PARAMETER SELECTION STRATEGY FOR SPARSITY ENFORCING PRIOR MODELS

Mircea Dumitru

Laboratoire des signaux et systèmes, CNRS – CentraleSupélec – Université Paris-Saclay, France

ABSTRACT

The Bayesian framework and heavy tailed distributions expressed as continuous Gaussian scale mixtures have been used intensively in the sparsity context. The Posterior Mean corresponding iterative algorithms strongly depend on the parameter selection. We propose a parameter selection strategy based on the link of the mixing and prior distribution. We compare it with other parameter selection strategies for three prior models obtained as particular cases of the Generalized Hyperbolic distribution and show that the proposed parameter selection strategy seem to be more suitable for the sparsity context.

Index Terms— continuous Gaussian scale mixture, parameter selection, sparsity, Generalized Inverse Gaussian, inverse problems

1. INTRODUCTION

In this paper we consider heavy tailed distributions to model sparse structures. In particular, we focus on such distributions expressed as continuous Gaussian scale mixtures with conjugate mixing distributions. This class of distributions is particularly used in signal processing applications using a linear model and performing the inversion in a Bayesian framework. The Normal Inverse Gaussian distribution was considered in [1], in a denoising application, in [2] and [3]. The Variance-Gamma distribution was considered in [4], in a radar application, in [5], in [6] and [7] in an audio application. Its particular case, the Laplace distribution, was considered in [8] and [9]. The Student-t distribution was considered in [10], in a chronobiological application, in [11] and indirectly in [12] in a block-sparse application. In [13], [14] and [15] those heavy tailed priors are treated as particular cases of the Generalized Hyperbolic distribution, with the corresponding Generalized Inverse Gaussian distribution as the mixing distribution. Typically, from the corresponding joint posterior distribution the unknowns of the linear model, together with the hyperparameters, are estimated via Bayesian point estimators, like Maximum *A Posteriori* (MAP) or Posterior Mean (PM), [15], [12], where the PM is performed via the variational Bayesian approximation (VBA) or using MCMC techniques. The corresponding iterative algorithms are very much alike, with slightly different expressions for the updating equations corresponding to the model variances (MAP) or hyperparameters, i.e. the posterior distribution modelling the variances, (PM and MCMC). In all three cases, the prior parameters considered for the mixing distribution are present in the iterative algorithms, therefore the crucial importance of the parameter selection (p.s.). Moreover, typically those prior parameters are used for initialization of the iterative algorithm. One p.s. strategy, considering the mixing distribution, is to use non-informative priors. In this strategy, the mixing distribution is considered and the parameters are set such that the mixing distribution is approaching the Jeffreys prior.

Two other parameters selection strategies are considered, one accounting for the prior distribution in the p.s. setting and the second one accounting for both the prior and mixing distribution, exploiting the link of their first and second order moments. The reconstruction performances of the three p. s. strategies for three heavy tailed continuous Gaussian scale mixtures are compared considering the PM via VBA corresponding iterative algorithms, using as measurement the confusion matrix of the false positives and false negatives with different degrees of relaxation.

We present quantitative results about the *sparsity enforcing* of the prior models and the p.s. strategies considered.

2. GENERALIZED HYPERBOLIC PM ALGORITHM

For the linear model,

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{\epsilon}, \tag{1}$$

the Generalized Hyperbolic prior model is considered, Eq. (2), via a zero-mean Normal distribution for $f_j \mid v_{f_j}$ and the variance v_{f_j} modelled as a Generalized Inverse Gaussian *(distribution, expressed via the modified Bessel function of the second degree $\mathcal{K}_p\left(\cdot\right)$

$$\begin{cases} p(f_j \mid v_{f_j}) = \mathcal{N}(f_j \mid 0,\ v_{f_j}) \\ p(v_{f_j} \mid \gamma^2, \delta^2, p) = \mathcal{GIG}(v_{f_j} \mid \gamma^2, \delta^2, p), \end{cases} \tag{2}$$

with the corresponding marginal $p(f_j \mid \gamma^2, \delta^2, p)$ the Generalized Hyperbolic distribution [14]. For $p = -\frac{1}{2}$, the prior model is the Normal Inverse Gaussian distribution, with zero location and asymmetry parameters. For $\delta \searrow 0$ and $p > 0$, the prior model corresponds to the Variance Gamma distribution, with zero location and asymmetry parameters. The corresponding mixing distribution is the Gamma $\mathcal{G}\left(v_{f_j} \mid p, \frac{\gamma^2}{2}\right)$ distribution (expressed in terms of shape and rate parameters). In particular, if $p = 1$ the prior model corresponds to the Laplace distribution, with zero location parameter and the corresponding Exponential mixing distribution. For $\gamma \searrow 0$ and $p < 0$, the prior model corresponds to the (two parameters) Student-t distribution. The corresponding mixing distribution is the Inverse Gamma $\mathcal{IG}\left(v_{f_j} \mid -p, \frac{\delta^2}{2}\right)$ distribution. Non-stationary noise model, i.e. different variances v_{ϵ_i} for each ϵ_i and also different variances v_{f_j} for each f_j are assumed for the linear model (1). The likelihood, prior and the variance priors write

$$\begin{cases} p(\boldsymbol{f} \mid \boldsymbol{v}_f) \propto \det(\boldsymbol{V}_f)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left\|\boldsymbol{V}_f^{-\frac{1}{2}}\boldsymbol{D}\boldsymbol{f}\right\|^2\right\} \\ p(\boldsymbol{v}_f \mid \gamma_f^2, \delta_f^2, p_f) \propto \prod_{j=1}^M v_{f_j}^{p_f-1} \exp\left\{-\frac{1}{2}\left(\gamma_f^2 \sum_{j=1}^M v_{f_j} \right.\right. \\ \qquad\qquad \left.\left. + \ \delta_f^2 \sum_{j=1}^M v_{f_j}^{-1}\right)\right\} \\ \boldsymbol{V}_f = \operatorname{diag}[\boldsymbol{v}_f] \ ; \ \boldsymbol{v}_f = \left[v_{f_1} \ldots v_{f_j} \ldots v_{f_M}\right] \\ p(\boldsymbol{g} \mid \boldsymbol{f}, \boldsymbol{v}_\epsilon) \propto \det(\boldsymbol{V}_\epsilon)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left\|\boldsymbol{V}_\epsilon^{-\frac{1}{2}}(\boldsymbol{g}-\boldsymbol{H}\boldsymbol{f})\right\|^2\right\} \\ p(\boldsymbol{v}_\epsilon \mid \gamma_\epsilon^2, \delta_\epsilon^2, p_\epsilon) \propto \prod_{i=1}^N v_{\epsilon_i}^{p_\epsilon-1} \exp\left\{-\frac{1}{2}\left(\gamma_\epsilon^2 \sum_{i=1}^N v_{\epsilon_i} \right.\right. \\ \qquad\qquad \left.\left. + \ \delta_\epsilon^2 \sum_{i=1}^N v_{\epsilon_i}^{-1}\right)\right\} \\ \boldsymbol{V}_\epsilon = \operatorname{diag}[\boldsymbol{v}_\epsilon] \ ; \ \boldsymbol{v}_\epsilon = \left[v_{\epsilon_1} \ldots v_{\epsilon_i} \ldots v_{\epsilon_N}\right], \end{cases} \tag{3}$$

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

$$\text{Posterior} \approx q(\boldsymbol{f}, \boldsymbol{v}_f, \boldsymbol{v}_\epsilon) = q_1(\boldsymbol{f}) \prod_{j=1}^M q_{2j}(v_{f_j}) \prod_{i=1}^N q_{3i}(v_{\epsilon_i}), \tag{4}$$

by minimizing the Kullback-Leibleir diveregence. Generally, the minimization leads to proportionalities between each separable q_i distribution and an exponential with arguments the expected value of the log posterior distribution with respect to all other separable distributions, i.e. $q_{/i}$ (see [16]):

$$q_i \propto \exp\left\{\mathbb{E}_{q_{/i}}[\ln \text{Posterior}]\right\}. \tag{5}$$

For $q_1(\boldsymbol{f})$ a (multivariate) Normal distribution is obtained, $\mathcal{N}\left(\boldsymbol{f} \,\middle|\, \widetilde{\boldsymbol{f}}, \widetilde{\boldsymbol{\Sigma}}\right)$ with

$$\widetilde{\boldsymbol{\Sigma}} = \left(\boldsymbol{H}^T \widetilde{\boldsymbol{V}}_\epsilon \boldsymbol{H} + \boldsymbol{D}^T \widetilde{\boldsymbol{V}}_f \boldsymbol{D}\right)^{-1} ; \ \widetilde{\boldsymbol{f}} = \widetilde{\boldsymbol{\Sigma}} \boldsymbol{H}^T \widetilde{\boldsymbol{V}}_\epsilon \boldsymbol{g} \tag{6}$$

using the notations,

$$\widetilde{\boldsymbol{V}}_f = \operatorname{diag}[\widetilde{\boldsymbol{v}}_f] ; \widetilde{\boldsymbol{v}}_f = [\ldots \widetilde{v}_{f_j} \ldots] ; \widetilde{v}_{f_j} = \mathbb{E}_{q_{2j}(v_{f_j})}\left[v_{f_j}^{-1}\right] \tag{7}$$

and the similar notations corresponding to $\widetilde{\boldsymbol{V}}_\epsilon$. The proportionality corresponding to $q_{2j}(v_{f_j})$ writes:

$$q_{2j}(v_{f_j}) \propto v_{f_j}^{\left(p_f - \frac{3}{2}\right)} \exp\left\{\frac{-1}{2}\left(\gamma_f^2 v_{f_j} + \left(\delta_f^2 + r_j(\widetilde{\boldsymbol{f}})\right) v_{f_j}^{-1}\right)\right\}, \tag{8}$$

with

$$r_j(\widetilde{\boldsymbol{f}}) = \left(\boldsymbol{D}_j \widetilde{\boldsymbol{f}}\right)^2 + \operatorname{Tr}\left[\boldsymbol{D}_j^T \boldsymbol{D}_j \widehat{\Sigma}\right] \tag{9}$$

and the corresponding distribution for q_{2j} is a $\mathcal{GIG}$ distribution, Eq. (10):

$$q_{2j}(v_{f_j}) = \mathcal{GIG}\left(v_{f_j} \,\middle|\, \gamma_f^2, \widetilde{\delta}_{f_j}^2, p_f - \frac{1}{2}\right), \widetilde{\delta}_{f_j}^2 = \delta_f^2 + r_j(\widetilde{\boldsymbol{f}}), \tag{10}$$

Similary, for $q_{3i}(v_{\epsilon_i})$:

$$q_{3i}(v_{\epsilon_i}) = \mathcal{GIG}\left(v_{\epsilon_i} \,\middle|\, \gamma_\epsilon^2, \widetilde{\delta}_{\epsilon_i}^2, p_\epsilon - \frac{1}{2}\right), \widetilde{\delta}_{\epsilon_i}^2 = \delta_\epsilon^2 + s_i(\widetilde{\boldsymbol{f}}), \tag{11}$$

with

$$s_i(\widetilde{\boldsymbol{f}}) = \left(g_i - \boldsymbol{H}_i \widetilde{\boldsymbol{f}}\right)^2 + \operatorname{Tr}\left[\boldsymbol{H}_i^T \boldsymbol{H}_i \widehat{\Sigma}\right] \tag{12}$$

Via Eq. (7) and (10)

$$\widetilde{v}_{f_j} = \mathbb{E}_{q_{2j}(v_{f_j})}\left[v_{f_j}^{-1}\right] = \frac{\gamma_f}{\widetilde{\delta}_{f_j}} \frac{\mathcal{K}_{p_f - \frac{3}{2}}\left(\gamma_f \widetilde{\delta}_{f_j}\right)}{\mathcal{K}_{p_f - \frac{1}{2}}\left(\gamma_f \widetilde{\delta}_{f_j}\right)} \tag{13}$$

for $j = 1, \ldots, M$ and with similar expressions for $\widetilde{v}_{\epsilon_i}$, $i = 1, \ldots, N$. The corresponding algorithm, with a $\mathcal{GIG}$ mixing distribution is presented in Algorithm (1).

3. PARAMETER SELECTION

The parameters selection resumes to the choice of the priors parameters. In this particular context, they are the same as the parameters corresponding to the mixing distributions. This particular context, gives the possibility to consider three p.s. strategies.

3.1. Non informative priors

First one is considering the mixing distribution and selects the parameters such that non-informative priors, [17], [18], [19] are modelling the variances, i.e. sets γ_f, δ_f, p_f such that $\mathcal{GIG}(v_{f_j} \mid \gamma_f^2, \delta_f^2, p_f)$ is non-informative. Setting both parameters *close to zero* for the $\mathcal{NIG}$ and $\mathcal{VG}$, $\mathcal{S}t$ prior models the mixing distribution approaches Jeffreys, Table (1), second row.

3.2. Prior form

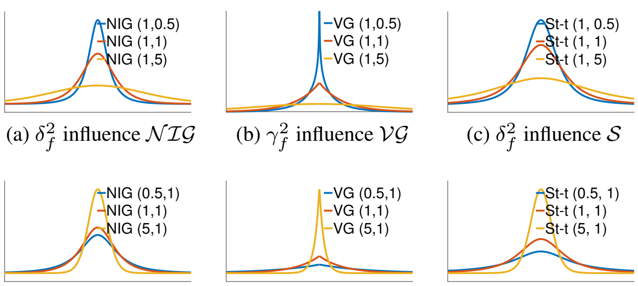
The second strategy is considering the prior distribution and selects the parameters such that the prior is concentrated around the mean, in order to enforce sparsity. The influence

Algorithm 1 PM via VBA - General Hyperbolic prior model, non-stationary errors model

Ensure: INITIALIZATION $\gamma_f,\ \delta_f,\ p_f,\ \gamma_\epsilon,\ \delta_\epsilon,\ p_\epsilon$

compute $\widetilde{v}_{f_j}^{(0)}$ via Eq. (13) using δ_f instead of $\widetilde{\delta}_{f_j}$ and then $\widetilde{V}_{f_j}^{(0)}$ via Eq. (7)
compute $\widetilde{v}_{\epsilon_i}^{(0)}$ using δ_ϵ instead of δ_{ϵ_i} and then $\widetilde{V}_{f_j}^{(0)}$
function VBAGH($\gamma_f, \delta_f, p_f, \gamma_\epsilon, \delta_\epsilon, p_\epsilon, M, N, NoIter$)
  for $n = 0$ to $NoIter$ **do**
    compute $\widetilde{\mathbf{\Sigma}}^{(n)}$ and $\widetilde{\boldsymbol{f}}^{(n)}$ via Eq. (6)
    for $j = 1$ to M **do**
      compute $r_j^{(n+1)}(\widetilde{\boldsymbol{f}})$ then $\widetilde{\delta}_{f_j}^{(n+1)}$ via Eq. (10)
      compute $\widetilde{v}_{f_j}^{(n+1)}$ via Eq. (13)
    end for
    compute $\widetilde{\boldsymbol{V}}_f{}^{(n+1)}$ via Eq.(7)
    for $i = 1$ to N **do**
      compute $s_i^{(n+1)}(\widetilde{\boldsymbol{f}})$ then $\widetilde{\delta}_{\epsilon_i}^{(n+1)}$ via Eq. (11)
      compute $\widetilde{v}_{\epsilon_i}^{(n+1)}$
    end for
    compute $\widetilde{\boldsymbol{V}}_\epsilon{}^{(n+1)}$ via Eq. (7)
  end for
  return $\widetilde{\boldsymbol{f}}^{(n)}, \widetilde{\mathbf{\Sigma}}^{(n)}, \widetilde{\boldsymbol{v}}_f{}^{(n+1)}, \widetilde{\boldsymbol{v}}_\epsilon^{(n+1)}$
end function

of each corresponding parameter for three prior models considered is presented in Fig. (1), and the parameter setting in Table (1), third row.



(a) δ_f^2 influence $\mathcal{NIG}$ (b) γ_f^2 influence $\mathcal{VG}$ (c) δ_f^2 influence $\mathcal{S}$

(d) γ_f^2 influence $\mathcal{NIG}$ (e) p_f influence $\mathcal{VG}$ (f) $-p_f$ influence $\mathcal{S}$

Fig. 1: Priors behaviour depending on the parameters.

3.3. Link between f_j and v_{f_j}

The third strategy is based on the link between f_j and its corresponding variance v_{f_j}, which in the sparsity context, during iterations, associates small variances v_{f_j} for the *close to zero* values f_j and significant variances for the *non close to zero* values of $\boldsymbol{f}$. Formally, this mechanism is modelled by Eq. (14),

$$\text{Var}_{\mathcal{P}}\left(f_j\right) = \mathbb{E}_{\mathcal{M}}\left(v_{f_j}\right), \tag{14}$$

where $\mathcal{P}$ and $\mathcal{M}$ denote the prior and the mixing distributions respectively. This p.s. strategy is expressing the parameters via the first and second order moments of the mixing distribution $\epsilon = \text{Var}_{\mathcal{P}}\left(f_j\right) = \mathbb{E}_{\mathcal{M}}\left(v_{f_j}\right)$ and $\omega = \text{Var}_{\mathcal{M}}\left(v_{f_j}\right)$, Eq. (15). Imposing a *close to zero* value for the prior distribution variance, $\epsilon \searrow 0$, assures a concentration around the zero mean of the prior distribution and implicitly, via Eq. (14), a small expected value for the mixing distribution. Therefore, setting a *close to zero* variance for the mixing distribution will impose a *sparse-like* structure for the variance $\boldsymbol{v}_f$.

$$\frac{\delta}{\gamma}\frac{\mathcal{K}_{p+1}(\gamma\delta)}{\mathcal{K}_p(\gamma\delta)} = \epsilon; \quad \frac{\delta^2}{\gamma^2}\frac{\mathcal{K}_{p+2}(\gamma\delta)}{\mathcal{K}_p(\gamma\delta)} = \omega + \epsilon^2. \tag{15}$$

For the $\mathcal{NIG}$ prior model, $-p = \frac{1}{2}$, using, $\mathcal{K}_{\frac{3}{2}}(x) = \mathcal{K}_{\frac{1}{2}}(x)\left(1 + \frac{1}{x}\right)$ leads to:

$$\gamma^2 = \omega^{-1},\ \delta^2 = \epsilon^2\omega^{-1}. \tag{16}$$

For the $\mathcal{VG}$ prior model, $\delta^2 \searrow 0$ and $p > 0$, using the asymptotic relation for small arguments for modified Bessel function:

$$p = \epsilon^2\omega^{-1},\ \frac{\gamma^2}{2} = \epsilon\omega^{-1}. \tag{17}$$

For the $\mathcal{S}t$ prior model, $\gamma^2 \searrow 0$ and $p < 0$, using the asymptotic relation for small arguments for modified Bessel function:

$$-p = 2 + \epsilon^2\omega^{-1},\ \frac{\delta^2}{2} = \epsilon\left(1 + \epsilon^2\omega^{-1}\right) \tag{18}$$

Table (1) presents the parameters corresponding to the three strategies and the five priors considered.

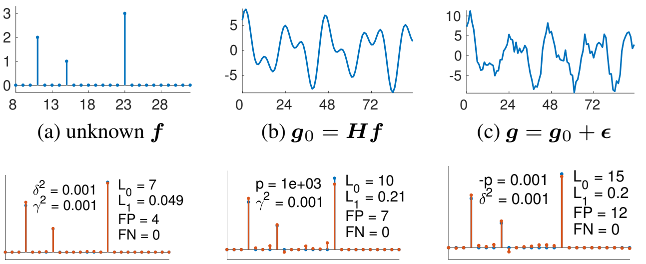
	$(\gamma^2, \delta^2, p = -\frac{1}{2})$	$(\gamma^2, \delta^2 \searrow 0, p > 0)$	$(\gamma^2 \searrow 0, \delta^2, p < 0)$
Mixing	—	$\mathcal{G}\left(v_{f_j} \mid p,\ \frac{\gamma^2}{2}\right)$	$\mathcal{IG}\left(v_{f_j} \mid -p,\ \frac{\delta^2}{2}\right)$
Jeffreys	$\gamma^2 \searrow 0,\ \delta^2 \searrow 0$	$p \searrow 0,\ \gamma^2 \searrow 0\quad (\mathcal{J})$	$-p \searrow 0,\ \delta^2 \searrow 0\quad (\mathcal{J})$
Prior	$\mathcal{NIG}\left(f_j \mid \gamma^2, \delta^2\right)$	$\mathcal{VG}\left(f_j \mid p, \gamma^2\right)$	$\mathcal{S}t\left(f_j \mid -p, \delta^2\right)$
Form	$\gamma^2 \gg 0,\ \delta^2 \searrow 0$	$p \gg 0,\ \gamma^2 \searrow 0$	$-p \gg 0,\ \delta^2 \searrow 0$
Link	$\gamma^2,\ \delta^2$ via Eq. (16)	$p,\ \gamma^2$ via Eq. (17)	$-p,\ \delta^2$ via Eq. (18)

Table 1: Mixing distribution $\mathcal{GIG}$ parameter's corresponding to different p.s. strategies.

4. SIMULATION RESULTS

We study the behaviour of the p.s. strategies considered, for the three sparsity enforcing priors considered. We consider a biological 1-D application, where the goal is to infer over the *sparse* corresponding periodic component vector $\boldsymbol{f}$, Fig (2a) of a short time series, Fig (2b) in the presence of noise, $\boldsymbol{g}$, Fig (2c). More details about the application and the limitations of classical approaches can be found in [10]. We consider this example for its small dimension, allowing a great number of simulations, corresponding to different values of the parameters. For each prior model with parameters corresponding to the three p.s. strategies estimation is obtain via

Algorithm (1). The second row of Fig. (2) shows a comparison between $\boldsymbol{f}$ and $\widehat{\boldsymbol{f}}$, corresponding to the three prior models, when the Jeffreys prior p.s. is used. For each estimation,



(a) unknown $\boldsymbol{f}$ (b) $\boldsymbol{g}_0 = \boldsymbol{H}\boldsymbol{f}$ (c) $\boldsymbol{g} = \boldsymbol{g}_0 + \boldsymbol{\epsilon}$

(d) $\widehat{\boldsymbol{f}}$ via $\mathcal{NIG}$ (e) $\widehat{\boldsymbol{f}}$ via $\mathcal{VG}$ (f) $\widehat{\boldsymbol{f}}$ via $\mathcal{S}t$

Fig. 2: Estimation via $\mathcal{NIG}$, $\mathcal{VG}$ and $\mathcal{S}t$ prior models corresponding to Jeffreys prior p.s. strategy for one realisation. ϵ corresponds to SNR= 10dB. For all three cases the corresponding parameters are both set to 0.001. τ = 0.005

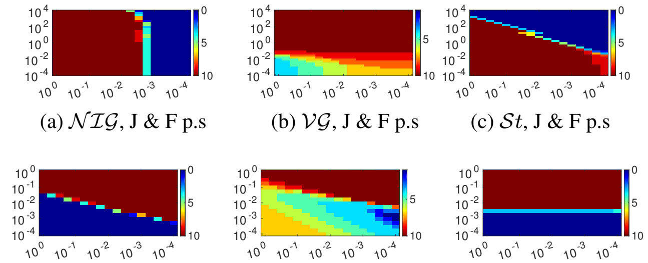
the L_0, L_1, False Positive (FP) and False Negative (FN) measures are considered. The measurements are considered w.r.t. the *strictly sparse* corresponding vector $\widehat{\boldsymbol{f}}^{\dagger}$, obtained via a threshold τ

$$\widehat{f}_j^{\dagger} = \begin{cases} 0, & \text{if } \widehat{f}_j < \tau \max_j \widehat{f}_j \\ \widehat{f}_j, & \text{else} \end{cases} \tag{19}$$

For the $\mathcal{NIG}$ prior model for the Jeffreys (J) p.s. strategy, both parameters have to be set *close to zero*. We consider the FP and FN corresponding values for $\gamma^2 = 10^{-k}, \delta^2 = 10^{-k}, k = \{0, 0.25, \dots 4\}$. For the Form (F) p.s. strategy, first parameter have to be set $\gg 0$ and the second one *close to zero*. We consider the FP and FN corresponding values for $\gamma^2 = 10^{k}, \delta^2 = 10^{-k}$. For the Link (L) p.s. strategy, the parameter are set via Eq. (16), with ϵ and ω set *close to zero*. We consider the FP and FN corresponding values for $\epsilon = 10^{-k}, \omega = 10^{-k}$. The same values are considered for the corresponding parameters of the $\mathcal{VG}$ and $\mathcal{S}t$ prior models, using their corresponding equations for the L p.s. strategy, i.e. (17) and (18) respectively. For measuring the *sparsity enforcement* corresponding to each prior model and p.s. strategy, we consider the confusion matrix $CM(r)$ depending on a relaxation degree r

$$CM(r) = \begin{cases} 0, FN + FP \leq r \\ 1, FN + FP > r \end{cases} \tag{20}$$

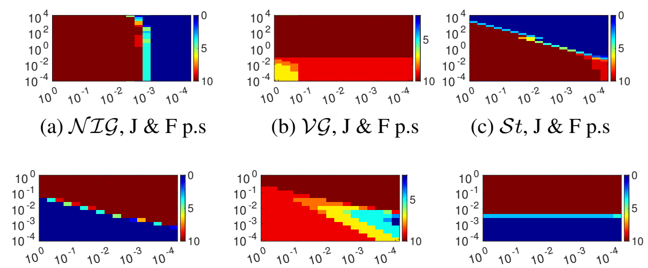
Figure (3), first row, presents the sum of confusion matrices corresponding to ten realisations of the measurements for $r = 3$, CM(3) for the three prior models using the Jeffreys and Form p.s. strategies. The x-axis depends on γ^2 and the y-axis depends δ^2, for the $\mathcal{NIG}$ prior model. The x-axis depends on p and the y-axis depends γ^2, for the $\mathcal{VG}$ prior model. The x-axis depends on $-p$ and the y-axis depends δ^2, for the $\mathcal{S}t$ prior model. The second row of the figure presents the corresponding results for the Link strategy. The x-axis depends on ϵ and the y-axis depends ω.



(a) $\mathcal{NIG}$, J & F p.s (b) $\mathcal{VG}$, J & F p.s (c) $\mathcal{S}t$, J & F p.s

(d) $\mathcal{NIG}$, Link p.s (e) $\mathcal{VG}$, Link p.s (f) $\mathcal{S}t$, Link p.s

Fig. 3: Sum of ten $CM(3)$ matrices: First row: Form & Jeffreys p.s. strategies for $\mathcal{NIG}$, $\mathcal{VG}$ and $\mathcal{S}t$. Second row: Link p.s. strategy for $\mathcal{NIG}$, $\mathcal{VG}$ and $\mathcal{S}t$. $\tau = 0005$.



(a) $\mathcal{NIG}$, J & F p.s (b) $\mathcal{VG}$, J & F p.s (c) $\mathcal{S}t$, J & F p.s

(d) $\mathcal{NIG}$, Link p.s (e) $\mathcal{VG}$, Link p.s (f) $\mathcal{S}t$, Link p.s

Fig. 4: Sum of ten $CM(2)$ matrices. $\tau = 0.005$.

5. CONLCUSIONS

Simulation results, for different values for the threshold τ show that the structure of CM is the same in terms of small CM values. This is also the case when the number of experiments is varied. The relaxation parameter r is compressing the area of small CM values as r is approaching 0. For $r = 0$, the CM measures the exact reconstruction in terms of sparsity. To goal of this paper was to compare how much is each prior model enforcing sparsity and not the exact reconstruction, which may strongly depend on the particularities of the applications. We notice that between the three priors considered the $\mathcal{NIG}$ and $\mathcal{S}t$ models seems to perform much better for sparsity enforcing. Also, the proposed p.s. strategy, accounting for both the mixing and prior distribution via (14) seems to work in all cases. Moreover, the Jeffreys prior p.s. strategy is working for the $\mathcal{NIG}$ but it seems not to be adapted for the $\mathcal{S}t$ prior model. However, considering the blue regions from the Form and Jeffreys CM, i.e. the regions with good sparsity enforcing, they correspond via (16), (17) and (18) to parameter setting obtained via the Link p.s. strategy.

6. REFERENCES

[1] H. Sadreazami, M. O. Ahmad, and M. N. S. Swamy, "Contourlet domain image denoising using Normal Inverse Gaussian distribution," in *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2014, pp. 1–4.

[2] T. A. Øigård and A. Hanssen, "The multivariate normal inverse Gaussian heavy-tailed distribution; simulation and estimation," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2002, pp. 1489–1492.

[3] T. A. Øigård, A. Hanssen, and R. E. Hansen, "The multivariate normal inverse Gaussian distribution: EM-estimation and analysis of synthetic aperture sonar data," in *2004 12th European Signal Processing Conference*, 2004, pp. 1433–1436.

[4] A. P. Doulgeris and T. Eltoft, "Scale Mixture of Gaussian Modelling of Polarimetric SAR Data," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1–12, 2010.

[5] S. Atapattu, C. Tellambura, and H. Jiang, "A Mixture Gamma Distribution to Model the SNR of Wireless Channels," *IEEE Transactions on Wireless Communications*, vol. 10, pp. 4193–4203, 2011.

[6] T. Eltoft, T. Kim, and T. Lee, "On the multivariate Laplace distribution," *IEEE Signal Processing Letters*, vol. 13, pp. 300–303, 2006.

[7] R. Tahmasbi and S. Rezaei, "A Soft Voice Activity Detection Using GARCH Filter and Variance Gamma Distribution," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 1129–1134, 2007.

[8] P. Garrigues and B. A. Olshausen, "Group Sparse Coding with a Laplacian Scale Mixture Prior," in *Advances in Neural Information Processing Systems 23*, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Eds., pp. 676–684. Curran Associates, Inc., 2010.

[9] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian Compressive Sensing Using Laplace Priors," *IEEE Transactions on Signal Processing*, vol. 19, pp. 53–63, 2010.

[10] M. Dumitru, A. Mohammad-Djafari, and S. S. Baghai, "Precise periodic components estimation for chronobiological signals through Bayesian Inference with sparsity enforcing prior," *EURASIP Journal on Bioinformatics and Systems Biology, Springer*, vol. 3, pp. 1–32, 2016.

[11] A. Mohammad-Djafari and M. Dumitru, "Bayesian sparse solutions to linear inverse problems with non-stationary noise with Student-t priors," *Digital Signal Processing*, vol. 47, pp. 128–156, 2015.

[12] M. Korki, J. Zhang, C. Zhang, and Zayyani H., "Iterative Bayesian Reconstruction of Non-IID Block-Sparse Signals," *IEEE Transactions on Signal Processing*, vol. 64, pp. 2906–2921, 2016.

[13] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A Variational Bayes Framework for Sparse Adaptive Estimation," *IEEE Transactions on Signal Processing*, vol. 62, pp. 4723–4736, 2014.

[14] D. Wraith and F. Forbes, "Location and scale mixtures of Gaussians with flexible tail behaviour: Properties, inference and application to multivariate clustering," *Computational Statistics & Data Analysis*, vol. 90, no. C, pp. 61–73, 2015.

[15] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian Group-Sparse Modeling and Variational Inference," *IEEE Transactions on Signal Processing*, vol. 62, pp. 2906–2921, 2014.

[16] V. Šmìdl and A. Quinn, *The Variational Bayes Method in Signal Processing*, Signal, Image and Speech Processing. Springer-Verlag Berlin Heidelberg, first edition, 2006.

[17] R. E. Kaas and L. Wasserman, "The Selection of Prior Distributions by Formal Rules," *Journal of the American Statistical Association*, vol. 91, pp. 1343–1370, 1996.

[18] H. Jeffreys, *Theory of Probability*, Oxford, third edition, 1961.

[19] J. M. Bernardo and Smith A. F. M., *Bayesian Theory*, Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA, 2000.