

Bayesian sparse solutions to linear inverse problems with non stationary noise with student-t priors

Ali Mohammad-Djafari and Mircea Dumitru

*Laboratoire des signaux et systèmes (L2S)
UMR 8506 CNRS-SUPELEC-UNIV PARIS SUD
plateau de Moulon, 3 rue Joliot-Curie, 91192 GIF-SUR-YVETTE Cedex, France*

Abstract

Bayesian approach has become a commonly used method for inverse problems arising in signal and image processing. One of the main advantages of the Bayesian approach is the possibility to propose unsupervised methods where the likelihood and prior model parameters can be estimated jointly with the main unknowns. In this paper, we propose to consider linear inverse problems in which the noise may be non stationary and where we are looking for a sparse solution. To consider both of these requirements, we propose to use Student-t prior model both for the noise of the forward model and the unknown signal or image. The main interest of the Student-t prior model is its Infinite Gaussian Scale Mixture (IGSM) property. Using the resulted hierarchical prior models we obtain a joint posterior probability distribution of the unknowns of interest (input signal or image) and their associated hidden variables. To be able to propose practical methods, we use either a Joint Maximum A Posteriori (JMAP) estimator or an appropriate Variational Bayesian Approximation (VBA) technique to compute the Posterior Mean (PM) values. The proposed method is applied in many inverse problems such as deconvolution, image restoration and computed tomography. In this paper, we show only some results in signal deconvolution and in periodic components determination of some biological signals related to dynamic circadian clock period determination for cancer studies.

Keywords:

Bayesian sparsity enforcing, Variational Bayesian Approximation (VBA), Student-t, Deconvolution, Periodic components estimation, Biological time series processing

1. Introduction

Many linear inverse problems such as signal deconvolution, image restoration, Computed Tomography (CT) image reconstruction, Fourier Synthesis (FS) inversion can be modelled as

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}, \quad (1)$$

where \mathbf{f} represents the unknown quantity: input signal $f(t)$ or original image $f(x, y)$; \mathbf{g} represents the measured data: output signal $g(t)$, blurred image $g(x, y)$ or projections $g_\phi(r)$ in CT; \mathbf{H} represents the forward matrix operator obtained from: the Impulse Response Function (IRF) of the measurement system $h(t)$, the Point Spread Function (PSF) of the imaging system $h(x, y)$ or the geometry of detector-object-detectors in CT; and finally, $\boldsymbol{\epsilon}$ represents the errors [1, 2, 3, 4, 5, 6].

When the system \mathbf{H} is known and we know \mathbf{f} , computing or predicting \mathbf{g} is the *forward problem*. When the input \mathbf{f} and the output \mathbf{g} are known, determining \mathbf{H} is called *identification*. When \mathbf{H} is known and \mathbf{g} is given, estimating \mathbf{f} is called *inversion* and the problem is called *inverse problem*. When \mathbf{H} is partially known, for example when the IRF in signal deconvolution or the PSF in image restoration depend on some unknown parameters, we have myopic or blind deconvolution problems.

Between the classical inverse problems arising in signal and image processing, we mention here a few examples:

– Deconvolution:

$$g(t) = h(t) * f(t) + \epsilon(t) = \int f(\tau) h(t - \tau) d\tau + \epsilon(t), \quad (2)$$

where, when discretized, we obtain the equation (1) and where: the vector \mathbf{f} contains the samples of $f(t)$; the vectors \mathbf{g} and $\boldsymbol{\epsilon}$ contain the samples of $g(t)$ and $\epsilon(t)$ and the $\mathbf{H}\mathbf{f}$ is equivalent to the convolution operation $h(t) * f(t)$. In this case, \mathbf{H} is a Toeplitz matrix which is entirely defined from the samples of the impulse response $h(t)$ denoted by the vector \mathbf{h} .

– Image restoration:

$$g(x, y) = \iint f(x', y') h(x - x', y - y') dx' dy' + \epsilon(x, y), \quad (3)$$

where $h(x, y)$ represents the Point Spread Function (PSF) of the imaging system [1, 2, 3, 4, 7, 5, 6]. Here too, when discretized, we obtain the equation (1) with: \mathbf{f} is a vector containing all the pixels of the image $f(x, y)$

scanned column by column; \mathbf{g} and ϵ contain the pixels of $g(x, y)$ and $\epsilon(x, y)$ and \mathbf{H} is a Toeplitz-Bloc-Toeplitz (TBT) matrix which is entirely defined from the pixels of the PSF $h(x, y)$.

– X ray Computed Tomography (CT) image reconstruction:

$$g(r, \phi) = \iint f(x', y') \delta(r - x \cos \phi - y \sin \phi) dx' dy' + \epsilon(r, \phi), \quad (4)$$

where $g(r, \phi)$ is the projection at angle ϕ and $f(x, y)$ represents the image to reconstruct [8, 9, 10, 11, 12, 13, 14]. Here too, when discretized, we obtain the equation (1) with: \mathbf{f} representing the pixels of the object $f(x, y)$ scanned column by column as in the previous example; \mathbf{g} contains the samples of projections $g(r, \phi)$ organized row by row for different angles and called sinogram and \mathbf{H} is a very sparse matrix which is entirely defined from the geometry of the tomographic system. In a simple straight line model, the elements H_{ij} is the length of the intersection of the ray i in the pixel j .

– Inverse Fourier series signal modelling problem:

$$g(t) = \int f(\nu) \exp [+j2\pi\nu t] d\nu + \epsilon(t), \quad (5)$$

where $g(t)$ represents a time series with periodic components $f(\nu)$. Here again, in the discretized version of equation (1): \mathbf{f} represents values of $f(\nu)$ for different frequencies, \mathbf{g} represents $g(t)$ and \mathbf{H} is the Discrete Fourier Transform (DFT) matrix.

– Fourier Synthesis inverse problem:

$$g(u, v) = \iint f(x', y') \exp [-j2\pi(ux + vy)] dx' dy' + \epsilon(u, v), \quad (6)$$

where $g(u, v)$ is the 2D Fourier Transform (FT) of the objection $f(x, y)$. Here again, in the discretized version of this equation, \mathbf{f} represents $f(x, y)$, \mathbf{g} represents $g(u, v)$ and \mathbf{H} is the 2D DFT matrix [15].

Many other examples can be given in Microwave imaging [16, 17], Ultrasound echography, Seismic imaging, Radio astronomy [18] Fluorescence imaging [100], Inverse scattering [19, 20, 21, 22], Eddy current non destructive testing [23], SAR imaging [24] etc.

In all these examples, the common inverse problem is to estimate \mathbf{f} from the observations of \mathbf{g} . In general, the inverse problems are ill-posed [25]. This

means that, in practice, the data \mathbf{g} alone is not sufficient to define an unique and satisfactory solution. Regularization theory and the Bayesian inversion have been successfully used for this task. See for example [33, 35, 36, 48] for quadratic and Tikhonov regularization, [27, 42, 47] for Total variation, [28, 29, 30, 34] for different entropy based regularization, [32, 40] for L_p and sparsity enforcing, [37, 38, 39, 43] for blind deconvolution and applications, [31, 41] for Cross Validation (CV) and generalized CV methods for determining the regularization parameter, and [26, 54] for Bernoulli-Gaussian models, [44] for Compress Sensing approach, [45, 46] for multichannel blind deconvolution, [49, 50] for nonlinear and space variant PSF, [51, 52] for document image restoration, [53] for joint restoration and segmentation.

In this paper, first we consider the linear inverse problem $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$ where we know that the noise $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_M]'$ is non stationary and that the input is sparse. Accounting for sparsity has been considered in many ways. The first one is by L_0 or L_1 regularization methods [55, 57, 56, 58, 59]. One of these methods which has become now the standard is LASSO [60]. The second way is via Bayesian inference using strict sparsity or sparsity enforcing priors. For the strict sparsity requirement, very often Bernoulli distribution is used: For example, Bernoulli-Gaussian [26, 54, 72], Bernoulli-Laplace [95], Bernoulli-Gamma, etc. For the sparsity enforcing, mainly three categories of priors have been considered and used very often: Generalized Gaussian (GG), Mixture models and heavy tailed probability laws such as Student-t. See [60, 61] and their references for a review of these priors.

To account for the non stationarity of the noise, a zero mean Gaussian model with unknown varying variance has been considered in [62] and a Cauchy-Gaussian model in [63]. To account for both of these two prior information, we propose to model the noise as a zero mean non stationary Gaussian with unknown variances $\{v_{\epsilon_i}, i = 1, \dots, M\}$, on which we assign Inverse Gamma priors and to enforce the sparsity, we propose to use the Student-t prior which is a heavy tailed probability law. The main advantage of Student-t is that, thanks to its Infinite Gaussian Scaled Mixture Model (IGSM) property, it can be used in a hierarchical Gaussian-Gamma model. In this way, in fact both for the non stationarity of the noise and for sparsity enforcing we have the same prior model structure: Gaussian with unknown variances on which we assign Inverse Gamma priors. In our knowledge, this combination is new and first communicated in a conference by the first author [101, 102]. However, the Bayesian framework with different priors both on the noise and on the solution goes back to 1950 with Gaussian [64] or

Poisson [65] for the noise and Gaussian for the solution. But more specific priors and in particular the Markovian model [66, 67], Non Gaussian priors [16] the hierarchical models [68] are more recent. The main difficulties in these methods have been more on the computational aspects. Beside the classical Gaussian approximation [69] and the MCMC methods [70, 71, 72], we may mention the more recent ones: the Approximate Bayesian Computation (ABC) [73, 74, 75, 76], Variational Bayesian Approximation (VBA) [77, 78] and Message Passing (MP) [79, 80, 81, 82] methods.

The rest of this paper is organized as follows: In the next section, the details of the above mentioned prior laws are given and the expression of the joint posterior law of all the unknowns is obtained. Then, successively are presented the Joint Maximum A posteriori (JMAP) and the Variational Bayesian Approximation (VBA) methods and algorithms. A comparison of their computational costs is also given. Some discussions on their theoretical and practical implementations are presented. In the simulation section, we show some results in deconvolution of sparse signals and another application of the proposed method in a study related to biological dynamic circadian cycle studies where the observed time series are modelled by an expansion of very limited number of periodic components. Finally, some conclusions are presented in the last section.

2. Proposed prior laws and the expression of the joint posterior law

Starting from $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon}$, where \mathbf{g} is a vector of the length M and \mathbf{f} is a vector of length N , to account for possible non stationarity of the noise, we propose to use:

$$p(\boldsymbol{\epsilon}|\mathbf{v}_\epsilon) = \mathcal{N}(\boldsymbol{\epsilon}|\mathbf{0}, \mathbf{V}_\epsilon) \quad \text{with} \quad \mathbf{V}_\epsilon = \text{diag}[\mathbf{v}_\epsilon], \quad (7)$$

where $\mathbf{v}_\epsilon = [v_{\epsilon_1}, \dots, v_{\epsilon_M}]'$ contain the unknown variances of the non stationary noise. To be able to estimate them, we assign an Inverse Gamma conjugate prior on v_{ϵ_i} :

$$p(\mathbf{v}_\epsilon) = \prod_i p(v_{\epsilon_i}) \quad \text{with} \quad p(v_{\epsilon_i}) = \mathcal{IG}(v_{\epsilon_i}|\alpha_{\epsilon_0}, \beta_{\epsilon_0}), \forall i. \quad (8)$$

From this, we can define the expression of the likelihood:

$$p(\mathbf{g}|\mathbf{f}, \mathbf{v}_\epsilon) = \mathcal{N}(\mathbf{g}|\mathbf{H}\mathbf{f}, \mathbf{V}_\epsilon) \quad \text{with} \quad \mathbf{V}_\epsilon = \text{diag}[\mathbf{v}_\epsilon]. \quad (9)$$

To account for the sparsity, as mentioned in the previous section, we use the Student-t model:

$$p(\mathbf{f}|\nu) = \prod_{j=1}^N \mathcal{S}(f_j|\nu), \quad (10)$$

where

$$\mathcal{S}(f_j|\nu) = \frac{1}{\sqrt{\pi\nu}} \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)} (1 + f_j^2/\nu)^{-(\nu+1)/2}. \quad (11)$$

Thanks to the Infinite Gaussian Scaled Mixture (IGSM) property of this probability law:

$$\mathcal{S}(f_j|\nu) = \int_0^\infty \mathcal{N}(f_j|0, 1/u_j) \mathcal{G}(u_j|\nu/2, \nu/2) du_j, \quad (12)$$

we propose to use the following hierarchical model

$$p(\mathbf{f}|\mathbf{v}_f) = \mathcal{N}(\mathbf{f}|\mathbf{0}, \mathbf{V}_f) \quad \text{with} \quad \mathbf{V}_f = \text{diag}[\mathbf{v}_f] \quad (13)$$

and

$$p(\mathbf{v}_f) = \prod_j p(v_{f_j}) \quad \text{with} \quad p(v_{f_j}) = \mathcal{IG}(v_{f_j}|\alpha_{f_0}, \beta_{f_0}), \forall j, \quad (14)$$

where $\mathbf{v}_f = [v_{f_1}, \dots, v_{f_N}]'$.

Remark: With this hierarchical model, we have a kind of generalization of the Student-t, called since after IGSM, which is now defined with two parameters:

$$\mathcal{S}(f_j|\alpha, \beta) = \int_0^\infty \mathcal{N}(f_j|0, 1/u_j) \mathcal{G}(u_j|\alpha, \beta) du_j, \quad (15)$$

The mean value of this probability distribution is evidently zero and its variance is given by

$$\text{Var}\{X\} = \int x^2 \mathcal{S}(x|\alpha, \beta) dx = 2\beta \left[\frac{\Gamma(\alpha+1/2)\Gamma(\alpha-1)}{\Gamma(\alpha)\Gamma(\alpha-1/2)} - 1 \right]. \quad (16)$$

For $\alpha = \beta = \nu/2$ we obtain the variance of classical Student-t which is $\frac{\nu}{\nu-2}$. This remark takes its importance when we need to fix the two parameters (α, β) for the initialization of the algorithms. In particular, to ensure the sparsity of the solution, we need to have a probability distribution which is concentrated enough around the zero and has enough heavy tails for the

range of the variation of the quantity of interest. In general, we want to have small values for the variances $v_j = 1/u_j$ when f_j is zero and when f_j has high value, we need to reach high values of v_j which means that the parameters of the Inverse Gamma model of v_j have to be fixed in such a way that its mean is close to zero and its variance in accordance to the dynamic of the variation of desired solution f_j . So, to fix a priori the parameters, we propose to choose

$$\begin{cases} \alpha = 2 + \zeta^2 \\ \beta = \sqrt{v_0} \zeta (1 + \zeta^2) \end{cases} \quad (17)$$

which gives

$$\begin{cases} E\{v_j\} = \frac{\beta}{\alpha-1} = \sqrt{v_0} \zeta \\ \text{Var}\{v_j\} = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)} = v_0 \end{cases} \quad (18)$$

where ζ is a small positive value to insure $\alpha > 2$ and also insures small value for $E\{v_j\}$ and the desired variance v_0 for v_j .

The Figure 1 presents the shape of this model for different values of the two parameters (α, β) compared to the normal distribution.

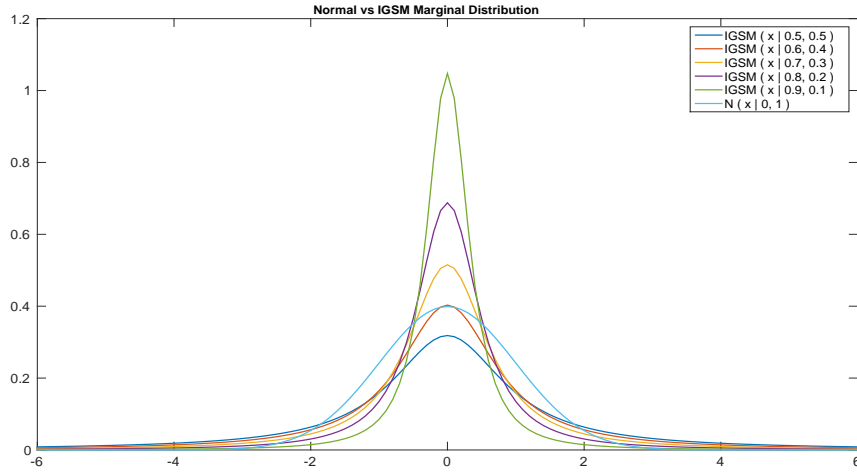


Figure 1: Shape of generalized Student-t for different values of the two parameters (α, β) compared to the normal distribution.

The global generative model described via the equations 1, 7, 8, 9, 13 and 14 is illustrated graphically in the Figure 2.

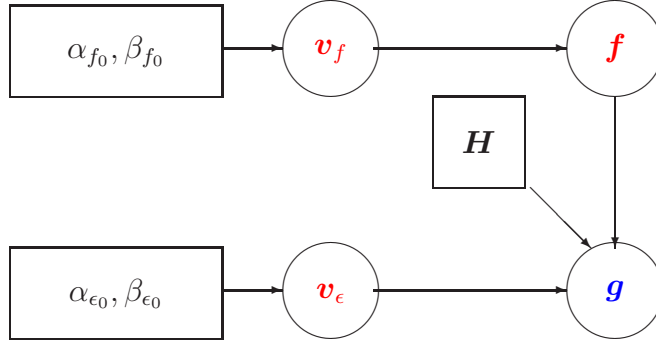


Figure 2: Graphical model linking different variables.

Using these prior laws, the joint posterior law of all the unknowns becomes

$$p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon | \mathbf{g}) \propto \frac{p(\mathbf{g} | \mathbf{v}_f, \mathbf{v}_\epsilon) p(\mathbf{f} | \mathbf{v}_f)}{p(\mathbf{v}_f | \alpha_0, \beta_0) p(\mathbf{v}_\epsilon | \alpha_{\epsilon_0}, \beta_{\epsilon_0})}. \quad (19)$$

From this point, at least two directions can be followed: first one is the JMAP solution:

$$\begin{aligned} (\hat{\mathbf{f}}, \hat{\mathbf{v}}_f, \hat{\mathbf{v}}_\epsilon) &= \arg \max_{(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon)} \{p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon | \mathbf{g})\} \\ &= \arg \min_{(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon)} \{J(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon)\}, \end{aligned} \quad (20)$$

where

$$J(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon) = -\ln p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon | \mathbf{g}). \quad (21)$$

The second is the Variational Bayesian Approximation (VBA) which mainly consists in first approximating $p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon | \mathbf{g})$ by a separable probability law, for example $q(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon) = q_1(\mathbf{f})q_2(\mathbf{v}_f)q_3(\mathbf{v}_\epsilon)$, and then, to use this for defining any estimators $\hat{\mathbf{f}}$, for \mathbf{f} , $\hat{\mathbf{v}}_f$ for \mathbf{v}_f and $\hat{\mathbf{v}}_\epsilon$ for \mathbf{v}_ϵ . In recent years, there was extensive works on VBA in Machine Learning community [83, 84, 85, 86, 87] and in general [88, 89, 90, 91, 92]. However, very few works have been done for inverse problems [93, 94, 95, 96, 97, 103].

In the following sections, we give details of these methods.

3. Joint MAP

The criterion to be optimized is:

$$J_{\text{JMAP}}(\mathbf{f}, \boldsymbol{\theta}) = -\ln p(\mathbf{g} | \mathbf{f}, \mathbf{v}_\epsilon) - \ln p(\mathbf{f} | \mathbf{v}_f) - \ln p(\mathbf{v}_f) - \ln p(\mathbf{v}_\epsilon), \quad (22)$$

which, using the above mentioned priors becomes:

$$\begin{aligned}
& J_{\text{JMAP}}(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon) \\
&= \sum_{i=1}^M \ln v_{\epsilon_i} + \frac{1}{2}(\mathbf{g} - \mathbf{H}\mathbf{f})' \mathbf{V}_\epsilon^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}) \\
&+ \sum_{j=1}^N \ln v_{f_j} + \frac{1}{2}\mathbf{f}' \mathbf{V}_f^{-1} \mathbf{f} \\
&+ \sum_{i=1}^M [(\alpha_{\epsilon_0} + 1) \ln v_{\epsilon_i} + \beta_{\epsilon_0}/v_{\epsilon_i}] \\
&+ \sum_{j=1}^N [(\alpha_{f_0} + 1) \ln v_{f_j} + \beta_{f_0}/v_{f_j}],
\end{aligned} \tag{23}$$

which can also be written as:

$$\begin{aligned}
& J_{\text{JMAP}}(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon) \\
&= \sum_{i=1}^M \left[\ln v_{\epsilon_i} + \frac{1}{2v_{\epsilon_i}}(g_i - [\mathbf{H}\mathbf{f}]_i)^2 \right] \\
&+ \sum_{j=1}^N \left[\ln v_{f_j} + \frac{1}{2v_{f_j}} f_j^2 \right] \\
&+ \sum_{i=1}^M [(\alpha_{\epsilon_0} + 1) \ln v_{\epsilon_i} + \beta_{\epsilon_0}/v_{\epsilon_i}] \\
&+ \sum_{j=1}^N [(\alpha_{f_0} + 1) \ln v_{f_j} + \beta_{f_0}/v_{f_j}].
\end{aligned} \tag{24}$$

One of the basic optimization algorithm for this optimization problem is an alternate optimization with respect to each of the arguments which is detailed in the following. In fact, when \mathbf{v}_ϵ and \mathbf{v}_f are fixed, the criterion as a function of \mathbf{f} is a quadratic one:

$$J_0(\mathbf{f}) = \frac{1}{2}(\mathbf{g} - \mathbf{H}\mathbf{f})' \mathbf{V}_\epsilon^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}) + \frac{1}{2}\mathbf{f}' \mathbf{V}_f^{-1} \mathbf{f}, \tag{25}$$

which has an analytical solution:

$$\mathbf{f} = (\mathbf{H}' \mathbf{V}_\epsilon^{-1} \mathbf{H} + \mathbf{V}_f^{-1})^{-1} \mathbf{H}' \mathbf{V}_\epsilon^{-1} \mathbf{g}. \tag{26}$$

where $\mathbf{V}_\epsilon = \text{diag}[\mathbf{v}_\epsilon]$ and $\mathbf{V}_f = \text{diag}[\mathbf{v}_f]$.

When \mathbf{f} is fixed, the criterion is separable in v_{ϵ_i} and in v_{f_j} and we obtain easily the expressions of the minimizers by putting equal to zero its derivative with respect to each v_{ϵ_i} or v_{f_j} :

$$v_{\epsilon_i} = \frac{\beta_{\epsilon_i}}{\alpha_{\epsilon_i}} \text{ with: } \alpha_{\epsilon_i} = \alpha_{\epsilon_0} + 3/2, \beta_{\epsilon_i} = \beta_{\epsilon_0} + \frac{1}{2}(g_i - [\mathbf{H}\mathbf{f}]_i)^2 \tag{27}$$

and

$$v_{f_j} = \frac{\beta_{f_j}}{\alpha_{f_j}} \text{ with: } \alpha_{f_j} = \alpha_{f_0} + 3/2, \beta_{f_j} = \beta_{f_0} + \frac{1}{2}(f_j)^2 \tag{28}$$

These relations are summarized in the following algorithm:

JMAP alternate optimization algorithm:

Initialization: $\alpha_{\epsilon_0}, \beta_{\epsilon_0}, \alpha_{f_0}, \beta_{f_0} \rightarrow v_{\epsilon}^{(0)} = \frac{\beta_{\epsilon_0}}{\alpha_{\epsilon_0}-1}, v_f^{(0)} = \frac{\beta_{f_0}}{\alpha_{f_0}-1}$

$\mathbf{v}_{\epsilon}^{(0)} = v_{\epsilon}^{(0)} \mathbf{1}, \mathbf{v}_f^{(0)} = v_f^{(0)} \mathbf{1},$

$\mathbf{V}_{\epsilon} = \text{diag} [\mathbf{v}_{\epsilon}^{(0)}], \mathbf{V}_f = \text{diag} [\mathbf{v}_f^{(0)}]$

Iterations:

Step 1:

$$\mathbf{f} = (\mathbf{H}'\mathbf{V}_{\epsilon}^{-1}\mathbf{H} + \mathbf{V}_f^{-1})^{-1}\mathbf{H}'\mathbf{V}_{\epsilon}^{-1}\mathbf{g}$$

Step 2:

$v_{\epsilon_i} = \frac{\beta_{\epsilon_i}}{\alpha_{\epsilon_i}}$ with:

$$\alpha_{\epsilon_i} = \alpha_{\epsilon_0} + 3/2, \quad \beta_{\epsilon_i} = \beta_{\epsilon_0} + \frac{1}{2}(\mathbf{g}_i - [\mathbf{H}\mathbf{f}]_i)^2$$

$\mathbf{V}_{\epsilon} = \text{diag} [\mathbf{v}_{\epsilon}]$

Step 3:

$v_{f_j} = \frac{\beta_{f_j}}{\alpha_{f_j}}$ with:

$$\alpha_{f_j} = \alpha_{f_0} + 3/2, \quad \beta_{f_j} = \beta_{f_0} + \frac{1}{2}(f_j)^2$$

$\mathbf{V}_f = \text{diag} [\mathbf{v}_f]$

(29)

We may note that, in fact, the implementation of this algorithm does not need any matrix inversion because the computation of $\mathbf{f} = (\mathbf{H}'\mathbf{V}_{\epsilon}^{-1}\mathbf{H} + \mathbf{V}_f^{-1})^{-1}\mathbf{H}'\mathbf{V}_{\epsilon}^{-1}\mathbf{g}$ in Step 1 can be done via the optimization of the following quadratic criterion:

$$J_0(\mathbf{f}) = \frac{1}{2}(\mathbf{g} - \mathbf{H}\mathbf{f})'\mathbf{V}_{\epsilon}^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}) + \frac{1}{2}\mathbf{f}'\mathbf{V}_f^{-1}\mathbf{f}, \quad (30)$$

which can be done by any appropriate gradient based algorithm. This step is however very important in particular for high dimensional data. In this paper we do not focus much more on this point. In practice, we used either a steepest gradient descent or conjugate gradient algorithms.

As we will see later, the main advantage of this approach is its low computational cost. The main drawback is in the fact that, at each iteration, the uncertainties associated to the output of each step are not accounted for. Also, theoretically, the JMAP estimation may not have all the necessary good characteristics of a Bayesian approach because it corresponds to the mode of the posterior. Theoretically, a better estimator is the Posterior Mean (PM). Its exact computation needs huge dimensional integration which

has only an analytical solution in the Gaussian case. In general, its approximate estimate can be done by MCMC methods which are very intensive in cost. Variational Bayesian Approximation (VBA) methods are alternatives to MCMC which can theoretically give posterior mean estimates with lower computational costs than MCMC methods. The main steps of this approach are presented in the next subsection.

4. Variational Bayesian Approximation (VBA)

VBA mainly consists in first approximating $p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon | \mathbf{g})$ by a separable probability law, for example $q(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon) = q_1(\mathbf{f})q_2(\mathbf{v}_f)q_3(\mathbf{v}_\epsilon)$, and then, to use this for defining any estimators $\hat{\mathbf{f}}$, for \mathbf{f} , $\hat{\mathbf{v}}_f$ for \mathbf{v}_f and $\hat{\mathbf{v}}_\epsilon$ for \mathbf{v}_ϵ . The main steps to find q is to use the Kullback-Leibler divergence $\text{Kl}(q : p)$ and to optimize it to find the expressions of $q_1(\mathbf{f})$, $q_2(\mathbf{v}_f)$ and $q_3(\mathbf{v}_\epsilon)$. Using an alternate optimization technique, we obtain:

$$\begin{cases} q_1(\mathbf{f}) & \propto \exp[\langle \ln p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon, \mathbf{g}) \rangle_{q_2 q_3}] \\ q_2(\mathbf{v}_f) & \propto \exp[\langle \ln p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon, \mathbf{g}) \rangle_{q_1 q_3}] \\ q_3(\mathbf{v}_\epsilon) & \propto \exp[\langle \ln p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon, \mathbf{g}) \rangle_{q_1 q_2}] \end{cases} \quad (31)$$

To obtain the expressions of $q_1(\mathbf{f})$, $q_2(\mathbf{v}_f)$ and $q_3(\mathbf{v}_\epsilon)$, we need to compute $\langle \ln p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon, \mathbf{g}) \rangle$ with respect to q_1 , q_2 and q_3 . Looking at the expression of $J(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon) = \ln p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\epsilon, \mathbf{g})$ in our case, we find very easily that if we choose $q_1(\mathbf{f})$ to be Gaussian, $q_2(\mathbf{v}_f)$ and $q_3(\mathbf{v}_\epsilon)$ to be products of Inverse Gamma, during the iterations these families are conserved due to the conjugate properties of Gamma and Gaussian:

$$\begin{cases} q_1(\mathbf{f}) & = \mathcal{N}(\mathbf{f} | \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f) \\ q_2(\mathbf{v}_f) & = \prod_j \mathcal{IG}(v_{f_j} | \alpha_{f_j}, \beta_{f_j}) \\ q_3(\mathbf{v}_\epsilon) & = \prod_i \mathcal{IG}(v_{\epsilon_i} | \alpha_{\epsilon_i}, \beta_{\epsilon_i}) \end{cases} \quad (32)$$

We then need to find appropriate update relations between the parameters $(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$, $(\alpha_{f_j}, \beta_{f_j})$ and $(\alpha_{\epsilon_i}, \beta_{\epsilon_i})$.

The details of these steps are given in the Appendix. The resulting algorithm is the following:

VBA alternate optimization algorithm:

Initialization: $\alpha_{\epsilon_0}, \beta_{\epsilon_0}, \alpha_{f_0}, \beta_{f_0} \rightarrow \mathbf{v}_\epsilon^{(0)} = \frac{\beta_{\epsilon_0}}{\alpha_{\epsilon_0}-1}, \quad \mathbf{v}_f^{(0)} = \frac{\beta_{f_0}}{\alpha_{f_0}-1}$

$\mathbf{v}_\epsilon^{(0)} = v_\epsilon^{(0)} \mathbf{1}, \mathbf{v}_f^{(0)} = v_f^{(0)} \mathbf{1},$
 $\mathbf{V}_\epsilon = \text{diag} [\mathbf{v}_\epsilon^{(0)}], \mathbf{V}_f = \text{diag} [\mathbf{v}_f^{(0)}]$

Iterations:

Step 1:

$q_1(\mathbf{f}) = \mathcal{N}(\mathbf{f} | \boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$ with:
 $\boldsymbol{\Sigma}_f = (\mathbf{H}' \mathbf{V}_\epsilon^{-1} \mathbf{H} + \mathbf{V}_f^{-1})^{-1}$
 $\boldsymbol{\mu}_f = \boldsymbol{\Sigma}_f \mathbf{H}' \mathbf{V}_\epsilon^{-1} \mathbf{g}$
 $\mathbf{f} = \boldsymbol{\mu}_f$

Step 2:

$q_2(v_{\epsilon_i}) = \mathcal{IG}(v_{\epsilon_i} | \alpha_{\epsilon_i}, \beta_{\epsilon_i})$ with: (33)

$\alpha_{\epsilon_i} = \alpha_{\epsilon_0} + \frac{1}{2}$
 $\beta_{\epsilon_i} = \beta_{\epsilon_0} + \frac{1}{2} \langle |g_i - [\mathbf{H} \mathbf{f}]_i|^2 \rangle$
with: $\langle \|\mathbf{g} - \mathbf{H} \mathbf{f}\|^2 \rangle = \|\mathbf{g} - \mathbf{H} \boldsymbol{\mu}_f\|^2 + \text{Tr} \{ \mathbf{H} \boldsymbol{\Sigma}_f \mathbf{H}' \}$
 $\langle v_{\epsilon_i} \rangle = \frac{\beta_{\epsilon_i}}{\alpha_{\epsilon_i}-1}$
 $\mathbf{V}_\epsilon = \text{diag} [\langle \mathbf{v}_\epsilon \rangle]$

Step 3:

$q_3(v_{f_j}) = \mathcal{IG}(v_{f_j} | \alpha_{f_j}, \beta_{f_j})$ with:
 $\alpha_{f_j} = \alpha_{f_0} + \frac{1}{2}$
 $\beta_{f_j} = \beta_{f_0} + \frac{1}{2} \langle |\mathbf{f}_j|^2 \rangle$
with: $\langle \|\mathbf{f}\|^2 \rangle = \|\boldsymbol{\mu}_f\|^2 + \text{Tr} \{ \boldsymbol{\Sigma}_f \}$
 $\langle v_{f_j} \rangle = \frac{\beta_{f_j}}{\alpha_{f_j}-1}$
 $\mathbf{V}_f = \text{diag} [\langle \mathbf{v}_f \rangle]$

We may note that, here, a costly step is the computation of $\boldsymbol{\Sigma}_f = (\mathbf{H}' \mathbf{V}_\epsilon^{-1} \mathbf{H} + \mathbf{V}_f^{-1})^{-1}$. However, we only need its diagonal elements for the computation of

$$\langle \|\mathbf{f}\|^2 \rangle = \|\boldsymbol{\mu}_f\|^2 + \text{Tr} \{ \boldsymbol{\Sigma}_f \} \quad (34)$$

in Step 3 and the computation of

$$\langle \|\mathbf{g} - \mathbf{H} \mathbf{f}\|^2 \rangle = \|\mathbf{g} - \mathbf{H} \boldsymbol{\mu}_f\|^2 + \text{Tr} \{ \mathbf{H} \boldsymbol{\Sigma}_f \mathbf{H}' \} \quad (35)$$

in Step 2.

If we can decompose $\Sigma_f = D'D$, then

$$\text{Tr} \{ \Sigma_f \} = \|D\mathbf{1}\|^2 \text{ and } \text{Tr} \{ H\Sigma_f H' \} = \|HD\mathbf{1}\|^2, \quad (36)$$

where $\mathbf{1}$ is a vector of ones entries. This hint gives the possibility to compute quantities needed in different steps in this algorithm.

5. Comparison between JMAP and VBA

Looking in the details on the two methods and as illustrated in the Figure 3, we see that:

In JMAP, during the iterations only the values are transmitted between the different steps without accounting for the uncertainties.

In VBA, during the iterations the probability laws are transmitted between the different steps thus accounting for uncertainties. In particular, we see that in steps 2 and 3, not only the value of (which is the expected value of $q(\mathbf{f})$) is transmitted but also its covariance matrix Σ_f . This process has many similarities with message passing methods. However, the computation of this covariance matrix is the main extra cost of VBA with respect to JMAP.

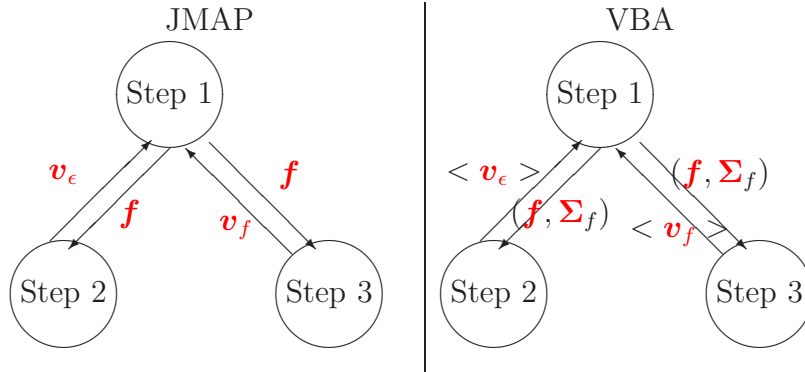


Figure 3: Comparison between JMAP and VBA;

From the theoretically point of view, we can not say too much about the convergence properties of these algorithms. The proposed JMAP is an alternate optimisation algorithm and so its convergence depends mainly on the convexity of the JMAP criterion with respect to all of its arguments. The proposed VBA is also an alternate optimisation algorithm in the space

of probability density functions. However, in the space of parameters, as mentioned before, the fact that the covariance matrix of \mathbf{f} is used in steps 2 and 3 makes a better theoretical property. A rigorous mathematical proof is not easy. However, in practice, both algorithms converge to a local minimum. The initialization is then important.

For initialization, we need to choose appropriately the hyper-parameters $(\alpha_{\epsilon_0}, \beta_{\epsilon_0})$ in such a way that \mathbf{v}_ϵ be fixed according to a reasonable prior knowledge of the variance of the noise and its variability along the observation process and $(\alpha_{f_0}, \beta_{f_0})$ in such a way that the corresponding prior (student-t) be well concentrated around the zero with heavy tailed scaled to the dynamic of the seeking solution. These points are discussed more extensively in the second application of the proposed method for periodic components estimation.

6. Simulations

We implemented these two algorithms for many linear inverse problems. Here, we give two examples in signal processing which are the deconvolution and periodic components estimation of biological time series problem. In both cases, the main idea is first to simulate sparse inputs \mathbf{f} and then generate data $\mathbf{g} = \mathbf{H}\mathbf{f} + \epsilon$ with different SNR and different realization of the noise ϵ . Then, applying different reconstruction algorithms, compute the estimate $\hat{\mathbf{f}}$ and compare with \mathbf{f} , for example using the following quantitative relative distances:

$$\delta_p f = \frac{\|\hat{\mathbf{f}} - \mathbf{f}\|_p^p}{\|\mathbf{f}\|_p^p}, \quad p = 1, 2. \quad (37)$$

$\delta_2 f$ is the Normalized Mean Square Error (NMSE) and $\delta_1 f$ is the Normalized Mean Absolute Error (NMAE). This simulation protocol is illustrated in Figure 4.

Also, when the original signal \mathbf{f} is sparse, we can try to obtain a sparse solution by thresholding $\hat{\mathbf{f}}$ to be able, for example, to give the Missing Values (MV):

$$\text{MV}(\hat{\mathbf{f}}, \mathbf{f}) = \begin{array}{l} \# \text{ of non zero value components in } \mathbf{f} \\ \text{not present in } \hat{\mathbf{f}} \end{array} \quad (38)$$

and the False Alarms (FA):

$$\text{FA}(\hat{\mathbf{f}}, \mathbf{f}) = \begin{array}{l} \# \text{ of zero value components in } \mathbf{f} \\ \text{which is not zero value in } \hat{\mathbf{f}} \end{array} \quad (39)$$

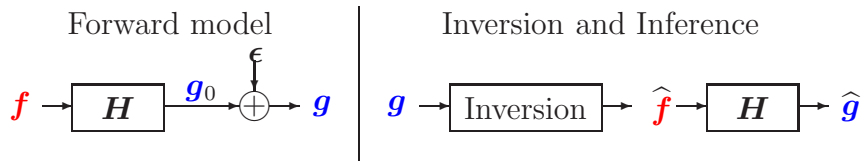


Figure 4: Forward model for simulation and Inversion or Inference. In simulation we can compare $\hat{\mathbf{f}}$ with \mathbf{f} and $\hat{\mathbf{g}}$ with \mathbf{g}_0 or \mathbf{g} . In real case, we do not have access to \mathbf{f} , we can only compare $\hat{\mathbf{g}}$ with \mathbf{g} .

as a measure of performances for the algorithm.

These performance measures can be computed in simulation, but for real applications these are not possible. One can then, compute $\hat{\mathbf{g}} = \mathbf{H}\hat{\mathbf{f}}$ and compare with \mathbf{g} using the following relative distances:

$$\delta_p g = \frac{\|\mathbf{g} - \hat{\mathbf{g}}\|_p^p}{\|\mathbf{g}\|_p^p}, \quad p = 1, 2. \quad (40)$$

$\delta_2 g$ is the Normalized Mean Square Residual Error (NMSRE) and $\delta_1 g$ is the Normalized Mean Absolute Residual Error (NMARE).

The whole protocol of the forward simulation and inversion and possibilities of comparison are illustrated in the Figure 4.

In simulation, methods with which reach lower values for NMSE, NMAE as well as MV and FA are preferred. In real case, we cannot have access to these quantities. We may then reject the methods which give high values for NMSRE and NMARE. However, having very low values for these quantities (over fitting) does not forcibly mean that the method is good. We may use Cross validation methods in this case.

In all these simulations, there are other quantities which are in general important to monitor:

- The evolution of the different quantities such as f_j , v_{f_j} and v_{ϵ_i} during the iterations;
- final values of v_{ϵ} and v_f ;
- final values of $\boldsymbol{\mu}_f$ and $\text{diag}[\boldsymbol{\Sigma}_f]$ which can be interpreted as the posterior means and variances of the solution.

In the following, we present two examples, one for a deconvolution and another for periodic components estimation of time series.

6.1. Deconvolution

For the deconvolution problem, we simulated an input $f(t)$, an Impulse Response Function (IRF) $h(t)$ and computed the ideal output $g_0(t) = h(t) * f(t)$ on which we added a non stationary noise $\epsilon(t)$ with slowly varying variances v_{ϵ_i} to obtain the simulated data $g(t) = g_0(t) + \epsilon(t)$. Figure 5 shows these signals.

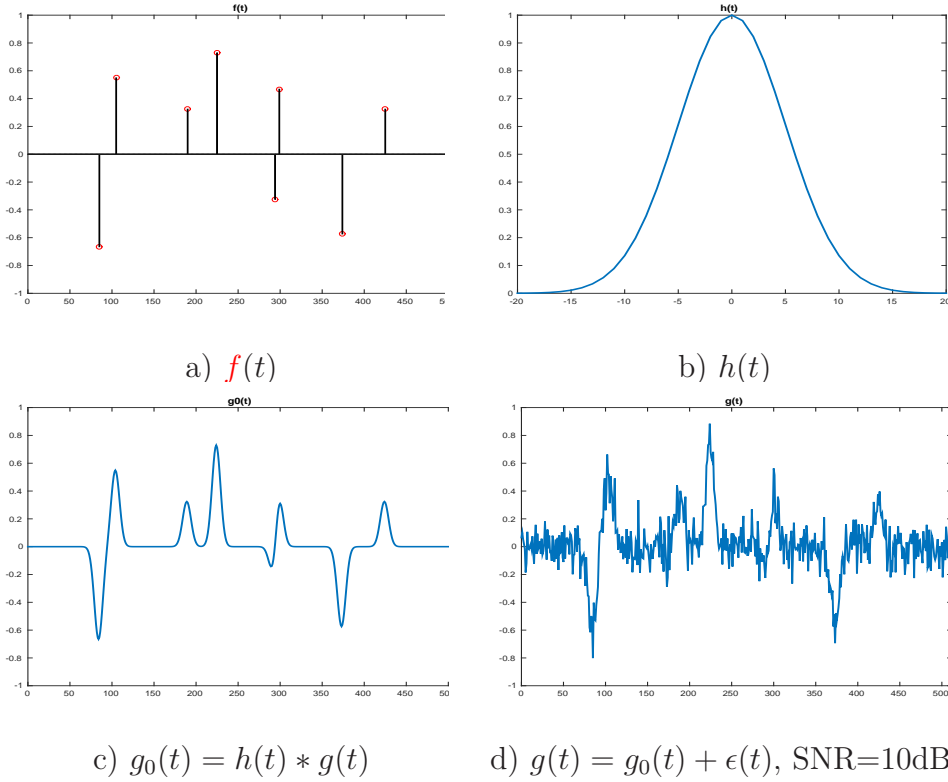


Figure 5: Data generation: a) Input $f(t)$, b) IRF $h(t)$, c) output $g_0(t) = h(t) * f(t)$ and d) noisy data $g(t) = g_0(t) + \epsilon(t)$.

From these data, we applied the JMAP and VBA algorithms and compared the results. As a reference, the results obtained by Lasso is also given. In Figure 6, we can see one of such results.

To show some other relative performances of of the proposed methods compared to Lasso, which is the most concurrent in deterministic methods, we present the Normalized Mean Absolute Errors (NMAE) and Normalized Mean Square Errors (NMSE) between f and \hat{f} and the Normalized Mean

Absolute Residual Errors (NMARE) and Normalized Mean Square Residual Errors (NMSRE) between \mathbf{g} and $\hat{\mathbf{g}}$ as a function of SNR. These results are presented in Figure 7.

As we can see, the performances depend on the criterion. To show the main advantage of the proposed method, we simulated the case where the noise variance is changing during the measurement process. As the proposed method is designed to estimate the noise variance, the performances are surely better than the methods which can not do that. This is shown in Figure 8.

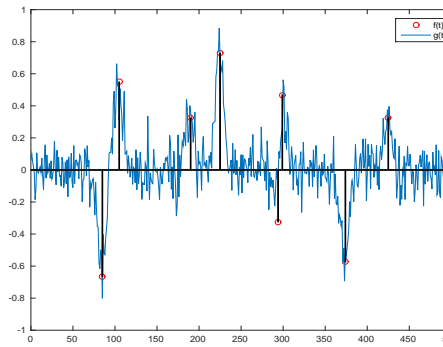
For this non-stationary case, also, the Figure 9 shows the NMSE, NMAE, NMSRE, NMARE, MFA and MV as in Figure 7 for the stationary case.

We did many other extensive simulations comparing the performances of these two proposed algorithms compared to more classical regularization based methods and in particular with L_2 or L_1 regularization criteria. In general, the results with L_2 regularization are not sparse, but those with L_1 are as good as the proposed method. However, in regularization methods, the results depend on the regularization parameter. Even if there are methods based on cross validation which can give a good value to use, but there is no easy way to measure the remaining uncertainty of the computed solution.

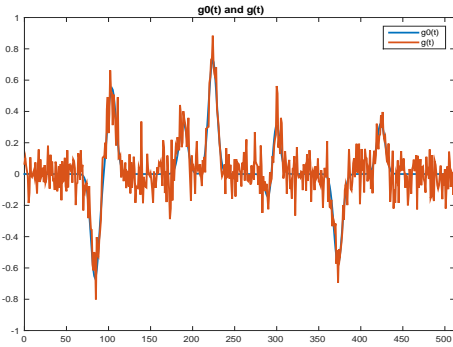
In the proposed Bayesian method, not only we have an unsupervised method, but also we have the possibilities to quantify the uncertainties, for example to put error bar on the solution. A typical result obtained by VBA is given in Figure 10. In this figure, the results are shown with error bars using the estimated diagonal elements of the posterior covariance matrix $\hat{\Sigma}$. First row shows the results for the stationary noise case and the second row for the non-stationary noise case.

As a final conclusion for these simulations, we see the superiority of the Bayesian approach, mainly in three main points:

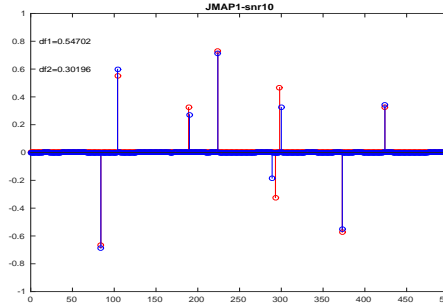
- i) the possibility to easily account for the non stationarity of the noise;
- ii) the possibility of estimating noise variances;
- ii) the possibility of putting error bars on the solution.



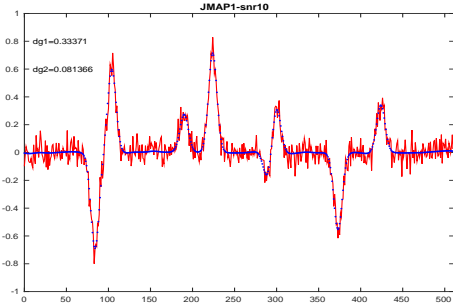
Original: a) $f(t)$ and $g(t)$



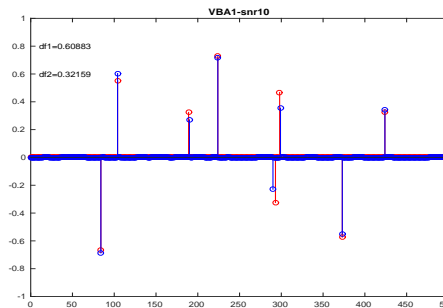
b) $g_0(t)$ and $g(t)$



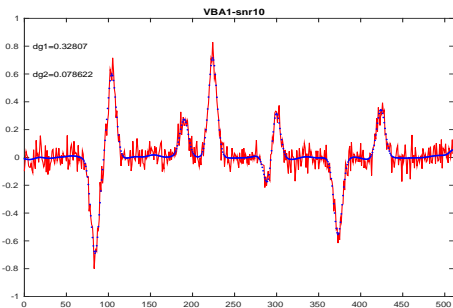
JMAP: c) $f(t)$ and $\hat{f}(t)$



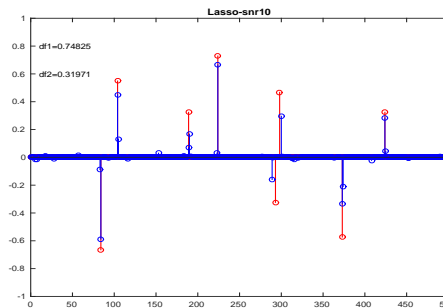
d) $g(t)$ and $\hat{g}(t)$



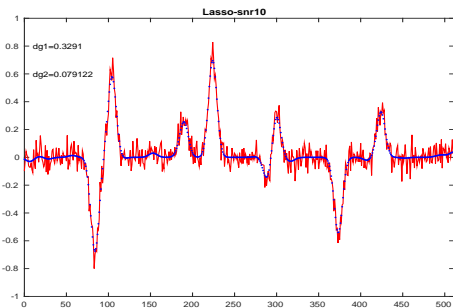
VBA: e) $f(t)$ and $\hat{f}(t)$



f) $g(t)$ and $\hat{g}(t)$



Lasso: g) $f(t)$ and $\hat{f}(t)$



18 h) $g(t)$ and $\hat{g}(t)$

Figure 6: A typical result obtained by JMAP, VBA which are compared with the results obtained by Lasso: First row: a) original $f(t)$, $g_0(t)$ and b) $g(t)$. Second row: c) $f(t)$ and $\hat{f}(t)$ and d) $g(t)$ and $\hat{g}(t)$ obtained by JMAP. Third row: e) $f(t)$ and $\hat{f}(t)$ and f) $g(t)$ and $\hat{g}(t)$ obtained by JMAP. Fourth row: g) $f(t)$ and $\hat{f}(t)$ and h) $g(t)$ and $\hat{g}(t)$ obtained by Lasso with optimal CV value for the regularization parameter. The relative distances between $f(t)$ and $\hat{f}(t)$ and between $g(t)$ and $\hat{g}(t)$ are given as the measures of performances.

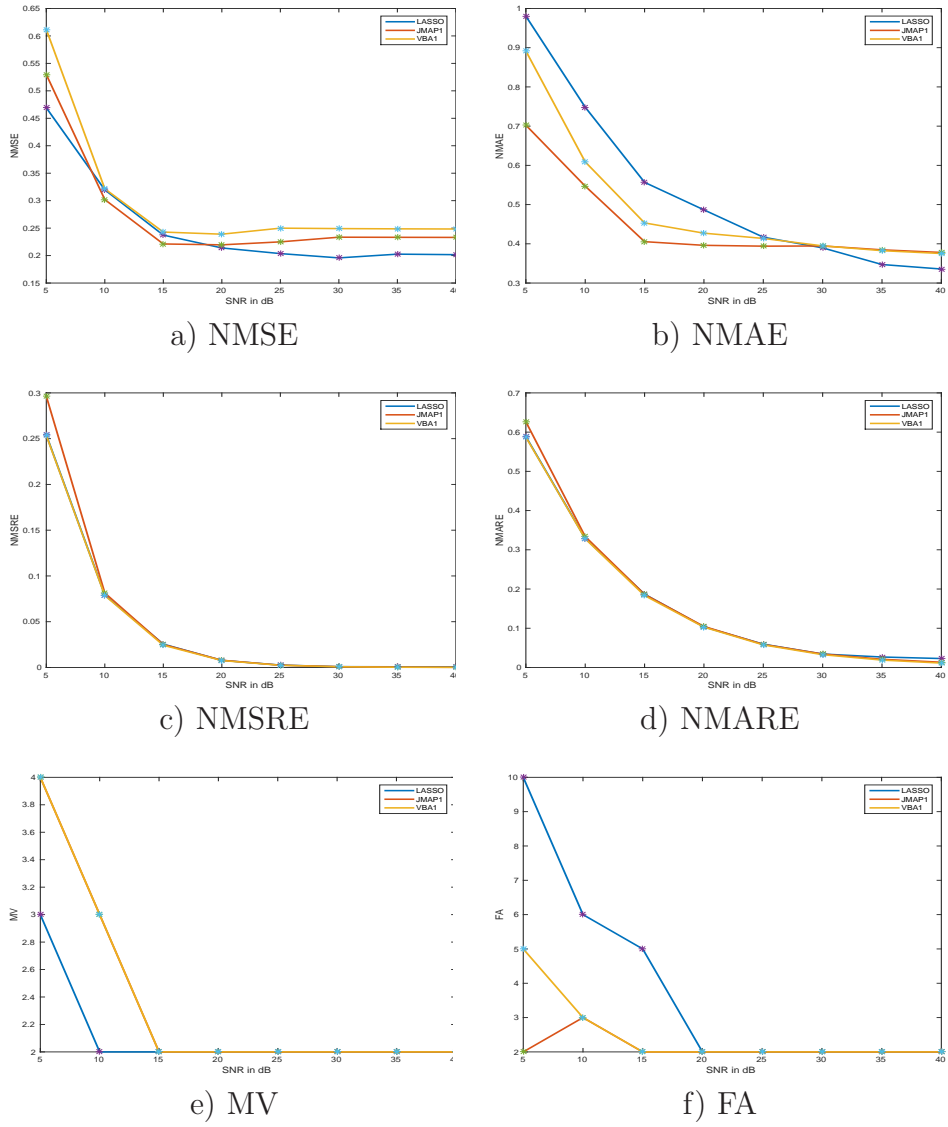


Figure 7: The relative distances between $f(t)$ and $\hat{f}(t)$ (NMSE and NMAE) and between $g(t)$ and $\hat{g}(t)$ (NMSRE and NMARE) and the number of Missing Values (MV) and the number of False Alarms (FA) when comparing $f(t)$ and $\hat{f}(t)$ are given for LASSO and the two proposed methods: JMAP and VBA.

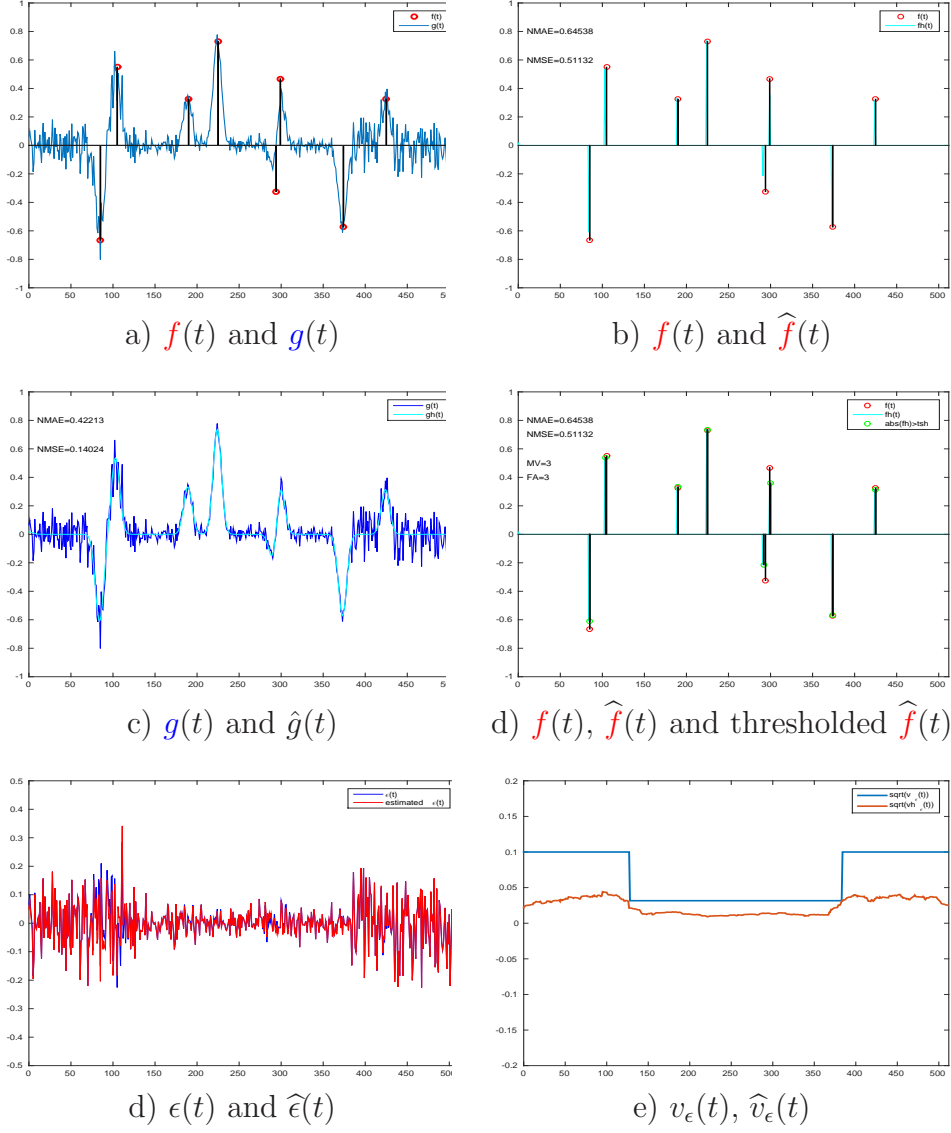


Figure 8: The results in this figure have the same legends as in Figure 6 but for the case where the noise is non stationary. The non stationarity of the noise is simulated by the changes of its variance during the measurements process. Noise level change is shown in d) and its standard deviation estimated values by the proposed VBA method is given in e).

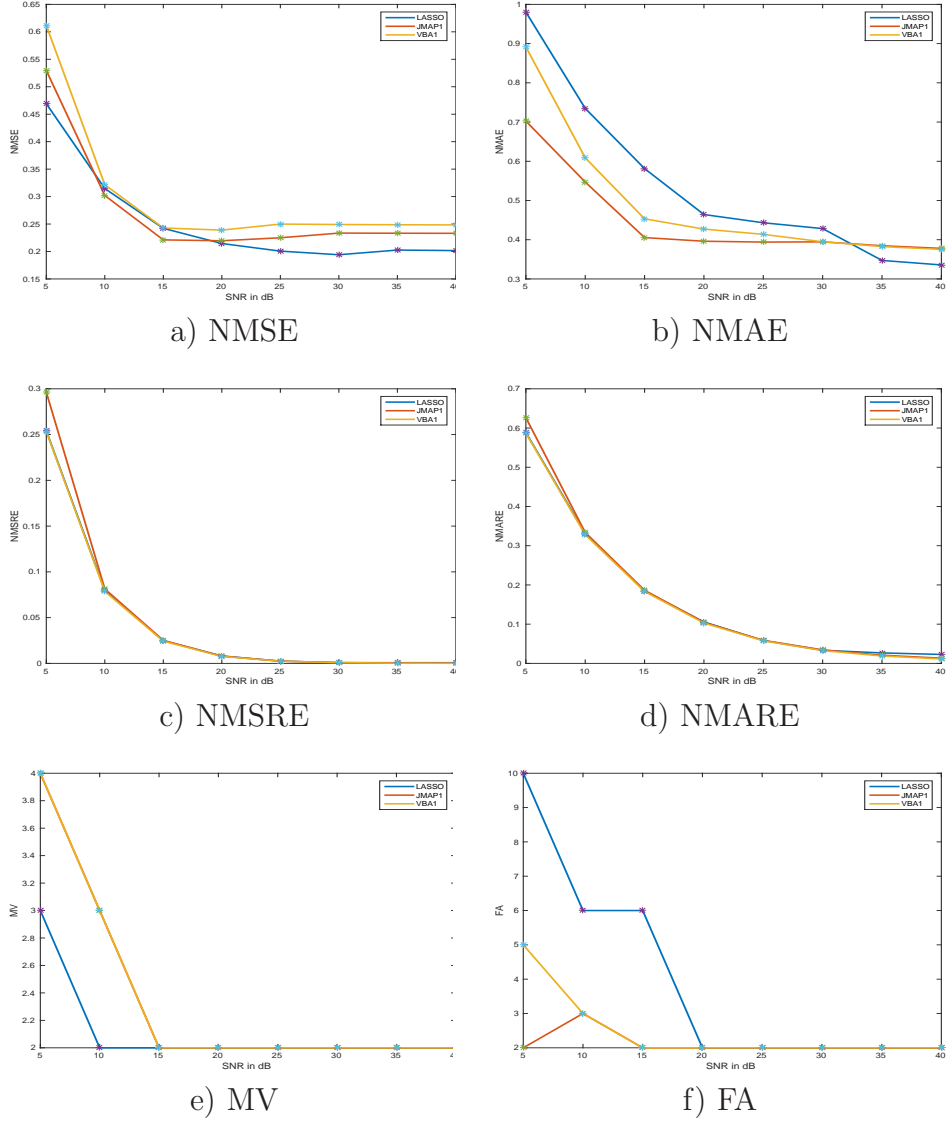


Figure 9: The results in this figure have the same legends as in Figure 7 but for the case where the noise is non stationary. The relative distances between $f(t)$ and $\hat{f}(t)$ (NMSE and NMAE) and between $g(t)$ and $\hat{g}(t)$ (NMSRE and NMARE) and the number of missing values (MV) and the number of false alarms (FA) when comparing $f(t)$ and $\hat{f}(t)$ are given for LASSO and the two proposed methods: JMAP and VBA.

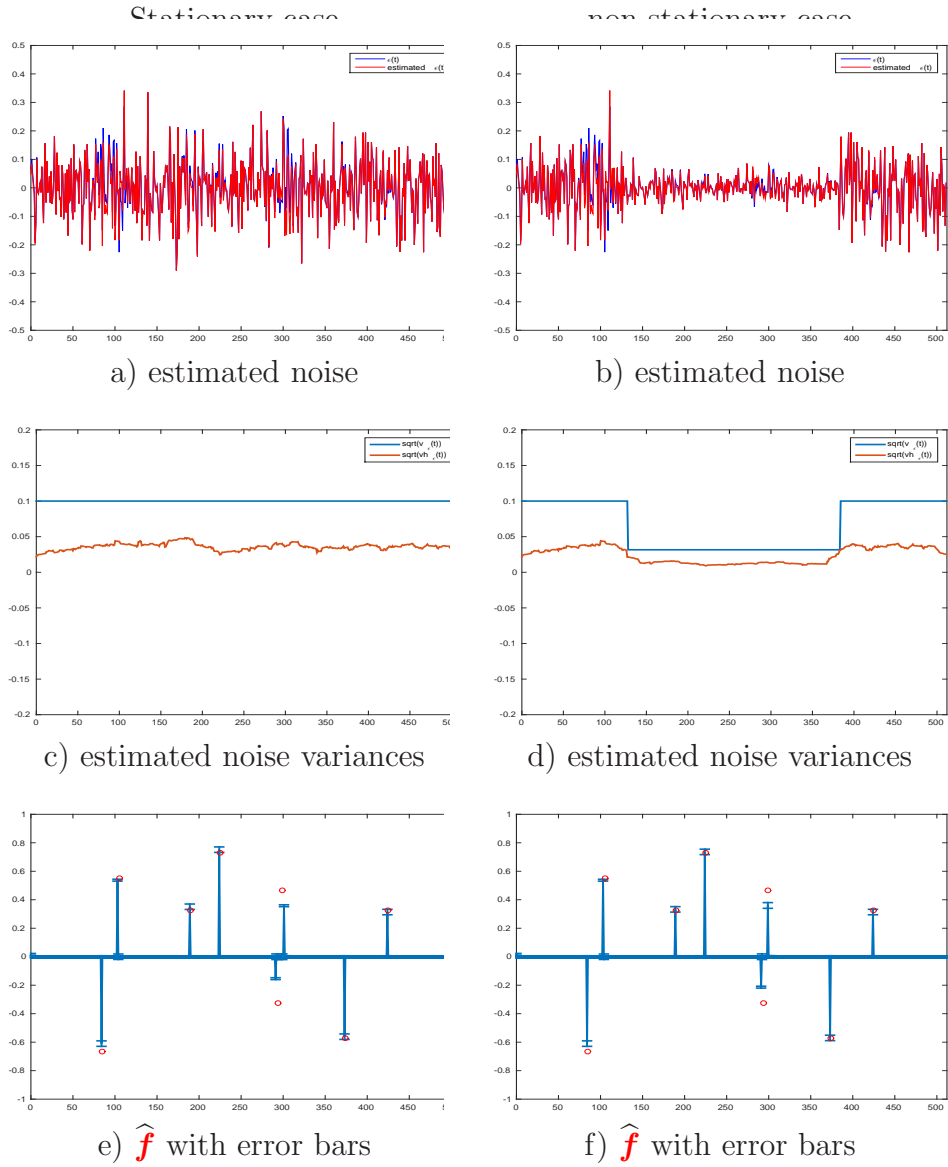


Figure 10: Stationary (left column) and non-stationary (right column) noise cases: First row: original and estimated noise, Second row: original and estimated noise variances Third row: reconstruction results with their associated error bars ($\pm 3\sigma$).

6.2. Periodic components estimation in biological time series

For the periodic components estimation of biological time series, here we show some simulation results modelling time series issued in cancer treatment experiments in the chronobiological context. The time series considered represents the photon absorption of mice inoculated with cancer and due to the tumor growth the time series presents some particularities: short length (10 days) and increasing trend. The challenge is to be able to decide the stability or instability of the periodic components, having as a prior information the sparse structure of the periodic component vector, i.e. the reduced number of clocks expressed. Since the circadian period is defined at 24 h, in order to analyse the stability or instability of the periodic components, we need a method that can analyse very short signals relative to the circadian period (4 days signals) and offer informations of the periods in a specific range with a precision of one hour. More precisely, for such signals we want informations concerning the periods inside the interval [8-32] (which represents the circadian domain plus the corresponding harmonics.)

$$g(t_m) = \sum_{n=1}^N f(p_n) e^{j2\pi \frac{1}{p_n} t_m} + \epsilon_m, \quad m \in \{1, \dots, M\} \quad (41)$$

where $g(t_m)$ represents the observed value at time t_m , p_n represents the n -th periodic component and ϵ_m accounts for errors, uncertainties as well as the measurement noise. With the notation $g(t_m) = g_m$ and $f(p_n) = f_n$, defining the vectors $\mathbf{f} = [f_1, f_2, \dots, f_N]^T$, $\mathbf{g} = [g_1, g_2, \dots, g_M]^T$ and $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \dots, \epsilon_M]^T$ we obtain the following model:

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\epsilon} \quad (42)$$

where the elements of the matrix $H_{mn} = e^{j2\pi \frac{1}{p_n} t_m}$. We may note that, as the data are real valued, f_n can be considered as a complex number $f_j = f_n^R + j f_n^I$ and we can also write the forward model as

$$g(t_m) = \sum_{n=1}^N f^R(p_n) \cos(2\pi \frac{1}{p_n} t_m) + f^I(p_n) \sin(2\pi \frac{1}{p_n} t_m) + \epsilon_m, \quad m \in \{1, \dots, M\} \quad (43)$$

Before going to the numerical experimentation, a complementary point for applying the proposed method for this case is that we assume the variances of the real parts f^R and imaginary part f^I are the same:

$$p(\mathbf{f}^R | \mathbf{v}_f) = p(\mathbf{f}^I | \mathbf{v}_f) \quad (44)$$

and that they are only conditionally independent

$$p(\mathbf{f}|\mathbf{v}_f) = p(\mathbf{f}^R|\mathbf{v}_f) p(\mathbf{f}^I|\mathbf{v}_f). \quad (45)$$

Note however that this does not mean that a priori they are independent, because they are conditionally independent.

For validating the proposed method, in a first step, we generated synthetic data. In the real case the theoretical \mathbf{f} is unknown, so the only possible comparison is between the available \mathbf{g} (representing the real data) and the estimated $\hat{\mathbf{g}} = \mathbf{H}\hat{\mathbf{f}}$ (obtained via the reconstruction done over the estimated $\hat{\mathbf{f}}$). An important step for validating the method is to consider signals with known corresponding periodic components vector, which gives the possibility to compare \mathbf{f} and the estimated $\hat{\mathbf{f}}$. We consider the following protocol:

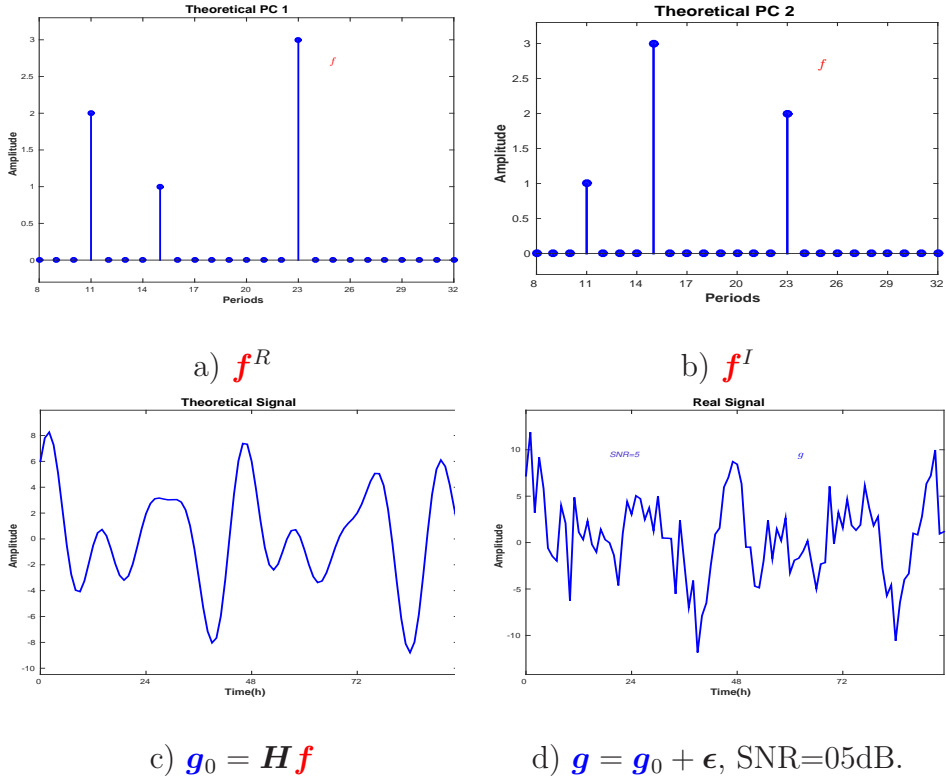


Figure 11: Real part \mathbf{f}^R and Imaginary part \mathbf{f}^I of \mathbf{f} , generated signal without and with noise.

(a) Generate a sparse amplitude periodic components vector \mathbf{f} , via \mathbf{f}^R and \mathbf{f}^I , Figure 11, (a) and (b). For the simulations used in this article, we

analysed a periodic components vector for the interval associated with the circadian domain and the most important corresponding harmonics, i.e. the interval [8, 32], with one hour precision [99, 98].

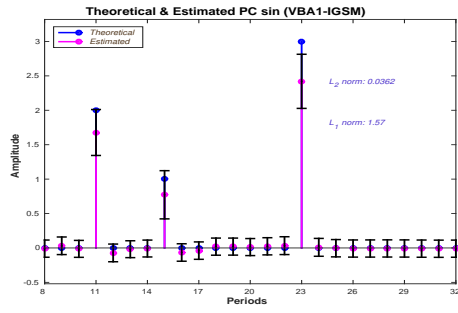
(b) Generate the corresponding signal $\mathbf{g}_0 = \mathbf{H}\mathbf{f}$ called the theoretical output (representing 4 days=96 hours length, simulating a body temperature measurement or a gene expression measurement or a rest-activity pattern measurement. In the analysis of the real data, we will present simulations corresponding to the last two, i.e. gene expression and rest-activity pattern measurements), Figure 11, (c).

(c) Add some noise $\mathbf{g} = \mathbf{g}_0 + \boldsymbol{\epsilon}$ to generate the data data for the inversion step. For the first simulation presented, we added a Gaussian noise, corresponding to SNR=05dB, 11, (d). We include during this simulation section the detailed cases of levels of noise corresponding to SNR=10dB, SNR=15dB, SNR=20dB, comparing the proposed method with the FFT method and also with LASSO. Also, the comparative study between the proposed method and LASSO depending on the SNR is presented, considering levels of noise between 05dB and 40dB.

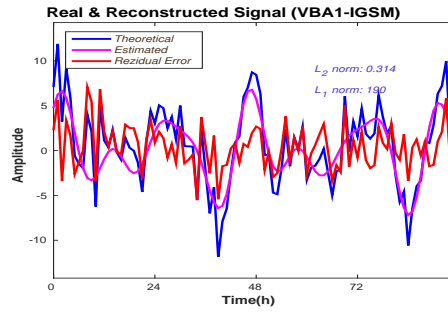
(d) We then applied the proposed method and compared the estimated $\hat{\mathbf{f}}^R$, $\hat{\mathbf{f}}^I$ and $\hat{\mathbf{f}}$ with the original \mathbf{f}^R , \mathbf{f}^I and $\hat{\mathbf{f}}$, Figures 12 (a), (c) and (e). The proposed method also indicates the variances corresponding to each estimated period from the PC vectors \mathbf{f}^R and \mathbf{f}^I . We note that via the proposed algorithm, after we applied VBA we concluded that \mathbf{f} is modelled by a multivariate Normal distribution, and the corresponding covariance matrix was estimated, $\hat{\boldsymbol{\Sigma}}$. We also compare the original signal \mathbf{g} and the theoretical output \mathbf{g}_0 with the reconstructed one $\hat{\mathbf{g}}$, Figure 12 (b) and (d). A comparison between the estimated noise the original one is presented in Figure 12, (f).

As a comparison, we decided to compare the performances of the proposed method with two classical methods: the classical FFT which is today the standard technique used in the chronobiological community and the Lasso which is now the standard L1 regularization, so a good candidate for a method enforcing the sparsity.

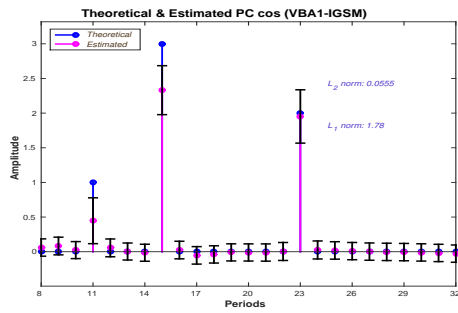
Figure 13 presents a comparison between the proposed method (d), Lasso (c) and FFT (b). The FFT is not offering the wanted precision, while the estimation corresponding to Lasso is far for being precise. In particular, the number of false alarms is 7, and the most important peak in the PC vector, corresponding to 23 hours is estimated as zero. For the proposed method, the estimated PC vector is a sparse one, and the only peaks estimated are



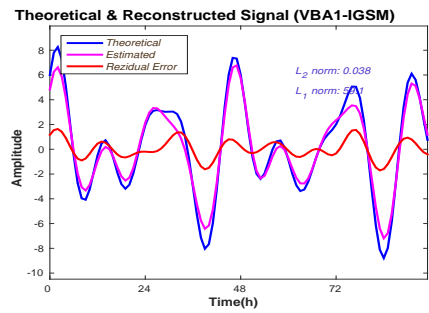
a) f_1 and \hat{f}_{1PM}



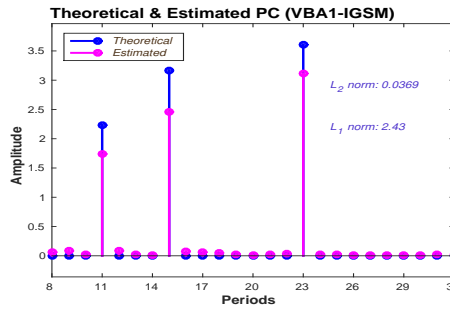
b) g and \hat{g}_{PM}



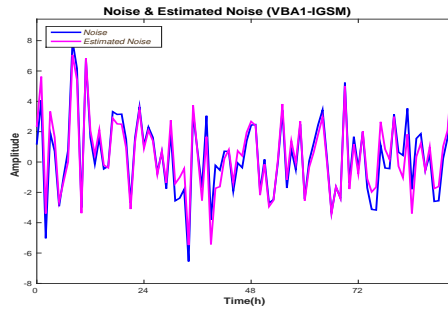
c) f_2 and \hat{f}_{2PM}



d) g_0 vs. \hat{g}_{PM}



e) f and \hat{f}_{PM}



f) ϵ vs. ϵ_{PM}

Figure 12: Comparison between PC and estimated PC, between g_0 , g and g_{PM} and between noise and estimated noise

the three ones that appear in the theoretical PC. In terms of the error, the normalized MSE corresponding to the proposed method is $NMSE= 0.03$, while for LASSO is $NMSE= 0.69$.

Figure 14 presents a comparison between the proposed method (d), Lasso

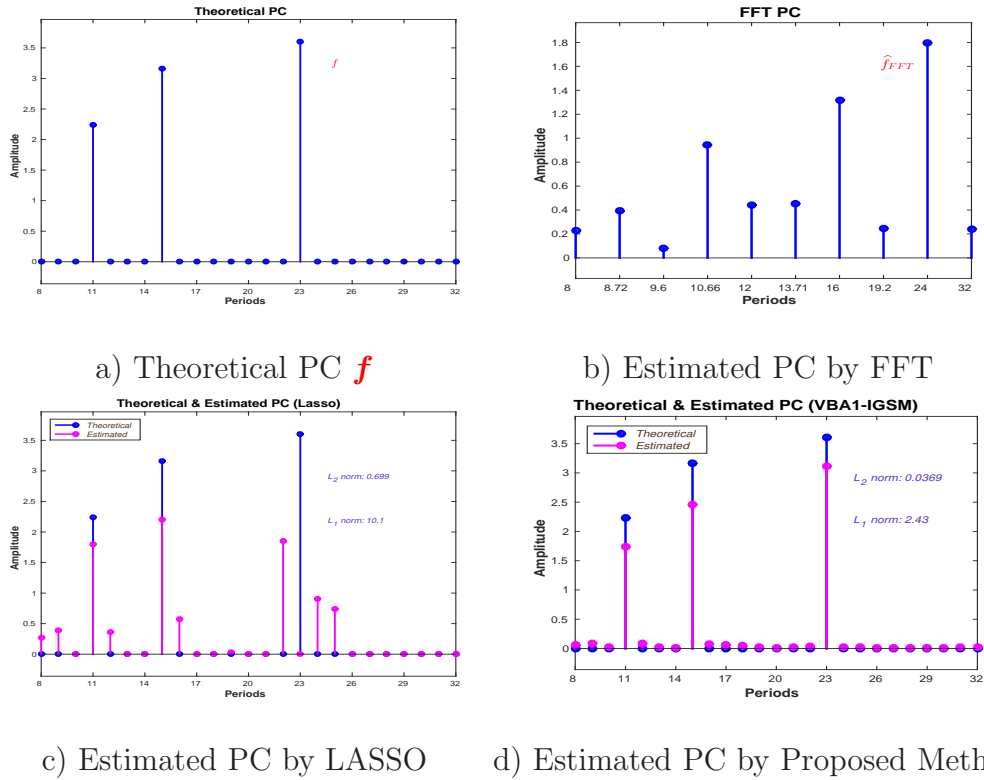


Figure 13: Comparison between proposed method vs. Lasso and FFT (Simulation corresponding to SNR=5dB).

(c) and FFT (b), corresponding to a noise level SNR= 10dB. We note that for the LASSO estimation, the number of false alarms is 6, while the dominant period of the PC vector, i.e. 23 hours, is inaccurate.

Figure 15 presents a comparison between the proposed method (d), Lasso (c) and FFT (b), corresponding to a noise level SNR= 15dB. We note that for the LASSO estimation, the number of false alarms is 3, while the dominant period of the PC vector, i.e. 23 hours, is inaccurate. This is also the case of the simulations corresponding to SNR= 20dB, Figure 15.

The behaviour of the proposed method compared to LASSO for different levels of noise is presented in Figure 17. We have considered levels of noise corresponding to SNR= {05, 10, 15, 20, 30, 40} dB. We compare the estimations corresponding to the two methods for f^I , f^R and f using as measurements of the errors the NMSE 17 (b), (d), (f) and MAE 17 (a), (c), (e).

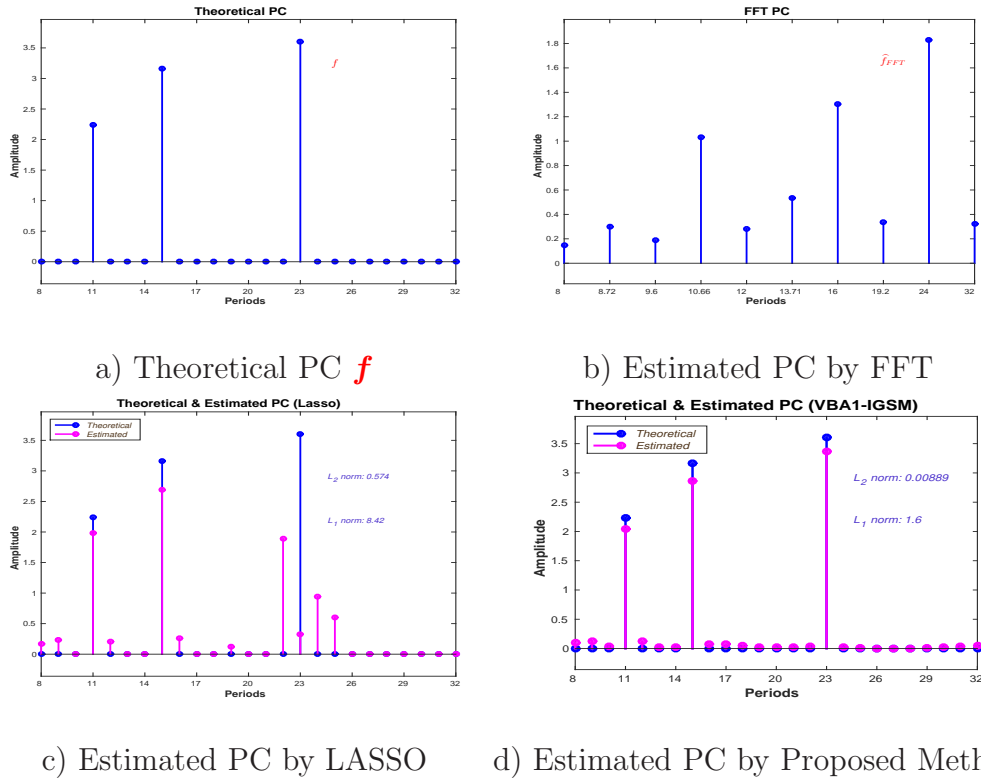


Figure 14: Comparison between proposed method versus Lasso and FFT (Simulation with SNR of 10dB).

The same analysis is presented for g and g_0 in Figure 18.

We may note that, for typical data we have, i.e.; $t_n = n\Delta t, n = 1, \dots, N$ with $\Delta t = 1$ hour, $N = 96$, $p_m = 8 : 32$ hours, the matrix \mathbf{H} which is of size 96×25 is very ill-contrionned. This is due to the fact that, the successive rows of the matrix are very close to each other. In DFT matrix, this is not the case, because the spacing are linear in frequencies but not in periods. Here, we are looking for a precision of 1 hour in period estimation between 8 and 32 hours.

The final step is applying the proposed method on real data. As mentioned, since the standard method in chronobiology today is FFT, we include the corresponding FFT results for comparison. The detailed biological explanation of the experiments are given in [98, 99]. From the signal processing point of view, the main objective is the estimation of a few periodic components of a signal observed for a very short period of time relative to the

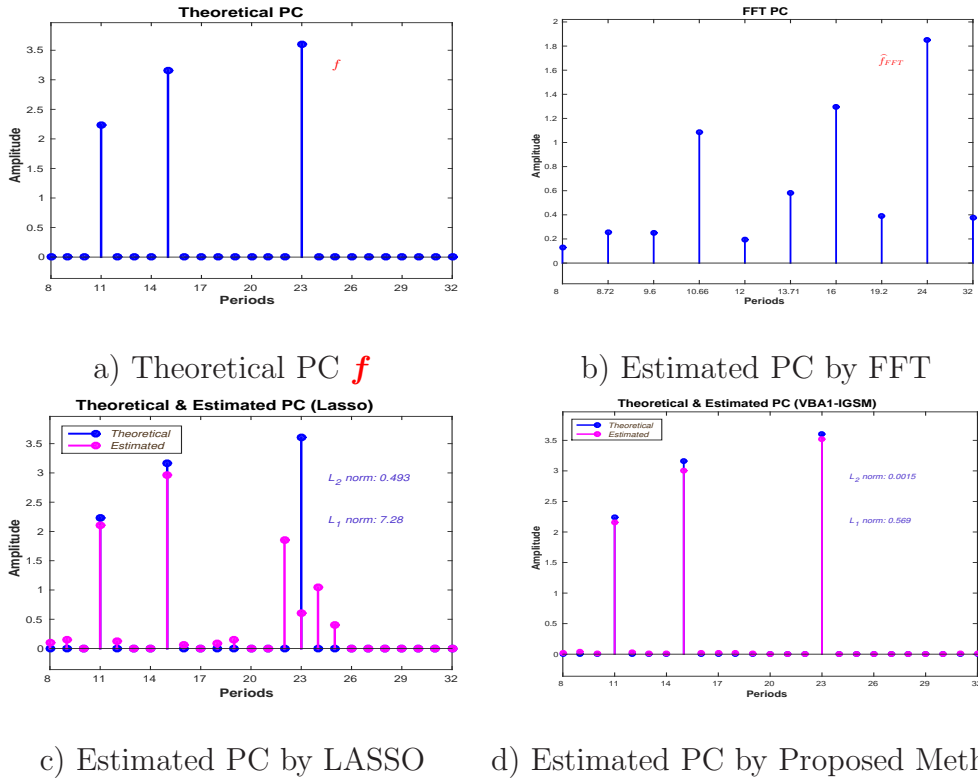


Figure 15: Comparison between proposed method versus Lasso and FFT (Simulation with SNR of 15dB).

prior knowledge of the circadian period (~ 24 hours) to study their evolution during the days with a precision of one hour. Knowing that this precision cannot be obtained with FFT based methods when observing the signal on the intervals of 4 days, we developed the proposed method.

For the results corresponding to real data, we consider data obtained in experiments in chronobiology for cancer treatment. The particular experiment presented is realized over mice, investigating the locomotor activity (rest-activity patterns) of KI/KI Per2::luc mouse, aged 10 weeks, singly housed in RT-BIO and synchronized with LD-12:12 (i.e. 12 hours of light, followed by 12 hours of dark, Light-Dark, LD). The particular signal considered represent the locomotor activity of the mouse, which is known to be rhythmic. After the LD part of the signal, the mouse is kept in total darkness (Dark-Dark, DD) for 3 days, corresponding to the before treatment part of the signal and then D-luciferin is loaded in subcutaneous implanted Alzet

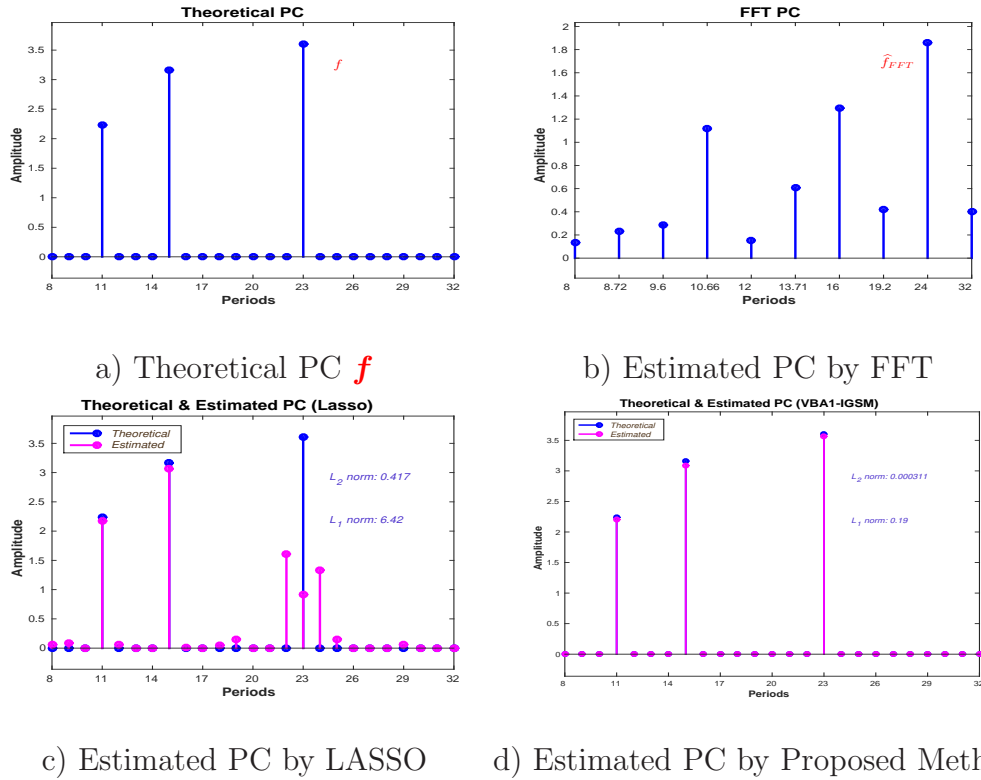
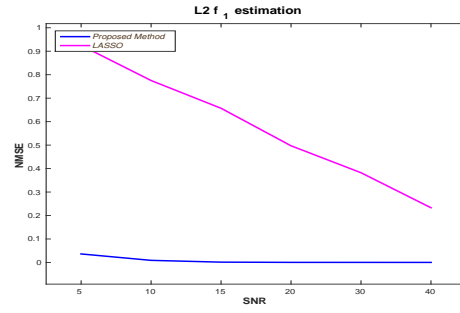
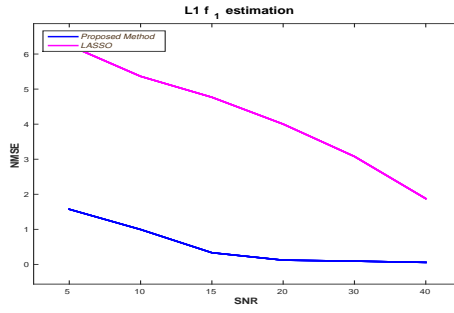


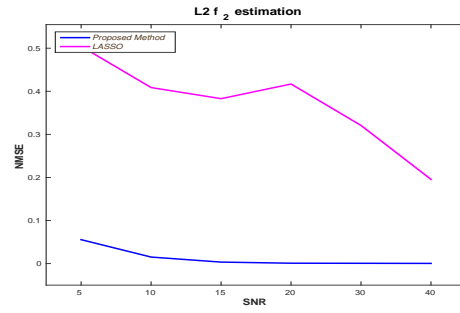
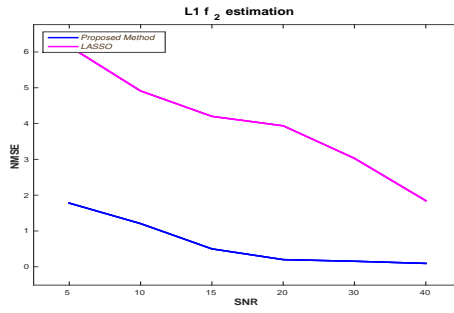
Figure 16: Comparison between proposed method versus Lasso and FFT (Simulation with SNR of 20dB).

pump [90mg/ml], recording for 5 days the Activity signal corresponding to the during treatment part of the signal. The last two days represents the during treatment part of the signal. During the DD segment, the locomotor activity might be perturbed, due to the absence of the Light-Day regime and due to the treatment effects. Figures (20) (a) present the raw data corresponding to the activity signal. The figure indicates the four segments of interest: the LD segment, the DD-before treatment segment, the DD-during segment and the DD-after treatment segment. Each segment is analysed using the proposed method and the FFT method, the standard method in chronobiology today. The raw data signal is sampled every minute. For the four segments studied, we consider mean-zero signals, normalized between $[-10:10]$ and sampled every hour. Figure (19) presents the raw data corresponding to the rest-activity pattern, Activity signal considered. The four segments of interest are presented in Figure (20) (a), (b), (c) and (d). As



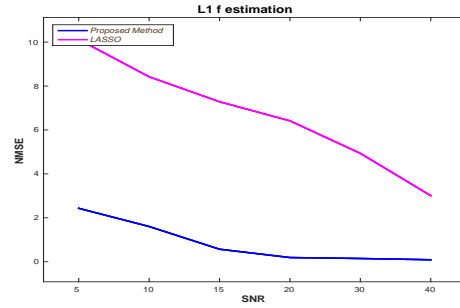
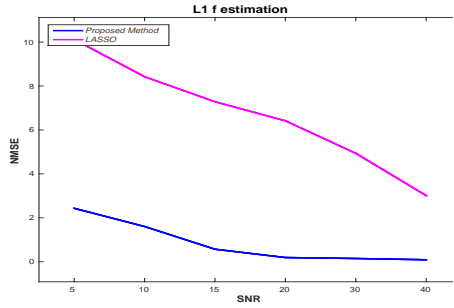
a) L_1 PC \hat{f}^R reconstruction

b) L_2 PC \hat{f}^R reconstruction



c) L_1 PC \hat{f}^I reconstruction

d) L_2 PC \hat{f}^I reconstruction



e) L_1 PC \hat{f} reconstruction

f) L_2 PC \hat{f} reconstruction

Figure 17: Comparison between PC and estimated PC, between g_0 , g and g_{PM} and between noise and estimated noise

mentioned, the context is the analysis of short signals relative to the prior knowledge of the ~ 24 hours. We note that for the experiment discussed

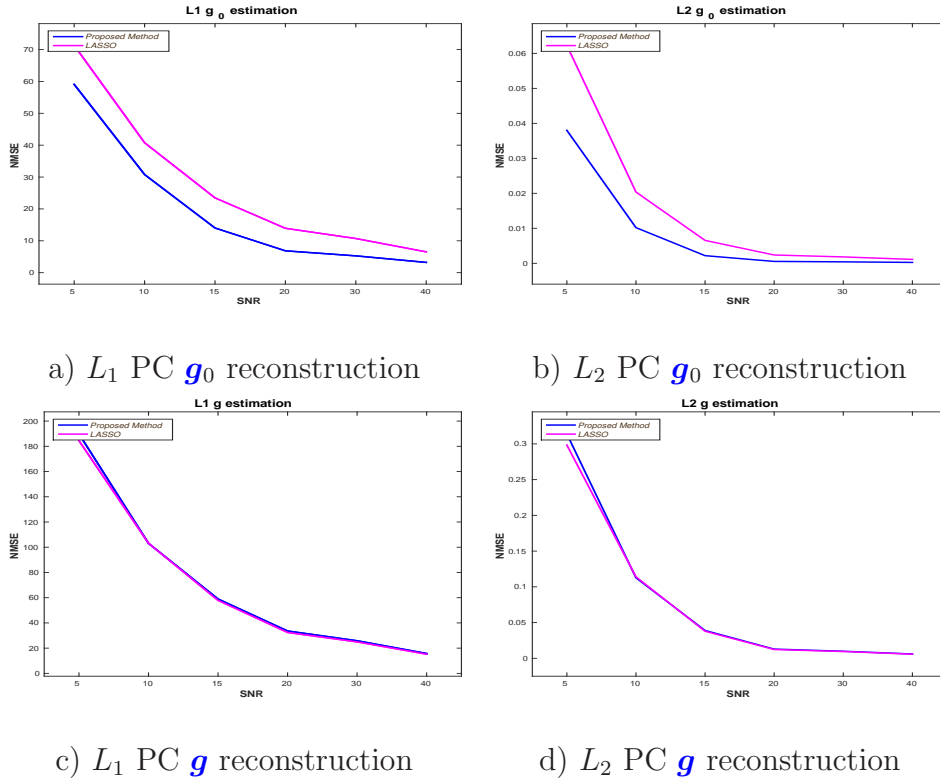
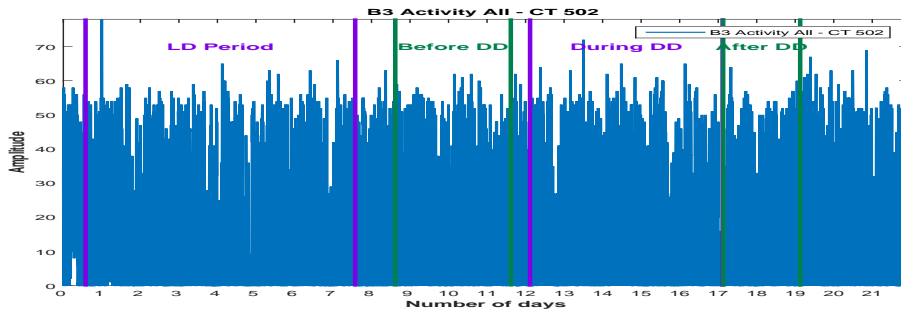


Figure 18: Comparison between PC and estimated PC, between g_0 , g and g_{PM} and between noise and estimated noise

here we need to analyse signals having a length of 3 days (Figure (20) (c), corresponding to the before treatment segment) or 2 days (Figure (20) (e), corresponding to the after treatment segment)

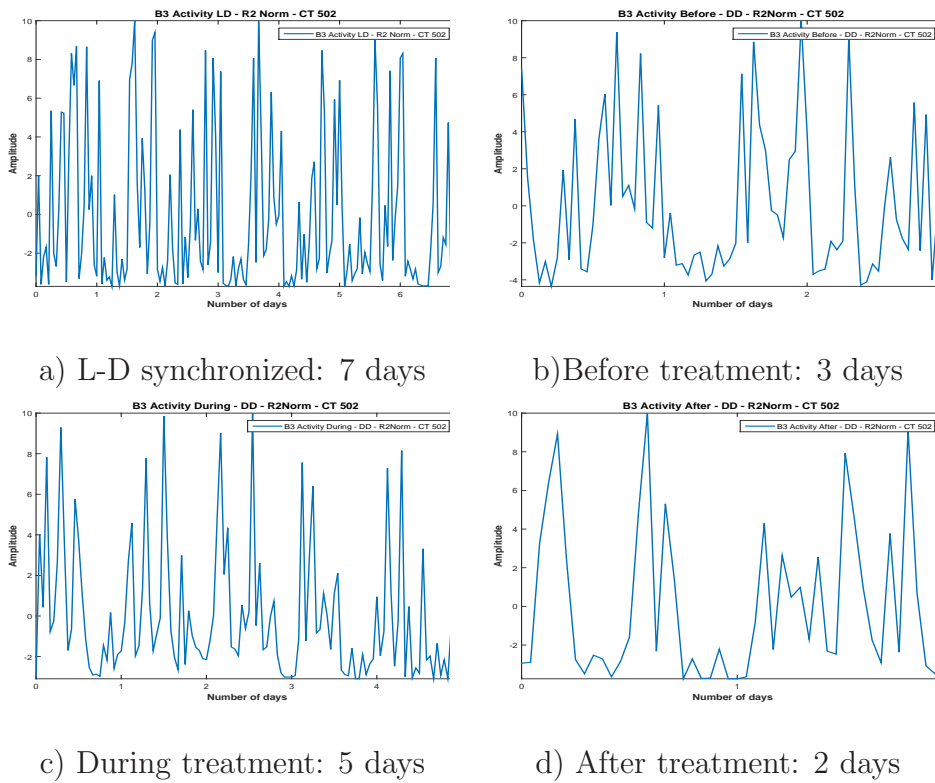
For the LD segment, 8 days are available. The PC vector corresponding to the LD segment is estimated using the Proposed Method and FFT, Figure (21):

The dominant period is estimated at 24 hours, via FFT. Moreover, the PC vector is difficult to be interpreted by the biologist, since the vector is not sparse and the peaks corresponding to the biological phenomena cannot be distinguished from the peaks due to the noise. Via the Proposed method, the PC vector is sparse, estimating only two non-zero peaks, and setting the dominant period at 25 hours. For analysing the stability of the dominant period we consider 4-days length signals (windows) from the available signal, with a shift of one day and compute the PC via FFT and the Proposed



a) Activity Signal: Raw data

Figure 19: Activity raw data



a) L-D synchronized: 7 days

b) Before treatment: 3 days

c) During treatment: 5 days

d) After treatment: 2 days

Figure 20: Activity raw data (a) and the corresponding parts,(b), (c), (d), normalized and one-hour sampled

method. The comparison is presented in Figure (22):

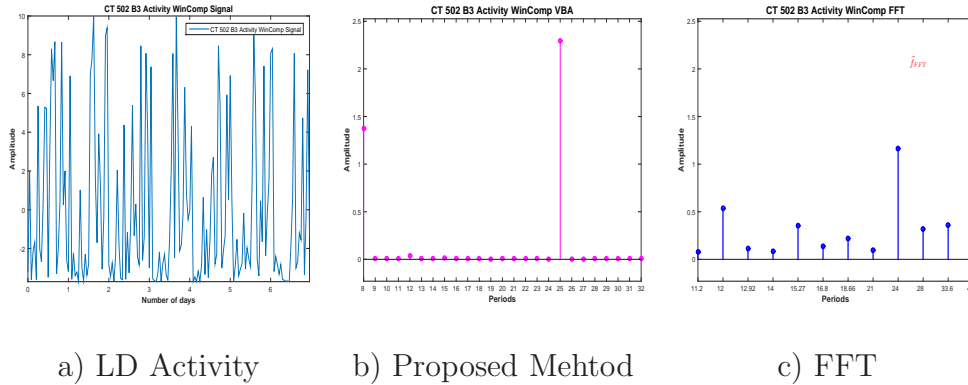


Figure 21: Considered signal (a) and the corresponding PC via VBA (b) and FFT (c)

Via FFT, all four windows considered present a 24 hours dominant periodicity, Figure (22) (c),(f),(i) and (l), leading to the conclusion that the rest-activity patterns are stable during the LD segment. Via the proposed method a variability of the dominant period is detected: for the first and fourth windows the dominant period is 25 hours while for the second and third windows the dominant period is 24 hours. The representation of the results corresponding to the two methods is presented in Figure (23).

The lines of matrices represents the PC vector corresponding to different windows. The colorbar indicates the numerical values of the amplitudes. The figure shows the stability and the variability detected by the FFT method, respectively the Proposed Method. We also note the sparse structure of PC vectors estimated via the Proposed Method. For the DD-before treatment segment, only 3 days of data are available. The PC vector is estimated using FFT and Proposed Method, Figure (24). In particular, the only two non-zero peaks that are present in the PC vector estimated via the Proposed Method correspond to two highest peaks from the PC vector estimated via FFT. For the DD-during treatment segment 5 days are available, Figure (25). For analysing the stability of the dominant period we consider 4-days length signals (windows) from the available signal, with a shift of one day and compute the PC via FFT and the Proposed method. The comparison is presented in Figure (26).

Via the Proposed Method, a variability is detected for the dominant period, while via the FFT method this variability is not detected. The representation of the results corresponding to the two methods is presented in Figure (27).

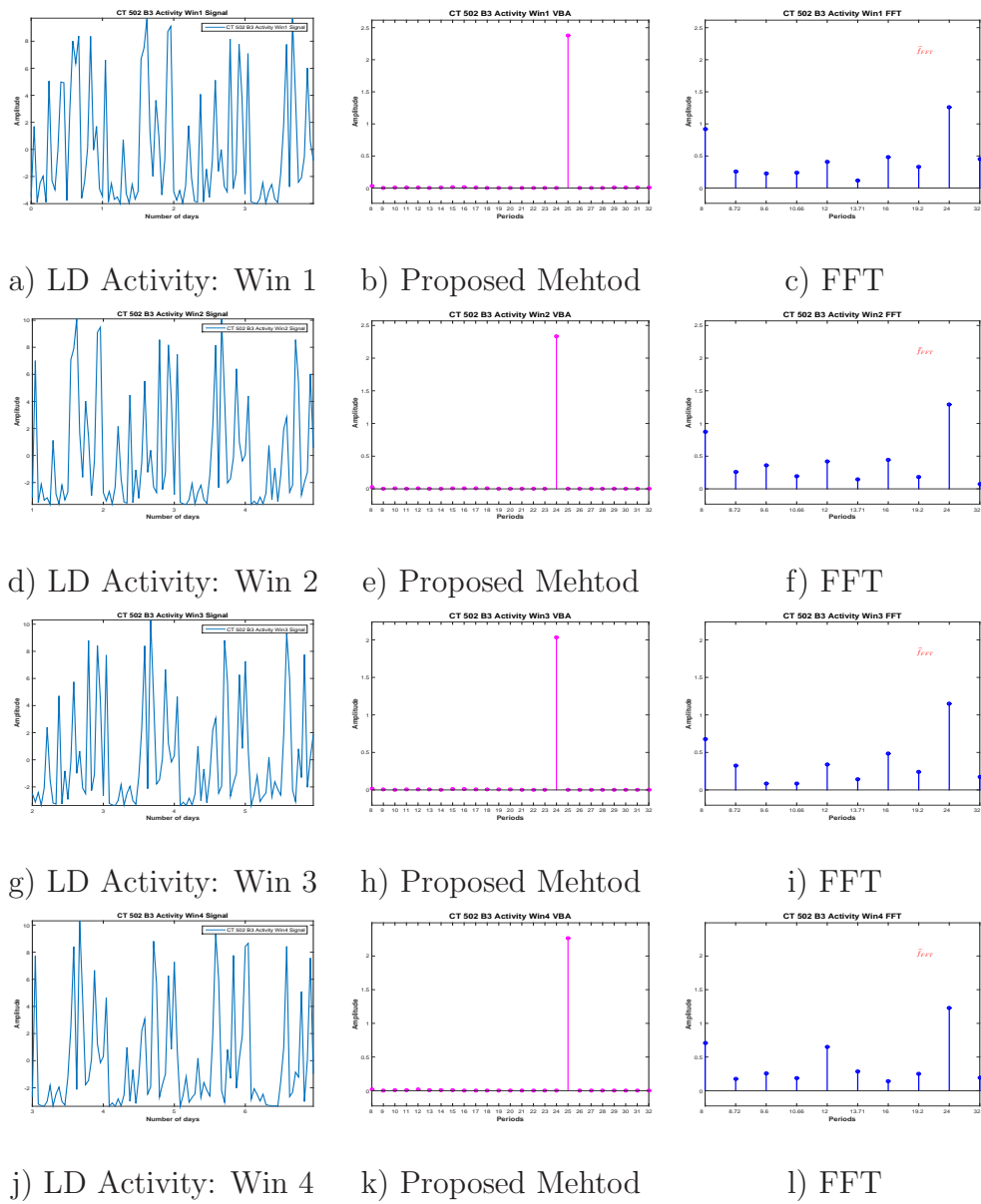
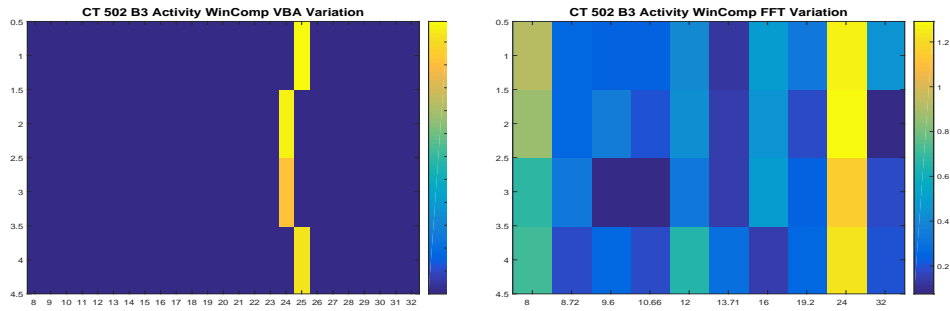


Figure 22: PC vector stability: estimation via Proposed Method and FFT for 4-days length signals

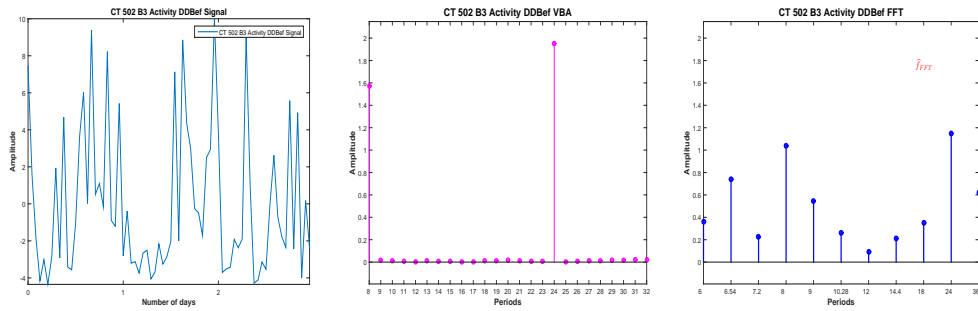
For the DD-after treatment segment, 2-days length data are available. We compare the Proposed Method with the FFT method in Figure (28). The circadian rhythm is perturbed during this phase. The result corresponding



a) Proposed Method

b) FFT

Figure 23: PC vector stability: Proposed Method (a) vs. FFT (b)



a) DD Activity: before

b) Proposed Mehtod

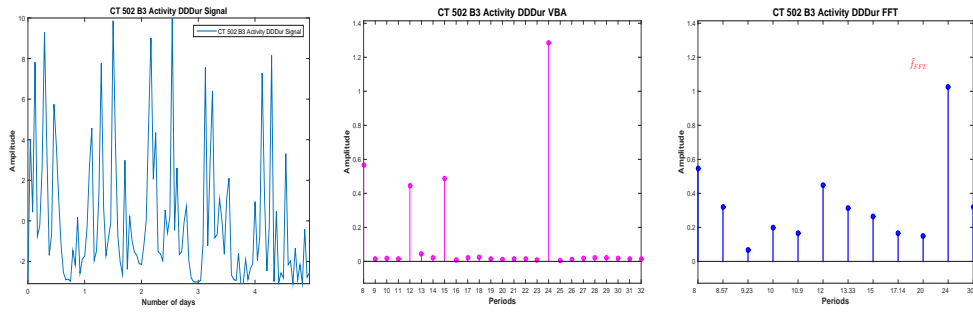
c) FFT

Figure 24: DD before signal (a) and the corresponding PC via Proposed Method (b) and FFT (c)

to the Proposed Method is consistent with the result obtained via FFT, the only three peaks associated with biological phenomena being the 3 highest peaks present in the PC vector estimated via FFT. All the other non-zero peaks that are present in the FFT estimated PC vector are associated with the noise in the Proposed Method estimation.

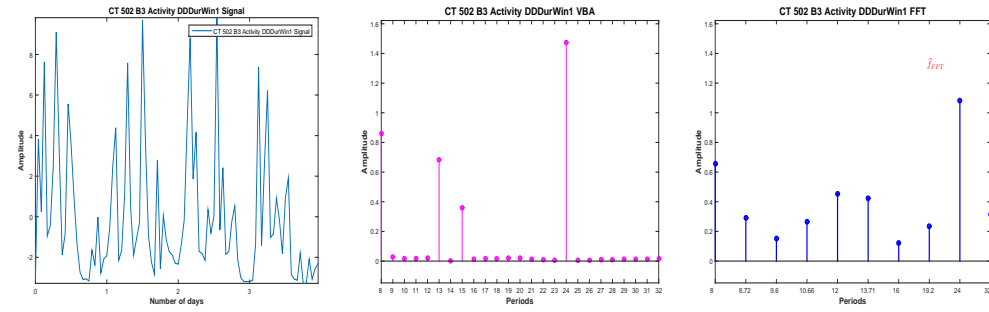
7. Conclusions

In this paper, we considered linear inverse problems and proposed appropriate prior models to account for non stationarity of the errors (forward modelling and measurement noise) and for sparsity enforcing of the input. A generalized Student-t probability density function and its IGSM equivalence

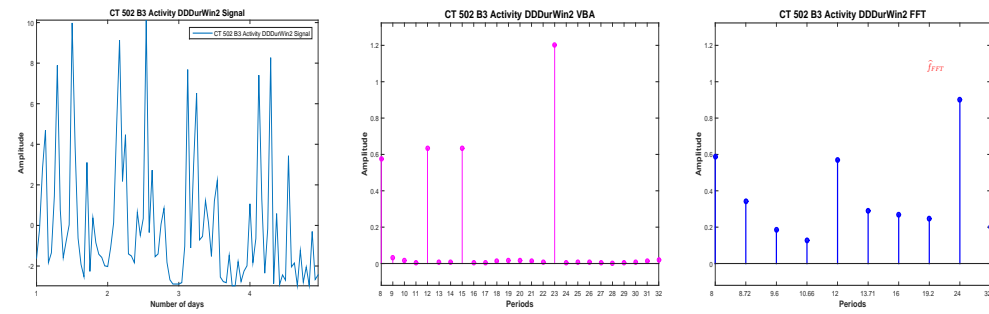


a) DD During: before b) Proposed Mehtod c) FFT

Figure 25: DD during signal (a) and the corresponding PC via VBA (b) and FFT (c)



a) Activity during: W1 b) Proposed Mehtod c) FFT



d) Activity during: W2 e) Proposed Mehtod f) FFT

Figure 26: PC Stability: Estimated PC by FFT and VBA for 4-days length signals, Activity DD, during

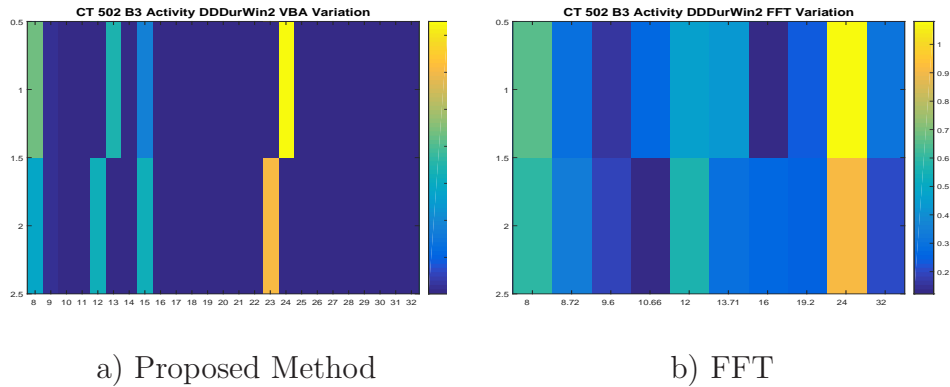


Figure 27: PC Stability: Proposed Method (a) vs. FFT (b)

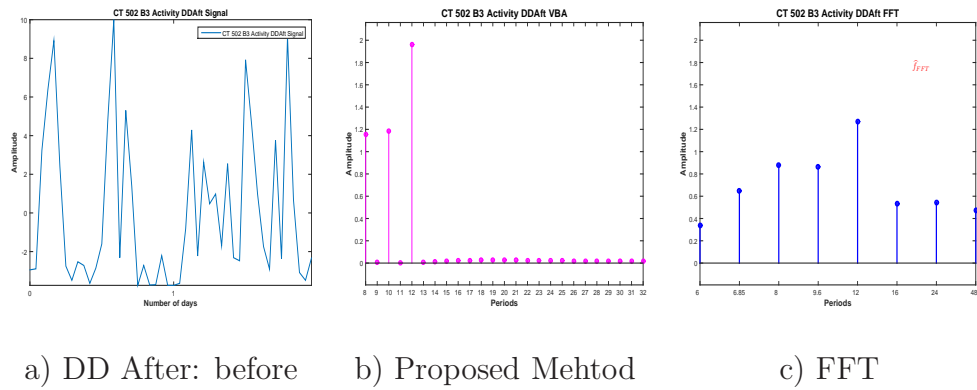


Figure 28: DD after signal (a) and the corresponding PC via Proposed Method (b) and FFT (c)

are used for both. Even if Student-t has been used previously for sparsity enforcing, its use for non stationarity is new. Using these prior laws, we obtained the expression of the joint posterior law of all the unknowns of the problem and proposed two main estimators: JMAP using an alternate optimization algorithm and and the Posterior Mean computed via appropriate VBA method. Even if the MCMC methods are the standard methods for Bayesian computation, we prefer to use VBA which is more effective for practical applications. The advantages of VBA compared to MCMC methods have been noticed by many authors [79, 93, 94, 96, 91, 21, 103, 101, 102]. Indeed, it is crucial to consider the implementation issues, in particular, for great dimensional problems in imaging science such as image deconvolution

or 3D computed tomography. However, in this paper, to show some of the performances of the proposed methods, we considered two signal processing inverse problems: Impulse input deconvolution and periodic components estimation in biological time series.

In the presented chrono-biological application, we have showed the importance of the precision in the estimation of the periodic component around 24 hours. For the FFT based method which is the standard one in the chrono-biology community, the result has not enough resolution to see small changes (a few hours) in the period during the days. The results obtained by the proposed method could show these variations which are very important to biological studies.

To summarize, here we give the highlights of this paper:

- Bayesian inference for linear inverse problems
- Non-stationary noise is modelled via Gaussian with unknown varying variances
- A generalized Student-t prior model is proposed for enforcing sparsity
- Details of two Bayesian algorithms: JMAP and VBA are presented and their performances are compared to the most concurrent method which is Lasso in two applications: parse signal deconvolution and in periodic components estimation in biological time series.
- In the deconvolution, we only compared the performances in simulation, but in the second application, not only we showed the performances in simulation but also for real data obtained in a chrono-biological experiments for cancer research.

Appendix A.

- With respect to \mathbf{f} :

$$\begin{aligned}
\frac{\partial \mathcal{L}(\mathbf{f}, \hat{\mathbf{z}}, \hat{\mathbf{v}}_{\mathbf{f}})}{\partial \mathbf{f}} = 0 &\Leftrightarrow \frac{\partial}{\partial \mathbf{f}} \left(\|\mathbf{V}_{\boldsymbol{\epsilon}}^{-\frac{1}{2}}(\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 + \|(\mathbf{V}_{\mathbf{f}})^{-\frac{1}{2}}\mathbf{f}\|^2 \right) = 0 \\
&\Leftrightarrow -\mathbf{H}^T \mathbf{V}_{\boldsymbol{\epsilon}}^{-1}(\mathbf{g} - \mathbf{H}\mathbf{f}) + \mathbf{V}_{\mathbf{f}}^{-1}\mathbf{f} = 0 \\
&\Leftrightarrow \left[\mathbf{H}^T \mathbf{V}_{\boldsymbol{\epsilon}}^{-1} \mathbf{H} + \mathbf{V}_{\mathbf{f}}^{-1} \right] \mathbf{f} = \mathbf{H}^T \mathbf{V}_{\boldsymbol{\epsilon}}^{-1} \mathbf{g} \\
&\Rightarrow \hat{\mathbf{f}}_{JMAP} = \left[\mathbf{H}^T \mathbf{V}_{\boldsymbol{\epsilon}}^{-1} \mathbf{H} + \mathbf{V}_{\mathbf{f}}^{-1} \right]^{-1} \mathbf{H}^T \mathbf{V}_{\boldsymbol{\epsilon}}^{-1} \mathbf{g}
\end{aligned}$$

- With respect to v_{ϵ_i} , $i \in \{1, 2, \dots, N\}$:

$$\begin{aligned}
\frac{\partial \mathcal{L}(\hat{\mathbf{f}}, \mathbf{v}_{\boldsymbol{\epsilon}}, \hat{\mathbf{v}}_{\mathbf{f}})}{\partial v_{\epsilon_i}} = 0 &\Leftrightarrow \frac{\partial}{\partial v_{\epsilon_i}} \left(\left(\alpha_{\epsilon_i 0} + \frac{3}{2} \right) \ln v_{\epsilon_i} \right. \\
&\quad \left. + \left[\beta_{\epsilon_i 0} + \frac{1}{2} (g_i - \mathbf{H}_i \mathbf{f})^2 \right] v_{\epsilon_i}^{-1} \right) = 0 \\
&\Leftrightarrow \left(\alpha_{\epsilon_i 0} + 1 + \frac{1}{2} \right) v_{\epsilon_i} - \left(\beta_{\epsilon_i 0} + \frac{1}{2} (g_i - \mathbf{H}_i \mathbf{f})^2 \right) = 0 \\
&\Rightarrow \hat{v}_{\epsilon_i JMAP} = \frac{\beta_{\epsilon_i 0} + \frac{1}{2} (g_i - \mathbf{H}_i \mathbf{f})^2}{\alpha_{\epsilon_i 0} + 1 + \frac{1}{2}}
\end{aligned}$$

- With respect to v_f , $j \in \{1, 2, \dots, M\}$:

$$\begin{aligned}
\frac{\partial \mathcal{L}(\hat{\mathbf{f}}, \hat{\mathbf{v}}_\epsilon, \mathbf{v}_f)}{\partial v_{f_j}} = 0 &\Leftrightarrow \frac{\partial}{\partial v_{f_j}} \left(\frac{1}{2} \ln \det(\mathbf{V}_f) + \frac{1}{2} \|\mathbf{V}_f\|^{-\frac{1}{2}} \mathbf{f}\|^2 \right. \\
&\quad \left. + (\alpha_{f_0} + 1) \ln v_{f_j} + \beta_{f_0} v_{f_j}^{-1} \right) = 0 \\
&\Leftrightarrow \frac{\partial}{\partial v_{f_j}} \left(\left[\alpha_{f_0} + 1 + \frac{1}{2} \right] \ln v_{f_j} + \left[\beta_{f_0} + \frac{f_j^2}{2} \right] v_{f_j}^{-1} \right) = 0 \\
&\Leftrightarrow \left(\alpha_{f_0} + 1 + \frac{1}{2} \right) v_{f_j} - \left(\beta_{f_0} + \frac{f_j^2}{2} \right) v_{f_j}^{-2} = 0 \\
&\Rightarrow \widehat{v_{f_j}}_{JMAP} = \frac{\beta_{f_0} + \frac{f_j^2}{2}}{\alpha_{f_0} + 1 + \frac{1}{2}}
\end{aligned}$$

Appendix B.

The analytical expression of the logarithm:

$$\begin{aligned}
\ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g}) &= -\frac{1}{2} \ln \det(\mathbf{V}_\epsilon) - \frac{1}{2} \|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 - \frac{1}{2} \ln \det(\mathbf{V}_f) \\
&\quad - \frac{1}{2} \|\mathbf{V}_f^{-\frac{1}{2}} \mathbf{f}\|^2 - \sum_{i=1}^N (\alpha_{\epsilon_i 0} + 1) \ln v_{\epsilon_i} - \sum_{i=1}^N \beta_{\epsilon_i 0} v_{\epsilon_i}^{-1} \\
&\quad - \sum_{j=1}^M (\alpha_{f_j 0} + 1) \ln v_{f_j} - \sum_{j=1}^M \beta_{f_j 0} v_{f_j}^{-1} + C
\end{aligned} \tag{B.1}$$

- **Expression of $q_1(\mathbf{f})$:**

The proportionality relation concerning $q_1(\mathbf{f})$ refers to \mathbf{f} , so in the expression of $\ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g})$ all the terms free of \mathbf{f} can be regarded as constants:

$$\begin{aligned}
\left\langle \ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g}) \right\rangle_{q_2(\mathbf{v}_\epsilon) q_3(\mathbf{v}_f)} &= \left\langle C - \frac{1}{2} \|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 - \right. \\
&\quad \left. - \frac{1}{2} \|\mathbf{V}_f^{-\frac{1}{2}} \mathbf{f}\|^2 \right\rangle_{q_2(\mathbf{v}_\epsilon) q_3(\mathbf{v}_f)}
\end{aligned}$$

leading to:

$$\begin{aligned} \langle \ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g}) \rangle_{q_2(\mathbf{v}_\epsilon) q_3(\mathbf{v}_f)} = & C - \frac{1}{2} \left\langle \|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 \right\rangle_{q_2(\mathbf{v}_\epsilon)} - \\ & - \frac{1}{2} \left\langle \|\mathbf{V}_f^{-\frac{1}{2}} \mathbf{f}\|^2 \right\rangle_{q_3(\mathbf{v}_f)} \end{aligned} \quad (\text{B.2})$$

Considering the notations corresponding to \mathbf{V}_ϵ and denoting the i -th line of the matrix \mathbf{H} with \mathbf{H}_i , $i \in \{1, 2, \dots, N\}$, we write:

$$\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f}) = \left[v_{\epsilon_1}^{-1/2} (g_1 - \mathbf{H}_1 \mathbf{f}) \dots v_{\epsilon_N}^{-1/2} (g_N - \mathbf{H}_N \mathbf{f}) \right]^T \quad (\text{B.3})$$

so the norm is written:

$$\|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 = \sum_{i=1}^N v_{\epsilon_i}^{-1} (g_i - \mathbf{H}_i \mathbf{f})^2 \quad (\text{B.4})$$

Introducing the notations:

$$\widetilde{v_{\epsilon_i}^{-1}} = \int v_{\epsilon_i}^{-1} q_{2i}(v_{\epsilon_i}) dv_{\epsilon_i}; \widetilde{\mathbf{v}_\epsilon^{-1}} = \left[\widetilde{v_{\epsilon_1}^{-1}} \dots \widetilde{v_{\epsilon_i}^{-1}} \dots \widetilde{v_{\epsilon_N}^{-1}} \right]^T; \widetilde{\mathbf{V}_\epsilon^{-1}} = \text{diag} \left(\widetilde{v_{\epsilon_i}^{-1}} \right) \quad (\text{B.5})$$

we can write:

$$\left\langle \|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 \right\rangle_{q_2(\mathbf{v}_\epsilon)} = \sum_{i=1}^N \widetilde{v_{\epsilon_i}^{-1}} (g_i - \mathbf{H}_i \mathbf{f})^2 = \left\| \left(\widetilde{\mathbf{V}_\epsilon^{-1}} \right)^{1/2} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right\|^2 \quad (\text{B.6})$$

Introducing the notation

$$\widetilde{v_{f_j}^{-1}} = \int v_{f_j}^{-1} q_{3j}(v_{f_j}) dv_{f_j}; \widetilde{\mathbf{v}_f^{-1}} = \left[\widetilde{v_{f_1}^{-1}} \dots \widetilde{v_{f_j}^{-1}} \dots \widetilde{v_{f_M}^{-1}} \right]^T; \widetilde{\mathbf{V}_f^{-1}} = \text{diag} \left(\widetilde{v_{f_j}^{-1}} \right) \quad (\text{B.7})$$

we can write:

$$\left\langle \|\mathbf{V}_f^{-\frac{1}{2}} \mathbf{f}\|^2 \right\rangle_{q_3(\mathbf{v}_f)} = \left\| \left(\widetilde{\mathbf{V}_f^{-1}} \right)^{\frac{1}{2}} \mathbf{f} \right\|^2 \quad (\text{B.8})$$

Finally from (B.2), (B.6) and (B.8), for the expression of the logarithm $\langle \ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g}) \rangle_{q_2(\mathbf{v}_\epsilon) q_3(\mathbf{v}_f)}$ we have:

$$\begin{aligned} \left\langle \ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g}) \right\rangle_{q_2(\mathbf{v}_\epsilon) q_3(\mathbf{v}_f)} = & C - \frac{1}{2} \left\| \left(\widetilde{\mathbf{V}_\epsilon^{-1}} \right)^{1/2} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right\|^2 \\ & - \frac{1}{2} \left\| \left(\widetilde{\mathbf{V}_f^{-1}} \right)^{\frac{1}{2}} \mathbf{f} \right\|^2 \end{aligned} \quad (\text{B.9})$$

and via the first proportionality and the notation:

$$J(\mathbf{f}) = \left\| \left(\widetilde{\mathbf{V}}_{\boldsymbol{\epsilon}}^{-1} \right)^{1/2} (\mathbf{g} - \mathbf{H}\mathbf{f}) \right\|^2 + \left\| \left(\widetilde{\mathbf{V}}_{\mathbf{f}}^{-1} \right)^{1/2} \mathbf{f} \right\|^2 \quad (\text{B.10})$$

the probability $q_1(\mathbf{f})$ can be expressed by the following proportionality:

$$q_1(\mathbf{f}) \propto \left\{ -\frac{1}{2} J(\mathbf{f}) \right\} \quad (\text{B.11})$$

The criterion $J(\mathbf{f})$ introduced in equation (B.10) is quadratic in \mathbf{f} . Equation (B.11) establish a proportionality relation between $q_1(\mathbf{f})$ and an exponential function having as argument a quadratic criterion. This leads to the following

The probability distribution function $q_1(\mathbf{f})$ is a multivariate Normal Distribution.

Of course, the mean is given by the solution that minimize the criterion $J(\mathbf{f})$ i.e. the solution of the equation $\frac{\partial J(\mathbf{f})}{\partial \mathbf{f}} = 0$ (and in particular, this is the same criterion that arrived in the MAP estimation technique for \mathbf{f} , with some the formal differences):

$$\frac{\partial J(\mathbf{f})}{\partial \mathbf{f}} = 0 \Rightarrow \widehat{\mathbf{f}}_{PM} = \left(\mathbf{H}^T \widetilde{\mathbf{V}}_{\boldsymbol{\epsilon}}^{-1} \mathbf{H} + \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1} \right)^{-1} \mathbf{H}^T \widetilde{\mathbf{V}}_{\boldsymbol{\epsilon}}^{-1} \mathbf{g} \quad (\text{B.12})$$

The corresponding covariance matrix is computed by identification. On one hand we have the following relation:

$$\mathcal{N}(\mathbf{f} | \widehat{\mathbf{f}}_{PM}, \widehat{\boldsymbol{\Sigma}}) \propto \left(\det(\widehat{\boldsymbol{\Sigma}}) \right)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{f} - \widehat{\mathbf{f}}_{PM})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{f} - \widehat{\mathbf{f}}_{PM}) \right\} \quad (\text{B.13})$$

On the other hand, we have the following proportionality, given by equation (B.11):

$$\mathcal{N}(\mathbf{f} | \widehat{\mathbf{f}}_{PM}, \widehat{\boldsymbol{\Sigma}}) \propto q_1(\mathbf{f}) \propto \exp \left\{ -\frac{1}{2} J(\mathbf{f}) \right\} \quad (\text{B.14})$$

So, the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ must respect the following relation:

$$(\mathbf{f} - \widehat{\mathbf{f}}_{PM})^T \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{f} - \widehat{\mathbf{f}}_{PM}) \equiv J(\mathbf{f}), \quad (\text{B.15})$$

where the sign \equiv represents an equality between the two terms until a free- \mathbf{f} term. If we consider the covariance matrix

$$\widehat{\boldsymbol{\Sigma}} = \left(\mathbf{H}^T \widetilde{\mathbf{V}}_{\boldsymbol{\epsilon}}^{-1} \mathbf{H} + \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1} \right)^{-1} \quad (\text{B.16})$$

we have the following equalities:

$$\begin{aligned}
(\mathbf{f} - \widehat{\mathbf{f}}_{PM})^T \widehat{\Sigma}^{-1} (\mathbf{f} - \widehat{\mathbf{f}}_{PM}) &= (\mathbf{f} - \widehat{\Sigma} \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g})^T \widehat{\Sigma}^{-1} (\mathbf{f} - \widehat{\Sigma} \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g}) \\
&= (\mathbf{f}^T - \mathbf{g}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} \widehat{\Sigma}) (\widehat{\Sigma}^{-1} \mathbf{f} - \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g}) \\
&= \mathbf{f}^T (\mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} + \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1}) \mathbf{f} - 2 \mathbf{f}^T \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g} + C,
\end{aligned} \tag{B.17}$$

where we have used the equality $\mathbf{f}^T \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g} = \mathbf{g}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} \mathbf{f}$, as a consequence of the fact that one term is the transpose of the other and the term is a scalar. We used $\widehat{\Sigma} = \widehat{\Sigma}^T$ and $\mathbf{g}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} (\mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} + \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1})^{-1} \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g}$ was viewed as a constant C . We also have the following equalities:

$$\begin{aligned}
J(\mathbf{f}) &= \left\| \left(\widetilde{\mathbf{V}}_{\epsilon}^{-1} \right)^{1/2} (\mathbf{g} - \mathbf{H} \mathbf{f}) \right\|^2 + \left\| \left(\widetilde{\mathbf{V}}_{\mathbf{f}}^{-1} \right)^{1/2} \mathbf{f} \right\|^2 = \\
&= (\mathbf{g}^T - \mathbf{f}^T \mathbf{H}^T) \widetilde{\mathbf{V}}_{\epsilon}^{-1} (\mathbf{g} - \mathbf{H} \mathbf{f}) + \mathbf{f}^T \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1} \mathbf{f} \\
&= \mathbf{f}^T (\mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} + \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1}) \mathbf{f} - 2 \mathbf{f}^T \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g} + C.
\end{aligned} \tag{B.18}$$

Equations (B.17) and (B.18) shows that equality imposed in (B.15) is verified with the covariance matrix defined as in (B.16). So, for the Normal distribution $\mathcal{N}(\mathbf{f} | \widehat{\mathbf{f}}, \widehat{\Sigma})$ proportional to $q_1(\mathbf{f})$ we have the following parameters:

$$q_1(\mathbf{f}) = \mathcal{N}(\mathbf{f} | \widehat{\mathbf{f}}_{PM}, \widehat{\Sigma}), \begin{cases} \widehat{\mathbf{f}}_{PM} = (\mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} + \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1})^{-1} \mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{g} \\ \widehat{\Sigma} = (\mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon}^{-1} \mathbf{H} + \widetilde{\mathbf{V}}_{\mathbf{f}}^{-1})^{-1} \end{cases} \tag{B.19}$$

• Expression of $q_{2i}(v_{\epsilon_i})$

The proportionality relation concerning $q_{2i}(v_{\epsilon_i})$ refers to v_{ϵ_i} so in the expres-

sion of $\ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g})$ all the terms free of v_{ϵ_i} can be regarded as constants:

$$\begin{aligned} \left\langle \ln p(\mathbf{f}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g}) \right\rangle_{q_1(\mathbf{f}) q_{2-i}(v_{\epsilon_i}) q_3(\mathbf{v}_f)} &= C - \frac{1}{2} \langle \ln \det(\mathbf{V}_\epsilon) \rangle_{q_{2-i}(v_{\epsilon_i})} \\ &\quad - (\alpha_{\epsilon_i 0} + 1) \ln v_{\epsilon_i} \\ &\quad - \frac{1}{2} \left\langle \|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 \right\rangle_{q_1(\mathbf{f}) q_{2-i}(v_{\epsilon_i})} \\ &\quad - \beta_{\epsilon_i 0} v_{\epsilon_i}^{-1} \end{aligned} \quad (\text{B.20})$$

For the first integral, it is trivial to verify:

$$\langle \ln \det(\mathbf{V}_\epsilon) \rangle_{q_{2-i}(v_{\epsilon_i})} = C + \ln v_{\epsilon_i} \quad (\text{B.21})$$

For the second integral, we have the following development:

$$\left\langle \|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 \right\rangle_{q_1(\mathbf{f}) q_{2-i}(v_{\epsilon_i})} = \left\langle \|\widetilde{\mathbf{V}}_{\epsilon-i}^{-1 \frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 \right\rangle_{q_1(\mathbf{f})} \quad (\text{B.22})$$

where we have introduced the following notations:

$$\widetilde{\mathbf{v}}_{\epsilon-i}^{-1} = \left[\widetilde{v}_{\epsilon_1}^{-1} \dots \widetilde{v}_{\epsilon_{i-1}}^{-1} v_{\epsilon_i}^{-1} \widetilde{v}_{\epsilon_{i+1}}^{-1} \dots \widetilde{v}_{\epsilon_N}^{-1} \right]^T ; \quad \widetilde{\mathbf{V}}_{\epsilon-i}^{-1} = \text{diag} \left(\widetilde{\mathbf{v}}_{\epsilon-i}^{-1} \right) \quad (\text{B.23})$$

Again, using the fact that $q_1(\mathbf{f})$ is a multivariate Normal Distribution we have:

$$\left\langle \|\widetilde{\mathbf{V}}_{\epsilon-i}^{-1 \frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 \right\rangle_{q_1(\mathbf{f})} = \|\widetilde{\mathbf{V}}_{\epsilon-i}^{-1 \frac{1}{2}} (\mathbf{g} - \mathbf{H}\widehat{\mathbf{f}}_{PM})\|^2 + \text{Tr} \left(\mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon-i}^{-1} \mathbf{H} \widehat{\Sigma} \right) \quad (\text{B.24})$$

and considering as constants all terms free of v_{ϵ_i} we have:

$$\|\widetilde{\mathbf{V}}_{\epsilon-i}^{-1 \frac{1}{2}} (\mathbf{g} - \mathbf{H}\widehat{\mathbf{f}}_{PM})\|^2 = C + v_{\epsilon_i}^{-1} \left(g_i - \mathbf{H}_i \widehat{\mathbf{f}}_{PM} \right)^2 \quad (\text{B.25})$$

and

$$; \quad \text{Tr} \left(\mathbf{H}^T \widetilde{\mathbf{V}}_{\epsilon-i}^{-1} \mathbf{H} \widehat{\Sigma} \right) = C + v_{\epsilon_i}^{-1} \mathbf{H}_i \widehat{\Sigma} \mathbf{H}_i^T \quad (\text{B.26})$$

where \mathbf{H}_i is the line i of the matrix \mathbf{H} , so we can conclude:

$$\left\langle \|\mathbf{V}_\epsilon^{-\frac{1}{2}} (\mathbf{g} - \mathbf{H}\mathbf{f})\|^2 \right\rangle_{q_1(\mathbf{f}) q_{2-i}(v_{\epsilon_i})} = C + \left[\mathbf{H}_i \widehat{\Sigma} \mathbf{H}_i^T + \left(g_i - \mathbf{H}_i \widehat{\mathbf{f}}_{PM} \right)^2 \right] v_{\epsilon_i}^{-1} \quad (\text{B.27})$$

From (B.20) via (B.21) and (B.27) we get:

$$\begin{aligned} \left\langle \ln p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g}) \right\rangle_{q_1(\mathbf{f}) q_{2-i}(\mathbf{v}_{\epsilon_i}) q_3(\mathbf{v}_f)} &= C - \left(\alpha_{\epsilon_i 0} + 1 + \frac{1}{2} \right) \ln v_{\epsilon_i} \\ &\quad \left(\beta_{\epsilon_i 0} + \frac{1}{2} \left[\mathbf{H}_i \widehat{\Sigma} \mathbf{H}_i^T \right. \right. \\ &\quad \left. \left. + \left(\mathbf{g}_i - \mathbf{H}_i \widehat{\mathbf{f}}_{PM} \right)^2 \right] \right) v_{\epsilon_i}^{-1} \end{aligned}$$

from which we can establish the proportionality corresponding to $q_{2i}(\mathbf{v}_{\epsilon_i})$:

$$q_{2i}(\mathbf{v}_{\epsilon_i}) \propto v_{\epsilon_i}^{-(\alpha_{\epsilon_i 0} + 1 + \frac{1}{2})} \exp \left\{ \frac{-1}{v_{\epsilon_i}} \left(\beta_{\epsilon_i 0} + \frac{1}{2} \left[\mathbf{H}_i \widehat{\Sigma} \mathbf{H}_i^T + \left(\mathbf{g}_i - \mathbf{H}_i \widehat{\mathbf{f}}_{PM} \right)^2 \right] \right) \right\} \quad (\text{B.28})$$

Equation (B.28) leads to following

The probability distribution function $q_{3i}(\mathbf{v}_{\epsilon_i})$ is an Inverse Gamma Distribution, with the parameters α_{ϵ_i} and β_{ϵ_i} :

We can write:

$$q_{2i}(\mathbf{v}_{\epsilon_i}) = \mathcal{IG}(v_{\epsilon_i} | \alpha_{\epsilon_i}, \beta_{\epsilon_i}), \begin{cases} \alpha_{\epsilon_i} = \alpha_{\epsilon_i 0} + \frac{1}{2} \\ \beta_{\epsilon_i} = \beta_{\epsilon_i 0} + \frac{1}{2} \left[\mathbf{H}_i \widehat{\Sigma} \mathbf{H}_i^T + \left(\mathbf{g}_i - \mathbf{H}_i \widehat{\mathbf{f}}_{PM} \right)^2 \right] \end{cases} \quad (\text{B.29})$$

• Expression of $q_{3j}(\mathbf{v}_{f_j})$

The proportionality relation concerning $q_{3j}(\mathbf{v}_{f_j})$ refers to \mathbf{v}_{f_j} , so in the expression of $\ln p(\mathbf{f}, \mathbf{z}, \mathbf{v}_\epsilon, \mathbf{v}_f | \mathbf{g})$ all the terms free of \mathbf{v}_{f_j} can be regarded as constants. Considering all \mathbf{v}_{f_j} free terms as constants it is easy to verify:

$$\left\langle \ln \det(\mathbf{V}_f) \right\rangle_{q_{3-j}(\mathbf{v}_{f_j})} = C + \ln v_{f_j} \quad (\text{B.30})$$

For the second integral:

$$\left\langle \left\| \left(\mathbf{V}_f \right)^{-\frac{1}{2}} \mathbf{f} \right\|^2 \right\rangle_{q_1(\mathbf{f}) q_{3-j}(\mathbf{v}_{f_j})} = \left\langle \left\| \left(\widetilde{\mathbf{V}}_{f_{-i}}^{-1} \right)^{\frac{1}{2}} \mathbf{f} \right\|^2 \right\rangle_{q_1(\mathbf{f})} \quad (\text{B.31})$$

where we have introduced the notations:

$$\widetilde{\mathbf{v}}_{f_{-i}}^{-1} = \left[\widetilde{v}_{f_1}^{-1} \ \dots \ \widetilde{v}_{f_{i-1}}^{-1} \ v_{f_i}^{-1} \ v_{f_{i+1}}^{-1} \ \dots \ \widetilde{v}_{f_N}^{-1} \right]^T ; \quad \widetilde{\mathbf{V}}_{f_{-i}}^{-1} = \text{diag} \left(\widetilde{\mathbf{v}}_{f_{-i}}^{-1} \right) \quad (\text{B.32})$$

Considering the fact that $q_1(\mathbf{f})$ was established a multivariate Normal Distribution, we have:

$$\begin{aligned} \left\langle \left\| \left(\widetilde{\mathbf{V}_{\mathbf{f}^{-i}}^{-1}} \right)^{\frac{1}{2}} \mathbf{f} \right\|^2 \right\rangle_{q_1(\mathbf{f})} &= \left\| \left(\widetilde{\mathbf{V}_{\mathbf{f}^{-i}}^{-1}} \right)^{\frac{1}{2}} \widehat{\mathbf{f}}_{PM} \right\|^2 + \text{Tr} \left(\widetilde{\mathbf{V}_{\mathbf{f}^{-i}}^{-1}} \widehat{\boldsymbol{\Sigma}} \right) = \\ &= C + v_{f_i}^{-1} \left(\widehat{f}_{jPM}^2 + \widehat{\boldsymbol{\Sigma}}_{jj} \right) \end{aligned} \quad (\text{B.33})$$

Via (B.30) and (B.33) we get:

$$\begin{aligned} \left\langle \ln p(\mathbf{f}, \mathbf{v}_{\boldsymbol{\epsilon}}, \mathbf{v}_{\mathbf{f}} | \mathbf{g}) \right\rangle_{q_1(\mathbf{f}) q_2(\mathbf{v}_{\boldsymbol{\epsilon}}) q_{3-j}(\mathbf{v}_{f_j})} &= - \left(\alpha_{f_j0} + \frac{1}{2} + 1 \right) \ln v_{f_j} \\ &\quad - \left(\beta_{f_j0} + \frac{1}{2} \left(\widehat{f}_{jPM}^2 + \widehat{\boldsymbol{\Sigma}}_{jj} \right) \right) v_{f_j}^{-1} \end{aligned} \quad (\text{B.34})$$

from which we can establish the proportionality corresponding to $q_4(\mathbf{v}_{f_j})$:

$$q_{3j}(\mathbf{v}_{f_j}) \propto v_{f_j}^{-(\alpha_{f_j0} + \frac{1}{2} + 1)} \exp \left\{ - \left[\beta_{f_j0} + \frac{1}{2} \left(\widehat{f}_{jPM}^2 + \widehat{\boldsymbol{\Sigma}}_{jj} \right) \right] v_{f_j}^{-1} \right\} \quad (\text{B.35})$$

Equation (B.35) leads to the following

The probability distribution function $q_4(\mathbf{v}_{\mathbf{f}})$ is an Inverse Gamma Distribution, with the parameters α_{f_j} and β_{f_j} :

$$q_{3j}(\mathbf{v}_{f_j}) = \mathcal{IG} \left(v_{f_j} | \alpha_{f_j}, \beta_{f_j} \right), \begin{cases} \alpha_{f_j} = \alpha_{f_j0} + \frac{1}{2} \\ \beta_{f_j} = \beta_{f_j0} + \frac{1}{2} \left(\widehat{f}_{jPM}^2 + \widehat{\boldsymbol{\Sigma}}_{jj} \right) \end{cases} \quad (\text{B.36})$$

Expressions (B.19), (B.29) and (B.36) resumes the distributions families and the corresponding parameters for $q_1(\mathbf{f})$, $q_{2i}(\mathbf{v}_{\boldsymbol{\epsilon}_i})$, $i \in \{1, 2, \dots, N\}$ and $q_{3j}(\mathbf{v}_{f_j})$, $j \in \{1, 2, \dots, M\}$. However, the parameters corresponding to the multivariate Normal distribution are expressed via $\widetilde{\mathbf{V}_{\boldsymbol{\epsilon}^{-1}}}$ and $\widetilde{\mathbf{V}_{\mathbf{f}^{-1}}}$ (and by extension all elements forming the three matrices $\widetilde{v_{\boldsymbol{\epsilon}_i}^{-1}}$, $i \in \{1, 2, \dots, N\}$ and $\widetilde{v_{f_j}^{-1}}$, $j \in \{1, 2, \dots, M\}$).

• **Computation of $\widetilde{\mathbf{V}_{\boldsymbol{\epsilon}^{-1}}}$, $\widetilde{\mathbf{V}_{\mathbf{f}^{-1}}}$:**

For a Inverse Gamma Distribution with parameters α and β , $\mathcal{IG}(x|\alpha, \beta)$, the following relation holds:

$$\langle x^{-1} \rangle_{\mathcal{IG}(x|\alpha, \beta)} = \frac{\alpha}{\beta}$$

The prove of the above relation is done by direct computation, using the analytical expression of the Inverse Gamma Distribution:

$$\begin{aligned} \langle x^{-1} \rangle_{\mathcal{IG}(x|\alpha, \beta)} &= \int x^{-1} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left\{-\frac{\beta}{x}\right\} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \int \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^{-(\alpha+1)-1} \exp\left\{-\frac{\beta}{x}\right\} dx = \\ &= \frac{\alpha}{\beta} \underbrace{\int \mathcal{IG}(x|\alpha+1, \beta) dx}_1 = \frac{\alpha}{\beta} \end{aligned}$$

Since $q_{2i}(v_{\epsilon_i})$, $i \in \{1, 2, \dots, N\}$ and $q_{3j}(v_{f_j})$, $j \in \{1, 2, \dots, M\}$ are Inverse Gamma Distributions, with parameters α_{ϵ_i} and β_{ϵ_i} , $i \in \{1, 2, \dots, N\}$ respectively α_{f_j} and β_{f_j} , $j \in \{1, 2, \dots, M\}$ we can express the expectancies $\widetilde{v_{\epsilon_i}^{-1}}$ and $\widetilde{v_{f_j}^{-1}}$ via the parameters of the two Inverse Gamma Distributions using the result above:

$$\widetilde{v_{\epsilon_i}^{-1}} = \frac{\alpha_{\epsilon_i}}{\beta_{\epsilon_i}} \quad ; \quad \widetilde{v_{f_j}^{-1}} = \frac{\alpha_{f_j}}{\beta_{f_j}} \quad (\text{B.37})$$

Using the notation introduced in (B.5) and (B.7) we obtain:

$$\widetilde{\mathbf{V}_{\epsilon}^{-1}} = \begin{bmatrix} \frac{\alpha_{\epsilon_1}}{\beta_{\epsilon_1}} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\alpha_{\epsilon_i}}{\beta_{\epsilon_i}} & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \frac{\alpha_{\epsilon_N}}{\beta_{\epsilon_N}} \end{bmatrix} = \widetilde{\mathbf{V}_{\epsilon}^{-1}} \quad ; \quad \widetilde{\mathbf{V}_{f}^{-1}} = \begin{bmatrix} \frac{\alpha_{f_1}}{\beta_{f_1}} & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \frac{\alpha_{f_j}}{\beta_{f_j}} & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & \frac{\alpha_{f_M}}{\beta_{f_M}} \end{bmatrix} = \widetilde{\mathbf{V}_{f}^{-1}} \quad (\text{B.38})$$

In equation, (B.38) we have introduced other notations for $\widetilde{\mathbf{V}_{f}^{-1}}$ and $\widetilde{\mathbf{V}_{\epsilon}^{-1}}$. All three values were expressed during the model via unknown expectancies, but in this point we arrive to expression that don't contain any more integrals to be computed. Therefore, the new notations represents the final expressions for the density functions q that depends only on numerical hyperparameters, set in the prior modelling.

References

- [1] B. Hunt, A theorem on the difficulty of numerical deconvolution, *IEEE Trans. AU-20* (1972) 94–95.
- [2] T. S. Huang, P. M. Narendra, Image restoration by singular value decomposition., *Appl Opt* 14 (9) (1975) 2213–2216.
- [3] T. S. Huang, D. A. Barker, S. P. Berger, Iterative image restoration., *Appl Opt* 14 (5) (1975) 1165–1168.
- [4] N. N. Abdelmalek, T. Kasvand, J. P. Croteau, Image restoration for space invariant point spread functions., *Appl Opt* 19 (7) (1980) 1184–1189.
- [5] W. Souidene, K. Abed-Meraim, A. Beghdadi, A new look to multichannel blind image deconvolution., *IEEE Trans Image Process* 18 (7) (2009), 1487–1500.
- [6] A. Levin, Y. Weiss, F. Durand, W. T. Freeman, Understanding blind deconvolution algorithms., *IEEE Trans. PAMI*, July 2011.
- [7] S. Gulam Razul, W. Fitzgerald, C. Andrieu, Bayesian model selection and parameter estimation of nuclear emission spectra using RJMCMC, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 497 (2-3) (2003) 492–510.
- [8] S. Deans, *The Radon Transform and some of its applications*, wiley interscience Edition, 1983.
- [9] G. Beylkin, The inversion problem and applications of the generalized radon transform, *Commun. Pure Appl. Math.* 37 (1984) 579–599.
- [10] C. Cai, A. Mohammad-Djafari, S. Legoupil, T. Rodet, Bayesian data fusion and inversion in x-ray Multi-Energy computed tomography, in: 2011 18th IEEE International Conference on Image Processing (IEEE ICIP2011), Brussels, Belgium, 2011, pp. 1405–1408.
- [11] G. Herman, A. Kuba, L. MyiLibrary, *Advances in discrete tomography and its applications*, Vol. 1, Birkhäuser Boston:, 2007.

- [12] G. Herman, A. Kuba (Eds.), *Discrete Tomography: Foundations, Algorithms, and Applications*, Birkhäuser, 2009.
- [13] K. Batenburg, J. Sijbers, DART: A practical reconstruction algorithm for discrete tomography, *Image Processing, IEEE Transactions on* 20 (9) (2011) 2542–2553.
- [14] S. Roux, H. Leclerc, F. Hild, Tomographic reconstruction of binary fields, in: *Journal of Physics: Conference Series*, Vol. 386, IOP Publishing, 2012, p. 012014.
- [15] A. Mohammad-Djafari, G. Demoment, Maximum entropy fourier synthesis with application to diffraction tomography, *Applied Optics* 26 (1987) 1745–1754.
- [16] M. Nguyen, A. Mohammad-Djafari, Bayesian approach with the maximum entropy priors in image reconstruction from the microwave scattered field data, *IEEE Trans. on Medical Imaging*, Vol. 13, Num. 2, pp 254-262.
- [17] O. Féron, B. Duchêne, A. Mohammad-Djafari, Microwave imaging of piecewise constant objects in a 2D-TE configuration, *International Journal of Applied Electromagnetics and Mechanics* 26 (6) (2007) 167–174.
- [18] E. Kuruoglu, A. Tonazzini, L. Bianchi, Source separation in noisy astrophysical images modelled by markov random fields, in: *Image Processing, 2004. ICIP'04. 2004 International Conference on*, Vol. 4, IEEE, 2004, pp. 2701–2704.
- [19] H. Carfantan, A. Mohammad-Djafari, A Bayesian framework for nonlinear diffraction tomography, in: *IEEE EURASIP Workshop on Nonlinear Signal and Image Processing NSIP'97*, Mackinac island, MI, USA, 1997.
- [20] O. Féron, B. Duchêne, A. Mohammad-Djafari, Microwave imaging of inhomogeneous objects made of a finite number of dielectric and conductive materials from experimental data, *Inverse Problems* 21 (6) (2005) 95–115.
- [21] H. Ayasso, A. Mohammad-Djafari, Joint NDT image restoration and segmentation using Gauss–Markov–Potts prior models and variational

- bayesian computation, *IEEE Transactions on Image Processing* 19 (9) (2010) 2265–2277.
- [22] L. Gharsalli, A. Hacheme, D. Bernard, A. Mohammad-Djafari, Microwave tomography for breast cancer detection within a variational bayesian approach, in: *IEEE European Signal Processing Conference, EUSIPCO*, Vol. nc, Marrakech, Morocco, pp.ID1569743387, 2013.
 - [23] M. Nikolova, A. Mohammad-Djafari, Eddy current tomography using a binary Markov model, *Signal Processing* 49 (1996) 119–132.
 - [24] A. Achim, E. E. Kuruoglu, J. Zerubia, SAR image filtering based on the heavy-tailed rayleigh model, *IEEE Transactions on Image Processing* 15 (9) (2006) 2686–2693.
 - [25] J. Hadamard, Sur les problèmes aux dérivées partielles et leur signification physique, *Princeton Univ. Bull.* 13.
 - [26] Y. Goussard, G. Demoment, J. Idier, A new algorithm for iterative deconvolution of sparse spike trains, *Proc. Int. Conf. ASSP* (1990) 3.
 - [27] T. F. Chan, C. K. Wong, Total variation blind deconvolution., *IEEE Trans Image Process* 7 (3) (1998) 370–375.
 - [28] R. Wiggins, Minimum entropy deconvolution, *IEEE Proceedings Intern. Symp. Computer-Aided Seismic Analysis and Discrimination*.
 - [29] T. Deeming, Deconvolution and reflection coefficient estimation using a generalized minimum entropy principle, *Proc. 51st Ann. Meeting of the Soc. of Exploration Geophysicists* (1981) 1.
 - [30] D. Donoho, On minimum entropy deconvolution, *Applied Time Series Analysis II* (1981) 1.
 - [31] G. Wahba, Constrained regularization for ill-posed linear operator equations with applications in meteorology and medicine, in *Statistical Decision Theory and Related Topics III* New-York: Academic (1982) 383–417.
 - [32] R. Yarlagadda, J. Bednar, T. Watt, Fast algorithms for l_p deconvolution, *IEEE Transactions on Acoustics Speech and Signal Processing ASSP-33*.

- [33] G. Demoment, R. Reynaud, Fast minimum-variance deconvolution, *IEEE Transactions on Acoustics Speech and Signal Processing ASSP-33* (1985) 1324–1326.
- [34] A. Mohammad-Djafari, G. Demoment, Image restoration and reconstruction using entropy as a regularization functional, *Maximum Entropy and Bayesian Methods in Science and Engineering 2* (1988) 341–355.
- [35] T. J. Holmes, Blind deconvolution of quantum-limited incoherent imagery: maximum-likelihood approach., *J Opt Soc Am A* 9 (7) (1992) 1052–1061.
- [36] S. U. Pillai, B. Liang, Blind image deconvolution using a robust GCD approach., *IEEE Trans Image Process* 8 (2) (1999) 295–301.
- [37] M. K. Ng, R. J. Plemmons, S. Qiao, Regularization of FIR blind image deconvolution., *IEEE Trans Image Process* 9 (6) (2000) 1130–1134.
- [38] S. Jefferies, K. Schulze, C. Matson, K. Stoltenberg, E. K. Hege, Blind deconvolution in optical diffusion tomography., *Opt Express* 10 (1) (2002) 46–53.
- [39] S. Fiori, Fast fixed-point neural blind-deconvolution algorithm., *IEEE Trans Neural Netw* 15 (2) (2004) 455–459.
- [40] L. Wei, L. Hua-ming, Q. Pei-wen, Sparsity enhancement for blind deconvolution of ultrasonic signals in nondestructive testing application., *Rev Sci Instrum* 79 (1) (2008) 014901.
- [41] H. Liao, M. K. Ng, Blind deconvolution using generalized cross-validation approach to regularization parameter estimation., *IEEE Trans Image Process* 20 (3) (2011) 670–680.
- [42] W. Zuo, Z. Lin, A generalized accelerated proximal gradient approach for total-variation-based image restoration., *IEEE Trans Image Process* 20 (10) (2011) 2748–2759.
- [43] A. G. Marrugo, M. Sorel, F. Sroubek, M. S. Milln, Retinal image restoration by means of blind deconvolution., *J Biomed Opt* 16 (11) (2011) 116016.

- [44] M. Rostami, O. Michailovich, Z. Wang, Image deblurring using derivative compressed sensing for optical imaging application., *IEEE Trans Image Process* 21 (7) (2012) 3139–3149.
- [45] F. Sroubek, G. Cristbal, J. Flusser, A unified approach to superresolution and multichannel blind deconvolution., *IEEE Trans Image Process* 16 (9) (2007) 2322–2332.
- [46] F. Sroubek, P. Milanfar, Robust multichannel blind deconvolution via fast alternating minimization., *IEEE Trans Image Process* 21 (4) (2012) 1687–1700.
- [47] L. Yan, H. Fang, S. Zhong, Blind image deconvolution with spatially adaptive total variation regularization., *Opt Lett* 37 (14) (2012) 2778–2780.
- [48] L. Yan, H. Liu, S. Zhong, H. Fang, Semi-blind spectral deconvolution with adaptive Tikhonov regularization., *Appl Spectrosc* 66 (11) (2012) 1334–1346.
- [49] Y.-W. Tai, X. Chen, S. Kim, S. J. Kim, F. Li, J. Yang, J. Yu, Y. Matsushita, M. S. Brown, Nonlinear camera response functions and image deblurring: theoretical analysis and practice., *IEEE Trans Pattern Anal Mach Intell* 35 (10) (2013) 2498–2512.
- [50] X. Zhu, P. Milanfar, Removing atmospheric turbulence via space-invariant deconvolution., *IEEE Trans Pattern Anal Mach Intell* 35 (1) (2013) 157–170.
- [51] H. Pan, T. Blu, An iterative linear expansion of thresholds for l_1 -based image restoration., *IEEE Trans Image Process* 22 (9) (2013) 3715–3728.
- [52] T. Lelore, F. Bouchara, Fair: a fast algorithm for document image restoration., *IEEE Trans Pattern Anal Mach Intell* 35 (8) (2013) 2039–2048.
- [53] H. Ayasso, 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, 2009*, pp. 1297–1300.

- [54] F. Champagnat, Y. Goussard, J. Idier, Unsupervised deconvolution of sparse spike trains using stochastic approximation, *IEEE Trans. signal processing* 44 (12) (1996) 2988-2998.
- [55] I. Daubechies, M. Defrise, C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint., *Comm. Pure Appl. Math* 57 (2004) 1413–1457.
- [56] J. A. Tropp, A. C. Gilbert, M. J. Strauss, Algorithms for simultaneous sparse approximation. part i: Greedy pursuit, *Signal Processing, special issue "sparse approximations in signal and image processing"* 86 (2006) 572–588.
- [57] J. A. Tropp, Algorithms for simultaneous sparse approximation. part ii: Convex relaxation, *Signal Processing, special issue "Sparse approximations in signal and image processing"* 86 (2006) 589–602.
- [58] E. J. Candès, M. Wakin, S. Boyd, Enhancing sparsity by reweighted l_1 minimization, *Journal of Fourier Analysis and Applications* 14 (2008) 877–905.
- [59] N. Polson, J. Scott., Shrink globally, act locally: sparse Bayesian regularization and prediction, *Bayesian Statistics* 9.
- [60] M. Tipping, Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research* 1 (2001) 211–244.
- [61] A. Mohammad-Djafari, Bayesian approach with prior models which enforce sparsity in signal and image processing, *EURASIP Journal on Advances in Signal Processing Special issue on Sparse Signal Processing* (2012) 2012:52.
- [62] N. Chen, D. Giannakis, R. Herbei, A. J. Majda, An MCMC algorithm for parameter estimation in signals with hidden intermittent instability, *SIAM ASA Journal on Uncertainty Quantification* 2 (1) (2014) 647–669.
- [63] Y. Chen, E. Kuruoglu and H. Cheung So", Optimum linear regression in additive CauchyGaussian noise, *Signal Processing* 106 (2015) 312318.
- [64] K. M. Hanson, G. W. Wechsung, Bayesian approach to limited-angle reconstruction in computed tomography 73 (1983) 1501–1509.

- [65] N. Mascarenhas, C. Santos, P. Cruvinel, Transmission tomography under poisson noise using the anscombe transformation and wiener filtering of the projections, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 423 (2) (1999) 265–271.
- [66] S. Geman, D. Geman, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images., *IEEE Trans Pattern Anal Mach Intell* 6 (6) (1984) 721–741.
- [67] J. Idier, Y. Goussard, Markov modeling for bayesian multi-channel deconvolution, *Proceedings of IEEE ICASSP* (1990) 2.
- [68] G. Gindi, M. Lee, A. Rangarajan, Z. I., Bayesian reconstruction of functional images using anatomical information as priors, *IEEE Trans. on Medical Imaging MI-12* (4) (1993) 670–680.
- [69] H. Rue, S. Martino, N. Chopin, Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion), *Journal of the Royal Statistical Society, Series B* 71 (2) (2009) 319–392.
- [70] W. Gilks, S. Richardson, D. Spiegelhalter, *Monte Carlo Markov Chains in practice*, Chapman and Hall, 1996.
- [71] P. Del Moral, A. Doucet, A. Jasra, An adaptive sequential monte carlo method for approximate bayesian computation, *Tech. rep.*, Imperial College, London (2009).
- [72] D. Ge, J. Idier, E. Le Carpentier, A new MCMC algorithm for blind Bernoulli- Gaussian deconvolution, in: *Proceedings of EUSIPCO: Septembre 2008; Lausanne, Suisse, 2008*.
- [73] M. Beaumont, J. Cornuet, J. Marin, C. Robert, Adaptive approximate bayesian computation, *Biometrika* 96 (4) (2009) 983–990.
- [74] M. Blum, Approximate bayesian computational: a non-parametric perspective, *Journal of the American Statistical Association* 491 (2010) 1178–1187.

- [75] M. Blum, O. François, Non-linear regression models for approximate bayesian computation, *Statistics and Computing* 20 (1) (2010) 63–73.
- [76] J. Marin, P. Pudlo, C. Robert, R. Ryder, Approximate bayesian computational methods, arxiv:1101.0955.
- [77] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, L. Zdeborová, Statistical-physics-based reconstruction in compressed sensing, *Phys. Rev. X* 2 (2012) 021005.
- [78] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, L. Zdeborová, Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices, *Journal of Statistical Mechanics: Theory and Experiment* 2012 (08) (2012) P08009.
- [79] M. Beal, Variational Algorithms for Approximate Bayesian Inference, Ph.D. thesis, Gatsby Computational Neuroscience Unit, University College London (2003).
- [80] J. Winn, C. M. Bishop, T. Jaakkola, Variational message passing, *Journal of Machine Learning Research* 6 (2005) 661–694.
- [81] J. Parker, V. Cevher, P. Schniter, Compressive sensing under matrix uncertainties: An approximate message passing approach, in: *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on, 2011*, pp. 804 –808.
- [82] P. Schniter, S. Rangan, Compressive Phase Retrieval via Generalized Approximate Message Passing, in: *Proceedings of Allerton Conference on Communication, Control, and Computing, 2012*.
- [83] C. Archambeau, M. Verleysen, Robust bayesian clustering., *Neural Netw* 20 (1) (2007) 129–138.
- [84] A. Honkela, H. Valpola, Variational learning and bits-back coding: an information-theoretic view to bayesian learning., *IEEE Trans Neural Netw* 15 (4) (2004) 800–810.
- [85] V. P. Oikonomou, D. I. Fotiadis, A bayesian approach for the estimation of ar coefficients from noisy biomedical data., *Conf Proc IEEE Eng Med Biol Soc 2007* (2007) 3270–3273.

- [86] D. Shutin, C. Zechner, S. R. Kulkarni, H. V. Poor, Regularized variational bayesian learning of echo state networks with delay&sum readout., *Neural Comput* 24 (4) (2012) 967–995.
- [87] K. Watanabe, S. Watanabe, Stochastic complexities of general mixture models in variational bayesian learning., *Neural Netw* 20 (2) (2007) 210–219.
- [88] J. Gao, Robust l_1 principal component analysis and its Bayesian variational inference., *Neural Comput* 20 (2) (2008) 555–572.
- [89] J. Daunizeau, K. J. Friston, S. J. Kiebel, Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models., *Physica D* 238 (21) (2009) 2089–2118.
- [90] K. Watanabe, S. Akaho, S. Omachi, M. Okada, Variational bayesian mixture model on a subspace of exponential family distributions., *IEEE Trans Neural Netw* 20 (11) (2009) 1783–1796.
- [91] G. Chantas, N. P. Galatsanos, R. Molina, A. K. Katsaggelos, Variational bayesian image restoration with a product of spatially weighted total variation image priors., *IEEE Trans Image Process* 19 (2) (2010) 351–362.
- [92] K. Watanabe, M. Okada, K. Ikeda, Divergence measures and a general framework for local variational approximation., *Neural Netw* 24 (10) (2011) 1102–1109.
- [93] R. Molina, J. Mateos, A. K. Katsaggelos, Blind deconvolution using a variational approach to parameter, image, and blur estimation., *IEEE Trans Image Process* 15 (12) (2006) 3715–3727.
- [94] S. D. Babacan, R. Molina, A. K. Katsaggelos, Parameter estimation in tv image restoration using variational distribution approximation., *IEEE Trans Image Process* 17 (3) (2008) 326–339.
- [95] J. A. Dobrosotskaya, A. L. Bertozzi, A wavelet-Laplace variational technique for image deconvolution and inpainting., *IEEE Trans Image Process* 17 (5) (2008) 657–663.

- [96] S. D. Babacan, R. Molina, A. K. Katsaggelos, Variational Bayesian blind deconvolution using a total variation prior., *IEEE Trans Image Process* 18 (1) (2009) 12–26.
- [97] S. Babacan, J. Wang, R. Molina, A. Katsaggelos, Bayesian blind deconvolution from differently exposed image pairs., *IEEE Trans Image Process* 19 (11).
- [98] V. Pasquale Roche, A. Mohamad-Djafari, F. I. Pasquale, A. Karabou, A. Gorbach, F. Lvi, Thoracic surface temperature rhythms as circadian biomarkers for cancer chronotherapy, *International Early Online* (2013) 1–12.
- [99] X. Li, A. Mohammad-Djafari, M. Dumitru, S. Dulong, E. Filipski, S. Siffroi-Fernandez, A. Mteyrek, F. Scaglione, C. Guettier, F. Delaunay, F. Lévi, A circadian clock transcription model for the personalization of cancer chronotherapy. *Cancer research* 73: 24. 7176-7188 Dec(2013).
- [100] S. C. Davis, H. Dehghani, J. Wang, S. Jiang, B. W. Pogue and K. D. Paulsen, Image-guided diffuse optical fluorescence tomography implemented with Laplacian-type regularization”, *Opt. Express*, vo. 15, num. 7, pp 4066–4082.
- [101] A. Mohammad-Djafari, Bayesian Blind Deconvolution using a Student-t prior model and Variational Bayesian Approximation, paper 2, Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields, 22-23 August 2014, Koc University, Istanbul, Turkey. http://conf-scoop.org/INCT-2014/2.Djafari_INCT.pdf
- [102] Bayesian Blind Deconvolution of images comparing JMAP, EM and BVA with a Student-t a priori model, paper 9, Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields, 22-23 August 2014, Koc University, Istanbul, Turkey. http://conf-scoop.org/ACV-2014/9.Djafari_ACV.pdf
- [103] S. D. Babacan, R. Molina, M. N. Do and A. K. Katsaggelos, Bayesian Blind Deconvolution with General Sparse Image Priors, *European Conference on Computer Vision (ECCV)*, Firenze, Italy, October 2012