Inverse Problems in Imaging Systems and General Bayesian Inversion Framework

Ali Mohammad-Djafari
Signal and System lab.
Mixed Branch of Research No.8506 (CNRS, Supélec, University of South Paris)
Supélec, Plateau de Moulon, 3 Joliot-Curie, st.
91192 GIF-SUR-YVETTE Cedex, France

Abstract: In this paper, first a great number of inverse problems which arise in instrumentation, in computer imaging systems and in computer vision are presented. Then a common general forward modeling for them is given and the corresponding inversion problem is presented. Then, after showing the inadequacy of the classical analytical and least square methods for these ill posed inverse problems, a Bayesian estimation framework is presented which can handle, in a coherent way, all these problems. One of the main steps, in Bayesian inversion framework is the prior modeling of the unknowns. For this reason, a great number of such models and in particular the compound hidden Markov models are presented. Then, the main computational tools of the Bayesian estimation are briefly presented. Finally, some particular cases are studied in detail and new results are presented.

Key words: Inverse problem, Bayesian framework, Large system, Markov model

1. INTRODUCTION

Inverse problems arise in many applications in science and engineering. The main reason is that, very often we want to measure the distribution of an non-observable quantity $f(r)$ from the observation of another quantity $g(s)$ which is related to it and accessible to the measurement. The mathematical relation which gives $g(s)$ when $f(r)$ is known is called forward problem:

$$g(s) = [Hf(r)](s) + \in(s) \qquad (1)$$

where H is the forward model. In this relation, r and s may represent either time t, position on a line x, position on a surface $r = (x, y)$, position in space $r = (x, y, z)$ or any combinations of them.

This forward model is often non linear, but it can be linearized. So, in this paper, we only consider the linear model, which in its general form, can be written as

$$g(s) = \int h(r,s) f(r)\, dr + \in(s) \qquad (2)$$

Where $h(r,s)$ represents the measuring system response and $\in(s)$ all the errors (modeling, linearization and the other unmodelled errors often called noise). In this paper, we assume that the forward model is known perfectly, or at least, known excepted a few number of parameters. The inverse problem is then the task of going back from the observed quantity $g(s)$ to $f(r)$. The main difficulty is that, very often these problems are ill-posed, in position to the forward problems which are well-posed as defined by Hadamard [1]. A problem is mathematically well-posed if the problem has a solution (existence), if the solution exists (uniqueness), and if the solution is stable (stability).

[Downloaded from jiaeee.com on 2026-11-20]

A problem is then called ill-posed if any of these conditions are not satisfied [2].

In this paper, we will only consider the algebraic methods of inversion where, in a first step the forward problem is discretized, i.e.,, the integral equation is approximated by a sum and the input f, the output g and the errors $\in$ are assumed to be well represented by the finite dimensional vectors f, g and $\in$ such that:

$$g_i = \sum_{j=1}^{n} H_{ij} f_j + \in_i \quad , \quad i = 1,...,n \rightarrow g = Hf + \in \qquad (3)$$

Where $g_i = g(s_i), \in_i = \in (s_i), f_j = f(r_j)$ and $H_{ij} = h(r_j, s_i)$ or in a more general case

$$\begin{aligned} g_i &= \langle \phi_i(s), g(s) \rangle = \int \phi_i(s) g(s)\, ds \\ \in_i &= \langle \phi_i(s), \in (s) \rangle = \int \phi_i(s) \in (s)\, ds \\ f_i &= \langle \psi_{j(s)}, f(r) \rangle = \int \psi_i(r) f(r)\, dr \end{aligned} \qquad (4)$$

where $\phi_i(s)$ and $\psi_j(s)$ are appropriate basis function in their corresponding spaces which means that, we assume

$$\begin{aligned} g(s) &\cong \sum_{i=1}^{m} g_i \phi_i(s) \\ \in (s) &\cong \sum_{i=1}^{m} \in_i \phi_i(s) \\ f(r) &\cong \sum_{j=1}^{n} f_j \psi_i(r) \\ H_{ij} &\cong \langle \phi_i(s), \psi_j(s) \rangle = \int\int \psi_i(r) \phi_i(s)\, dr\, ds \end{aligned} \qquad (5)$$

But, before going further in details of the inversion methods, we are going to present a few examples.

1.1. 1D Signals

Any instrument such as a thermometer which tries to measure a non-directly measurable quantity $f(t)$ (here the time variation of the temperature) transforms it to the time variation of a measurable quantity $g(t)$ (here the length of the liquid in the thermometer). A perfect instrument has be at least linear. Then the relation between the output $g(t)$ and the input $f(t)$ is:

$$g(t) = \int h(t,t') f(t')\, dt + \in (t) \qquad (6)$$

where $h(t,t')$ the instrument's response. If this response is invariant in time, then we have a convolution forward model:

$$g(t) = \int h(t-t') f(t')\, dt + \in (t) \qquad (7)$$

and the corresponding inverse problem is called deconvolution.

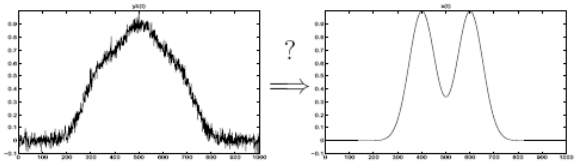


Fig.1: Deconvolution of ID signals.

The convolution equation (7) can also be written

$$g(t) = \int h(\tau) f(t-\tau)\, d\tau + \in (t) \qquad (8)$$

which is obtained by change of variable $t - t' = \tau$. Assuming the sampling interval of f, h and g to be equal to 1, the discretized version of the deconvolution equation can then be written:

$$g(i) = \sum_{k} h(k) f(i-k) + \in (i) \ , \quad i = 1,...,T \qquad (9)$$

which can be written in the general vector-matrix form:

$$g = Hf + \in \qquad (10)$$

where g and f contains samples of the output $g(t)$ and the input $f(t)$ and the matrix H, in this case, is a Toeplitz matrix with a generic ligne composed of the samples of the impulse response $h(t)$. The Toeplitz property is thus identified to the time invariance property of the system response (convolution forward problem).

1.2. Image Restoration

In this paper, we consider more the case of bivariate signals or images. As an example, when the unknown and measured quantities are images, we have:

$$g(r) = \int h(r-r') f(r')\, dr' + \in (r) \qquad (11)$$

and if the system response is spatially invariant, we have

[Downloaded from jiaeee.com on 2026-11-20]

$$g(r) = \int h(r - r')\, f(r') + \in (r) \qquad (12)$$

The case of denoising is the particular case where the point spread function (psf) $h(r)$ is $h(r) = \delta(r)$:

$$g(r) = f(r) + \in (r) \qquad (13)$$

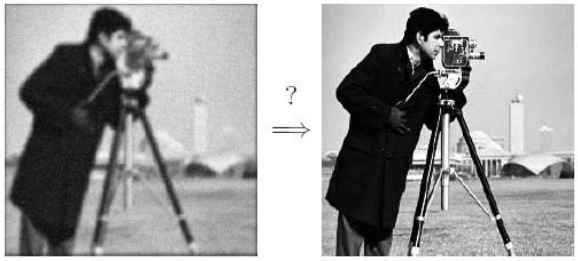


Fig.2: Image restoration as an inverse problem

The discretized version of the 2D deconvolution equation can also be written as $g = Hf + \in$ where g and f contains, respectively, the rasterized samples of the output $g(r)$ and the input $f(r)$, and the matrix H in this case, is a huge dimensional Toeplitz-Bloc-Toeplitz (TBT) matrix with a generic bloc-ligne composed of the samples of the point spread function (PSF) $h(r)$. The TBC property is thus identified to the space invariance property of the system response (2D convolution forward problem). For more details on the structure of this matrix refer to the book [3] and the papers [4, 5, 6].

1.3. Image Reconstruction in Computed Tomography

In previous examples, $g(s)$ and $f(r)$ where defined in the same space. The case of image reconstruction in X ray computed tomography (CT) is interesting, because the observed data $g(s)$ and the unknown image $f(r)$ are defined in different spaces. The usual forward model in CT is shown in Figure (1.3).

In 2D case, the relation between the image to be reconstructed $f(x, y)$ and the projection data $g(r, \phi) = g_\phi(r)$ is given by the Radon transform:

$$g(r,\phi) = \int_{L_{r,\phi}} f(x,y)\,dl + \in (r,\phi) = \iint f(x,y)\,\delta(r - x\cos\phi - y\sin\phi)\,dxy + \in (r,\phi) \qquad (14)$$

The discretized version of this forward equation can also be written as $g = Hf + Q$ where $g = [g_1, \cdots, g_k]$ contains samples of projection data $g(r, \phi_k)$ for different angles ϕ_k, $k = 1, \cdots, K$, $f = \{f(r),\ r \in R\}$ contains the image pixels put in a vector and the elements H_{ij} of the matrix H, in this case, represents the length of the i-th ray in the j-th pixel. This matrix is a very sparse matrix with great number of zero valued elements [7, 8].

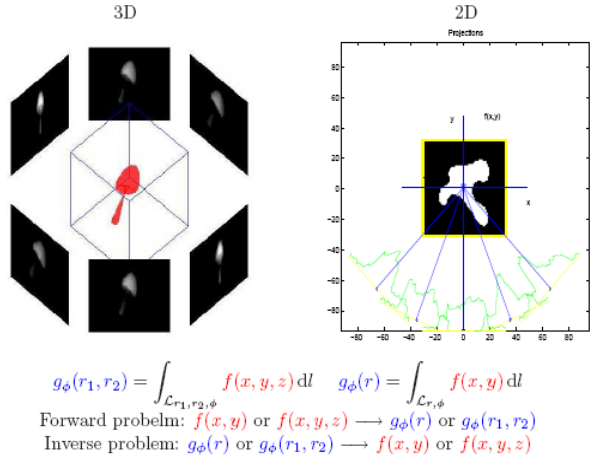


Fig.3: 2D and 3D ray computed tomography

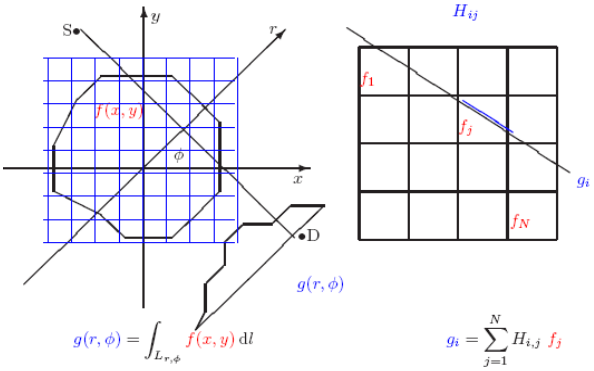


Fig.4: Discretized 2D X ray coputed tomography

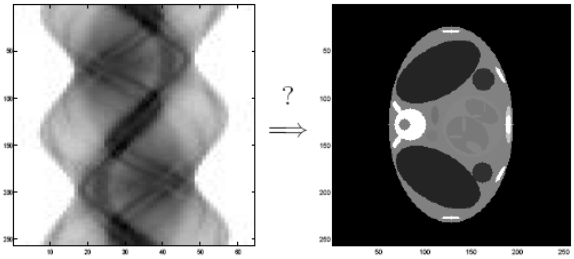


Fig.5:Inverse problem of image reconstruction in x ray computed tomography

[Downloaded from jiaeee.com on 2026-11-20]

1.4 Time varying imaging systems

When the observed and unknown quantities depend on space r and time t, we have

$$g(r,t) = \int h(r-r',t-t')\, f(r',t')\, dr'dt' + \in (r,t) \qquad (15)$$

If the point spread function of the imaging system does not depend on time, then we have

In this case, t can also be considered as an index:

$$g_t(r) = \int h(r-r')\, f_t(r')dr' + \in_t (r) \qquad (17)$$

$$g(r,t) = \int h(r-r')\, f(r',t)\, dr' + \in (r,t) \qquad (16)$$

One example of such problem is the video image restoration shown in Figure (1-6).

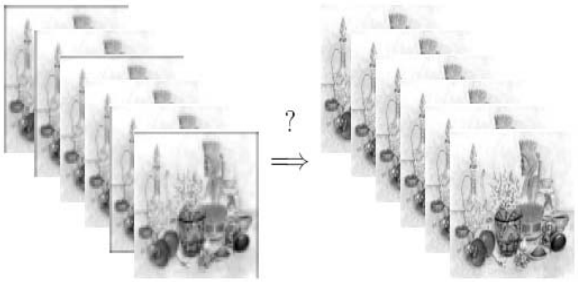


Fig.6: inverse problem of video image restoration

The discretized version of this inverse problem can be written as

$$g_t = Hf_t + \in_t \qquad (18)$$

Where g_t and f_t contains samples of the ouput $g_t(r)$ and the intput $f_t(r)$ and the matrix H, in this case, is again a Toeplitz-Bloc-Toeplitz (TBT) matrix with a generic bloc-ligne composed of the samples of the point spread function (PSF) $h(r)$.

1.5. Multi Inputs Multi Outputs Inverse problems

Multi Inputs Multi Outputs (MIMO) imaging systems can be modeled as:

$$g_i(s) = \sum_{j=1}^{N} \int h_{ij}(s,r)\, f_j(r)\, dr + \in_i (r)\ , \quad i = 1,...,N \qquad (19)$$

1.5.1. MIMO Sources Localization and Estimation

One such example is the case where n radio sources $\{f_j(t),\ j = 1,\cdots,n\}$ emitting in the same time are received by m receivers $\{g_i(t),\ i = 1,\cdots,m\}$, each one receiving a linear combination of delayed and degraded versions of original waves:

$$g_i(t) = \sum_{j=1}^{N} \int h_{ij}(t-t')\, f_j(t'-\tau_{ij})\, dt' + \in_i (t)\ , \quad i = 1,...,N \qquad (20)$$

Where $h_{ij}(t)$ is the impulse response of the channel between the i-th receiver and the j-th source. The discretized version of this inverse problem can be written as

$$g_i = H_{i,j} f_j + \in_i \qquad (21)$$

Where g_i and f_j contains samples of the output $g_i(t)$ and the input $f_t(t)$ and the matrices $H_{i,j}$ are Toeplitz matrices described by the impulse

$$g_i(r) = \sum_j \int h_{ij}(r-r')\, f_j(r')dr' + \in_i (r) \qquad (22)$$

responses $h_{i,j}(t)$.

1.5.2. MIMO Deconvolution

A MIMO image restoration problem is :
and one such example is the case of color image restoration where each color component can be considered as an input.

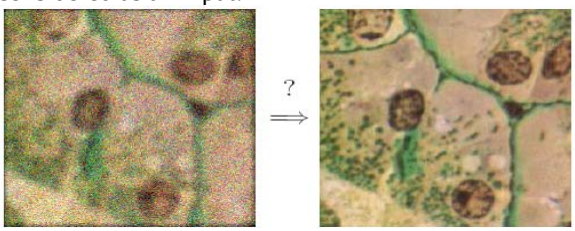


Fig.7: Color image restoration as an example of MMO inverse problem

1.6. Source Separation

A particular case of a MIMO inverse problem is the blind source separation (BSS):

$$g_i(r) = \sum_j \int A_{ij} h_{ij}(r-r')\, f_j(r')\, dr' + \in_i (r) \qquad (23)$$

and a more particular one is the case of instantaneous mixing:

[Downloaded from jiaeee.com on 2026-11-20]

$$g_i(r) = \sum_j A_{ij} f_j(r) + \in_i (r) \qquad (24)$$

The particularity of these problems is that the the mixing matrix $A = \{A_{ij}\}$ is also unknown.

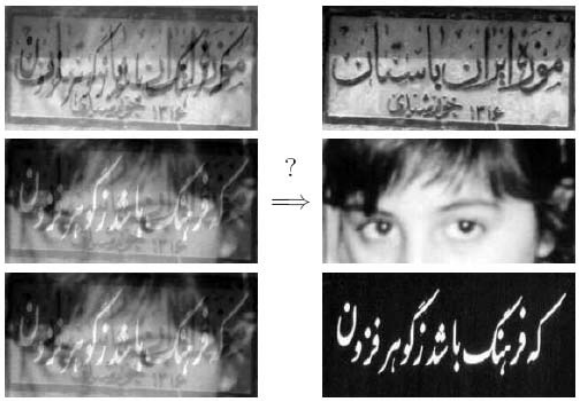


Fig.8: Blind image separation

1.7. Multi Inputs Single Output Inverse problems

A Multi Inputs Single Output (MISO) system is a particular case of MIMO when we have only one input:

$$g_i(s) = \sum_j \int h_i(s,r) f(r)\, dr + \in_i (r) \qquad (25)$$

1.7.1. MIMO sources localization and Estimation

One example of MISO inverse problem is a non destructive testing (NDT) for detection and evaluation of the defect created due to an impact on a surface of an object using microwave imaging where two images are obtained when a rectangular waveguide scans this surface two times. In the first scan the rectangular waveguide is oriented in shorter side and in the second case in longer side. By this way, two images $g_i(r),\ i = 1,2$ are obtained, each has to be considered as the output of a linear system with the same input $f(r)$ and two different channels. This is a MISO linear and invariant systems.

1.7.2. Image Super-resolution as a MISO Inverse Problem

Another MISO system is the case of Super-Resolution (SR) imaging using a few Low Resolution (LR) images obtained by low cost cameras:

$$g_i(s) = \sum_j \int h_i(s,r) f(r)\, dr + \in_i (r) \qquad (26)$$

where g_i are the LR images and f is the desired High Resolution (HR) image. The functions h_i represent a combination of at least three operations: i) a low pass filtering effect, ii) a movement (translational or with rotation and zooming effects) of the camera and iii) a sub-sampling.

The following figure shows one such situation.

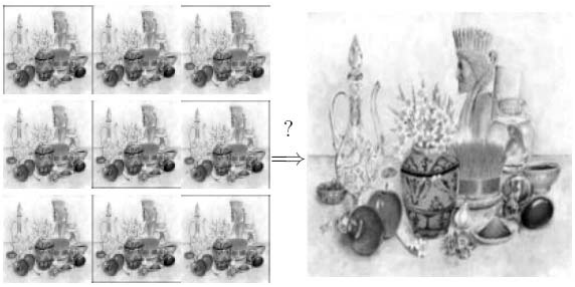


Fig.9: SR problem where a series of LR image are used to construct a HR image

The discretized version of this inverse problem can be written as

$$g_i = H_{i,j} f + \in_i \qquad (27)$$

Where g_i and f contains samples of the output $g_i(t)$ and the input $f_i(t)$ and the matrices $H_{i,j}$ are Toeplitz matrices described by the impulse responses $h_{i,j}(t)$.

1.8. Multi Modality in CT Imaging Systems

Using different modalities has become a main tool in imaging systems where to explore the internal property of a body one can use X rays, ultrasounds, microwaves, infra-red, magnetic resonance, etc. As an example, in X ray imaging, the observed radiographies give some information on the voluminal distribution of the material density inside the object while the ultrasound echography gives information on the changing positions (contours) of ultrasound properties inside the object. One can then want to use both techniques and use a kind of data

[Downloaded from jiaeee.com on 2026-11-20]

fusion to obtain a higher quality of images of the body. An example of such situation is given in (1.8).

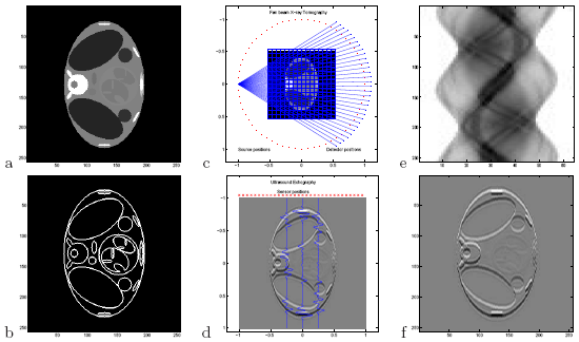


Fig.10: multi modality in CT imaging systems (a) Original object, (b) Contours of the different homogenous regions, (c) Data geometry in X ray tomography, (d) Data acquisition geometry in ultrasound echography, (e) Observed data (sonogram) in X ray tomography, (f) Observed data in ultrasound echography.

1.9. Fusion of X-Ray and Ultrasound Echography

An example of multimodality and data fusion in CT is the use of X ray radiographic data and the ultrasound echographic data is shown in Figure (11) and for more details on this application see [9, 10, 11, 12].

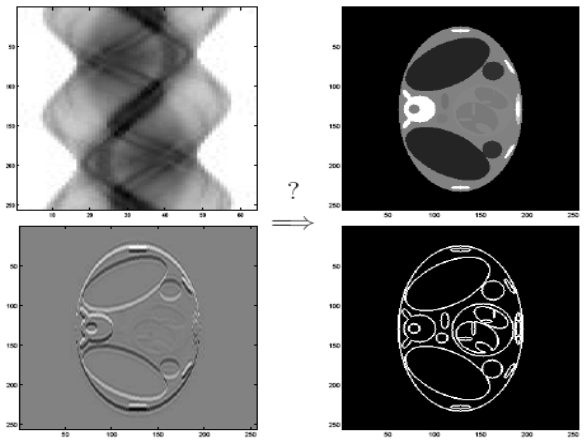


Fig.11: Inverse problem of X ray and ultrasound data fusion

2. Basics of Deterministic Inversion Methods

To illustrate the basics of the inversion methods, we start by considering the case of a Single Input Single Output (SISO) linear system:

$$g = Hf + \in \tag{28}$$

The idea can be easily extended to the case of MISO or MIMO. For an extend details to these methods refer to [13, 14].

2.1. Match Filtering

First assume that the errors and measurement noise are negligible and that we could choose the basis functions ϕ_i and ψ_i could be chosen in such a way that the matrix H is square $(m = n)$ and self-djoint $(H'H = 1)$ (un unrealistic hypothesis). Then, the solution to the problem would be:

$$\hat{f} = H'g \tag{29}$$

This solution has been used in many cases. For example in deconvolution, this solution is called Matching filtering. The main reason is that, in a deconvolution problem, the matrix H is a Toeplitz matrix, so is its transpose H'. The forward matrix operation Hf corresponds to a convolution $\text{conv}(h, f)$. The adjoint matrix operation $H'g$ then also corresponds to a convolution $\text{conv}(h, g)$ where $\tilde{h}(t) = \hat{h}(-t)$.

Another example is in computed tomography (CT) where the projection data in each angle direction g_i is related to the image f through a projecting matrix in that direction H_i such that we can write:

$$\begin{bmatrix} g_1 \\ \vdots \\ g_K \end{bmatrix} = \begin{bmatrix} H_1 \\ \vdots \\ H_K \end{bmatrix} f + \begin{bmatrix} \in_1 \\ \vdots \\ \in_K \end{bmatrix} \tag{30}$$

And the adjoint operation:

$$\hat{f} = H'g = \sum_{k=1}^{K} H'_k g \tag{31}$$

corresponds to what is called back-projection.

However, as it is mentioned, the hypothesis made here are unrealistic.

[Downloaded from jiaeee.com on 2026-11-20]

2.2. Direct Inversion

The next step is just to assume that the forward matrix is invertible. Then, one can try to define the solution as:

$$\hat{f} = H^{-1} g \tag{32}$$

But, in practice, this also is an illusion, because, even if the matrix H is mathematically invertible, it is, very often, very ill-conditioned. This means that small errors on the data δg will generate great errors δ on the solution. This method, in deconvolution, corresponds to the analytical method of inverse filtering, which is, in general, unstable.

In other applications, the main difficulty is that, very often the matrix H is even not square, i.e., $(m \neq n)$, because the number of the measured data m may not be equal to the number of parameters n describing the unknown function f in (5).

2.3. Least Square and Generalized Inversion

For the case where $m \rangle n$, a solution will be the least square (LS) defined as:

$$\hat{f} = \arg\min_{f} \left\{ \left\| g - Hf \right\|^2 \right\} \tag{33}$$

which results to the normal equation:

$$\left[H'H \right] \hat{f} = H'g \tag{34}$$

and if the matrix $H'H$ is inversable (rang $(H'H) = n$), then the solution is given by:

$$\hat{f} = \left[H'H \right]^{-1} H'g \tag{35}$$

When $m \langle n$, the problem may have an infinite number of solutions. So, we may choose one of them by requesting some particular a priori property, for example to have minimum norm. The mathematical problem is then:

$$\hat{f} = \arg \min_{\{Hf = g\}} \left\{ \left\| f \right\|^2 \right\} \tag{36}$$

or written differently

$$\min imize \quad \left\| f \right\|^2 \quad subject \ \ to \ \ Hf = g \tag{37}$$

The solution is obtained via the Lagrange multiplier method which, in this case, results to

$$\begin{bmatrix} I & -H^t \\ H & 0 \end{bmatrix} \begin{pmatrix} f \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ g \end{pmatrix} \tag{38}$$

which gives

$$\hat{f}_2 = H^t (HH^t)^{-1} g \tag{39}$$

if HH^t is invertible.

The main difficulty in these methods is that the solution, in general, is too sensitive to the error in the data due to the ill conditioning of the matrices to be inverted.

2.4. Regularization Methods

The main idea in regularization theory is that a stable solution to an ill-posed inverse problem can not be obtained only by minimizing a distance between the observed data and the output of the model, as it is for example, in LS methods. A general framework is then to define the solution of the problem as the minimizer of a compound criterion such as:

$$\hat{f} = \arg\min_{f} \{ J(f) \} \tag{40}$$

with

$$J(f) = \Delta_1(g, Hf) + \lambda \Delta_2(f, f_0) \tag{41}$$

where Δ_1 and Δ_2 are two distances, the first defined in the observed quantity space and the second in the unknown quantity space. λ is the regularization parameter which regulates the compromise with the two terms and f_0 is an a priori solution. An example of such criterion is:

$$J(f) = \left\| g - Hf \right\|^2 + \lambda \left\| f - f_0 \right\|^2 \tag{42}$$

which results to

$$\hat{f} = f_0 + \left[H'H + \lambda I \right]^{-1} H' (g - Hf_0) \tag{43}$$

We may note that the condition number of the matrix to be inverted here can be controlled by appropriately choosing the value of the regularization parameter λ.

[Downloaded from jiaeee.com on 2026-11-20]

Even if the methods based on regularization approach have been used with success in many applications, three main open problems still remains:
i) Determination of the regularization parameter, ii) The arguments for choosing the two distances Δ_1 and Δ_2 and iii) Quantification of the uncertainties associated to the obtained solutions. Even if there have been a lot of works trying to answer to these problems and there are effective solutions such as the L-curve or the Croos Validation for the first, the two others are still open problems. The Bayesian estimation framework, as we will see, can give answers to them [15].

3. Bayesian Estimation Framework

To illustrate the basics of the Bayesian estimation framework, let first consider the simple case of SISO system $g = Hf + \in$ where we assume that H is known.
In a general Bayesian estimation framework, the forward model is used to define the likelihood function $p(g|f,\theta_1)$ and we have to translate our prior knowledge about the unknowns f through a prior probability law $p(f|\theta_2)$ and then use the Bayes rule to find an expression for $p(f|g,\theta)$.

$$p(f|g,\theta) = \frac{p(g|f,\theta_1)p(f|\theta_2)}{p(g|\theta)} \tag{44}$$

where $p(g|f,\theta_1)$ is the likelihood whose expression is obtained from the forward model and assumption on the errors $\in$, $\theta = (\theta_1, \theta_2)$ represents all the hyper-parameters (parameters of the likelihood and priors) of the problem and

$$p(g|\theta) = \int p(g|f,\theta_1)p(f|\theta_2)\,df \tag{45}$$

is called the evidence of the model.

When the expression of $p(f|g,\theta)$ is obtained, we can use it to define any estimates for f. Two usual estimators are the maximum a posteriori (MAP)

$$\hat{f} = \arg\max_f \{p(f|g,\theta)\} \tag{46}$$

and the Mean Square Error (MSE) estimator which corresponds to the posterior mean

$$\hat{f} = \int f\ p(f|g,\theta)df \tag{47}$$

Unfortunately only for the linear problems and the Gaussian laws where $p(f|g,\theta)$ is also Gaussian we have analytical solutions for these two estimators. For almost all other cases, the first one needs an optimization algorithm and the second an integration one. For example, the relaxation methods can be used for the optimization and the MCMC algorithms can be used for expectation computations. Another difficult point is that the expressions of $p(g|f,\theta_1)$ and $p(f|\theta_2)$ and thus the expression of $p(f|g,\theta)$ depend on the hyper-parameters θ which, in practical applications, have also to be estimated either in a supervised way using the training data or in an unsupervised way. In both cases, we need also to translate our prior knowledge on them through a prior probability $p(\theta)$.
Thus, one of the main steps in any inversion method for any inverse problem is modeling the unknowns. In probabilistic methods and in particular in the Bayesian approach, this step becomes the assignment of the probability law $p(f|\theta_1)$. This point, as well as the assignment of $p(\theta)$, are discussed the next two subsections.

3.1 Simple case of Gaussian models

Let consider as a first example the simple case where ϵ and f are assumed to be Gaussian:

$$\begin{aligned} p(g|f,\delta_\in^2) &= N(0, R_\in = \sigma_\in^2 I_0) \\ &\propto \exp\left[-\frac{1}{2\sigma_\in^2}\in^t\in\right] \\ p(f|\sigma_f^2, P_0) &= N(f_0, R_f = \sigma_f^2 P_0) \\ &\propto \exp\left[-\frac{1}{2\sigma_f^2}(f-f_0)^t P_0^{-1}(f-f_0)\right] \end{aligned} \tag{48}$$

Then, it is easy to show that:

[Downloaded from jiaeee.com on 2026-11-20]

$$
\begin{aligned}
p(g|f,\sigma_\epsilon^2) &= N(Hf,\sigma_\epsilon^2 I_0) \\
&\propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}(g-Hf)^t(g-Hf)\right] \\
p(g|\sigma_\epsilon^2,\sigma_f^2,P_0) &= N(Hf_0, HR_fH^t+R_\epsilon) \\
p(g,f|\sigma_\epsilon^2,\sigma_f^2,P_0) &\propto \exp\left[-\frac{1}{2\sigma_\epsilon^2}(g-Hf)^t(g-Hf)-\frac{1}{2\sigma_f^2}(f-f_0)^t P_0^{-1}(f-f_0)\right]
\end{aligned} \tag{49}
$$

and

$$
p(f|g,\sigma_\epsilon^2,\sigma_f^2,P_0)=\frac{p(g,f|\sigma_\epsilon^2,\sigma_f^2,P_0)}{p(g|\sigma_\epsilon^2,\sigma_f^2,P_0)}=N(\hat{f},\hat{p}) \tag{50}
$$

With

$$
\begin{cases}
\hat{f} = f_0 + R_f H^t (HR_f H^t + R_e)^{-1}(g - Hf_0) \\
\quad = \hat{R}H^t R_e^{-1}(g - Hf_0), \\
\hat{P} = R_f - R_f H^t (HR_f H^t + R_e)^{-1} HR_f \\
\quad = (R_f^{-1} + H^t R_e^{-1} H)^{-1}
\end{cases} \tag{51}
$$

When $f_0 = 0$ and noting by $\lambda = \dfrac{\sigma_\epsilon^2}{\sigma_f^2}$, these relations write:

$$
\begin{cases}
\hat{f} = (H^t H + \lambda P_0^{-1})^{-1} H^t g = \hat{P} H^t g \\
\hat{P} = \sigma_e^2 (H^t H + \lambda P_0^{-1})^{-1}
\end{cases} \tag{52}
$$

It is noted that, in this case, all the point estimators such as the MAP, the posterior mean or posterior median are the same and can be obtained by:

$$
\hat{f} = \arg\max_f \{p(f \mid g)\} = \arg\min_f \{-\ln p(f \mid g)\} = \arg\min_f \{J(f)\} \tag{53}
$$

with

$$
J(f) = \| g - Hg \|^2 + \lambda (f^t P_0^{-1} f) \tag{54}
$$

Three particular cases are of interest:

- $P_0 = I$. This is the case where f_j are assumed centered, Gaussian and i.i.d.:

$$
p(f) \propto \exp\left[-\frac{1}{2\sigma_f^2}\sum_j f_j^2\right] \propto \exp\left[-\frac{1}{2\sigma_f^2}\| f \|^2\right] \tag{55}
$$

- $P_0 = CC^t$. This is the case where f_j are assumed centered, Gaussian but correlated. The vector $\mathbf{f}$ is then considered to be obtained by:

$$
f = C\xi \tag{56}
$$

with C corresponds to a moving average (MA) filtering and $p(\xi) = N(0,1)$. In this case, we have:

$$
p(f) \propto \exp\left[-\frac{1}{2\sigma_f^2}\sum_j [Cf]_j^2\right] \propto \exp\left[-\frac{1}{2\sigma_f^2}\| Cf \|^2\right] \tag{57}
$$

- $p_0 = \left(D^t D\right)^{-1} = \left(I - A\right)^{-1}$. This is the case where f_j are assumed centered, Gaussian and autoregressive:

$$
f = Af + \xi \tag{58}
$$

with A a matrix obtained from the AR coefficients and $p(\xi) = N(0,1)$. In this case, we have

$$
p(f) \propto \exp\left[-\frac{1}{2\sigma_f^2}\| Df \|^2\right] \tag{59}
$$

A particular case of AR model is the first order Markov chain

$$
p(f_j \mid f_j) = N(f_{j-1}, \sigma_f^2) \tag{60}
$$

with corresponding A and $D = I - A$ matrices

$$
A = \begin{bmatrix} 0 & 0 & & 0 \\ 1 & 0 & & 0 \\ 0 & 1 & 0 & \\ & & & \\ 0 & & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 1 & 0 & & 0 \\ -1 & 1 & & 0 \\ 0 & -1 & 1 & \\ & & & \\ 0 & & -1 & 1 \end{bmatrix} \tag{61}
$$

which give the possibility to write

$$
p(f) \propto \exp\left[-\frac{1}{2\sigma_f^2}\| Df \|^2\right] \propto \exp\left[-\frac{1}{2\sigma_f^2}\sum_j (f_j - f_{j-1})^2\right] \tag{62}
$$

[Downloaded from jiaeee.com on 2026-11-20]

These particular cases give us the possibility to extend the prior model to other more sophisticated non-Gaussian models which can be classified in three groups:

- Separable:

$$p(f) \propto \exp\left[-\alpha \sum_j \phi(f_j)\right] \tag{63}$$

where ϕ is any positive valued function.

- Simple Markovian:

$$p(f) \propto \exp\left[\sum_j \phi(f_j - f_{j-1})\right] \tag{64}$$

where ϕ is any positive valued function called potential function of the Markovian model.

- Compound Markovian:

$$p(f \mid c) \propto \exp\left[-\alpha \sum_j \phi(f_j - f_{j-1}, c_j)\right] \tag{65}$$

where ϕ is any positive valued function whose expression depends on the hidden variable c.
Some examples of the ϕ expressions used in many applications are:

$$\phi(t) = \left\{ t^2;\ |t|^\beta,\ 1 \le \beta \le 2;\quad -t\ln t + 1,\ t > 0;\quad \min(t^2, 1);\ \frac{-1}{1+t^2} \right\} \tag{66}$$

These equations can easily be extended for the case of multi-sensor case.

However, even if a Gaussian model for the noise is acceptable, this model is rarely realistic for most real word signals or images. Indeed, very often, a signal or an image can be modeled locally by a Gaussian, but its energy or amplitude can be modulated, i.e.; piecewise homogeneous and Gaussian [16, 17, 18]. To find an appropriate model for such cases, we introduce hidden variables and in particular hidden Markov modeling (HMM). In the following, we first give a summary description of these models and then we will consider the general case of MIMO systems with prior HMM modeling.

3.2. Modeling Using Hidden Variables

3.2.1 Signal and Images with Energy Modulation

A simple model which can capture the variance modulated signal or images is [19, 17, 20].

$$p(f_j \mid d_j, \lambda) = N(0, 2d_j) \quad , \quad p(d_j \mid \lambda) = \varsigma(3/2, \lambda) \tag{67}$$

where G is a Gamma distribution. It is then easy to show the following relations:

$$p(f, d \mid \lambda) \propto \exp\left[-\lambda \sum_j \left(\frac{f^2}{4d_j^2} + d_j\right)\right]$$

$$p(g, f) \propto \exp\left[\frac{-1}{2\sigma_e^2}\left\|g - Hf\right\|^2\right] \tag{68}$$

and

$$p(f, d \mid g) \propto exp\left[-J(f, d)\right] \tag{69}$$

$$J(f, d) = \frac{1}{2\sigma_e^2}\left\|g - Hf\right\|^2 + \lambda \sum_j \left(\frac{f_j^2}{4d_j^2} + d_j\right)$$

If we try to find the joint MAP estimate of the unknowns (f, d) by optimization successively with respect to f when d is fixed and with respect to d when f is fixed, we obtain the following iterative algorithm:

$$\hat{f} = (\sigma_e^{-2} H^t H + 2\lambda D)^{-1} H^t g$$
$$D = diag\left[1/(4d_j^2),\ j = 1, \ldots, n\right] \tag{70}$$
$$\hat{d}_j = f_j / 2$$

3.2.2. Amplitude Modulated Signals

To illustrate this with applications in telecommunication signal and image processing, we consider the case of a Gaussian signal modulated with a two level or binary signal. A simple model which can capture the variance modulated signal or images is

$$p(f_j \mid z_j, \lambda) = N(z_j, 2/\lambda)$$

[Downloaded from jiaeee.com on 2026-11-20]

$$z_j \in \{m_1 = 0,\ m_2 = 1\} \qquad (71)$$
$$p(z_j = m_k) = (1/2),\ k = 1,\ldots,K = 2$$

It is then easy to show the following:

$$p(f_j \mid \lambda) = (1/2)[N(0, 2/\lambda) + N(1, 2/\lambda)]$$
$$= \sum_{k=1}^{K} (1/2) N(m_k, \sigma_k^2 = 2/\lambda)$$
$$p(f \mid z, \lambda) \propto exp\left[-\lambda \sum_j (f_j - z_j)^2\right]$$
$$p(f_j \mid z_j, \lambda) \propto exp\left[-\lambda (f_j - z_j)^2\right]$$
$$p(z \mid f, \lambda) \propto exp\left[-\lambda \sum_j (z_j - f_j)^2\right] \qquad (72)$$
$$p(z_j = k \mid f_j, \lambda) \propto \left[-\lambda (z_j - f_j)^2\right]$$
$$p(g \mid f, \sigma_e^2) \propto exp\left[\frac{-1}{2\sigma_e^2} \|g - Hf\|^2\right]$$

and

$$p(f, z \mid g, \sigma_e^2, \lambda) \propto exp[-J(f, z)]$$
$$J(f, z) = \frac{1}{2\sigma_e^2} \|g - Hf\|^2 + \lambda \|f - z\|^2 \qquad (73)$$
$$+ ln(1/2) \sum_k \sum_j \delta(z_j - m_k)$$

where $z = [z_1, \cdots, z_N]'$.

Again, trying to obtain the JMAP estimate $(\hat{f}, \hat{z})$ by optimizing successively $J(f, z)$ with respect to f and z we obtain:

$$\hat{f} = (\sigma_e^{-2} H^t H + \lambda I)^{-1} [H^t g + \lambda z]$$
$$\hat{z}_j = \begin{cases} 1 & f_j > z \\ 0 & f_j < a \end{cases} \qquad (74)$$

where the threshold a is a function of λ .

3.2.3. Gaussians Mixture Model

The previous model can be generalized to the general mixture of Gaussians. We then have the following relations:

$$p(f_j \mid z_j = k, m_k, \upsilon_k) = N(m_k, \upsilon_k = 2/\lambda_k)$$
$$p(z_j = k) = \pi_k \quad z_j \in \{1,\ldots,K\}$$

$$P(f_j \mid \pi_k, m_k, \upsilon_k) = \sum_{k=1}^{K} \pi_k N(m_k, \upsilon_k)$$
$$= \sum_{k=1}^{K} p(z_j = k) N(m_k, \upsilon_k) \qquad (75)$$
$$p(f \mid z, m, \lambda) \propto exp\left[-\sum_k \sum_{\{j: z_j = k\}} \lambda_k (f_j - m_k)^2\right]$$
$$\propto exp\left[-\sum_j \sum_k \lambda_k \delta(z_j - k)(f_j - m_k)^2\right]$$
$$p(g \mid f, \sigma_e^2, m) \propto exp\left[\frac{-1}{2\sigma_e^2} \|g - Hf\|^2\right]$$

and thus

$$p(z \mid f, m, \lambda) \propto p(f \mid z, m, \lambda) \prod_k \pi_k^{\sum_j \delta(z_j - k)}$$
$$\propto exp\left[-\sum_j \sum_k \left[\lambda_k \delta(z_j - k)(f_j - m_k)^2 + ln\, \pi_k\right]\right]$$
$$p(z_j = k \mid f, m, \lambda) \propto \left[-\lambda_k (f_j - m_k)^2 + ln\, \pi_k\right] \qquad (76)$$
$$p(f, z \mid g, \sigma_e^2, m, \lambda) \propto exp\left[-J(f, z)\right]$$

and

$$p(f, z \mid g, \sigma_e^2, m, \lambda) \propto exp[-J(f, z)]$$

with

$$J(f, z) = \frac{1}{2\sigma_e^2} \|g - Hf\|^2 + \sum_k \sum_{\{j, z_j = k\}} \lambda_k (f_j - m_k)^2$$
$$+ \sum_k ln(\pi_k) \sum_j \delta(z_j - m_k)$$
$$= \frac{1}{2\sigma_e^2} \|g - Hf\|^2 + \sum_k \lambda_k \|f_k - m_k 1\|^2$$
$$+ \sum_k n_k\, ln(\pi_k) \qquad (77)$$

where $m = \{m_1, \cdots, m_k\}$, $\lambda = \{\lambda_1, \cdots, \lambda_k\}$, $n_k = p_j \delta(z_j - k)$ is the number of samples f_j which are in the class $z_j = k$ and $f_k = \{f_j : z_j = k\}$. For more details and applications of such modeling see [21, 22, 23, 24].

3.2.4. Mixture of Gauss-Markov Model

In the previous model, we assumed that the samples in each class are independent. Here, we extend this to a Markovian model:

$$p(f_j \mid z_j = k, z_{j-1} \neq k,\ f_{j-1}, m_k, \upsilon_k) = N(m_k, \upsilon_k)$$
$$p(f_j \mid z_j = k, z_{j-1} = k,\ f_{j-1}, m_k, \upsilon_k) = N(f_{j-1}, \upsilon_k)$$
$$p(z_j = k) = \pi_k \quad z_j \in \{1,\ldots,K\} \qquad (78)$$

[Downloaded from jiaeee.com on 2026-11-20]

which can be written in a more compact way if we introduce $q_j = 1-\delta(z_j - z_{j-1})$ by

$$p(f_j \mid g_j, f_{j-1}, m_k, \upsilon_k) = N(q_j m_k + (1-q_j) f_{j-1}, \upsilon_k \quad (79)$$

which results in:

$$p(f \mid z,m,\lambda) \propto exp\left[-\sum_j \sum_k \lambda_k \delta(z_j - k)\left[f_j - (q_j m_k + (1-q_j) f_{j-1})\right]^2\right]$$
$$\propto exp\left[-\sum_j \sum_k \lambda_k \delta(z_j - k)\left[(1-q_j)(f_j - g_i)(f_j - f_{j-1})^2 + q_j (f_j - m_k)^2\right]\right] \quad (80)$$

and when combined with

$$p(g \mid f, \sigma_e^2) \propto exp\left[\frac{-1}{2\sigma_e^2}\|g - Hf\|^2\right]$$

gives:

$$p(f, z \mid g, \sigma_e^2, m, \lambda) \propto exp\left[-J(f, z)\right]$$

with

$$J(f,z) = \frac{1}{2\sigma_e^2}\|g - Hf\|^2 + \sum_j \sum_k \lambda_k \delta(z_j - k)\left[f_j - (q_j m_k + (1-q_j) f_{j-1})\right]^2 + \sum_k n_k \ln(\pi_k) \quad (81)$$
$$= \frac{1}{2\sigma_e^2}\|g - Hf\|^2 + \sum_j (1-q_j)(\tilde{f}_j - \tilde{f}_{j-1})^2 + \sum_k n_k \ln(\pi_k)$$
$$= \frac{1}{2\sigma_e^2}\|g - Hf\|^2 + \left\|GD\tilde{f}\right\|^2 + \sum_k n_k \ln(\pi_k)$$

where $\hat{f}_j = \lambda z_j (f_j - m z_j)$, D is the first order finite difference matrix and Q is a matrix with q_i as its diagonal elements.

A particular case of this model is of great interest: $m_k = 0, \forall k$ and $\lambda_k = \lambda, \ \forall k$. Then, we have:

$$p(f_j \mid q_j, f_{j-1}, m_k, \upsilon_k) = N((1-g_j) f_{j-1}, \upsilon_k)$$
$$p(f \mid q,m,\lambda) \propto exp\left[-\sum_j \lambda\left[f_j - (1-q_j) f_{j-1}\right]^2\right]$$
$$\propto exp\left[-\lambda \sum_j \left[(1-q_j)(f_j - f_{j-1})^2 + q_j f_j^2\right]\right] \quad (82)$$

and

$$p(f, q \mid g, \sigma_e^2, m, \lambda) \propto exp\left[-J(f,q)\right]$$

with

$$J(f,q) = \frac{1}{2\sigma_e^2}\|g - Hf\|^2 + \lambda \sum_j \left[(1-q_j)(f_{j-1})^2 + q_j f_j^2\right] \quad (83)$$
$$= \frac{1}{2\sigma_e^2}\|g - Hf\|^2 + \lambda\|QDf\|^2 + \sum_k n_k \ln(\alpha_k)$$

Where $n_k = \sum_j q_j$ is the number of discontinuities (length of the contours in the case of an image) $\alpha_k = p(q_i = 1)$ and $1 - \alpha_k = P(q_j = 0)$.

In all these mixture models, we assumed z_j independent with $P(z = k) = \prod_k$. However, z_j corresponds to the label of the sample f_j. It is then better to put a Markovian structure on it to capture the fact that, in general, when the neighboring samples of f_j have all the same label, then it must be more probable that this sample has the same label. This feature can be modeled via the Potts-Markov modeling of the classification labels z_j. In the next section, we use this model, and at the same time, we extend all the previous models to 2D case for applications in image processing and to MIMO applications.

3.3. Mixture and Hidden Markov Models for Images

In image processing applications, the notions of contours and regions are very important. In the following, we note by $r = (x, y)$ the position of a pixel and by $f(r)$ its gray level or by $f(r) = \{f_1(r), ..., f_N(r)\}$ its color or spectral components. In classical RGB color representation $N = 3$, but in hyper-spectral imaging N may be more than one hundred. When the observed data are also images we note them by $g(r) = \{g_1(r), ..., g_M(r)\}$. For any image $f_j(r)$ we note by $q_j(r)$, a binary valued hidden variable, its contours and by $z_j(r)$, a discrete value hidden variable representing its region labels. We focus here on images with homogeneous regions and use the mixture models of the previous section with an additional Markov model for the hidden variable $z_j(r)$.

[Downloaded from jiaeee.com on 2026-11-20]

3.3.1 Homogeneous regions modeling

In general, any image $f_j(r),\ r \in R$ is composed of a finite set k_j of homogeneous regions R_{jk} with given labels $z_j(k) = K,\ K = 1,...,k_j$ such that $R_{jk} = \{r : z_j(k) = K\}$, $R_j = U_k R_{jk}$ and the corresponding pixel values $f_{jk} = \{f_j(r) : r \in R_{jk}\}$ and $f_j = U_k f_{jk}$. The Hidden Markov modeling (HMM) is a very general and efficient way to model appropriately such images.

The main idea is to assume that all the pixel values $f_{jk} = \{f_j(r) : r \in R_{jk}\}$ of a homogeneous region k follow a given probability law, for example a Gaussian $N(m_{jk}1, \sum_{jk})$ where **1** is a generic vector of ones of the size n_{jk} the number of pixels in region k.

In the following, we consider two cases:

- The pixels in a given region are assumed iid:

$$p(f_j(r) \mid z_j(r) = k) = N(m_{jk}, \sigma^2_{j,k}), \quad k = 1,...,K_j \quad (84)$$

and thus

$$p(f_{jk} \mid z_j(r) = k) = p(f_j(r), r \in R_{jk}) = N(m_{jk}1, \sigma^2_{j,k} I) \quad (85)$$

This corresponds to the classical separable and monovariate mixture models.

- The pixels in a given region are assumed to be locally dependent:

$$p(f_{jk} \mid z_j(r) = k) = p(f_j(r), r \in R_{jk}) = N(m_{jk}1, \sum_{jk}) \quad (86)$$

where $\sum_{jk}$ is an appropriate covariance matrix. This corresponds to the classical separable but multivariate mixture models.

In both cases, the pixels in different regions are assumed to be independent:

$$p(f_j) = \prod_{k=1}^{K_j} p(f_{jk}) = \prod_{k=1}^{K_j} N(m_{jk}1, \sum_{jk}) \quad (87)$$

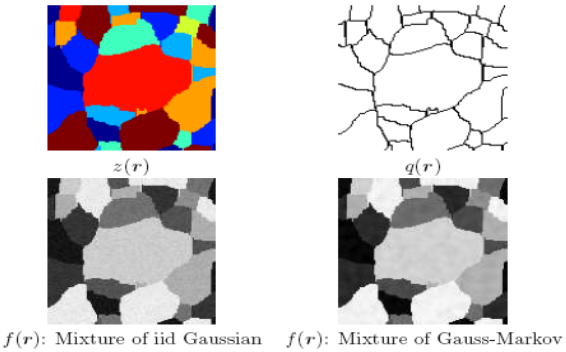


Fig.12: mixture and hidden Markov models for images

3.3.2. Modeling the Labels

Noting that all the models (84), (85) and (86) are conditioned on the value of $z_j(r) = k$, they can be rewritten in the following general form

$$p(f_{jk}) = \sum_k P(z_j(r) = k) N(m_{jk}, \sum_{jk}) \quad (88)$$

where either $\sum_{jk}$ is a diagonal matrix $\sum_{jk} = \sigma_{jk}{}^2 I$ or not. Now, we need also to model the vector variables $z_j = \{z_j(r),\ r \in R\}$. Here also, we can consider two cases:

- Independent Gaussian Mixture model (IGM), where $\{z_j(r),\ r \in R\}$ are assumed to be independent and

$$P(z_j(r) = k) = pk, \text{ with } \sum_k pk = 1, \quad p(z_j) = \prod_k pk \quad (89)$$

- Contextual Gaussian Mixture model (CGM), where $z_j = \{z_j(r),\ r \in R\}$ are assumed to be Markovian:

$$p(z_j) \propto exp\left[\alpha \sum_{r \in R} \sum_{r \in v(r)} \delta(z_j(r) - z_j(s))\right] \quad (90)$$

which is the Potts Markov random field (PMRF). The parameter α controls the mean value of the regions' sizes.

[Downloaded from jiaeee.com on 2026-11-20]

3.3.3. Hyper-parameters Prior Law

The final point before obtaining an expression for the posterior probability law of all the unknowns, i.e., $p\left(\underline{f},\theta\middle|\underline{g}\right)$ is to assign a prior probability law $p\left(\underline{\theta}\right)$ to the hyper-parameters θ. Even if this point has been one of the main discussing points between Bayesian and classical statistical research community, and still there are many open problems, we choose here to use the conjugate priors for simplicity. The conjugate priors have at least two advantages: 1) they can be considered as a particular family of a differential geometry based family of priors [25, 26, 27] and 2) they are easy to use because the prior and the posterior probability laws stay in the same family. In our case, we need to assign prior probability laws to the means m_{jk}, to the variances $\sigma_{jk}^{\ 2}$ or to the covariance matrices $\sum_{jk}$ and also to the covariance matrices of the noises $\in_i$ of the likelihood functions. The conjugate priors for the means m_{jk} are in general the Gaussians $N\left(m_{j_{k0}},\sigma_{j_{ko}}^{\ 2}\right)$, those of variances $\sigma_{jk}^{\ 2}$ are the inverse Gammas $IG\left(\alpha_0,\beta_0\right)$ and those for the covariance matrices $\sum_{jk}$ are the inverse Wishart's $Iw(\alpha_0,\Lambda_0)$.

3.3.4. Expressions of Likelihood, Prior and Posterior Laws

We now have all the elements for writing the expressions of the posterior laws. We are going to summarizes them here:

- Likelihood:

$$p\left(\underline{g}\middle|\underline{f},\underline{\theta}\right)=\prod_{i=1}^{M}p\left(\underline{g}\middle|\underline{f},\sum_{\in i}\right)=\prod_{i=1}^{M}N\left(\underline{g}-\underline{f},\sum_{\in i}\right)$$

where we assumed that the noises $\in_i$ are independent, centered and Gaussian with covariance matrices $\sum_{\in i}$ which, hereafter, are also assumed to be diagonal $\sum_{\in i}=\sigma_{\in i}^{\ 2}I$.

- HMM for the images:

$p\left(\underline{f}\middle|\underline{z},\underline{\theta}\right)=\prod_{j=1}^{N}p\left(f_j\middle|z_j,m_j,\sum_j\right)$ where we used $\underline{z}=\left\{z_j,\ j=1,...,N\right\}$ and where we assumed that $f_j\middle|z_j$ are independent.

- PMRF for the labels:

$$p\left(\underline{z}\right)\propto\prod_{j=1}^{N}\exp\left[\alpha\sum_{r\in R}\ \sum_{s\in v(r)}\delta\left(z_j(r)-z_j(s)\right)\right]$$

where we used the simplified notation $p\left(z_j\right)=P\left(Z_j(r)=z(r),\ r\in R\right)$ and where we assumed $\left\{z_j,\ j=1,...,N\right\}$ are independent.

- Conjugate priors for the hyper-parameters:

$$p(m_{jk})=N(m_{jk0},\sigma_{jk0}^2),\qquad p(\sigma_{jk}^2)=\tau\varsigma(\alpha_{j0},\beta_{j0}),$$
$$p(\sum_{jk})=\tau W(\alpha_{j0},\Lambda_{j0}),\qquad p(\sigma_{ei})=\tau\varsigma(\alpha_{j0},\beta_{j0})$$

- Joint posterior law of $\underline{f}$, $\underline{z}$ and $\underline{\theta}$.

$$p(\underline{f},\underline{z},\underline{\theta}\mid\underline{g})\propto(\underline{g}\mid\underline{f},\theta_1)p(\underline{f}\mid\underline{z},\theta_2)p(\underline{z}\mid\theta_2)p(\underline{\theta})$$

3.4. Bayesian Estimators and Computational Methods

The expression of this joint posterior law is, in general, known up to a normalization factor. This means that, if we consider the Joint Maximum *A* Posteriori (JMAP) estimate:

$$(\hat{\underline{f}},\hat{\underline{f}},\hat{\underline{\theta}})=arg\ \max_{(\underline{f},\underline{z},\underline{\theta})}\left\{p(\underline{f},\underline{z},\underline{\theta}\mid\underline{g})\right\}\qquad(91)$$

we need a global optimization algorithm, but if we consider the Minimum Mean Square Estimator (MMSE) or equivalently the Posterior Mean (PM) estimates, then we need to compute this factor which needs huge dimensional integrations. There are however three main approaches to do Bayesian computation:

- Laplase approximation: When the posterior law is unimodal, it is reasonable to approximate it with an equivalent Gaussian which allows then to do all computations analytically. Unfortunately, very often, $p\left(\underline{f},\underline{z},\underline{\theta}\middle|\underline{g}\right)$ as a function of $\underline{f}$ only may be Gaussian, but as a function of $\underline{z}$ or $\underline{\theta}$ is not. So, in general, this approximation method can not be used for all variables.

- Variational and mean field approximation: The main idea behind this approach is to approximate the joint posterior $p\left(\underline{f},\underline{z},\underline{\theta}\middle|\underline{g}\right)$ with another simpler distribution $q\left(\underline{f},\underline{z},\underline{\theta}\middle|\underline{g}\right)$ for which the computations can be done. A first step simpler distribution $q\left(\underline{f},\underline{z},\underline{\theta}\middle|\underline{g}\right)$ is a separable ones:

[Downloaded from jiaeee.com on 2026-11-20]

$$q(\underline{f},\underline{z},\underline{\theta}\,|\,\underline{g}) = q_1(\underline{f})q_2(\underline{z})q_3(\underline{\theta}) \qquad (92)$$

In this way, at least reduces the integration computations to the product of three separate ones. This process can again be applied to any of these three distributions, for example $q_1(\underline{f}) = \prod_j q_{1j}(f_j)$. With the Gaussian mixture modeling we proposed, $q_1(\underline{f})$ can be chosen to be Gaussian, $q_2(\underline{z})$ to be separated to two parts $q_{1B}(\underline{z})$ and $q_{1W}(\underline{z})$ where the pixels of the images are separated in two classes B and W as in a checker board. This is thanks the properties of the proposed Potts-Markov model with the four nearest neighborhood which gives the possibility to use $q_{1B}(\underline{z})$ and $q_{1W}(\underline{z})$ separately. For $q_3(\underline{\theta})$ very often we also choose a separable distribution which use the conjugate properties of the prior distributions.

• Markov Chain Monte Carlo (MCMC) sampling which gives the possibly to explore the joint posterior law and compute the necessary posterior mean estimates. In our case, we propose the general MCMC Gibbs sampling algorithm to estimate $\underline{f}$, $\underline{z}$ and $\underline{\theta}$ by first separating the unknowns in two sets $p\left(\underline{f},\underline{z}\middle|\underline{\theta},g\right)$ and $p\left(\underline{\theta}\middle|\underline{f},\underline{z},\underline{g}\right)$. Then, we separate again the first set in two subsets $p\left(\underline{f}\middle|\underline{z},\underline{\theta},\underline{g}\right)$ and $p\left(\underline{z}\middle|\underline{\theta},\underline{g}\right)$. Finally, when possible, using the separability along the channels, separate these two last terms in $p\left(f_j\middle|z_j,\theta_j,g_j\right)$ and $p\left(z_j\middle|\theta_j,g_j\right)$. The general scheme is then, using these expressions, to generates samples $\underline{f}^{(n)}$, $\underline{z}^{(n)}$, $\underline{\theta}^{(n)}$ from the joint posterior law $p\left(\underline{f},\underline{z},\underline{\theta}\middle|\underline{g}\right)$ and after the convergence of the Gibbs samplers, to compute their mean and to use them as the posterior estimates.

In this paper we are not going to detail these methods. However, in the following we propos to examine some particular cases through a few case studies in relation to image restoration, image fusion and joint segmentation, blind image separation.]

4. Case Studies

4.1. Single Channel Image Denoising and Restoration

The simplest example of inversion is a single channel image denoising and restoration when the PSF of the imaging system is given. The forward model for this problem is

$$g(r) = h(r) * f(r) + e(r), \quad r \in R \text{ or } g = Hf + e \qquad (93)$$

where the denoising case corresponds to the case where $h(r) = \sigma(r)$ and $H = 1$. Assuming the noise to be centered, white and Gaussian with known variance σ_ϵ^2, we have

$$p(g\,|\,f) = N(Hf, \Sigma_e) \quad with \quad \Sigma_e = \sigma_e^2 I \qquad (94)$$

The priors for this case can be summarized as follows:

$$p(f(r)\,|\,z(r) = k) N(m_k, \sigma_k^2), \quad k = 1,...,K \qquad (95)$$

$$p(z) = p(z(r), r \in R) \propto \left[\alpha \sum_{r \in R} \sum_{r \in v(r)} \delta(z(r) - z(s))\right] \qquad (96)$$

where

$$f_k = \{f(r) : r \in R_k\}, \quad R_k = \{r : z(r) = k\}$$

$$p(f_k\,|\,z(r) = k) = N(m_k 1_k, \Sigma_k) \quad with \quad \Sigma_k = \sigma_k^2 I_k$$

$$p(f\,|\,z) = \prod_k N(m_k 1_k, \Sigma_k) = N(m_z, \Sigma_z) \quad with$$

$$m_z = [m_1 1_1', ..., m_K 1_K']', \quad \Sigma_z = diag[\Sigma_1, ..., \Sigma_K]$$

$$p(m_k) = N(m_{k0}, \sigma_{k0}^2) \qquad (97)$$

$$p(\sigma_k^2) = \tau\varsigma(\alpha_{k0}, \beta_{k0}), \quad p(\sigma_e^2) = \tau\varsigma(\alpha_0^e, \beta_0^e)$$

and the posterior probability laws we need to implement an MCMC like algorithm are:

$$p(f\,|\,z,\theta,g) = N(\hat{f}, \hat{\Sigma})$$

With $\hat{\Sigma} = (H^t \Sigma_e^{-1} H + \Sigma_z^{-1})^{-1}$ (98)

And $\hat{f} = \hat{\Sigma}(H^t \Sigma_e^{-1} g + \Sigma_z^{-1} m_z)$

$$p(z\,|\,g,\theta) \propto p(g\,|\,z,\theta) p(z) \qquad (99)$$

$$p(g\,|\,z,\theta) = N(Hm_z, \Sigma_g) \text{ with } \Sigma_g = H\Sigma_z H^t + \Sigma_e)$$

and the posterior probabilities of the hyper-parameters are:

$$p(m_k\,|\,z,f) = N(\mu_k, \upsilon_k^2) \text{ with } \upsilon_k^2 = \left(\frac{n_k}{\sigma_k^2} + \frac{1}{\sigma_{k_0}^2}\right)^{-1}, \quad \mu_k = \upsilon_k^2 \left(\frac{n_k \bar{f}_k}{\sigma_k^2} + \frac{m_{k_0}}{\sigma_{k_0}^2}\right)$$

[Downloaded from jiaeee.com on 2026-11-20]

$$p(\sigma_k^2 \mid f, z) = \tau\varsigma(\alpha_k, \beta_k) \text{ with } \alpha_k = \alpha_{k_0} + \frac{n_k}{2} \text{ and } \beta_k = \beta_{k_0} + \frac{n_k \bar{s}_k}{2}$$

$$\text{where } \bar{f}_k = \frac{1}{n_k} \sum_{r \in R_k} f_i(r), \quad \bar{s}_k = \sum_{r \in R_k} (f(r) - m_k)^2$$

$$p(\sigma_e^2 \mid f, g) = \tau\varsigma(\alpha^e, \beta^e) \text{ with } \alpha^e = \frac{n}{2} + \alpha_0^2 \text{ and } \beta^e = \frac{1}{2}\|g - Hf\|^2 + \beta_0^e$$

n_k = number of pixels in R_k and n = total number of pixels.

Here, we show two examples of simulations: the first in relation with image denoising and the second in relation with image deconvolution. In both cases, we have chosen the same input image $f(r)$. In the first case, we only has added a Gaussian noise and in the second case, we first blurred it with box car PSF of size 7×7 pixels and added a Gaussian noise. Fig. (13) shows the original image, its contours and its regions. Fig. (14) shows the observed noisy image and the results obtained by the proposed method. Remember that, in this method, we have also the estimated contours and region labels as byproducts. Fig. (15) shows the observed blurred and noisy image and the results obtained by the proposed restoration method.

For other inverse problems which can be modeled as a SISO model and where such Bayesian approach has been used refer to [28].

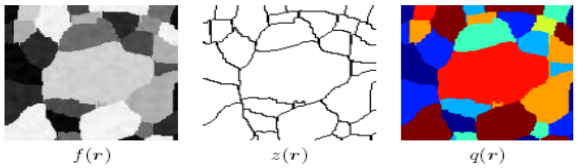


Fig.13: Original image, its contour and its region labels used for image denoising and image restoration

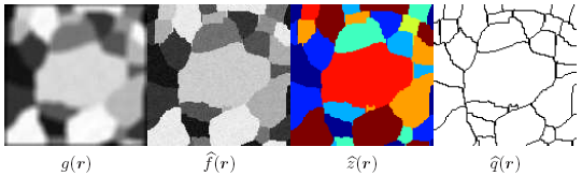


Fig.14: Observed noisy image and the results of the proposed denoising method.

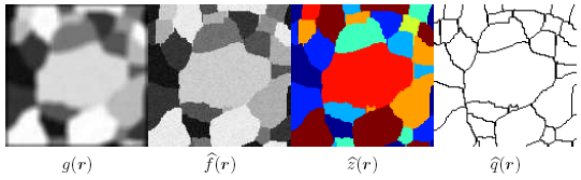


Fig.15: Observed noisy image and the results of the proposed restoration method.

4.2. Registered Images Fusion and Joint Segmentation

Here, each observed image $g_i(r)$ (or equivalently g_i) is assumed to be a noisy version of the unobserved real image $f_i(r)$ (or equivalently f_i)

$$g_i(r) = f_i(r) + e_i(r),\ r \in R, \text{ or } g_i = f_i + e_i, \quad i = 1,...,M \quad (100)$$

which gives

$$p(g_i \mid f_i) = N(f_i, \Sigma_{ei}) \text{ with } \Sigma_{ei} = \sigma_{ei}^2 I \quad (101)$$

and

$$p(\underline{g} \mid \underline{f}) = \prod_i p(g_i \mid f_i) \text{ with} \quad (102)$$

and all the unobserved real images $f_i(r),\ i = 1,...,M$ are assumed to have a common segmentation $z(r)$ (or equivalently z) which is modeled by a discrete value Potts Random Markov Field (PRMF). Then, using the same notations as in previous case, we have the following relations:

$$p(f_i(r) z(r) = k) = N(m_{ik}, \sigma_{ik}^2), \quad k = 1,...,K$$

$$f_{ik} = \{f_i(r) : r \in R_k\}, \quad R_k = \{r : z(r) = k\}$$

$$p(f_{ik} \mid z(r) = k) = N(m_{ik} 1_k, \Sigma_{ik}) \text{ with } \Sigma_{ik} = \sigma_{ik}^2 I_k$$

$$p(z) = p(z(r), r \in R) \propto exp\left[\alpha \Sigma_{r \in R} \Sigma_{s \in v(r)} \delta(z(r) - z(s))\right]$$

$$p(f_i \mid z) = N(m_{zi}, \Sigma_{zi}) \text{ with}$$

$$m_{zi} = [m_{i1} 1_1', ..., m_{iK} 1_K']', \quad \Sigma_{zi} = diag[\Sigma_{i1}, ..., \Sigma_{iK}]$$

$$p(m_{ik}) = N(m_{ik0}, \sigma_{ik0}^2)$$

$$p(\sigma_{ik}^2) = \tau\varsigma(\alpha_{i0}, \beta_{i0}), \quad p(\sigma_{ei}^2) = \tau\varsigma(\alpha_{i0}^2, \beta_{i0}^e)$$

$$p(\underline{f} \mid z) = \prod_i p(f_i \mid z)$$

and all the conditional and posterior probability laws we need to implement the proposed Bayesian methods are summarized here:

$$p(f_i \mid z, \theta_i, g_i) = N(\hat{f}_i, \hat{\Sigma}_i)$$

With

$$\hat{\Sigma}_i = (\Sigma_{ei}^{-1} + \Sigma_z^{-1})^{-1}, \quad \hat{f}_i = \hat{\Sigma}_i (\Sigma_{ei}^{-1} g_i + \Sigma_{zi}^{-1} m_{zi})$$

[Downloaded from jiaeee.com on 2026-11-20]

$$p(z \mid \underline{g}, \theta) \propto \left(\prod_i p(g_i \mid z, \theta_i)\right) p(z(r), r \in R) \text{ with}$$

$$p(g_i \mid z, \theta_i) = N(m_{zi}, \Sigma_{gi}) \text{ with } \Sigma_{gi} = \Sigma_{zi} + \Sigma_{ei}$$

$$p(m_{ik} \mid f_i, z, \sigma_{ik}^2) = N(\mu_{ik}, \upsilon_{ik}^2)$$

$$\mu_{ik} = \upsilon_{ik}^2 \left(\frac{m_{i0}}{\sigma_{i0}^2} + \frac{n_k \bar{f}_{ik}}{\sigma_{ik}^2}\right), \qquad \upsilon_{ik}^2 = \left(\frac{1}{\sigma_{i0}^2} + \frac{n_k}{\sigma_{ik}^2}\right)^{-1}$$

$$p(\sigma_{ik}^2 \mid f_i, z) = \tau\varsigma(\alpha_{ik}, \beta_{ik}) \text{ with } \alpha_{ik} = \sigma_{i0} + \frac{n_k}{2}, \ \beta_{ik} = \beta_{i0} + \frac{\bar{s}_i}{2}$$

$$\text{where} \quad \bar{f}_{ik} = \frac{1}{n_k} \Sigma_{r \in R_k} f_i(r), \qquad \bar{s}_i = \Sigma_{r \in R_k} (f_i(r) - m_{ik})^2$$

$$p(\sigma_{ei}^2 \mid f_i, g_i) = \tau\varsigma(\alpha_i^e, \beta_i^e) \text{ with } \alpha_i^e = \frac{n}{2} + \alpha_{i0}^e, \quad \beta_i^e = \frac{1}{2}\|g_i - f_i\|^2 + \beta_{i0}^e$$

For more details on this model and its application in medical image fusion as well as in image fusion for security systems see [29, 30].

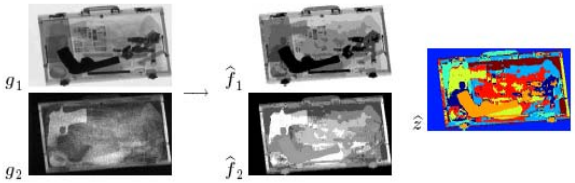


Fig.16: Image fusion and segmentation of two images from a security system measurement.

4.3. Joint Segmentation of Hyper-spectral Images

The proposed model is the same as the model of the previous section except for the last equation of the forward model which assumes that the pixels in similar regions of different images are independent. For hyper-spectral images, this hypothesis is not valid and we have to account for their correlations. This work is under consideration.

4.4. Segmentation of a Video Sequence of Images

Here, we can not assume that all the images in the video sequence have the same segmentation labels. However, we may use the segmentation obtained in an image as an initialization for the segmentation of next image. For more details on this model and to see a typical result see.

4.5. Joint Segmentation and Separation of Instantaneous Mixed Images

Here, the additional difficulty is that we also have to estimate the mixing matrix A. For more details on this model and to see some typical result in joint segmentation and separation of images see [27, 31, 32, 33, 34, 35].

5. Conclusion

In this paper we first showed that many image processing problems can be presented as inverse problems by modeling the relation of the observed image to the unknown desired features explicitly. Then, we presented a very general forward modeling for the observations and a very general probabilistic modeling of images through a hidden Markov modeling (HMM) which can be used as the main basis for many image processing problems such as: 1) simple or multi channel image restoration, 2) simple or joint image segmentation, 3) multi-sensor data and image fusion, 4) joint segmentation of color or hyper-spectral images and 5) joint blind source separation (BSS) and segmentation. Finally, we presented detailed forward models, prior and posterior probability law expressions for the implementation of MCMC algorithms for a few cases of those problems showing typical results which can be obtained using these methods.

REFERENCES

[1] J. Hadamard, “Sur les problmes aux drives partielles et leur signification physique,” Princeton Univ. Bull., vol. 13, 1901.

[2] G. Demoment, “D´econvolution des signaux,” Cours de l’´Ecole sup´erieure d’´electrit´e 3086, 1985.

[3] H. C. Andrews and B. R. Hunt, Digital Image Restoration, Prentice-Hall, Englewood Cliffs, nj, 1977.

[4] B. R. Hunt, “A matrix theory proof of the discrete convolution theorem,” IEEE Trans. Automat. Contr., vol. AC-19, pp. 285–288, 1971.

[5] B. R. Hunt, “A theorem on the difficulty of numerical deconvolution,” IEEE Trans. Automat. Contr., vol. AC-20, pp. 94–95, 1972.

[6] B. R. Hunt, “Deconvolution of linear systems by constrained regression and its relationship to the Wiener theory,” IEEE Trans. Automat. Contr.,vol. AC-17, pp. 703–705, 1972.

[7] A. Mohammad-Djafari, “Binary polygonal shape image reconstruction from a small number of projections,” Elektrik, vol. 5, no. 1, pp. 127–138,1997.

[8] A. Mohammad-Djafari and C. Soussen, “Compact object reconstruction,” in Discrete

[Downloaded from jiaeee.com on 2026-11-20]

Tomography: Foundations, Algorithms and Applications, G. T. Herman and A. Kuba, Eds., chapter 14, pp. 317–342. Birkhauser, Boston, MA, 1999.
[9] A. Mohammad-Djafari, “Bayesian approach with hierarchical markov modeling for data fusion in image reconstruction applications,” in Fusion 2002,
7-11 Jul., Annapolis, Maryland, USA, July 2002.
[10] A. Mohammad-Djafari, “Fusion of x ray and geometrical data in computed tomography for non destructive testing applications,” in Fusion 2002, 7-11 Jul., Annapolis, Maryland, USA, July 2002.
[11] A. Mohammad-Djafari, “Hierarchical markov modeling for fusion of x ray radiographic data and anatomical data in computed tomography,” in Int.
Symposium on Biomedical Imaging (ISBI 2002), 7-10 Jul., Washington DC, USA, July 2002.
[12] A. Mohammad-Djafari, “Fusion bay´esienne de donn´ees en imagerie x et ultrasonore,” in GRETSI 03, France, Sep. 2003.
[13] A. Mohammad-Djafari, “Solving inverses problems: From deterministic to probabilistic approaches,” in Seminar in Electrical Eng. Dept. of Purdue University, in, Dec. 1997.
[14] A. Mohammad-Djafari, N. Qaddoumi, and R. Zoughi, “A blind deconvolution approach for resolution enhancement of near-field microwave images,” in Mathematical modeling, Bayesian estimation and Inverse prob-lems, SPIE 99, Denver, Colorado, USA, F. Prˆeteux, A. Mohammad-Djafari, and E. Dougherty, Eds., 1999, vol. 3816, pp. 274–281.
[15] A. Mohammad-Djafari, J.-F. Giovannelli, G. Demoment, and J. Idier,“Regularization, maximum entropy and probabilistic methods in mass spectrometry data processing problems,” Int. Journal of Mass Spectrometry, vol. 215, no. 1-3, pp. 175–193, Apr. 2002.
[16] M. Nikolova, J. Idier, and A. Mohammad-jafari, “Inversion of largesupport ill-posed linear operators using a piecewise Gaussian mrf,” IEEE Trans. Image Processing, vol. 7, no. 4, pp. 571–585, Apr. 1998.
[17] J. Idier, A. Mohammad-Djafari, and G. Demoment, “Regularization methods and inverse problems: an information theory standpoint,” in 2nd In-ternational Conference on Inverse Problems in Engineering, Le Croisic,France, June 1996, pp. 321–328.
[18] J. Idier, Ed., Approche baysienne pour les problmes inverses, Trait IC2, Srie traitement du signal et de l’image, Herms, Paris, 2001.
[19] J. Idier, “Convex half-quadratic criteria and interacting auxiliary variables for image restoration,” IEEE Trans. Image Processing, vol. 10, no. 7, pp.
1001–1009, July 2001.
[20] J. Idier, Probl`emes inverses en restauration de signaux et d’images, Habilitation diriger des recherches, Universit´e de Paris-Sud, Orsay, France,
July 2000.
[21] H. Snoussi and A. Mohammad-Djafari, “Bayesian source separation with mixture of Gaussians prior for sources and Gaussian prior for mixture coefficients,” in Bayesian Inference and Maximum Entropy Methods, A. Mohammad-Djafari, Ed., Gif-sur-Yvette, France, July 2000, Proc. of
MaxEnt, pp. 388–406, Amer. Inst. Physics.
[22] Hichem Snoussi AND Ali Mohammad-Djafari, “Fast joint separation and segmentation of mixed images,” Journal of Electronic Imaging, vol. 13, no.
2, pp. 349–361, April 2004.
[23] Hichem Snoussi AND Ali Mohammad-Djafari, “Bayesian unsupervised learning for source separation with mixture of gaussians prior,” Journal of VLSI Signal Processing Systems, vol. 37, no. 2/3, pp. 263–279, June/July 2004.
[24] Mahieddine Ichir AND Ali Mohammad-Djafari, “Hidden markov models for blind source separation,” IEEE Trans. on Signal Processing, vol. 15, no. 7, pp. 1887–1899, Jul 2006.
[25] H. Snoussi and A. Mohammad-Djafari, “Information Geometry and Prior Selection.,” in Bayesian Inference and Maximum Entropy Methods,
C. Williams, Ed. MaxEnt Workshops, Aug. 2002, pp. 307–327, Amer. Inst. Physics.
[26] H. Snoussi, Bayesian approach to source separation. Applications in im-agery, Ph.D. thesis, University of Paris–Sud, Orsay, France, september
2003.

[Downloaded from jiaeee.com on 2026-11-20]

[27] H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images," Journal of Electronic Imaging, vol. 13, no. 2, pp.349–361, Apr. 2004.
[28] A. Mohammad-Djafari, "Bayesian approach for inverse problems in optics," in SPIE03, USA, Sep. 2003.
[29] O. F´eron and A. Mohammad-Djafari, "Image fusion and joint segmentation using an MCMC algorithm," Journal of Electronic Imaging, vol. 14, no. 2,pp. paper no. 023014, Apr 2005.
[30] O. F´eron, D. B., and A. Mohammad-Djafari, "Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data," Inverse Problems, vol. 21, no. 6, pp.95–115, Dec 2005.
[31] A. Mohammadpour, O. Feron, and A. Mohammad-Djafari, "Bayesian segmentation of hyperspectral images," in BAYESIAN INFERENCE AND MAXIMUM ENTROPY METHODS IN SCIENCE AND ENGINEER-ING: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering. 2004, vol. 735, pp. 541–548,AIP.
[32] A. Mohammad-Djafari and A. Mohammadpour, "Hyperspectral image processing using a bayesian classification approach," in Proceedings of PSIP 2005, Physics in Signal and Image Processing. 2005, pp. 245–250, PSIP 2005, Physics in Signal and Image Processing.
[33] Nadia Bali AND Ali Mohammad-Djafari, "Joint dimensionality reduction, classification and segmentation of hyperspectral images," in ICIP 2006. Oct. 2006, ICIP06, October 8-11, Atlanta, GA, USA.
[34] Nadia Bali AND Ali Mohammad-Djafari, "Hierarchical markovian models for joint classification, segmentation and data reduction of hyperspectral images," in ESANN 2006. Sep. 2006, ESANN 2006, September 4-8, Belgium.
[35] Nadia Bali AND Ali Mohammad-Djafari, "Hierarchical markovian models for hyperspectral image segmentation," in ICPR 2006. Aug. 2006, ICPR06, Aug. 20-24, Hong Gong.

[Downloaded from jiaeee.com on 2026-11-20]