



**HAL**  
open science

## Model selection in the sparsity context for inverse problems in Bayesian framework

Mircea Dumitru, Li Wang, Ali Mohammad-Djafari, Nicolas Gac

► **To cite this version:**

Mircea Dumitru, Li Wang, Ali Mohammad-Djafari, Nicolas Gac. Model selection in the sparsity context for inverse problems in Bayesian framework. 37th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Jul 2017, Jarinu, Brazil. 10.1007/978-3-319-91143-4\_15 . hal-01568318

**HAL Id: hal-01568318**

**<https://hal.science/hal-01568318v1>**

Submitted on 25 Jul 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Model selection in the sparsity context for inverse problems in Bayesian framework

Mircea Dumitru and Li Wang and Ali Mohammad-Djafari and Nicolas Gac

**Abstract** The Bayesian approach is considered for inverse problems with a typical forward model accounting for errors and *a priori* sparse solutions. Solutions with sparse structure are enforced using heavy tailed prior distributions. The particular case of such prior expressed via normal variance mixtures with conjugate laws for the mixing distribution is the main interest of this paper. Such a prior is considered in this paper, namely the Student-t distribution. Iterative algorithms are derived via posterior mean estimation. The mixing distribution parameters appear in updating equations and are also used for the initialization. For the choice of mixing distribution parameters, three model selection strategies are considered: *i*) parameters approximating the mixing distribution with Jeffrey law, i.e. keeping the mixing distribution well defined but as close as possible to the Jeffreys priors, *ii*) based on prior distribution form, fixing the parameters corresponding to the form inducing the most sparse solution and *iii*) based on sparsity mechanism, fixing the hyperparameters using the statistical measures of the mixing and prior distribution. For each strategy of model selection, the theoretical advantages and drawbacks are discussed and the corresponding simulations are reported for a 1D direct sparsity application in a biomedical context.

## 1 Introduction

In this paper, we compare three model selection strategies for a particular context of the Bayesian approach for inverse problems. More precisely, we consider a linear model describing the forward problem and the available prior information the sparse structure of the unknown. The sparse structure is modeled via heavy tailed priors ( $\mathcal{P}$ ), well known in the literature for enforcing sparsity [7], [1], [3]. The

---

Mircea Dumitru, Li Wang, Ali Mohammad-Djafari, Nicolas Gac  
Laboratoire des signaux et systèmes, CNRS – CentraleSupélec – Université Paris-Saclay, 3, Rue Joliot-Curie, 91192 Gif sur Yvette, France, e-mail: surname.name@lss.supelec.fr

particular class of priors considered in this article is the zero mean normal variance mixtures. The unknowns are estimated using the Posterior Mean (PM) estimation via Variational Bayesian Approximation (VBA), [10], [2]. Typically, the initialization of the derived iterative algorithm is done using the hyperparameters [4], i.e. the mixing distributions ( $\mathcal{M}$ ) parameters. Therefore, the model selection is a crucial step in such algorithms. In this specific context, three different strategies for model selection are considered and compared for the particular Student-t prior case.

We consider the linear forward model,

$$g = Hf + \varepsilon, \quad (1)$$

where  $g$  represents the  $N \times 1$  observed data,  $H$  represents a  $N \times M$  measurement matrix,  $f$  represents the unknown *sparse* signal and  $\varepsilon$  accounts for measurement and modeling errors. In this paper, the sparsity is accounted in the Bayesian hierarchical prior models framework, using sparsity enforcing prior distributions to model  $f_j$ ,  $j \in \{1, \dots, M\}$ , [7], [8]. For computational reasons we consider heavy tailed distributions expressed via zero mean normal variance mixtures with conjugate priors as mixing distributions,

$$\begin{cases} p(f_j | v_{f_j}) = \mathcal{N}(f_j | 0, v_{f_j}) \\ p(v_{f_j} | \xi_f) = \mathcal{M}(v_{f_j} | \xi_f) \end{cases}, \quad (2)$$

where  $\xi_f$  represents the parameters of the mixing distribution. In this article, the Inverse Gamma distribution ( $\mathcal{IG}$ ), corresponding respectively to the two parameters Student-t ( $\mathcal{St}$ ) distribution will be considered for simulations results and comparisons between the different model selection strategies considered. However, the framework is general and can be used for other sparsity enforcing priors expressed as normal variance mixture, e.g. the Normal-Inverse Gaussian ( $\mathcal{NIG}$ ) distribution and the Variance-Gamma ( $\mathcal{VG}$ ) distribution. The *non-stationary independent Gaussian uncertainties model* is assumed with conjugate priors modeling the variances,

$$\begin{cases} p(\varepsilon_i | v_{\varepsilon_i}) = \mathcal{N}(\varepsilon_i | 0, v_{\varepsilon_i}) \\ p(v_{\varepsilon_i} | \xi_\varepsilon) = \mathcal{M}(v_{\varepsilon_i} | \xi_\varepsilon) \end{cases}, \quad (3)$$

where  $\xi_\varepsilon$  represents the parameters of the mixing distribution. For the derived iterative algorithms, the parameters of the *posterior* mixing distributions  $\hat{\xi}_f$  and  $\hat{\xi}_\varepsilon$  modeling variances  $v_{f_j}$  and  $v_{\varepsilon_i}$ , have updating expressions depending on  $\xi_f$  and  $\xi_\varepsilon$ . The model selection i.e. the choice of *prior* mixing parameters  $\xi_f$  and  $\xi_\varepsilon$  is therefore crucial in the context of non-supervised algorithms. In practice, such algorithms can be obtained by considering non-informative prior mixing distributions, i.e. considering the Jeffreys prior as the mixing distribution (more exactly, conserving the *conjugate prior* setting, using the conjugate prior with parameters values  $\xi_f$  and  $\xi_\varepsilon$  such that the corresponding mixing prior is close to Jeffreys prior). This approach was successfully used in [9].

Two other model selection approaches accounting for the sparsity particular context

and the specific sparsity enforcing priors used are considered. The first one is based on the form of the prior distribution, i.e. a model selection strategy considering the parameters for which the prior distribution is as concentrated as possible around the mean. For the second one, we first show that the variance of the posterior distribution  $\text{Var}_{[\mathcal{D}]}(f_j)$ , modeling  $f$  is linked with the expectation of the  $\mathcal{M}$  distribution  $E_{[\mathcal{M}]}$ , modeling the corresponding variance  $v_f$ . More precisely

$$\text{Var}_{[\mathcal{D}]}(f_j) = E_{[\mathcal{M}]}(v_{f_j}). \quad (4)$$

Secondly, we consider a small variance for the prior distribution, i.e.  $\text{Var}_{[\mathcal{D}]}(f_j) = \varepsilon \searrow 0$  in order to impose a model that is concentrating the points  $f_j$  around the zero mean. Clearly, doing this, via Eq. (4), the expectation of the mixture distribution has the same value and doing the same for the mixture distribution variance,  $\text{Var}_{[\mathcal{M}]}(v_{f_j}) = \omega \searrow 0$  will impose a sparse structure for  $v_f$ , with small values  $v_{f_j}$  corresponding to the small values  $f_j$  and *significant* values  $v_{f_j}$  corresponding to the *significant* values  $f_j$ . A sparse structure is therefore enforced by:

1. considering a heavy tailed prior distribution.
2. setting a small variance for prior distribution,  $\text{Var}_{[\mathcal{D}]}(f_j) = \varepsilon \searrow 0$ .
3. enforcing a sparse structure for  $v_f$  by setting a small variance also for the mixing distribution  $\text{Var}_{[\mathcal{M}]}(v_{f_j}) = \omega \searrow 0$ .

The rest of the paper is organized as follows. Sec. (2) is introducing the general hierarchical prior model, setting the context of the particular class of sparsity enforcing prior used, and presenting the normal variance mixtures considered during paper and their behaviour depending on the parameters. The corresponding PM algorithms are developed in Sec. (3). Empirical evaluations of performances and comparisons between the results corresponding to the two approaches for modeling the hyperparameters are presented in Sec. (4). Conclusions are drawn in Sec. (5).

## 2 Hierarchical prior models based on normal variance mixtures

The framework of the hierarchical prior model discussed in this paper, Fig. (1), is based on sparsity enforcing prior distributions expressed as marginals of normal variance mixtures and non-stationary independent Gaussian uncertainties (noise) model with conjugate priors modeling the variances. The posterior distribution writes

$$p(f, v_\varepsilon, v_f | g) \propto \mathcal{N}(g | Hf, v_\varepsilon) \mathcal{N}(f | 0, v_f) \mathcal{M}(v_{f_j} | \xi_f) \mathcal{M}(v_{\varepsilon_i} | \xi_\varepsilon). \quad (5)$$

In this specific framework, the product of the two conditional distributions  $p(g | f, v_\varepsilon) = \mathcal{N}(g | Hf, v_\varepsilon)$  and  $p(f | v_f) = \mathcal{N}(f | 0, v_f)$  is common to the posterior distribution, while the differences are induced by the choice of the mixing distributions  $p(v_{f_j} | \xi_f) = \mathcal{M}(v_{f_j} | \xi_f)$  and  $p(v_{\varepsilon_i} | \xi_\varepsilon) = \mathcal{M}(v_{\varepsilon_i} | \xi_\varepsilon)$ . We consider in the fol-

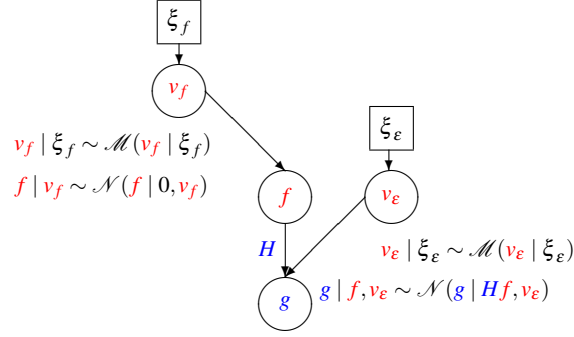


Fig. 1: Hierarchical prior model for forward model, Eq. (1).

lowing the particular case of the Student-t prior expressed as a normal variance mixture.

## 2.1 Inverse Gamma mixing distribution

In Eq. (2), the Inverse Gamma is considered as the mixing distribution,  $\mathcal{M}(v_{f_j} | \xi_f) = \mathcal{IG}(v_{f_j} | \alpha_f, \beta_f)$ , with the probability density function given by:

$$\mathcal{IG}(v_{f_j} | \alpha_f, \beta_f) = \frac{\beta_f^{\alpha_f}}{\Gamma(\alpha_f)} v_{f_j}^{-\alpha_f-1} \exp\left(-\frac{\beta_f}{v_{f_j}}\right), \quad \alpha_f > 0, \beta_f > 0, \quad (6)$$

where  $\Gamma(\cdot)$  denotes the Gamma function. The corresponding hyperparameters are  $\xi_f = (\alpha_f, \beta_f)$  and the corresponding prior  $p(f_j | \alpha_f, \beta_f)$  is a two-parameter  $\mathcal{St}$  distribution:

$$p(f_j | \alpha_f, \beta_f) = \frac{\Gamma(\alpha_f + \frac{1}{2})}{\sqrt{2\pi\beta_f}\Gamma(\alpha_f)} \left(1 + \frac{f_j^2}{2\beta_f}\right)^{-(\alpha_f + \frac{1}{2})} = \mathcal{St}(f_j | \alpha_f, \beta_f). \quad (7)$$

$\alpha_f = \beta_f = v_f/2$  corresponds to the standard  $\mathcal{St}$  form, [4]. The expectation of mixing distribution  $\mathcal{IG}$  (equal to the variance of the  $\mathcal{St}$  distribution) and the variance of the mixing distribution  $\mathcal{IG}$  are given by

$$\mathbb{E}_{[\mathcal{IG}]}(v_{f_j}) = \text{Var}_{[\mathcal{St}]}(f_j) = \frac{\beta_f}{\alpha_f - 1}; \quad \text{Var}_{[\mathcal{IG}]}(v_{f_j}) = \frac{\beta_f^2}{(\alpha_f - 1)^2 (\alpha_f - 2)}, \quad (8)$$

with  $\alpha_f > 1$  for the first equality and  $\alpha_f > 2$  for the second one. This model gives the possibility to consider a heavy tailed distribution to model the sparse structure of  $f$ . It is expressed via the Normal distribution and a conjugate prior, which has great computational advantages, guaranteeing the same family distributions for the pos-

terior distributions. The choice of the (prior) parameters  $\alpha_f$  and  $\beta_f$  plays a crucial role. Different approaches can be considered to chose the parameters.

1. An approach based on the prior distribution form, imposing a small value for  $\beta_f$  and a large value for  $\alpha_f$ . Establishing how small or how large the parameters should be setted is difficult.
2. *Close to Jeffreys* prior, setting both parameters close to zero. As before, the same difficulty of establishing *how close to zero* the parameters should be fixed is encountered.
3. Using Eq. (8) to fix the parameters depending on the mixing and prior distribution moments. The relation between  $\alpha_f$  and  $\beta_f$  parameters and the moments  $\varepsilon$  and  $\omega$  is given by

$$\alpha_f = 2 + \frac{\varepsilon^2}{\omega} ; \beta_f = \varepsilon \left( 1 + \frac{\varepsilon^2}{\omega} \right). \quad (9)$$

This model selection is based on the data characteristics, i.e. the sparse structure. However, the same difficulty appears for establishing *how small* should  $\varepsilon$  and  $\omega$  be.

Table (1) resumes the three strategies considered for model selection.

Parameters	Mixing distribution	Prior distribution	Moments
$\alpha_f \searrow 0; \beta_f \searrow 0$	$\mathcal{I}\mathcal{G}(\mathbf{v}_{f_j}   \alpha_f, \beta_f)$	$\mathcal{S}t(\mathbf{f}_j   \alpha_f, \beta_f)$	not defined
Simulates the Jeffreys prior and has the advantages of a conjugate prior distribution but is difficult to measure <i>how close</i> to 0 should the two parameters should be chosen.			
$\alpha_f \nearrow \infty; \beta_f \searrow 0$	$\mathcal{I}\mathcal{G}(\mathbf{v}_{f_j}   \alpha_f, \beta_f)$	$\mathcal{S}t(\mathbf{f}_j   \alpha_f, \beta_f)$	not defined
Advantages of a conjugate prior law, parameters chosen in accordance with their influence on the distribution form but difficult to measure <i>how close</i> to $\infty$ (0) should $\alpha_f$ ( $\beta_f$ ) be fixed			
$\alpha_f = 2 + \frac{\varepsilon^2}{\omega}; \beta_f = \varepsilon \left( 1 + \frac{\varepsilon^2}{\omega} \right)$	$\mathcal{I}\mathcal{G}(\mathbf{v}_{f_j}   \alpha_f, \beta_f)$	$\mathcal{S}t(\mathbf{f}_j   \alpha_f, \beta_f)$	defined
Advantages of a conjugate prior law, parameters chosen in accordance with data structure, considers the moments of the $\mathcal{I}\mathcal{G}$ and $\mathcal{S}t$ laws; same difficulties as above for $\varepsilon$ and $\omega$			

Table 1: Student-t sparsity enforcing prior: model selection strategies

### 3 PM estimation via VBA

The PM estimation is considered via Variational Bayesian Approximation (VBA). The posterior distribution is first approximated with a separable one,

$$p(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\varepsilon | \mathbf{g}) \approx q(\mathbf{f}, \mathbf{v}_f, \mathbf{v}_\varepsilon | \mathbf{g}) = q_1(\mathbf{f}) \prod_{j=1}^M q_{2j}(\mathbf{v}_{f_j}) \prod_{i=1}^N q_{3i}(\mathbf{v}_{\varepsilon_i}), \quad (10)$$

by minimizing the Kullback-Leibler divergence. Proportionality relations for each separable distribution are obtained. It can be shown that  $q_1(\mathbf{f})$  is a multivariate

Normal distribution,

$$q_1(f) = \mathcal{N}(f | \hat{f}, \hat{\Sigma}); \hat{f} = (H^T \tilde{V}_\varepsilon H + \tilde{V}_f)^{-1} H^T \tilde{V}_\varepsilon g, \quad \hat{\Sigma} = (H^T \tilde{V}_\varepsilon H + \tilde{V}_f)^{-1}, \quad (11)$$

using the notations

$$\begin{aligned} \tilde{v}_{\varepsilon_i} &= \left\langle v_{\varepsilon_i}^{-1} \right\rangle_{q_{3i}(v_{\varepsilon_i})}, i \in \{1, \dots, N\}; \tilde{v}_\varepsilon = [\dots \tilde{v}_{\varepsilon_i} \dots]^T; \tilde{V}_\varepsilon = \text{diag}[\tilde{v}_\varepsilon], \\ \tilde{v}_{f_j} &= \left\langle v_{f_j}^{-1} \right\rangle_{q_{2j}(v_{f_j})}, j \in \{1, \dots, M\}; \tilde{v}_f = [\dots \tilde{v}_{f_j} \dots]^T; \tilde{V}_f = \text{diag}[\tilde{v}_f]. \end{aligned} \quad (12)$$

In general,  $q_{2j}(v_{f_j})$  and  $q_{3i}(v_{\varepsilon_i})$  belong to the same family as the  $\mathcal{M}$  distribution.

In particular, for the  $\mathcal{S}t$  prior,  $q_{2j}(v_{f_j})$  are  $\mathcal{S}G(v_{f_j} | \alpha_f + \frac{1}{2}, \hat{\beta}_{f_j})$  and  $q_{3i}(v_{\varepsilon_i})$  are  $\mathcal{S}G(v_{\varepsilon_i} | \alpha_\varepsilon + \frac{1}{2}, \hat{\beta}_{\varepsilon_i})$  distributions, with the analytical expressions of  $\beta$  parameters given by:

$$\hat{\beta}_{f_j} = \beta_f + \frac{1}{2} (\hat{\Sigma}_{jj} + f_j^2); \hat{\beta}_{\varepsilon_i} = \beta_\varepsilon + \frac{1}{2} [H_i \hat{\Sigma} H_i^T + (g_i - H_i \hat{f})^2] \quad (13)$$

The parameters corresponding to the multivariate Normal distribution are expressed via  $\tilde{V}_f$  and  $\tilde{V}_\varepsilon$  (and by extension all elements forming the two matrices  $\tilde{v}_{f_j}$ ,  $j \in \{1, 2, \dots, M\}$  and  $\tilde{v}_{\varepsilon_i}$ ,  $i \in \{1, 2, \dots, N\}$ , Eq. (12)). The following relation holds:

$$\langle x^{-1} \rangle_{\mathcal{S}G(x|\alpha, \beta)} = \frac{\alpha}{\beta} \quad (14)$$

## 4 Simulation results

The forward model Eq. (1) is considered for a 1-D application in biology where a short time series of gene expressions is modeled as

$$g(t) = \sum_{j=1}^M f_{1j} \cos\left(\frac{2\pi}{p_j} t\right) + f_{2j} \sin\left(\frac{2\pi}{p_j} t\right) + \varepsilon(t), \quad (15)$$

where  $p_j \in [8, \dots, 32]$ ,  $t = 1, \dots, N$ , and the objective is to find  $f_1 = [f_{1j}, \dots, f_{1M}]$  and  $f_2 = [f_{2j}, \dots, f_{2M}]$ . This relation can be written as:

$$g = Hf + \varepsilon = [H_1 | H_2] \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} + \varepsilon = H_1 f_1 + H_2 f_2 + \varepsilon. \quad (16)$$

The objective is a precise estimation of the periodic component (PC) vectors ( $f_1$  and  $f_2$ , considered between 8 and 32 hours) corresponding to a short (relative to the a priori dominant period) signal  $g$  (considered for four days, sampled every hour). The biological prior information is the reduced number of non-zero periods, i.e. the

sparse structure of the PC vector. More details about the application and limitations of the classical methods can be found in [5] and [4].

The synthetic PC vectors  $f_1$  and  $f_2$  are presented in Fig. (2a) and Fig. (2b), the corresponding biological signal in Fig. (2d) and the added noise (corresponding to 10dB) in Fig. (2c). We consider the PM iterative algorithm, corresponding to the

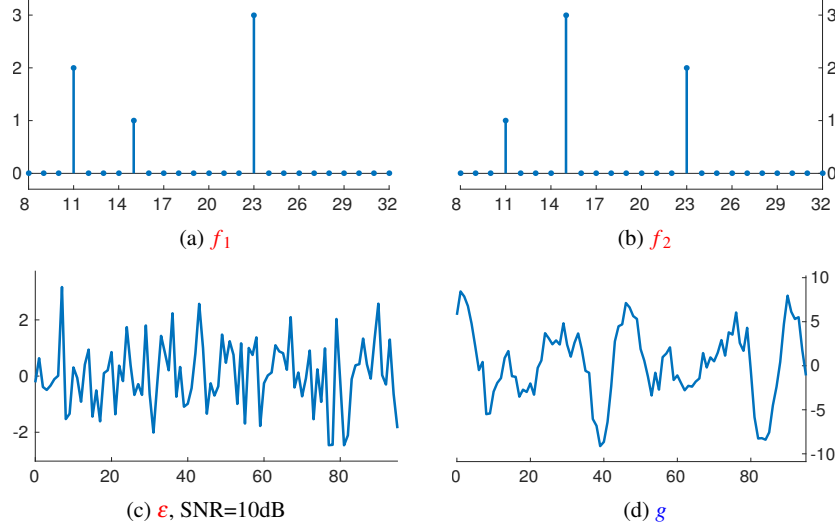


Fig. 2: Synthetic data: 4-days noisy signal  $g$  sampled every hour and corresponding PC vectors  $f_1$  and  $f_2$ .

Student-t prior model. The three different model selection strategies are considered. The algorithm initialization is done via the hyperparameters, i.e. the  $\mathcal{M}$  distribution parameters, modeling respectively  $v_f$  and  $v_\varepsilon$ . In this paper we are interested in testing which model selection strategy is most adequate in enforcing the sparsity, for this specific prior model and this specific application. The results discussed in this section are limited to this aspect. For this purpose, the simulations are done assuming  $E_{[\mathcal{M}]}(v_\varepsilon)$  and  $Var_{[\mathcal{M}]}(v_\varepsilon)$  to be known, and then the noise hyperparameters, namely the  $\mathcal{M}$  parameters modeling  $v_\varepsilon$ , are derived, depending on the noise model, via Eq. (9).

The estimated PC vectors  $\hat{f}_1$  and  $\hat{f}_2$ , corresponding to the  $\mathcal{S}t$  prior, are compared with synthetic data, in Fig. (3).

1. Results corresponding to the Jeffreys prior model selection are reported in Fig. (3a) and (3b). The results correspond to  $\alpha_f = \beta_f = 10^{-3}$ . Clearly, the estimation results are far from synthetic input. More important, for this model selection the prior distribution does not enforce sparsity very well. Indeed, for the the first periods of the PC vector the estimation is rather good with the zero values well estimated, but for the second part of the PC vector the sparsity is not enforced.



2. Results corresponding to the model selection based on the  $\mathcal{S}t$  prior distribution are reported in Fig. (3c) and (3d). The results correspond to  $\alpha_f = 10^3, \beta_f = 10^{-3}$ . In this case, the result is too sparse: all PC vector values are estimated as zero values. Those particular values correspond for a *strong* prior, which does not account for data. We will see that this model selection strategy can lead to very good estimation results. In particular, for  $\alpha_f = 10^1, \beta_f = 10^{-1}$  both estimations  $\hat{f}_1$  and  $\hat{f}_2$  are precise.
3. Finally, the results corresponding to the model selection based on the sparsity mechanism are reported in Fig. (3e) and (3f). The results correspond to  $\varepsilon = \omega = 10^{-3}$  (see Eq. (9)).

The reconstruction results corresponding to different hyperparameters values, for each of the three model selection strategies, are reported in Table (2). We consider  $k \in \{1, 2, 3, 4, 5\}$  and the following values corresponding to each model selection strategy:

1. for Jeffreys like model selection, we consider  $\alpha_f = \beta_f = 10^{-k}$ .
2. for model selection based on the  $\mathcal{S}t$  form, we consider  $\alpha_f = 10^k; \beta_f = 10^{-k}$ .
3. for model selection based on sparsity mechanism,  $\alpha_f$  and  $\beta_f$  are defined via Eq. (9), using  $\varepsilon = \text{Var}_{[\mathcal{S}t]} = \text{E}_{[\mathcal{S}t]}$  and  $\omega = \text{Var}_{[\mathcal{S}t]}$ . We consider  $\varepsilon = \omega = 10^{-k}$ .

Table (2) reports the quality reconstruction not in the sense of some numerical measure, like  $L_1$  or  $L_2$  reconstruction errors but rather if the results are as sparse as the synthetic inputs  $f_1$  and  $f_2$ . We notice that the model selection strategy based on the

	Jeffreys like $\alpha_f = \beta_f = 10^{-k}$	$\mathcal{S}t$ form $\alpha_f = 10^k; \beta_f = 10^{-k}$	Sparsity mechanism $\varepsilon = \omega = 10^{-k}$ , Eq. (9)
$k = 1$	✗	✓	✓
$k = 2$	✗	✗	✓
$k = 3$	✗	✗	✓
$k = 4$	✗	✗	✗
$k = 5$	✗	✗	✗

Table 2: m. s. for  $\mathcal{S}t$  prior model: qualitative estimation depending on the numerical values for  $\alpha_f$  and  $\beta_f$  parameters.

sparsity mechanism, for the  $\mathcal{S}t$  prior model is more flexible. Good results can be achieved using the  $\mathcal{P}$  distribution form. In this case, having to set a small value for one hyperparameter ( $\beta_f$ ) and a significant value for the other ( $\alpha_f$ ) is a difficult task, since each of both hyperparameters are influencing the  $\mathcal{P}$  distribution form, enforcing sparsity to much. Model selection strategy based on the sparsity mechanism is using the statistical measures of the unknown of the model and its corresponding variance, which generally can be approximately inferred in each application. Moreover, some preliminary results are indicating a strong influence of  $\varepsilon$  and a weak influence of  $\omega$  in the model selection, so reducing it to only one parameter.

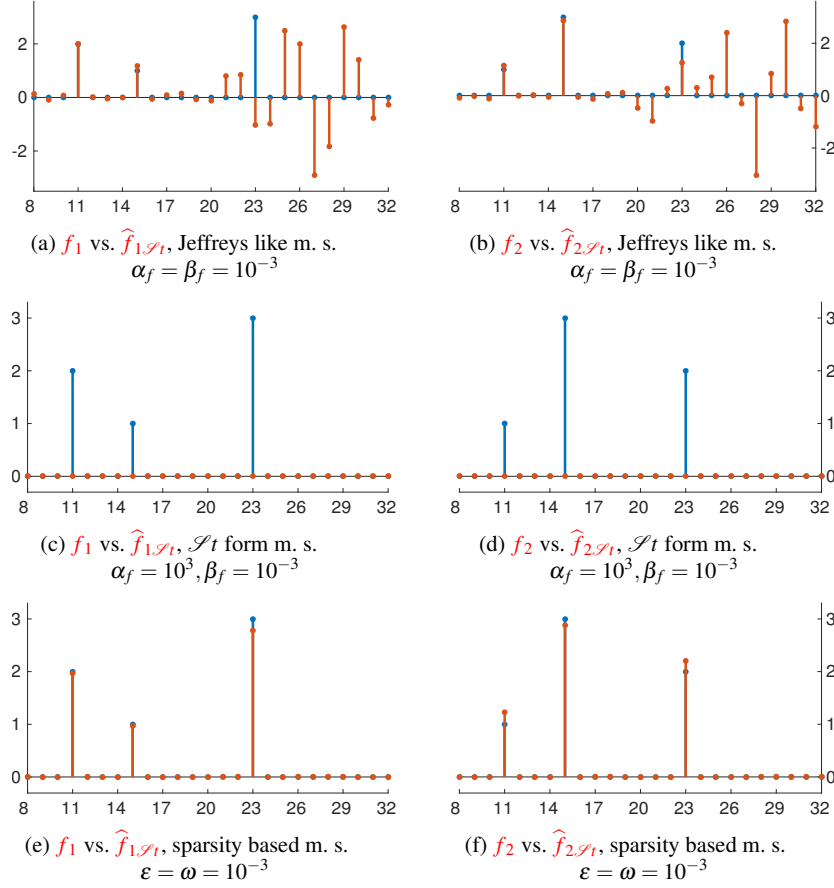


Fig. 3: Comparison between the data  $f_1, f_2$  and  $\hat{f}_{1\mathcal{S}_t}, \hat{f}_{2\mathcal{S}_t}$ . Three model selection strategies,  $\mathcal{S}_t$  prior model: Jeffreys like, ( $\alpha_f = \beta_f = 10^{-3}$ , (3a) and (3b)), based on  $\mathcal{S}_t$  form, ( $\alpha_f = 10^3, \beta_f = 10^{-3}$ , (3c) and (3d)) and sparsity based, ( $\alpha_f = 2 + \frac{\varepsilon^2}{\omega}, \beta_f = \varepsilon \left(1 + \frac{\varepsilon^2}{\omega}\right), \varepsilon = \omega = 10^{-3}$ , (3e) and (3f)).

## 5 Conclusion

In the Bayesian framework, using heavy-tailed distributions in order to enforce sparsity, we have compared the PM iterative algorithms reconstruction results corresponding to a specific sparsity enforcing law (Student-t) corresponding to three different model selection strategies in terms of sparsity enforcing.

The model selection strategy based on the sparsity mechanism can lead to good results in terms of sparsity enforcing and seems to be more flexible than the other two strategies considered. This model selection strategy is based on the assignment of

the hyperparameters using the statistical measures of the prior and mixing distributions,  $\varepsilon$  and  $\omega$ . The interval for selecting  $\varepsilon$  and  $\omega$  is rather large in all three cases. Some preliminary results indicate a weak dependency between  $\omega$  the reconstruction results in terms of sparsity enforcing.

The sparsity is not enforced when the model selection strategy based on Jeffreys priors is used. The model selection strategy based on the prior distribution form can give good reconstruction results in terms of sparsity enforcing but in this case the interval seems to be rather small.

Clearly, the results strongly depend on the application and on the specific prior law (induced by the choice of the mixing distribution). For future work, those strategies will be compared for other sparsity enforcing distributions and other applications, [6]. Also, a key concept is the *sparsity rate* (SR). Another perspective of this work is to study a possible relation between the SR and the model selection, more precisely a link between the model selection and SR.

We mention that in this paper we have measured the reconstruction results in terms of sparsity enforcing. Evidently, this is just the first step in a much more detailed analysis, accounting also for different reconstruction measures, like  $L_1$ ,  $L_2$  reconstruction errors, false positives, etc. This paper reports the preliminary results corresponding to the best model selection strategies for the Student-t prior model in terms of sparsity enforcing.

## References

1. S. Derin Babacan, Shinichi Nakajima, and Minh N. Do. Bayesian Group-Sparse Modeling and Variational Inference. *IEEE Trans. Signal Process.*, 14:2906–2921, 2014.
2. Matthew J. Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, University College London, 2003.
3. Nicolas Dobigeon, Alfred O. Hero, and Jean-Yves Tournet. Hierarchical Bayesian Sparse Image Reconstruction with Application to MRFM. *IEEE Transactions on Image Processing*, 18:2059–2070, 2009.
4. Mircea Dumitru. *A Bayesian Approach for Periodic Components Estimation for Chronobiological Signals*. PhD thesis, Université Paris-Saclay, Mar 2016.
5. Mircea Dumitru, Ali Mohammad-Djafari, and Baghai Simona S. Precise periodic components estimation for chronobiological signals through Bayesian Inference with sparsity enforcing prior. *EURASIP Journal on Bioinformatics and Systems Biology, Springer*, 3, 2016.
6. Mircea Dumitru, Li Wang, Nicolas Gac, and Ali Mohammad-Djafari. Computed tomography reconstruction based on a hierarchical model and variational Bayesian method. In *2017 IEEE Int. Conf. on Image, Processing (ICIP)*. IEEE, 2017.
7. Ali Mohammad-Djafari. Bayesian approach with prior models which enforce sparsity in signal and image processing. *EURASIP Journal on Advances in Signal Processing*, 52, 2012.
8. Ali Mohammad-Djafari and Mircea Dumitru. Bayesian sparse solutions to linear inverse problems with non-stationary noise with Student-t priors. *Digital Signal Processing*, 47:128–156, 2015.
9. François Orieux, Jean-François Giovannelli, and Thomas Rodet. Bayesian estimation of regularization and point spread function parameters for wiener–hunt deconvolution. *J. Opt. Soc. Am. A*, 27(7):1593–1607, Jul 2010.
10. Václav Šmídl and Anthony Quinn. *The Variational Bayes Method in Signal Processing*. Signals and Communication Technology. Springer–Verlag Berlin Heidelberg, 1st edition, 2006.