Unsupervised Bayesian Convex Deconvolution Based on a Field With an Explicit Partition Function

Jean-François Giovannelli

Abstract—This paper proposes a non-Gaussian Markov field with a special feature: an *explicit* partition function. To the best of our knowledge, this is an original contribution. Moreover, the explicit expression of the partition function enables the development of an *unsupervised* edge-preserving convex deconvolution method. The method is fully Bayesian, and produces an estimate in the sense of the posterior mean, numerically calculated by means of a Monte-Carlo Markov chain technique. The approach is particularly effective and the computational practicability of the method is shown on a simple simulated example.

Index Terms—Bayesian statistics, convex potentials, deconvolution, hyperparameters estimation, Monte-Carlo Markov chain, partition function, regularization, unsupervised estimation.

I. INTRODUCTION

THE research concerning regularization for ill-posed inverse problems was first carried out by Phillips *et al.* in the sixties and are compiled in [1]. For the specific problem of deconvolution they lead to the contributions of Hunt [2] based on toroidal models and fast implementation by the fast Fourier transform (FFT). These methods rely on quadratic penalization, i.e., Gaussian laws in a Bayesian framework. The solutions thus formulated are linear w.r.t. the data and numerically efficient. However, their resolution is limited: the capability to properly restore sharp edges is limited.

At the beginning of the 1980s, in order to overcome these limitations, Geman and Geman [3] (see also [4]) introduced a much superior Markovian field including hidden variables [5]. The hidden variables (also referred to as dual or auxiliary variable) are binary and interactive variables modeling sharp edges and closed contours. The data processing then relies on a detection-estimation strategy and allows the recovery of distinct zones with abrupt changes. The calculation of the solution in the sense of the maximum *a posteriori* is based on a simulated annealing algorithm which requires intensive numerical computations. For the sake of computational efficiency in some cases, Geman and Reynolds [6] and then Geman and Yang [7] introduced auxiliary (also referred to as dual) variables: the sampling of a correlated non-Gaussian field reduces to the sampling of a correlated Gaussian field for one part and to the sampling of a

separable field for the other. Furthermore, the construction proposed by [7] is founded on the work of Hunt and the toroidal models: the sampling of the correlated Gaussian field reduces to the sampling of an inhomogeneous white Gaussian field followed by an FFT. The proposal below takes advantage of this construction.

The case of fields with convex potential [8]–[13] (see also [14] and [15]) was laid down in the nineties as fulfilling a compromise between the quality of the reconstructed images and the computational burden. In this framework, a particular attention has been paid to the case of $L_2 - L_1$ potentials [9]–[13]: a quadratic behavior around the origin and a linear behavior at large values allow edge preservation. In this context, the constructions of [6] and [7] respectively led to two algorithms: Arthur and Legend [16] (see also [17]). The work presented here concerns this type of potential.

With such potentials, the regularized solutions usually necessitate the adjustment of three hyperparameters: two parameters to control the law for the image and one parameter to control the law for the noise. Several attempts are dedicated to the question of hyperparameter estimation and the investigated solutions are frequently based on statistical approaches: (approximated or pseudo) likelihood, Bayesian strategies, EM and SEM algorithms, etc. The reader may consult papers such as [18]–[24] and reference books such as [25, part VI], [26, Ch. 7], or [27, Ch. 8]. These approaches are potentially very powerful but they come up against a major difficulty: the partition function of existing *a priori* fields depends on hyperparameters and is not explicitly given.

The first novelty of the paper lies in the fact that it proposes a new random field with an explicit partition function. To this end, the paper build an original type of compound (toroidal) field with $L_2 - L_1$ potential. The work is largely inspired by the Bayesian interpretation of dual variables in terms of location mixture of Gaussian proposed by [28]. Moreover, it is also inspired by [29] (itself based on the contributions of Hunt [2] and Geman and Yang [7]). However, none of these contributions put forward the idea of a field with an explicit partition function. Afterward, the paper proposes a second novelty: a full Bayesian unsupervised (i.e., including hyperparameter estimation) edge-preserving convex deconvolution method, thanks to the knowledge of the partition function. It is based on a *posterior* law for the whole set of unknown parameters (including hyperparameters) and a minimum mean square error strategy.

The paper is presented in the following manner. Section II introduces the notations and states the problem. Section III is devoted to the construction of the proposed field, and Section IV proposes its use for image deconvolution and demonstrates the

numerical practicability. Conclusions and perspectives are delivered in Section V. Most of the calculations are explained in Appendices I–VIII.

II. NOTATION AND PROBLEM STATEMENT

Work is carried out on $P \times P$ real images, with $N = P^2$ pixels, represented in a matrix form. a_{pq} denotes the generic element of the matrix $\boldsymbol{A}$, $N_2(\boldsymbol{A}) = \sum |a_{pq}|^2$ its squared norm and $\overset{\circ}{\boldsymbol{A}}$ its FFT 2-D. The transformation is normalized: the Parseval relationship is written as $N_2(\boldsymbol{A}) = N_2(\overset{\circ}{\boldsymbol{A}})$ and the sum of the pixels is $\sum a_{pq} = P \overset{\circ}{a}_{00}$. The symbols $\star$ and $\otimes$ respectively represent the circular convolution and the Schur product (termwise) of matrices. If $\boldsymbol{F}$ represents a circular filter and $\boldsymbol{I}$ an input object, the output is written $\boldsymbol{O} = \boldsymbol{F} \star \boldsymbol{I}$ in the spatial domain resulting in $\overset{\circ}{\boldsymbol{O}} = \overset{\circ}{\boldsymbol{F}} \otimes \overset{\circ}{\boldsymbol{I}}$ in the Fourier domain. If $\overset{\circ}{f}_{pq} \neq 0$ for all p, q, the associated filter is invertible.

In the subsequent developments about deconvolution, $\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{H}$, and $\boldsymbol{N}$, respectively, denote the observed data, the unknown object, the convolution matrix and the observation noise. With these notations, the observation equation is written

$$\boldsymbol{Y} = \boldsymbol{H} \star \boldsymbol{X} + \boldsymbol{N}. \tag{1}$$

The deconvolution problem consists in recovering the unknown object $\boldsymbol{X}$ given the observed data $\boldsymbol{Y}$ and given the observation model $\boldsymbol{H}$. The ill-posedness of the problem has been well identified for several decades and the problem is nowadays often tackled in a Bayesian framework using Markov priors. In a Gibbs form, the prior law writes

$$f_{\mathcal{X}}[\boldsymbol{X}] = K_{\mathcal{X}}^{-1} \exp[-\Phi_{\boldsymbol{\theta}}(\boldsymbol{X})]$$

where $K_{\mathcal{X}}$ is the partition function (normalizing constant) and $\Phi_{\boldsymbol{\theta}}$ is the Gibbs energy controlled by a set of parameters (such as variance, threshold, scale, correlation length, etc.) collected in a vector $\boldsymbol{\theta}$. The general methodology is well known: the solution is determined from the *a posteriori* law and a point estimate can be chosen as the mean or the maximizer, for instance. Anyway, the posterior law (and the point estimates) depends upon hyperparameters notably on the parameters of the prior $\boldsymbol{\theta}$. The inference about these parameters can be attempted in a statistical framework whose keystone is an exact and explicit likelihood function (in an usual sense or in a posterior sense). This function is itself founded on a complete expression for the prior law including the partition function as it depends on $\boldsymbol{\theta}$. It is given as a large dimension integral

$$K_{\mathcal{X}} = K_{\mathcal{X}}(\boldsymbol{\theta}) = \int_{\mathbb{R}^N} \exp[-\Phi_{\boldsymbol{\theta}}(\boldsymbol{X})] \mathrm{d}\boldsymbol{X}. \tag{2}$$

It is a commonplace to say that $K_{\mathcal{X}}(\boldsymbol{\theta})$ can be explicitly given for two well-known classes of (continuous state) field:

i) Φ is quadratic, i.e., the field is Gaussian;
ii) Φ is separable, i.e., the field is white.

In other cases, and especially for nonseparable and non-Gaussian fields, the theoretical calculation and the numerical computation of (2) are desperate tasks [25, p. 281]

and they have never been achieved.[1] However, its achievement is made possible and simple in the next section, for a special nonseparable and non-Gaussian field.

III. PRIOR FIELD WITH PARTITION FUNCTION

Taking advantage of i) and ii) the proposed random field is a compound field involving two variables: a pixel variable noted as $\mathcal{X}$ and an auxiliary (or dual or hidden) variable noted as $\mathcal{B}$. The joint law for $(\mathcal{X}, \mathcal{B})$ is defined by the law of $\mathcal{X} \mid \mathcal{B}$ for one part and by the law of $\mathcal{B}$ for the other part. The former is a Gaussian component [case i)] and the latter is a separable component (case ii)].

A. Toroidal Gaussian Field for $\mathcal{X} \mid \mathcal{B}$

Let us consider two matrices $\boldsymbol{B}$ and $\boldsymbol{F}$ with $\overset{\circ}{f}_{pq} \neq 0$ for all p, q and the toroidal (circular shift invariant) Gaussian field with a density parametrized in the form

$$f_{\mathcal{X} \mid \mathcal{B}}[\boldsymbol{X} \mid \boldsymbol{B}] = K_{\mathcal{X} \mid \mathcal{B}}^{-1} \exp[-\gamma_{\mathrm{d}} N_2(\boldsymbol{F} \star \boldsymbol{X} - \boldsymbol{B})/2] \tag{3}$$

where $\gamma_{\mathrm{d}} > 0$ is an inverse variance. The matrix $\boldsymbol{F}$ designs the field structure and especially the neighborhood system and the form of the cliques. In the Fourier domain, the potential is separable and naturally develops in two forms

$$\begin{aligned} N_2(\boldsymbol{F} \star \boldsymbol{X} - \boldsymbol{B}) &= N_2(\overset{\circ}{\boldsymbol{F}} \otimes \overset{\circ}{\boldsymbol{X}} - \overset{\circ}{\boldsymbol{B}}) \\ &= \sum_{pq} | \overset{\circ}{f}_{pq} \overset{\circ}{x}_{pq} - \overset{\circ}{b}_{pq} |^2 \\ &= \sum_{pq} | \overset{\circ}{f}_{pq} |^2 | \overset{\circ}{x}_{pq} - \overset{\circ}{b}_{pq} / \overset{\circ}{f}_{pq} |^2 \end{aligned}$$

which has three essential consequences for the following developments.

1) The law for $\overset{\circ}{\mathcal{X}}$ is separable and each $\overset{\circ}{X}_{pq}$ is Gaussian with mean $\overset{\circ}{b}_{pq} / \overset{\circ}{f}_{pq}$ and inverse variance $\gamma_{\mathrm{d}} | \overset{\circ}{f}_{pq} |^2$. As a result, the sampling of $\mathcal{X}$ reduces to the sampling of an inhomogeneous white Gaussian noise followed by an FFT 2-D.

2) The change of variable $\overline{\mathcal{X}} = \boldsymbol{F} \star \mathcal{X}$ is invertible, $\overline{\mathcal{X}}$ is white and each $\bar{X}_{pq}$ is Gaussian with mean b_{pq} and common inverse variance γ_{d}.

3) The partition function $K_{\mathcal{X} \mid \mathcal{B}}$ is easily tractable in the Fourier domain thanks to a change of variable

$$\begin{aligned} K_{\mathcal{X} \mid \mathcal{B}} &= \int_{\mathbb{R}^N} \exp[-\gamma_{\mathrm{d}} N_2(\boldsymbol{F} \star \boldsymbol{X} - \boldsymbol{B})/2] \mathrm{d}\boldsymbol{X} \\ &= \left[\gamma_{\mathrm{d}}^{N/2} \ (2\pi)^{-N/2} \ \prod | \overset{\circ}{f}_{pq} | \right]^{-1} \end{aligned}$$

and does not depend on $\boldsymbol{B}$.

In relation to existing works such as [7], [16], and [27]–[29], the main idea here is simply to focus on the case where the change of variable $\overline{\mathcal{X}} = \boldsymbol{F} \star \mathcal{X}$ is invertible (point 2 above) that is to say the number of cliques and the number of pixels are equal.

---

[1]The partition function is, however, known for the Ising field [30]. It is a binary field out of the scope of the developed work.

Remark 1: The partition function $K_{\mathcal{X}|\mathcal{B}}$ does not depend on $\boldsymbol{B}$ as a counterpart of a limitation: the number of cliques and the number of pixels are equal. As an illustration of the limitation, let us point out that $K_{\mathcal{X}|\mathcal{B}}$ depends on $\boldsymbol{B}$ for a field based on horizontal cliques plus vertical cliques (the number of cliques is greater than the number of pixels).

B. Compound Field

A separable and homogeneous field is then introduced for the auxiliary variable $\mathcal{B}$ with a density $f_{\mathcal{B}}[\boldsymbol{B}]$, product of the $f_B[b_{pq}]$. The joint density is written as $f_{\mathcal{X},\mathcal{B}}[\boldsymbol{X},\boldsymbol{B}] = f_{\mathcal{X}|\mathcal{B}}[\boldsymbol{X}|\boldsymbol{B}]f_{\mathcal{B}}[\boldsymbol{B}]$ and the marginal law is obtained by integrating the auxiliary variables

$$f_{\mathcal{X}}[\boldsymbol{X}] = \int_{\mathbb{R}^N} f_{\mathcal{X}|\mathcal{B}}[\boldsymbol{X}|\boldsymbol{B}]f_{\mathcal{B}}[\boldsymbol{B}]\mathrm{d}\boldsymbol{B}.$$

Since the partition function $K_{\mathcal{X}|\mathcal{B}}$ does not depend on $\boldsymbol{B}$, the calculations can be achieved

$$
\begin{aligned}
&f_{\mathcal{X}}[\boldsymbol{X}]\\
&= K_{\mathcal{X}|\mathcal{B}}^{-1} \int_{\mathbb{R}^N} f_{\mathcal{B}}[\boldsymbol{B}]\exp[-\gamma_{\mathrm{d}}N_2(\boldsymbol{F}\star\boldsymbol{X}-\boldsymbol{B})/2]\,\mathrm{d}\boldsymbol{B}\\
&= K_{\mathcal{X}|\mathcal{B}}^{-1} \prod_{pq}\int_{\mathbb{R}} f_B[b_{pq}]\exp[-\gamma_{\mathrm{d}}(\bar{x}_{pq}-b_{pq})^2/2]\mathrm{d}b_{pq}
\end{aligned}
$$

which involves a separable convolution product.

Remark 2: The proposed construction is possible for any probability density function f_B. In this sense, it is possible to design a large class of potential functions.

Thus, a wide range of law is available, but the convex potential case is the one of interest here, as mentioned in the introduction. So, the following property is of importance.

Property 1: For any log-concave probability density function f_B, the probability density function $f_{\mathcal{X}}$ is log-concave [31, Theorem 7], [32].

C. Laplace Law for Auxiliary Variables

The following developments are dedicated to the case of auxiliary variables under a Laplace law suggested by [28].

As mentioned by [28] itself, among the Huber-like distributions, such a Laplace-convolved-Gauss probability will have two main advantages: i) the convolution involved in the marginal law $f_{\mathcal{X}}$ (Section III-B) will be made explicit and ii) the sampling of auxiliary variables (Section IV-C) will be directly feasible thanks to the inversion of the cumulative density function $F_{\mathcal{B}|\mathcal{X}}$. The Laplace law is written in the form

$$f_{\mathcal{B}}[\boldsymbol{B}] = K_{\mathcal{B}}^{-1}\exp[-\gamma_{\mathrm{b}}N_1(\boldsymbol{B})/2] \qquad (4)$$

where $\gamma_{\mathrm{b}} > 0$ is a scale parameter, and $N_1(\boldsymbol{B}) = \sum|b_{pq}|$ is the L_1 norm. The partition function is simply calculated thanks to separability

$$K_{\mathcal{B}} = \int_{\mathbb{R}^N}\exp[-\gamma_{\mathrm{b}}N_1(\boldsymbol{B})/2]\mathrm{d}\boldsymbol{B} = [\gamma_{\mathrm{b}}/4]^{-N}.$$

According to (3) and (4), the joint density for $(\mathcal{X},\mathcal{B})$ takes the form

$$f_{\mathcal{X},\mathcal{B}}[\boldsymbol{X},\boldsymbol{B}] = K_{\mathcal{X},\mathcal{B}}^{-1}\exp[-\gamma_{\mathrm{d}}N_2(\boldsymbol{F}\star\boldsymbol{X}-\boldsymbol{B}) + \gamma_{\mathrm{b}}N_1(\boldsymbol{B})/2] \qquad (5)$$

and the partition function is explicit: $K_{\mathcal{X},\mathcal{B}} = K_{\mathcal{X}|\mathcal{B}}K_{\mathcal{B}}$.

The marginal law for $\mathcal{X}$ involves the 1-D convolution of a Gaussian density and a Laplacian density

$$
\begin{aligned}
&f_{\mathcal{X}}[\boldsymbol{X}]\\
&= K_{\mathcal{X},\mathcal{B}}^{-1}\prod_{pq}\int_{\mathbb{R}}\exp[-\gamma_{\mathrm{d}}(\bar{x}_{pq}-b_{pq})^2\\
&\qquad + \gamma_{\mathrm{b}}|b_{pq}|/2]\mathrm{d}b_{pq}\\
&= K_{\mathcal{X},\mathcal{B}}^{-1}\prod_{pq} I(+\infty,\bar{x}_{pq},\gamma_{\mathrm{d}},\gamma_{\mathrm{b}})
\end{aligned}
$$

where I is defined in Appendix II. Thus, the potential function φ appears

$$f_{\mathcal{X}}[\boldsymbol{X}] = K_{\mathcal{X},\mathcal{B}}^{-1}\quad\exp\left[-\sum_{pq}\varphi(\bar{x}_{pq})/2\right]$$

with

$$\varphi(x) = -2\log I(+\infty,x,\gamma_{\mathrm{d}},\gamma_{\mathrm{b}}). \qquad (6)$$

It is named the log-erf potential and it is shown in Fig. 3. The details of the calculations concerning this potential are given in Appendix III. Concerning the first derivative, one has

$$\varphi'(0) = 0 \quad\text{and}\quad \varphi'(+\infty) = \gamma_{\mathrm{b}}$$

and concerning the second derivative at origin, one has

$$\varphi''(0) = \frac{\gamma_{\mathrm{b}}^2}{2}[(\eta\sqrt{\pi}\,\mathrm{erfcx}[\eta])^{-1} - 1]$$

with $\eta = \gamma_{\mathrm{b}}/\sqrt{8\gamma_{\mathrm{d}}}$ ($\mathrm{erfcx}[\cdot]$ is given in Appendix I). As expected (see Property 1), this is a convex potential. It is a $\mathrm{L}_2 - \mathrm{L}_1$ potential which can be reconciled with other more common $\mathrm{L}_2 - \mathrm{L}_1$ potentials (Huber, log-cosh, hyperbolic, fair function). In the case of the Huber potential

$$x \mapsto \lambda\begin{cases} x^2, & \text{if } |x| \leq s\\ 2s|x| - s^2, & \text{if } |x| \geq s\end{cases} \qquad (7)$$

by identifying the second derivatives at zero and the slopes at infinity, one has

$$\lambda = \frac{1}{2}\varphi''(0) \quad\text{and}\quad s = \frac{\varphi'(+\infty)}{\varphi''(0)}. \qquad (8)$$

Compared log-erf and Huber potentials and their derivatives are shown in Fig. 3. Using the expansions (14) and (13) of Appendix I, two limit cases can be identified, according to the value of the ratio η

$$\text{for } \eta \gg 1: \lambda \simeq \gamma_{\mathrm{d}}; \quad s \simeq \frac{\gamma_{\mathrm{b}}}{2\gamma_{\mathrm{d}}}$$

$$\text{for } \eta \ll 1: \lambda \simeq \gamma_{\mathrm{b}}\sqrt{\frac{\gamma_{\mathrm{d}}}{2\pi}}; \quad s \simeq \sqrt{\frac{\pi}{2\gamma_{\mathrm{d}}}}.$$

In the two limit cases, on a log-log scale, there is linear behavior of λ and s as a function of γ_{b}, for a fixed γ_{d} (see Fig. 4). The intersection of the two linear behaviors can be identified as a
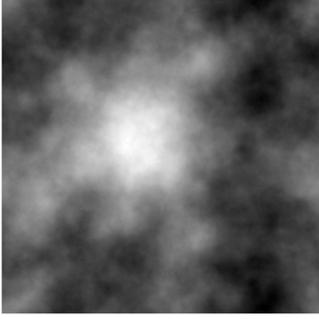
Fig. 1. Sample of the field, with $\gamma_{\mathrm{d}} = \gamma_{\mathrm{b}} = 1$ (ε is also set to 1).
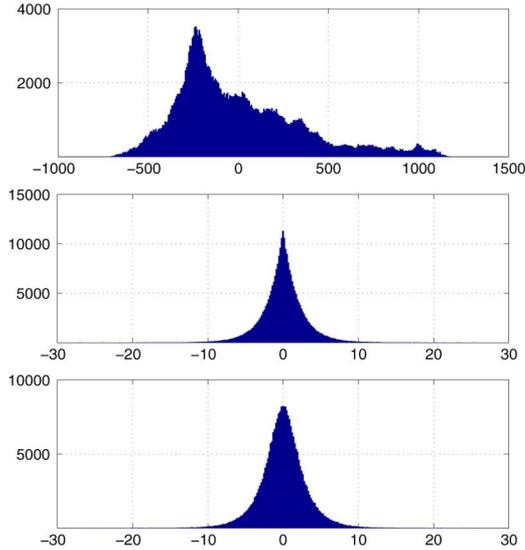

Fig. 2. Histograms (image at Fig. 1). From top to bottom: Histogram of image pixels X, histogram of auxiliary variables B, and histogram of differences $\bar{X}$.
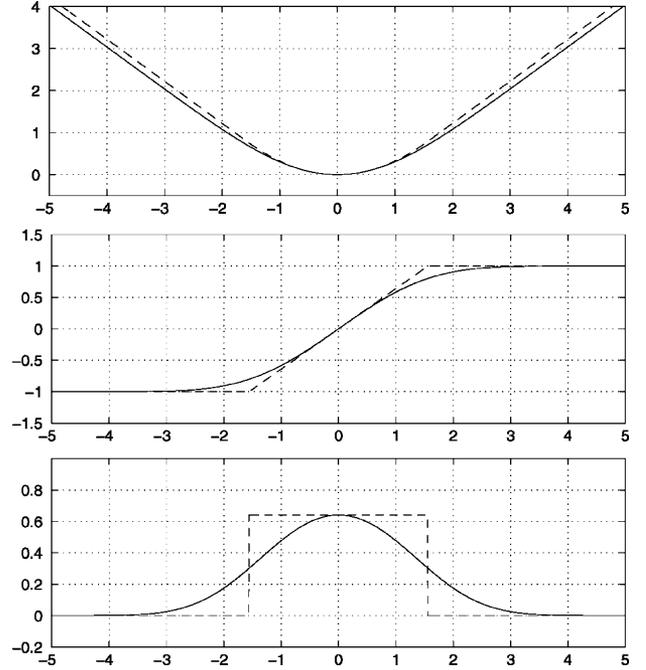

Fig. 3. From top to bottom: Potential function, first and second derivative. Solid line: log-erf potential $\varphi(x)$ of (6) and dotted line: corresponding Huber potential of (7). The potential parameters are $\gamma_{\mathrm{d}} = \gamma_{\mathrm{b}} = 1$, and, hence, the equivalent Huber parameters are $\lambda \simeq 0.32$ and $s \simeq 1.56$, according to (8).
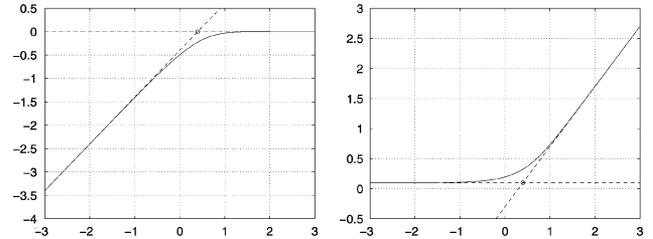

Fig. 4. λ and s as a function of γ_{b}, for a fixed $\gamma_{\mathrm{d}}(\gamma_{\mathrm{d}} = 1)$ on a log-log scale. As expected, the plot essentially shows two linear behaviors and a critical case for $\gamma_{\mathrm{b}} = \sqrt{2\pi}(\log_{10}\sqrt{2\pi} \sim 0.4)$.

critical behavior for $\gamma_{\mathrm{b}} = \sqrt{2\pi\gamma_{\mathrm{d}}}$. The critical value will be used for the initialization of simulations of Section IV-D (see also Appendix VIII).

D. Practical Case

In practice, the field is based on a 3×3 Laplacian filter, defined by $[0, 1, 0; 1, -4, 1; 0, 1, 0]$ and represented by the matrix $\boldsymbol{D}$. At null frequency, one has $\overset{\circ}{d}_{00} = 0$ and as a consequence the mean level of the image is not managed. So, an extra parameter is introduced to drive the mean level: it is denoted by $\varepsilon(\varepsilon \geq 0)$ and the characteristic matrix $\boldsymbol{F}$ is set to $\boldsymbol{F}_\varepsilon = \boldsymbol{D} + \varepsilon$.

Remark 3: If $\varepsilon = 0$, the field cannot be normalized and each clique is formed from the four nearest neighbors (cross-like clique). If $\varepsilon \neq 0$, the field can be normalized and each clique is spread out over the entire image.

The following developments take $\varepsilon > 0$ and the partition function of the joint field writes:

$$K_{\mathcal{X},\mathcal{B}}^{-1} = \delta\varepsilon\gamma_{\mathrm{d}}^{N/2}\gamma_{\mathrm{b}}^{N}, \quad \text{with}$$
$$\delta = (32\pi)^{-N/2} \prod_{\substack{(p,q) \\ \neq (0,0)}} |\overset{\circ}{d}_{pq}|.$$

Fig. 1 gives a sample of the field with $\gamma_{\mathrm{d}} = \gamma_{\mathrm{b}} = 1$ and Fig. 2 gives histograms of the image pixels, the auxiliary variables B (a Laplace histogram) and the differences $\bar{X}$ (an over-Gaussian histogram).

Remark 4: It is noteworthy that the marginal model $\mathcal{X}$ is homogeneous, but the conditional model $\mathcal{X} \mid \mathcal{B}$ is nonhomogeneous (except if all the b_{pq} are equal).

IV. DECONVOLUTION

As a result of the previous section, a new random field is now available with a special feature: an explicit (and simple) partition function. In this section, the field serves as a prior in a deconvolution method whose specificity is to be unsupervised (i.e., including hyperparameter estimation). More precisely, the method relies on a full Bayesian framework and the solution is determined from an *a posteriori* law based on an *a priori* law (given as follows) for the object, the noise, and the hyperparameters.

A. Prior Choices

1) Object Law: The *a priori* field is defined in the previous section. The joint density for $(\mathcal{X}, \mathcal{B})$ is given by (5) and it is driven by three parameters: $\gamma_\mathrm{d}, \gamma_\mathrm{b}$, and ε.

2) Noise Law: This work is founded on the usual case of zero-mean white Gaussian noise with inverse variance denoted γ_n. The density is written

$$f_\mathcal{N}(\boldsymbol{N}) = (2\pi)^{-N/2} \quad \gamma_\mathrm{n}^{N/2} \exp[-\gamma_\mathrm{n}\, N_2(\boldsymbol{N})/2].$$

3) Hyperparameter Law: Four parameters are to be managed: $\gamma_\mathrm{n}, \gamma_\mathrm{d}, \gamma_\mathrm{b}$, and ε. The three parameters of major importance are $\boldsymbol{\gamma} = [\gamma_\mathrm{n}, \gamma_\mathrm{d}, \gamma_\mathrm{b}]$; the fourth parameter ε drives the prior mean level of the image and it is considered as a nuisance parameter. Anyway, very few is *a priori* known about these parameters and the idea is to use noninformative or diffuse and separable priors.

- The proposed prior law for the three parameters $\gamma_\mathrm{n}, \gamma_\mathrm{d}$ and γ_b is a conjugate law. It is a gamma law (see (15), Appendix IV) with parameters respectively denoted $(\alpha_\mathrm{n}, \beta_\mathrm{n}), (\alpha_\mathrm{d}, \beta_\mathrm{d})$, and $(\alpha_\mathrm{b}, \beta_\mathrm{b})$. It allows for easy computations with the posterior law. Moreover, it includes diffuse and noninformative prior: the uniform prior and the Jeffrey's prior are obtained as limit cases for $(\alpha, \beta) = (1, \infty)$ and for $(\alpha, \beta) = (0, \infty)$ respectively.
- The last parameter ε is considered as a nuisance parameter and the proposed strategy resorts to integration out. The desired prior law is a Dirac law, so that no information is accounted for about the mean level of the image (it is set on the basis of observed data only). Formally, in a first step, a uniform density over $[0, M_\varepsilon]$ is introduced and in a second step the limit law for $M_\varepsilon \to 0$ is considered.

B. Joint Law

Thus, the joint law is established for $(\mathcal{Y}, \mathcal{X}, \mathcal{B}, \mathcal{C}, \mathcal{E})$

$$
\begin{aligned}
&f_{\mathcal{Y},\mathcal{X},\mathcal{B},\mathcal{C},\mathcal{E}}(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}, \varepsilon)\\
&= \delta' \gamma_\mathrm{n}^{\alpha_\mathrm{n}-1+N/2} \gamma_\mathrm{d}^{\alpha_\mathrm{d}-1+N/2}\\
&\quad \times \gamma_\mathrm{b}^{\alpha_\mathrm{b}-1+N} \varepsilon M_\varepsilon^{-1} \mathbb{1}_{[0,M_\varepsilon]}(\varepsilon)\\
&\quad \times \exp-\{Q_\varepsilon/2 + \gamma_\mathrm{n}/\beta_\mathrm{n} + \gamma_\mathrm{d}/\beta_\mathrm{d} + \gamma_\mathrm{b}/\beta_\mathrm{b}\}
\end{aligned}
$$

where $\delta' = (2\pi)^{-N/2}\delta/\beta_\mathrm{n}^{\alpha_\mathrm{n}} \Gamma[\alpha_\mathrm{n}]\beta_\mathrm{b}^{\alpha_\mathrm{b}} \Gamma[\alpha_\mathrm{b}]\beta_\mathrm{d}^{\alpha_\mathrm{d}} \Gamma[\alpha_\mathrm{d}]$ is a normalization constant and Q_ε is part of the Co-logarithm of the density involving $\boldsymbol{F}_\varepsilon$

$$Q_\varepsilon = \gamma_\mathrm{n} N_2(\boldsymbol{Y} - \boldsymbol{H} \star \boldsymbol{X}) + \gamma_\mathrm{d} N_2(\boldsymbol{F}_\varepsilon \star \boldsymbol{X} - \boldsymbol{B}) + \gamma_\mathrm{b} N_1(\boldsymbol{B}).$$

The *a posteriori* density is formed for $\mathcal{X}, \mathcal{B}, \mathcal{C}$, and $\mathcal{E}$, given $\mathcal{Y}$ thanks to the Bayes rule

$$
\begin{aligned}
&f_{\mathcal{X},\mathcal{B},\mathcal{C},\mathcal{E}|\mathcal{Y}}(\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}, \varepsilon|\boldsymbol{Y})\\
&= \frac{f_{\mathcal{X},\mathcal{B},\mathcal{C},\mathcal{E},\mathcal{Y}}(\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}, \varepsilon, \boldsymbol{Y})}{\int_{\boldsymbol{X},\boldsymbol{B},\boldsymbol{\gamma},\varepsilon} f_{\mathcal{X},\mathcal{B},\mathcal{C},\mathcal{E},\mathcal{Y}}(\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}, \varepsilon, \boldsymbol{Y})\,\mathrm{d}\boldsymbol{X}\,\mathrm{d}\boldsymbol{B}\,\mathrm{d}\boldsymbol{\gamma}\,\mathrm{d}\varepsilon}
\end{aligned}
$$

and it is parametrized by the (α, β) and M_ε. Then, ε is integrated out and the law for $\mathcal{X}, \mathcal{B}, \mathcal{C}$ given $\mathcal{Y}$ writes

$$f_{\mathcal{X},\mathcal{B},\mathcal{C}|\mathcal{Y}}(\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}|\boldsymbol{Y}) = \int_\varepsilon f_{\mathcal{X},\mathcal{B},\mathcal{C},\mathcal{E}|\mathcal{Y}}(\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}, \varepsilon|\boldsymbol{Y})\,\mathrm{d}\varepsilon.$$

It is also parametrized by the (α, β) and M_ε, so, the limit is set when M_ε tends to 0. The detail of the calculations is given in Appendix V and it is shown that a probability density function is obtained if the mean level of the object is observed, i.e., $\overset{\circ}{h}_{00} \neq 0$.

C. Posterior Law and Posterior Mean

Thus, the total posterior law can be deduced for all the unknown parameters $(\mathcal{X}, \mathcal{B}, \mathcal{C})$ given the observed data $\mathcal{Y}$

$$
\begin{aligned}
&f_{\mathcal{X},\mathcal{B},\mathcal{C}|\mathcal{Y}}(\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}|\boldsymbol{Y})\\
&\propto \gamma_\mathrm{n}^{\alpha_\mathrm{n}-1+N/2} \quad \gamma_\mathrm{d}^{\alpha_\mathrm{d}-1+N/2} \quad \gamma_\mathrm{b}^{\alpha_\mathrm{b}-1+N}\\
&\quad \times \exp-\{Q_0/2 + \gamma_\mathrm{n}/\beta_\mathrm{n} + \gamma_\mathrm{d}/\beta_\mathrm{d} + \gamma_\mathrm{b}/\beta_\mathrm{b}\} \quad (9)
\end{aligned}
$$

where Q_0 involves $\boldsymbol{F}_0 = \boldsymbol{D}$

$$Q_0 = \gamma_\mathrm{n} N_2(\boldsymbol{Y} - \boldsymbol{H} \star \boldsymbol{X}) + \gamma_\mathrm{d} N_2(\boldsymbol{D} \star \boldsymbol{X} - \boldsymbol{B}) + \gamma_\mathrm{b} N_1(\boldsymbol{B}).$$

In practice, the chosen point estimate is the posterior mean (i.e., the minimum mean square error). Its calculation is performed by means of Monte-Carlo Markov chain stochastic sampling algorithm [25], [33]: auxiliary variables, object, and hyperparameters are successively sampled given the other in a Gibbs strategy.

1) Sampling Auxiliary Variables: The sampling of auxiliary variables is delicate but can be directly done. It is based on the inversion of the cumulative density function (cdf) $F_{\mathcal{B}|\mathcal{X}}$. It is sufficient to uniformly sample u in $[0, 1]$ and to compute $b = F_{\mathcal{B}|\mathcal{X}}^{-1}(u)$. The calculations can be found in Appendix VI.

2) Sampling Object: The object is a toroidal Gaussian field and the $\overset{\circ}{X}_{pq}$ are independent with mean $\overset{\circ}{\mu}_{pq}$ and inverse variance $\overset{\circ}{\nu}_{pq}$ (see calculations in Appendix VII)

$$\overset{\circ}{\nu}_{pq} = \gamma_\mathrm{n} |\overset{\circ}{h}_{pq}|^2 + \gamma_\mathrm{d} |\overset{\circ}{d}_{pq}|^2 \qquad (10)$$

$$\overset{\circ}{\mu}_{pq} = [\gamma_\mathrm{n} \overset{\circ}{h}_{pq}^{*}\overset{\circ}{y}_{pq} + \gamma_\mathrm{d} \overset{\circ}{d}_{pq}^{*}\overset{\circ}{b}_{pq}]/\overset{\circ}{\nu}_{pq} \qquad (11)$$

where superscript $*$ stands for the complex conjugate. Thus, the sampling is reduced to the sampling of an inhomogeneous white Gaussian noise followed by an FFT 2-D.

3) Sampling Hyperparameters: Each parameter $\gamma_\mathrm{n}, \gamma_\mathrm{d}$, and γ_b follows a gamma[2] law derived form (9) (see Appendix IV) with respective parameters α and β

$$
\begin{aligned}
\alpha &= \alpha_\mathrm{n} + N/2 \quad \text{and}\\
\beta^{-1} &= \beta_\mathrm{n}^{-1} + N_2(\boldsymbol{Y} - \boldsymbol{H} \star \boldsymbol{X})/2\\
\alpha &= \alpha_\mathrm{d} + N/2 \quad \text{and}\\
\beta^{-1} &= \beta_\mathrm{d}^{-1} + N_2(\boldsymbol{D} \star \boldsymbol{X} - \boldsymbol{B})/2\\
\alpha &= \alpha_\mathrm{b} + N \quad \text{and}\\
\beta^{-1} &= \beta_\mathrm{b}^{-1} + N_1(\boldsymbol{B})/2.
\end{aligned}
$$

The description of the method and the algorithm are now complete and synthesized in Table II. The remainder of this section illustrates the implementation practicability.

---

[2]The sampling of the Gamma variables is achieved using the Matlab function `gamrnd`.

	Data	PM	CPM	CPM (best γ_b)	CPM (best γ_d)	CPM (best γ_n)	MAP-LogErf	MAP-Huber
Dist. L2	11.62%	3.93%	3.94%	3.87%	3.85%	3.81%	5.56%	5.63%
Dist. L1	35.47%	19.47%	19.50%	18.92%	19.40%	19.18%	20.68%	20.84%

TABLE II
DETAILED ALGORITHM (PSEUDO-CODE)

```
Initialize
  • N = 1, Delta = inf
  • HatX = Data
  • GamN, GamD, GamB  (Annex VIII)

Repeat
  • Sample step
     – Auxiliary variables B (Sect. IV-C.1)
     – Object X (Sect. IV-C.2)
     – Hyperparameter GamN, GamD, GamB (Sect. IV-C.3)

  • Update
     – N = N+1
     – Delta = ( HatX - X ) / N
     – HatX = ( (N-1)*HatX + 1*X ) / N
Until Delta<eps
```
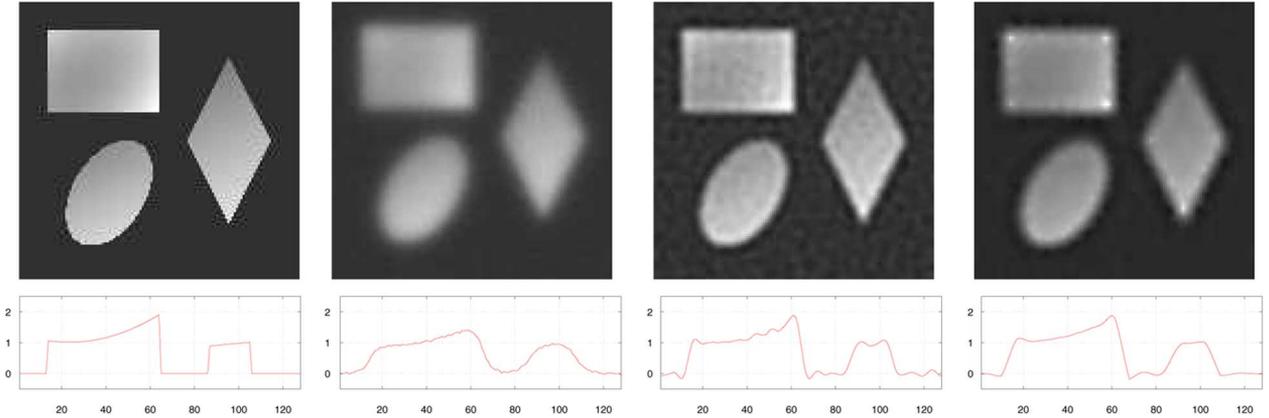

Fig. 5. From left to right: Original image $X^\star$, observed data Y, deconvolved image $\widehat{X}_{\mathrm{PM}}$, and deconvolved image $\widehat{X}_{\mathrm{MAP}}$. At the top: Gray-level images; at the bottom: profile of the 100th row (which encroaches on both the rectangle and the rhombus). In order to evaluate the relative dynamics in each case, all the images are shown in the same grayscale between -0.5 and 2. The four shown profiles are also presented between -0.5 and 2.

D. Computation Feasibility

This part illustrates the previous developments and it only aims at demonstrating the numerical practicability of the method. It is built on a deliberately simple image $X^\star$ appropriate in order to evaluate the capabilities and the limitations of the proposed approach: the image is set up from homogeneous zones separated by sharp edges (see Fig. 5, on the left). It is a 128×128 image composed of a black background and three objects with gray levels gradually changing between 0.7 and 2.1. The difference between neighboring pixels varies between 0 and 2.1 in absolute value. Regarding the Laplacian of the image, $\bar{X} = F \star X$, the set of $\bar{X}_k$ can be split in two sets: 94% of the $\bar{X}_k$ are less than 2.10^{-4} (inside homogeneous zones) and 6% of the $\bar{X}_k$ are greater than 3.10^{-2} (located around edges). No value is between 2.10^{-4} and 3.10^{-2}.

The impulse response of the system is Gaussian shaped with six pixels width at half-maximum, the noise variance is 2.10^{-2} and the resulting observed image Y is shown in Fig. 5 (in the second column). The resolution is clearly degraded and details
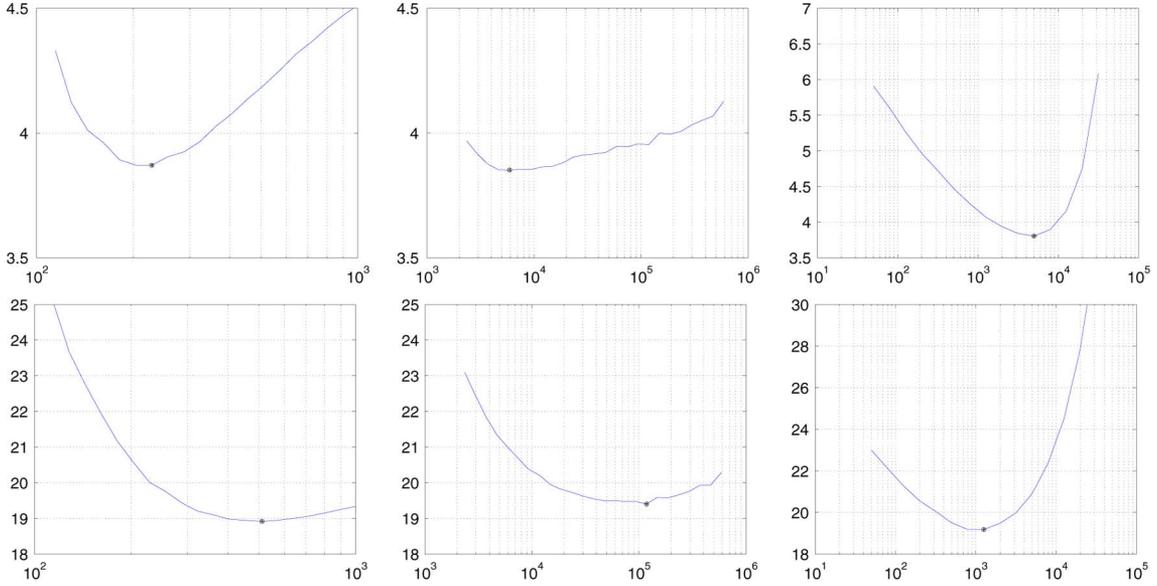
Fig. 6. Distances between the true image $X^\star$ and conditional posterior mean $\widehat{X}_{\mathrm{CPM}}$ as a function of the parameters $\gamma_{\mathrm{b}}, \gamma_{\mathrm{d}}$, and γ_{n}, around the *posterior* mean value $\widehat{\gamma}$. From left to right: Error is shown as a function of $\gamma_{\mathrm{b}}, \gamma_{\mathrm{d}}$, and γ_{n}. Top row gives L2 distance and bottom row gives L1 distances. The black dots give the minimum distances reported in Table I.

of the edges are no longer visible (neither on the gray-level image nor on the shown profile). The dynamic is also strongly affected, notably at about the 64th sample of the shown profile.

The procedure is initialized by the empirical least-squares hyperparameters given in Appendix VIII. The object X is initialized by the observed data (and there is no need to initialize the auxiliary variables). Moreover, practically, the (α_*, β_*) are set to $(0, \infty)$ corresponding to the Jeffrey's prior.

The proposed algorithm[3] generates samples of the *a posteriori* law $f_{\mathcal{X}, \mathcal{B}, \mathcal{C} \mid \mathcal{Y}}(X, B, \gamma \mid Y)$. Practically, the algorithm behaves as expected: the stationary law is attained after a burn-in time (about 200 iterations) and remains in a steady behavior. The empirical mean of the generated images is recursively computed and the algorithm is stopped when its variation becomes smaller than a given value T (in quadratic norm). In the presented example $T = 5.10^{-4}$, the algorithm produced 953 iterations and computation time was 47 s.

The resulting generated hyperparameters $\gamma_{\mathrm{b}}, \gamma_{\mathrm{d}}$, and γ_{n} are shown in Fig. 7. The left part of the figure shows the 953 iterates of the three parameters: after about 200 iterations the three parameters are stabilized and seem to be under the stationary law of the chain. The empirical mean value (approximating the *posterior* mean) of the parameters, respectively, are $\widehat{\gamma_{\mathrm{b}}} = 2.88\,10^2, \widehat{\gamma_{\mathrm{d}}} = 5.91\,10^4$, and $\widehat{\gamma_{\mathrm{n}}} = 1.99\,10^3$. The iterates are also shown on the right hand side of Fig. 7 as histograms: they are clearly very concentrated around the *posterior* mean (with small variance), i.e., the marginal law for the hyperparameters are quasi-Dirac distributions.

Considering the numerical value, in the sense of (8), the equivalent regularization parameter is $\widehat{\lambda} = 2.17\,10^1$ and the equivalent threshold is $\widehat{s} = 6.67\,10^{-3}$. It is noticeable that the threshold value correctly split the $\bar{X}_k$ in two sets (less than

[3]The proposed algorithm has been implemented with the computing environment Matlab on a PC, with a 2-GHz AMD-Athlon CPU, and 512 MB of RAM. Code is $\sim$100 lines long.
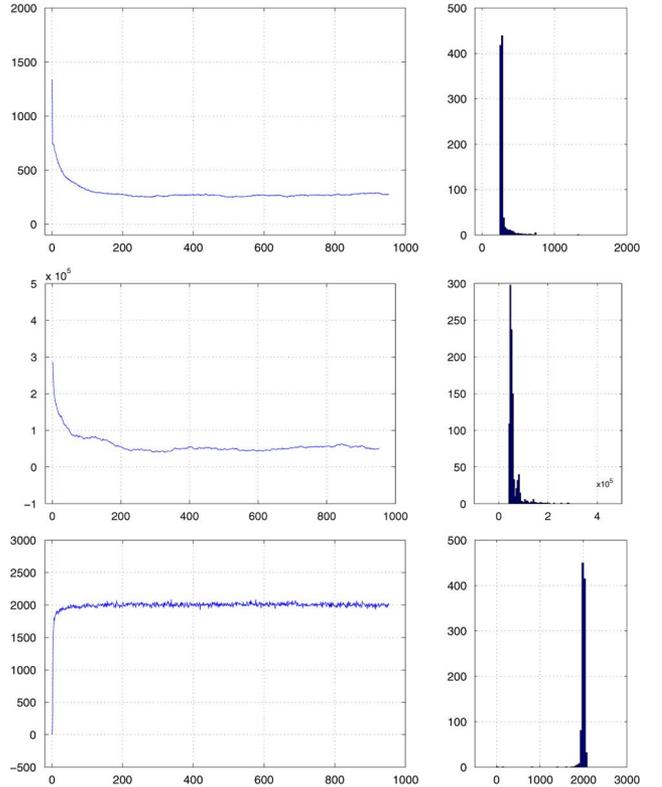

Fig. 7. Monte-Carlo Markov chain for the three hyperparameters generated by the proposed Gibbs sampler. From top to bottom: $\gamma_{\mathrm{b}}, \gamma_{\mathrm{d}}$, and γ_{n}. The left part of the figure shows the samples as a function of iteration index and the right part of the figure shows the samples as histograms.

2.10^{-4}—greater than 3.10^{-2}). The point is that the method automatically adjusts hyperparameters to correctly separate the $\bar{X}_k$. This is a first argument in favor of the proposed strategy in order to tune the threshold of an $\mathrm{L}_2 - \mathrm{L}_1$ Gibbs potential. The resulting image is shown in Fig. 5 (on the third column). The

effect of deconvolution is notable on the image in gray level, as well as on the shown profile. The three objects are correctly positioned, the orders of magnitude are respected and the zero level is correctly reconstructed: it can be seen on the entire image and in particular on the shown profile. The dynamic is also correctly restored: this aspect is notable on the shown profile around the maximum (64th sample). The true dynamic occupies the range 0–1.9 whereas the dynamic of the observed data scarcely exceeds 0–1.4: the proposed method restores the dynamic to 0–1.88 that is to say 99% of the original variation.

A global quantitative comparison has been achieved by evaluating i) the distance between original image $X^\star$ and observed data Y and ii) the distance between original image $X^\star$ and estimated image $\widehat{X}_{\mathrm{PM}}$. The considered distances are normalized L2 and L1 distances. The main results are listed in Table I, first and second columns and show an improvement by a factor 2.95 (11.62% to 3.93%) for L2 distance and a factor 1.82 (35.47% to 19.47%) for L1 distance.

In order to deepen the numerical study, a second estimate has been computed: $\widehat{X}_{\mathrm{CPM}}$ the conditional *posterior* mean (CPM), i.e., the mean of the conditional *posterior* law $f_{\mathcal{X},\mathcal{B}\,|\,\mathcal{Y},\mathcal{C}}(X, B\,|\,Y, \gamma)$. $\widehat{X}_{\mathrm{CPM}}$ is clearly a function of the hyperparameters γ and a twofold evaluation is proposed.

- The first estimate is the one obtained with $\gamma = \widehat{\gamma}$. Practically, the marginal estimate $\widehat{X}_{\mathrm{PM}}$ and the conditional estimate $\widehat{X}_{\mathrm{CPM}}(\widehat{\gamma})$ are quasi-equal; this is due to the fact that the marginal law for the hyperparameters are quasi-Dirac distributions. Quantitatively, regarding L2 distances, the PM produces 3.93% whereas the CPM produces 3.94%; regarding L1 distances, the PM produces 19.47% whereas the CPM produces 19.50%. In both cases, the modification is almost imperceptible.

- The measurement of errors has also been explored for the CPM as a function $\gamma_{\mathrm{b}}, \gamma_{\mathrm{d}}$, and γ_{n}, around the *posterior* mean $\widehat{\gamma}$. Results are given in Fig. 6: in each case, smooth variation of distances is observed when varying parameters and an optimum is visible. It is reported on Table I and shows almost imperceptible modification: optimization of the hyperparameters (based on the true image $X^\star$) allows negligible improvement (smaller than 0.1% for L2 error and smaller than 0.5% for L1 error). So, the main conclusion is that, the unsupervised proposed approach is a relevant tool in order to tune parameters: it works (without the knowledge of the true image), as well as an optimized approach (based on the knowledge of the true image).

Finally, a third estimate has been computed: the Maximum *A posteriori* (MAP). It has been computed for the log-erf and the Huber potentials. Both of them have been computed with equivalent hyperparameters (given above): $(\widehat{\gamma_{\mathrm{b}}}, \widehat{\gamma_{\mathrm{d}}}, \widehat{\gamma_{\mathrm{n}}})$ for the log-erf potential and $(\widehat{\lambda}, \widehat{s})$ for the Huber potential. The two MAP solutions (log-erf and Huber) are visually indiscernible: this is expected from so similar potential. The results are presented in Fig. 5, right column: the estimated map suffers from cross-like artifact, due to the cross-like structure of the neighborhood system. Quantitatively speaking, the measurements of errors are given on Table I: log-erf and Huber produce almost similar errors. Moreover, the errors are greater than the one produced by the PM and the CPM.

The restoration is nevertheless imperfect and of limited resolution: the sharp edges remain slightly smoothed and limited in amplitude. The ringing effect also affects the quality of the deconvolved image. This diagnostic is long awaited in the framework of convex deconvolution. Anyway, the important point is not so much the property of the deconvolved image itself (intrinsic of any convex deconvolution) but the (new) practical capability to automatically tune the hyperparameters. Moreover, the potential improvement is certainly wide considering more heavy-tailed law for the auxiliary variables, as explained in the next section.

V. CONCLUSION

This paper presents a twofold novelty in the field of statistical image reconstruction and restoration.

1) The partition function is explicitly given for a specific non-Gaussian Markov field, with an $L_2 - L_1$ Gibbs potential. It is built as a compound field involving: an auxiliary variable following a separable Laplace distribution and a pixel variable following a Gaussian distribution given the auxiliary variable.

2) An unsupervised deconvolution method is deduced, based on the exact likelihood taking advantage of the knowledge of the partition function. The method is fully Bayesian, and the point estimate is the posterior mean computed thanks to a Monte-Carlo Markov chain technique.

The paper focuses on the deconvolution problem, but it is also possible to deal with simpler questions than deconvolution: parameter estimation from direct observation of the field, edge enhancement or denoising.

Moreover, the paper relies on Gaussian noise, but the case of non-Gaussian noise is also envisaged, in particular the use of robust norms to reject abnormal data (outliers). To this end, a separable version of the $L_2 - L_1$ proposed field could be suitable as a law for noise measurement. The proposed method can be directly applied in the case of large support operator, e.g., reconstruction problems such as Fourier synthesis [34].

The proposed methodology also remains valid for other linear model and the required modification concerns the sampling of the object. It remains Gaussian but its sampling is no longer possible in a single step for the entire image by FFT 2-D. The Gibbs sampling techniques constitute an adapted tool but the calculation time would be (maybe dramatically) extended. For nonlinear problems, the law for the object is no longer Gaussian and a case by case study is required.

Concerning the *a priori* field, other laws for auxiliary variables are certainly desirable. The possible improvements are numerous considering more heavy-tailed law in order to overcome the limitation of the convex deconvolution. The methodology still remains valid but the difficulty then concerns the sampling of the auxiliary variables. The direct sampling by inversion of the cumulative density function may not be possible, however, the rejection or the Hastings–Metropolis algorithms could be used to overcome this difficulty.

In the case of myopic deconvolution, it is also conceivable to estimate (part of) the parameters of the observation system. Here again, a case by case study is necessary, but the delicate question of the system parameter sampling can probably be tackled by means of rejection or Hastings–Metropolis algorithms.

APPENDIX I
ERF, ERFC, ERFCX

The erf function is defined for $x \in \mathbb{R}$ by

$$\mathrm{erf}[x] = \frac{2}{\sqrt{\pi}} \int_0^x \exp[-t^2]\,\mathrm{d}t \qquad (12)$$

and ierf denotes the reciprocal function. Elsewhere, $\mathrm{erfc}[x] = 1 - \mathrm{erf}[x]$ and $\mathrm{erfcx}[x] = \exp[x^2]\mathrm{erfc}[x]$. Concerning the latter, there are the following expansions:

$$\mathrm{erfcx}[x] \underset{+\infty}{\sim} [1 - x^{-2}]/(x\sqrt{\pi}) \qquad (13)$$

$$\mathrm{erfcx}[x] \underset{0}{\sim} 1 - 2x/\sqrt{\pi} \qquad (14)$$

and the derivative $\mathrm{erfcx}[x]' = 2\,x\,\mathrm{erfcx}[x] - 2/\sqrt{\pi}$.

APPENDIX II
GAUSS AND LAPLACE CONVOLUTION

Considering the calculations, a large part of the proposed developments is based on the convolution of a Gaussian function and a Laplacian function.

A. Preliminary Calculi

For $x_0 \geq 0$ and $x \in \mathbb{R}$, write

$$J(x_0, x, d, b) = \int_0^{x_0} \exp[-\{d(y - x)^2 + by\}/2]\mathrm{d}y$$

simply written as $J(x_0, x)$ when there is no ambiguity. On rewriting the argument of the exponential, we have

$$d(y - x)^2 + by = d[(y - \tilde{x})^2 + (x^2 - \tilde{x}^2)]$$

with $\tilde{x} = x - b/2d$. The change of variable $t = (y - \tilde{x})\sqrt{d/2}$, yields

$$J(x_0, x) = \sqrt{\pi/2d}\exp[b^2/8d]\exp[-bx/2]$$
$$\times \{\mathrm{erf}[\tilde{x}\sqrt{d/2}] - \mathrm{erf}[(\tilde{x} - x_0)\sqrt{d/2}]\}$$

where the function erf is defined by (12). In particular, one has $J(0, x) = 0$ and

$$J(+\infty, x) = \sqrt{\pi/2d}\exp[b^2/8d]\exp[-bx/2]\mathrm{erfc}[-\tilde{x}\sqrt{d/2}].$$

B. Convolution

For $x_0, x \in \mathbb{R}$, write

$$I(x_0, x, d, b) = \int_{-\infty}^{x_0} \exp[-\{d(y - x)^2 + b|y|\}/2]\mathrm{d}y$$

written simply as $I(x_0, x)$ when there is no ambiguity. By the change of variables $y' = -y, y' = by$, and $y' = y\sqrt{d}$, it is shown that

$$I(+\infty, x, d, b) = I(+\infty, -x, d, b)$$
$$I(x_0, x, d, b) = I(bx_0, bx, d/b^2, 1)/b$$
$$I(x_0, x, d, b) = I(x_0\sqrt{d}, x\sqrt{d}, 1, b/\sqrt{d})/\sqrt{d}.$$

It can thus be deduced that

$$I(x_0, x, d, b) = J(+\infty, -x, d, b) - J(-x_0, -x, d, b)$$
$$I(x_0, x, d, b) = J(+\infty, -x, d, b) + J(x_0, x, d, b)$$

respectively for $x_0 < 0$ and $x_0 \geq 0$. These relationships are useful for the study of the potential function (next Appendix) and for the inversion of the cdf of $\mathcal{B}\,|\,\mathcal{X}$ (Appendix VI).

APPENDIX III
LOG-ERF POTENTIAL FUNCTION

According to the results of the previous Appendix the potential function of the marginal field $\mathcal{X}$, (6), Section III-C is written

$$\varphi(x) = -2\log I(+\infty, x, \gamma_\mathrm{d}, \gamma_\mathrm{b}).$$

By putting $\chi(x) = \exp[\gamma_\mathrm{b} x/2]\mathrm{erfc}[(\rho + x)\sqrt{\gamma_\mathrm{d}/2}]$, $\rho = \gamma_\mathrm{b}/2\gamma_\mathrm{d}$, the potential function can be written

$$\varphi(x) = -2\log[\chi(x) + \chi(-x)]$$

up to additive constants. The derivation shows that

$$\chi'(x) = \frac{\gamma_\mathrm{b}}{2}\chi(x) - \sqrt{\frac{2\gamma_\mathrm{d}}{\pi}}\exp[-\gamma_\mathrm{d}(\rho^2 + x^2)/2]$$

$$\chi''(x) = \frac{\gamma_\mathrm{b}}{2}\chi'(x) + \sqrt{\frac{2\gamma_\mathrm{d}^3}{\pi}}x\exp[-\gamma_\mathrm{d}(\rho^2 + x^2)/2]$$

and it can easily be deduced that

$$\varphi'(x) = -2\frac{\chi'(x) - \chi'(-x)}{\chi(x) + \chi(-x)} = -\gamma_\mathrm{b}\frac{\chi(x) - \chi(-x)}{\chi(x) + \chi(-x)}$$

and, in particular

$$\varphi'(0) = 0 \quad \text{and} \quad \varphi'(+\infty) = \gamma_\mathrm{b}.$$

Moreover, concerning the second derivative at origin

$$\varphi''(0) = -2\frac{\chi''(0)}{\chi(0)} = -\gamma_\mathrm{b}\frac{\chi'(0)}{\chi(0)}$$
$$= \frac{\gamma_\mathrm{b}^2}{2}[(\eta\sqrt{\pi}\mathrm{erfcx}[\eta])^{-1} - 1]$$

with $\eta = \gamma_\mathrm{b}/\sqrt{8\gamma_\mathrm{d}}$.

APPENDIX IV
GAMMA PROBABILITY DENSITY FUNCTION

The gamma probability density function is parametrized by $\alpha > 0$ and $\beta > 0$ in the form

$$f_\gamma(x; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma[\alpha]}x^{\alpha - 1}\exp[-x/\beta]\mathbb{1}_{\mathbb{R}_+}(x) \qquad (15)$$

where $\mathbb{1}_{\mathbb{R}_+}$ is the indicator function of $\mathbb{R}_+$. The expected value is $\alpha\beta$, the variance is $\alpha\beta^2$ and it is maximal for $x = \beta(\alpha - 1)$ in the case $\alpha > 1$.

APPENDIX V
INTEGRATION OF HYPERPARAMETER

A. Preliminary Result

Given a function $f : \mathbb{R} \to \mathbb{R}_+, \mathcal{C}^\infty$ and assume that $g(x) = x f(x)$ can be integrated. By integrating from 0 to M the Taylor expansion of $g(x)$ at origin, one shows that

$$\frac{1}{M^2} \int_0^M x f(x) \, \mathrm{d}x \xrightarrow[M=0]{} \frac{1}{2} f(0). \qquad (16)$$

Then, give a function $\psi : \mathbb{R}^Q \times \mathbb{R} \to \mathbb{R}_+$ such that $\varepsilon \psi(\boldsymbol{u}, \varepsilon)$ can be integrated over $\mathbb{R}^{Q+1}$. By using (16), it can be seen that

$$\frac{\int_0^M \varepsilon \psi(\boldsymbol{u}, \varepsilon) \, \mathrm{d}\varepsilon}{\int_{\boldsymbol{v}} \int_0^M \eta \psi(\boldsymbol{v}, \eta) \, \mathrm{d}\eta \, \mathrm{d}\boldsymbol{v}} \xrightarrow[M=0]{} \frac{\psi(\boldsymbol{u}, 0)}{\int_{\boldsymbol{v}} \psi(\boldsymbol{v}, 0) \, \mathrm{d}\boldsymbol{v}} \qquad (17)$$

provided that $\psi(\boldsymbol{v}, 0)$ can still be integrated over $\mathbb{R}^Q$.

B. Posterior Law

The *a posteriori* law (Section IV-B) for $\mathcal{X}, \mathcal{B}, \mathcal{C}$ and $\mathcal{E}$ given $\mathcal{Y}$ (parametrized by the coefficient M_ε) is written, after simplification by $\delta' M_\varepsilon$

$$\frac{\int_{\varepsilon=0}^{M_\varepsilon} \psi(\boldsymbol{u}, \varepsilon) \mathrm{d}\varepsilon}{\int_{\boldsymbol{X}\boldsymbol{B}\boldsymbol{\gamma}} \int_{\varepsilon=0}^{M_\varepsilon} \psi(\boldsymbol{v}, \varepsilon) \mathrm{d}\varepsilon \mathrm{d}\boldsymbol{\gamma} \mathrm{d}\boldsymbol{X} \mathrm{d}\boldsymbol{B}}$$

where $\boldsymbol{u}$ represents all the parameters $\boldsymbol{X}, \boldsymbol{B}, \boldsymbol{\gamma}$, and

$$\psi(\boldsymbol{u}, \varepsilon) = \delta' \gamma_{\mathrm{n}}^{\alpha_{\mathrm{n}}-1+N/2} \quad \gamma_{\mathrm{d}}^{\alpha_{\mathrm{d}}-1+N/2} \quad \gamma_{\mathrm{b}}^{\alpha_{\mathrm{b}}-1+N} \varepsilon$$
$$\times \exp -\{Q_\varepsilon/2 + \gamma_{\mathrm{n}}/\beta_{\mathrm{n}} + \gamma_{\mathrm{d}}/\beta_{\mathrm{d}} + \gamma_{\mathrm{b}}/\beta_{\mathrm{b}}\}.$$

To apply the relationship (17), it is sufficient to ensure that $\psi(\boldsymbol{v}, 0)$ can be integrated. Since the norms in $\mathbb{R}^N$ are equivalent, $k \in \mathbb{R}_+$ can be found such as $N_1(\boldsymbol{B}) < k N_2(\boldsymbol{B})$ for all $\boldsymbol{B}$. Thus, the integrand can be majored by a Gaussian integrand and convergence ensured if and only if $\overset{\circ}{h}_{00} \neq 0$.

In the limit, when $M_\varepsilon \to 0$, we have the result (9).

APPENDIX VI
INVERSION OF $\mathcal{B} \,|\, \mathcal{X}$ CDF

The sampling of auxiliary variables (Section IV-C) given the object is based on the inversion of cdf of $\mathcal{B} \,|\, \mathcal{X}$. For $u \in [0, 1]$

$$u = F_{\mathcal{B}|\mathcal{X}}(b) = \int_{-\infty}^b f_{\mathcal{B}|\mathcal{X}} = \frac{I(b, \bar{b}, \gamma_{\mathrm{d}}, \gamma_{\mathrm{b}})}{I(+\infty, \bar{b}, \gamma_{\mathrm{d}}, \gamma_{\mathrm{b}})} \qquad (18)$$

is to be resolved. In order to solve this equation, write $\rho = \gamma_{\mathrm{b}}/2\gamma_{\mathrm{d}}$

$$\theta_- = \exp[+\gamma_{\mathrm{b}} \bar{b}/2] \mathrm{erfc}[(\rho + \bar{b}) \sqrt{\gamma_{\mathrm{d}}.2}]$$
$$\theta_+ = \exp[-\gamma_{\mathrm{b}} \bar{b}/2] \mathrm{erfc}[(\rho - \bar{b}) \sqrt{\gamma_{\mathrm{d}}/2}]$$

and $\theta = \theta_- + \theta_+$. Moreover, $s = F_{\mathcal{B}|\mathcal{X}}(0) = \theta_-/\theta$ and $\bar{u} = \theta u - \theta_-$. Equation (18) is resolved differently depending on whether $u \leq s$ (i.e., $b \leq 0$) or $u \geq s$ (i.e., $b \geq 0$) and yields

$$b = \bar{b} + \rho - \mathrm{ierf}\{T_-\} \sqrt{2/\gamma_{\mathrm{d}}}, \quad \text{if} \quad u \leq s$$
$$b = \bar{b} - \rho - \mathrm{ierf}\{T_+\} \sqrt{2/\gamma_{\mathrm{d}}}, \quad \text{if} \quad u \geq s$$

where $\mathrm{ierf}\{\cdot\}$ is defined in Appendix I and

$$T_- = \mathrm{erf}[(\bar{b} + \rho) \sqrt{\gamma_{\mathrm{d}}/2}] - \bar{u} \exp[-\gamma_{\mathrm{b}} \bar{b}/2]$$
$$T_+ = \mathrm{erf}[(\bar{b} - \rho) \sqrt{\gamma_{\mathrm{d}}/2}] - \bar{u} \exp[+\gamma_{\mathrm{b}} \bar{b}/2].$$

Thus, it is possible to sample $\mathcal{B} \,|\, \mathcal{X}$ simply from u uniformly distributed over $[0, 1]$.

APPENDIX VII
CONDITIONAL POSTERIOR LAW FOR $\mathcal{X}$

The posterior law $(\mathcal{X}, \mathcal{B}, \mathcal{C} \,|\, \mathcal{Y})$ given by (9) in Section IV-C involves

$$Q_0 = \gamma_{\mathrm{n}} N_2(\boldsymbol{Y} - \boldsymbol{H} \star \boldsymbol{X}) + \gamma_{\mathrm{d}} N_2(\boldsymbol{D} \star \boldsymbol{X} - \boldsymbol{B}) + \gamma_{\mathrm{b}} N_1(\boldsymbol{B})$$

and the conditional posterior law $(\mathcal{X} \,|\, \mathcal{B}, \mathcal{C}, \mathcal{Y})$ required to sample object in Section IV-C2 involves

$$Q_{0|} = \gamma_{\mathrm{n}} N_2(\boldsymbol{Y} - \boldsymbol{H} \star \boldsymbol{X}) + \gamma_{\mathrm{d}} N_2(\boldsymbol{D} \star \boldsymbol{X} - \boldsymbol{B}).$$

In the Fourier domain

$$Q_{0|} = \gamma_{\mathrm{n}} N_2(\overset{\circ}{\boldsymbol{Y}} - \overset{\circ}{\boldsymbol{H}} \otimes \overset{\circ}{\boldsymbol{X}}) + \gamma_{\mathrm{d}} N_2(\overset{\circ}{\boldsymbol{D}} \otimes \overset{\circ}{\boldsymbol{X}} - \overset{\circ}{\boldsymbol{B}})$$
$$= \sum_{pq} \gamma_{\mathrm{n}} |\overset{\circ}{y}_{pq} - \overset{\circ}{h}_{pq} \overset{\circ}{x}_{pq}|^2 + \gamma_{\mathrm{d}} |\overset{\circ}{d}_{pq} \overset{\circ}{x}_{pq} - \overset{\circ}{b}_{pq}|^2$$

that is to say a separable summation. Moreover, it can be rewritten and identified to a sum of quadratic terms

$$Q_{0|} = \sum_{pq} \overset{\circ}{\nu}_{pq} |\overset{\circ}{x}_{pq} - \overset{\circ}{\mu}_{pq}|^2$$

with $\overset{\circ}{\nu}_{pq}$ and $\overset{\circ}{\mu}_{pq}$ given in (10) and (11).

APPENDIX VIII
EMPIRICAL LEAST SQUARES HYPERPARAMETERS

The initialization of the algorithm is based on second order statistics of the analyzed data, in the Fourier domain. Considering the structure of the *a priori* field and the noise, for all (p, q), such as $\overset{\circ}{f}_{pq} \neq 0$ one has

$$\overset{\circ}{X}_{pq} = \sqrt{r_{\mathrm{d}}} \, \overset{\circ}{G}_{pq} / \overset{\circ}{f}_{pq} \quad \text{and} \quad \overset{\circ}{N}_{pq} = \sqrt{r_{\mathrm{n}}} \, \overset{\circ}{G}'_{pq}$$

where $\overset{\circ}{G}_{pq}$ and $\overset{\circ}{G}'_{pq}$ are two independent zero-mean white Gaussian noise with unitary variance. Moreover, considering the observation (1), one also has $\overset{\circ}{Y}_{pq} = \overset{\circ}{h}_{pq} \overset{\circ}{X}_{pq} + \overset{\circ}{N}_{pq}$. With $\overset{\circ}{Z}_{pq} = |\overset{\circ}{Y}_{pq}|^2$ and $\overset{\circ}{r}_{pq} = |\overset{\circ}{h}_{pq}| / |\overset{\circ}{f}_{pq}|$, we have

$$\mathrm{E}[\overset{\circ}{Z}_{pq}] = r_{\mathrm{d}} \, \overset{\circ}{r}_{pq} + r_{\mathrm{n}}.$$

Thus, the parameters r_{d} and r_{n} can be selected at the minimum of the least squares criterion

$$J(r_{\mathrm{d}}, r_{\mathrm{n}}) = \sum_{pq} (\overset{\circ}{z}_{pq} - r_{\mathrm{d}} \, \overset{\circ}{r}_{pq} + r_{\mathrm{n}})^2.$$

It is found that

$$r_{\mathrm{d}} = \frac{\gamma - \alpha\delta}{\beta - \alpha^2} \quad \text{and} \quad r_{\mathrm{n}} = \frac{\beta\delta - \alpha\gamma}{\beta - \alpha^2}$$

with $N\alpha = \sum \overset{\circ}{r}_{pq}$, $N\beta = \sum \overset{\circ}{r}_{pq}^{2}$, $N\gamma = \sum \overset{\circ}{r}_{pq}\overset{\circ}{z}_{pq}$, and $N\delta = \sum \overset{\circ}{z}_{pq}$. These values for r_{d} and r_{n} are used to initialize the proposed algorithm (Section IV-C): $\gamma_{\mathrm{d}} = 1/r_{\mathrm{d}}$ and $\gamma_{\mathrm{n}} = 1/r_{\mathrm{n}}$. The third parameter γ_{b} is initialized at the critical value: $\gamma_{\mathrm{b}} = \sqrt{2\pi\gamma_{\mathrm{d}}}$.

REFERENCES

[1] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*. Washington, DC: Winston, 1977.

[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[4] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: The MIT Press, 1987.

[5] J. Idier, "Regularization tools and models for image and signal reconstruction," in *Proc. 3nd Int. Conf. Inverse Problems in Engng.*, Port Ludlow, WA, Jun. 1999, pp. 23–29.

[6] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[7] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.

[8] C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.

[9] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm," *IEEE Trans. Med. Imag.*, vol. 9, no. 1, pp. 84–93, Mar. 1990.

[10] L. Rudin, S. Osher, and C. Fatemi, "Nonlinear total variation based noise removal algorithm," *Phys. D*, vol. 60, pp. 259–268, 1992.

[11] H. R. Künsch, "Robust priors for smoothing and image restoration," *Ann. Inst. Statist. Math.*, vol. 46, no. 1, pp. 1–19, 1994.

[12] J. A. O'Sullivan, "Roughness penalties on finite domains," *IEEE Trans. Image Process.*, vol. 4, no. 9, pp. 1258–1268, Sep. 1995.

[13] J. A. Fessler, H. Erdoğan, and W. B. Wu, "Exact distribution of edge-preserving MAP estimators for linear signal models with gaussian measurement noise," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1049–1055, Jun. 2000.

[14] M. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[15] J.-L. Starck, F. Murtagh, P. Querre, and F. Bonnarel, "Entropy and astronomical data analysis: Perspectives from multiresolution analysis," *Astron. Astrophys.*, vol. 368, pp. 730–746, 2001.

[16] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.

[17] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.

[18] Z. Zhou, R. M. Leahy, and J. Qi, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," *IEEE Trans. Image Process.*, vol. 6, no. 6, pp. 844–861, Jun. 1997.

[19] M. T. de Figueiredo and J. M. N. Leitao, "Unsupervised image restoration and edge location using compound Gauss–Markov fields and the MDL principle," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1089–1102, Aug. 1997.

[20] S. S. Saquib, C. A. Bouman, and K. D. Sauer, "ML parameter estimation for Markov random fields with applications to Bayesian tomography," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1029–1044, Jul. 1998.

[21] 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.*, vol. 8, no. 7, pp. 954–963, Jul. 1999.

[22] R. Molina, A. K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 231–246, Feb. 1999.

[23] A. D. Lanterman, U. Grenander, and M. I. Miller, "Bayesian segmentation via asymptotic partition functions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 4, pp. 337–347, Apr. 2000.

[24] V. Pascazio and G. Ferraiuolo, "Statistical regularization in linearized microwave imaging through MRF-based MAP estimation: Hyperparameter estimation and image computation," *IEEE Trans. Image Process.*, vol. 12, no. 5, pp. 572–582, May 2003.

[25] G. Winkler, *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Berlin, Germany: Springer-Verlag, 2003.

[26] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Tokyo, Japan: Springer-Verlag, 2001.

[27] J. Idier, Ed., *Approche bayésienne pour les problèmes inverses*. Paris, France: Traité IC2, Série traitement du signal et de l'image, Hermès, 2001.

[28] F. Champagnat and J. Idier, "A connection between half-quadratic criteria and EM algorithm," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 709–712, Sep. 2004.

[29] A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia, "Hyperparameter estimation for satellite image restoration by a MCMC maximum likelihood method," *Pattern Recognit.*, vol. 35, no. 2, pp. 341–352, 2002.

[30] L. Onsager, "A two-dimensional model with an order-disorder transition," *Phys. Rev.*, vol. 65, no. 3/4, pp. 117–149, Feb. 1944.

[31] A. Prèkopa, "On logarithmic concave measures and functions," *Acta Sci. Math.*, vol. 34, pp. 335–343, 1973.

[32] I. A. Ibragimov, "On the composition of unimodal distributions," *Theory Probab. Appl.*, vol. 1, 1956.

[33] C. P. Robert and G. Casella, *Monte-Carlo Statistical Methods*, ser. Springer Texts in Statistics. New York: Springer, 2000.

[34] J.-F. Giovannelli and A. Coulais, "Positive deconvolution for superimposed extended source and point sources," *Astron. Astrophys.*, vol. 439, pp. 401–412, 2005.

Jean-François Giovannelli was born in Béziers, France, in 1966. He graduated from the École Nationale Supérieure de l'Électronique et de ses Applications in 1990 and the Doctorat degree in physics from the Université Paris-Sud, Orsay, France, in 1995.

He is presently an Assistant Professor with the Département de Physique, Université Paris-Sud, and a Researcher with the Laboratoire des Signaux et Systèmes (CNRS-Supélec-UPS), Gif-sur-Yvette, France. He is interested in regularization and Bayesian methods for inverse problems in signal and image processing. His application fields essentially concern astronomical, medical, and geophysical imaging.