A HIERARCHICAL BAYESIAN MODEL FOR FRAME REPRESENTATION

Lotfi Chaâri[1,2], Jean-Christophe Pesquet[1], Jean-Yves Tourneret[3], Philippe Ciuciu[2] and Amel Benazza-Benyahia[4]

[1] Université Paris-Est, IGM and UMR-CNRS 8049, 77454 Marne-la-Vallée cedex, France {chaari,pesquet}@univ-mlv.fr

[2] CEA/DSV/I[2]BM/Neurospin, CEA Saclay, Bbt. 145, Point Courrier 156, 91191 Gif-sur-Yvette cedex, France philippe.ciuciu@cea.fr

[3] University of Toulouse, IRIT/ENSEEIHT/TSA 31071 Toulouse, France jean-yves.tourneret@enseeiht.fr

[2] URISA, SUP'COM Cité Technologique des Communications, 2083, Tunisia benazza.amel@supcom.rnu.tn

ABSTRACT

In many signal processing problems, it may be fruitful to represent the signal under study in a redundant linear decomposition called a frame. If a probabilistic approach is adopted, it becomes then necessary to estimate the hyper-parameters characterizing the probability distribution of the frame coefficients. This problem is difficult since in general, the frame synthesis operator is not bijective and consequently, the frame coefficients are not directly observable. In this work, a hierarchical Bayesian model is introduced to solve this problem. A hybrid MCMC algorithm is subsequently proposed to sample from the derived posterior distribution. We show that through classical Bayesian estimators, this algorithm allows us to determine these hyper-parameters, as well as the frame coefficients in applications to image denoising with uniform noise.

***Index Terms*—** frame representations, Bayesian estimation, MCMC, hyper-parameter estimation, sparsity, wavelets.

1. INTRODUCTION

In signal/image processing, data representation is a crucial operation in many problems like reconstruction, restoration or compression. In this respect, many decompositions have been proposed in order to obtain suitable signal representations in other domains than the original spatial or temporal one. The traditional Fourier and discrete cosine transforms provide a good frequency localization of information, but at the expense of a poor spatial or temporal localization. To improve localization both in the spatial/temporal and frequency domains, the wavelet transform (WT) was introduced as a powerful tool in the 1980's. Many wavelet-like basis decompositions have been subsequently proposed offering differents features. For instance, we can mention wavelet packets or grouplet bases [1]. To further improve signal representations, redundant linear decomposition families called *frames* have become the focus of many works in the last decade. For the sake of clarity, it must be pointed out that the term frame [2] is understood in the sense of Hilbert space theory and not in the sense of some recent works like [3].

The main advantage of frames lies in their flexibility to capture local features of the signal. Hence, they result in sparser representations as shown in the image processing literature on curvelets [2] or dual-trees [4]. A major difficulty when using frame representations in a statistical framework is however to estimate the parameters of the frame coefficient probability distribution. Actually, since frame synthesis operators are generally not injective, even if the signal is perfectly known, the determination of its frame coefficients is an underdetermined problem. In this paper we propose a hierarchical Bayesian approach to estimate these parameters. This approach allows us to deal with any desirable distribution for the frame coefficients, in particular non necessarily log-concave priors promoting sparsity. However, we will focus here on generalized Gaussian (GG) priors in order to provide accurate models of sparse signals [5]. In addition, our method can be applied to noisy data when imprecise measurements of the signal are only available. Our work takes advantage of the developments in Markov Chain Monte Carlo (MCMC) algorithms [6]. These algorithms have been investigated for signal/image processing problems with sparsity constraints [7, 8]. In particular, we will consider Hybrid MCMC algorithms [9, 10] combining Metropolis-Hastings (MH) and Gibbs moves to sample according to some posterior distribution of interest. MCMC algorithms and WT have been jointly investigated in some works dealing with signal denoising in a Bayesian framework [11, 12]. However, in contrast with the present work where overcomplete frame representations are considered, these works are limited to wavelet bases for which the hyper-parameter estimation problem is much easier to handle.

This paper is organized as follows. We first begin by giving a brief overview on frames in Section 2. In Section 3, we formulate the statistical problem, before describing the proposed hierarchical Bayesian method. To illustrate the effectiveness of our algorithm, experiments on both synthetic and natural data are presented in Section 4. Finally some conclusions are drawn in Section 5.

2. BACKGROUND

In the following, we will consider real-valued digital signals/images of size L as elements of the Hilbert space $\mathbb{R}^L$ endowed with the classical Euclidean norm $\|\cdot\|$. Let K be

This work was supported by grants from Région Ile de France.

an integer greater than or equal to L. A family of vectors $(\boldsymbol{e}_k)_{1\le k\le K}$ in the finite-dimentional space $\mathbb{R}^L$ is a frame when there exists a constant μ in $]0,+\infty[$ such that

$$\forall \boldsymbol{x} \in \mathbb{R}^L, \qquad \mu\|\boldsymbol{x}\|^2 \le \sum_{k=1}^{K} |\langle \boldsymbol{x} \mid \boldsymbol{e}_k\rangle|^2. \tag{1}$$

If the inequality becomes an equality, $(\boldsymbol{e}_k)_{1\le k\le K}$ is called a *tight* frame. The bounded linear frame analysis operator is defined as

$$F\colon \mathbb{R}^L \to \mathbb{R}^K : \boldsymbol{x} \mapsto (\langle \boldsymbol{x} \mid \boldsymbol{e}_k\rangle)_{1\le k\le K}. \tag{2}$$

The adjoint synthesis frame operator is therefore given by

$$F^*\colon \mathbb{R}^K \to \mathbb{R}^L : (\xi_k)_{1\le k\le K} \mapsto \sum_{k=1}^{K} \xi_k \boldsymbol{e}_k. \tag{3}$$

Note that F is injective whereas F^* is surjective. When $F^{-1} = F^*$, $(\boldsymbol{e}_k)_{k\in\mathbb{K}}$ is an orthonormal basis. A simple example of a redundant frame is the union of $M > 1$ orthonormal bases. In this case, the frame is tight with $\mu = M$ and thus, we have $F^*F = M\mathrm{I}$ where I is the identity operator.

3. PROPOSED APPROACH

3.1. Problem statement

An observed signal $\boldsymbol{y} \in \mathbb{R}^L$ can be written according to its frame representation (FR) involving coefficients $\boldsymbol{x} \in \mathbb{R}^K$ as

$$\boldsymbol{y} = F^*\boldsymbol{x} + \boldsymbol{n}, \tag{4}$$

where $\boldsymbol{n}$ is the error between the observed signal $\boldsymbol{y}$ and its FR $F^*\boldsymbol{x}$. This error is modeled by imposing that $\boldsymbol{x}$ belongs to the closed convex set $C_\delta = \{\boldsymbol{x} \in \mathbb{R}^K \mid N(\boldsymbol{y} - F^*\boldsymbol{x}) \le \delta\}$ where $\delta \in [0,\infty[$ is some bound and $N(\cdot)$ can be any norm on $\mathbb{R}^L$. In a probabilistic setting, $\boldsymbol{y}$ and $\boldsymbol{x}$ are assumed to be realizations of random vectors $\boldsymbol{Y}$ and $\boldsymbol{X}$. In this context, our goal is to characterize the probability distribution of $\boldsymbol{X}$ given $\boldsymbol{Y}$, by considering some parametric probabilistic model and by estimating the associated hyper-parameters.
When F is bijective and $\delta = 0$, this estimation can be performed by inverting the transform so as to deduce $\boldsymbol{x}$ from $\boldsymbol{y}$ and by resorting to standard estimation techniques on $\boldsymbol{x}$. However, as mentioned in Section 2, for redundant frames, F^* is not bijective, which makes the hyper-parameter estimation problem difficult. In what follows, we develop an MCMC algorithm to solve this problem.

3.2. Hierarchical Bayesian model

In a Bayesian framework, we first need to define prior distributions for the frame coefficients. For instance, this prior may be chosen so as to promote the sparsity of the representation. Let $f(\boldsymbol{x}|\boldsymbol{\theta})$ be the probability density function (pdf) of the frame coefficients that depends on an unknown hyper-parameter vector $\boldsymbol{\theta}$ and $f(\boldsymbol{\theta})$ is the prior pdf for this hyper-parameter vector. In compliance with the observation model (4), the conditional pdf $f(\boldsymbol{y}|\boldsymbol{x})$ is assumed to be a *uniform* distribution on the closed convex set $D_\delta = \{\boldsymbol{y} \in \mathbb{R}^L \mid N(\boldsymbol{y} - F^*\boldsymbol{x}) \le \delta\}$, where $\delta > 0$. Denoting by $\boldsymbol{\Theta}$ the random variable associated with the hyper-parameter vector $\boldsymbol{\theta}$ and using the hierarchical structure between $\boldsymbol{Y}, \boldsymbol{X}$ and $\boldsymbol{\Theta}$, the conditional distribution of $(\boldsymbol{X}, \boldsymbol{\Theta})$ given $\boldsymbol{Y}$ can be written as

$$f(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x})f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta}), \tag{5}$$

where $\propto$ means "proportional to". In this paper, we assume that frame coefficients are a priori independent with marginal GG distributions. The scale and shape parameters associated with the k-th component X_k of the frame coefficient vector $\boldsymbol{X}$ are denoted by $\gamma_k > 0$ and $\beta_k > 0$, respectively. However, some frame coefficients are assumed to have the same distributions. By splitting the frame coefficients into G different groups (e.g., associated with the different subbands), characterized by the hyper-parameter vector $\boldsymbol{\theta}_g = (\beta_g, \gamma_g)$, the prior on the frame coefficient can be expressed as

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{g=1}^{G} \left[\left(\frac{\beta_g}{2\gamma_g^{1/\beta_g}\Gamma(1/\beta_g)} \right)^{n_g} \exp\left(\frac{-1}{\gamma_g} \sum_{k\in S_g} |x_k|^{\beta_g} \right) \right] \tag{6}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\dots,\boldsymbol{\theta}_G)$, $\Gamma(.)$ is the Gamma function and S_g denotes the set of elements of the g-th group containing n_g elements.

To complete the hierarchical Bayesian model, the following improper hyperprior was retained

$$f(\boldsymbol{\theta}) = \prod_{g=1}^{G} [f(\gamma_g)f(\beta_g)] \propto \prod_{g=1}^{G} \left[\frac{1}{\gamma_g} 1_{\mathbb{R}^+}(\gamma_g) 1_{[0,3]}(\beta_g) \right], \tag{7}$$

where for a set A, $1_A(\xi) = 1$ if $\xi \in A$ and 0 otherwise. This hyperprior is of practical interest since Jeffrey's prior on γ_g is non-informative, and the interval $[0,3]$ covers all possible values of β_g usually encountered in practical applications.
Bayesian estimators such as the maximum a posteriori (MAP) or the minimum mean square error (MMSE) estimators associated with the posterior distribution (5) have no simple closed-form expressions. The alternative developed in this paper is to use MCMC methods to draw samples of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ from the full posterior pdf in (5), and estimate the unknown model parameters $\boldsymbol{x}$ and hyper-parameters $\boldsymbol{\theta}$ accordingly.

3.3. Sampling strategy

Sampling according to the posterior distribution (5) will be performed by using a hybrid Gibbs sampler which iteratively generates samples according to $f(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ and $f(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})$.

3.3.1. Frame coefficient sampling

Straightforward calculations yield the conditional distribution

$$f(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \propto 1_{C_\delta}(\boldsymbol{x}) \prod_{g=1}^{G} \exp\left(\frac{-1}{\gamma_g} \sum_{k\in S_g} |x_k|^{\beta_g} \right) \tag{8}$$

which is nothing but the product of GG distributions truncated on C_δ. Sampling directly according to this truncated distribution is not easy to perform since the adjoint frame operator F^* is usually of large dimension. For this reason, we propose here to use an MH move, whose performance will strongly depend on the proposal distribution that we will design carefully to improve the acceptance ratio.

We denote as $\boldsymbol{x}^{(i)}$ the i-th accepted sample of the algorithm and $q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})$ the proposal that is used to generate a candidate at iteration i. The main difficulty for choosing $q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})$ stems from the fact that it must guarantee that $\boldsymbol{x} \in C_\delta$ as mentioned in Section 3.1, while yielding a tractable expression of $q(\boldsymbol{x}^{(i-1)} \mid \boldsymbol{x})/q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})$.

For this reason, we propose to exploit the algebraic properties of frame representations. More precisely, we can write $\boldsymbol{x} = \boldsymbol{x}_H + \boldsymbol{x}_{H^\perp}$, where $\boldsymbol{x}_H$ and $\boldsymbol{x}_{H^\perp}$ are realizations of random vectors taking their values in $H = \mathrm{Ran}(F)$ and $H^\perp = [\mathrm{Ran}(F)]^\perp = \mathrm{Null}(F^*)$, respectively.[1] The proposal distribution used in this paper allows us to generate samples $\boldsymbol{x}_H \in H$ and $\boldsymbol{x}_{H^\perp} \in H^\perp$. More precisely, a separable form of the proposal pdf will be considered as $q(\boldsymbol{x} \mid \boldsymbol{x}^{(i)}) = q\left(\boldsymbol{x}_H \mid \boldsymbol{x}_H^{(i-1)}\right) q\left(\boldsymbol{x}_{H^\perp} \mid \boldsymbol{x}_{H^\perp}^{(i-1)}\right)$, where $\boldsymbol{x}_H^{(i-1)} \in H$, $\boldsymbol{x}_{H^\perp}^{(i-1)} \in H^\perp$ and $\boldsymbol{x}^{(i-1)} = \boldsymbol{x}_H^{(i-1)} + \boldsymbol{x}_{H^\perp}^{(i-1)}$.

If we consider the decomposition $\boldsymbol{x} = \boldsymbol{x}_H + \boldsymbol{x}_{H^\perp}$, sampling $\boldsymbol{x}$ in C_δ is equivalent to sampling $\boldsymbol{\lambda} \in \overline{C}_\delta$, where $\overline{C}_\delta = \{\boldsymbol{\lambda} \in \mathbb{R}^L | N(\boldsymbol{y} - F^*F\boldsymbol{\lambda}) \leq \delta\}$. Indeed, we can write $\boldsymbol{x}_H = F\boldsymbol{\lambda}$ where $\boldsymbol{\lambda} \in \mathbb{R}^L$ and, since $\boldsymbol{x}_{H^\perp} \in \mathrm{Null}(F^*)$, $F^*\boldsymbol{x} = F^*F\boldsymbol{\lambda}$. Sampling $\boldsymbol{\lambda}$ in $\overline{C}_\delta$ can be easily performed, e.g., by generating $\boldsymbol{u}$ from the uniform distribution on the ball $B_{\boldsymbol{y},\delta}$ and by taking $\boldsymbol{\lambda} = (F^*F)^{-1}\boldsymbol{u}$.

To make the sampling of $\boldsymbol{x}_H$ at iteration i more efficient, it is interesting to take the sampled value at the previous iteration $\boldsymbol{x}_H^{(i-1)}$ into account. Similarly to random walk generation techniques, we proceed by randomly generating $\boldsymbol{u}$ in $B_{\hat{\boldsymbol{u}}^{(i-1)},\eta}$ where $\eta \in]0,\delta[$ and $\hat{\boldsymbol{u}}^{(i-1)} = P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) + \boldsymbol{y}$, P being the projection [2] onto the ball $B_{\boldsymbol{0},\delta-\eta}$. This allows us to draw a vector $\boldsymbol{x}_H = F(F^*F)^{-1}\boldsymbol{u}$ in C_δ such that $N(\boldsymbol{u} - \boldsymbol{u}^{(i-1)}) \leq 2\eta$.

Once we have simulated $\boldsymbol{x}_H = F\boldsymbol{\lambda}$, $\boldsymbol{x}_{H^\perp}$ has to be sampled as an element of $H^\perp$. We propose to sample $\boldsymbol{x}_H$ by drawing a variable $\boldsymbol{z}$ distributed according to the Gaussian distribution $\mathcal{N}(\boldsymbol{x}^{(i-1)}, \sigma_{\boldsymbol{x}}^2 \boldsymbol{I})$ and by projecting $\boldsymbol{z}$ onto $H^\perp$ as $\boldsymbol{x}_{H^\perp} = \Pi_{H^\perp}\boldsymbol{z}$, where $\Pi_{H^\perp} = \mathrm{I} - F(F^*F)^{-1}F^*$ is the orthogonal projection operator onto $H^\perp$.

3.3.2. *Hyper-parameter sampling*

Instead of sampling $\boldsymbol{\theta}$ according to $f(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})$, we propose to iteratively sample according to $f(\gamma_g|\beta_g,\boldsymbol{x},\boldsymbol{y})$ and $f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y})$. Due to the new parametrization introduced in (6), it is easy to show that $f(\gamma_g|\beta_g,\boldsymbol{x},\boldsymbol{y})$ is the pdf of the inverse gamma distribution $\mathcal{IG}\left(\frac{n_g}{\beta_g}, \sum_{k\in S_g} |x_k|^{\beta_g}\right)$ that is easy to sample. Conversely, it is more difficult to sample according to the pdf $f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y})$ which is given by

$$f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y}) \propto \frac{\beta_g^{n_g} 1_{[0,3]}(\beta_g)}{\gamma_g^{n_g/\beta_g}\left[\Gamma\left(1/\beta_g\right)\right]^{n_g}} \exp\left(\frac{-1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\right). \tag{9}$$

Consequently, this sampling step is achieved by using an MH move with a Gaussian proposal distribution $q(\beta_g \mid \beta_g^{(i-1)})$ truncated on $[0,3]$ with standard deviation $\sigma_{\beta_g} = 0.05$. The resulting sampler is summarized in Algorithm 1.

Algorithm 1 Proposed MCMC algorithm

Initialize with some $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_g^{(0)})_{1\leq g\leq G} = (\gamma_g^{(0)}, \beta_g^{(0)})_{1\leq g\leq G}$ and $\boldsymbol{x}^{(0)} \in C_\delta$. Set $i = 1$.

repeat

Sampling of $\boldsymbol{x}$:

- Given $\boldsymbol{x}^{(i-1)}$ generate $x_H^{(i)}$ and $x_{H^\perp}^{(i)}$.
- Compute the acceptance ratio $r(\widetilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)}) = \dfrac{f(\widetilde{\boldsymbol{x}}^{(i)}|\boldsymbol{\theta}^{(i-1)},\boldsymbol{y})\, q(\boldsymbol{x}^{(i-1)}|\widetilde{\boldsymbol{x}}^{(i)})}{f(\boldsymbol{x}^{(i-1)}|\boldsymbol{\theta}^{(i-1)},\boldsymbol{y})\, q_\eta(\widetilde{\boldsymbol{x}}^{(i)}|\boldsymbol{x}^{(i-1)})}$ and accept the proposed candidate $\widetilde{\boldsymbol{x}}^{(i)}$ with probability $\min\{1, r(\widetilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)})\}$.

Sampling of $\boldsymbol{\theta}$:

for $g = 1$ to G **do**

- Generate $\gamma_g^{(i)} \backsim \mathcal{IG}\left(\frac{n_g}{\beta_g^{(i-1)}}, \sum_{k\in S_g} |x_k^{(i)}|^{\beta_g^{(i-1)}}\right)$.

- Simulate $\beta_g^{(i)}$ as follows:

- Generate $\widetilde{\beta}_g^{(i)} \backsim q(\cdot \mid \beta_g^{(i-1)})$
- Compute the ratio $r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)}) = \frac{f(\widetilde{\beta}_g^{(i)}|\gamma_g^{(i)},\boldsymbol{x}^{(i)},\boldsymbol{y})q(\beta_g^{(i-1)}|\widetilde{\beta}_g^{(i)})}{f(\beta_g^{(i-1)}|\gamma_g^{(i)},\boldsymbol{x}^{(i)},\boldsymbol{y})q(\widetilde{\beta}_g^{(i)}|\beta_g^{(i-1)})}$ and accept the proposed candidate with the probability $\min\{1, r(\widetilde{\beta}_g^{(i)}, \beta_g^{(i-1)})\}$.

end for

until Convergence

4. EXPERIMENTAL RESULTS

To show the effectiveness of our algorithm, experiments have been carried out on synthetic and natural images. As a frame representation, we have used the union of two 2D separable wavelet bases $\mathcal{B}_1$ and $\mathcal{B}_2$ using Daubechies and shifted Daubechies filters of length 8 and 4, respectively. In the following, the sup-norm is used for $N(\cdot)$ in (4).

4.1. Experiments on synthetic data

To generate a synthetic signal, we synthesize wavelet frame coefficients $\boldsymbol{x}$ from known prior distributions.

Let $\boldsymbol{x}_1 = (a_1, (h_{1,j}, v_{1,j}, d_{1,j})_{1\leq j\leq 2})$ and $\boldsymbol{x}_2 = (a_2, (h_{2,j}, v_{2,j}, d_{2,j})_{1\leq j\leq 2})$ be the sequences of wavelet basis coefficients generated in $\mathcal{B}_1$ and $\mathcal{B}_2$ respectively, where a, h, v, d stand for approximation, horizontal, vertical and diagonal coefficients, and j is the resolution level. The number of groups (i.e., the number of subbands) G is therefore equal

[1] We recall that the range of F is $\mathrm{Ran}(F) = \{\boldsymbol{x} \in \mathbb{R}^K | \exists \boldsymbol{y} \in \mathbb{R}^L, F\boldsymbol{y} = \boldsymbol{x}\}$ and the null space of F^* is $\mathrm{Null}(F^*) = \{\boldsymbol{x} \in \mathbb{R}^K | F^*\boldsymbol{x} = \boldsymbol{0}\}$.

[2] It is here defined for every vector $\boldsymbol{a} \in \mathbb{R}^L$ as $P(\boldsymbol{a}) = \boldsymbol{a}$ if $N(\boldsymbol{a}) \leq \delta - \eta$ and $\frac{\delta-\eta}{N(\boldsymbol{a})}\boldsymbol{a}$ otherwise.

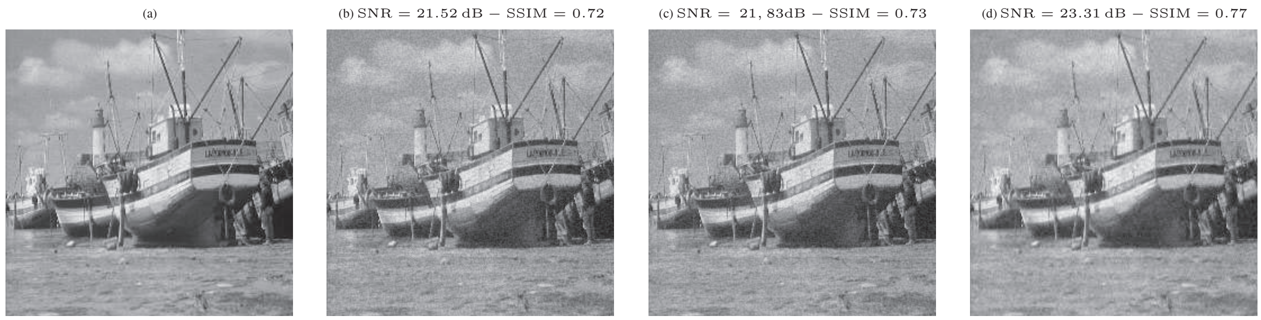

Fig. 1. Original image of size 256×256 (a), noisy image (b), denoised images using Wiener filtering (c) and our algorithm (d).

Table 1. NMSEs for the estimated hyper-parameters.

	a_1	a_2	$h_{1,1}$	$h_{2,1}$	$v_{1,1}$	$v_{2,1}$	$d_{1,1}$	$d_{2,1}$	$h_{1,2}$	$h_{2,2}$	$v_{1,2}$	$v_{2,2}$	$d_{1,2}$	$d_{2,2}$
β	0.039	0.010	0.012	0.015	0.022	0.025	0.011	0.029	0.021	0.016	0.020	0.013	0.023	0.011
α	0.023	0.028	0.030	0.025	0.026	0.031	0.044	0.023	0.026	0.034	0.019	0.022	0.031	0.040

to 14. Wavelet frame coefficients in each subband are generated from GG distributions in accordance with the chosen priors. Each parameter β_g is distributed according to a uniform distribution over $[0, 3]$ whereas a Jeffrey's prior is assigned to each parameter γ_g. After generating the hyper-parameters from their prior distributions, a set of frame coefficients is randomly generated to synthesize the observed data. The hyper-parameters are then supposed unknown and sampled using the proposed algorithm with $\delta = 10^{-4}$. They are finally estimated by computing the posterior mean of the generated samples (following the MMSE principle). Having reference values, the normalized mean square errors (NMSEs) related to the estimation of each hyper-parameter belonging to a given subband are computed from 30 Monte Carlo runs. The NMSE values are reported in Table 1. These error values indicate the good performance of the proposed algorithm for estimating the hyper-parameters.

4.2. Application to image denoising

In this experiment, we are interested in recovering an image from its noisy observation corrupted by a noise $\boldsymbol{n}$ uniformly distributed over the ball $[-\delta, \delta]^{256\times256}$ with $\delta = 20$. Denoising is performed by using the MMSE estimate which is computed from the wavelet frame coefficients sampled by our algorithm. The adjoint frame operator is then applied to recover the denoised image.The denoised images using Wiener filtering and our algorithm are depicted in Figs. 4 (c) and 4 (d), whereas the original and noisy images are shown in Figs. 4 (a) and 4 (b). It is clear that, when dealing with uniform noise, the proposed approach outperforms the Wiener filtering and provides good denoising results. Signal to noise ratio (SNR) and structural similarity (SSIM) values are also given in Fig. 4 in order to quantitatively evaluate the improvement brought by the proposed denoising technique.

5. CONCLUSION

We proposed an MCMC algorithm to estimate the frame coefficients and their hyper-parameters for signals/images based on a noisy observation. Our experiments showed that an accurate estimation of the frame coefficients and their hyper-parameters can be performed, which may be useful in many statistical signal/image processing problems. In our future work, we plan to apply this approach to wavelet based restoration and reconstruction problems.

6. REFERENCES

[1] S. Mallat, "Geometrical grouplets," *Appl. Comput. Harm. Anal.*, vol. 26, no. 2, pp. 161–180, 2009.

[2] E. J. Candès and D. L. Donoho, "Recovering edges in ill-posed inverse problems: Optimality of curvelet frames," *Ann. of Stat.*, vol. 30, no. 3, pp. 784–842, 2002.

[3] F. Destrempes, J.-F. Angers, and M. Mignotte, "Fusion of hidden Markov random field models and its Bayesian estimation," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2920–2935, 2006.

[4] C. Chaux, L. Duval, and J.-C. Pesquet, "Image analysis using a dual tree M-band wavelet transform," *IEEE Trans. on Image Process.*, vol. 15, no. 8, pp. 2397–2412, 2006.

[5] M. Seeger, S. Gerwinn, and M. Bethge, "Bayesian inference for sparse generalized linear models," *Mach. Lear.*, vol. 4701, pp. 298–309, 2007.

[6] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2004.

[7] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2059–2070, 2009.

[8] T. Blumensath and M. E. Davies, "Monte Carlo methods for adaptive sparse approximations of Time-Series," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4474–4486, 2007.

[9] S. Zeger and R. Karim, "Generalized linear models with random effects; A Gibbs sampling approach," *J. of the Am. Stat. Assoc.*, vol. 86, no. 413, pp. 79–86, 1991.

[10] L. Tierney, "Markov chains for exploring posterior distributions," *Ann. of Stat.*, vol. 22, no. 4, pp. 1701–1762, 1994.

[11] P. Mueller and B. Vidakovic, "MCMC methods in wavelet shrinkage: Non-equally spaced regression, density and spectral density estimation," *Bayesian Inference in Wavelet Based Models*, pp. 187–202, 1999, Springer-Verlag, New York.

[12] G. Heurta, "Multivariate Bayes wavelet shrinkage and applications," *J. of Appl. Stat.*, vol. 32, no. 5, pp. 529–542, 2005.