SUMMARIZING POSTERIOR DISTRIBUTIONS IN SIGNAL DECOMPOSITION PROBLEMS WHEN THE NUMBER OF COMPONENTS IS UNKNOWN

Alireza Roodaki, Julien Bect, and Gilles Fleury

E3S—SUPELEC Systems Sciences
Dept. of Signal Processing and Electronic Systems, SUPELEC, Gif-sur-Yvette, France.
Email: {alireza.roodaki, julien.bect, gilles.fleury}@supelec.fr

ABSTRACT

This paper addresses the problem of summarizing the posterior distributions that typically arise, in a Bayesian framework, when dealing with signal decomposition problems with unknown number of components. Such posterior distributions are defined over union of subspaces of differing dimensionality and can be sampled from using modern Monte Carlo techniques, for instance the increasingly popular RJ-MCMC method. No generic approach is available, however, to summarize the resulting variable-dimensional samples and extract from them component-specific parameters.

We propose a novel approach to this problem, which consists in approximating the complex posterior of interest by a "simple"—but still variable-dimensional—parametric distribution. The distance between the two distributions is measured using the Kullback-Leibler divergence, and a Stochastic EM-type algorithm, driven by the RJ-MCMC sampler, is proposed to estimate the parameters. The proposed algorithm is illustrated on the fundamental signal processing example of joint detection and estimation of sinusoids in white Gaussian noise.

Index Terms— Bayesian inference; Posterior summarization; Trans-dimensional MCMC; Label-switching; Stochastic EM.

1. INTRODUCTION

Nowadays, owing to the advent of Markov Chain Monte Carlo (MCMC) sampling methods [1], Bayesian data analysis is considered as a conventional approach in machine learning, signal and image processing, and data mining problems—to name but a few. Nevertheless, in many applications, practical challenges remain in the process of extracting, from the generated samples, quantities of interest to summarize the posterior distribution.

Summarization consists, loosely speaking, in providing a few simple yet interpretable parameters and/or graphics to the end-user of a statistical method. For instance, in the case of a scalar parameter with a unimodal posterior distribution, measures of location and dispersion (e.g., the empirical mean and the standard deviation, or the median and the interquartile range) are typically provided in addition to a graphical summary of the distribution (e.g., a histogram or a kernel density estimate). In the case of multimodal distributions summarization becomes more difficult but can be carried out using, for instance, the approximation of the posterior by a Gaussian Mixture Model (GMMs) [2].

This paper addresses the problem of summarizing posterior distributions in the case of trans-dimensional problems, i.e. "the problems in which the number of things that we don't know is one of the things that we don't know" [3]. The problem of signal decomposition when the number of components is unknown is an important example of such problems. Let $\mathbf{y} = (y_1, y_2, \ldots, y_N)^t$ be a vector of N observations, where the superscript t stands for vector transposition. In signal decomposition problems, the model space is a finite or countable set of models, $\mathcal{M} = \{\mathcal{M}_k,\, k \in \mathcal{K}\}$, where $\mathcal{K} \subset \mathbb{N}$ is an index set. It is assumed here that, under $\mathcal{M}_k$, there are k components with vectors of component-specific parameters $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}^k$, where $\boldsymbol{\Theta} \subset \mathbb{R}^d$. In a Bayesian framework, a joint posterior distribution is obtained through Bayes' formula for the model index k and the vector of component-specific parameters, after assigning prior distributions on them :

$$f\,(k,\, \boldsymbol{\theta}_k) \;\propto\; p\,(\mathbf{y} \mid k,\, \boldsymbol{\theta}_k)\, p\,(\boldsymbol{\theta}_k \mid k)\, p\,(k)\,,$$

where $\propto$ indicates proportionality. This joint posterior distribution, defined over a union of subspaces of differing dimensionality, completely describes the information (and the associated uncertainty) provided by the data $\mathbf{y}$ about the candidate models and the vector of unknown parameters.

1.1. Illustrative example: sinusoid detection

In this example, it is assumed that under $\mathcal{M}_k$, $\mathbf{y}$ can be written as a linear combination of k sinusoids observed in white Gaussian noise. The unknown component-specific parameters are $\boldsymbol{\theta}_k = \{\mathbf{a}_k, \boldsymbol{\omega}_k, \boldsymbol{\phi}_k\}$, where $\mathbf{a}_k$, $\boldsymbol{\omega}_k$ and $\boldsymbol{\phi}_k$ are the vectors of amplitudes, radial frequencies and phases, respectively. We use the hierarchical model, prior distributions, and Reversible Jump MCMC (RJ-MCMC) sampler [3] proposed in [4] for this problem; the interested reader is thus referred to [3, 4] for more details.

Figure 1 represents the posterior distributions of both the number of components k and the sorted[1] radial frequencies $\boldsymbol{\omega}_k$ given k obtained using the RJ-MCMC sampler. Each row is dedicated to one value of k, for $2 \leq k \leq 4$; observe that, other models have negligible posterior probabilities. In the experiment, the observed signal of length $N = 64$ consists of three sinusoids with amplitudes $\mathbf{a}_k = (20, 6.32, 20)^t$ and radial frequencies $\boldsymbol{\omega}_k = (0.63, 0.68, 0.73)^t$. The $\mathrm{SNR} \triangleq \frac{\|\mathbf{D}_k.\mathbf{a}_k\|^2}{N\sigma^2}$ is set to the moderate value of 7 dB, where $\mathbf{D}_k$ is the design matrix and σ^2 is the noise variance.

Roughly speaking, two approaches co-exist in the literature for such situations: Bayesian Model Selection (BMS) and Bayesian Model Averaging (BMA). The BMS approach ranks models according to their posterior probabilities $p(k|\mathbf{y})$, selects one model, and then summarizes the posterior under the (fixed-dimensional) selected model. This is at the price of loosing valuable information

[1] Owing to the invariance of both the likelihood and the prior under permutation of the components, component-specific marginal posteriors are all equal: this is the "label-switching" phenomenon [5, 6, 7]. Identifiability constraints (such as sorting) are the simplest way of dealing with this issue.

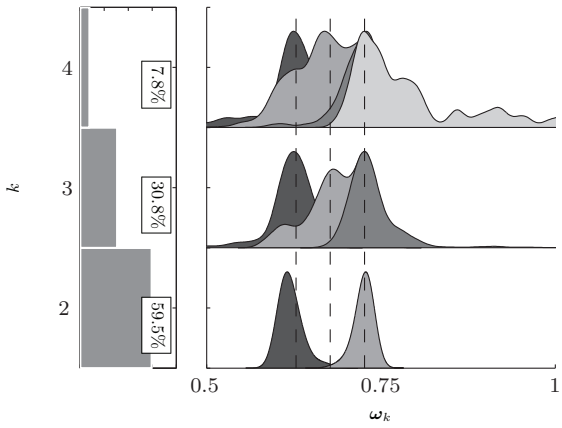

Fig. 1: Posteriors of k (left) and sorted radial frequencies, $\boldsymbol{\omega}_k$, given k (right). The true number of components is three. The vertical dashed lines in the right figure locate the true radial frequencies.

provided by the other (discarded) models. For instance, in the example of Figure 1, all information about the small—and therefore harder to detect—middle component is lost by selecting the most *a posteriori* probable model $\mathcal{M}_2$. The BMA approach consists in reporting results that are averaged over all possible models; it is, therefore, not appropriate for studying component-specific parameters, the number of which changes in each model[2].

More information concerning these two approaches can be found in [3] and references therein. To the best of our knowledge, no generic method is currently available, that would allow to summarize the information that is so easily read on Figure 1 for this very simple example: namely, that *there seem to be three sinusoidal components in the observed noisy signal, the middle one having a smaller probability of presence than the others*.

1.2. Outline of the paper

In this paper, we propose a novel approach to summarize the posterior distributions over variable-dimensional subspaces that typically arise in signal decomposition problems with an unknown number of components. It consists in approximating the complex posterior distribution with a parametric model (of varying-dimensionality), by minimization of the Kullback-Leibler (KL) divergence between the two distributions. A Stochastic EM (SEM)-type algorithm [8], driven by the output of an RJ-MCMC sampler, is used to estimate the parameters of the approximate model.

Our approach shares some similarities with the relabeling algorithms proposed in [6, 7] to solve the "label switching" problem, and also with the EM algorithm used in [9] in the context of adaptive MCMC algorithms (both in a *fixed*-dimensional setting). The main contribution of this paper is the introduction of an original variable-dimensional parametric model, which allows to tackle directly the difficult problem of approximating a distribution defined over a union of subspaces of differing dimensionality—and thus provides a first solution to the "trans-dimensional label-switching" problem, so to speak.

The paper is organized as follows. Section 2 introduces the proposed model and stochastic algorithm. Section 3 illustrates the approach using the sinusoid detection example already discussed in the introduction. Finally, Section 4 concludes the paper and gives directions for future work.

[2]See, however, the intensity plot provided in Section 3 (middle plot on Figure 4) as an example of a BMA summary related to a component-specific parameter.

2. PROPOSED ALGORITHM

Let F denote the target posterior distribution, defined on the variable-dimensional space $\mathbb{X} = \bigcup_{k=0}^{k_{\max}} \{k\} \times \boldsymbol{\Theta}^k$. We assume that F admits a probability density function (pdf) f, with respect to the kd-dimensional Lebesgue measure on each $\{k\} \times \boldsymbol{\Theta}^k$, $k \in \mathcal{K}$. To keep things simple, we also assume that $\boldsymbol{\Theta} = \mathbb{R}^d$.

Our objective is to approximate the exact posterior density f using a "simple" parametric model $q_{\boldsymbol{\eta}}$, where $\boldsymbol{\eta}$ is the vector of parameters defining the model. The pdf $q_{\boldsymbol{\eta}}$ will *also* be defined on the variable-dimensional space $\mathbb{X}$ (i.e., it is not a fixed-dimensional approximation as in the BMS approach). We assume that a Monte Carlo sampling method is available, e.g. a RJ-MCMC sampler [3], to generate M samples from f, which we denote by $\mathbf{x}^{(i)} = \left(k^{(i)}, \boldsymbol{\theta}^{(i)}_{k^{(i)}}\right)$, for $i = 1, \ldots, M$.

2.1. Variable-dimensional parametric model

Let us describe the proposed parametric model from a generative point of view. As in a traditional GMM, we assume that there is a certain number L of "Gaussian components" in the (approximate) posterior, each generating a d-variate Gaussian vector with mean $\boldsymbol{\mu}_l$ and covariance matrix $\boldsymbol{\Sigma}_l$, $1 \leq l \leq L$.

An $\mathbb{X}$-valued random variable $\mathbf{x} = (k, \boldsymbol{\theta}_k)$, with $0 \leq k \leq L$, is generated as follows. First, each of the L components can be either present or absent according to a binary indicator variable $\xi_l \in \{0, 1\}$. These Bernoulli variables are assumed to be independent, and we denote by $\pi_l \in (0; 1]$ the "probability of presence" of the l^{th} Gaussian component. Second, given the indicator variables, $k = \sum_{l=1}^{L} \xi_l$ Gaussian vectors are generated by the Gaussian components that are present ($\xi_l = 1$) and randomly arranged in a vector $\boldsymbol{\theta}_k$.

We denote by $q_{\boldsymbol{\eta}}$ the pdf of the random variable $\mathbf{x}$ that is thus generated, with $\boldsymbol{\eta}_l = (\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l, \pi_l)$ the vector of parameters of the l^{th} Gaussian component, $1 \leq l \leq L$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_L)$.

Remark. In contrast with GMMs, where only one component is present at a time (i.e., $k = 1$ in our notations), there is no constraint here on the sum of the probabilities of presence.

2.2. Estimating the model parameters

One way to fit the parametric distribution $q_{\boldsymbol{\eta}}(\mathbf{x})$ to the posterior $f(\mathbf{x})$ is to minimize the KL divergence of f from $q_{\boldsymbol{\eta}}$, denoted by $D_{KL}(f(\mathbf{x}) \| q_{\boldsymbol{\eta}}(\mathbf{x}))$. Thus, we define the criterion to be minimized as

$$\mathcal{J}(\boldsymbol{\eta}) \triangleq D_{KL}\left(f(\mathbf{x}) \| q_{\boldsymbol{\eta}}(\mathbf{x})\right) = \int_{\mathbb{X}} f(\mathbf{x}) \log \frac{f(\mathbf{x})}{q_{\boldsymbol{\eta}}(\mathbf{x})} \, \mathrm{d}\mathbf{x}.$$

Using samples generated by the RJ-MCMC sampler, this criterion can be approximated as

$$\mathcal{J}(\boldsymbol{\eta}) \simeq \hat{\mathcal{J}}(\boldsymbol{\eta}) = -\frac{1}{M} \sum_{i=1}^{M} \log\left(q_{\boldsymbol{\eta}}(\mathbf{x}^{(i)})\right) + C.$$

where C is a constant that does not depend on $\boldsymbol{\eta}$. One should note that minimizing $\hat{\mathcal{J}}(\boldsymbol{\eta})$ amounts to estimating $\boldsymbol{\eta}$ such that

$$\hat{\boldsymbol{\eta}} = \operatorname{argmax}_{\boldsymbol{\eta}} \sum_{i=1}^{M} \log\left(q_{\boldsymbol{\eta}}(\mathbf{x}^{(i)})\right). \quad (1)$$

At the r^{th} iteration,

S-step draw allocation vectors $\mathbf{z}^{(i,r)} \sim p\left(\cdot \,|\, \mathbf{x}^{(i)}, \hat{\boldsymbol{\eta}}^{(r-1)}\right)$, for $i = 1, \ldots, M$.

M-step estimate $\hat{\boldsymbol{\eta}}^{(r)}$ such that

$$\hat{\boldsymbol{\eta}}^{(r)} = \operatorname{argmax}_{\boldsymbol{\eta}} \sum_{i=1}^{M} \log p\left(\mathbf{x}^{(i)}, \mathbf{z}^{(i,r)} \,|\, \boldsymbol{\eta}\right).$$

Fig. 2: SEM algorithm.

Now, we assume that each element of the i^{th} observed sample $\mathbf{x}_j^{(i)}$, for $j = 1, \ldots, k^i$, has arisen from one of the L Gaussian components contained in $q_{\boldsymbol{\eta}}$. At this point, it is natural to introduce allocation vectors $\mathbf{z}^{(i)}$ corresponding to the i^{th} observed sample $\mathbf{x}^{(i)}$, for $i = 1, \ldots, M$, as latent variables. The element $\mathbf{z}_j^{(i)} = l$ indicates that $\mathbf{x}_j^{(i)}$ is allocated to the l^{th} Gaussian component.

Hence, given the allocation vector $\mathbf{z}^{(i)}$ and the parameters of the model $\boldsymbol{\eta}$, the conditional distribution of the observed samples, i.e., the model's likelihood, is

$$p(\mathbf{x}^{(i)} \,|\, \mathbf{z}^{(i)}, \boldsymbol{\eta}) = \prod_{j=1}^{k^{(i)}} \mathcal{N}(\mathbf{x}_j^{(i)} \,|\, \boldsymbol{\mu}_{\mathbf{z}_j^{(i)}}, \boldsymbol{\Sigma}_{\mathbf{z}_j^{(i)}}).$$

It turns out that the EM-type algorithms, which have been used in similar works [6, 7, 9], are not appropriate for solving this problem, as computing the expectation in the E-step is intricate. More explicitly, in our problem the computational burden of the summation in the E-step over the set of all possible allocation vectors $\mathbf{z}$ increases very rapidly with k. In fact, even for moderate values of k, say, $k = 10$, the summation is far too expensive to compute as it involves $k! \approx 3.6\,10^6$ terms. In this paper, we propose to use SEM [8] , a variation of the EM algorithm in which the E-step is substituted with stochastic simulation of the latent variables from their conditional posterior distributions given the previous estimates of the unknown parameters. In other words, for $i = 1, \ldots, M$, the allocation vectors $\mathbf{z}^{(i)}$ are drawn from $p(\cdot \,|\, \mathbf{x}^{(i)}, \hat{\boldsymbol{\eta}}^{(r)})$. This step is called the Stochastic (S)-step. Then, these random samples are used to construct the so-called pseudo-completed likelihood which reads

$$p\left(\mathbf{x}^{(i)}, \mathbf{z}^{(i)} \,|\, \boldsymbol{\eta}\right) = \prod_{j=1}^{k^{(i)}} \mathcal{N}\left(\mathbf{x}_j^{(i)} \,|\, \boldsymbol{\mu}_{\mathbf{z}_j^{(i)}}, \boldsymbol{\Sigma}_{\mathbf{z}_j^{(i)}}\right) \times \frac{\mathbb{1}_{\mathcal{Z}}(\mathbf{z}^{(i)})}{k^{(i)}!} \prod_{l=1}^{L} \pi_l^{\boldsymbol{\xi}_l^{(i)}} (1 - \pi_l)^{(1-\boldsymbol{\xi}_l^{(i)})}, \quad (2)$$

where $\mathcal{Z}$ is the set of all allocation vectors and $\boldsymbol{\xi}_l^{(i)} = 1$ if and only if there is a $j \in \{1, \ldots, k^{(i)}\}$ such that $\mathbf{z}_j^{(i)} = l$. The term $k^{(i)}!$ in (2) comes from the random permutation of components' labels. The proposed SEM-type algorithm for our problem is described in Figure 2.

Direct sampling from $p(\cdot \,|\, \mathbf{x}^{(i)}, \hat{\boldsymbol{\eta}}^{(r)})$, as required by the S-step, is unfortunately not feasible. Instead, since

$$p(\mathbf{z}^{(i)} \,|\, \mathbf{x}^{(i)}, \hat{\boldsymbol{\eta}}^{(r)}) \propto p(\mathbf{x}^{(i)}, \mathbf{z}^{(i)} \,|\, \hat{\boldsymbol{\eta}}^{(r)})$$

can be computed up to a normalizing constant, we devised an Independent Metropolis-Hasting (I-MH) algorithm to construct a Markov chain with $p(\mathbf{z}^{(i)} \,|\, \mathbf{x}^{(i)}, \hat{\boldsymbol{\eta}}^{(r)})$ as its stationary distribution.

2.3. Robustified algorithm

Preliminary experiments with the model and method described in the previous sections proved to be disappointing. To understand why, it must be remembered that the pdf $q_{\boldsymbol{\eta}}$ we are looking for is only an *approximation* (hopefully a good one) of the true posterior f. For instance, for high values of k, the posterior typically involves a diffuse part which can not properly represented by the parametric model (this can be seen quite clearly for $k = 4$ on Figure 1). Therefore, for any $\boldsymbol{\eta}$, some samples generated by the RJ-MCMC sampler are *outliers* with respect to $q_{\boldsymbol{\eta}}$ (i.e., the true posterior can be considered as a *contaminated* version of $q_{\boldsymbol{\eta}}$) which causes problems when using a maximum likelihood-type estimate such as (1).

These robustness issues were solved, in this paper, using two modifications of the algorithm (only in the one-dimensional case up to now). First, robust estimates [10] of the means and variances of a Gaussian distribution, based on the median and the interquartile range, are used instead of the empirical means and variances in the M-step. Second, a Poisson process component (with uniform intensity) is added to the model, in order to account for the diffuse part of the posterior and allow for a number L of Gaussian components which is smaller than the maximum observed $k^{(i)}$.

Remark. Similar robustness concerns are widespread in the clustering literature; see, e.g., [11] and the references therein.

3. RESULTS

In this section, we will investigate the capability of the proposed algorithm for summarizing variable-dimensional posterior distributions. We emphasize again that the output of the trans-dimensional Monte Carlo sampler, e.g. RJ-MCMC in this paper, is considered as the observed data for our algorithm. Regarding the fact that in this paper we provide results for the sinusoids' radial frequencies, the proposed parametric model consists of univariate Gaussian components. In other words, the space of component-specific parameters $\boldsymbol{\Theta} = (0; \pi) \subset \mathbb{R}$. But we believe that our algorithm is not limited to the problems with one-dimensional component-specific parameters. Therefore, in this section, it is assumed that each Gaussian component has a mean μ, a variance s^2, and a probability of presence π to be estimated.

Before launching the algorithm, first, we need to initialize the parametric model. It is natural to deduce the number L of Gaussian components from the posterior distribution of k. Here, we set it to the 90^{th} percentile to keep all the probable models in the play. To initialize the Gaussian components' parameters, i.e. μ and s^2, we used the robust estimates of the posterior of the sorted radial frequencies given $k = L$.

We ran the "robustified" stochastic algorithm introduced in Section 2 on the specific example shown in Figure 1, for 50 iterations, with $L = 3$ Gaussian components (the posterior probability of $\{k \leq 3\}$ is approximately 90.3%). Figure 3 illustrates the evolution of model parameters $\boldsymbol{\eta}$ together with the criterion $\mathcal{J}$. Two substantial facts can be deduced from this figure; first, the decreasing behavior of the criterion $\mathcal{J}$, which is almost constant after the 10^{th} iteration. Second, the convergence of the parameters of parametric model, esp. means μ and probabilities of presence π, though using a naive initialization procedure. Indeed after the 40^{th} iteration there is no significant move in the parameter estimates. Table 1 presents the summaries provided by the proposed method along with the ones obtained using the BMS approach. Contrary to BMS, the method that we proposed has enabled us to benefit from the information of all probable models to give summaries about the middle harder to

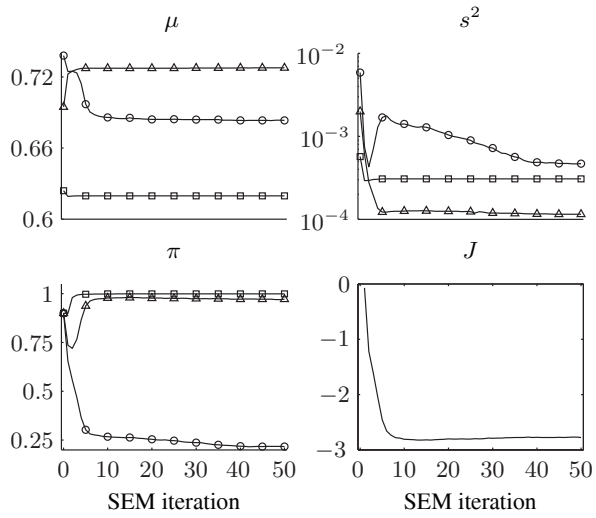

Fig. 3: Performance of the proposed summarizing algorithm on the sinusoid detection example. There are three Gaussian components in the model.

Comp	μ	s	π	μ_{BMS}	s_{BMS}
1	0.62	0.017	1	0.62	0.016
2	0.68	0.021	0.22	—	—
3	0.73	0.011	0.97	0.73	0.012

Table 1: Summaries of the variable-dimensional posterior distribution shown in Figure 1; The proposed approach vs. BMS.

detect component. Turning to the results of our approach, it can be seen that the estimated means are compatible with the true radial frequencies. Furthermore, the estimated probabilities of presence are consistent with uncertainty of them in the variable-dimensional posterior shown in Figure 1. Note the small estimated standard deviations which indicate our robustifying strategies have been useful.

The pdf's of the estimated Gaussian components are shown in Figure 4 (top). Comparing with the posterior of sorted radial frequencies shown in Figure 1, it can be inferred that the proposed algorithm has managed to remove the label-switching phenomenon in a variable-dimensional problem. Furthermore, the intensity plot of the allocated samples to the point process component is depicted in Figure 4 (bottom). This presents the outliers in the observed samples which cannot be described by the Gaussian components. Note that without the point process component these outliers would be allocated to the Gaussian components which can, consequently, yield in a significant deterioration of parameter estimates.

4. CONCLUSION

In this paper, we have proposed a novel algorithm to summarize posterior distributions defined over union of subspaces of differing dimensionality. For this purpose, a variable-dimensional parametric model has been designed to approximate the posterior of interest. The parameters of the approximate model have been estimated by means of a SEM-type algorithm, using samples from the posterior f generated by an RJ-MCMC algorithm. Modifications of our initial SEM-type algorithm have been proposed, in order to cope with the lack of robustness of maximum likelihood-type estimates. The relevance of the proposed algorithm, both for summarizing and for relabeling variable-dimensional posterior distributions, has been illustrated on the problem of detecting and estimating sinusoids in Gaussian white noise.

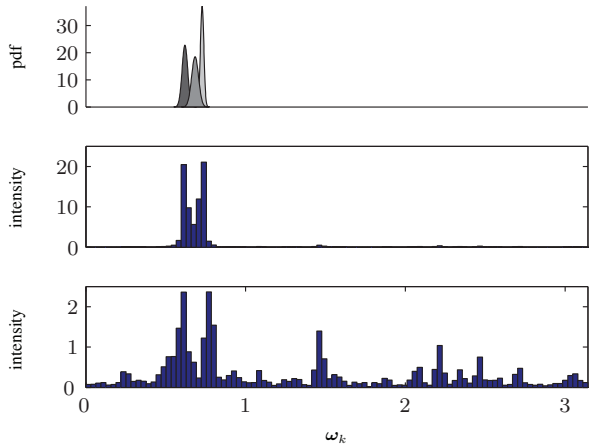

Fig. 4: The pdf of fitted Gaussian components (top), the histogram intensity of all radial frequencies samples (middle), and the histogram intensity of the allocated samples to the Poisson point process component (bottom).

We believe that this algorithm can be used in the vast domain of signal decomposition and mixture model analysis to enhance inference in trans-dimensional problems. For this purpose, generalizing the proposed algorithm to the multivariate case and analyzing its convergence properties is considered as future work. Another important point would be to use a more reliable initialization procedure.

References

[1] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods (second edition)*, Springer Verlag, 2004.

[2] M. West, "Approximating posterior distributions by mixture," *J. Roy. Stat. Soc. B Met.*, vol. 55, no. 2, pp. 409–422, 1993.

[3] P. J. Green, "Reversible jump MCMC computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[4] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2667–2676, 1999.

[5] S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components," *J. Roy. Stat. Soc. B Met.*, vol. 59, no. 4, pp. 731–792, 1997.

[6] M. Stephens, "Dealing with label switching in mixture models," *J. Roy. Stat. Soc. B Met.*, pp. 795–809, 2000.

[7] M. Sperrin, T. Jaki, and E. Wit, "Probabilistic relabelling strategies for the label switching problem in bayesian mixture models," *Stat. and Comput.*, vol. 20, pp. 357–366, 2010.

[8] G. Celeux and J. Diebolt, "The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem," *Comp. Statis. Quaterly*, vol. 2, pp. 73–82, 1985.

[9] Y. Bai, R. V. Craiu, and A. F. Di Narzo, "Divide and conquer: a mixture-based approach to regional adaptation for MCMC," *J. Comput. Graph. Stat.*, , no. 0, pp. 1–17, 2011.

[10] P. J. Huber and E. M. Ronchetti, *Robust statistics (2nd Edition)*, Wiley., 2009.

[11] R.N. Davé and R. Krishnapuram, "Robust clustering methods: a unified view," *IEEE Trans. Fuzzy Sys.*, vol. 5, no. 2, pp. 270–293, 1997.