Bayesian Approach with Maximum Entropy Priors to Imaging Inverse Problems. Part I : Foundations *

Ali Mohammad–Djafari
Laboratoire des Signaux et Systèmes (CNRS–ESE–UPS)
École Supérieure d'Électricité
Plateau de Moulon, 91192 Gif–sur–Yvette Cedex, France.

Abstract

This paper is the first of two papers on Bayesian approach with maximum entropy priors to imaging inverse problems. In this paper we propose a method based on the Maximum *a posteriori* (MAP) Bayesian approach with Maximum Entropy (ME) priors to solve the linear system of equations which is obtained after the discretisation of the integral equations which arise in various tomographic image restoration and reconstruction problems. The main difficulty with the Bayesian approach in real applications is the assignment of the prior laws. We discuss in detail about this main problem which is how to assign a prior probability law to an image to reflect our *a priori* knowledge about it. We give arguments how to choose the form of the prior law and, we propose a family of prior laws depending on two parameters obtained by using the ME principle and some arguments of scale invariance property. We present then some well–posed Bayesian problem statements and give the procedure to solve them. We will see that in a practical problem we have to estimate not only an image but also infer about the parameters of the prior law which are called the hyperparameters of the problem. We give brief descriptions of the different methods for the estimation of the hyperparameters from the data and propose a method to estimate simultaneously these hyperparameters and the pixel values of the image based on the generalized likelihood criterion. Some simulation results are presented to show the performances of the proposed method. In the second part we present some of the imaging applications in which we used the proposed method. These applications are image restoration, microwave tomographic image reconstruction and Eddy current non destructive control of conducting media.

*submitted to *IEEE Trans. on Image Processing*, August 1993. rejected on 1995.

1 Introduction

We address a class of image reconstruction and restoration problems which is to solve the integral equations of the form:

$$g_{ij} = \int_D f(\boldsymbol{r}')\, h_{ij}(\boldsymbol{r}')\, \mathrm{d}\boldsymbol{r}' + b_{ij}, \quad i,j = 1,\cdots,M, \tag{1}$$

where $\boldsymbol{r}' \in \mathbb{R}^2$, $f(\boldsymbol{r}')$ is the object (image reconstruction problems) or the original image (image restoration problems), g_{ij} are the measured data (the projections in image reconstruction or the degraded image in image restoration problems), b_{ij} are the measurement noise and $h_{ij}(\boldsymbol{r}')$ are known functions which depend only on the measurement system. In the second part of this paper we show the generality of this relation in many signal and image restoration and reconstruction applications.

In all these applications we have to solve the following ill-posed problem: how to estimate the function $f(x,y)$ from some finite set of measured data which may also be noisy, because there is no experimental measurement device, even the most elaborate, which could be entirely free from uncertainty, the simplest example being the finite precision of the measurements.

The numerical solution of these equations needs a discretisation procedure which can be done by a quadrature method. The linear system of equations resulting from the discretisation of an ill-posed problem is, in general, very ill-conditioned if not singular [1]. So the problem is to find a unique and stable solution for this linear system. The general methods which enable us to find a unique and stable solution to an ill-posed problem by introducing an *a priori* information on the solution are called regularisation [2, 3, 1]. The *a priori* information can be either in a deterministic form (positivity) or in a probabilistic form (mean values, variance values or in general some constraints on the probability density functions).

In many of these image reconstruction and restoration problems $f(x,y)$ represents the spatial distribution of a positive quantity (for example the conductivity or the density of the matter inside the examined object) [4, 5]. So, when discretised these problems can be described by the following discrete problem:

> Estimate a positive vector $\boldsymbol{x} \in \mathbb{R}^n_+$ (representing the pixel intensities in an image) given a vector of measurements $\boldsymbol{y} \in \mathbb{R}^m$ (representing, for example, either a degraded image pixel values in restoration problems or the projections values in reconstruction problems) and a linear transformation $\boldsymbol{A}$ relating them by:
>
> $$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \tag{2}$$
>
> where $\boldsymbol{b}$ represents the discretisation errors and the measurement noise which is supposed to be zero-mean and additive.

In this paper we propose to use the Bayesian approach to find a regularised solution to this problem. This idea is not new [6, 7, 8, 9, 10, 11, 12, 13, 14, 1, 15, 16, 17, 18], however we have to note that the Bayesian theory only gives us a framework for the formulation of the inverse problem, not a solution of it. The main difficulty is, in general, before the application of the Bayes' formula, *i.e.*; how to formulate appropriately the problem and how to assign the direct probabilities. The use of ME principle to assign the direct probabilities also is not new [19, 20, 21, 22, 23, 24, 25, 26], but using ME principle to assign the direct probabilities and then using the Bayesian estimation theory is less usual [10, 27, 15, 28, 29, 30]. In fact, ME has also been used directly in image restoration and reconstruction [31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51].

Keeping these facts in mind, we propose the following organization to this paper:

In section 2. we give a brief description and the basic ideas of the Bayesian approach. In section 3. we give a short description of the ME principle as a tool for assigning the probability laws from incomplete knowledge about a physical quantity. In section 4. we present the whole procedure of solving an inverse problem for a very known case which is the Gaussian case. This is very helpful to illustrate the approach which is used for the proposed method. In section 5. we present the outline of the proposed method. In section 6. we discuss about the assignment of a prior law to an image. We then show how different prior probability laws are deduced from different *a priori* knowledge about the image when the principle of ME is used. We consider then some particular interesting cases in which the prior laws are the members of a two–parameters family of ME laws. In section 7. we present different problems to solve, starting by the simplest to the more and more difficult problems which arise in real applications. We will see that in a practical problem we have also to infer about the parameters of the prior law which are called the hyperparameters of the problem and are not known in practical problems. We give then brief descriptions of the different methods of the estimation of the hyperparameters from the data. In section 8. we give a detailed description of the proposed method, and finally, in section 9. we give some simulation results to illustrate the performances of the method.

In the second part of the paper we give some application examples in which the proposed method has been used effectively.

2 Bayesian approach framework

A general Bayesian approach involves the following steps:

- Assign a probability law $p(\boldsymbol{x})$ to the unknown parameters to translate our incomplete *a priori* information (prior beliefs or prior knowledge) about these parameters;
- Assign a probability law to the measured data $p(\boldsymbol{y}|\boldsymbol{x})$ to translate the lack of total precision of the model and the inevitable existence of the measurement noise;
- Use the Bayes rule to combine these by calculating the posterior law $p(\boldsymbol{x}|\boldsymbol{y})$ of the unknown parameters;
- Infer a solution $\widehat{\boldsymbol{x}}$ for the problem using a decision rule.
- Caracterize the proposed solution by calculating for example posterior variances, covariances, error bars, etc.

Note that in this approach :

- We are able to solve inverse problems which are described by a finite number of unknown parameters.
- Assigning a probability law to un unknown parameter X does not necessarily mean that this parameter is a random quantity for whome the probability is a limit to its realization frequency. The probability is just a measure of our confidence to that parameter value. For example, if we assume to know that the nominal value of quantity is m and the variations around this nominal value is σ, then we can assign a Gaussian probability law to it to represente this knowledge. If we had more knowledge the choice will different. In any case, this does not forcibly mean that X is a random quantity. However, in a special application we can judge the goodness of hypothesis and the resulting choice of the probability law when dealing with real data.
- If it is more common to assign $p(\boldsymbol{y}|\boldsymbol{x})$ to the measured data to take account of the noise, the assignement of $p(\boldsymbol{x})$ has been less usual and more controversial [25]. In our point of view the both problems are logically equivalent. The main problem is that our prior knowledge is incomplet and does not lead to a unique choice for them. One can reduce the choices for the direct probability laws $p(\boldsymbol{y}|\boldsymbol{x})$ and $p(\boldsymbol{x})$ by asking to satisfy (directly or undirectly) some properties such as scale invariance.
- In some cases these probability laws can be assigned from the physics of the image generation process and the measurement process. This is for example the case of the Positron Emission Tomography (PET) where from the physiques of the measuring system it is known that a Poisson law is more appropriate than a Gaussian law for $p(\boldsymbol{y}|\boldsymbol{x})$.

- When our prior knowledge is or can be assumed to be in the form of some constraints on the probability distribution function (pdf) of the unknown parameters then we can then use the ME principle, to choose one prior law between all the possible probability laws satisfying these constraints [21, 22, 24].

For now, suppose now that we are able to assign the probability densities $p(\boldsymbol{y}|\boldsymbol{x})$ and $p(\boldsymbol{x})$. We can then use the Bayes' rule to find $p(\boldsymbol{x}|\boldsymbol{y})$:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x}) \tag{3}$$

and use the maximum a posteriori (MAP) estimation rule to give a solution to the problem, i.e.

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{y}|\boldsymbol{x})\, p(\boldsymbol{x})\} \tag{4}$$

Other estimators are possible. In fact, all we want to know is resumed in the posterior law. In general, one can construct a Bayesian estimator by defining a cost (or utility) function $C(\widehat{\boldsymbol{x}}, \boldsymbol{x})$ and by minimizing its mean value

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{z}} \left\{\mathrm{E}_{X|Y}\left\{C(\boldsymbol{z}, \boldsymbol{x})\right\}\right\} = \arg\min_{\boldsymbol{z}} \left\{\int C(\boldsymbol{z}, \boldsymbol{x}) p(\boldsymbol{x}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{x}\right\}. \tag{5}$$

The three classical estimators are :

- Maximum *a posteriori* (MAP) estimator :

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\}, \tag{6}$$

  is obtained when defining $C(\widehat{\boldsymbol{x}}, \boldsymbol{x}) = 1 - \delta(\widehat{\boldsymbol{x}} - \boldsymbol{x})$.

- Posterior mean (PM) estimator :

$$\widehat{\boldsymbol{x}} = \mathrm{E}_{X|Y}\{\boldsymbol{x}\} = \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{y})\, \mathrm{d}\boldsymbol{x}, \tag{7}$$

  is obtained when defining $C(\widehat{\boldsymbol{x}}, \boldsymbol{x}) = (\widehat{\boldsymbol{x}} - \boldsymbol{x})^t(\widehat{\boldsymbol{x}} - \boldsymbol{x})$, and

- Maximum of the Marginal Posterior (MMP) estimator :

$$\widehat{\boldsymbol{x}} = \arg\max_{x_j} \{p(x_j|\boldsymbol{y})\}, \tag{8}$$

  where $p(x_j|\boldsymbol{y})$ is the marginal posterior law which is related to $p(\boldsymbol{x}|\boldsymbol{y})$ by :

$$p(x_j|\boldsymbol{y}) = \int \cdots \int p(\boldsymbol{x}|\boldsymbol{y})\, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \cdot \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n \tag{9}$$

  and which is obtained when defining $C(\widehat{\boldsymbol{x}}, \boldsymbol{x}) = 1 - \sum_{j=1}^n \delta(\widehat{x}_j - x_j)^2$.

In general, these estimators are not linear functions of the observations $\boldsymbol{y}$. One exception is the Gaussian case. In fact, as we will see in section 4, if the relation between the observations $\boldsymbol{y}$ and the unknown variables $\boldsymbol{x}$ is linear and the direct probability laws $p(\boldsymbol{x})$ and $p(\boldsymbol{y}|\boldsymbol{x})$ are both Gaussian, then the posterior law is also Gaussian and all these estimators are equivalent and linear function of the observations $\boldsymbol{y}$.

Now, before going further in advance, let us come back in steps 1 and 2, *i.e.*; before the application of the Baye's rule. The main difficulty is that, in practical applications, we are not given directly $p(\boldsymbol{y}|\boldsymbol{x})$ and $p(\boldsymbol{x})$, and the main problem is how to assign them. In the next section we will see how the ME principle can help us to do this.

3 Maximum entropy principle as a tool for assigning probability laws

The main idea behind the Maximum Entropy (ME) principle is that, if we have not enough information about a quantity to assign it a probability law $p(\boldsymbol{x})$, we can choose, between all the probability laws satisfying the constraints defined by our prior knowledge the one which has maximum entropy. Doing this, we remain the most noncommital to the information (in the sense of Shannon) not given to us [26].

Mathematically, the ME principle can be used if this knowledge can be stated as some constraints on $p(\boldsymbol{x})$. In general these constraints are not sufficient to determine uniquely $p(\boldsymbol{x})$. Then, between all probability laws which satisfy these constraints, we choose the one which has maximum entropy [26]. The mathematical problem can be stated as follows:

Given the K constraints:

$$\mathrm{E}\{\phi_k(\boldsymbol{x})\} = \int \phi_k(\boldsymbol{x})\, p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} = d_k, \quad k = 1, \cdots, K \tag{10}$$

where $\phi_k(\boldsymbol{x})$ are known functions, determine $p(\boldsymbol{x})$ which maximizes the entropy:

$$H = -\int p(\boldsymbol{x}) \ln p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{11}$$

The solution is classically given by:

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left[-\sum_{k=1}^{K} \lambda_k\, \phi_k(\boldsymbol{x})\right] \tag{12}$$

where Z is the partition function which is given by the normalization constraint:

$$Z(\lambda_1, \lambda_2, \cdots, \lambda_k) = \int \exp\left[-\sum_{k=1}^{K} \lambda_k\, \phi_k(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x} \tag{13}$$

and the Lagrange multipliers λ_k, $k = 1, \cdots, K$ are determined by the constraints (10) by solving the following system of equations [23] :

$$-\frac{\partial \ln Z(\lambda_1, \cdots, \lambda_k)}{\partial \lambda_k} = d_k, \quad k = 1, \cdots, K. \tag{14}$$

which can also be written as :

$$G_k(\lambda_1, \cdots, \lambda_k) = \int \phi_k(\boldsymbol{x}) \exp\left[-\sum_{k=0}^{K} \lambda_k\, \phi_k(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x} = d_k,, \quad k = 0, \cdots, K, \tag{15}$$

with $\phi_0(\boldsymbol{x}) = 1$ and $d_0 = 1$. The resolution of this system can be done by either the Gauss–Newton-Raphson method or by the Newton-Raphson method. For an implementation of last one using MATLAB programming language see [52].

4 The Gaussian case

To show the whole procedure of solving the inverse problems in this Bayesian approach, let us consider a classical case, *i.e.*; *the Gaussian case.*

Let us assume that we have only an approximate information about the noise variance σ^2 and some global information about the object. More precisely, assume that what we know about $\boldsymbol{b}$ and $\boldsymbol{x}$ can be resumed as follows:

- Information about $\boldsymbol{b}$:

$$\left\{ \begin{array}{l} \mathrm{E}\{\boldsymbol{b}\} = 0 \\ \mathrm{E}\left\{\boldsymbol{b}\boldsymbol{b}^t\right\} = \boldsymbol{R}_b = \sigma_b^2 \boldsymbol{I} \end{array} \right. \tag{16}$$

  where $\boldsymbol{R}_b$ is the covariance matrix of $\boldsymbol{b}$.

- Information about $\boldsymbol{x}$:

$$\left\{ \begin{array}{l} \mathrm{E}\{\boldsymbol{x}\} = \boldsymbol{x}_0 \\ \mathrm{E}\{(\boldsymbol{x}-\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)^t\} = \boldsymbol{R}_x = \sigma_x^2 \boldsymbol{P}_0 \end{array} \right. \tag{17}$$

  where $\boldsymbol{x}_0$ is the mean value of $\boldsymbol{x}$ and $\boldsymbol{R}_x$ is the covariance matrix.

Now, using the ME principle we can assign the direct probability laws $p(\boldsymbol{x})$, $p(\boldsymbol{b})$, and $p(\boldsymbol{y}|\boldsymbol{x})$. First our knowledge about the $\boldsymbol{b}$ will give us:

$$\boldsymbol{b} \sim \mathcal{N}(0, \boldsymbol{R}_b) \longrightarrow p(\boldsymbol{b}) \propto \exp\left[-\frac{1}{2}\boldsymbol{b}^t \boldsymbol{R}_b^{-1} \boldsymbol{b}\right] \tag{18}$$

and consequently, using the modele 2 with known $\boldsymbol{A}$, we can deduce $p(\boldsymbol{y}|\boldsymbol{x})$:

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{R}_b) \longrightarrow p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^t \boldsymbol{R}_b^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})\right]. \tag{19}$$

Our knowledge about $\boldsymbol{x}$ will give us:

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}_0, \boldsymbol{R}_x) \longrightarrow p(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{R}_x^{-1}(\boldsymbol{x}-\boldsymbol{x}_0)\right]. \tag{20}$$

Now, using the Bayes' formula, we have:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left[-\frac{1}{2}J(\boldsymbol{x})\right] \tag{21}$$

with

$$J(\boldsymbol{x}) = (\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^t \boldsymbol{R}_b^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) + (\boldsymbol{x}-\boldsymbol{x}_0)^t \boldsymbol{R}_x^{-1}(\boldsymbol{x}-\boldsymbol{x}_0), \tag{22}$$

which is a quadratic function of $\boldsymbol{x}$. Consequently, $p(\boldsymbol{x}|\boldsymbol{y})$ is a Gaussian probability density function, and it is easy to show that its mean $\widehat{\boldsymbol{x}}$, and its covariance matrix $\boldsymbol{P}$ are given by:

$$\boldsymbol{x}|\boldsymbol{y} \sim \mathcal{N}(\widehat{\boldsymbol{x}}, \boldsymbol{P}) \quad \text{with} \quad \left\{ \begin{array}{lcl} \widehat{\boldsymbol{x}} &=& \boldsymbol{x}_0 + \boldsymbol{P}\boldsymbol{A}^t \boldsymbol{R}_b^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0) \\ \boldsymbol{P} &=& \left(\boldsymbol{A}^t \boldsymbol{R}_b^{-1}\boldsymbol{A} + \boldsymbol{R}_x^{-1}\right)^{-1} \end{array} \right. \tag{23}$$

If now we replace $\boldsymbol{R}_b = \sigma_b^2 \boldsymbol{I}$ and $\boldsymbol{R}_x = \sigma_x^2 \boldsymbol{P}_0$ we obtain:

$$\begin{array}{lcl} \widehat{\boldsymbol{x}} &=& \left(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1}\left(\boldsymbol{A}^t\boldsymbol{y} + \lambda \boldsymbol{P}_0^{-1}\boldsymbol{x}_0\right) \\ &=& \boldsymbol{x}_0 + \boldsymbol{R}_x\boldsymbol{A}^t(\boldsymbol{A}\boldsymbol{R}_x\boldsymbol{A}^t + \lambda \boldsymbol{I})^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}_0) \\ \boldsymbol{P} &=& \sigma_b^2\left(\boldsymbol{A}^t\boldsymbol{A} + \lambda \boldsymbol{P}_0^{-1}\right)^{-1} \end{array} \tag{24}$$

with $\lambda = \sigma_b^2/\sigma_x^2$. Note also the following properties :

- In this special Gaussian case $\widehat{\boldsymbol{x}}$ is both the posterior mean value of $\boldsymbol{x}$ and the argument which maximizes the posterior law:

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\} = \arg\min_{\boldsymbol{x}} \{J(\boldsymbol{x})\} = \mathrm{E}_{X|Y}\{\boldsymbol{x}\} \tag{25}$$

- Minimization of the criterion $J(\boldsymbol{x})$ which can also be written in the form:

$$J(\boldsymbol{x}) = ||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||^2_{\boldsymbol{R}_b} + \lambda||\boldsymbol{x} - \boldsymbol{x}_0||^2_{\boldsymbol{R}_x} \tag{26}$$

  can be considered as a regularisation procedure to the inverse problem (2). Indeed, the Bayesian approach will give us here a new interpretation of the regularisation parameter in terms of the signal to noise ratio, *i.e.*; $\lambda = \sigma_b^2/\sigma_x^2$.

- The solution $\widehat{\boldsymbol{x}}$ is a linear function of the data $\boldsymbol{y}$. This is due to the fact that the problem is linear and all the probability laws are Gaussian. In general however, $\widehat{\boldsymbol{x}}$ is not a linear function of the data $\boldsymbol{y}$. We have asked ourselves if it is possible to find a class of probability laws for $p(\boldsymbol{x})$ and for $p(\boldsymbol{y}|\boldsymbol{x})$ who result to Bayesian estimators with some weaker properties than linearity. In particular we derived the class of probability laws which result to Bayesian estimators who have *scale invariance* property [53, 54, 55].

  may not need that the solution be a linear function of the data $\boldsymbol{y}$, but the scale invariance is the minimum property which is needed. We will see that the solutions obtained by the proposed method satisfy this property.

5 Outline of the proposed method

As we have seen, the main difficulty in applying the Bayesian approach in practical applications is in the first step which is assigning the direct probabilities $p(\boldsymbol{x})$ and $p(\boldsymbol{y}|\boldsymbol{x})$. In fact, $p(\boldsymbol{y}|\boldsymbol{x})$ is related to the noise probability law, and $p(\boldsymbol{x})$ is a prior law on $\boldsymbol{x}$. Assigning $p(\boldsymbol{y}|\boldsymbol{x})$ is more commonly understood and accepted but assigning $p(\boldsymbol{x})$ is more unusual and still misunderstood.

What we propose in this paper is to use the Bayesian framework and the ME principle to solve inverse problems arising in image reconstruction and restoration applications. Formally, we follow exactly the same steps as in the last section; we do the same hypothesis about the noise but not the same about the image $\boldsymbol{x}$. This is because, in many imaging applications, the image $\boldsymbol{x}$ is positive and the Gaussian prior law is no more very appropriate. We propose to use the ME principle to derive a family of prior laws with special properties which can be used more appropriately in imaging applications.

So we discuss first in detail how to assign a prior law $p(\boldsymbol{x})$ which can reflect our prior knowledge about the image. We will show that, with some global constraints on the image $\boldsymbol{x}$, *i.e.*; knowing that $x_j > 0$ and two statistical constraints which are the expectation of two functions $\phi_1(\boldsymbol{x})$ and $\phi_2(\boldsymbol{x})$, $p(\boldsymbol{x})$ is in the form:

$$p(\boldsymbol{x}) \propto \exp\left[-\lambda\phi_2(\boldsymbol{x}) - \mu\phi_1(\boldsymbol{x})\right] \quad \text{with} \quad \phi_2(\boldsymbol{x}) = \sum_{j=1}^{n} H(x_j) \quad \text{and} \quad \phi_1(\boldsymbol{x}) = \sum_{j=1}^{n} S(x_j) \tag{27}$$

We discuss then about the possible forms of the functions S and H which insure a scale invariance property and show that (S, H) can be one of the following:

$$\Big\{S(x), H(x)\Big\} = \Big\{(x^{r_1}, x^{r_2}), (x^{r_1}, \ln x), (x^{r_1}, x^{r_1}\ln x), (\ln x, \ln^2 x)\Big\}. \tag{28}$$

We consider then the case where $S(x) = x$, and so where $H(x)$ is one of the following:

$$\left\{ H(x) \right\} = \left\{ x^r, \ln x, x \ln x \right\}. \tag{29}$$

With the relations (19) and (27) the estimation problem (25) is then equivalent to:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}>0} \left\{ J(\boldsymbol{x}; \lambda, \mu) = Q(\boldsymbol{x}) + \lambda\phi_2(\boldsymbol{x}) + \mu\phi_1(\boldsymbol{x}) \right\} \tag{30}$$

which can also be considered as the solution of a regularisation problem in which (λ, μ) are the regularisation parameters (hyperparameters). Then we address the two main difficulties in real applications which are:

- When the hyperparameters (λ, μ) are given how to optimise the criterion (30) ?
- How to determine the hyperparameter values (λ, μ) from the available data $\boldsymbol{y}$?

For the first, we propose a slightly modified conjugate gradient method and for the second we propose a method based on the generalized maximum likelihood (GML) to estimate the hyperparameters $\boldsymbol{\theta} = (\lambda, \mu)$, and the solution $\boldsymbol{x}$ by the joint maximisation:

$$(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{\boldsymbol{x}>0, \boldsymbol{\theta}\in\Theta} \left\{ p(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) \right\} = \arg\min_{\boldsymbol{x}>0, \boldsymbol{\theta}\in\Theta} \left\{ J(\boldsymbol{x}, \boldsymbol{\theta}) \right\} \tag{31}$$

which is implemented iteratively, *i.e.*;

$$\widehat{\boldsymbol{x}}^{(k+1)} = \arg\min_{\boldsymbol{x}>0} \left\{ J\left(\boldsymbol{x}, \widehat{\boldsymbol{\theta}}^{(k)}\right) \right\} \tag{32}$$

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \arg\min_{\boldsymbol{\theta}\in\Theta} \left\{ J\left(\widehat{\boldsymbol{x}}^{(k)}, \boldsymbol{\theta}\right) \right\} \tag{33}$$

Unfortunately it is not easy to insure that this procedure will converge in all cases and when it converges it is not sure that it converges to the optimal solution. However, in practical applications the method has given satisfaction as we could see from the simulation results. A better way should be the marginalised maximum likelihood (MML) method [56, 57], but, unfortunately, its implementation in our image applications is very difficult and needs a calculus time which is not reasonable for our applications. However, we are working on its implementation by the EM (Expectation-Maximisation) algorithm and a comparison between the two methods GML and MML is currently being studied.

To resume, the main contributions of this paper are the following:

- we give arguments how to choose the form of the prior law $p(\boldsymbol{x})$ and we derive and propose three families of two parameter prior laws suitable for use in imaging applications.
- The form of $p(\boldsymbol{x})$ will explain the different expressions of the entropy used in image restoration and reconstruction problems.
- we present a practical method to estimate iteratively the hyperparameters and the pixel values of the object.

6 Determining the form of the prior law

$p(\boldsymbol{x})$ is a prior law on $\boldsymbol{x}$. It must be as general and noninformative as possible, *i.e.*; it must reflect our prior knowledge about $\boldsymbol{x}$. In image reconstruction and restoration problems, we know for example that $x_j > 0$, and may have some statistical *a priori* knowledge about the object $\boldsymbol{x}$ such as the mean values, correlations, etc. Thus in general, we have very limited pieces of information (number of constraints) on $p(\boldsymbol{x})$. The main problem is how to determine $p(\boldsymbol{x})$ to reflect this information. In the following we give two different viewpoints which give the same results.

A Classical statistics viewpoint :

We want to determine $p(\boldsymbol{x})$ from a finite set of statistical observations (direct or indirect) on $\boldsymbol{x}$. So we limit ourselves to a parametric representation of $p(\boldsymbol{x})$ with a few number of parameters. The estimation is done from some finite scalar observation functionals $\phi_k(\boldsymbol{x}), k = 1, \cdots, K$ on the image.

First, we want that the $\{\phi_k(\boldsymbol{x}), k = 1, \cdots, K\}$ be the sufficient statistics for the prior law $p(\boldsymbol{x})$. This will conduct us to the familly of the generalised exponential laws:

$$p(\boldsymbol{x}) = \exp\left[-\sum_{k=1}^{K} \lambda_k \phi_k(\boldsymbol{x})\right] \tag{34}$$

Then, we assume that we cannot have any *a priori* information about the correlations in $\boldsymbol{x}$. This hypothesis limits $\phi_k(\boldsymbol{x})$ to be in the form:

$$\phi_k(\boldsymbol{x}) = \sum_{j=1}^{n} \phi_j(x_j), \quad k = 1, \cdots, K. \tag{35}$$

Our next hypothesis is that we do not want distinguish *a priori* any region against any other in the object which must be found. This means that the pixels are *a priori* interchangeable so that $p(\boldsymbol{x})$ must be symmetric in x_j . This limits us to choose:

$$\phi_k(\boldsymbol{x}) = \sum_{j=1}^{n} \phi(x_j), \quad k = 1, \cdots, K \tag{36}$$

So, these hypothesis conduct us to:

$$p(\boldsymbol{x}) = \exp\left[-\sum_{k=1}^{K} \lambda_k \sum_{j=1}^{n} \phi(x_j)\right] = \prod_{j=1}^{n} \exp\left[-\sum_{k=1}^{K} \lambda_k \, \phi(x_j)\right] = \prod_{j=1}^{n} p(x_j) \tag{37}$$

We can remark that this means that the pixels are considered to be *a priori* independent.

For practical reasons we have limited ourselves to a solution with two parameters ($K = 2$) for $p(\boldsymbol{x})$ and chose two scalar observation functions:

$$\phi_1(\boldsymbol{x}) = \sum_{j=1}^{n} S(x_j), \quad \phi_2(\boldsymbol{x}) = \sum_{j=1}^{n} H(x_j) \tag{38}$$

then we have:

$$p(\boldsymbol{x}) = \prod_{j=1}^{N} p(x_j), \quad \text{with} \quad p(x_j) = \frac{1}{Z_j} \exp\left[-\lambda H(x_j) - \mu S(x_j)\right] \tag{39}$$

or, equivalently:

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left[-\lambda \phi_1(\boldsymbol{x}) - \mu \phi_2(\boldsymbol{x})\right] \tag{40}$$

It sill remains how to choose the functions $S(x_i)$ and $H(x_i)$. We will discuss about it in a few minutes.

The Maximum Entropy viewpoint :

In this case we use the ME principle to directly determine the form of $p(\boldsymbol{x})$. We suppose that the only *a priori* knowledge that we dispose about the object is in the form:

$$\begin{cases} \mathrm{E}\left\{\phi_1(\boldsymbol{x})\right\} = s \\ \mathrm{E}\left\{\phi_2(\boldsymbol{x})\right\} = h \end{cases} \tag{41}$$

where $\phi_1(\boldsymbol{x})$ and $\phi_2(\boldsymbol{x})$ are two known functions (for example those defined in (38) in the preceding section). With these two global constraints, the ME principle gives us an exponential probability density function, the same as in (40):

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left[-\lambda \phi_1(\boldsymbol{x}) - \mu \phi_2(\boldsymbol{x})\right] \tag{42}$$

The parameters (λ, μ) are related to (s, h). They are obtained by calculating the partition function $Z(\lambda, \mu)$:

$$Z(\lambda, \mu) = \int \exp\left[-\lambda \phi_1(\boldsymbol{x}) - \mu \phi_2(\boldsymbol{x})\right] \mathrm{d}\boldsymbol{x} \tag{43}$$

and by solving the following system of equations:

$$\begin{cases} \dfrac{-\partial \ln Z(\lambda, \mu)}{\partial \mu} = s \\ \dfrac{-\partial \ln Z(\lambda, \mu)}{\partial \lambda} = h \end{cases} \tag{44}$$

What can be noted here is that, in classical statistics viewpoint we had to choose the form of the prior law using the concept of sufficient statistics, while in this case we used the maximum entropy concept to do the task. However what is important is the result: a prior law depending on two parameters and two known functions. We have not yet discussed about these functions. In the next section we give some arguments how to choose them.

6.1 How to choose H and S.

Without any other restriction, $H(x_j)$ and $S(x_j)$ in (38) can be any function, but we want to be able to estimate the parameters λ and μ from some statistics, either direct observations on the image $\boldsymbol{x}$ or indirect observations $\boldsymbol{y}$ which are linearly related to $\boldsymbol{x}$. In a linear inverse problem the measurement scale is arbitrary, so the knowledge of s and h in equations (41–44) must give us a prior law $p(\boldsymbol{x})$ which must be independent of the scale factor of the image. In other words, when we change the scale of the image $U = kX$, the form of the prior law must stay the same and the parameters of the transformed image must only depend on the parameters of the original image and the scale factor k. This point is more extensively discussed in related works [54, 55].

This argument which we call it *the scale invariance axiom*, can be described mathematically as follows: Given the constraints:

$$\begin{cases} \mathrm{E}_X\left\{\phi_1(\boldsymbol{x})\right\} = s_1 \\ \mathrm{E}_X\left\{\phi_2(\boldsymbol{x})\right\} = h_1 \end{cases} \tag{45}$$

The ME prior law for the image $\boldsymbol{x}$ is:

$$p_X(\boldsymbol{x};\lambda_1,\mu_1) \propto \exp\left[-\lambda_1\phi_1(\boldsymbol{x}) - \mu_1\phi_2(\boldsymbol{x})\right] \tag{46}$$

Now given the same kind of constraints on an image $U = kX$:

$$\left\{\begin{array}{l} \mathrm{E}_U\left\{\phi_1(\boldsymbol{u})\right\} = s_k \\ \mathrm{E}_U\left\{\phi_2(\boldsymbol{u})\right\} = h_k \end{array}\right. \tag{47}$$

The ME prior law for the image $\boldsymbol{u}$ is:

$$p_U(\boldsymbol{u};\lambda_k,\mu_k) \propto \exp\left[-\lambda_k\,\phi_1(\boldsymbol{u}) - \mu_k\,\phi_2(\boldsymbol{u})\right] \tag{48}$$

Now, the scale invariance property becomes :

$$\forall k > 0, \quad p_U(\boldsymbol{u},\lambda_k,\mu_k) = \frac{1}{k^n}\,p_X\left(\frac{\boldsymbol{u}}{k},\lambda_1,\mu_1\right). \tag{49}$$

Noting that,

$$p_X(\boldsymbol{x};\lambda_1,\mu_1) = \prod_{j=1}^{N} p_{X_j}(x_j), \quad \text{and} \quad p_U(\boldsymbol{u};\lambda_k,\mu_k) = \prod_{j=1}^{N} p_{U_j}(u_j) \tag{50}$$

the preceding relation becomes :

$$\forall k > 0, \quad p_{U_j}(u_j,\lambda_k,\mu_k) = \frac{1}{k}\,p_{X_j}\left(\frac{u_j}{k},\lambda_1,\mu_1\right) \tag{51}$$

So we must have:

$$\forall k > 0,\ \forall x, \quad \exp\left[\lambda_k\,H(kx) + \mu_k\,S(kx)\right] \propto \exp\left[\lambda_1\,H(x) + \mu_1\,S(x)\right] \tag{52}$$

or, equivalently:

$$\forall k > 0,\ \forall x, \quad \lambda_k\,H(kx) + \mu_k\,S(kx) = \lambda_1\,H(x) + \mu_1\,S(x) + \eta_k \tag{53}$$

Using this axiom, we have shown [5] that the only choices for H and S are:

$$\left\{\Big(S(x),H(x)\Big)\right\} = \left\{(x^{r_1},x^{r_2}),(x^{r_1},\ln x, x>0),(x^{r_1},x\ln x, x>0),(\ln x,\ln^2 x, x>0)\right\} \tag{54}$$

In the special case where $S(x) = x$, the only choices for H are:

$$\left\{x^r,\ln x, x\ln x\right\} \tag{55}$$

What is interesting in this result is that we find different expressions of the entropy priors. To be more explicit, first consider the case where $S(x) = x$ and $H(x) = x\ln x$. In this case the prior law $p(\boldsymbol{x})$ is

$$p(\boldsymbol{x}) = \frac{1}{Z}\exp\left[-\lambda\sum_{j=1}^{N} x_j\ln x_j - \mu\sum_{j=1}^{N} x_j\right] \tag{56}$$

which is equivalent to the *entropic* prior probability density function (*pdf*) used by Skilling *et all.* [58].

Consider now the case $S(x) = x$ and $H(x) = x^2$. In this case we have

$$p(\boldsymbol{x}) = \frac{1}{Z}\exp\left[-\lambda\sum_{j=1}^{N} x_j^2 - \mu\sum_{j=1}^{N} x_j\right] \tag{57}$$

which is a multivariate Gaussian *pdf*.

Finally, in the case where $S(x) = x$ and $H(x) = \ln x$ we have

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp \left[-\lambda \sum_{j=1}^{N} \ln x_j - \mu \sum_{j=1}^{N} x_j \right] \tag{58}$$

which is a multivariate Gamma *pdf*.

It is interesting to study the forms of some of these probability laws in function of their parameter values. For this, note that, in all cases we have :

$$p(\boldsymbol{x}) = \prod_{j=1}^{N} p(x_j), \quad \text{with} \quad p(x_j) = \frac{1}{Z_j} \exp\left[-\lambda H(x_j) - \mu S(x_j)\right]. \tag{59}$$

Now consider the following cases:

- Case 1: $S(x) = x$ and $H(x) = x^2, x \in \mathbb{R}, \lambda > 0,\ \mu \in \mathbb{R}$

  In this case we have:

$$p(x) = \frac{1}{Z(\lambda,\mu)} \exp\left[-\lambda x^2 - \mu x\right] = \frac{1}{Z(\lambda,\mu)} \exp\left[-\lambda \left(x + \frac{\mu}{2\lambda}\right)^2\right] \tag{60}$$

  which is the Gaussian probability density function, and if we note by $e(\lambda,\mu) = \mathrm{E}\{x\}$ (the mean) and by $v(\lambda,\mu) = \mathrm{E}\left\{(x-e)^2\right\}$ (the variance), we have the following relations:

$$\left\{ \begin{array}{l} v = \dfrac{1}{2\lambda} \\ e = -\dfrac{\mu}{2\lambda} \end{array} \right. \longrightarrow \left\{ \begin{array}{l} \lambda = \dfrac{1}{2v} \\ \mu = -\dfrac{e}{v} \end{array} \right. \tag{61}$$

- Case 2: $S(x) = x$ and $H(x) = x^2, x > 0, \lambda > 0,\ \mu > 0$

  In this case we have:

$$p(x) = \frac{1}{Z(\lambda,\mu)} \exp\left[-\lambda x^2 - \mu x\right], \quad x > 0 \tag{62}$$

  which is a truncated Gaussian probability density function. Note that, in this case we have not any explicit relation between (λ,μ) and (e,v).

- Case 3: $S(x) = x$ and $H(x) = \ln x, x > 0, \lambda < 1,\ \mu > 0$

  In this case $p(x)$ is in the form:

$$p(x) = A x^{-\lambda} \exp\left[-\mu x\right] \tag{63}$$

  which is a Gamma probability density function and we have the following relations:

$$\left\{ \begin{array}{l} e = \dfrac{1-\lambda}{\mu} \\ v = \dfrac{1-\lambda}{\mu^2} \end{array} \right. \longrightarrow \left\{ \begin{array}{l} \lambda = \dfrac{v - e^2}{v} \\ \mu = \dfrac{e}{v} \end{array} \right. \tag{64}$$

- Case 4: $S(x) = x$ and $H(x) = x \ln x, x > 0, \lambda > 0,\ \mu \in \mathbf{R}$
  In this case $p(x)$ is in the form:

$$p(x) = A \exp\left[-\lambda\, x \ln x - \mu\, x\right] \tag{65}$$

  and we have not analytic solutions to $Z(\lambda, \mu)$ and it is impossible to establish an analytic relation between (λ, μ) and (e, v).

An interesting property of these probability density functions is that, by construction, they are scale invariant, *i.e.*; if the *pdf* of X is

$$p_X(x, \lambda_1, \mu_1) = \frac{1}{Z_1} \exp\left[-\lambda_1\, H(x) - \mu_1\, S(x)\right], \tag{66}$$

then the *pdf* of $U = kX$ will be

$$p_U(u, \lambda_k, \mu_k) = \frac{1}{Z_k} \exp\left[-\lambda_k\, H(u) - \mu_k\, S(u)\right], \tag{67}$$

where $\{Z_k, \lambda_k, \mu_k\}$ are related to $\{Z_1, \lambda_1, \mu_1\}$ by a linear relation

$$Z_k = k Z_1, \quad \text{and} \quad \begin{pmatrix} \lambda_k \\ \mu_k \end{pmatrix} = \boldsymbol{A}_k \begin{pmatrix} \lambda_1 \\ \mu_1 \end{pmatrix}, \tag{68}$$

where $\boldsymbol{A}_k$ is a 2×2 matrix depending only on the scale factor k. This property is very interesting, because, if we are given the samples $\{u_1, \cdots, u_n\}$ and we are asked to estimate the parameters $\{\lambda_k, \mu_k\}$ we can normalize the samples by a factor k to obtain $\{x_1, \cdots, x_n\}$ and estimate the parameters $\{\lambda_1, \mu_1\}$ and then calculate the parameters $\{\lambda_k, \mu_k\}$ from them.

We are going now to explain how one can estimate the parameters (λ, μ) from the direct observations $\boldsymbol{x}$ or indirect data $\boldsymbol{y}$.

For the estimation of the parameters (λ, μ) from the direct observations $\boldsymbol{x}$, we have showed in preceding works [59, 60, 61] that there are some close relations between the two classical methods of parameter estimation of a parametric law which are the maximum likelihood (ML) method and the moment method (MM) and the calculus of the Lagrange parameters in the maximum entropy (ME) determination of the same law. Let us resume this in the following. First consider the following problem :

Problem P1: Assume that we are given the samples $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ of a random variable X, for which we assume to know the parametric *pdf* in the form :

$$p(x; \lambda, \mu) = \frac{1}{Z(\lambda, \mu)} \exp\left[-\lambda H(x) - \mu S(x)\right], \tag{69}$$

with $H(x)$ and $S(x)$ two known functions. The problem is now to estimate the parameters λ, μ from the samples $\boldsymbol{x}$. The two classical methods to solve this problem are :

- Method of moments (MM):
  The method of moments for estimating the parameters $\{\lambda, \mu\}$ from direct observations $\boldsymbol{x}$ consists of estimating $\mathrm{E}\{x\}$ and $\mathrm{E}\{x^2\}$ by their empirical values from the samples $\boldsymbol{x}$ and solving the following system of equations

$$\left\{ \begin{array}{l} G_1(\lambda, \mu) = \mathrm{E}\{x\} = \displaystyle\int x p(x; \lambda, \mu)\, \mathrm{d}x = \frac{1}{n} \sum_{j=1}^{N} x_j \\ G_2(\lambda, \mu) = \mathrm{E}\{x^2\} = \displaystyle\int x^2 p(x; \lambda, \mu)\, \mathrm{d}x = \frac{1}{n} \sum_{j=1}^{N} x_j^2 \end{array} \right. \tag{70}$$

In some special cases we can have analytical solutions to these equations.

- Method of maximum likelihood (ML) :

  This method can be resumed as follows: estimate (λ, μ) by maximizing the likelihood function $L(\lambda, \mu) = \prod_{j=1}^{n} p(x_i; \lambda, \mu)$, *i.e.*;

$$(\widehat{\lambda}, \widehat{\mu}) = \arg\max_{(\lambda,\mu)} \left\{ \prod_{j=1}^{n} p(x_i; \lambda, \mu) \right\}. \tag{71}$$

  Replacing the expression of $p(x; \lambda, \mu)$, it is easy then to show that the solution is obtained by solving :

$$\left\{ \begin{array}{l} -\dfrac{\partial \ln Z(\lambda, \mu)}{\partial \mu} = \displaystyle\int S(x) p(x; \lambda, \mu) \, \mathrm{d}x = \frac{1}{n} \sum_{j=1}^{N} H(x_j) \\ -\dfrac{\partial \ln Z(\lambda, \mu)}{\partial \lambda} = \displaystyle\int H(x) p(x; \lambda, \mu) \, \mathrm{d}x = \frac{1}{n} \sum_{j=1}^{N} S(x_j) \end{array} \right. \tag{72}$$

  We can remark that these two methods are equivalent in the case of exponential *pdf* family, *i.e.*; when $S(x) = x$ and $H(x) = x^2$.

Now, to see the relation with the ME method, consider the following problem :

Problem P2 : Assume that, the only thing that we know about the random variable X is the form :

$$\left\{ \begin{array}{l} \mathrm{E}\{S(X)\} = \mu_1 \\ \mathrm{E}\{H(X)\} = \mu_2 \end{array} \right. \tag{73}$$

and we want to determine $p(x; \lambda, \mu)$. The ME method gives the solution :

$$p(x; \lambda, \mu) = \frac{1}{Z(\lambda, \mu)} \exp\left[-\lambda H(x) - \mu S(x)\right], \tag{74}$$

where the Lagrange parameters are calculated by :

$$\left\{ \begin{array}{l} -\dfrac{\partial \ln Z(\lambda, \mu)}{\partial \mu} = \displaystyle\int S(x) p(x; \lambda, \mu) \, \mathrm{d}x = s \\ -\dfrac{\partial \ln Z(\lambda, \mu)}{\partial \lambda} = \displaystyle\int H(x) p(x; \lambda, \mu) \, \mathrm{d}x = h \end{array} \right. \tag{75}$$

Now, comparing the relations (70), (72) and (75), we can see that the two methods ML and ME are equivalent in the case of generalized exponential family *pdf*'s (H and S any function of x), and that the three methods become equivalent in the case of exponential family *pdf*'s (H and S polynomial functions of x).

Now, we are going to explain the next step which is the estimation of the parameters (λ, μ) from the indirect observations $\boldsymbol{y}$.

7 Different levels of problems to solve

Problem 1: A well-posed bayesian problem solving should be stated as follows:

Given $\boldsymbol{A}, \boldsymbol{y}, \sigma^2$, $\{\phi_1(\boldsymbol{x}) = \sum H(x_j),\ \phi_2(\boldsymbol{x}) = \sum S(x_j)\}$ and $\{\lambda, \mu\}$, estimate $\boldsymbol{x}$.

The solution is immediate; given σ^2 we have

$$p(\boldsymbol{y}|\boldsymbol{x}) \propto \exp\left[-Q(\boldsymbol{x})\right], \quad \text{with} \quad Q(\boldsymbol{x}) = \frac{1}{2\sigma_b^2}[\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}]^t[\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}] \tag{76}$$

and given $\{H(x), S(x)\}$ and $\{\lambda, \mu\}$ we have

$$p(\boldsymbol{x}) \propto \exp\left[-\lambda\phi_1(\boldsymbol{x}) - \mu\phi_2(\boldsymbol{x})\right] \tag{77}$$

and the MAP solution is

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} \{p(\boldsymbol{x}|\boldsymbol{y})\} = \arg\min_{\boldsymbol{x}} \{J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \lambda\phi_1(\boldsymbol{x}) + \mu\phi_2(\boldsymbol{x})\} \tag{78}$$

Problem 2: The next well-posed problem is

Given $\boldsymbol{A}, \boldsymbol{y}, \sigma^2$, $\{H(x),\, S(x)\}$ and $\{s = \mathrm{E}\{S(x)\},\, h = \mathrm{E}\{H(x)\}\}$, estimate $\boldsymbol{x}$.

The solution is also immediate; first using the ME method determine $\{\lambda, \mu\}$ from $\{s, h\}$ using (75) :

$$(s, h) \quad \stackrel{ME}{\longrightarrow} \quad (\lambda, \mu), \tag{79}$$

and then calculate the solution as in the preceding case.
Problem 3: The third problem is

Given $\boldsymbol{A}, \boldsymbol{y}, \sigma^2$ and $\{H(x), S(x)\}$, and a typical image $\boldsymbol{t}$, estimate $\boldsymbol{x}$ and the parameters $\{\lambda, \mu\}$.

Here, the problem is a little more complicated. To solve it, first, using the typical image $\boldsymbol{x}_t$ we have to estimate the parameters $\{\lambda, \mu\}$ of his probability law. As we have seen in the preceding section, this can be done either by the MM (70) or ML (72) methods. Then, the next step which is the estimation of $\boldsymbol{x}$ is the same as in the two preceding problems.

Noting that we have

$$p(\boldsymbol{t}; \lambda, \mu) = \frac{1}{Z(\lambda, \mu)} \exp\left[-\lambda\phi_1(\boldsymbol{t}) - \mu\phi_2(\boldsymbol{t})\right] = \prod_{j=1}^{N} p(t_j, \lambda, \mu), \tag{80}$$

with

$$p(t_j, \lambda, \mu) = \frac{1}{Z_j(\lambda, \mu)} \exp\left[-\lambda H(t_j) - \mu S(t_j)\right], \tag{81}$$

it is easy to show the following

- In the method of moments (MM) we have to solve:

$$\left\{ \begin{array}{l} G_1(\lambda, \mu) = \displaystyle\int t p(t; \lambda, \mu)\, \mathrm{d}t = \frac{1}{n} \sum_{j=1}^{N} t_j \\ G_2(\lambda, \mu) = \displaystyle\int t^2 p(t; \lambda, \mu)\, \mathrm{d}t = \frac{1}{n} \sum_{j=1}^{N} t_j^2 \end{array} \right. \tag{82}$$

- In the method of maximum likelihood (ML)we have to solve:

$$\left\{\begin{array}{l} G_1(\lambda,\mu) = \int S(t)p(t;\lambda,\mu)\,\mathrm{d}t = \sum_{j=1}^N H(t_j) \\ G_2(\lambda,\mu) = \int H(t)p(t;\lambda,\mu)\,\mathrm{d}t = \sum_{j=1}^N S(t_j) \end{array}\right. \tag{83}$$

Problem 4: The fourth problem which is a more realistic one in practical applications is

Given $\boldsymbol{A},\boldsymbol{y},\sigma^2$, and $\{H(x),S(x)\}$, estimate $\boldsymbol{x}$ and the parameters $\{\lambda,\mu\}$.

Here, the problem is more complicated because we have to estimate the parameters $\{\lambda,\mu\}$ from the indirect observations $\boldsymbol{y}$. Two classical methods in statistics for this problem are

- Method of Maximum Marginalised Likelihood (MML) :

  In this method the hyperparameters $\boldsymbol{\theta}=(\lambda,\mu)$ are estimated by

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left\{L(\boldsymbol{\theta}) = \int p(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{x}\right\} \tag{84}$$

  Unfortunately the calculus of the integral $L(\boldsymbol{\theta})$ is never possible (excepted the Gaussian case), but in some cases, with some conditions of regularity for $L(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, one can obtain the solution using the Expectation–Maximisation (EM) algorithm [62], which in our case, is given by :

$$\left\{\begin{array}{lll} \text{E:} & Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)}) & = \mathrm{E}_{X|Y;\widehat{\theta}^{(k)}}\{\ln p(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta})\} \\ \text{M:} & \widehat{\boldsymbol{\theta}}^{(k+1)} & = \arg\max_{\boldsymbol{\theta}}\left\{Q(\boldsymbol{\theta};\widehat{\boldsymbol{\theta}}^{(k)})\right\} \end{array}\right. \tag{85}$$

  However, even the implementation of this algorithm is not easy in our case, because the E–step needs the calculation of the posterior law and the calculus of expectation integral which is, a N–tuple integration in general, and needs the calculation N simple integrals in our case. We will come a little more in details on this remark.

- Method of Generalized Maximum Likelihood (GML) :

  The main idea behind this approach is to consider the hyperparameters $\boldsymbol{\theta}$ on the same level that the other parameters $\boldsymbol{x}$ and try to estimate them simultaneously by:

$$(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{x},\boldsymbol{\theta})}\{p(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\theta}) = p(\boldsymbol{x};\boldsymbol{\theta})\,p(\boldsymbol{y}|\boldsymbol{x})\} \tag{86}$$

  This optimization problem can be implemented by successively maximizing with respect to $\boldsymbol{\theta}$ and to $\boldsymbol{x}$:

$$\left\{\begin{array}{lll} \widehat{\boldsymbol{\theta}}^{(k)} & = & \arg\max_{\boldsymbol{\theta}}\left\{p(\widehat{\boldsymbol{x}}^{(k)},\boldsymbol{y};\boldsymbol{\theta})\right\} \\ \widehat{\boldsymbol{x}}^{(k+1)} & = & \arg\max_{\boldsymbol{x}}\left\{p(\boldsymbol{x}|\boldsymbol{y};\widehat{\boldsymbol{\theta}}^{(k)})\right\} \end{array}\right. \tag{87}$$

  Note that the first equation is equivalent to

$$\widehat{\boldsymbol{\theta}}^{(k)} = \arg\max_{\boldsymbol{\theta}} \left\{ p(\widehat{\boldsymbol{x}}^{(k)};\boldsymbol{\theta}) \right\} \tag{88}$$

which can be interpreted as a maximum likelihood estimator of $\boldsymbol{\theta}$ if $\boldsymbol{x}^{(k)}$ could be considered as a sample of the prior law $p(\boldsymbol{x};\boldsymbol{\theta})$.

It has been shown [57] that, the GML criterion (86) may not have a maximum, so this method has not all the good properties of the MML method. However, the implementation of the MML is not easy in our case, so we have chosen the GML method which we present in this paper which will be detailed in the next section. Another point is that, even if the criterion has a maximum, it is not sure that the iterative algorithme of (87) converges to that maximum. Some cautions are necessary when implementing this method. For now, let us briefly compare these two methods in our case.

Noting back $\boldsymbol{\theta} = \{\lambda, \mu\}$ and replacing $p(\boldsymbol{x};\lambda,\mu)$ from (40) and $p(\boldsymbol{y}|\boldsymbol{x})$ from (19) in the equations (84), (85), (86) and (87) we can make the following comparison :

- **MML :** When using the EM algorithm (85) to solve the MML problem (84), at iteration $(k+1)$ we have to estimate

$$\left(\widehat{\lambda};\widehat{\mu}\right)^{(k+1)} = \arg\max_{(\lambda;\mu)} \left\{ Q\left((\lambda;\mu), \left(\widehat{\lambda};\widehat{\mu}\right)^{(k)} \right) \right\} \tag{89}$$

In our case, this will lead to the following system of nonlinear equations :

$$\left\{ \begin{array}{lcl} -\dfrac{\partial \ln Z\,(\lambda;\mu)}{\partial \lambda_1} & = & \dfrac{1}{N}\displaystyle\int \phi_1(\boldsymbol{x}) p(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} = \dfrac{1}{N}\sum_{j=1}^{N}\int H(x_j)p(x_j|\boldsymbol{y})\,\mathrm{d}x_j \\ -\dfrac{\partial \ln Z\,(\lambda;\mu)}{\partial \lambda_2} & = & \dfrac{1}{N}\displaystyle\int \phi_2(\boldsymbol{x}) p(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} = \dfrac{1}{N}\sum_{j=1}^{N}\int S(x_j)p(x_j|\boldsymbol{y})\,\mathrm{d}x_j \end{array} \right. \tag{90}$$

- **GML :** At iteration $(k+1)$ of the iterative optimization algorithm (87) we have to estimate

$$\left(\widehat{\lambda};\widehat{\mu}\right)^{(k+1)} = \arg\max_{(\lambda;\mu)} \left\{ p(\widehat{\boldsymbol{x}}_{MAP}^{(k)};\lambda;\mu) \right\} \tag{91}$$

To do this we have to solve the following system of nonlinear equations :

$$\left\{ \begin{array}{lcl} -\dfrac{\partial \ln Z\,(\lambda;\mu)}{\partial \lambda} & = & \dfrac{1}{N}\phi_1\left(\widehat{\boldsymbol{x}}_{MAP}^{(k)}\right) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} H(\widehat{x}_{MAP_j}^{(k)}) \\ -\dfrac{\partial \ln Z\,(\lambda;\mu)}{\partial \mu} & = & \dfrac{1}{N}\phi_2\left(\widehat{\boldsymbol{x}}_{MAP}^{(k)}\right) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} S(\widehat{x}_{MAP_j}^{(k)}) \end{array} \right. \tag{92}$$

Thus, comparing (90) and (92), we can say that if $p(\boldsymbol{x}|\boldsymbol{y})$ is very concentrated around the $\widehat{\boldsymbol{x}}_{MAP}$, *i.e.*; $p(\boldsymbol{x}|\boldsymbol{y}) \approx \delta(\boldsymbol{x} - \widehat{\boldsymbol{x}}_{MAP})$, then the integrals of the RHS of the equation (90) will be equivalent to RHS of (92) and the two methods will give the same numerical result. But this is not true in general.

The GML method has not, in general, the good properties of an estimation method. In fact it has been shown that, the GML criterion may not have even a local maximum [57]. However GML is easy to implement, because it does not need any integration. This is the reason why we have chosen this method for practical applications. We will discuss more about this in the next section.

Problem 5: The fifth problem is

Given $\boldsymbol{A}, \boldsymbol{y}$, and $\{H(x), S(x)\}$, estimate $\boldsymbol{x}$ and the parameters σ_b and $\{\lambda, \mu\}$.

This problem can also be handled as in the preceding one. As the whole estimation problem satisfies the scale invariance axiom, it is easy to separate the problem to two parts: estimation of $\{\lambda, \mu\}$ and the estimation of σ_b. We have recently shown [63] that, the GML criterion, in this case may not have even a local maximum if we keep free all these three parameters. Fixing σ_b may not eliminate this uncertainty. A solution may be to keep fixed the signal to noise ratio σ_b^2/σ_x^2. Noting that σ_x^2 is related to (λ, μ), this means that some functional constraint on the values of (λ, μ) can be helpful to insure that the GML criterion have a global maximum. We are still working on this and we will report on this work in near future.

8 Proposed method

In this paper we considered the problem 4, *i.e.*; the following problem :

Given $\boldsymbol{A}, \boldsymbol{y}, \sigma^2$ and the functions $(S(x), H(x))$, estimate $\boldsymbol{x}$ and the hyperparameters $\boldsymbol{\theta} = [\lambda, \mu]$.

As we discussed in the last section the proposed method is based on the GML method which can be resumed as follows:

$$\left\{ \begin{array}{rcl} \int H(x) p(x; \lambda, \mu) \, \mathrm{d}x & = & \frac{1}{N} \sum_{j=1}^{N} H(\widehat{x}_{MAP_j}^{(k)}) \\ \int S(x) p(x; \lambda, \mu) \, \mathrm{d}x & = & \frac{1}{N} \sum_{j=1}^{N} S(\widehat{x}_{MAP_j}^{(k)}) \end{array} \right. \tag{93}$$

where

$$\widehat{x}_{MAP}^{(k)} = \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{x}|\boldsymbol{y}; \widehat{\lambda}^{(k-1)}, \widehat{\mu}^{(k-1)}) \right\} \tag{94}$$

is the MAP estimate at the iteration $(k-1)$ when the hyperparameters are assumed to have the values $\left(\widehat{\lambda}^{(k-1)}, \widehat{\mu}^{(k-1)}\right)$.

In a preceding work, we had replaced the parameter estimation step (93), which can be considered as as the ML method of estimating (λ, μ) from the observation at the iteration $(k-1)$, *i.e.*; $\widehat{\boldsymbol{x}}^{(k-1)}$, by the MM (method of moments). The reason for this was that in some cases, as we have seen in section 6, it is possible to obtain analytical and explicit relations between (λ, μ) and (e, v) the mean and the variance of the estimated image at the iteration $(k-1)$. So, the calculus cost of the estimation of the hyperparameters is reduced to the calculation of the mean and the variance of the image at each iteration.

Summary of the Method : The method described above is then the following:

- The algorithm is initialized by either an approximate solution obtained previously or by:

$$\widehat{\boldsymbol{x}}^{(1)} = \frac{1}{\sum a_{1j}^2} \boldsymbol{A}^t \boldsymbol{y} \tag{95}$$

- A first approximation of the hyperparameters $(\lambda, \mu)^{(1)}$ is calculated using $\widehat{\boldsymbol{x}}^{(1)}$. This can be done either by the MM or ML methods.

  The MM method has the advantage in the cases where we have explicit relations between (λ, μ) and (e, v). This occurs in case 1: $(H = x^2, x \in \mathbf{R})$, and in case 3: $(H = \ln x, x > 0)$ of the section 6.

The ML method is applicable in all cases, but we have to calculate the integrals and solve the system of nonlinear equations (52) in each iteration.

Some constraints are applied to these hyperparameters to avoid the difficulty due to the fact that the GML criterion may not be a concave function of them. The algorithm will stop either when a maximum number of iterations is reached or when the relative changes of the hyperparameters values are less than some ϵ. In fact, we search a stationary point of the GML criterion.

- A modified conjugate gradient algorithm is used to solve approximately:

$$\widehat{\boldsymbol{x}}^{(k+1)} = \arg\max_{\boldsymbol{x}} \left\{ p(\boldsymbol{x}|\boldsymbol{y}; \widehat{\lambda}^{(k)}, \widehat{\mu}^{(k)}) \right\} = \arg\min_{\boldsymbol{x}} \left\{ J(\boldsymbol{x}, \lambda^{(k)}, \mu^{(k)}) \right\} \tag{96}$$

We have reported on this algorithm elsewhere [14, 1]. The criterion $J(\boldsymbol{x})$ is a convex function of $\boldsymbol{x}$ subject to some range constraints on the hyperparameters (λ, μ). There are two types of constraints on (λ, μ): the one due to the integrability and normality constraints of the prior law $p(x; \lambda, \mu)$, and, the other to insure the convergence of the gradient or conjugate-gradient algorithm. The conjunction of these constraints can be resumed in the following table.

Cases	$H(x)$	$S(x)$	x domain	λ	μ
Case 1	x^2	x	$\mathbf{R}$	$\lambda > 0$	$\mu \in \mathbf{R}$
Case 2	x^2	x	$x > 0$	$\lambda < 0$	$\mu > 0$
Case 3	$\ln x$	x	$x > 0$	$\lambda < 0$	$\mu > 0$
Case 4	$-x \ln x$	x	$x > 0$	$\lambda > 0$	$\mu \in \mathbf{R}$

Table 1. Conditions on the (λ, μ).

- After some iterations, a new estimate of the hyperparameters $(\lambda, \mu)^{(k+1)}$ is calculated and we continue until some stop criterion is achieved.

9 Simulation results

The main objective of these simulations is to show how the method works and to verify experimentally the convergence of the method. For this, we have first created four synthetic images (O1, O2, O3 and O4) (Figure 1), then, for each of them we displayed the histogram (Figure 2).

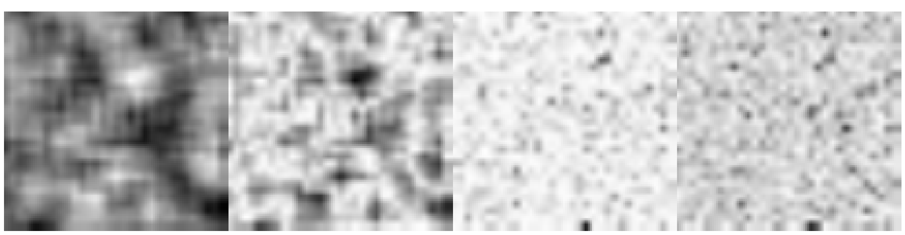

Figure 1. Four synthetic images O1, O2, O3 and O4.

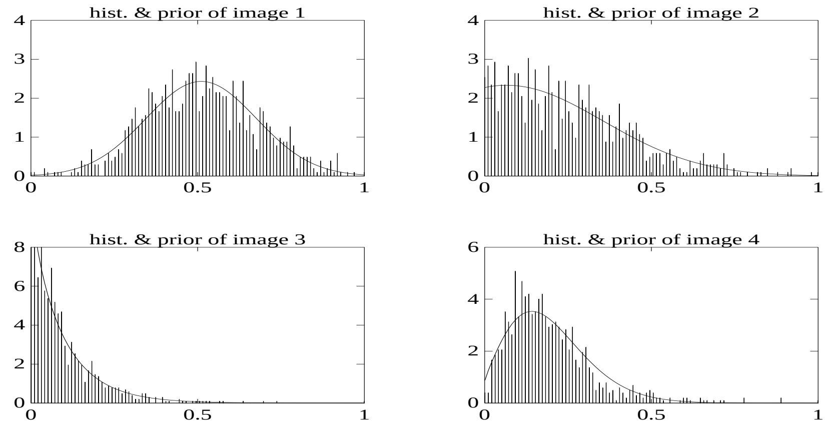

Figure 2. Histograms and the estimated prior laws.

According to the histogram, we choosed a special parametric prior law between the four cases described in section 6. and estimated their parameters by maximum likelihood (ML) method. According to the histograms of the images, we assigned

- a Gaussian law ($H(x) = x^2$, $S(x) = x$) to the first image,
- a truncated Gaussian law ($H(x) = x^2$, $S(x) = x$, $x > 0$) to the second one,
- a Gamma law ($H(x) = \ln x, S(x) = x, x > 0$) to the third one, and, finally,
- a law specified by ($H(x) = -x \ln x$, $S(x) = x$, $x > 0$) to the last one.

Figure 3 shows the histograms and the estimated prior laws and Table 2. summarizes the results of parameter estimation.

images	$H(x)$	$S(x)$	x domain	λ_1	λ_2
O1	x^2	x	$\mathbb{R}$	18.1	-18.5
O2	x^2	x	$x > 0$	5.94	-0.78
O3	$\ln x$	x	$x > 0$	0.0	9.47
O4	$-x \ln x$	x	$x > 0$	-9.93	9.54

Table 2. Prior laws and their parameters estimated by ML method.

In a next step, we created degraded images by blurring them with a Gaussian point spread function (PSF) and added a Gaussian noise with a given variance. The signal to noise ratio was fixed to 20dB. Figure 3 shows the degraded images.

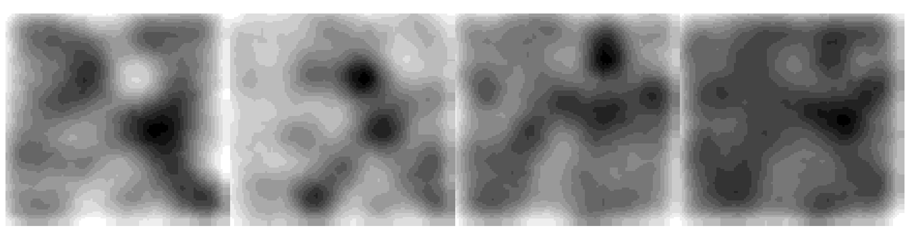

Figure 3. Degraded images.

In the final step, using the GML method we restored these images and simultaneously estimated the hyperparameters (λ, μ). Figure 4 shows the restored images and the Table 2. summarizes the hyperparameter estimation results.

images	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$J(\boldsymbol{x})$	$Q(\boldsymbol{x})$	D
O1	18.1	-17.4	5.7e4	6.2e4	2.8e-2
O2	8.17	-2.12	2.4e4	2.5e4	1.4e-1
O3	-3.54	13.7	1.9e5	1.8e5	4.3e-1
O4	-15.9	-1.91	9.8e5	1.0e6	2.3e-1

Table 3. Results of hyperparameter estimation $(\widehat{\lambda}_1, \widehat{\lambda}_2)$, the final values of the criterion $J(\boldsymbol{x})$, $Q(\boldsymbol{x})$ and D the relative distance between the original and estimated images.

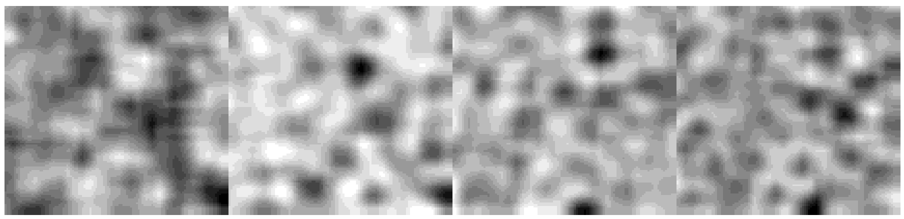

Figure 4. Restored images.

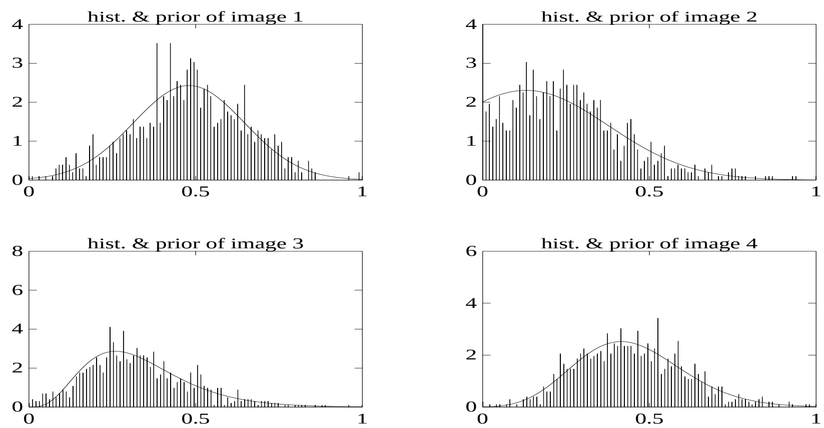

Figure 5. Histograms and the estimated prior laws after restoration.

The main objective of these simulations was to show the performances of the proposed method. The main conclusion is that, the method seems to converge practically even if its theoretical convergence is not proved. However, the final values of the hyperparameters are in some cases very far from their values estimated directly from the original objects. In fact this is normal, but the final estimation of the solution (reconstructed object) is not very sensible to the values of these hyperparameters. The reconstructed objects are more sensible on the choice of form of the prior law than the values of its parameters.

In Part II of this paper [?] we will give some results obtained by the proposed method in two image reconstruction applications.

10 Conclusions

In this paper we proposed a method based on the maximum *a posteriori* (MAP) Bayesian approach to solve the linear inverse problems (integral equations) which arise in various signal and image restoration and reconstruction problems.

A Bayesian approach is a coherent way for solving inverse problems because it allows us to take both the uncertainty on the data and the *a priori* information on the solution into account.

One major difficulty, however, is the determination of the direct probability laws (prior law of the image and the noise). We used the ME principle and a scale invariance property to achieve this task. We then considered a case where we assumed to know only the noise variance and some global constraints on the image, and applied the proposed approach. The resulting Bayesian MAP estimation procedure can be interpreted as a regularisation procedure with a criterion which depends on three hyperparameters $(\sigma_b^2, \lambda, \mu)$.

In real applications two problems arise:

- how to minimize effectively the criterion when the hyperparameters are given?
- how to determine the hyperparameters $(\sigma_b^2, \lambda, \mu)$ from the data.

For the first problem we proposed a modified conjugate gradient method. The details of the implementation of the algorithm is found elsewhere. There is not any theoretical difficulty, because, the criterion is convex, and subject to some range–constraints on the hyperparameters the convergence of the algorithm is proved [64].

For the second problem we discussed about the available statistical methods, and proposed a method based on the GML approach which make it possible to determine iteratively the hyperparameters and the solution. However, the existence of a maximum (even local) to the GML criterion is not insured theoretically. However, the proposed method seems to work practically, but we continue to work on this problem. In the proposed method, we fixed the σ_b^2 and estimated (λ, μ). According to a very preliminary and recent work [63], it seems better to fix some signal to noise ratio σ_b^2/σ_x^2. Noting that σ_x^2 is related to (λ, μ), this means that some functional constraint on the values of (λ, μ) can be helpful to insure the convexity of the GML criterion. We will report on this work in near future.

References

[1] G. Demoment, "Image reconstruction and restoration : Overview of common estimation structure and problems," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 37, pp. 2024–2036, Dec. 1989.

[2] Titterington, "General structure of regularization procedures in image reconstruction," *Å*, vol. 144, pp. 381–387, 1985.

[3] T. D. M., "Common structure of smoothing techniques in statistics," *International Statistical Review*, vol. 53, pp. 141–170, 1985.

[4] A. C. Kak, "Computerized tomography with x-ray, emission, and ultrasound sources," *Proceedings of the IEEE*, vol. 67, pp. 1245–1272, September 1979.

[5] A. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging.* IEEE Press, ieee press ed., 1987.

[6] Box and Tiao, *Bayesian Inference in Statistical Analysis.* 1973.

[7] G. Herman and A. Lent, "A computer implementation of a bayesian analysis of image reconstruction," *Inform. Contr.*, vol. 31, pp. 364–384, 1976.

[8] B. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Transactions on Communications*, vol. C-26, pp. 219–229, 1977.

[9] G. Herman, H. Hurwitz, A. Lent, and H. Lung, "On the bayesian approach to image reconstruction," *Inform. Contr.*, vol. 42, pp. 60–71, 1979.

[10] H. Trussell, "The relationship between image restoration by the maximum a posteriori method and maximum entropy method," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-28, pp. 114–117, 1980.

[11] B. Frieden, "Statistical models for the image restoration problem," *Computer Graphics and Image Processing*, vol. 12, pp. 40–59, 1980.

[12] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, p. 2, 1984.

[13] I. Csiszár, "An extended maximum entropy principle and a bayesian justification," in *Bayesian Statistics 2* (J. M. Bernardo, M. H. DeGroot, and A. F. M. Smith, eds.), pp. 83–98, North-Holland: Elsevier Science Publishers, 1985.

[14] A. Mohammad-Djafari and G. Demoment, "Image restoration and reconstruction using entropy as a regularization functional," in *Maximum Entropy and Bayesian Methods in Science and Engineering* (G. J. Erikson and S. C. Ray, eds.), vol. 2, (Dordrecht, The Netherlands), pp. 341–355, MaxEnt Workshops, Kluwer Academic Publishers, 1988.

[15] A. Mohammad-Djafari and G. Demoment, "Maximum entropy and bayesian approach in tomographic image reconstruction and restoration," in *Maximum Entropy and Bayesian Methods* (S. J., ed.), (Dordrecht, The Netherlands), pp. 195–201, MaxEnt Workshops, Kluwer Academic Publishers, 1989.

[16] J. Besag, "Digital image processing : Towards bayesian image analysis," *Journal of Applied Statistics*, vol. 16, no. 3, pp. 395–407, 1989.

[17] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified em algorithm," *IEEE Transactions on Medical Imaging*, vol. 9, pp. 84–93, Mar. 1990.

[18] G. Gindi, M. Lee, A. Rangarajan, and Z. I., "Bayesian reconstruction of functional images using anatomical information as priors," *IEEE Transactions on Medical Imaging*, vol. MI-12, no. 4, pp. 670–680, 1993.

[19] E. Jaynes, "Information theory and statistical mechanics i," *Physical review*, vol. 106, pp. 620–630, 1957.

[20] E. Jaynes, "Information theory and statistical mechanics ii," *Physical review*, vol. 108, pp. 171–190, 1957.

[21] E. Jaynes, "Prior probabilities," *IEEE Transactions on Systems Science and Cybernetics*, vol. SSC-4, pp. 227–241, Sept. 1968.

[22] E. Jaynes, "Where do we stand on maximum entropy ?," in *The Maximum Entropy Formalism* (R. Levine and M. Tribus, eds.), Cambridge (MA): M.I.T. Press, 1978.

[23] N. Agmon, Y. Alhassid, and D. Levine, "An algorithm for finding the distribution of maximal entropy," *Journal of Computational Physics*, vol. 30, pp. 250–258, 1979.

[24] E. Jaynes, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, vol. 70, pp. 939–952, 1982.

[25] E. Jaynes, "Where do we go from here?," in *Maximum-Entropy and Bayesian Methods in Inverse Problems* (J. C.R. Smith & T. Grandy, ed.), pp. 21–58, 1985.

[26] Rosenkrantz, *E.T. Jaynes : Papers on Probability, Statistics and Statistical Physics.* Dordrecht, The Netherlands: Kluwer Academic Publishers, reidel ed., 1982.

[27] B. Frieden, "Maximum-probable restoration of photon-limited images," *Applied Optics*, vol. 26, no. 9, pp. 1755–1764, 1987.

[28] A. Mohammad-Djafari, "Maximum entropy and linear inverse problems; a short review," in *Maximum Entropy and Bayesian Methods* (A. Mohammad-Djafari and G. Demoment, eds.), (Dordrecht, The Netherlands), pp. 253–264, The 12th Int. MaxEnt Workshops, Paris, France, Kluwer Academic Publishers, 1992.

[29] M. Nguyen and A. Mohammad-Djafari, "Bayesian maximum entropy image reconstruction from the microwave scattered field data," in *Maximum Entropy and Bayesian Methods* (A. Mohammad-Djafari and G. Demoment, eds.), (Dordrecht, The Netherlands), pp. 253–264, The 12th Int. MaxEnt Workshops, Paris, France, Kluwer Academic Publishers, 1992.

[30] M. Nikolova and A. Mohammad-Djafari, "Maximum entropy image reconstruction in eddy current tomography," in *Maximum Entropy and Bayesian Methods, Proc. of the 12th Int. MaxEnt Workshops* (A. Mohammad-Djafari and G. Demoment, eds.), (Dordrecht, The Netherlands), pp. 253–264, The 12th Int. MaxEnt Workshops, Paris, France, Kluwer Academic Publishers, 1992.

[31] B. Frieden, "Restoring with maximum likelihood and maximum entropy," *Journal of the Optical Society of America*, vol. 62, pp. 511–518, Apr. 1972.

[32] R. Kikuchi and B. Soffer, "Maximum entropy image restoration. i. the entropy expression," *Journal of the Optical Society of America*, vol. 67, pp. 1656–1665, Dec. 1977.

[33] S. Wernecke and L. D'Addario, "Maximum entropy image reconstruction," *IEEE Transactions on Computers*, vol. C-26, pp. 351–364, Apr. 1977.

[34] B. Frieden, "Restoring with maximum entropy.iii.poisson sources and backgrounds," *Journal of the Optical Society of America*, vol. 68, pp. 93–103, 1978.

[35] S. F. Gull and G. J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686–690, 1978.

[36] G. Minerbo, "Ment: a maximum entropy algorithm for reconstructing a source from projection data," *Computer Graphics and Image Processing*, vol. 10, pp. 48–68, 1979.

[37] C. Smith, "A dual method for maximum entropy restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, pp. 411–414, 1979.

[38] R. Bryan and J. Skilling, "Deconvolution by maximum entropy as illustrated by application to the jet of m87," *Monthly Notices of the Royal Astronomical Society*, vol. 191, pp. 69–79, 1980.

[39] G. Daniell and S. Gull, "Maximum entropy algorithm applied to image enhancement," *Proceedings of the IEE*, vol. 127E, pp. 170–172, 1980.

[40] S. Burch, S. Gull, and J. Skilling, "Image restoration by a powerful maximum entropy method," *Computer Vision and Graphics and Image Processing*, vol. 23, pp. 113–128, 1983.

[41] S. Sibisi, "Two-dimensional reconstructions from one-dimensional data by maximum entropy," *Nature*, vol. 301, pp. 134–136, 1983.

[42] S. Gull and J. Skilling, "Maximum entropy method in image processing," *Proceedings of the IEE*, vol. 131-F, pp. 646–659, 1984.

[43] J. Skilling and S. Gull, "Maximum entropy method in image processing," *Proceedings of the IEE*, vol. 131, pp. 646–659, 1984.

[44] J. Skilling and S. Gull, "Maximum entropy method in image processing," *Proceedings of the IEE*, vol. 131, pp. 646–659, 1984.

[45] J. Navaza, "On the maximum entropy estimate of electron density function," *Acta Crystallographica*, vol. A–41, pp. 232–244, 1985.

[46] R. Narayan and R. Nityananda, "Maximum entropy image restoration in astronomy," *Ann. Rev. Astron. Astrophys.*, vol. 24, pp. 127–170, 1986.

[47] A. Mohammad-Djafari and G. Demoment, "Maximum entropy diffraction tomography," in *Proceedings of IEEE ICASSP*, (Tokyo, Japan), pp. 1–34, 1986.

[48] X. Zhuang, E. Ostevold, and R. M. Haralick, "A differential equation approach to maximum entropy image restoration," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 2, pp. 208–218, Feb. 1987.

[49] A. Mohammad-Djafari and G. Demoment, "Maximum entropy reconstruction in x ray and diffraction tomography," *IEEE Transactions on Medical Imaging*, vol. 7, no. 4, pp. 345–354, 1988.

[50] Buck and Macaulay, "Linear inversion by the method of maximum entropy," in *Maximum Entropy and Bayesian Methods* (J. Skilling, ed.), (Dordrecht, The Netherlands), Kluwer Academic Publishers, 1989.

[51] N. Dusaussoy and I. Abdou, "The extended ment algorithm : A maximum entropy type algorithm using prior knowledge for computerized tomography," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1164–1180, 1991.

[52] A. Mohammad-Djafari, "A matlab program to calculate the maximum entropy distributions," in *Maximum Entropy and Bayesian Methods in Science and Engineering* (C. Smith, G. Erikson, and P. Neudorfer, eds.), (Dordrecht, The Netherlands), pp. 221–233, The 11th Int. MaxEnt Workshops, Seattle, USA, Kluwer Academic Publishers, 1991.

[53] A. Mohammad-Djafari and J. Idier, "Scale invariant bayesian estimators for linear inverse problems," in *Proc. of the First ISBA meeting*, (San Fransisco, USA), Aug. 1993.

[54] S. Brette, J. Idier, and A. Mohamad-Djafari, "Scale invariant markov models for bayesian inversion of linear inverse problems.," in *Proceedings of 14th Int. MaxEnt Workshop* (J. Skilling, ed.), (Cambridge, England), MaxEnt, Aug. 1994.

[55] S. Brette, J. Idier, and A. Mohamad-Djafari, "Scale invariant markov models for linear inverse problems," in *Proc. of the Second Int. Meeting of ISBA*, (Allicante, Spain), June 1994.

[56] M. Miller and D. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and toeplitz constrained covariances," *Proceedings of the IEEE*, vol. 75, pp. 892–906, July 1987.

[57] E. Gassiat, F. Monfront, and Y. Goussard, "On simultaneous signal estimation and parameter identification using a generalized likelihood approach," *IEEE Transactions on Information Theory*, vol. IT-38, pp. 157–162, 1992.

[58] J. Skilling, *Maximum-Entropy and Bayesian Methods.* Dordrecht, The Netherlands: Kluwer Academic Publisher, 1988.

[59] A. Mohammad-Djafari and G. Demoment, "Estimating priors in maximum entropy image processing," in *Proceedings of IEEE ICASSP*, pp. 2069–2072, IEEE, 1990.

[60] A. Mohammad-Djafari and J. Idier, "Maximum entropy prior laws of images and estimation of their parameters," in *Maximum Entropy and Bayesian Methods in Science and Engineering* (T. Grandy, ed.), (Dordrecht, The Netherlands), MaxEnt Workshops, Kluwer Academic Publishers, 1990.

[61] A. Mohammad-Djafari and J. Idier, "Maximum likelihood estimation of the lagrange parameters of the maximum entropy distributions," in *Maximum Entropy and Bayesian Methods in Science and Engineering* (C. Smith, G. Erikson, and P. Neudorfer, eds.), (Dordrecht, The Netherlands), pp. 131–140, MaxEnt Workshops, Seattle 1991, USA, Kluwer Academic Publishers, 1991.

[62] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society B*, vol. 39, pp. 1–38, 1977.

[63] S. Brette and A. Mohammad-Djafari, "Approche bayésienne des problèmes inverses," rapport de stage du dea-ats, LSS, Gif-sur-Yvette, France, 1993.

[64] B. Eldarwiche and A. Mohammad-Djafari, "Maximum d'entropie en reconstruction d'image," rapport de stage, Ecole Polytechnique, LSS, Gif-sur-Yvette, France, 1992.