specific image structures. Unfortunately, it is difficult to perform these analyses in large-scale imaging problems because of the cost involved in calculating probabilities w.r.t. $p(\boldsymbol{x}|\boldsymbol{y})$ when N is large (see [13], [14] for more details about state-of-the-art Bayesian computation methods). Consequently, most imaging methods do not quantify uncertainty.

Following the recent work [11], here we propose to address this difficulty by leveraging probability concentration phenomena and the model's underlying convex geometry to formulate Bayesian hypothesis tests as convex problems, that we then efficiently solve by using a scalable proximal primal-dual algorithm [5], [6], [9]. The method described in this paper is mainly based on our recent work [15], where a simpler projection onto convex sets (POCS) algorithm was used.

Quantifying the uncertainty associated with a particular structure is of great interest for astronomers, in the context of astronomical imaging. Related methods are important to provide the foundational statistical information on the estimated image, which would help to validate particular emissions appearing in the reconstructed sky image (e.g. compact sources). In particular, radio-interferometric imaging aims to observe an area of the sky at high angular resolution through an array of antennas. Next generation telescope, namely Square Kilometre Array (SKA), will provide giga-pixel images, with unprecedented dynamic range [16]–[18]. These big datasets necessitate scalable uncertainty quantification methods. In this work, we provide illustrative examples to show the potential of the proposed

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

[1] $\Gamma_0(\mathbb{R}^N)$ denotes the set of convex, proper, lower semi-continuous functions from $\mathbb{R}^N$ to $]-\infty,+\infty]$.

ISBN 978-90-827970-1-5 © EURASIP 2018

approach in this context. In particular, we focus on small compact structures which can correspond in practice to, e.g. point sources, or slightly extended sources.

The remainder of the paper is organized as follows. In Section II we introduce the class of Bayesian models considered. Then Section III presents proposed uncertainty quantification method. Numerical results are reported in Section IV. Finally, we conclude in Section V.

II. Problem description

In this work, we consider that $\boldsymbol{w}$, in the inverse problem (1), is a realization of a zero-mean random noise with a bounded energy, i.e. $\|\boldsymbol{w}\|_2 \leqslant \varepsilon$ with $\varepsilon > 0$ and $\|.\|_2$ denoting the usual Euclidean norm. Consequently, we choose the associated log-likelihood term appearing in (2) to be the indicator function[2] of the ℓ_2 ball, centred in $\boldsymbol{y}$ with radius $\varepsilon > 0$ (denoted by $\mathcal{B}_2(\boldsymbol{y}, \varepsilon)$):

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad f(\boldsymbol{x}, \boldsymbol{y}) = \iota_{\mathcal{B}_2(\boldsymbol{y},\varepsilon)}(\boldsymbol{\Phi}\boldsymbol{x}). \tag{3}$$

Moreover, for the log-prior we use the function

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad g(\boldsymbol{x}) = \iota_{[0,+\infty[^N}(\boldsymbol{x}) + \lambda\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1, \tag{4}$$

where $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$ is an orthonormal sparsity basis (e.g. wavelet transform [19]), and $\lambda > 0$ is a regularization parameter. This prior enforces positivity of pixel values and promotes sparsity in the representation induced by $\boldsymbol{\Psi}$. The resulting MAP estimate $\boldsymbol{x}^\dagger$, given by Problem (2) with (3)-(4), can be efficiently computed by using primal-dual proximal methods [5], [6], [9].

III. Proposed approach

A. Hypothesis test

The proposed uncertainty quantification method takes the form of a Bayesian hypothesis test on the degree of support for specific image structures appearing in the MAP estimate $\boldsymbol{x}^\dagger$ (e.g., lesions in medical imaging, or celestial sources in astronomical imaging). To define this test we postulate the two following hypotheses:

H_0 : The structure of interest is ABSENT in the true image,

H_1 : The structure of interest is PRESENT in the true image.

These hypotheses split the set of images $\mathbb{R}^N$ onto two regions. We denote by $\mathcal{S}$ the region containing all images *without the structure*; the images *containing the structure* belong to the complementary $\mathbb{R}^N \setminus \mathcal{S}$. Note that $\mathsf{P}(H_0|\boldsymbol{y}) = \mathsf{P}(\overline{\boldsymbol{x}} \in \mathcal{S}|\boldsymbol{y})$.

The null hypothesis H_0 is rejected with significance α if

$$\mathsf{P}(H_0|\boldsymbol{y}) \leqslant \alpha,$$

or equivalently if $\mathsf{P}(H_1|\boldsymbol{y}) > 1-\alpha$. Here we assess this probability by comparing the convex set $\mathcal{S}$ with the region of the solutions space where most of the posterior probability mass of $\overline{\boldsymbol{x}}|\boldsymbol{y}$ lies. This region is known as a *posterior credible set* in the Bayesian literature [11]. By definition, $\mathcal{C}_\alpha$ is a posterior credible set with confidence level $(1-\alpha)$, if $\mathsf{P}(\boldsymbol{x} \in \mathcal{C}_\alpha|\boldsymbol{y}) = 1-\alpha$. Identifying exact credible regions is too computationally expensive in imaging problems because of the high dimensionality involved. To overcome this difficulty, [11] proposes a conservative credible region $\widetilde{\mathcal{C}}_\alpha$, in the sense that $\mathsf{P}(\boldsymbol{x} \in \widetilde{\mathcal{C}}_\alpha|\boldsymbol{y}) \geqslant 1-\alpha$, that can be computed as a by-product of MAP estimation. Precisely, for any $\alpha \in]4\exp(-N/3), 1[$, $\widetilde{\mathcal{C}}_\alpha$ is the set given

$$\widetilde{\mathcal{C}}_\alpha = \big\{\boldsymbol{x} \in [0,+\infty[^N \;\big|\; \boldsymbol{\Phi}\boldsymbol{x} \in \mathcal{B}_2(\boldsymbol{y},\varepsilon) \text{ and } \boldsymbol{\Psi}\boldsymbol{x} \in \mathcal{B}_1(\boldsymbol{0}, \widetilde{\eta}_\alpha/\lambda)\big\}, \tag{5}$$

where $\widetilde{\eta}_\alpha = g(\boldsymbol{x}^\dagger) + N\big(\sqrt{16\log(3/\alpha)/N} + 1\big)$, and where $\mathcal{B}_1(\boldsymbol{0}, \widetilde{\eta}_\alpha/\lambda)$ denotes the ℓ_1 ball, centred in $\boldsymbol{0}$, with radius $\widetilde{\eta}_\alpha/\lambda$. Observe that $\widetilde{\mathcal{C}}_\alpha$ is convex. This property will play a central role in the proposed methodology.

Intuitively, if $\mathcal{S} \cap \widetilde{\mathcal{C}}_\alpha = \varnothing$ then $\mathcal{S} \subset \mathbb{R}^N \setminus \widetilde{\mathcal{C}}_\alpha$. And because $\mathsf{P}(\overline{\boldsymbol{x}} \in \widetilde{\mathcal{C}}_\alpha|\boldsymbol{y}) \geqslant 1-\alpha$, if $\mathcal{S} \cap \widetilde{\mathcal{C}}_\alpha = \varnothing$, then $\mathsf{P}(\overline{\boldsymbol{x}} \in \mathcal{S}|\boldsymbol{y}) \leqslant \alpha$. Consequently, the Bayesian hypothesis test described above can be reformulated as follows:

Theorem 3.1 (Thm. 3.2 in [15]): For every $\alpha \in]4\exp(-N/3), 1[$, if $\mathcal{S} \cap \widetilde{\mathcal{C}}_\alpha = \varnothing$, then $\mathsf{P}(H_0|\boldsymbol{y}) \leqslant \alpha$.

This theorem ensures that, if there is no intersection between the sets $\mathcal{S}$ and $\widetilde{\mathcal{C}}_\alpha$, then H_0 is rejected at level α. Conversely, if $\mathcal{S} \cap \widetilde{\mathcal{C}}_\alpha \neq \varnothing$ we then fail to reject H_0 at level α.

B. Minimization problem

In this work, we focus on compact (i.e., spatially localized) image structures. To define mathematically this type of structures, we need to introduce the selection matrix $\boldsymbol{M} \in \{0,1\}^{N_{\boldsymbol{M}}\times N}$, such that, for an image $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{M}\boldsymbol{x}$ corresponds to the region where the structure of interest is localized. Then, the set $\mathcal{S}$ is defined as follows. Firstly, we impose $\mathcal{S}$ to be a subset of the intensity images of $\mathbb{R}^N$, using a positivity constraint. In addition, to smooth the area of the structure, we use a linear inpainting technique to fill the pixels in $\boldsymbol{M}\boldsymbol{x}$ with the information contained in the other pixels $\boldsymbol{M}^c\boldsymbol{x}$ of the image $\boldsymbol{x}$, where $\boldsymbol{M}^c \in \{0,1\}^{(N-N_{\boldsymbol{M}})\times N}$. Formally, we use the constraint

$$\boldsymbol{M}\boldsymbol{x} = \boldsymbol{L}\boldsymbol{M}^c\boldsymbol{x} + \boldsymbol{\tau}, \text{ with } \boldsymbol{\tau} \in [-\tau, +\tau]^{N_{\boldsymbol{M}}}, \tag{6}$$

where $\boldsymbol{L} \in \mathbb{R}^{N_M\times(N-N_M)}$ is an inpainting operator and $\tau = 10^{-2}\|\boldsymbol{L}\boldsymbol{M}^c\boldsymbol{x}^\dagger\|_2/N_{\boldsymbol{M}}$. This inpainting can potentially amplify the energy in $\boldsymbol{M}\boldsymbol{x}$, and lead to artificial structures. To prevent this, we constrain the energy of the area of interest $\boldsymbol{M}\boldsymbol{x}$ to be smaller than $\theta = (1+10^{-2})\|\boldsymbol{L}\boldsymbol{M}^c\boldsymbol{x}^\dagger\|_2$. Mathematically, this constraint can be written as

$$\boldsymbol{M}\boldsymbol{x} \in \mathcal{B}_2(\boldsymbol{0}, \theta). \tag{7}$$

The resulting set $\mathcal{S}$ of images not containing the compact structure of interest is given by

$$\mathcal{S} = \big\{\boldsymbol{x} \in [0,+\infty[^N \;\big|\; \big(\boldsymbol{M} - \boldsymbol{L}\boldsymbol{M}^c\big)\boldsymbol{x} \in [-\tau,+\tau]^{N_{\boldsymbol{M}}} \text{ and } \boldsymbol{M}\boldsymbol{x} \in \mathcal{B}_2(\boldsymbol{0},\theta)\big\}. \tag{8}$$

It is interesting to note that if $\mathrm{dist}(\mathcal{S}, \widetilde{\mathcal{C}}_\alpha) > 0$, where $\mathrm{dist}(\mathcal{S}, \widetilde{\mathcal{C}}_\alpha) = \inf\Big\{\|\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{x}_{\mathcal{C}_\alpha}\|_2 \;:\; \big(\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}\big) \in \mathcal{S}\times\widetilde{\mathcal{C}}_\alpha\Big\}$,

[2]For a non-empty, closed convex subset $\mathcal{B}$ of $\mathbb{R}^N$, for every $\boldsymbol{x} \in \mathbb{R}^N$, $\iota_{\mathcal{B}}(\boldsymbol{x})$ is equal to 0 if $\boldsymbol{x} \in \mathcal{B}$, and $+\infty$ otherwise.

ISBN 978-90-827970-1-5 © EURASIP 2018

then we conclude that $\mathcal{S} \cap \widetilde{\mathcal{C}}_\alpha = \varnothing$, and consequently H_0 is rejected with significance α. Otherwise, if $\text{dist}(\mathcal{S}, \widetilde{\mathcal{C}}_\alpha) = 0$, we deduce that $\mathcal{S} \cap \widetilde{\mathcal{C}}_\alpha \neq \varnothing$. By combining the definition of the distance, and the definitions of the sets $\widetilde{\mathcal{C}}_\alpha$ and $\mathcal{S}$, respectively given in (5) and (8), the problem of determining if $\mathcal{S} \cap \widetilde{\mathcal{C}}_\alpha = \varnothing$ can be equivalently reformulated as follows:

$$\underset{(\boldsymbol{x}_{\mathcal{S}}, \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}) \in [0,+\infty[^N \times [0,+\infty[^N}{\text{minimize}} \ \frac{\gamma}{2} \|\boldsymbol{x}_{\mathcal{S}} - \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}\|_2^2$$
$$\text{s.t.} \ \begin{cases} \overline{\boldsymbol{L}} \boldsymbol{x}_{\mathcal{S}} \in [-\tau, +\tau]^{N_{\mathcal{M}}}, \ \boldsymbol{M} \boldsymbol{x}_{\mathcal{S}} \in \mathcal{B}_2(\boldsymbol{0}, \theta) \\ \boldsymbol{\Phi} \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha} \in \mathcal{B}_2(\boldsymbol{y}, \varepsilon), \ \boldsymbol{\Psi} \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha} \in \mathcal{B}_1(\boldsymbol{0}, \widetilde{\eta}_\alpha / \lambda) \end{cases} \tag{9}$$

with $\gamma > 0$ and $\overline{\boldsymbol{L}} = \boldsymbol{M} - \boldsymbol{L}\boldsymbol{M}^c$.

C. Proposed algorithm

To solve this minimization problem, we leverage the primal-dual forward-backward algorithm [5], [6], [9], described in Algorithm 1. At each iteration $k \in \mathbb{N}$, in addition to minimize the distance between $\boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k+1)}$ and $\boldsymbol{x}_{\mathcal{S}}^{(k+1)}$, projections are performed onto the different sets defined by the constraints appearing in problem (9). The projection of $\boldsymbol{x} \in \mathbb{R}^N$ onto a non-empty closed convex subset $\mathcal{C}$ of $\mathbb{R}^N$ is given by $\Pi_{\mathcal{C}}(\boldsymbol{x}) = \underset{\widetilde{\boldsymbol{x}} \in \mathcal{C}}{\text{argmin}} \ \|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\|_2$ [10]. In Algorithm 1, the 5 first equations are dedicated to the update of variable $\boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)}$, while the 5 last equations are dedicated to the update of variable $\boldsymbol{x}_{\mathcal{S}}^{(k)}$. The dual variables $\boldsymbol{v}_1^{(k)}$ and $\boldsymbol{v}_2^{(k)}$ are related to the primal variable $\boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)}$, while $\boldsymbol{u}_1^{(k)}$ and $\boldsymbol{u}_2^{(k)}$ are related to $\boldsymbol{x}_{\mathcal{S}}^{(k)}$. In particular, $\boldsymbol{v}_1^{(k)}$ and $\boldsymbol{v}_2^{(k)}$ are associated with the ℓ_2 ball and the ℓ_1 ball constraints induced by $\widetilde{\mathcal{C}}_\alpha$, respectively. Similarly, $\boldsymbol{u}_1^{(k)}$ and $\boldsymbol{u}_2^{(k)}$ are associated with the constraints (6) and (7), respectively. According to [9], if the following condition holds:

$$\frac{1}{\zeta} - \left(\mu_{1,1} \|\boldsymbol{\Phi}\|_S^2 + \mu_{1,2} \|\boldsymbol{\Psi}\|_S^2 + \mu_{2,1} \|\overline{\boldsymbol{L}}\|_S^2 + \mu_{2,2}\right) > \gamma, \tag{10}$$

where $\|.\|_S$ denotes the spectral norm, then the sequence $\left(\boldsymbol{x}_{\mathcal{S}}^{(k)}, \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)}\right)_{k \in \mathbb{N}}$ converges to a solution $\left(\boldsymbol{x}_{\mathcal{S}}^{\ddagger}, \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{\ddagger}\right)$ to problem (9).

Note that the choice of γ does not act on the solution to the minimization problem. However, this parameter can be used in practice to accelerate the convergence of the algorithm.

D. Uncertainty quantification parameter

By definition, the solution $\left(\boldsymbol{x}_{\mathcal{S}}^{\ddagger}, \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{\ddagger}\right)$ generated by Algorithm 1 satisfies

$$\|\boldsymbol{x}_{\mathcal{S}}^{\ddagger} - \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{\ddagger}\|_2 = \text{dist}(\mathcal{S}, \widetilde{\mathcal{C}}_\alpha). \tag{11}$$

To relate this quantity to the intensity of the structure of interest, it is useful to introduce the image $\boldsymbol{x}_{\mathcal{S}}^{\dagger} \in \mathcal{S}$. This image corresponds to a modified version of the MAP estimate $\boldsymbol{x}^{\dagger}$, where the structure has been removed. Formally, we define $\boldsymbol{M}^c \boldsymbol{x}_{\mathcal{S}}^{\dagger} = \boldsymbol{M}^c \boldsymbol{x}^{\dagger}$ and $\boldsymbol{M} \boldsymbol{x}_{\mathcal{S}}^{\dagger} = \boldsymbol{L}\boldsymbol{M}^c \boldsymbol{x}^{\dagger}$. Then, $\|\boldsymbol{x}_{\mathcal{S}}^{\dagger} - \boldsymbol{x}^{\dagger}\|_2$ is related to the intensity of the structure of interest, as it appears in the MAP estimate. To put these quantities on a meaningful scale we define the normalized parameter

$$\rho_\alpha = \frac{\|\boldsymbol{x}_{\mathcal{S}}^{\ddagger} - \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{\ddagger}\|_2}{\|\boldsymbol{x}_{\mathcal{S}}^{\dagger} - \boldsymbol{x}^{\dagger}\|_2}. \tag{12}$$

This parameter guarantees the structure's minimum intensity at level $1 - \alpha$, and its intensity as measured on the MAP estimate. In addition, if $\rho_\alpha > 0$, then the hypothesis H_0 is rejected.

Note that, in practice, we consider that Algorithm 1 has converged when all the constraints in (9) are satisfied (up to a tolerance), and the following stopping criteria are satisfied:

$$\begin{cases} \|\boldsymbol{x}_{\mathcal{S}}^{(k+1)} - \boldsymbol{x}_{\mathcal{S}}^{(k)}\|_2 \leqslant 10^{-5} \|\boldsymbol{x}_{\mathcal{S}}^{(k+1)}\|_2, \\ \|\boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k+1)} - \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)}\|_2 \leqslant 10^{-5} \|\boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k+1)}\|_2. \end{cases} \tag{13}$$

These stopping criteria does not allow to reach exactly $\rho_\alpha = 0$. Consequently, in our simulations we will consider that H_0 is rejected when $\rho_\alpha > 2\%$, to allow for numerical errors.

Algorithm 1 Primal-dual algorithm to solve (9).

Initialization: Let $\boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(0)} \in [0,+\infty[^N$, $\boldsymbol{x}_{\mathcal{S}}^{(0)} \in [0,+\infty[^N$, $\boldsymbol{v}_1^{(0)} \in \mathbb{R}^M$, $\boldsymbol{v}_2^{(0)} \in \mathbb{R}^N$, $\boldsymbol{u}_1^{(0)} \in \mathbb{R}^{N_{\mathcal{M}}}$ and $\boldsymbol{u}_2^{(0)} \in \mathbb{R}^{N_{\mathcal{M}}}$.
Let $(\mu_{1,1}, \mu_{1,2}, \mu_{2,1}, \mu_{2,2}, \zeta) \in [0,+\infty[^5$.
Iterations:
for $k = 0, 1, \ldots$

$$\widetilde{\boldsymbol{v}}_1^{(k)} = \boldsymbol{v}_1^{(k)} + \mu_{1,1} \boldsymbol{\Phi} \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)}$$
$$\boldsymbol{v}_1^{(k+1)} = \widetilde{\boldsymbol{v}}_1^{(k)} - \mu_{1,1} \Pi_{\mathcal{B}_2(\boldsymbol{y}, \varepsilon)}\left(\mu_{1,1}^{-1} \widetilde{\boldsymbol{v}}_1^{(k)}\right)$$
$$\widetilde{\boldsymbol{v}}_2^{(k)} = \boldsymbol{v}_2^{(k)} + \mu_{1,2} \boldsymbol{\Psi}^\dagger \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)}$$
$$\boldsymbol{v}_2^{(k+1)} = \widetilde{\boldsymbol{v}}_2^{(k)} - \mu_{1,2} \Pi_{\mathcal{B}_1(\boldsymbol{0}, \widetilde{\eta}_\alpha/\lambda)}\left(\mu_{1,2}^{-1} \widetilde{\boldsymbol{v}}_2^{(k)}\right)$$
$$\boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k+1)} = \Pi_{[0,+\infty[^N}\Big((1 - \gamma\zeta) \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)} + \gamma\zeta \boldsymbol{x}_{\mathcal{S}}^{(k)} - \zeta \boldsymbol{\Phi}^\dagger (2\boldsymbol{v}_1^{(k+1)} - \boldsymbol{v}_1^{(k)}) - \zeta \boldsymbol{\Psi} (2\boldsymbol{v}_2^{(k+1)} - \boldsymbol{v}_2^{(k)})\Big)$$
$$\widetilde{\boldsymbol{u}}_1^{(k)} = \boldsymbol{u}_1^{(k)} + \mu_{2,1} \overline{\boldsymbol{L}} \boldsymbol{x}_{\mathcal{S}}^{(k)}$$
$$\boldsymbol{u}_1^{(k+1)} = \widetilde{\boldsymbol{u}}_1^{(k)} - \mu_{2,1} \Pi_{[-\tau, +\tau]^{N_{\mathcal{M}}}}\left(\mu_{2,1}^{-1} \widetilde{\boldsymbol{u}}_1^{(k)}\right)$$
$$\widetilde{\boldsymbol{u}}_2^{(k)} = \boldsymbol{u}_2^{(k)} + \mu_{2,2} \boldsymbol{M} \boldsymbol{x}_{\mathcal{S}}^{(k)}$$
$$\boldsymbol{u}_2^{(k+1)} = \widetilde{\boldsymbol{u}}_2^{(k)} - \mu_{2,2} \Pi_{\mathcal{B}_2(\boldsymbol{0}, \theta)}\left(\mu_{2,2}^{-1} \widetilde{\boldsymbol{u}}_2^{(k)}\right)$$
$$\boldsymbol{x}_{\mathcal{S}}^{(k+1)} = \Pi_{[0,+\infty[^N}\Big((1 - \gamma\zeta) \boldsymbol{x}_{\mathcal{S}}^{(k)} + \gamma\zeta \boldsymbol{x}_{\widetilde{\mathcal{C}}_\alpha}^{(k)} - \zeta \overline{\boldsymbol{L}}^\dagger (2\boldsymbol{u}_1^{(k+1)} - \boldsymbol{u}_1^{(k)}) - \zeta \boldsymbol{M}^\dagger (2\boldsymbol{u}_2^{(k+1)} - \boldsymbol{u}_2^{(k)})\Big)$$

IV. Application to radio-interferometry

In the context of radio-interferometric imaging, each measurement $y_m \in \mathbb{C}$, called *visibility*, is acquired by an antenna pair, at a given time integration $t \in \{1, \ldots, T\}$, in the Fourier domain of the intensity image of interest $\overline{\boldsymbol{x}}$. Visibilities are related to an undersampled selection of the Fourier coefficients of the intensity image degraded by antenna gains [20]. At each time integration t, $n_a(n_a - 1)/2$ complex visibilities are acquired by the n_a antenna pairs. In total, the measurement vector $\boldsymbol{y}$ consists in $M = T n_a (n_a - 1)/2$ visibilities. In our simulations, we consider the antenna configuration of the *Very Large Array* (VLA) telescope, constituted by $n_a = 27$ antennas, with different values of $T \in \{10, \ldots, 500\}$. An example of $|\boldsymbol{y}|$, for $T = 50$ is shown in Fig. 1 (top). In this image, the

ISBN 978-90-827970-1-5 © EURASIP 2018

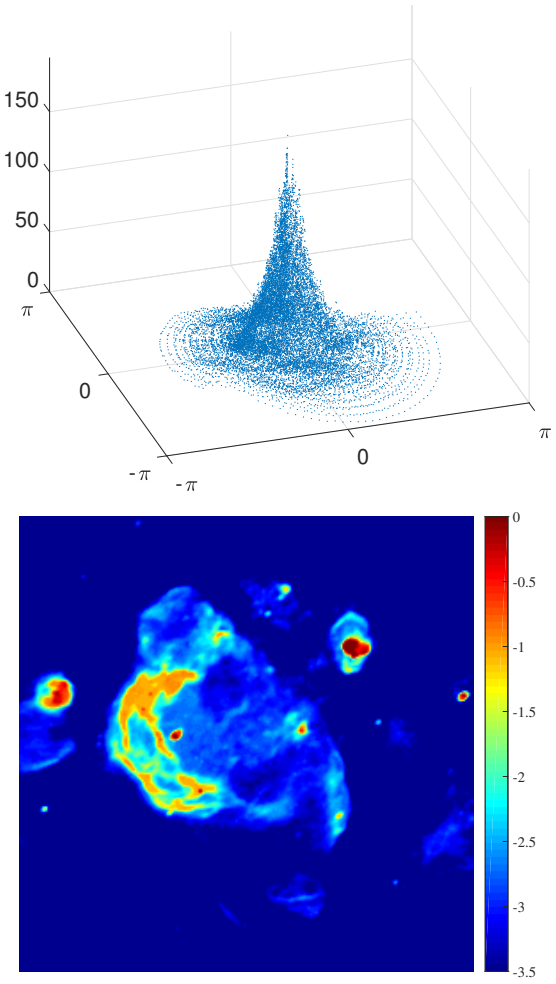

Fig. 1. Top: Visibility magnitudes $|\boldsymbol{y}|$ considering the VLA telescope, $T = 50$ integrations, and a time interval of 12 hours, with normalized continuous u-v plane. Bottom: Original image of the W28 supernova, in log-scale.

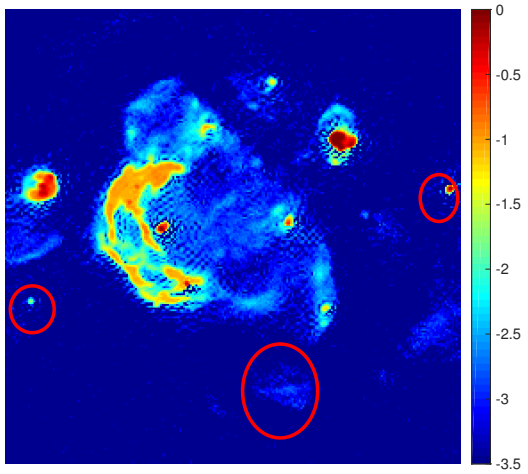

Fig. 2. MAP estimate obtained considering $T = 50$ integrations. The three structures highlighted in red show, from left to right, *Struct. 1*, *Struct. 2* and *Struct. 3*.

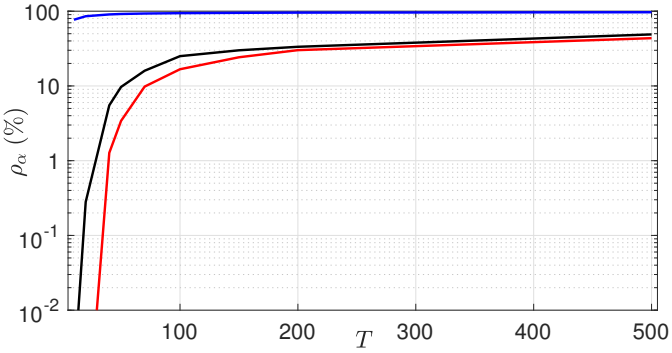

Fig. 3. Curves representing the values of ρ_α for $\alpha = 1\%$, in percentage, in log-scale, as a function of $T \in \{10, \ldots, 500\}$, for the three different structures of interest. *Struct. 1*: red curve, *Struct. 2*: black curve, and *Struct. 3*: blue curve

continuous Fourier space (also called *u-v plane*) is normalized such that the spatial frequencies belong to $[-\pi, \pi]^2$. We are interested in estimating the original sky brightness distribution for the W28 supernova image ($N = 256 \times 256$), displayed in Fig. 1 (bottom) with a log scale. The radio-interferometric imaging problem can be formulated following model (1) [17], [21], [22], where the observation matrix $\boldsymbol{\Phi}$ corresponds to an undersampling non-uniform Fourier operator [23].

The MAP estimate of $\overline{\boldsymbol{x}}$ is obtained by solving problem (2). The value of ε is chosen according to the noise level. In radio interferometry, the residual norm being distributed according to a χ^2 distribution with $2M$ degrees of freedom, we use $\varepsilon = \sigma(2M + 2\sqrt{4M})^{1/2}$ to impose that ε^2 is 2 standard deviations above the mean of the χ^2 distribution. In our simulations, we consider the noise variance to be equal to $\sigma^2 = 0.01$. Concerning the prior, we set $\boldsymbol{\Psi}$ to the Daubechies wavelet transform Db8 [19], [24].

We propose to quantify the uncertainty of 3 different structures appearing in the MAP estimate. Fig. 2 shows the MAP estimate obtained when considering $T = 50$ integrations. The three structures of interest are highlighted in red in this image and denoted, from left to right, *Struct. 1*, *Struct. 2* and *Struct. 3*. *Struct. 1* and *Struct. 3* represent point sources with different amplitudes, while *Struct. 2* is a low amplitude slightly extended source. In our simulations, we consider $\alpha = 1\%$.

The results obtained using the proposed uncertainty quantification method are summarized in Fig. 3. It shows three curves, representing the values of ρ_α in percentage (in log-scale) as a function of the time integration $T \in \{10, \ldots, 500\}$, for the three considered structures. In particular, the red curve is associated with *Struct. 1*, the black curve with *Struct. 2*, and the blue curve with *Struct. 3*. Note that, in Figure 3, the curves of *Struct. 1* and *Struct. 2* are truncated for display purpose. Their values of ρ_α reach $10^{-9}\%$ and $10^{-3}\%$ at $T = 10$, respectively. The three considered structures belonging to original image $\overline{\boldsymbol{x}}$, it is not surprising to see that their corresponding ρ_α increase with T (i.e., with the number of measurements). In other words, the structure's minimum energy tends to be higher when the number of measurement increases. For *Struct. 3*, we have $\rho_\alpha \geqslant 70\%$ for all the considered values of $T \in \{10, \ldots, 500\}$. In this case,

ISBN 978-90-827970-1-5

at least 70% of the structure's minimum energy is guaranteed at level $1-\alpha = 99\%$. Therefore, H_0 is always rejected for *Struct. 3*, for all the considered measurement configurations. Concerning both *Struct. 1* and *Struct. 2*, the conclusion are more nuanced. They will depend on the the tolerance chosen to consider that $\rho_\alpha \sim 0$. In our case, we have chosen that when $\rho_\alpha > 2\%$, H_0 is rejected. Then, for *Struct. 1* (resp. *Struct. 2*), for $T \geqslant 45$ (resp. $T \geqslant 35$) we conclude that H_0 is rejected, while for $T < 45$ (resp. $T < 35$) it can be deduced that H_0 cannot be rejected. The uncertainty related to these two structures is thus more dependent of T.

V. Conclusions

In this work, we have presented a new Bayesian uncertainty quantification method, combining the recent Bayesian inference results presented in [11] and advanced optimization techniques. In particular, we propose to perform Bayesian hypothesis testing on the MAP estimate, using a primal-dual algorithm. This method is mainly based on our recent work presented in [15], where a POCS approach was adopted. We have shown through simulation results that the proposed approach can be used for uncertainty quantification of particular structures appearing on radio sky images. We emphasize that quantifying the uncertainty associated with a particular image structure is of great interest for astronomers; for example, to validate particular emissions appearing in the reconstructed sky image. In radio-interferometric imaging, the next generation telescope Square Kilometre Array (SKA) will provide giga-pixel images with unprecedented dynamic range [16]–[18], and which will require uncertainty quantification methods that can scale efficiently. The proposed algorithmic structure leverages optimization tools which are known to be scalable and adapted to solve such large scale minimization problems [2], [3], [9], [17].

References

[1] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[2] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2010.

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

[4] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical imaging and vision*, vol. 40, no. 1, pp. 120–145, 2011.

[5] 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.

[6] Bang Công Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Advances in Computational Mathematics*, vol. 38, no. 3, pp. 667–681, 2013.

[7] P. Ochs, Y. Chen, T. Brox, and T. Pock, "iPiano: inertial proximal algorithm for non-convex optimization," *SIAM J. Imag. Sci.*, vol. 7, no. 2, pp. 1388–1419, 2014.

[8] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function," *J. Optim. Theory Appl.*, vol. 162, no. 1, Jul. 2014.

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

[10] Heinz H Bauschke and Patrick L Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer, 2017.

[11] M. Pereyra, "Maximum-a-posteriori estimation with bayesian confidence regions," *SIAM Journal on Imaging Sciences*, vol. 10, no. 1, pp. 285–302, 2017.

[12] X. Cai, M. Pereyra, and J. D. McEwen, "Uncertainty quantification for radio interferometric imaging: Ii. map estimation," Tech. Rep., Nov. 2017, arXiv:1711.04819.

[13] Alain Durmus, Eric Moulines, and Marcelo Pereyra, "Efficient Bayesian computation by proximal Markov chain Monte Carlo: when Langevin meets Moreau," *arXiv:1612.07471*, 2016.

[14] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. O. Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 224–241, 2016.

[15] A. Repetti, M. Pereyra, and Y. Wiaux, "Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimization," Tech. Rep., Jan. 2018.

[16] M. P. Van Haarlem, M. W. Wise, A. W. Gunst, G. Heald, J. P. McKean, J. W. T. Hessels, A. G. De Bruyn, R. Nijboer, J. Swinbank, R. Fallows, et al., "LOFAR: The low-frequency array," *Astronomy & Astrophysics*, vol. 556, pp. A2, 2013.

[17] 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.

[18] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio, "The Square Kilometre Array," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1482–1496, 2009.

[19] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, Burlington, MA, 2rd edition, 2009.

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

[21] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, and P. Vandergheynst, "Compressed sensing for radio interferometry: prior-enhanced basis pursuit imaging techniques," in *SPARS 09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[22] A. Repetti, J. Birdi, A. Dabbech, and Y. Wiaux, "Non-convex optimization for self-calibration of directiondependent effects in radio interferometric imaging," *Mon. Not. R. Astron. Soc.*, vol. 470, no. 4, pp. 3981–4006, Oct. 2017.

[23] F. A. Fessler and B. P. Sutton, "Nonuniform fast fourier transforms using min-max interpolation," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 560–574, 2003.

[24] I. Daubechies, *Ten lectures on wavelets*, CBS-NSF Regional Conferences in Applied Mathematics, 61, SIAM, Philadelphia, 1992.

ISBN 978-90-827970-1-5