Bayesian estimation of regularization and point spread function parameters for Wiener–Hunt deconvolution

François Orieux,[1,*] Jean-François Giovannelli,[2] and Thomas Rodet[1]

[1]*Laboratoire des Signaux et Systèmes (CNRS–SUPELEC–Univ. Paris-Sud 11), SUPELEC, Plateau de Moulon,*
3 rue Joliot-Curie, 91 192 Gif-sur-Yvette, France
[2]*Laboratoire d'Intégration du Matériau au Système (CNRS–ENSEIRB–Univ. Bordeaux 1–ENSCPB),*
351 cours de la Liberation, 33405 Talence, France
**Corresponding author: orieux@lss.supelec.fr*

Received October 21, 2009; revised March 5, 2010; accepted April 15, 2010;
posted April 30, 2010 (Doc. ID 118863); published June 9, 2010

This paper tackles the problem of image deconvolution with joint estimation of point spread function (PSF) parameters and hyperparameters. Within a Bayesian framework, the solution is inferred via a global *a posteriori* law for unknown parameters and object. The estimate is chosen as the posterior mean, numerically calculated by means of a Monte Carlo Markov chain algorithm. The estimates are efficiently computed in the Fourier domain, and the effectiveness of the method is shown on simulated examples. Results show precise estimates for PSF parameters and hyperparameters as well as precise image estimates including restoration of high frequencies and spatial details, within a global and coherent approach. © 2010 Optical Society of America

OCIS codes: 100.1830, 100.3020, 100.3190, 150.1488.

1. INTRODUCTION

Image deconvolution has been an active research field for several decades, and recent contributions can be found in papers such as [1–3]. Examples of application are medical imaging, astronomy, nondestructive testing, and, more generally, imagery problems. In these applications, degradations induced by the observation instrument limit the data resolution, while the need of precise interpretation can be of major importance. For example, this is particularly critical for long-wavelength astronomy (see e.g., [4]). In addition, the development of a high-quality instrumentation system must be completed rationally at an equivalent level of quality in the development of data processing methods. Moreover, even for poor performance systems, the restoration method can be used to bypass instrument limitations.

When the deconvolution problem is ill-posed, a possible solution relies on regularization, i.e., introduction of information in addition to the data and the acquisition model [5,6]. As a consequence of regularization, deconvolution methods are specific to the class of image in accordance with the introduced information. From this standpoint, the present paper is dedicated to relatively smooth images encountered for numerous applications in imagery [4,7,8]. The second-order consequence of ill-posedness and regularization is the need to balance the compromise between different sources of information.

In the Bayesian approach [1,9], information about unknowns is introduced by means of probabilistic models. Once these models are designed, the next step is to build the *a posteriori* law, given the measured data. The solution is then defined as a representative point of this law; the two most representative points are (1) the maximizer and (2) the mean. From a computational standpoint, the first leads to a numerical optimization problem and the latter leads to a numerical integration problem. However, the resulting estimate depends on two sets of variables in addition to the data:

1. First, the estimate naturally depends on the response of the instrument at work, namely, the point spread function (PSF). The literature is devoted to deconvolution predominantly in the case of known PSF. In contrast, the present paper is devoted to the case of unknown or poorly known PSF, and there are two main strategies to tackle its estimation from the available data set (without extra measurements):

   (i) In most practical cases, the instrument can be modeled using physical operating description. It is thus possible to find the equation for the PSF, at least in a first approximation. This equation is usually driven by a relatively small number of parameters. It is a common case in optical imaging, where a Gaussian-shaped PSF is often used [10]. It is also the case in other fields: interferometry [11], magnetic resonance force microscopy [12], fluorescence microscopy [13], etc. Nevertheless, in real experiments, the parameter values are unknown or imperfectly known and need to be estimated or adjusted in addition to the image of interest: the problem is called *myopic* deconvolution.

   (ii) The second strategy forbids the use of the parametric PSF deduced from the physical analysis, and the PSF then naturally appears in a non-parametric form. Practically, the non-parametric PSF is unknown or imperfectly known and needs to be estimated in addition to the image of interest: the problem is referred to as *blind* deconvolution, for example in interferometry [14–17].

From an inference point of view, the difficulty of both myopic and blind problems lies in the possible lack of information, resulting in ambiguity between image and PSF, even in the noiseless case. In order to resolve the ambiguity, information must be added [3,18], and it is crucial to make inquiries based on any available source of information. To this end, the knowledge of the parametric PSF represents a valuable means to structure the problem and possibly resolve the degeneracies. Moreover, due to the instrument design process, a nominal value as well as an uncertainty are usually available for the PSF parameters. In addition, from a practical and algorithmic standpoint, the myopic case, i.e., the case of parametric PSF, is often more difficult due to the non-linear dependence of the observation model with respect to the PSF parameters. In contrast, the blind case, i.e., the case of non-parametric PSF, yields a simpler practical and algorithmic problem since the observation model remains linear w.r.t. the unknown elements given the object. Despite the superior technical difficulty, the present paper is devoted to the myopic format since it is expected to be more efficient than the blind format from an information standpoint. Moreover, the blind case has been extensively studied, and a large number of papers are available [19–21], while the myopic case has been less investigated, though it is of major importance.

2. Second, the solution depends on the probability law parameters called hyperparameters (means, variances, parameters of correlation matrix, etc.). These parameters adjust the shape of the laws, and at the same time they tune the compromise between the information provided by the *a priori* knowledge and the information provided by the data. In real experiments, their values are unknown and need to be estimated: the problem is called *unsupervised* deconvolution.

For both families of parameters (PSF parameters and hyperparameters), two approaches are available. In the first one, the parameter values are empirically tuned or estimated in a preliminary step (with maximum likelihood [7] or calibration [22] for example); then the values are used in a second step devoted to image restoration given the parameters. In the second step, the parameters and the object are jointly estimated [2,19].

For the myopic problem, Jalobeanu *et al.* [23] address the case of a symmetric Gaussian PSF. The width parameter and the noise variance are estimated in a preliminary step by maximum likelihood. A recent paper [24] addresses the estimation of a Gaussian blur parameter, as in our experiment, with an empirical method. They found the Gaussian blur parameter by minimizing the absolute value of the second derivative (Laplacian) of the restored images.

The present paper addresses the myopic and unsupervised deconvolution problem. We propose a new method that jointly estimates the PSF parameters, the hyperparameters, and the image of interest. It is built in a coherent and global framework based on an extended *a posteriori* law for all the unknown variables. The posterior law is obtained via the Bayes rule, founded on *a priori* laws: Gaussian for image and noise, uniform for PSF parameters, and gamma or Jeffreys for hyperparameters.

Regarding the image prior law, we have paid special attention to the parameterization of the covariance matrix in order to facilitate law manipulations such as integration, conditioning, and hyperparameter estimation. The possible degeneracy of the *a posteriori* law in some limit cases is also studied.

The estimate is chosen as the mean of the posterior law and is computed using Monte Carlo simulations. To this end, Monte Carlo Markov chain (MCMC) algorithms [25] enable one to draw samples from the posterior distribution despite its complexity and especially the non-linear dependence w.r.t. the PSF parameters.

The paper is structured in the following manner. Section 2 presents the notation and describes the problem. The three following sections describe our methodology: first the Bayesian probabilistic models are detailed in Section 3, then a proper posterior law is established in Section 4; and an MCMC algorithm to compute the estimate is described in Section 5. Numerical results are presented in Section 6. Finally, Section 7 is devoted to the conclusion and perspectives.

2. NOTATION AND CONVOLUTION MODEL

Consider N-pixel real square images represented in lexicographic order by vector $\boldsymbol{x} \in \mathbb{R}^N$, with generic elements x_n. The forward model is written

$$\boldsymbol{y} = \boldsymbol{H}_{\boldsymbol{w}}\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the vector of data, $\boldsymbol{H}_{\boldsymbol{w}}$ a convolution matrix, $\boldsymbol{x}$ the image of interest, and $\boldsymbol{\epsilon}$ the modelization errors or the noise. Vector $\boldsymbol{w} \in \mathbb{R}^P$ stands for the PSF parameters, such as width or orientation of a Gaussian PSF.

The matrix $\boldsymbol{H}_{\boldsymbol{w}}$ is block-circulant with circulant-block (BCCB) for computational efficiency of the convolution in the Fourier space. The diagonalization [26] of $\boldsymbol{H}_{\boldsymbol{w}}$ is written $\boldsymbol{\Lambda}_{\boldsymbol{H}}=\boldsymbol{F}\boldsymbol{H}_{\boldsymbol{w}}\boldsymbol{F}^{\dagger}$, where $\boldsymbol{F}$ is the unitary Fourier matrix and $\dagger$ is the transpose conjugate symbol. The convolution in the Fourier space, is then

$$\mathring{\boldsymbol{y}} = \boldsymbol{\Lambda}_{\boldsymbol{H}}\mathring{\boldsymbol{x}} + \mathring{\boldsymbol{\epsilon}}, \tag{2}$$

where $\mathring{\boldsymbol{x}}=\boldsymbol{F}\boldsymbol{x}$, $\mathring{\boldsymbol{y}}=\boldsymbol{F}\boldsymbol{y}$ and $\mathring{\boldsymbol{\epsilon}}=\boldsymbol{F}\boldsymbol{\epsilon}$ are the 2D discrete Fourier transform (DFT-2D) of image, data, and noise, respectively.

Since $\boldsymbol{\Lambda}_{\boldsymbol{H}}$ is diagonal, the convolution is computed with a term-wise product in the Fourier space. There is a strict equivalence between a description in the spatial domain [Eq. (1)] and in the Fourier domain [Eq. (2)]. Consequently, for coherent description and computational efficiency, all the developments are equally done in the spatial space or in the Fourier space.

For notational convenience, let us introduce the component at null frequency $\mathring{x}_0 \in \mathbb{R}$ and the vector of component at non-null frequencies $\mathring{\boldsymbol{x}}_* \in \mathbb{C}^{N-1}$ so that the whole set of components is written $\mathring{\boldsymbol{x}}=[\mathring{x}_0,\mathring{\boldsymbol{x}}_*]$.

Let us denote $\mathbb{1}$ the vector of N components equal to $1/N$, so that $\mathbb{1}^{\mathrm{t}}\boldsymbol{x}$ is the empirical mean level of the image. The Fourier components are the $\mathring{\mathbb{1}}_n$, and we have $\mathring{\mathbb{1}}_0=1$ and

$\mathring{\mathbb{1}}_n = 0$ for $n \neq 0$. Moreover, $\boldsymbol{\Lambda}_{\mathbb{1}} = \boldsymbol{F}\mathbb{1}\mathbb{1}^{\mathrm{t}}\boldsymbol{F}^{\dagger}$ is a diagonal matrix with only one non-null coefficient at null frequency.

3. BAYESIAN PROBABILISTIC MODEL

This section presents the prior law for each set of parameters. Regarding the image of interest, in order to account for smoothness, the law introduces high-frequency penalization through a differential operator on the pixel. A conjugate law is proposed for the hyperparameters, and a uniform law is considered for the PSF parameters.

Moreover, we have paid special attention to the image prior law parameterization. In Subsection 3.A we present several parameterizations in order to facilitate law manipulations such as integration, conditioning, and hyperparameter estimation. Moreover, the correlation matrix of the image law may become singular in some limit cases, resulting in a degenerated prior law [when $p(\boldsymbol{x}) = 0$ for all $\boldsymbol{x} \in \mathbb{R}^N$]. Based on this parameterization, in Section 4 we study the degeneracy of the posterior in relation to the parameters of the prior law.

A. Image Prior Law

The probability law for the image is a Gaussian field with a given precision matrix $\boldsymbol{P}$ parameterized by a vector γ. The pdf reads

$$p(\boldsymbol{x}|\boldsymbol{\gamma}) = (2\pi)^{-N/2}\det[\boldsymbol{P}]^{1/2}\exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathrm{t}}\boldsymbol{P}\boldsymbol{x}\right]. \qquad (3)$$

For computational efficiency, the precision matrix is designed (or approximated) in a toroidal manner, and it is diagonal in the Fourier domain $\boldsymbol{\Lambda_P} = \boldsymbol{FPF}^{\dagger}$. Thus, the law for $\boldsymbol{x}$ also is written

$$p(\boldsymbol{x}|\boldsymbol{\gamma}) = (2\pi)^{-N/2}\det[\boldsymbol{F}]\det[\boldsymbol{\Lambda_P}]^{1/2}\det[\boldsymbol{F}^{\dagger}] \times \exp\left[-\frac{1}{2}\boldsymbol{x}^{\mathrm{t}}\boldsymbol{F}^{\dagger}\boldsymbol{\Lambda_P}\boldsymbol{F}\boldsymbol{x}\right], \qquad (4)$$

$$= (2\pi)^{-N/2}\det[\boldsymbol{\Lambda_P}]^{1/2}\exp\left[-\frac{1}{2}\mathring{\boldsymbol{x}}^{\dagger}\boldsymbol{\Lambda_P}\mathring{\boldsymbol{x}}\right], \qquad (5)$$

and it is sometimes referred to [27] as a Whittle approximation (see also [28], p. 133) for the Gaussian law. The filter obtained for fixed hyperparameters is also the Wiener–Hunt filter [29], as described in Subsection 5.A, below.

This paper focuses on smooth images and thus on positive correlation between pixels. It is introduced by high-frequency penalty using any circulant differential operator: pth differences between pixels, Laplacian and Sobel, among others. The differential operator is denoted by $\boldsymbol{D}$ and its diagonalized form by $\boldsymbol{\Lambda_D} = \boldsymbol{FDF}^{\dagger}$. Then, the precision matrix is written $\boldsymbol{P} = \gamma_1\boldsymbol{D}^{\mathrm{t}}\boldsymbol{D}$, and its Fourier counterpart is written

$$\boldsymbol{\Lambda_P} = \gamma_1\boldsymbol{\Lambda}_{\boldsymbol{D}}^{\dagger}\boldsymbol{\Lambda_D} = \mathrm{diag}(0, \gamma_1|\mathring{d}_1|^2, \ldots, \gamma_1|\mathring{d}_{N-1}|^2), \qquad (6)$$

where γ_1 is a positive scale factor, diag builds a diagonal matrix from elementary components, and $\mathring{d}_n$ is the nth DFT-2D coefficient of $\boldsymbol{D}$.

Under this parameterization of $\boldsymbol{P}$, the first eigenvalue is equal to zero, corresponding to the absence of penalty for the null frequency $\mathring{x}_0$, i.e., no information is accounted for about the empirical mean level of the image. As a consequence, the determinant vanishes, $\det[\boldsymbol{P}] = 0$, resulting in a degenerated prior. To manage this difficulty, several approaches have been proposed.

Some authors [2,30] still use this prior despite its degeneracy, and this approach can be analyzed in two ways:

1. On the one hand, this prior can be seen as a non-degenerated law for $\mathring{\boldsymbol{x}}_*$, the set of non-null frequency components only. In this format, the prior does not affect any probability for the null frequency component, and the Bayes rule does not apply to this component. Thus, this strategy yields an incomplete posterior law, since the null frequency is not embedded in the methodology.
2. On the other hand, this prior can be seen as a degenerated prior for the whole set of frequencies. The application of the Bayes rule is then somewhat confusing due to degeneracy. In this format, the posterior law cannot be guaranteed to remain non-degenerated.

In any case, neither of the two standpoints yields a posterior law that is both non-degenerated and addresses the whole set of frequencies.

An alternative parameterization relies on the energy of $\boldsymbol{x}$. An extra term $\gamma_0\boldsymbol{I}$, tuned by $\gamma_0 > 0$, in the precision matrix [31], introduces information for all the frequencies including $\mathring{x}_0$. The precision matrix is written

$$\boldsymbol{\Lambda_P} = \gamma_0\boldsymbol{I} + \gamma_1\boldsymbol{\Lambda}_{\boldsymbol{D}}^{\dagger}\boldsymbol{\Lambda_D} = \mathrm{diag}(\gamma_0, \gamma_0 + \gamma_1|\mathring{d}_1|^2, \ldots, \gamma_0 + \gamma_1|\mathring{d}_{N-1}|^2) \qquad (7)$$

with a determinant

$$\det[\boldsymbol{\Lambda_P}] = \prod_{n=0}^{N-1}(\gamma_0 + \gamma_1|\mathring{d}_n|^2). \qquad (8)$$

The obtained Gaussian prior is not degenerated and undoubtedly leads to a proper posterior. Nevertheless, the determinant Eq. (8) is not separable in γ_0 and γ_1. Consequently, the conditional posterior for these parameters is not a classical law, and future development will be more difficult. Moreover, the non-null frequencies $\mathring{\boldsymbol{x}}_*$ are controlled by two parameters, γ_0 and γ_1:

$$p(\mathring{\boldsymbol{x}}|\gamma_0, \gamma_1) = p(\mathring{x}_0|\gamma_0)p(\mathring{\boldsymbol{x}}_*|\gamma_0, \gamma_1). \qquad (9)$$

The proposed approach to manage the degeneracy relies on the addition of a term for the null frequency only, $\boldsymbol{\Lambda}_{\mathbb{1}} = \mathrm{diag}(1, 0, \ldots, 0)$:

$$\boldsymbol{\Lambda_P} = \gamma_0\boldsymbol{\Lambda}_{\mathbb{1}}^{\dagger}\boldsymbol{\Lambda}_{\mathbb{1}} + \gamma_1\boldsymbol{\Lambda}_{\boldsymbol{D}}^{\dagger}\boldsymbol{\Lambda_D}\,. = \mathrm{diag}(\gamma_0, \gamma_1|\mathring{d}_1|^2, \ldots, \gamma_1|\mathring{d}_{N-1}|^2). \qquad (10)$$

The determinant has a separable expression,

$$\det[\boldsymbol{\Lambda_P}] = \gamma_0\gamma_1^{N-1}\prod_{n=1}^{N-1}|\mathring{d}_n|^2, \qquad (11)$$

i.e., the precision parameters have been factorized. In addition, each parameter controls a different set of frequencies:

$$p(\mathring{\boldsymbol{x}}|\gamma_0,\gamma_1) = p(\mathring{x}_0|\gamma_0)p(\mathring{\boldsymbol{x}}_*|\gamma_1),$$

where γ_0 drives the empirical mean level of the image $\mathring{x}_0$ and γ_1 drives the smoothness $\mathring{\boldsymbol{x}}_*$ of the image. With the Fourier precision structure of Eq. (10), we have the non-degenerated prior law for the image that addresses separately all the frequencies with a factorized partition function w.r.t. (γ_0,γ_1):

$$p(\boldsymbol{x}|\gamma_0,\gamma_1) = (2\pi)^{-N/2}\prod_{n=1}^{N-1}|\mathring{d}_n|\gamma_0^{1/2}\gamma_1^{(N-1)/2} \times \exp\left[-\frac{\gamma_0}{2}\|\mathring{x}_0\|^2 - \frac{\gamma_1}{2}\|\boldsymbol{\Lambda}_{\boldsymbol{D}*}\mathring{\boldsymbol{x}}_*\|^2\right], \tag{12}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{D}*}$ is obtained from $\boldsymbol{\Lambda}_{\boldsymbol{D}}$ without the first line and column. The next step is to write the *a priori* law for the noise in an explicit form and the other parameters, including the law parameters $\boldsymbol{\gamma}$ and the instrument parameters $\boldsymbol{w}$.

B. Noise and Data Laws

From a methodological standpoint, any statistic can be included for errors (measurement and model errors). It is possible to account for correlations in the error process or to account for a non-Gaussian law, e.g., Laplacian law, generalized Gaussian law, or other laws based on a robust norm. In the present paper, the noise is modeled as zero-mean white Gaussian vector with unknown precision parameter γ_ϵ:

$$p(\epsilon|\gamma_\epsilon) = (2\pi)^{-N/2}\gamma_\epsilon^{N/2}\exp\left[-\frac{\gamma_\epsilon}{2}\|\boldsymbol{\epsilon}\|^2\right]. \tag{13}$$

Consequently, the likelihood for the parameters given the observed data is written

$$p(\boldsymbol{y}|\boldsymbol{x},\gamma_\epsilon,\boldsymbol{w}) = (2\pi)^{-N/2}\gamma_\epsilon^{N/2}\exp\left[-\frac{\gamma_\epsilon}{2}\|\boldsymbol{y}-\boldsymbol{H}_{\boldsymbol{w}}\boldsymbol{x}\|^2\right]. \tag{14}$$

It naturally depends on the image $\boldsymbol{x}$, on the noise parameter γ_ϵ, and the PSF parameters $\boldsymbol{w}$ embedded in $\boldsymbol{H}_{\boldsymbol{w}}$. It clearly involves a least-squares discrepancy that can be rewritten in the Fourier domain: $\|\boldsymbol{y}-\boldsymbol{H}_{\boldsymbol{w}}\boldsymbol{x}\|^2=\|\mathring{\boldsymbol{y}}-\boldsymbol{\Lambda}_{\boldsymbol{H}}\mathring{\boldsymbol{x}}\|^2$.

C. Hyperparameter Law

A classical choice for the hyperparameter law relies on the conjugate prior [32]: the conditional posterior for the hyperparameters is in the same family as its prior. It results in practical and algorithmic facilities: update of the laws amounts to update of a small number of parameters.

The three parameters γ_0, γ_1, and γ_ϵ are precision parameters of Gaussian laws Eq. (12) and (14); a conjugate law for these parameters is the gamma law (see Appendix B). Given parameters (α_i,β_i), for $i=0$, 1 or ϵ, the pdf reads

$$p(\gamma_i) = \frac{1}{\beta_i^{\alpha_i}\Gamma(\alpha_i)}\gamma_i^{\alpha_i-1}\exp(-\gamma_i/\beta_i), \quad \forall\ \gamma_i \in [0,+\infty[. \tag{15}$$

In addition to computational efficiency, the law allows for non-informative priors. With specific parameter values, one obtains two improper non-informative priori: the Jeffreys law $p(\gamma)=1/\gamma$ and the uniform law $p(\gamma)=\mathcal{U}_{[0,+\infty[}(\gamma)$ with (α_i,β_i) set to $(0,+\infty)$ and $(1,+\infty)$, respectively. The Jeffreys law is a classical law for the precisions and is considered non-informative [33]. This law is also invariant to power transformations: the law of γ^n [33,34] is also a Jeffreys law. For these reasons development is done using the Jeffreys law.

D. PSF Parameter Law

Regarding the PSF parameters $\boldsymbol{w}$, we consider that the instrument design process or a physical study provides a nominal value $\bar{\boldsymbol{w}}$ with uncertainty $\boldsymbol{\delta}$, that is, $\boldsymbol{w}\in[\bar{\boldsymbol{w}}-\boldsymbol{\delta},\bar{\boldsymbol{w}}+\boldsymbol{\delta}]$. The "Principle of Insufficient Reason" [33] leads to a uniform prior on this interval:

$$p(\boldsymbol{w}) = \mathcal{U}_{\bar{\boldsymbol{w}},\boldsymbol{\delta}}(\boldsymbol{w}), \tag{16}$$

where $\boldsymbol{\mathcal{U}}_{\bar{\boldsymbol{w}},\boldsymbol{\delta}}$ is a uniform pdf on $[\bar{\boldsymbol{w}}-\boldsymbol{\delta},\bar{\boldsymbol{w}}+\boldsymbol{\delta}]$. Nevertheless, within the proposed framework, the choice is not limited and other laws, such as Gaussian, are possible. In any case, other choices do not allow easier computation because of the non-linear dependency of the observation model w.r.t. PSF parameters.

4. PROPER POSTERIOR LAW

At this point, the prior law of each parameter is available: the PSF parameters, the hyperparameters, and the image. Thus, the joint law for all the parameters is built by multiplying the likelihood Eq. (14) and the *a priori* laws Eq. (12), (15), and (16):

$$p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w},\mathring{\boldsymbol{y}}) = p(\mathring{\boldsymbol{y}}|\mathring{\boldsymbol{x}},\gamma_\epsilon,\boldsymbol{w})p(\mathring{\boldsymbol{x}}|\gamma_0,\gamma_1)p(\gamma_\epsilon)p(\gamma_0)p(\gamma_1)p(\boldsymbol{w}) \tag{17}$$

and explicitly

$$p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w},\mathring{\boldsymbol{y}}) = \frac{(2\pi)^{-N}\prod_{n=1}^{N-1}|\mathring{d}_n|}{\beta_\epsilon^{\alpha_\epsilon}\Gamma(\alpha_\epsilon)\beta_0^{\alpha_0}\Gamma(\alpha_0)\beta_1^{\alpha_1}\Gamma(\alpha_1)}\gamma_\epsilon^{\alpha_\epsilon+N/2-1}\gamma_0^{\alpha_0-1/2}\gamma_1^{\alpha_1+(N-1)/2-1} \times\exp\left[-\frac{\gamma_\epsilon}{\beta_\epsilon}-\frac{\gamma_0}{\beta_0}-\frac{\gamma_1}{\beta_1}\right]\boldsymbol{\mathcal{U}}_{\bar{\boldsymbol{w}},\boldsymbol{\delta}}(\boldsymbol{w}) \times\exp\left[-\frac{\gamma_\epsilon}{2}\|\mathring{\boldsymbol{y}}-\boldsymbol{\Lambda}_{\boldsymbol{H}}\mathring{\boldsymbol{x}}\|^2-\frac{\gamma_0}{2}|\mathring{x}_0|^2-\frac{\gamma_1}{2}\|\boldsymbol{\Lambda}_{\boldsymbol{D}}\mathring{\boldsymbol{x}}\|^2\right]. \tag{18}$$

According to the Bayes rule, the *a posteriori* law reads

$$p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w}|\mathring{\boldsymbol{y}}) = \frac{p(\mathring{\boldsymbol{x}},\gamma_\epsilon,\gamma_0,\gamma_1,\boldsymbol{w},\mathring{\boldsymbol{y}})}{p(\mathring{\boldsymbol{y}})}, \tag{19}$$

where $p(\mathring{\boldsymbol{y}})$ is a normalization constant

$$p(\mathring{\boldsymbol{y}}) = \int p(\mathring{\boldsymbol{y}},\mathring{\boldsymbol{x}},\boldsymbol{\gamma},\boldsymbol{w})\mathrm{d}\mathring{\boldsymbol{x}}\mathrm{d}\boldsymbol{\gamma}\mathrm{d}\boldsymbol{w}. \tag{20}$$

As described before, setting $\gamma_0=0$ leads to degenerated prior and joint laws. However, when the observation system preserves the null frequency, γ_0 can be considered a nuisance parameter. In addition, only prior information on the smoothness is available.

In Bayesian framework, a solution to eliminate the nuisance parameters is to integrate them out in the *a posteriori* law. According to our parameterization in Subsection 3.A, the integration of γ_0 is the integration of a gamma law. Application of Appendix B.1 on γ_0 in the *a posteriori* law Eq. (19) provides

$$p(\mathring{\boldsymbol{x}}, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{\boldsymbol{y}}) = \frac{p(\mathring{x}_0) p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0)}{\int p(\mathring{x}_0) p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0) \mathrm{d}\gamma_\epsilon \mathrm{d}\gamma_1 \mathrm{d}\boldsymbol{w} \mathrm{d}\mathring{\boldsymbol{x}}_* \mathrm{d}\mathring{x}_0} \tag{21}$$

with

$$p(\mathring{x}_0) = \int p(\mathring{x}_0 | \gamma_0) p(\gamma_0) \mathrm{d}\gamma_0 = \left(1 + \frac{\beta_0 \mathring{x}_0^2}{2}\right)^{-\alpha_0 - 1/2}. \tag{22}$$

Now the parameter is integrated, and the parameters α_0 and β_0 are set to remove the null frequency penalization. Since we have $\alpha_0 > 0$ and $\beta_0 > 0$ we get $(1+\beta_0 \mathring{x}_0^2/2)^{-\alpha_0-1/2} \leqslant 1$, and the joint law is majored:

$$\left(1 + \frac{\beta_0 \mathring{x}_0^2}{2}\right)^{-\alpha_0 - 1/2} p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0) \leqslant p(\mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}_*, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{x}_0). \tag{23}$$

Consequently, by the dominated convergence theorem [35], the limit of the law with $\alpha_0 \to 1$ and $\beta_0 \to 0$ can be placed under the integral sign at the denominator. Then the null-frequency penalization $p(\mathring{x}_0)$ from the numerator and denominator are removed. This is equivalent to the integration of the γ_0 parameter under a Dirac distribution (see Appendix B). The equation is simplified, and the integration with respect to $\mathring{x}_0$ in the denominator Eq. (20)

$$\int_{\mathbb{R}} p(\mathring{\boldsymbol{y}} | \mathring{\boldsymbol{x}}, \gamma_\epsilon, \boldsymbol{w}) p(\mathring{\boldsymbol{x}}_* | \gamma_1) p(\gamma_1, \gamma_\epsilon, \boldsymbol{w}) \mathrm{d}\mathring{x}_0 \propto \int_{\mathbb{R}} p(\mathring{y}_0 | \mathring{x}_0, \gamma_\epsilon, \boldsymbol{w}) \mathrm{d}\mathring{x}_0 \tag{24}$$

$$\propto \int_{\mathbb{R}} \exp\left[-\frac{\gamma_\epsilon}{2}(\mathring{y}_0 - \mathring{h}_0 \mathring{x}_0)^2\right] \mathrm{d}\mathring{x}_0 \tag{25}$$

converges if and only if $\mathring{h}_0 \neq 0$: the null frequency is observed. If this condition is met, Eq. (21) with $\beta_0 = 0$ and $\alpha_0 = 1$ is a proper posterior law for the image, the precision parameters, and the PSF parameters. In other words, if the average is observed, the degeneracy of the *a priori* law is not transmitted to the *a posteriori* law.

Then, the obtained *a posteriori* law is written

$$p(\mathring{\boldsymbol{x}}, \gamma_\epsilon, \gamma_1, \boldsymbol{w} | \mathring{\boldsymbol{y}}) = \frac{p(\mathring{\boldsymbol{x}}, \gamma_\epsilon, \gamma_1, \boldsymbol{w}, \mathring{\boldsymbol{y}})}{p(\mathring{\boldsymbol{y}})}$$

$$\propto \gamma_\epsilon^{\alpha_\epsilon + N/2 - 1} \gamma_1^{\alpha_1 + (N-1)/2 - 1} \mathcal{U}_{\bar{w},\delta}(\boldsymbol{w}) \exp\left[-\frac{\gamma_\epsilon}{2}\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}} \mathring{\boldsymbol{x}}\|^2 - \frac{\gamma_1}{2}\|\boldsymbol{\Lambda}_{\boldsymbol{D}*} \mathring{\boldsymbol{x}}_*\|^2\right] \exp\left[-\frac{\gamma_\epsilon}{\beta_\epsilon} - \frac{\gamma_1}{\beta_1}\right]. \tag{26}$$

Finally, inference is done on this law Eq. (26). If the null frequency is not observed, or information must be added, the previous Eq. (19) can be used.

5. POSTERIOR MEAN ESTIMATOR AND LAW EXPLORATION

This section presents the algorithm to explore the posterior law Eq. (19) or Eq. (26) and to compute an estimate of the parameters. For this purpose, Monte Carlo Markov chain is used to provide samples. First, the obtained samples are used to compute different moments of the law. Afterward, they are also used to approximate marginal laws as histograms. These two representations are helpful for analyzing the *a posteriori* law, the structure of the available information, and the uncertainty. They are used in Subsection 6.C.2 to illustrate the mark of the ambiguity in the myopic problem.

Here, the samples of the *a posteriori* law are obtained by a Gibbs sampler [25,36,37]: it consists in iteratively sampling the conditional posterior law for a set of parameters given the other parameters (obtained at previous iteration). Typically, the sampled laws are the law of $\mathring{\boldsymbol{x}}$, γ_i, and $\boldsymbol{w}$. After a burn-in time, the complete set of samples are under the joint *a posteriori* law. The next three subsections present each sampling step.

A. Sampling the Image

The conditional posterior law of the image is a Gaussian law:

$$\mathring{\boldsymbol{x}}^{(k+1)} \sim p(\mathring{\boldsymbol{x}} | \mathring{\boldsymbol{y}}, \gamma_\epsilon^{(k)}, \gamma_0^{(k)}, \gamma_1^{(k)}, \boldsymbol{w}^{(k)}) \tag{27}$$

$$\sim \mathcal{N}(\boldsymbol{\mu}^{(k+1)}, \boldsymbol{\Sigma}^{(k+1)}). \tag{28}$$

The covariance matrix is diagonal and is written

$$\boldsymbol{\Sigma}^{(k+1)} = (\gamma_\epsilon^{(k)} |\boldsymbol{\Lambda}_{\boldsymbol{H}}^{(k)}|^2 + \gamma_0^{(k)} |\boldsymbol{\Lambda}_{\mathbb{1}}|^2 + \gamma_1^{(k)} |\boldsymbol{\Lambda}_{\boldsymbol{D}}|^2)^{-1}, \tag{29}$$

and the mean

$$\boldsymbol{\mu}^{(k+1)} = \gamma_\epsilon^{(k)} \boldsymbol{\Sigma}^{(k+1)} \boldsymbol{\Lambda}_{\boldsymbol{H}}^{\dagger(k)} \mathring{\boldsymbol{y}}, \tag{30}$$

where † is the transpose conjugate symbol. The vector $\boldsymbol{\mu}^{(k+1)}$ is the regularized least-square solution at the current iteration (or the Wiener–Hunt filter). Clearly, if the null frequency is not observed, $\mathring{h}_0 = 0$ and if $\gamma_0 = 0$, the covariance matrix $\boldsymbol{\Sigma}$ is not invertible and the estimate is not defined as described Section 4.

Finally, since the matrix is diagonal, the sample $\mathring{\boldsymbol{x}}^{(k+1)}$ is obtained by a term-wise product of $\boldsymbol{F\epsilon}$ (where $\boldsymbol{\epsilon}$ is white Gaussian) with the standard deviation matrix $(\boldsymbol{\Sigma}^{(k+1)})^{1/2}$

followed by the addition of the mean $\boldsymbol{\mu}^{(k+1)}$, also computed with term-wise products Eq. (30). Consequently, the sampling of the image is effective even with a high-dimensional object.

B. Sampling Precision Parameters

The conditional posterior laws of the precisions are gamma corresponding to their prior law with parameters updated by the likelihood

$$\gamma_i^{(k+1)} \sim p(\gamma_i | \mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}^{(k+1)}, \boldsymbol{w}^{(k)}) \tag{31}$$

$$\sim \mathcal{G}(\gamma_i | \alpha_i^{(k+1)}, \beta_i^{(k+1)}). \tag{32}$$

For γ_ϵ, γ_0, and γ_1 the parameters law are, respectively,

$$\alpha_\epsilon^{(k+1)} = \alpha_\epsilon + N/2 \quad \text{and} \quad \beta_\epsilon^{(k+1)} = \left(\beta_\epsilon^{-1} + \frac{1}{2}\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}}^{(k)}\mathring{\boldsymbol{x}}^{(k+1)}\|^2\right)^{-1}, \tag{33}$$

$$\alpha_0^{(k+1)} = \alpha_0 + 1/2 \quad \text{and} \quad \beta_0^{(k+1)} = \left(\beta_0^{-1} + \frac{1}{2}(\mathring{x}_0^{(k+1)})^2\right)^{-1}, \tag{34}$$

$$\alpha_1^{(k+1)} = \alpha_1 + (N-1)/2 \quad \text{and} \quad \beta_1^{(k+1)} = \left(\beta_1^{-1} + \frac{1}{2}\|\boldsymbol{\Lambda}_{\boldsymbol{D}}\mathring{\boldsymbol{x}}^{(k+1)}\|^2\right)^{-1}. \tag{35}$$

In the case of the Jeffreys prior, the parameters are

$$\alpha_\epsilon^{(k+1)} = N/2 \quad \text{and} \quad \beta_\epsilon^{(k+1)} = 2/\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}}^{(k)}\mathring{\boldsymbol{x}}^{(k+1)}\|^2, \tag{36}$$

$$\alpha_0^{(k+1)} = 1/2 \quad \text{and} \quad \beta_0^{(k+1)} = 2/(\mathring{x}_0^{(k+1)})^2, \tag{37}$$

$$\alpha_1^{(k+1)} = (N-1)/2 \quad \text{and} \quad \beta_1^{(k+1)} = 2/\|\boldsymbol{\Lambda}_{\boldsymbol{D}}\mathring{\boldsymbol{x}}^{(k+1)}\|^2. \tag{38}$$

Remark 1: *If the a posteriori law Eq. (26) without* γ_0 *is considered, there is no need to sample this parameter [Eqs. (34) and (37) are not useful] and* $\gamma_0^{(k)} = 0$ *in Eq. (29).*

C. Sample PSF Parameters

The conditional law for PSF parameters is written

$$\boldsymbol{w}^{(k+1)} \sim p(\boldsymbol{w} | \mathring{\boldsymbol{y}}, \mathring{\boldsymbol{x}}^{(k+1)}, \gamma_\epsilon^{(k+1)}) \tag{39}$$

$$\propto \exp\left[-\frac{\gamma_\epsilon^{(k+1)}}{2}\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},\boldsymbol{w}}\mathring{\boldsymbol{x}}^{(k+1)}\|^2\right], \tag{40}$$

where parameters $\boldsymbol{w}$ are embedded in the PSF $\boldsymbol{\Lambda}_{\boldsymbol{H}}$. This law is not standard and is intricate: no algorithm exists for direct sampling and we use the Metropolis–Hastings method to bypass this difficulty. In the Metropolis–Hastings algorithm, a sample $\boldsymbol{w}_{\mathrm{p}}$ is proposed and accepted with a certain probability. This probability depends on the ratio between the likelihood of the proposed value and the likelihood of the current value $\boldsymbol{w}^{(k)}$. In practice, in the independent form described in appendix C, with prior law as proposition law, it is divided into several steps.

1. PROPOSITION: Sample a proposition:

$$\boldsymbol{w}_{\mathrm{p}} \sim p(\boldsymbol{w}) = \mathcal{U}_{[\mathbf{a}\ \mathbf{b}]}(\boldsymbol{w}). \tag{41}$$

2. PROBABILITY OF ACCEPTANCE: Calculate the criterion:

$$J(\boldsymbol{w}^{(k)}, \boldsymbol{w}_{\mathrm{p}}) = \frac{\gamma_\epsilon^{(k+1)}}{2}(\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},\boldsymbol{w}^{(k)}}\mathring{\boldsymbol{x}}^{(k+1)}\|^2 - \|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},\boldsymbol{w}_{\mathrm{p}}}\mathring{\boldsymbol{x}}^{(k+1)}\|^2). \tag{42}$$

3. UPDATE: Sample $t \sim \mathcal{U}_{[0\ 1]}$ and take

$$\boldsymbol{w}^{(k+1)} = \begin{cases} \boldsymbol{w}_{\mathrm{p}} & \text{if } \log t < J \\ \boldsymbol{w}^{(k)} & \text{otherwise} \end{cases}. \tag{43}$$

D. Empirical Mean

The sampling of $\mathring{\boldsymbol{x}}$, $\boldsymbol{\gamma}$, and $\boldsymbol{w}$ is repeated iteratively until the law has been sufficiently explored. These samples $[\mathring{\boldsymbol{x}}^{(k)}, \boldsymbol{\gamma}^{(k)}, \boldsymbol{w}^{(k)}]$ follow the global *a posteriori* law of Eq. (19). By the large-numbers law, the estimate, defined as the posterior mean, is approximated by

$$\hat{\boldsymbol{x}} = \boldsymbol{F}^\dagger \mathbb{E}[\mathring{\boldsymbol{x}}] \approx \boldsymbol{F}^\dagger \left[\frac{1}{K}\sum_{k=0}^{K-1} \mathring{\boldsymbol{x}}^{(k)}\right]. \tag{44}$$

As described by Eq. (44), to obtain an estimate of the image in the spatial space, all the computations are achieved recursively in the Fourier space with a single IFFT at the end. An implementation example in pseudo code is described Appendix D.

6. DECONVOLUTION RESULTS

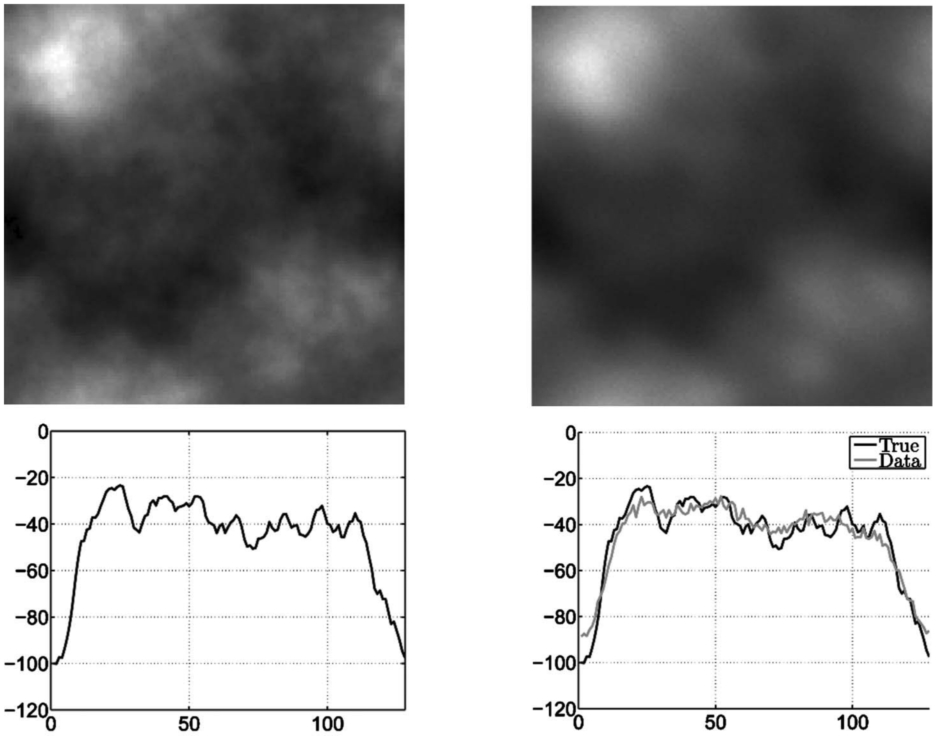
This section presents numerical results obtained by the proposed method. In order to completely evaluate the method, true value of all parameters $\boldsymbol{x}$, $\boldsymbol{w}$, γ_ϵ but also γ_1, γ_0 is needed. In order to achieve this, an entirely simulated case is studied: image and noise are simulated under their respective prior laws Eq. (12) and (13) with given values of γ_0, γ_1, and γ_ϵ. Thanks to this protocol, all experimental conditions are controlled and the estimation method is entirely evaluated.

The method has also been applied in different conditions (lower signal-to-noise ratio, broader PSF, different and realistic (non-simulated) images, and so on) and showed similar behavior. However, in the case of realistic images, since the true value of the hyperparameters γ_0 and γ_1 is unknown, the evaluation cannot be complete.

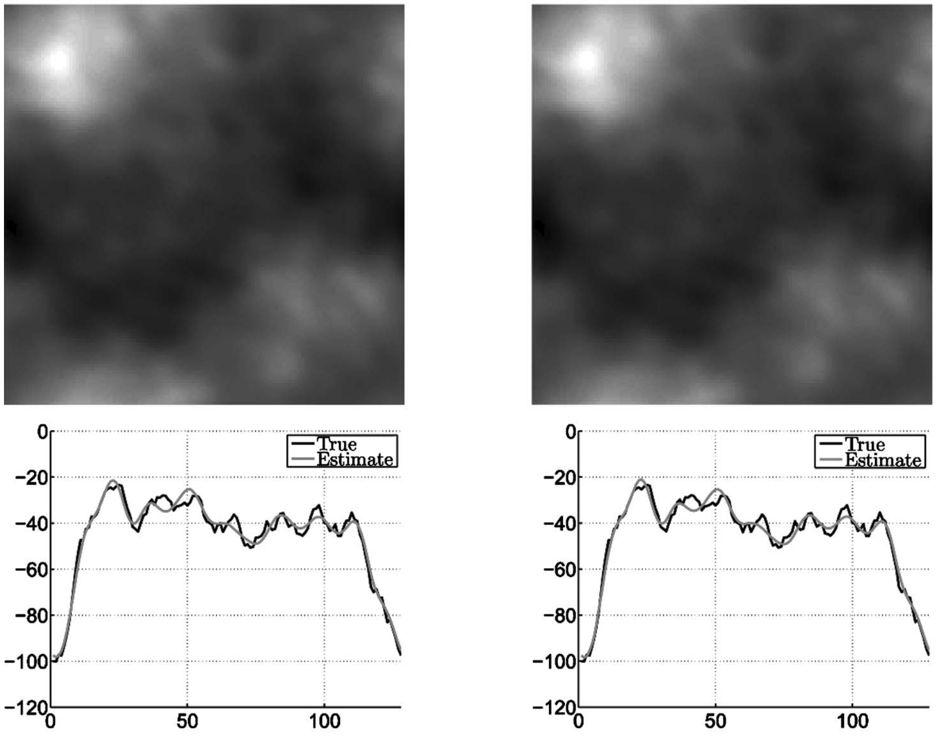
A. Practical Experimental Conditions

Concretely, a 128×128 image is generated in the Fourier space as the product of a complex white Gaussian noise and the *a priori* standard deviation matrix $\boldsymbol{\Sigma} = (\gamma_0 \boldsymbol{\Lambda}_1^\dagger \boldsymbol{\Lambda}_1 + \gamma_1 \boldsymbol{\Lambda}_{\boldsymbol{D}}^\dagger \boldsymbol{\Lambda}_{\boldsymbol{D}})^{-1/2}$, given by Eq. (10). The chosen matrix $\boldsymbol{\Lambda}_{\boldsymbol{D}}$ results from the FFT-2D of the Laplacian operator $[0\ 1\ 0;\ 1\ {-4}\ 1;\ 0\ 1\ 0]/8$, and the parameter values are $\gamma_0 = 1$ and $\gamma_1 = 2$.

These parameters provide the image shown in Fig. 1(a): it is an image with smooth features similar to a cloud. Pixels have numerical values between -100 and 150, and the profile line 68 shows fluctuations around a value of -40.



(a) True image (b) Data



(c) Non-myopic (d) Myopic

Fig. 1. (a) 128×128 sample of the *a priori* law for the object with $\gamma_0 = 1$ and $\gamma_1 = 2$. (b) Data computed with the PSF shown in Fig. 2. (c) and (d) Estimates with non-myopic and the myopic estimate, respectively. Profiles correspond to the 68th line of the image.

The *a priori* law for the hyperparameters are set to the non-informative Jeffreys law by fixing the (α_i, β_i) to $(0, +\infty)$, as explained in Subsection. 3.C. In addition, the PSF is obtained in the Fourier space by discretization of a normalized Gaussian shape,

$$\mathring{h}(\nu_\alpha, \nu_\beta) = \exp[-2\pi^2(\nu_\alpha^2(w_\alpha \cos^2\varphi + w_\beta \sin^2\varphi) + \nu_\beta^2(w_\alpha \sin^2\varphi + w_\beta \cos^2\varphi) + 2\nu_\alpha\nu_\beta \sin\varphi \cos\varphi(w_\alpha - w_\beta))], \qquad (45)$$

with frequencies $(\nu_\alpha, \nu_\beta) \in [-0.5;0.5]^2$. This low-pass filter, illustrated in Fig. 2, is controlled by three parameters:

- Two width parameters w_α and w_β set to 20 and 7, respectively. Their *a priori* laws are uniform: $p(w_\alpha) = \mathcal{U}_{[19\ 21]}(w_\alpha)$ and $p(w_\beta) = \mathcal{U}_{[6\ 8]}(w_\alpha)$ corresponding to an uncertainty of about 5% and 15% around the nominal value (see Subsection 3.D).
- A rotation parameter φ set to $\pi/3$. The *a priori* law is also uniform: $p(\varphi) = \mathcal{U}_{[\pi/4\ \pi/2]}(\varphi)$, corresponding to 50% uncertainty.

Then, the convolution is computed in the Fourier space, and the data are obtained by adding a white Gaussian noise with precision $\gamma_\epsilon = 0.5$. Data are shown Fig. 1(b): they are naturally smoother than the true image, and the small fluctuations are less visible and are corrupted by the noise. The empirical mean level of the image is correctly observed (the null frequency coefficient of $\boldsymbol{H_w}$ is $\mathring{h}_0 = 1$) so the parameter γ_0 is considered a nuisance parameter. Consequently it is integrated out under a Dirac distribution (see Section 4). This is equivalent to fix its value to 0 in the algorithm of Appendix D, line 4.

Finally, the method is evaluated on two different situations.

1. The unsupervised and non-myopic case: the parameters $\boldsymbol{w}$ are known. Consequently, there is no M.H. step (Subsection 5.C): lines 9–16 are ignored in the algorithm of Appendix D, and $\boldsymbol{w}$ is set to its true value. To obtain sufficient law exploration, the algorithm is run until the difference between two successive empirical means is less than 10^{-3}. In this case, 921 samples are necessary, and they are computed in approximately 12 s on a processor at 2.66 GHz with Matlab,
2. The unsupervised and myopic case: all the parameters are estimated. To obtain sufficient law exploration, the algorithm is run until the difference between two successive empirical means is less than 5×10^{-5}. In this case, 18, 715 samples are needed, and they are computed in approximately 7 min.

Remark: *The algorithm has also been run for up to 1,000,000 samples, in both cases, without perceptible qualitative changes.*

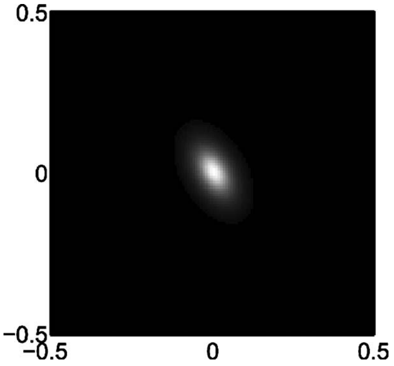


Fig. 2. PSF with $w_\alpha = 20$, $w_\beta = 7$ and $\varphi = \pi/3$. The x axis and y axis are reduced frequency.

B. Estimation Results

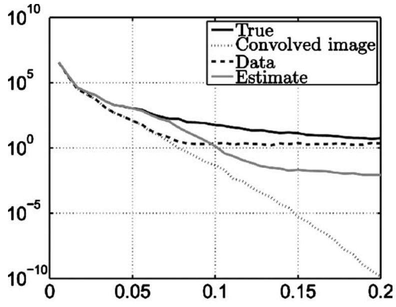
1. Images

The two results for the image are given Figs. 1(c) and 1(d) for the non-myopic and the myopic cases, respectively.

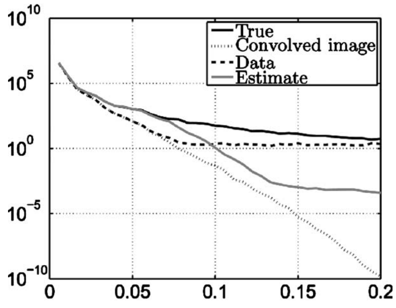
The effect of deconvolution is notable on the image, as well as on the shown profile. The object is correctly positioned, the orders of magnitude are respected, and the mean level is correctly reconstructed. The image is restored, more details are visible, and the profiles are closer matches to the true image than are the data. More precisely, pixels 20–25 of the 68th line of the image in Fig. 1 show the restoration of the original dynamic, whereas it is not visible in the data. Between pixels 70 and 110, fluctuations not visible in the data are also correctly restored.

In order to visualize and study the spectral contents of the images, the circular average of the empirical power spectral density is considered and is called "spectrum" hereafter. The subjacent spectral variable is a radial frequency f, such as $f^2 = \nu_\alpha^2 + \nu_\beta^2$. The spectra of the true object, data, and restored object are shown Figs. 3(a) and 3(b) in the non-myopic and myopic cases, respectively. It is clear that the spectrum of the true image is correctly retrieved, in both cases, up to the radial frequency $f \approx 0.075$. Above this frequency, noise is clearly dominant and information about the image is almost lost. In other words, the method produces correct spectral equalization in the properly observed frequency band. The result is expected from the Wiener–Hunt method, but the achievement is the joint estimation of hyperparameter and instrument parameters in addition to the correct spectral equalization.

Concerning a comparison between non-myopic and myopic cases, there is no visual difference. The spectra in



(a) Non-Myopic



(b) Myopic

Fig. 3. Circular average of the empirical power spectral density of the image, the convolved image, the data (convolved image corrupted by noise), and the estimates, in radial frequency with the y axis in logarithmic scale. The x axis is the radial frequency.

Table 1. Error e [Eq. (46)] and Averaged Standard Deviation $\hat{\sigma}$ of the Posterior Image Law[a]

	Data	Non-myopic	Myopic	Best
Error (e)	11.092 %	6.241 %	6.253 %	6.235 %
$\hat{\sigma}$ of $\boldsymbol{x}$ law	–	3.16	3.25	–

[a] The "Best" error has been obtained with the knowledge of the true image.

Figs. 3(a) and 3(b) in the non-myopic and myopic cases, respectively, are visually indistinguishable. This is also the case when comparing Figs. 1(c) and 1(d) and especially, the 68th line. From a more precise quantitative evaluation, a slight difference is observed is and detailed below.

In order to quantify performance, a normalized Euclidean distance

$$e = \|\boldsymbol{x} - \boldsymbol{x}^*\|/\|\boldsymbol{x}^*\| \tag{46}$$

between an image $\boldsymbol{x}$ and the true image $\boldsymbol{x}^*$ is considered. It is computed between the true image and the estimated images as well as between the true image and the data. Results are reported in Table 1 and confirm that the deconvolution is effective with an error of approximately 6% in the myopic case compared with 11% for the data. Both non-myopic and myopic deconvolution reduce error by a factor of 1.7 with respect to the observed data.

Regarding a comparison between the non-myopic and myopic cases, the errors are almost the same, with a slightly lower value for the non-myopic case, as expected. This difference is consistent with the intuition: more information is injected in the non-myopic case through the true PSF parameter values.

2. Hyperparameters and Instrument Parameters

Concerning the other parameters, their estimates are close to the true values and are reported in Table 2. The γ_ϵ estimate is very close to the true value with $\hat{\gamma}_\epsilon=0.49$ instead of 0.5 in the two cases. The error for the PSF parameters are 0.35%, 2.7%, and 1.9% for w_α, w_β, and ϕ, respectively. The value of γ_1 is underestimated in the two cases, with approximately 1.7 instead of 2. All the true values fall in the $\hat{\mu}\pm 3\hat{\sigma}$ interval.

In order to deepen the numerical study, the paper evaluates the capability of the method to accurately select the best values for hyperparameters and instrument parameters. To this end, we compute the estimation error Eq. (46) for a set of "exhaustive" values of the parameters $[\gamma_\epsilon, \gamma_1, w_\alpha, w_\beta, \phi]$. The protocol is the following: (1) choose a new value for a parameter (γ_ϵ, for example) and fix the other parameters to the value provided by our algorithm, (2) compute the Wiener–Hunt solution (Subsection 5.A), and (3) compute the error index.

Results are reported in Fig. 4. In each case, smooth variation of error is observed when varying hyperparameters and instrument parameters, and unique optimum is visible. In this way, one can find the values of the parameters that provide the best Wiener–Hunt solution when the true image $\boldsymbol{x}^*$ is known. This is reported in Table 1 and shows almost imperceptible improvement: optimization of the parameters (based on the true image $\boldsymbol{x}^*$) allows negligible improvement (smaller than 0.02% as reported in Table 1).

Thus the main conclusion is that the proposed unsupervised and myopic approach is a relevant tool for tuning parameters: it works (without knowledge of the true image) as well as being an optimal approach (based on the knowledge of the true image).

C. A Posteriori Law Characteristics

This section describes the *a posteriori* law using histograms, means and variances of the parameters. The sample histograms, Figs. 5 and 6, provide an approximation of the marginal posterior law for each parameter. Tables 1 and 2 report the variance for the image and law parameters respectively and thus allow to quantify the uncertainty.

1. Hyperparameter Characteristics

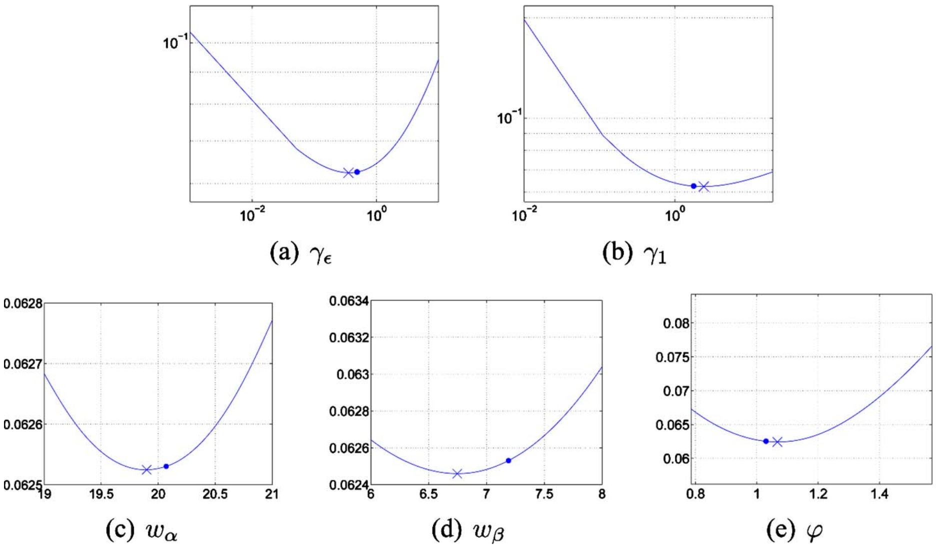
The histograms for γ_ϵ and γ_1, Fig. 5, are concentrated around a mean value in both non-myopic and myopic cases. The variance for γ_ϵ is lower than the one for γ_1, and it can be explained as follows.

The observed data are directly affected by noise (present at the system output), whereas they are indirectly affected by the object (present at the system input). The convolution system damages the object and not the noise: as a consequence, the parameter γ_ϵ (which drives noise law) is more reliably estimated than γ_1 (which drives object law).

A second observation is the smaller variance for γ_1 in the non-myopic case, Fig. 5(c), than in the myopic case, Fig. 5(d). This is the consequence of the addition of information in the non-myopic case w.r.t. the myopic one, through the value of the PSF parameters. In the myopic case, the estimates are based on the knowledge of an interval for the values of the instrument parameters, whereas in the non-myopic case, the estimates are based on the true values of the instrument parameters.

Table 2. Quantitative Evaluation: True and Estimated Values of Hyperparameters and PSF Parameters

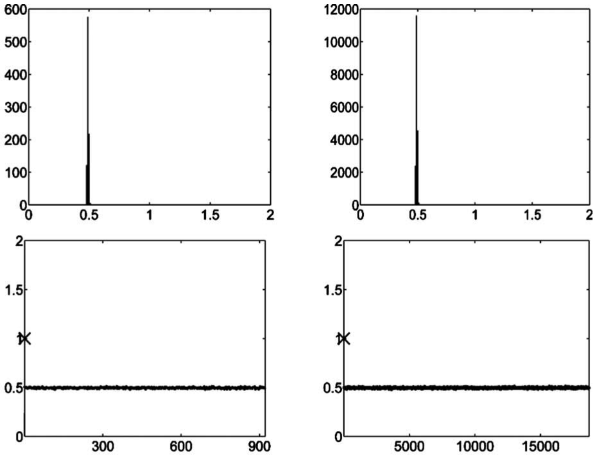
		$\hat{\gamma}_\epsilon\pm\hat{\sigma}$	$\hat{\gamma}_1\pm\hat{\sigma}$	$\hat{w}_\alpha\pm\hat{\sigma}$	$\hat{w}_\beta\pm\hat{\sigma}$	$\hat{\varphi}\pm\hat{\sigma}$
Case	True Value	0.5	2	20	7	1.05 ($\pi/3$)
Non-myopic	Estimate	0.49 ± 0.0056	1.78 ± 0.14	–	–	–
	Error	2.0 %	11 %	–	–	–
Myopic	Estimate	0.49 ± 0.0056	1.65 ± 0.15	20.07 ± 0.53	7.19 ± 0.38	1.03 ± 0.04
	Error	2.0 %	18 %	0.35 %	2.7 %	1.9 %



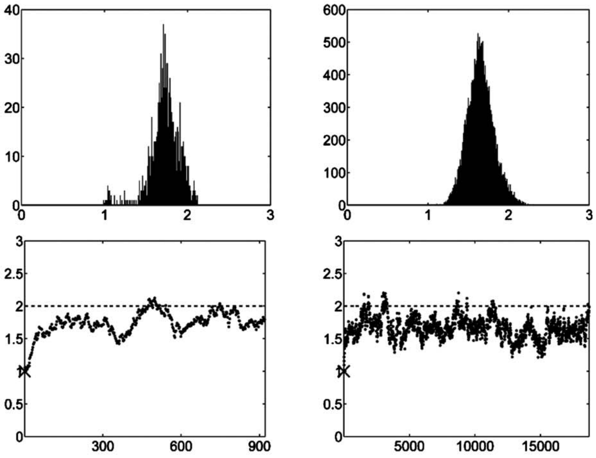
(a) γ_ϵ (b) γ_1

(c) w_α (d) w_β (e) φ

Fig. 4. (Color online) Computation of the best parameters in the sense e, Eq. (46). The symbol × is the minimum and the dot is the estimated value with our approach. The y axes of γ_ϵ and γ_1 are in logarithmic scale.



(a) γ_ϵ for non-myopic case (b) γ_ϵ for myopic case



(c) γ_1 for non-myopic case (d) γ_1 for myopic case

Fig. 5. Histograms and chains for the non-myopic case [(a) and (c)] and the myopic case [(b) and (d)] for γ_ϵ and γ_1, respectively. The symbol × on the y axes localizes the initial value, and the dashed line corresponds to the true value. The x axes are the iteration's index for the chains (bottom of figures) and the parameter value for the histograms (top of figures).

2. PSF Parameter Characteristics

Figure 6 gives histograms for the three PSF parameters and their appearance is quite different from the one for the hyperparameters. The histograms for w_α and w_β, Figs. 6(a) and 6(b), are not as concentrated as the one of Fig. 5 for hyperparameters. Their variances are quite large with respect to the interval of the prior law. In contrast, the histogram for the parameter φ, Fig. 6(c), has the smallest variance. It is analyzed as a consequence of a larger sensitivity of the data w.r.t. the parameter φ than w.r.t. the parameters w_α and w_β. In an equivalent manner, the observed data are more informative about the parameter φ than about the parameters w_α and w_β.

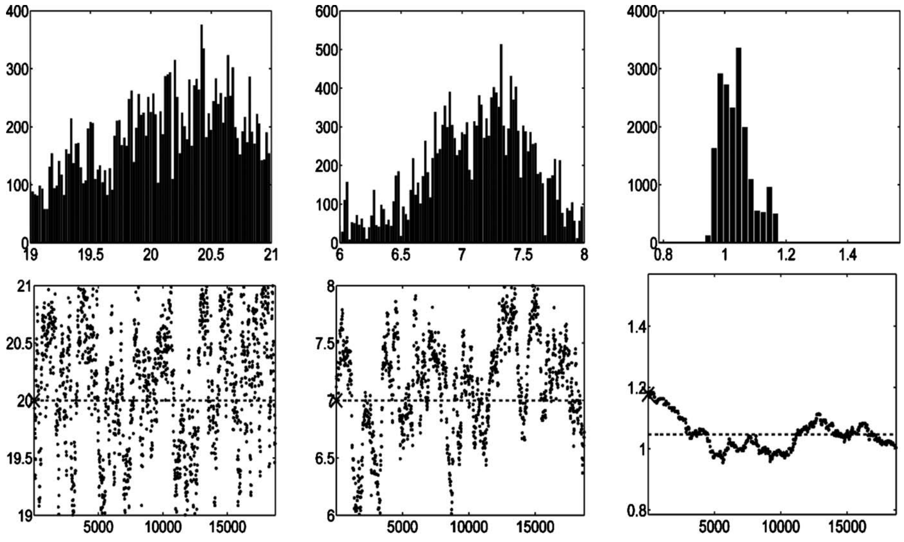
3. Mark of the Myopic Ambiguity

Finally, a correlation between parameters (γ_1, w_α) and (γ_1, w_β) is visible on their joint histograms, Fig. 7. It can be interpreted as a consequence of the ambiguity in the primitive myopic deconvolution problem, in the following manner: the parameters γ_1 and $\boldsymbol{w}$ both participate in the interpretation of the spectral content of data, γ_1, as a scale factor and $\boldsymbol{w}$ as a shape factor. An increase of w_α or w_β results in a decrease of the cutoff frequency of the observation system. In order to explain the spectral content of a given data set, the spectrum of the original image must contain more high frequencies, i.e., a smaller γ_1. This is also observed on the histogram illustrated, Fig. 7(a)

D. MCMC Algorithm Characteristics

Globally, the chains of Figs. 5 and 6, have a Markov feature (correlated) and explore the parameter space. They have a burn-in period followed by a stationary state. This characteristic has always been observed regardless of the initialization. For fixed experimental conditions, the stationary state of multiple runs was always around the same value. Considering different initializations, the only visible change is on the length of the burn-in period.

More precisely, the chain of γ_ϵ is concentrated in a small interval, the burn-in period is very short (fewer

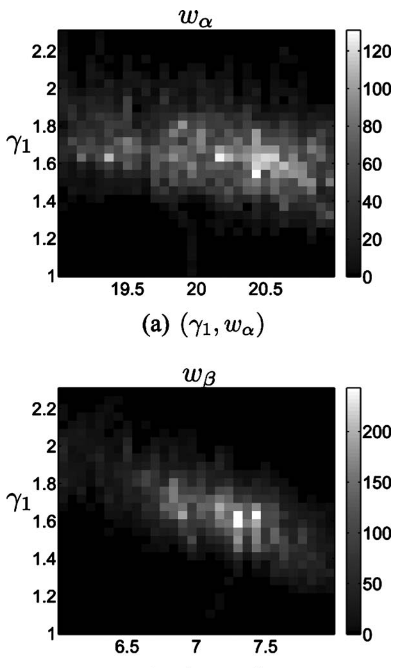


(a) w_α (b) w_β (c) φ

Fig. 6. Histogram and chain for the PSF parameters (a) w_α, (b) w_β, and (c) φ. The symbol × on the y axes localizes the initial value, and the dashed line corresponds to the true value. The x axis for the histograms and the y axis of the chain are limits of the *a priori* law.

than 10 samples), and its evolution seems independent of the other parameters. The chain of γ_1 has a larger exploration, the burn-in period is longer (approximately 200 samples), and the histogram is larger. This is in accordance with the analysis of Subsection 6.C.1.

Regarding PSF parameters, the behavior is different for (w_α, w_β) and φ. The chain of the two width parameters has a very good exploration of the parameter space with a quasi-instantaneous burn-in period. Conversely, the chain of φ is more concentrated, and its burn-in period is approximately 4,000 samples. This is also in accordance with previous analysis (Subsection 6.C.2).



(a) (γ_1, w_α)

(b) (γ_1, w_β)

Fig. 7. Joint histograms for the couple (a) (γ_1, w_α) and (b) (γ_1, w_β). The x and y axes are the parameter values.

Table 3. Acceptance Rate

	Parameter		
	w_α	w_β	φ
Acceptance rate	14.50 %	9.44 %	2.14 %

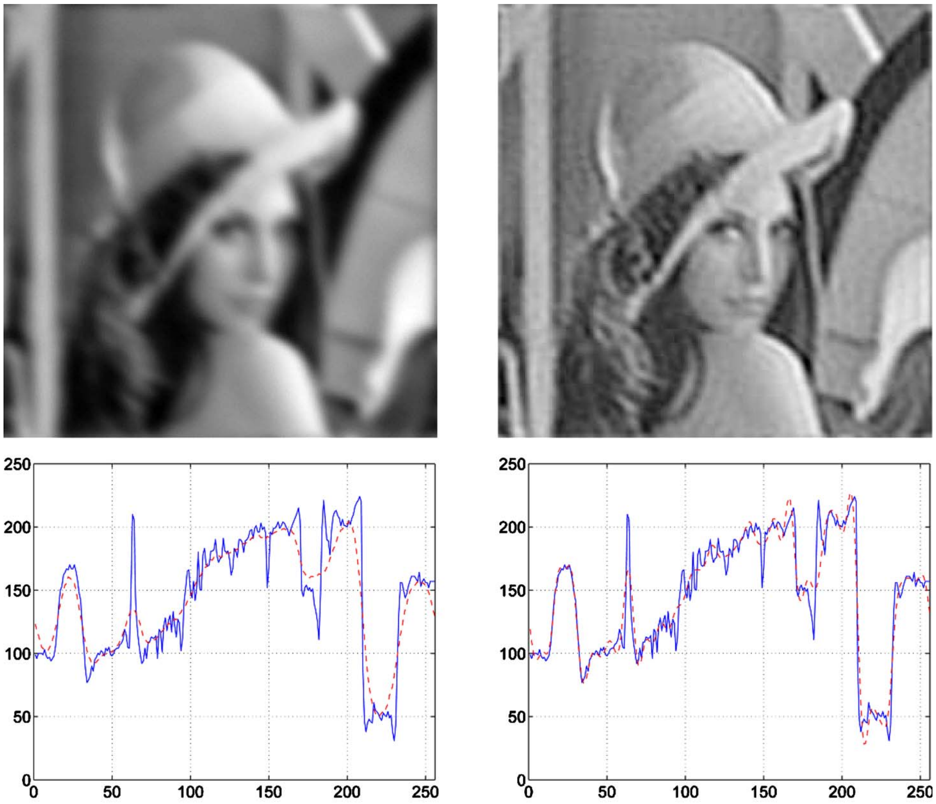
Acceptance rates in the Metropolis–Hastings algorithm are reported in Table 3: they are quite small, especially for the rotation parameter. This is due to the structure of the implemented algorithm: an independent Metropolis–Hastings algorithm with the prior law as a proposition law. The main advantage of this choice is its simplicity, but as a counterpart, a high rejection rate is observed due to a large *a priori* interval for the angle parameter. A future study will be devoted to the design of a more accurate proposition law.

E. Robustness of Prior Image Model

Figure 8 illustrates the proposed method on a more realistic image with heterogeneous spatial structures. The original is the Lena image, and the data have been obtained with the same Gaussian PSF and also with corruption by white Gaussian noise. Figure 8(b) shows that the restored image is closer to the true one than are the data. Smaller structures are visible and edges are sharper, for example, around pixel 200. The estimated parameters are $\hat{\gamma}_\epsilon = 1.98$ while the true value is $\gamma_\epsilon^* = 2$. Concerning the PSF parameters, the results are $\widehat{w_\alpha} = 19.3$, $\widehat{w_\beta} = 7.5$ and $\hat{\varphi} = 1.15$, while the true values are, respectively, $w_\alpha^* = 20$, $w_\beta^* = 7$ and $\varphi^* = 1.05$ as in Subsection 6.B. Here again, the estimated PSF parameters are close to the true values, giving an initial assessment of the capability of the method in a more realistic context.

7. CONCLUSION AND PERSPECTIVES

This paper presents a new global and coherent method for myopic and unsupervised deconvolution of relatively smooth images. It is built within a Bayesian framework



(a) Data (b) Estimated image

Fig. 8. (Color online) (a) Observed image and (b) restored image. Profiles correspond to the 68th line. The solid curve is the true profile, and the dashed curve correspond to (a) data and (b) estimated profiles.

and a proper extended *a posteriori* law for the PSF parameters, the hyperparameters, and the image. The estimate, defined as the posterior mean, is computed by means of an MCMC algorithm in less than a few minutes.

Numerical assessment attests that the parameters of the PSF and the parameters of the prior laws are precisely estimated. In addition, results also demonstrate that the myopic and unsupervised deconvolved image is closer to the true image than are the data and shows true restored high frequencies as well as spatial details.

The paper focuses on a linear invariant model often encountered in astronomy, medical imaging, and nondestructive testing and especially in optical problems. Noninvariant linear models can also be considered in order to address other applications such as spectrometry [4] or fluorescence microscopy [13]. The loss of the invariance property precludes entirely Fourier-based computations, but the methodology remains valid and practicable. In particular, it is possible to draw samples of the image by means of an optimization algorithm [38].

Gaussian law, related to L_2 penalization, is known for possible excessive sharp edges penalization in the restored object. The use of convex L_2-L_1 penalization [39–41] or non-convex L_2-L_0 penalization [42] can overcome this limitation. In these cases a difficulty occurs in the development of myopic and unsupervised deconvolution: the partition function of the prior law for the image is in intricate or even unknown dependency w.r.t. the parameters [1,7,43]. However a recent paper [41] overcomes the difficulty, resulting in an efficient unsupervised deconvolution, and we plan to extend this work for the myopic case.

Regarding noise, Gaussian likelihood limits robustness to outliers or aberrant data, and it is possible to appeal to a robust law such as Huber penalization in order to bypass the limitation. Nevertheless, the partition function for the noise law is again difficult or impossible to manage, and it is possible to resort to the idea proposed in [41] to overcome the difficulty.

Finally, estimation of the parameters of the correlation matrix (cutoff frequency, attenuation coefficients, etc.) is possible within the same methodological framework. This could be achieved for the correlation matrix of the object or the noise. As for the PSF parameters, the approach could rely on an extended *a posteriori* law, including the new parameters and a Metropolis–Hastings sampler.

APPENDIX A: LAW IN FOURIER SPACE

For a Gaussian vector $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the law for $\mathring{\boldsymbol{x}} = \boldsymbol{F}\boldsymbol{x}$ (the FFT of $\boldsymbol{x}$) is also Gaussian, whose first two moments are the following:

- The mean is

$$\mathring{\boldsymbol{\mu}} = \mathbb{E}[\mathring{\boldsymbol{x}}] = \boldsymbol{F}\mathbb{E}[\mathring{\boldsymbol{x}}] = \boldsymbol{F}\boldsymbol{\mu}. \tag{A1}$$

- The covariance matrix is

$$\mathring{\boldsymbol{\Sigma}} = \mathbb{E}[(\mathring{\boldsymbol{x}} - \mathring{\boldsymbol{\mu}})(\mathring{\boldsymbol{x}} - \mathring{\boldsymbol{\mu}})^{\dagger}] = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\dagger}. \tag{A2}$$

Moreover, if the covariance matrix $\boldsymbol{\Sigma}$ is circulant, it is written

$$\mathring{\boldsymbol{\Sigma}} = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^{\dagger} = \Lambda_{\boldsymbol{\Sigma}}, \tag{A3}$$

i.e., the covariance matrix $\mathring{\boldsymbol{\Sigma}}$ is diagonal.

APPENDIX B: GAMMA PROBABILITY DENSITY

1. Definition

The gamma pdf for $\gamma > 0$, with given parameter $\alpha > 0$ and $\beta > 0$, is written

$$\mathcal{G}(\gamma|\alpha,\beta) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\gamma^{\alpha-1}\exp(-\gamma/\beta). \tag{B1}$$

Table 4 gives three limit cases for (α,β). The following properties hold:

- The mean is $\mathbb{E}_{\mathcal{G}}[\gamma] = \alpha\beta$.
- The variance is $\mathbb{V}_{\mathcal{G}}[\gamma] = \alpha\beta^2$.
- The maximizer is $\beta(\alpha-1)$ if and only if $\alpha > 1$.

2. Marginalization

First consider an N-dimensional zero-mean Gaussian vector with a given precision matrix $\gamma\boldsymbol{\Gamma}$ with $\gamma > 0$. The pdf reads

$$p(\boldsymbol{x}|\gamma) = (2\pi)^{-N/2}\gamma^{N/2}\det[\boldsymbol{\Gamma}]^{1/2}\exp[-\gamma\boldsymbol{x}^{\mathrm{t}}\boldsymbol{\Gamma}\boldsymbol{x}/2]. \tag{B2}$$

So consider the conjugate pdf for γ as a gamma law with parameters (α,β) (see Appendix B.1). The joint law for $(\boldsymbol{x},\gamma)$ is the product of the pdf given by Eqs. (B1) and (B2): $p(\boldsymbol{x},\gamma) = p(\boldsymbol{x}|\gamma)p(\gamma)$. The marginalization of the joint law is known [44]:

$$\begin{aligned} p(\boldsymbol{x}) &= \int_{\mathbb{R}_+} p(\boldsymbol{x}|\gamma)p(\gamma)\mathrm{d}\gamma \\ &= \frac{\beta^{N/2}\det[\boldsymbol{\Gamma}]^{1/2}\Gamma(\alpha+N/2)}{(2\pi)^{N/2}\Gamma(\alpha)}\left(1+\frac{\beta\boldsymbol{x}^{\mathrm{t}}\boldsymbol{\Gamma}\boldsymbol{x}}{2}\right)^{-\alpha-N/2}, \end{aligned} \tag{B3}$$

which is an N-dimensional t-Student law of 2α degrees of freedom with a $\beta\Gamma$ precision matrix. Finally, the conditional law reads

$$p(\gamma|\boldsymbol{x}) = \frac{(2\pi)^{-N/2}\det[\boldsymbol{\Gamma}]^{1/2}}{\beta^{\alpha}\Gamma(\alpha)}\gamma^{\alpha+N/2-1}\exp[-\gamma\boldsymbol{x}^{\mathrm{t}}\boldsymbol{\Gamma}\boldsymbol{x}/2 + 1/\beta]. \tag{B4}$$

Thanks to conjugacy, it is also a gamma pdf with parameters $\bar{\alpha},\bar{\beta}$ given by $\bar{\alpha} = \alpha + N/2$ and $\bar{\beta}^{-1} = \beta^{-1} + 2/(\boldsymbol{x}^{\mathrm{t}}\boldsymbol{\Gamma}\boldsymbol{x})$.

Table 4. Specific Laws Obtained As Limit of the Gamma PDF

Law	α	β
Jeffreys	0	$+\infty$
Uniform	1	$+\infty$
Dirac	–	0

APPENDIX C: METROPOLIS–HASTINGS ALGORITHM

The Metropolis–Hastings algorithm provides samples of a target law $f(\boldsymbol{w})$ that cannot be directly sampled but can be evaluated, at least up to a multiplicative constant. Using the so called "instrument law" $q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)})$, samples of the target law are obtained by the following iterations:

1. Sample the proposition $\boldsymbol{w}_{\mathrm{p}} \sim q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)})$.
2. Compute the probability:

$$\rho = \min\left\{\frac{f(\boldsymbol{w}_{\mathrm{p}})}{f(\boldsymbol{w}^{(t)})}\frac{q(\boldsymbol{w}^{(t)}|\boldsymbol{w}_{\mathrm{p}})}{q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)})},1\right\}. \tag{C1}$$

3. Take

$$\boldsymbol{w}^{(t+1)} = \begin{cases} \boldsymbol{w}_{\mathrm{p}} & \text{with } \rho \text{ probability} \\ \boldsymbol{w}^{(t)} & \text{with } 1-\rho \text{ probability} \end{cases}. \tag{C2}$$

At convergence, the samples follow the target law $f(\boldsymbol{w})$ [25,36]. When $q(\boldsymbol{w}_{\mathrm{p}}|\boldsymbol{w}^{(t)}) = q(\boldsymbol{w}_{\mathrm{p}})$ the algorithm is named independent Metropolis–Hastings. In addition, if the instrument law is uniform, the acceptance probability becomes simpler:

$$\rho = \min\left\{\frac{f(\boldsymbol{w}_{\mathrm{p}})}{f(\boldsymbol{w}^{(t)})},1\right\}. \tag{C3}$$

APPENDIX D

Pseudo-code algorithm. `gamrnd`, `rand` and `randn` draw samples of the gamma-variable, uniform-variable, and zero-mean unit-variance white complex Gaussian vector, respectively.

```
Initialization of $[\overset{o}{\boldsymbol{x}}^{(0)}, \gamma^{(0)}, \boldsymbol{w}^{(0)}, k=0]$
repeat
    % Sample of $\overset{o}{\boldsymbol{x}}$
    $\Sigma \leftarrow \gamma_{\epsilon}^{(k)} |\boldsymbol{\Lambda}_{\boldsymbol{H}}|^2 + \gamma_0^{(k)} |\boldsymbol{\Lambda}_{\mathbb{1}}|^2 + \gamma_1^{(k)} |\boldsymbol{\Lambda}_{\boldsymbol{D}}|^2$
    $\boldsymbol{\mu} \leftarrow \gamma_{\epsilon}^{(k)} \Sigma^{-1} \boldsymbol{\Lambda}_{\boldsymbol{H}}^{*} \overset{o}{\boldsymbol{y}}$
    $\overset{o}{\boldsymbol{x}}^{(k)} \leftarrow \boldsymbol{\mu} + \Sigma^{-1/2}.*\mathrm{randn}$
    % Sample of $\gamma$
    $\gamma_{\epsilon}^{(k)} \leftarrow \mathrm{gamrnd}(\alpha_{\epsilon}, \beta_{\epsilon})$
    $\gamma_1^{(k)} \leftarrow \mathrm{gamrnd}(\alpha_1, \beta_1)$
    $\gamma_0^{(k)} \leftarrow \mathrm{gamrnd}(\alpha_0, \beta_0)$
    % Sample of $\boldsymbol{w}$
    $\boldsymbol{w}_{\mathrm{p}} \leftarrow \mathrm{rand}*(\boldsymbol{a}-\boldsymbol{b})+\boldsymbol{a}$
    $J \leftarrow \gamma_{\epsilon}(\|\overset{o}{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}} \overset{o}{\boldsymbol{x}}\|^2 - \|\overset{o}{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H},w_{\mathrm{p}}} \overset{o}{\boldsymbol{x}}\|^2)/2$
    if $\log(\mathrm{rand}) < \min\{J, 0\}$ then
        $\boldsymbol{w}^{(k)} \leftarrow w_{\mathrm{p}}$
        $\boldsymbol{\Lambda}_{\boldsymbol{H}} \leftarrow \boldsymbol{\Lambda}_{\boldsymbol{H},w_{\mathrm{p}}}$
    else
        $\boldsymbol{w}^{(k)} \leftarrow \boldsymbol{w}^{(k-1)}$
    end if
    % Empirical mean
    $k \leftarrow k+1$
    $\overset{o}{\boldsymbol{x}}^{(k)} \leftarrow \Sigma_i \overset{o}{\boldsymbol{x}}^{(i)}/k$
until $|\bar{\boldsymbol{x}}^{(k)} - \bar{\boldsymbol{x}}^{(k-1)}|/|\bar{\boldsymbol{x}}^{(k)}| \leqslant \mathrm{criterion}$
```

ACKNOWLEDGMENTS

The authors would like to thank Professor Alain Abergel of Laboratoire Institute d'Astrophysique Spatial at Université Paris-Sud 11, France, for fruitful discussions and constructive suggestions. The authors are also grateful to Cornelia Vacar, Laboratoire d'Intégration du Matériau au Système (IMS), Univ. Bordeaux, for carefully reading the paper.

REFERENCES

1. J. Idier, ed., *Bayesian Approach to Inverse Problems* (Wiley, 2008).
2. R. Molina, J. Mateos, and A. K. Katsaggelos, "Blind deconvolution using a variational approach to parameter, image, and blur estimation," IEEE Trans. Image Process. **15**, 3715–3727 (2006).
3. P. Campisi and K. Egiazarian, eds., *Blind Image Deconvolution* (CRC Press, 2007).
4. T. Rodet, F. Orieux, J.-F. Giovannelli, and A. Abergel, "Data inversion for over-resolved spectral imaging in astronomy," IEEE J. Sel. Top. Signal Process. **2**, 802–811 (2008).
5. A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems* (Winston, 1977).
6. S. Twomey, "On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature," J. Assoc. Comput. Mach. **10**, 97–101 (1962).
7. A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia, "Hyperparameter estimation for satellite image restoration by a MCMC maximum likelihood method," Pattern Recogn. **35**, 341–352 (2002).
8. J. A. O'Sullivan, "Roughness penalties on finite domains," IEEE Trans. Image Process. **4**, 1258–1268 (1995).
9. G. Demoment, "Image reconstruction and restoration: overview of common estimation structure and problems," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-37**, 2024–2036 (1989).
10. P. Pankajakshani, B. Zhang, L. Blanc-Féraud, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, "Blind deconvolution for thin-layered confocal imaging," Appl. Opt. **48**, 4437–4448 (2009).
11. E. Thiébaut and J.-M. Conan, "Strict *a priori* constraints for maximum likelihood blind deconvolution," J. Opt. Soc. Am. A **12**, 485–492 (1995).
12. N. Dobigeon, A. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," IEEE Trans. Image Process. (2009).
13. B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," Appl. Opt. **46**, 1819–1829 (2007).
14. L. Mugnier, T. Fusco, and J.-M. Conan, "MISTRAL: a myopic edge-preserving image restoration method, with application to astronomical adaptive-optics-corrected long-exposure images," J. Opt. Soc. Am. A **21**, 1841–1854 (2004).
15. E. Thiébaut, "MiRA: an effective imaging algorithm for optical interferometry," Proc. SPIE **7013**, 70131-I (2008).
16. T. Fusco, J.-P. Véran, J.-M. Conan, and L. M. Mugnier, "Myopic deconvolution method for adaptive optics images of stellar fields," Astron. Astrophys. Suppl. Ser. **134**, 193 (1999).
17. J.-M. Conan, L. Mugnier, T. Fusco, V. Michau, and G. Rousset, "Myopic deconvolution of adaptive optics images by use of object and point-spread function power spectra," Appl. Opt. **37**, 4614–4622 (1998).
18. A. C. Likas and N. P. Galatsanos, "A variational approach for Bayesian blind image deconvolution," IEEE Trans. Image Process. **52**, 2222–2233 (2004).
19. T. Bishop, R. Molina, and J. Hopgood, "Blind restoration of blurred photographs via AR modelling and MCMC," in *Proceedings of 15th IEEE International Conference on Image Processing, 2008, ICIP 2008* (IEEE Signal Processing Society, 2008).
20. E. Y. Lam and J. W. Goodman, "Iterative statistical approach to blind image deconvolution," J. Opt. Soc. Am. A **17**, 1177–1184 (2000).
21. Z. Xu and E. Y. Lam, "Maximum *a posteriori* blind image deconvolution with Huber–Markov random-field regularization," Opt. Lett. **34**, 1453–1455 (2009).
22. M. Cannon, "Blind deconvolution of spatially invariant image blurs with phase," IEEE Trans. Acoust., Speech, Signal Process. **24**, 58–63 (1976).
23. A. Jalobeanu, L. Blanc-Feraud, and J. Zerubia, "Estimation of blur and noise parameters in remote sensing," in *Proceedings of 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002)* (IEEE Signal Processing Society, 2002), Vol. 4, pp. 3580–3583.
24. F. Chen and J. Ma, "An empirical identification method of Gaussian blur parameter for image deblurring," IEEE Trans. Signal Process. (2009).
25. C. P. Robert and G. Casella, *Monte-Carlo Statistical Methods*, Springer Texts in Statistics (Springer, 2000).
26. B. R. Hunt, "A matrix theory proof of the discrete convolution theorem," IEEE Trans. Autom. Control **AC-19**, 285–288 (1971).
27. M. Calder and R. A. Davis, "Introduction to Whittle (1953) 'The analysis of multiple stationary time series'," Breakthroughs in Statistics **3**, 141–148 (1997).
28. P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods* (Springer-Verlag, 1991).
29. B. R. Hunt, "Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory," IEEE Trans. Autom. Control **AC-17**, 703–705 (1972).
30. K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis* (Academic, 1992), Chap. 2, pp. 36–43.
31. C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," IEEE Trans. Image Process. **2**, 296–310 (1993).
32. D. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ. Press, 2003).
33. R. E. Kass and L. Wasserman, "The selection of prior dis-

tributions by formal rules," J. Am. Stat. Assoc. **91**, 1343–1370 (1996).

34. E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge Univ. Press, 2003).
35. S. Lang, *Real and Functional Analysis* (Springer, 1993).
36. P. Brémaud, *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues*, Texts in Applied Mathematics 31 (Springer, 1999).
37. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).
38. P. Lalanne, D. Prévost, and P. Chavel, "Stochastic artificial retinas: algorithm, optoelectronic circuits, and implementation," Appl. Opt. **40**, 3861–3876 (2001)
39. H. R. Künsch, "Robust priors for smoothing and image restoration," Ann. Inst. Stat. Math. **46**, 1–19 (1994).
40. P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," IEEE Trans. Image Process. **6**, 298–311 (1997).
41. J.-F. Giovannelli, "Unsupervised Bayesian convex deconvolution based on a field with an explicit partition function," IEEE Trans. Image Process. **17**, 16–26 (2008).
42. D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," IEEE Trans. Image Process. **4**, 932–946 (1995).
43. X. Descombes, R. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood," IEEE Trans. Image Process. **8**, 954–963 (1999).
44. G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis* (Addison-Wesley, 1972).