BAYESIAN INFERENCE WITH HIERARCHICAL PRIOR MODELS FOR INVERSE PROBLEMS IN IMAGING SYSTEMS

Ali Mohammad-Djafari

Laboratoire des signaux et systèmes (L2S)
CNRS-SUPELEC-UNIV PARIS SUD
Plateau de Moulon, 91192 Gif-sur-Yvette, FRANCE

ABSTRACT

Bayesian approach is nowadays commonly used for inverse problems. Simple prior laws (Gaussian, Generalized Gaussian, Gauss-Markov and more general Markovian priors) are common in modeling and in their use in Bayesian inference methods. But, we need still more appropriate prior models which can account for non stationnarities in signals and for the presence of the contours and homogeneous regions in images. Recently, we proposed a family of hierarchical prior models, called Gauss-Markov-Potts, which seems to be more appropriate for many applications in Imaging systems such as X ray Computed Tomography (CT) or Microwave imaging in Non Destructive Testing (NDT). In this tutorial paper, first some bacgrounds on the Bayesian inference and the tools for assignment of priors and doing efficiently the Bayesian computation is presented. Then, more specifically hieararchical models and particularly the Gauss-Markov-Potts family of prior models are presented. Finally, their real applications in image restoration, in different practical Computed Tomography (CT) or other imaging systems are presented.

1. INTRODUCTION

Bayesian inference and estimation has nowadays become a common tool in many data, signal and image processing. Even if, the basics of this approach is now well understood, in practice, there exists three main difficulties for its application. The first is assigning priors, the second is summarizing the posterior and finaly, the third is doing the final computations. In this tutorial, first some basic backgrounds are presented, then, the inverse problems approach to data, signal and image processing is presented. To illustrate in detail the three aforementioned steps, the linear inverse problems are considered.

As this paper is a tutorial one and it should be used as a support for the tutorial, it should be self-contained. Thus, first some background materials and tools are presented briefly and progressively, some new or at least state of the art materials follows:

- In section 2 the methods for assigning a probability law on a quantity which can be observed directly and estimation of its associated parameters are presented. Here, we considered the Maximum Entropy (ME) method, the Maximum Likelihood (ML) method and the Parametric and Non Parametric Bayesian methods.

- In section 3, first a very brief presentation of the invese problems focusing on the linear models for signal and image processing is given. Then, the different steps of the Bayesian approach for them, i.e., assigning the likelihood term, assigning the priors and finding the expression of the posterior law and, finally, doing the computations are presented.

- In section 4 the family of Gauss-Markov-Potts priors is presented.

- In section 5 the problem of hyper parameters estimation is considered and the different classical methods such as Joint MAP, MCMC and Marginalization and Expectation-Maximization (EM) methods are presented.

In section 6 the Bayesian Variational Approximation (BVA) method is presented. Then, focusing on the estimation of the hyperparameters, it is shown that this approach has also JMAP, and EM as particular cases. A comparison of these three methods with their relative advantages and drawbacks are presented.

- Section 7 is focussed on the Mixture of Gaussians priors.

- Section 8 is focussed on the Gauss-Markov-Potts model and the associated Bayesian computational tools such as MCMC and BVA. Then, references on the applications of this class of priors are given in different applications.

- Section 9 summarizes the main conclusions and
- In section 10 references, deliberately limited to the coauthors (past and present PhD students) and collaborators, are presented. The readers can refer to the references of these papers for more references.

2. ASSIGNING A PROBABILITY LAW TO A QUANTITY WHEN OBSERVED DIRECTLY

First consider, the direct observation of a quantity (variable f). Assume that we observed $\boldsymbol{f} = \{f_1, \cdots, f_N\}$ and we want to assign it a probability law. Here, we may mention four main approaches:
- Maximum Entropy approach,
- Maximum Likelihood approach,
- Parametric Bayesian approach, and
- Non Parametric Bayesian approach.

2.1. Maximum Entropy approach

The main idea in this approach is to extract (to compute) from the data a few moments

$$\mathrm{E}\left\{\phi_k(f)\right\} = \frac{1}{N}\sum_{j=1}^{N}\phi_k(f_j) = d_k, \quad k = 1, \cdots, K \quad (1)$$

The selection of $\phi_k(.)$ and their number K are arbitray (prior knowledge), for example, the arithmetic moments where $\phi_k(x) = x^k$, harmonic means $\phi_k(x) = e^{j\omega_k x}$ or any other polynomial or geometrical functions. The next step is to select $p(f)$ which has its entropy

$$H = -\int p(f)\ln p(f)\ \mathrm{d}f \quad (2)$$

maximum subject to the constraints

$$\mathrm{E}\left\{\phi_k(f)\right\} = \int \phi_k(f)\, p(f)\ \mathrm{d}f = d_k, \quad k = 1, \cdots, K. \quad (3)$$

The solution to this linearly constrained optimization is easily obtained using the Lagrangian technics. It is given by:

$$p(f) = \frac{1}{Z}\exp\left[\sum_{k=1}^{K}\lambda_k\phi_k(f)\right] = \exp\left[\sum_{k=0}^{K}\lambda_k\phi_k(f)\right] \quad (4)$$

which can also be written as

$$p(f) = \exp\left[\sum_{k=0}^{K}\lambda_k\phi_k(f)\right] \text{ with } \phi_0 = 1 \text{ and } \lambda_0 = -\ln Z \quad (5)$$

where

$$Z = \exp\left[-\lambda_0\right] = \int \exp\left[\sum_{k=1}^{K}\lambda_k\phi_k(f)\right]\ \mathrm{d}f \quad (6)$$

and where $\lambda_k, k = 1, \cdots, K$ are obtained from the K constraints and Z from the normailty $\int p(f)\ \mathrm{d}f = 1$.

Now, assuming that the data are observed independently from each other, we have

$$p(\boldsymbol{f}) = \prod_{j=1}^{N} p(f_j) = \frac{1}{Z^N}\exp\left[\sum_{j=1}^{N}\sum_{k=1}^{K}\lambda_k\phi_k(f_j)\right]. \quad (7)$$

For more details on Maximum Entropy based methods refer to [1, 2, 3, 4] and their cited references.

2.2. Maximum Likelihood approach

In this approach, first a parametric family $p(f_j|\boldsymbol{\theta})$ is choson (Prior knowledge). Then, assuming that the data are observed independently from each other, the likelihood is defined

$$p(\boldsymbol{f}|\boldsymbol{\theta}) = \prod_{j=1}^{N} p(f_j|\boldsymbol{\theta}) \quad (8)$$

and the Maximum Likelihood estimte is defined as to be

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left\{p(\boldsymbol{f}|\boldsymbol{\theta})\right\} = \arg\min_{\boldsymbol{\theta}}\left\{-\sum_{j=1}^{N}\ln p(f_j|\boldsymbol{\theta})\right\} \quad (9)$$

It is shown that, for generalized exponential families, there is a direct link between ME and ML methods [5].

2.3. Parametric Bayesian approach

In this approach too, first a parametric family $p(f_j|\boldsymbol{\theta})$ is chosen (Prior knowledge). Then the likelihood is defined as in the previous case. The main difference is that, here a prior law $p(\boldsymbol{\theta}|\boldsymbol{\phi}_0)$ is also assigned to the parameters and then, using the Bayes rule:

$$p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{\phi}_0) = \frac{p(\boldsymbol{f}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\boldsymbol{\phi}_0)}{p(\boldsymbol{f}|\boldsymbol{\phi}_0)} \quad (10)$$

the expression of the posterior law is obtained, from which, we can infer on $\boldsymbol{\theta}$, using for example the Maximum A posteriori (MAP) estimate

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\left\{p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{\phi}_0)\right\} \quad (11)$$

or the Posterior Mean (PM)

$$\widehat{\boldsymbol{\theta}} = \int \boldsymbol{\theta}\, p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{\phi}_0)\ \mathrm{d}\boldsymbol{\theta}. \quad (12)$$

When a value for $\boldsymbol{\theta}$ is found, the probability law $p(\boldsymbol{f}|\boldsymbol{\theta})$ is determined. A main question here is how to assign the prior $p(\boldsymbol{\theta}|\boldsymbol{\phi}_0)$? There are a few different approaches: Conjugate priors, Reference priors, Jeffreys prior, ... For some discussion, see [6, 1, 7, 8, 3, 9].

2.4. Non Parametric Bayesian approach

In the classical parametric Bayesian approach, first a parametric family $p(f_j|\boldsymbol{\theta})$, for example a finite mixture of Gaussians

$$p(f_j|\boldsymbol{\theta}) = \sum_{k=1}^{K}\alpha_k\mathcal{N}(f_j|\mu_k, v_k) \text{ with } \sum_{k=1}^{K}\alpha_k = 1 \quad (13)$$

and then the parameters $\boldsymbol{\theta} = \{\alpha_k, \mu_k, v_k\} = (\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{v})$ are estimated. Here, the number of components of the mixture is fixed in advance. One simple way to present the Non Parametric modeling is to consider the same mixture model, but leaving the number of components to be estimated from the data. Another way, more mathematically presented is to consider the desired probability law as a function on wich we want to assign a probability law. Here, Dirichlet process which is a Discrete process accompanied with a continuous function (often Gaussian shape) can be used [10, 11, 12, 13]

3. BAYESIAN APPROACH FOR INVERSE PROBLEMS

3.1. Inverse problems

In many generic inverse problems in signal and image processing, the problem can be described as follows: Infer on

an unknown signal $f(t)$ from an observed signal $g(t')$ related between them through an operator $\mathcal{H} : f(t) \mapsto g(t)$. When this operator is linear, we can write:

$$g(t') = \int h(t, t')\, f(t)\, \mathrm{d}t \tag{14}$$

A very specific example is the deconvolution problem where $h(t, t') = h(t - t')$:

$$g(t') = \int h(t - t')\, f(t)\, \mathrm{d}t \tag{15}$$

The same relations can be written in image processing, for the general case

$$g(\boldsymbol{r}') = \int h(\boldsymbol{r}, \boldsymbol{r}')\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \tag{16}$$

where $\boldsymbol{r} = (x, y)$ and $\boldsymbol{r}' = (x', y')$ and the particular case of the image restoration which is:

$$g(\boldsymbol{r}') = \int h(\boldsymbol{r} - \boldsymbol{r}')\, f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}. \tag{17}$$

A third example is the Radon Transform

$$g(r, \phi) = \iint \delta(r - x\cos\phi - y\sin\phi)\, f(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{18}$$

which is used in Computed Tomography (CT). It is easy to see that if we note by $\boldsymbol{r}' = (r, \phi)$ and by $\boldsymbol{r} = (x, y)$, this relation is a particular case of (16).

When these relations are linear and we discretize them (using any moment method), we arrive to the relation:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{19}$$

where $\boldsymbol{f} = [f_1, \cdots, f_n]'$ represents the unknowns, $\boldsymbol{g} = [g_1, \cdots, g_m]'$ the observed data, $\boldsymbol{\epsilon} = [\epsilon_1, \cdots, \epsilon_m]'$ the errors of modelling and measurement and $\boldsymbol{H}$ the matrix of the system response.

3.2. Basics of the Bayesian approach

From this point, the main objective is to infer on $\boldsymbol{f}$ given the forward model (19), the data $\boldsymbol{g}$ and the matrix $\boldsymbol{H}$. By being Bayesian, we mean to use the Bayes rule:

$$p(\boldsymbol{f}|\boldsymbol{g}) = \frac{p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f})}{p(\boldsymbol{g})} \propto p(\boldsymbol{g}|\boldsymbol{f})\, p(\boldsymbol{f}) \tag{20}$$

to obtain what is called the posterior law $p(\boldsymbol{f}|\boldsymbol{g})$ from the likelihood $p(\boldsymbol{g}|\boldsymbol{f})$ and the prior $p(\boldsymbol{f})$. This posterior law combines the knowledge coming from the forward model and data (likelihood) and the prior knowledge.

However, to be able to use the Bayesian approach, first we need to assign $p(\boldsymbol{g}|\boldsymbol{f})$ and $p(\boldsymbol{f})$. Then, we can obtain the expression of the posterior law. Finally, we can infer on $\boldsymbol{f}$ using this posterior law.

3.3. Assigning the likelihood $p(\boldsymbol{g}|\boldsymbol{f})$

This step uses the forward model: $\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}$ and some prior knowledge about the error term $\boldsymbol{\epsilon}$. In fact, if we can assign a probability law $p(\boldsymbol{\epsilon})$, then, we can deduce the likelihood term $p(\boldsymbol{g}|\boldsymbol{f})$.

To assign $p(\boldsymbol{\epsilon})$ the things are more usual. Very often, a Gaussian prior is assigned because $\boldsymbol{\epsilon}$ is assumed to be centered, white and the only accessible and reasonable engineering quantity that we may know on it is its energy or power level: Signal to Noise Ratio (SNR). In terms of probability law its variance v_ϵ. Then, either using the Maximum Entropy Principle (MEP) or just the "common sens", we assign a Gaussian law:

$$p(\boldsymbol{\epsilon}) = \mathcal{N}(\boldsymbol{\epsilon}|0, v_\epsilon \boldsymbol{I}) \tag{21}$$

Now, using the Forward model (19) and this prior, we can write the expression of the forward likelihood

$$p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, v_\epsilon \boldsymbol{I}) \propto \exp\left[-\frac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] \tag{22}$$

Many other modeling for the likelihood are possible.

3.4. Assigning the prior $p(\boldsymbol{f})$

The next important step is to assign a prior to the unknown $\boldsymbol{f}$. Here too, different approaches can be used. The objective is to assign a prior law $p(\boldsymbol{f}|\boldsymbol{\theta})$ in such a way to translate our incomplete prior knowledge on $\boldsymbol{f}$.

3.4.1. Simple separable priors

A few examples of the prior knowledge we may have are:

Ex01: The signal (samples $\boldsymbol{f}$) we are looking for is the variation of the temperature at a given position over time t. It can go up or down around some nominal value f_0. However, its variation can not be too far from the nominal value. We may fixe a variance v_0 to consider this point. The two values f_0 and v_0 are given (we call them later the hyper parameters).

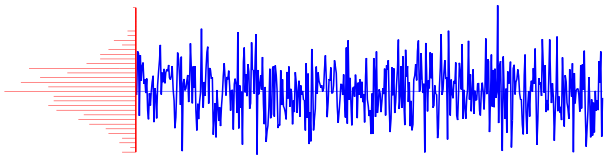
Ex02: The signal we are looking for is the distribution of the conductivity in a material. It is a positive quantity. We may also be able to fixe a mean f_0 and a variance v_0.

Ex03: The signal we are looking for is the distribution of the proportions of some material inside a body. Its value is in the interval $[0, 1]$.

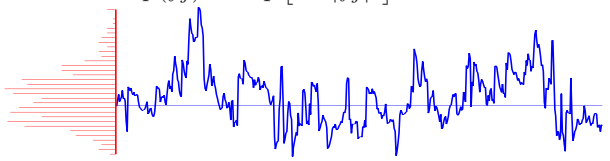
Ex04: The signal we are looking for looks like the implusions. The values can be positive or negative, very often near to zero, but it can also take great values.

Let see what we can propose for these examples:
For Ex01, we can use a Gaussian prior law

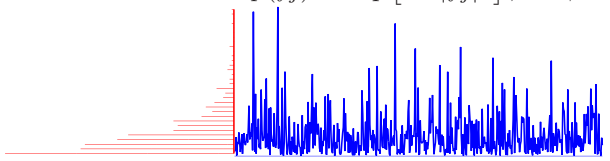
$$p(f_j) \propto \exp\left[-\frac{1}{2v_0}|f_j - f_0|^2\right] \tag{23}$$



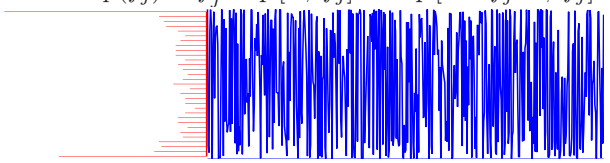
Gaussian: $p(f_j) \propto \exp\left[-\alpha|f_j|^2\right]$



Generalized Gaussian: $p(f_j) \propto \exp\left[-\alpha|f_j|^\beta\right], 1 \leq \beta \leq 2$



Gamma: $p(f_j) \propto f_j^\alpha \exp\left[-\beta f_j\right] \propto \exp\left[\alpha \ln f_j - \beta f_j\right]$



Beta: $p(f_j) \propto f_j^\alpha (1-f_j)^\beta \propto \exp\left[\alpha \ln f_j + \beta \ln(1-f_j)\right]$

Fig. 1. Separable prior laws: Gaussian, Generalized Gaussian, Gamma and Beta

For Ex02, we can use a Gamma prior law

$$p(f_j) \propto f_j^{\alpha_0} \exp\left[-\beta_0 f_j\right] \propto \exp\left[-\alpha_0 \ln f_j - \beta_0 f_j\right] \tag{24}$$

where α_0 and β_0 can be obtained from f_0 and v_0.
For Ex03, we can use a Beta prior law

$$p(f_j) \propto f_j^{\alpha_0} (1-f_j)^{\beta_0} \propto \exp\left[-\alpha_0 \ln f_j - \beta_0 \ln(1-f_j)\right] \tag{25}$$

where α_0 and β_0 can be obtained from f_0 and v_0.
For Ex04, we can use a Generalized Gaussian prior law

$$p(f_j) \propto \exp\left[-\alpha_0|f_j|^{\beta_0}\right] \tag{26}$$

where α_0 and β_0 can be obtained from f_0 and v_0.
We will call these families of prior laws as simple separable prior laws because we assume that these expressions are valid for all j and that we do not a priori know about any interactions (dependencies) between them. So, we have

$$p(\boldsymbol{f}) = \prod_j p(f_j). \tag{27}$$

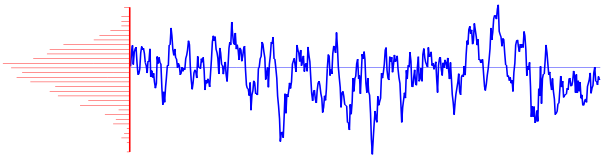
Figure 1 shows typical examples of these signals.

3.4.2. *Simple Markovian priors*

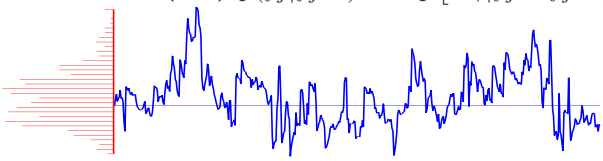
Now, let consider other cases.

Ex05 The signal we are looking for is the same as in EX01, but now, we have some extra information: The variation of the temperature can not be too fast. The two successive samples value are not independent.

Ex06 The signal we are looking for is the same as in EX05, but now, we have some extra information: In the room, there is an inhomogenious material. In some places the variation of temperature is fast, in some others slower.



Gauss-Markov (GM): $p(f_j|f_{j-1}) \propto \exp\left[-\gamma|f_j - f_{j-1}|^2\right]$



Generalized GM: $p(f_j|f_{j-1}) \propto \exp\left[-\gamma|f_j - f_{j-1}|^\beta\right]$

Fig. 2. Gauss-Markov and Generalized Gauss-Markov prior laws

For Ex05, we can use a Gauss-Markov prior law

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_{j=1}^{N} |f_j - f_{j-1}|^2\right] \tag{28}$$

where γ_0 fixes the rate of the dependencies.
For Ex06, we can use a Generalized Gauss-Markov prior law

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_{j=1}^{N} |f_j - f_{j-1}|^\beta\right] \tag{29}$$

where γ_0 fixes the rate of the dependencies and β_0 can be fixed from some knowledge about the distribution of the isolation materials.
We call this family of priors Simple Markovian priors where the general expression can be written as:

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_{j=1}^{N} \phi(|f_j - f_{j-1})\right] \tag{30}$$

with different expressions for the potential function $\phi(.)$.

3.4.3. *Simple Markovian priors for images*

Now, let consider some cases with images.

Ex05b $\boldsymbol{f}$ represents the pixel values of an image: $\boldsymbol{f} = \{f(\boldsymbol{r}), \boldsymbol{r} = (x,y) \in \mathcal{R}\}$, where $\mathcal{R}$ represents the surface of the image. $f(\boldsymbol{r})$ represents for example the temperature at the position $\boldsymbol{r} = (x,y)$. We know that the temperature at that position is not independent of the its neighbors positions $\boldsymbol{r}' \in \mathcal{N}(\boldsymbol{r})$.

Ex06b This is the image version of Ex06.

For Ex05b, we can use a Gauss-Markov prior law

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} |f(\boldsymbol{r}) - f(\boldsymbol{r}')|^2\right] \quad (31)$$

where γ_0 fixes the rate of the dependencies.
For Ex06b, we can use a Generalized Gauss-Markov prior law

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{N}(\boldsymbol{r})} |f(\boldsymbol{r}) - f(\boldsymbol{r}')|^\beta\right] \quad (32)$$

3.4.4. Hierarchical priors with hidden variables

Let now consider other examples.

Ex07 The signal we are looking for represents the reflection coefficient (for example inside a well in geophysical applications). So, its values are very often zero. When it is not zero, it can be positive or negative but not very far from zero.

Ex08 The signal we are looking for is a spectrum (the distribution of energies concentrated in some frequencies). Its values are very often zero and when not equal to zero, it is always positive.

For Ex07 we can use a Bernoulli-Gaussian model

$$\begin{cases} p(f_j|q_j) \propto \exp\left[-\frac{1}{2v_0}(1-q_j)|f_j|^2\right] \\ p(q_j = 1) = \alpha, \quad p(q_j = 0) = 1 - \alpha \end{cases} \quad (33)$$

which gives:

$$\begin{cases} p(\boldsymbol{f}|\boldsymbol{q}) \propto \exp\left[-\frac{1}{2v_0}\sum_{j=1}^N (1-q_j)|f_j|^2\right] \\ p(\boldsymbol{q}) \propto \alpha^{\sum_{j=1}^N \delta(q_j)}(1-\alpha)^{\sum_{j=1}^N \delta(1-q_j)}, \end{cases} \quad (34)$$

where $\sum_{j=1}^N \delta(q_j) = n_1$ is the number of ones and $\sum_{j=1}^N \delta(1-q_j) = n_0 = N - n_1$ is the number of zeros in the Bernoulli sequence $\boldsymbol{q} = [q_1, \cdots, q_N]'$.
For Ex08 we can use a Bernoulli-Gamma model:

$$\begin{cases} p(f_j|q_j) \propto \exp\left[-(1-q_j)(\alpha_0 \ln f_j + \beta_0 f_j)\right] \\ p(q_j = 1) = \alpha, \quad p(q_j = 0) = 1 - \alpha \end{cases} \quad (35)$$

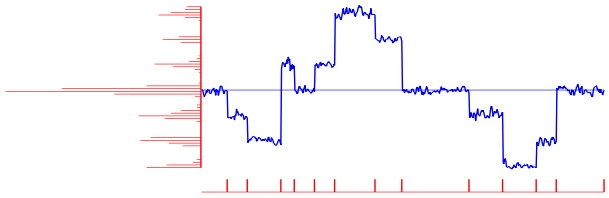
The Bernoulli variable q_j can be considered as a binary valued hidden variable. Other models for both $p(\boldsymbol{f}|\boldsymbol{q})$ and $p(\boldsymbol{q})$ are possible. For example a Gauss-Markov-Bernoulli model:

$$\begin{cases} p(\boldsymbol{f}|\boldsymbol{q}) & \propto \exp\left[-\frac{1}{2v_0}\sum_{j=1}^N (1-q_j)|f_j - f_{j-1}|^2\right] \\ p(\boldsymbol{q}) & \propto \alpha^{\sum_{j=1}^N \delta(q_j)}(1-\alpha)^{\sum_{j=1}^N \delta(1-q_j)}, \end{cases} \quad (36)$$

which is also called Piecewise Gaussian Model (PWG). Another example a Gauss-Markov-Ising Model (GMIM):

$$\begin{cases} p(\boldsymbol{f}|\boldsymbol{q}) & \propto \exp\left[-\frac{1}{2v_0}\sum_{j=1}^N (1-q_j)|f_j - f_{j-1}|^2\right] \\ p(\boldsymbol{q}) & \propto \exp\left[\gamma_0 \delta(q_j - q_{j-1})\right] \end{cases} \quad (37)$$

The final example we consider here is the Gauss-Markov-Potts Model (GMPM):



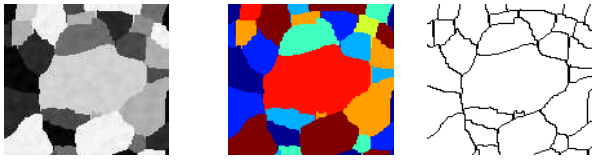
Piecewise Gaussians (contours hidden variables)
$p(f_j|q_j, f_{j-1}) = \mathcal{N}\left((1-q_j)f_{j-1}, \sigma_f^2\right)$

$p(f_j|z_j = k) = \mathcal{N}\left(m_k, \sigma_k^2\right)$ & z_j Markovian
Mixture of Gaussians (regions labels hidden variables)

Fig. 3. Piecewise Gaussian and Gauss-Markov-Potts for 1D signals

4. GAUSS-MARKOV-POTTS PRIOR MODELS FOR IMAGES

The two last prior models have their most significance in image processing where the contours and regions are naturally introduced via the hidden variables $q(\boldsymbol{r})$ representing contours and $z(\boldsymbol{r})$ representing the labels of the regions.



$f(\boldsymbol{r})$ $z(\boldsymbol{r})$ $q(\boldsymbol{r})$

Fig. 4. An image $f(\boldsymbol{r})$, its region labels $z(\boldsymbol{r})$ and its contours $q(\boldsymbol{r})$.

The Gauss-Markov-Potts model take its real importance in image segmentation and in inverse problems of imaging systems in particular in Non Destructive Testing (NDT) systems where we know that the object under test is composed of a finite set of K of homogeneous materials. Thus, the image we are looking for is composed of homogeneous compact regions. Translating this prior knowledge in a probability model can be done very easily through the following:

$$p(f(\boldsymbol{r})|z(\boldsymbol{r}) = k, m_k, v_k) = \mathcal{N}(m_k, v_k) \quad (38)$$

which results to a Mixture of Gaussians model for the intensities $f(\boldsymbol{r})$:

$$p(f(\boldsymbol{r})) = \sum_k P(z(\boldsymbol{r}) = k)\,\mathcal{N}(m_k, v_k) \quad (39)$$

For the hidden variables $z(\boldsymbol{r})$ we have two options:

- Separable iid hidden variables: $p(\boldsymbol{z}) = \prod_{\boldsymbol{r}} p(z(\boldsymbol{r}))$
- Markovian hidden variables: $p(\boldsymbol{z})$ Potts-Markov: where

$$p(\boldsymbol{z}) \propto \exp\left[\gamma \sum_{\boldsymbol{r}\in\mathcal{R}} \sum_{\boldsymbol{r}'\in\mathcal{V}(\boldsymbol{r})} \delta(z(\boldsymbol{r}) - z(\boldsymbol{r}'))\right] \tag{40}$$

is the Potts-Markov model.

4.1. Summarizing families of prior laws

In general, we can distinguish three great classes of priors:

- Simple separable priors: The general form is

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \sum_{j=1}^{N} \phi(f_j)\right] \tag{41}$$

  where $\phi(x)$ are, in general, positive functions, for example
  - $\phi(x) = x^2$ which gives the Gaussian prior
  - $\phi(x) = |x|^\beta$ with $0 < \beta < 2$ which gives the Generalized Gaussian prior
  - $\phi(x) = \alpha \ln x + \beta x$ with $x > 0$ and $\alpha > 0, \beta > 0$ which gives the Gamma prior
  - $\phi(x) = \alpha \ln x + \beta \ln(1-x)$ with $0 < x < 1$ and $\alpha > 0, \beta > 0$ which gives the Beta prior.
- Simple Markovian priors: The general form is

$$p(\boldsymbol{f}) \propto \exp\left[-\gamma \Omega(\boldsymbol{f})\right] \tag{42}$$

  where $\Omega(\boldsymbol{f}) = \sum_{j=1}^{N} \sum_{i\in\mathcal{V}(j)} \phi(f_j, f_i)$ where $\mathcal{V}(j)$ represents the neighboring sites (samples in signals, pixels in images) of j. The positive function $\phi(.)$ is called potential function and $\Omega(\boldsymbol{f})$ the total energy.
- Hierarchical priors: Very often, in particular for non stationary signals or non homogeneous images, we may use hidden variables z_j which can be associated to any sample f_j to let define in a hierarchical way $p(f_j|z_j)\,p(z_j)$ or $p(\boldsymbol{f}|\boldsymbol{z})\,p(\boldsymbol{z})$. As an example, we consider:

$$\begin{cases} p(f_j|z_j) = \mathcal{N}(f_j|0, z_j) \rightarrow p(\boldsymbol{f}|\boldsymbol{z}) = \prod_j p(f_j|z_j) \\ p(z_j) = \mathcal{IG}(z_j|\alpha, \beta) \quad \rightarrow p(\boldsymbol{z}) = \prod_j p(z_j) \end{cases} \tag{43}$$

4.2. Bayesian estimation with simple priors

The Bayesian inference approach is based on the posterior law:

$$p(\boldsymbol{f}|\boldsymbol{g}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{\theta}_2)}{p(\boldsymbol{g}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{\theta}_2) \tag{44}$$

where the sign $\propto$ stands for "proportional to", $p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)$ is the likelihood, $p(\boldsymbol{f}|\boldsymbol{\theta}_2)$ the prior model, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are their corresponding parameters (often called the hyper-parameters of the problem) and $p(\boldsymbol{g}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is called the evidence of the model. This simple Bayesian approach processing is showed in the following scheme:

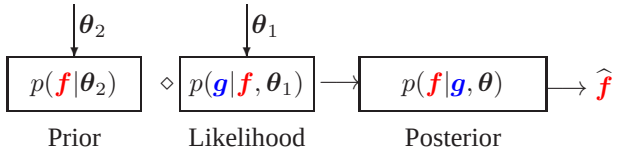


Fig. 5. Bayesian inference with simple priors

When both the likelihood and the prior are Gaussian, the posterior is also Gaussian and all the computations can be done analytically. This case is summarized in the following scheme:

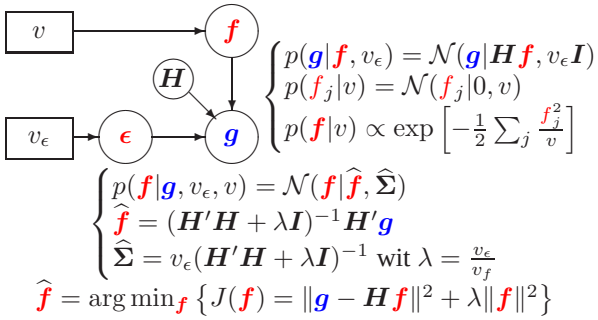


Fig. 6. Bayesian inference with simple priors

5. FULL BAYESIAN ESTIMATION WITH SIMPLE PRIORS

5.1. Joint posterior law

When the parameters $\boldsymbol{\theta}$ have to be estimated too, a prior $p(\boldsymbol{\theta}|\boldsymbol{\phi}_0)$ with fixed values for $\boldsymbol{\phi}_0$ is assigned to them and the expression of the joint posterior

$$p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}, \boldsymbol{\phi}_0) = \frac{p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\,p(\boldsymbol{f}|\boldsymbol{\theta}_2)\,p(\boldsymbol{\theta}|\boldsymbol{\phi}_0)}{p(\boldsymbol{g}|\boldsymbol{\phi}_0)} \tag{45}$$

is used to infer them jointly. This method is summarized in the following scheme:

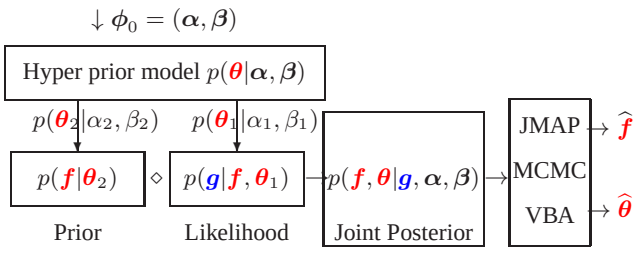


Fig. 7. Full Bayesian inference with simple priors

From the joint posterior, classically, three methods have been proposed: Joint Maximum A Posteriori (JMAP), MCMC methods, Marginalization and Expectation-Maximization (EM) methods which can all be considered as special cases

of Bayesian Variational Approximation (BVA) method [14, 15, 16, 17, 18].

5.2. Joint Maximum A Posteriori (JMAP)

The JMAP solution is defined as:

$$(\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\theta}}) = \arg\max_{(\boldsymbol{f},\boldsymbol{\theta})} \{p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}, \phi_0)\} \tag{46}$$

and one way to obtain it is an alternate optimization:

$$\left\{ \begin{array}{ll} \widehat{\boldsymbol{f}} & = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}, \widehat{\boldsymbol{\theta}}|\boldsymbol{g}, \phi_0)\right\} \\ \widehat{\boldsymbol{\theta}} & = \arg\max_{\boldsymbol{\theta}} \left\{p(\widehat{\boldsymbol{f}}, \boldsymbol{\theta}|\boldsymbol{g}, \phi_0)\right\} \end{array} \right. \tag{47}$$

5.3. MCMC

The main idea and objective of the MCMC methods are the exploration of the space of the solution by generating samples from the posterior law and thus being able to compute empirically the expected values $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{f}}$ of the unknowns. In general, Gibbs sampling method is used to successively sample from the conditionals $p(\boldsymbol{f}|\widehat{\boldsymbol{\theta}}, \boldsymbol{g}, \phi_0)$ and $p(\widehat{\boldsymbol{\theta}}|\boldsymbol{f}, \boldsymbol{g}, \phi_0)$. The main difficulties are:
- Convergence and great number of iterations needed
- Cost of the computations particularly in inverse problems.
The interested readers can refer to [9, 19, 20]

5.4. Marginalization and Expectation-Maximization (EM)

The main idea here is, first focus on the estimation of the hyper parameters $\boldsymbol{\theta}$ by marginalizing over $\boldsymbol{f}$:

$$p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{g}, \phi_0) = \int p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}, \phi_0)\, \mathrm{d}\boldsymbol{f}, \tag{48}$$

then estimating $\boldsymbol{\theta}$ by

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{g}, \phi_0)\} \tag{49}$$

The estimated value $\widehat{\boldsymbol{\theta}}$ can then be used for the estimation of $\boldsymbol{f}$.

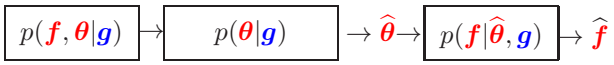


Fig. 8. Marginalization for estimation of hyper parameters.

The main difficulty here is that, in general, an analytical expression for $p(\boldsymbol{\theta}|\boldsymbol{f}, \boldsymbol{g}, \phi_0)$ can not be obtained. The Expectation-Maximization (EM) algorithm is an iterative technical to compute $\widehat{\boldsymbol{\theta}}$. As we will see in the below, all these methods can be considered as particular cases of the Bayesian Variational Approximation (BVA) methods, well known in statistical physics, but recently used for inverse problems.

6. BAYESIAN VARIATIONAL APPROXIMATION (BVA)

6.1. BVA basics

As we could see, either we have to do computations with the simple posterior $p(\boldsymbol{f}|\boldsymbol{g})$ or with the joint posterior $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ when the hyper parameters are not known, or as we will see later with $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ when we have to infer on the unknown of the interest $\boldsymbol{f}$, the hidden variables $\boldsymbol{z}$ and the hyper parameters $\boldsymbol{\theta}$. In all these cases, doing Bayesian computation (Optimization in MAP and JMAP or integration when posterior means are needed) may be very costly. The main idea behind BVA is to approximate these posterior laws by simpler ones, for example:
$p(\boldsymbol{f}|\boldsymbol{g})$ by $q(\boldsymbol{f}) = \prod_j q_j(f_j)$ or
$p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by $q(\boldsymbol{f}, \boldsymbol{\theta}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{\theta})$ or
$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ by $q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}) = q_1(\boldsymbol{f})\, q_2(\boldsymbol{z})\, q_3(\boldsymbol{z})$.

The main advantage then is to be able to do the computations much faster. However, these approximations have to be done using a criterion. The main criterion used is using the Kullback-Leibler divergence:

$$\mathrm{KL}(q : p) = \int q \ln \frac{q}{p} \tag{50}$$

which can be considered as a kind of differential geometry projection of p over a particular space $\mathcal{Q}$ of some parametric or nonparametric manifold of probability spaces. When $\mathcal{Q}$ is choosed to be the space of separable probability laws q_j, the approach is called Mean Field theory.

To illustrate the basic ideas and tools, let consider a random vector $\boldsymbol{X}$ and its probability density function $p(\boldsymbol{x})$ that we want to approximate by $q(\boldsymbol{x}) = \prod_j q_j(x_j)$. Using the KL criterion:

$$\begin{aligned} \mathrm{KL}(q : p) &= \int q(\boldsymbol{x}) \ln \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})}\, \mathrm{d}\boldsymbol{x} \\ &= \int q(\boldsymbol{x}) \ln q(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} - \int q(\boldsymbol{x}) \ln p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \\ &= \sum_j \int q_j(x_j) \ln q_j(x_j)\, \mathrm{d}x_j - \langle \ln p(\boldsymbol{x}) \rangle_q \\ &= \sum_j \int q_j(x_j) \ln q_j(x_j)\, \mathrm{d}x_j \\ &\qquad - \int q_j(x_j) < \ln p(\boldsymbol{x}) >_{q_{-j}}\, \mathrm{d}x_j \end{aligned} \tag{51}$$

where we used the notation

$$\langle \ln p(\boldsymbol{x}) \rangle_q = \int q(\boldsymbol{x}) \ln p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{52}$$

and $q_{-j}(\boldsymbol{x}) = \prod_{i \neq j} q_i(x_i)$.

From here, trying to find the solution q_i, we can use the flowing alternate optimization algorithm:

$$q_j(x_j) \propto \exp\left[< \ln p(\boldsymbol{x}) >_{q_{-j}}\right] \tag{53}$$

In the case of two variables $\boldsymbol{x} = [x_1, x_2]'$, we have:

$$\left\{ \begin{array}{l} q_1(x_1) \propto \exp\left[< \ln p(\boldsymbol{x}) >_{q_2(x_2)}\right] \\ q_2(x_2) \propto \exp\left[< \ln p(\boldsymbol{x}) >_{q_1(x_1)}\right] \end{array} \right. \tag{54}$$

Three different algorithms can be obtained depending on the choice of a particular family for $q_j(x_j)$:

- $q_1(x_1) = \delta(x_1 - \tilde{x}_1)$ and $q_2(x_2) = \delta(x_2 - \tilde{x}_2)$

$$\left\{\begin{array}{l} q_1(x_1) \propto p(x_1, x_2 = \tilde{x}_2) \\ q_2(x_2) \propto p(x_1 = \tilde{x}_1, x_2) \end{array}\right. \tag{55}$$

which becomes equivalent to JMAP:

$$(\widehat{x}_1, \widehat{x}_2) = \arg \max_{(x_1,x_2)} \{p(x_1, x_2)\} \tag{56}$$

by the following alternate optimization algorithm:

$$\left\{\begin{array}{l} \tilde{x}_1 = \arg\max_{x_1} \{p(x_1, x_2 = \tilde{x}_2)\} \\ \tilde{x}_2 = \arg\max_{x_2} \{p(x_1 = \tilde{x}_1, x_2)\} \end{array}\right. \tag{57}$$

The main drawback here is that the uncertainties of the x_1 is not used for the estimation of x_2 and the uncertainties of x_2 is not used for the estimation of x_1.

- $q_1(x_1) = \delta(x_1 - \tilde{x}_1)$ and $q_2(x_2)$ free form. In the same way, this time we obtain:

$$\left\{\begin{array}{l} Q(x_1, x_2 = \tilde{x}_2) =< \ln p(x_1 = \tilde{x}_1, x_2) >_{q_2(x_2)} \\ \tilde{x}_2 = \arg\max_{x_2} \{Q(x_1 = \tilde{x}_1, x_2)\} \end{array}\right. \tag{58}$$

which can be compared with the classical EM algorithm. Here, the uncertainties of the x_1 is used for the estimation of x_2 but the uncertainties of x_2 is not used for the estimation of x_1.

- both $q_1(x_1)$ and $q_2(x_2)$ have free form. The main difficulty here is that, at each iteration the expression of q_1 and q_2 may change. However, if $p(x_1, x_2)$ is in a generalized exponential family, the expressions of $q_1(x_1)$ and $q_2(x_2)$ will also be in the same family and we have only to update the parameters at each iteration. For some extensions and more details see [21].

6.2. BVA with simple prior models and hyper parameter estimation

Variational Bayesian Approximation (BVA) methods try to approximate $p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g})$ by a separable one $q(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}}, \boldsymbol{g})\ q_2(\boldsymbol{\theta}|\widetilde{\boldsymbol{f}}, \boldsymbol{g})$ and then using them for estimation [22, 23, 24, 25, 26, 27, 28, 29, 30].

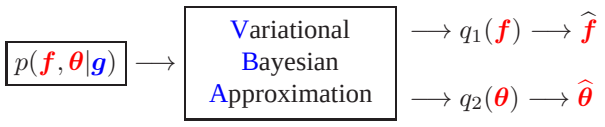


Fig. 9. BVA for the estimation of hyper parameters.

As we have seen it in previous section, different choices for the family of laws q_1 and q_2 result in different algorithms:

- Case 1 : $\longrightarrow$ Joint MAP

$$\left\{\begin{array}{l} \widehat{q}_1(\boldsymbol{f}|\widetilde{\boldsymbol{f}}) = \delta(\boldsymbol{f} - \widetilde{\boldsymbol{f}}) \\ \widehat{q}_2(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}) = \delta(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}) \end{array}\right. \rightarrow \left\{\begin{array}{l} \widetilde{\boldsymbol{f}} = \arg\max_{\boldsymbol{f}} \left\{p(\boldsymbol{f}, \widetilde{\boldsymbol{\theta}}|\boldsymbol{g})\right\} \\ \widetilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{p(\widetilde{\boldsymbol{f}}, \boldsymbol{\theta}|\boldsymbol{g})\right\} \end{array}\right. \tag{59}$$

- Case 2 : $\longrightarrow$ Bayesian EM

$$\left\{\begin{array}{l} \widehat{q}_1(\boldsymbol{f}) \propto p(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}}, \boldsymbol{g}) \\ \widehat{q}_2(\boldsymbol{\theta}|\widetilde{\boldsymbol{\theta}}) = \delta(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}) \end{array}\right. \rightarrow \left\{\begin{array}{l} Q(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}}) = \langle \ln p(\boldsymbol{f}, \boldsymbol{\theta}|\boldsymbol{g}) \rangle_{q_1(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}})} \\ \widetilde{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{Q(\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}})\right\} \end{array}\right. \tag{60}$$

- Case 3: Appropriate choice for inverse problems

$$\left\{\begin{array}{l} \widehat{q}_1(\boldsymbol{f}) \propto p(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \widetilde{\boldsymbol{\theta}})\, p(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}}) \\ \widehat{q}_2(\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\widetilde{\boldsymbol{f}}, \boldsymbol{g}) \propto p(\boldsymbol{g}|\widetilde{\boldsymbol{f}}, \boldsymbol{\theta})\, p(\boldsymbol{\theta}) \end{array}\right. \tag{61}$$

with appropriate choice of conjugate priors for $p(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}})$ and $p(\boldsymbol{\theta})$ the expressions of $\widehat{q}_1(\boldsymbol{f})$ will be in the same family as $p(\boldsymbol{f}|\widetilde{\boldsymbol{\theta}})$ and $\widehat{q}_2(\boldsymbol{\theta})$ will be in the same family as $p(\boldsymbol{\theta})$. Then, these iterations just become those of updating the parameters.

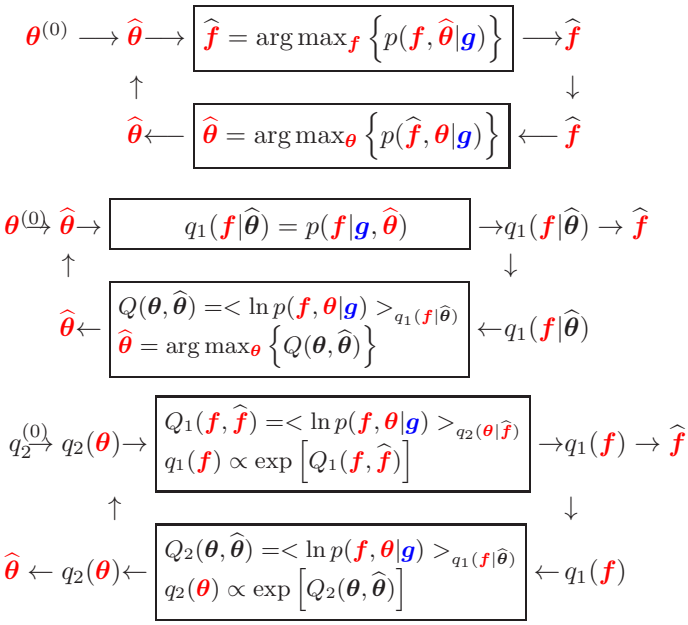


Fig. 10. Comparison between JMAP, EM and BVA.

To illustrate the differences between these three cases, we consider the following model:

$$\begin{array}{l} p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon) = \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, v_\epsilon \boldsymbol{I}) \propto \exp\left[-\frac{1}{2v_\epsilon}\|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2\right] \\ p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \\ p(f_j|v_j) = \mathcal{N}(f_j|0, v_j) \propto \exp\left[-\frac{1}{2}\frac{f_j^2}{v_j}\right] \\ p(\boldsymbol{f}|\boldsymbol{v}) = \mathcal{N}(\boldsymbol{f}|0, \text{diag}\,[v_1, \cdots, v_N]) \propto \exp\left[-\frac{1}{2}\sum_j \frac{f_j^2}{v_j}\right] \\ p(v_j|\alpha_0, \beta_0) = \mathcal{IG}(v_j|\alpha_0, \beta_0) \end{array} \tag{62}$$

which is illustrated in the following graphical scheme:

It is then easy to show the following relations:

$$\begin{array}{ll} p(\boldsymbol{f}, \boldsymbol{v}, v_\epsilon|\boldsymbol{g}) & \propto \mathcal{N}(\boldsymbol{g}|\boldsymbol{H}\boldsymbol{f}, v_\epsilon \boldsymbol{I})\mathcal{N}(\boldsymbol{f}|0, \text{diag}\,[\boldsymbol{v}]) \\ & \quad \mathcal{IG}(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0})\mathcal{IG}(v_j|\alpha_0, \beta_0) \end{array} \tag{63}$$

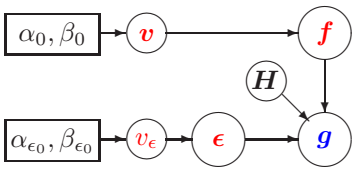


Fig. 11. Graphical model with Gaussian priors and hyper parameter estimation.

$$\begin{cases} p(\boldsymbol{f}|\boldsymbol{g}, v_\epsilon, v) = \mathcal{N}(\boldsymbol{f}|\widehat{\boldsymbol{f}}, \widehat{\boldsymbol{\Sigma}}) \\ \widehat{\boldsymbol{\Sigma}} = (\boldsymbol{H}'\boldsymbol{H} + \hat{v}_\epsilon \boldsymbol{V})^{-1} \text{ with } \boldsymbol{V} = \text{diag}\,[\hat{v}_j] \\ \widehat{\boldsymbol{f}} = \widehat{\boldsymbol{\Sigma}}\boldsymbol{H}'\boldsymbol{g} \\ p(v_\epsilon|\boldsymbol{g}, \boldsymbol{f}, \alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(v_\epsilon|\widehat{\alpha}_{\epsilon_0}, \widehat{\beta}_{\epsilon_0}) \\ p(v_j|\boldsymbol{g}, \boldsymbol{f}, \alpha_0, \beta_0) = \mathcal{IG}(v_j|\widehat{\alpha}_j, \widehat{\beta}_j) \\ \hat{v}_\epsilon = \frac{\widehat{\alpha}_{\epsilon_0}}{\widehat{\beta}_{\epsilon_0}} \\ \hat{v}_j = \frac{\widehat{\alpha}_j}{\widehat{\beta}_j} \end{cases} \tag{64}$$

$$\widehat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \left\{ J(\boldsymbol{f}) = \|\boldsymbol{g} - \boldsymbol{H}\boldsymbol{f}\|^2 + \langle v_\epsilon \rangle_q \sum_j \frac{f_j^2}{\langle v_j \rangle_q} \right\} \tag{65}$$

It is also easy to compute q_1 and q_2 in the VBA approximation. The following figure summarizes and compare JMAP and VBA.

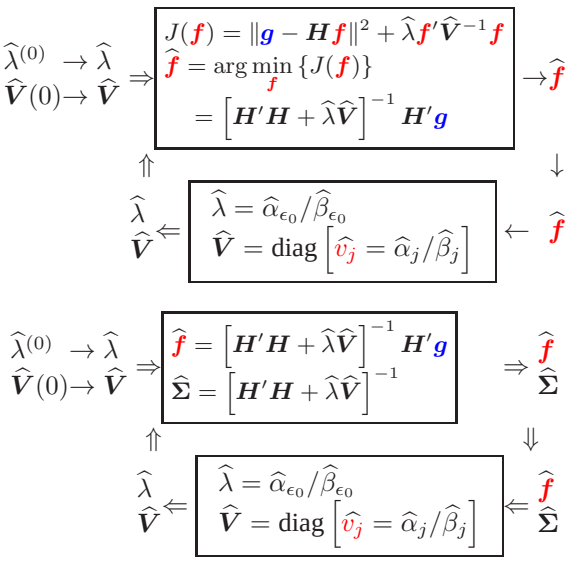


Fig. 12. Comparison between JMAP and VBA.

Two main differences are:

- In JMAP, the uncertainties of $\widehat{\boldsymbol{f}}$ ($\widehat{\boldsymbol{\theta}}$) are not transmitted to the estimation of $\widehat{\boldsymbol{\theta}}$ ($\widehat{\boldsymbol{f}}$). However, here, there is no need to compute the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ which costs a lot computational. In this case we have:

$$\begin{cases} \widehat{\alpha}_{\epsilon_0} = \alpha_{\epsilon_0} + M/2 \\ \widehat{\beta}_{\epsilon_0} = \beta_{\epsilon_0} + \frac{1}{2}\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}\|^2, \\ \widehat{\alpha}_j = \alpha_0 + N/2 \\ \widehat{\beta}_j = \beta_0 + \frac{1}{2}\|\widehat{\boldsymbol{f}}\|^2 \end{cases} \tag{66}$$

- In VBA the uncertainties $\widehat{\boldsymbol{\Sigma}}$ of $\widehat{\boldsymbol{f}}$ are transmitted to the estimation of $\widehat{\boldsymbol{\theta}}$ ($\widehat{\boldsymbol{f}}$). However, here, we have to compute this posterior covariance matrix $\widehat{\boldsymbol{\Sigma}}$ which costs a lot computational. In this case we have:

$$\begin{cases} \widehat{\alpha}_{\epsilon_0} = \alpha_{\epsilon_0} + M/2 \\ \widehat{\beta}_{\epsilon_0} = \beta_{\epsilon_0} + \frac{1}{2}\|\boldsymbol{g} - \boldsymbol{H}\widehat{\boldsymbol{f}}\|^2 + \text{Tr}\left\{\boldsymbol{H}'\boldsymbol{H}\right\}, \\ \widehat{\alpha}_j = \alpha_0 + N/2 \\ \widehat{\beta}_j = \beta_0 + \frac{1}{2}\|\widehat{\boldsymbol{f}}\|^2 + \frac{1}{2}\text{Tr}\left\{\widehat{\boldsymbol{\Sigma}}\right\} + \widehat{\boldsymbol{f}}'\widehat{\boldsymbol{f}} \end{cases} \tag{67}$$

6.3. BVA with hierarchical prior models

For hierarchical prior models with hidden variables $\boldsymbol{z}$, the problem becomes more complex, because we have to give the expression of the joint posterior law

$$p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) \propto p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta}_1)\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)\, p(\boldsymbol{z}|\boldsymbol{\theta}_3)\, p(\boldsymbol{\theta}) \tag{68}$$

and then approximate it by a separable one

$$q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\boldsymbol{g})\, q_2(\boldsymbol{z}|\boldsymbol{g})\, q_3(\boldsymbol{\theta}|\boldsymbol{g}) \tag{69}$$

and where the expressions of $q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{g})$ is obtained by minimizing the Kullback-Leibler divergence

$$\text{KL}(q : p) = \int q \ln \frac{q}{p} = \left\langle \ln \frac{q}{p} \right\rangle_q \tag{70}$$

It is then easy to show that $\text{KL}(q : p) = \ln p(\boldsymbol{g}) - \mathcal{F}(q)$ where $p(\boldsymbol{g}|\mathcal{M})$ is the likelihood of the model

$$p(\boldsymbol{g}) = \int\int\int p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\, \text{d}\boldsymbol{f}\, \text{d}\boldsymbol{z}\, \text{d}\boldsymbol{\theta} \tag{71}$$

with $p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g}) = p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{\theta})\, p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta})\, p(\boldsymbol{z}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$ and $\mathcal{F}(q)$ is the free energy associated to q defined as

$$\mathcal{F}(q) = \left\langle \ln \frac{p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})}{q(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta})} \right\rangle_q \tag{72}$$

So, for a given model, minimizing $\text{KL}(q : p)$ is equivalent to maximizing $\mathcal{F}(q)$ and when optimized, $\mathcal{F}(q^*)$ gives a lower bound for $\ln p(\boldsymbol{g})$.

Without any other constraint than the normalization of q, an alternate optimization of $\mathcal{F}(q)$ with respect to q_1, q_2 and q_3 results in

$$\begin{cases} q_1(\boldsymbol{f}) \propto \exp\left[-\langle \ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\rangle_{q_2(\boldsymbol{z})q_3(\boldsymbol{\theta})}\right], \\ q_2(\boldsymbol{z}) \propto \exp\left[-\langle \ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\rangle_{q_1(\boldsymbol{f})q_3(\boldsymbol{\theta})}\right] \\ q_3(\boldsymbol{\theta}) \propto \exp\left[-\langle \ln p(\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{g})\rangle_{q_1(\boldsymbol{f})q_2(\boldsymbol{z})}\right] \end{cases} \tag{73}$$

Note that these relations represent an implicit solution for $q_1(\boldsymbol{f})$, $q_2(\boldsymbol{z})$ and $q_3(\boldsymbol{\theta})$ which need, at each iteration, the expression of the expectations in the right hand of exponentials. If $p(\boldsymbol{g}|\boldsymbol{f}, \boldsymbol{z}, \boldsymbol{\theta}_1)$ is a member of an exponential family and if all the priors $p(\boldsymbol{f}|\boldsymbol{z}, \boldsymbol{\theta}_2)$, $p(\boldsymbol{z}|\boldsymbol{\theta}_3)$, $p(\boldsymbol{\theta}_1)$, $p(\boldsymbol{\theta}_2)$, and $p(\boldsymbol{\theta}_3)$ are conjugate priors, then it is easy to see that these expressions leads to standard distributions

for which the required expectations are easily evaluated. In that case, we may note

$$q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = q_1(\boldsymbol{f}|\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g})\, q_2(\boldsymbol{z}|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g})\, q_3(\boldsymbol{\theta}|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}};\boldsymbol{g}) \tag{74}$$

where the tilded quantities $\widetilde{\boldsymbol{z}}$, $\widetilde{\boldsymbol{f}}$ and $\widetilde{\boldsymbol{\theta}}$ are, respectively functions of $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\theta}})$, $(\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}})$ and $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}})$ and where the alternate optimization results to alternate updating of the parameters $(\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}})$ for q_1, the parameters $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{\theta}})$ of q_2 and the parameters $(\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}})$ of q_3. Finally, we may note that, to monitor the convergence of the algorithm, we may evaluate the free energy

$$\begin{aligned}\mathcal{F}(q) &= \langle \ln p(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{g}|\mathcal{M})\rangle_q + \langle -\ln q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\rangle_q \\ &= \langle \ln p(\boldsymbol{g}|\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta})\rangle_q + \langle \ln p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{\theta})\rangle_q + \langle \ln p(\boldsymbol{z}|\boldsymbol{\theta})\rangle_q \\ &\quad + \langle -\ln q(\boldsymbol{f})\rangle_q + \langle -\ln q(\boldsymbol{z})\rangle_q + \langle -\ln q(\boldsymbol{\theta})\rangle_q\end{aligned} \tag{75}$$

where all the expectations are with respect to q.

Other decompositions are also possible:

$$\begin{aligned} q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = \quad & \textstyle\prod_j q_{1j}(f_j|\widetilde{\boldsymbol{f}}_{(-j)},\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g}) \\ & \textstyle\prod_j q_{2j}(z_j|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}}(-j),\widetilde{\boldsymbol{\theta}};\boldsymbol{g}) \\ & \textstyle\prod_l q_{3l}(\theta_l|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}}_{(-l)};\boldsymbol{g}) \end{aligned} \tag{76}$$

or

$$\begin{aligned} q(\boldsymbol{f},\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{g}) = \quad & q_1(\boldsymbol{f}|\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}};\boldsymbol{g}) \\ & \textstyle\prod_j q_{2j}(z_j|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}}_{(-j)},\widetilde{\boldsymbol{\theta}};\boldsymbol{g}) \\ & \textstyle\prod_l q_{3l}(\theta_l|\widetilde{\boldsymbol{f}},\widetilde{\boldsymbol{z}},\widetilde{\boldsymbol{\theta}}_{(-l)};\boldsymbol{g}) \end{aligned} \tag{77}$$

Here, we consider this case and give some more details on it.

7. BAYESIAN VARIATIONAL APPROXIMATION WITH MIXTURE OF GAUSSIANS PRIORS

The mixture models are very commonly used as prior models. These models are summarized in the following.

7.1. Mixture of Gaussians (MoG) simple model

First we consider the simplest case where the number K and the proportions $\boldsymbol{\alpha} = \{\alpha_k, k = 1\cdots,K\}$ are known.

$$\begin{aligned} & p(z_j = k|\alpha_k) = \alpha_k, \quad \textstyle\sum_k \alpha_k = 1 \\ & p(f_j|z_j = k) = \mathcal{N}(f_j|m_{j_k}, v_{j_k}) \\ & p(m_{j_k}|m_0, v_0) = \mathcal{N}(m_{j_k}|m_0, v_0) \\ & p(v_{j_k}|\alpha_0, \beta_0) = \mathcal{IG}(v_{j_k}|\alpha 0, \beta_0) \\ & p(\boldsymbol{f}|\boldsymbol{z},\boldsymbol{m},\boldsymbol{v}) = \textstyle\prod_j \mathcal{N}(f_j|m_{z_j}, v_{z_j}) \\ & p(\boldsymbol{z}|\boldsymbol{\alpha}) = \alpha_k^{n_k} \text{ with } n_k = \textstyle\sum_j \delta(m_{z_j} - m_k) \\ & p(\boldsymbol{g}|\boldsymbol{f}, v_\epsilon) = \mathcal{N}(\boldsymbol{H}\boldsymbol{f}, v_\epsilon \boldsymbol{I}) \\ & p(v_\epsilon|\alpha_{\epsilon_0}, \beta_{\epsilon_0}) = \mathcal{IG}(\alpha_{\epsilon_0}, \beta_{\epsilon_0}) \end{aligned} \tag{78}$$

If the proportions are not known, we have to add a prior to it. The appropriate prior is the Dirichlet prior

$$p(\boldsymbol{\alpha}|\alpha_0) \propto \alpha_k^{\alpha_0}, \text{ with } \alpha_0 = 1/K \tag{79}$$

With these priors, it is then easy to find the expressions for the joint posterior law, all the conditionals necessary for MCMC or all the separable laws for VBA. We refer the authors to [21, 31, 32] for the details.

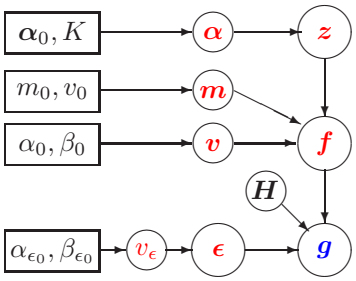
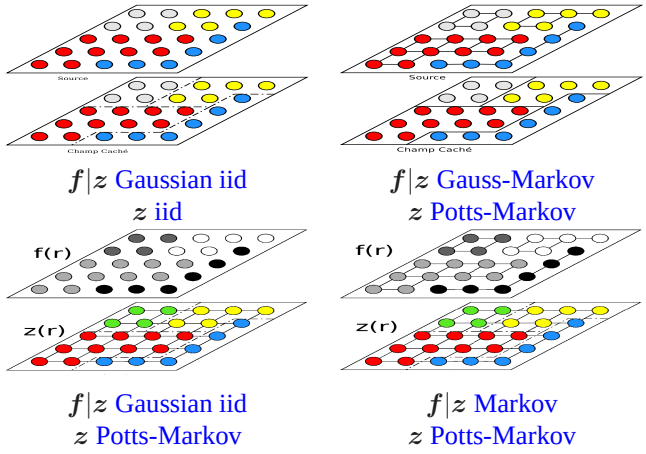


Fig. 13. Mixture of Gaussians prior model and its associated graphical model.

8. BAYESIAN VARIATIONAL APPROXIMATION WITH GAUSS-MARKOV-POTTS PRIORS

The main drawback of the MoG model of the previous section is that the spatial structure of the images is not considered. This can be done either by putting a Markovian model on $\boldsymbol{f}$ or on $\boldsymbol{z}$ or on both of them.

To summarize, with two variables $f(\boldsymbol{r})$ and $z(\boldsymbol{r})$, we can define four different models:



$\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid
$\boldsymbol{z}$ iid

$\boldsymbol{f}|\boldsymbol{z}$ Gauss-Markov
$\boldsymbol{z}$ Potts-Markov

$\boldsymbol{f}|\boldsymbol{z}$ Gaussian iid
$\boldsymbol{z}$ Potts-Markov

$\boldsymbol{f}|\boldsymbol{z}$ Markov
$\boldsymbol{z}$ Potts-Markov

Fig. 14. An image $f(\boldsymbol{r})$, its region labels $z(\boldsymbol{r})$ and its contours $q(\boldsymbol{r})$.

The first one is exactly the MoG of the previous section. The second one is a non homogeneous Markov model for $f(\boldsymbol{r})$ conditioned on $z(\boldsymbol{r})$. The third and the forth cases are of great interest. We called them Gauss-Markov-Potts prior models and used them extensively in different applications:

- Image segmentation and images fusion [33]
- Image restoration for NDT applications [34, 35]
- Computed Tomography (CT) for NDT applications [36, 37]
- Blind Sources Separation and Images separation [38, 39, 40, 41, 42, 43]
- Fourier Synthesis part of microwave imaging [44]
- Super Resolution Images [45, 46, 47]
- Microwave imaging for NDT [33, 48, 49]
- Optical Diffraction Tomography [50, 51]

- Synthetic Aperture Radar (SAR) imaging [52]
- Acoustical sources localization [53]

9. CONCLUSIONS

In this review paper, first the basics of the Bayesian estimation with different prior laws are presented. Then, the full Bayesian approach with hyper parameters estimation is considered. The different Bayesian computational approaches (JMAP, Marginalization and EM, MCMC and Variational Bayesian Approximation (VBA) are presented and compared. Focus is made more on the VBA method with hierarchical priors. A class of these hierarchical priors containing the Mixture of Gaussians (MoG) is considered. These priors are called Gauss-Markov-Potts. Finally, references on the successful use of these priors in different applications are given.

10. REFERENCES

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

[2] A. Mohammad-Djafari and G. Demoment, "Image restoration and reconstruction using entropy as a regularization functional," *Maximum Entropy and Bayesian Methods in Science and Engineering*, vol. 2, pp. 341–355, 1988.

[3] A. Mohammad-Djafari and G. Demoment, "Estimating priors in maximum entropy image processing," in *Proc. International Conference on Acoustics, Speech, and Signal Processing ICASSP-90*, pp. 2069–2072, 3–6 April 1990.

[4] A. Mohammad-Djafari and A. Mohammadpour, "On the estimation of a parameter with incomplete knowledge on a nuisance parameter," in *24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 735, pp. 533–540, AIP, 2004.

[5] A. Mohammad-Djafari, "Maximum likelihood estimation of the lagrange parameters of the maximum entropy distributions," *Maximum-entropy and Bayesian methods*, 1991.

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

[7] Zellner, "Maximal data information prior distributions," *in New developpements in the applications of bayesian methods, A. Aykac and C. Brumat éditeurs associés, North-Holland, Amsterdam*, pp. 211–232, 1977.

[8] S. Hill and J. Spall, "Shannon information–theoretic priors for state–space model parameter," in *Bayesian Analysis of Time Series and Dynamic Models*, pp. 509–524, (J. C. Spall, eds.), Marcel Deckker Inc., 1988.

[9] C. Robert, *L'analyse statistique bayésienne*. éditions économica, paris ed., 1987.

[10] E. Barat, C. Comtat, T. Dautremer, T. Montagu, M. D. Fall, A. Mohammad-Djafari, and R. Trébossen, "Nonparametric bayesian spatial reconstruction for positron emission tomography.," in *10th International meeting on fully three-dimensional image reconstruction in radiology and nuclear medecine*, (Beijing, China), 2009.

[11] M. D. Fall, E. Barat, A. Mohammad-Djafari, and C. Comtat, "Spatial emission tomography reconstruction using Pitman-Yor process," in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 1193, pp. 194–201, AIP, 2009.

[12] M. D. Fall, É. Barat, C. Comtat, T. Dautremer, T. Montagu, and A. Mohammad-Djafari, "A bayesian nonparametric model for dynamic (4d) pet," in *IEEE Medical Imaging Conference NSS/MIC.*, 2011.

[13] M. D. Fall, É. Barat, C. Comtat, T. Dautremer, T. Montagu, and A. Mohammad-Djafari, "A discrete-continuous bayesian model for emission tomography,," in *IEEE International Conference on Image Processing (ICIP)*, 2011.

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

[15] D. Rubin and D. Thayer, "EM algorithms for ML factor analysis," *Psychometrika*, vol. 47, no. 1, pp. 69–76, 1982.

[16] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of gibbs random fields using simulated annealing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-11, no. 8, pp. 799–813, 1989.

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

[18] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the em algorithm," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-36, no. 4, pp. 477–489, 1988.

[19] C. Robert, *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Verlag, 2007.

[20] C. P. Robert, "Mixtures of distributions: inference and estimation," *Markov chain Monte Carlo in practice*, vol. 441, p. 464, 1996.

[21] A. Mohammad-Djafari, "Approche variationnelle pour le calcul baysien dans les problmes inverses en imagerie," *Arxive*, vol. http://arxiv.org/abs/0904.4148, p. 31p, 2009.

[22] R. A. Choudrey, *Variational Methods for Bayesian Independent Component Analysis*. PhD thesis, University of Oxford, 2002.

[23] M. Beal, *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.

[24] A. C. Likas and N. P. Galatsanos., "A variational approach for bayesian blind image deconvolution," *IEEE Transactions on Signal Processing*, 2004.

[25] J. Winn, C. M. Bishop, and T. Jaakkola, "Variational message passing," *Journal of Machine Learning Research*, vol. 6, pp. 661–694, 2005.

[26] S. Chatzis and T. Varvarigou, "Factor analysis latent subspace modeling and robust fuzzy clustering using t-distributionsclassification of binary random patterns," *IEEE Trans. on Fuzzy Systems*, vol. 17, pp. 505–517, 2009.

[27] T. Park and G. Casella., "The Bayesian Lasso," *Journal of the American Statistical Association*, 2008.

[28] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, 2001.

[29] L. He, H. Chen, and L. Carin, "Tree-Structured Compressive Sensing With Variational Bayesian Analysis," *IEEE Signal. Proc. Let.*, vol. 17, no. 3, pp. 233–236, 2010.

[30] A. Fraysse and T. Rodet, "A gradient-like variational Bayesian algorithm," in *SSP 2011*, no. S17.5, (Nice, France), pp. 605–608, jun 2011.

[31] H. Ayasso, B. Duchêne, and A. Mohammad-Djafari, "A Bayesian approach to microwave imaging in a 3-D configuration," in *Proceeding of The 10th Workshop on Optimization and Inverse Problems in Electromagnetism*, (Ilmenau Allemagne), pp. 180–182, Spetember 2008.

[32] H. Ayasso and A. Mohammad-Djafari, "Joint NDT image restoration and segmentation using Gauss–Markov–Potts prior models and variational bayesian computation," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2265–2277, 2010.

[33] O. Féron and A. Mohammad-Djafari, "Image fusion and joint segmentation using an MCMC algorithm," *Journal of Electronic Imaging*, vol. 14, p. paper no. 023014, Apr 2005.

[34] H. Ayasso and A. Mohammad-Djafari, "Joint image restoration and segmentation using Gauss-Markov-Potts prior models and variational bayesian computation," in *Proceeding of the 15th IEEE International Conference on Image Processing, (ICIP)*, (Égypte), pp. 1297–1300, 2009.

[35] H. Ayasso and A. Mohammad-Djafari, "Variational Bayes with Gauss-Markov-Potts prior models for joint image restoration and segmentation.," in *proceedings of The International Conference on Computer Vision Theory and Applications (VISAPP) (VISAPP)*, (Funchal, Madeira Portugal), pp. 571–576, 2008.

[36] S. Fékih-Salem, A. Vabre, and A. Mohammad-Djafari, "Bayesian tomographic reconstruction of microsystems," in *Bayesian Inference and Maximum Entropy Methods, AIP Conf. Proc. 954* (K. et al. Knuth, ed.), pp. 372–380, MaxEnt Workshops, American Institute of Physics, July 2007.

[37] H. Ayasso, S. Fkih-Salem, and A. Mohammad-Djafari, "Variational Bayes approach for tomographic reconstruction," in *Proceedings of the 28th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, MaxEnt*, vol. 1073, (Sao Paulo Brésil), pp. 243–251, November 2008.

[38] H. Snoussi and A. Mohammad-Djafari, "Unsupervised learning for source separation with mixture of gaussians prior for sources and gaussian prior for mixture coefficients," in *Proc. IEEE Signal Processing Society Workshop Neural Networks for Signal Processing XI*, pp. 293–302, 10–12 Sept. 2001.

[39] H. Snoussi and A. Mohammad-Djafari, "Penalized maximum likelihood for multivariate gaussian mixture," in *Bayesian Inference and Maximum Entropy Methods* (R. L. Fry, ed.), pp. 36–46, MaxEnt Workshops, American Institute of Physics, Aug. 2002.

[40] H. Snoussi and A. Mohammad-Djafari, "Bayesian separation of HMM sources," in *Bayesian Inference and Maximum Entropy Methods* (R. L. Fry, ed.), pp. 77–88, MaxEnt Workshops, American Institute of Physics, Aug. 2002.

[41] H. Snoussi and A. Mohammad-Djafari, "Separation of mixed hidden Markov model sources," in *Bayesian Inference and Maximum Entropy Methods* (R. L. Fry, ed.), MaxEnt Workshops, American Institute of Physics, Aug. 2002.

[42] H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images," *Journal of Electronic Imaging*, vol. 13, pp. 349–361, April 2004.

[43] H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of Gaussians prior," *Journal of VLSI Signal Processing Systems*, vol. 37, pp. 263–279, June/July 2004.

[44] O. Féron, Z. Chama, and A. Mohammad-Djafari, "Reconstruction of piecewise homogeneous images from partial knowledge of their Fourier transform," in *MaxEnt04* (G. Erickson and Y. Zhai, eds.), (Garching, Germany), American Institute of Physics, august 2004.

[45] F. Humblot, *Détection de petits objets dans une image en utilisant les techniques de super-résolution*. Thèse, Université de Paris–Sud, Orsay, France, Sept. 2005.

[46] F. Humblot and A. Mohammad-Djafari, "Super-resolution and joint segmentation in bayesian framework," in *25th Inter. Workshop on Bayesian Inference and Maximum Entropy Methods (MaxEnt05). AIP Conference Proceedings* (K. Knuth, A. Abbas, R. Morris, and J. Castle, eds.), vol. 803, pp. 207–214, AIP, 2005.

[47] F. Humblot and A. Mohammad-Djafari, "Super-Resolution using Hidden Markov Model and Bayesian Detection Estimation Framework," *EURASIP Journal on Applied Signal Processing*, vol. Special number on Super-Resolution Imaging: Analysis, Algorithms, and Applications, pp. ID 36971, 16 pages, 2006.

[48] O. Féron, *Champs de Markov cachés pour les problèmes inverses. Application à la fusion de données et à la reconstruction d'images en tomographie micro-onde*. Thèse, Université de Paris–Sud, Orsay, France, Sept. 2006.

[49] O. Féron, B. Duchêne, and A. Mohammad-Djafari, "Microwave imaging of piecewise constant objects in a 2D-TE configuration," *International Journal of Applied Electromagnetics and Mechanics*, vol. 26, pp. 167–174, IOS Press 2007.

[50] H. Ayasso, B. Duchêne, and A. Mohammad-Djafari, "Une approche bayésienne de l'inversion en tomographie optique par diffraction," in *Interférences d'Ondes, Assemblée Générale du GDR Ondes*, (Paris, France), November 2009.

[51] H. Ayasso, B. Duchêne, and A. Mohammad-Djafari, "Bayesian estimation with Gauss–Markov–Potts priors in optical diffraction tomography," in *SPIE, Electronic Imaging (to appear)*, (San Francisco Airport, California, USA), January 2011.

[52] S. Zhu, A. Mohammad-Djafari, L. Xiang, and H. Wang, "A novel hierarchical bayesian method for sar image reconstruction," in *AIP Conference Proceedings*, vol. 1443, p. 222, 2012.

[53] N. Chu, J. Picheral, and A. Mohammad-Djafari, "A robust super-resolution approach with sparsity constraint for near-field wideband acoustic imaging," in *IEEE International Symposium on Signal Processing and Information Technology*, pp. 286–289, Bilbao, Spain, Dec.14-17,2011.