To link to this article: DOI: 10.1109/TSP.2010.2055562
http://dx.doi.org/10.1109/TSP.2010.2055562

To cite this version: Chaari, Lotfi and Pesquet, Jean-Christophe and Tourneret, Jean-Yves and Ciuciu, Philippe and Benazza-Benyahia, Amel (2010) *A Hierarchical Bayesian Model for Frame Representation.* IEEE Transactions on Signal Processing, vol. 58 (n° 11). pp. 5560-5571. ISSN 1053-587X

A Hierarchical Bayesian Model for Frame Representation

Lotfi Chaâri, *Student Member, IEEE*, Jean-Christophe Pesquet, *Senior Member, IEEE*, Jean-Yves Tourneret, *Senior Member, IEEE*, Philippe Ciuciu, *Member, IEEE*, and Amel Benazza-Benyahia

***Abstract*—In many signal processing problems, it is fruitful to represent the signal under study in a frame. If a probabilistic approach is adopted, it becomes then necessary to estimate the hyperparameters characterizing the probability distribution of the frame coefficients. This problem is difficult since in general the frame synthesis operator is not bijective. Consequently, the frame coefficients are not directly observable. This paper introduces a hierarchical Bayesian model for frame representation. The posterior distribution of the frame coefficients and model hyperparameters is derived. Hybrid Markov chain Monte Carlo algorithms are subsequently proposed to sample from this posterior distribution. The generated samples are then exploited to estimate the hyperparameters and the frame coefficients of the target signal. Validation experiments show that the proposed algorithms provide an accurate estimation of the frame coefficients and hyperparameters. Application to practical problems of image denoising in the presence of uniform noise illustrates the impact of the resulting Bayesian estimation on the recovered signal quality.**

***Index Terms*—Bayesian estimation, compressed sensing, frame representations, generalized Gaussian, hyperparameter estimation, MCMC, Metropolis Hastings, sparsity, wavelets.**

I. Introduction

DATA representation is a crucial operation in many signal and image processing applications. These applications include signal and image reconstruction [1], [2], restoration [3], [4] and compression [5], [6]. In this respect, many linear transforms have been proposed in order to obtain suitable signal representations in other domains than the original spatial or temporal ones. The traditional Fourier and discrete cosine transforms provide a good frequency localization, 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 [7]. Many wavelet-like basis decompositions have been subsequently proposed offering different features. For instance, we can mention the wavelet packets [8] or the grouplet bases [9]. To further improve signal representations, redundant linear decomposition families called *frames* have become the focus of many works during the last decade. For the sake of clarity, it must be pointed out that the term frame [10] is understood in the sense of Hilbert space theory and not in the sense of some recent works like [11].

The main advantage of frames lies in their flexibility to capture local features of the signal. Hence, they may result in sparse representations as shown in the literature on curvelets [10], contourlets [12], bandelets [13] or dual-trees [14] in image processing. However, a major difficulty when using frame representations (FRs) in a statistical framework is 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.

This paper studies a hierarchical Bayesian approach to estimate the frame coefficients and their hyperparameters. Although this approach is conceptually able to deal with any desirable distribution for the frame coefficients, we focus in this paper on generalized Gaussian (GG) priors. Note however that we do not restrict our attention to log-concave GG prior probability density functions (pdf), which may be limited for providing accurate models of sparse signals [15]. In addition, the proposed method can be applied to noisy data when imprecise measurements of the signal are only available. One of the contributions of this work is to address the case of uniform noise that has been used for bounded error measurement models [17]–[20] and to model quantization errors in data compression [16].

Our work takes advantage of the current developments in Markov chain Monte Carlo (MCMC) algorithms [21]–[23] that have already been investigated for instance in image separation [24], image restoration [25] and brain activity detection in functional MRI [26]. These algorithms have also been investigated for signal/image processing problems with sparsity constraints. These constraints may be imposed in the original space like in [27], where a sparse image reconstruction problem is assessed in the image domain. They may also be imposed on some redundant representation of the signal like in [28], where a time-series sparse coding problem is considered.

Hybrid MCMC algorithms [29], [30] combine Metropolis-Hastings (MH) [31] and Gibbs [32] moves to sample according to the posterior distribution of interest. MCMC algorithms and

Manuscript received November 06, 2009; accepted June 12, 2010. Date of publication June 28, 2010; date of current version October 13, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Wing-Kin Ma. This work was supported by grants from Région Ile de France and the Agence Nationale de la Recherche by Grant ANR-05-MMSA-0014-01.

L. Chaâri and J.-C. Pesquet are with the LIGM and UMR-CNRS 8049, Université Paris-Est, 77454 Marne-la-Vallée, France (e-mail: lotfi.chaari@univ-paris-est.fr; jean-christophe.pesquet@univ-paris-est.fr).

J.-Y. Tourneret is with the University of Toulouse, IRIT/ENSEEIHT/TSA, 31071 Toulouse, France (e-mail: jean-yves.tourneret@enseeiht.fr).

P. Ciuciu is with the CEA/DSV/I2BM/Neurospin, CEA Saclay, 91191 Gif-sur-Yvette cedex, France (e-mail: philippe.ciuciu@cea.fr).

A. Benazza-Benyahia is with the Ecole Supérieure des Communications de Tunis (SUP'COM-Tunis), Unité de Recherche en Imagerie Satellitaire et ses Applications (URISA) , Cité Technologique des Communications, 2083, Tunisia (e-mail: benazza.amel@supcom.rnu.tn).

Digital Object Identifier 10.1109/TSP.2010.2055562

WT have been jointly investigated in some works dealing with signal denoising under a Bayesian framework [24], [33]–[35]. However, in contrast with the present paper where overcomplete FRs are considered, these works are limited to wavelet bases for which the hyperparameter estimation problem is much easier to handle. Other interesting works concerning the use of MCMC methods for generating sparse representations [36], [37] assume Gaussian noise models, which may facilitate the derivation of the proposed sampler, especially when a mixture of Gaussians. Alternative Bayesian approaches have also been proposed in [38] and [39] for some specific forms of FRs.

This paper is organized as follows. Section II presents a brief overview on the concepts of frame and FR. The hierarchical Bayesian model proposed for FR is introduced in Section III. Two algorithms for sampling the posterior distribution are proposed in Section IV. To illustrate the effectiveness of these algorithms, experiments on both synthetic and real data are presented in Section V. In this section, applications to image recovery problems are also considered. Finally, some conclusions are drawn in Section VI.

II. Problem Formulation

A. The Frame Concept

In the following, we will consider real-valued digital signals of length L as elements of the Euclidean space $\mathbb{R}^L$ endowed with the usual scalar product and norm denoted as $\langle .|. \rangle$ and $||\cdot||$, respectively. Let K be an integer greater than or equal to L. A family of vectors $(\boldsymbol{e}_k)_{1\leq k\leq K}$ in the finite-dimensional space $\mathbb{R}^L$ is a frame when there exists a constant μ in $]0, +\infty[$ such that[1]

$$\forall \boldsymbol{y} \in \mathbb{R}^L, \quad \mu||\boldsymbol{y}||^2 \leq \sum_{k=1}^{K} |\, \langle \boldsymbol{y}|\boldsymbol{e}_k \rangle \,|^2. \tag{1}$$

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

$$\begin{aligned} F :& \mathbb{R}^L \to \mathbb{R}^K : \; \boldsymbol{y} \mapsto (\langle \boldsymbol{y}|\boldsymbol{e}_k \rangle)_{1\leq k\leq K} \\ F^* :& \mathbb{R}^K \to \mathbb{R}^L : \; (\xi_k)_{1\leq k\leq K} \mapsto \sum_{k=1}^{K} \xi_k \boldsymbol{e}_k. \end{aligned} \tag{2}$$

Note that F is injective whereas F^* is surjective. When $F^{-1} = F^*$, $(\boldsymbol{e}_k)_{k\in\mathbb{K}}$ is an orthonormal basis, where $\mathbb{K} = \{1, \ldots, K\}$. 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.

B. Frame Representation

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

$$\boldsymbol{y} = F^*\boldsymbol{x} + \boldsymbol{n} \tag{3}$$

[1]The classical upper bound condition is always satisfied in finite dimension. The frame condition here is also equivalent to saying that F has full rank L.

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 x belongs to the closed convex set

$$C_\delta = \{\boldsymbol{x} \in \mathbb{R}^K \mid N(\boldsymbol{y} - F^*\boldsymbol{x}) \leq \delta\} \tag{4}$$

where $\delta \in [0, \infty[$ is some error bound and $N(\cdot)$ can be any norm on $\mathbb{R}^L$.

In signal/image recovery problems, $\boldsymbol{n}$ is nothing but an additive noise that corrupts the measured data. In this paper, we will focus on the case of a bounded observation error modeled by uniform noise. By adopting a probabilistic approach, $\boldsymbol{y}$ 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 $\boldsymbol{X}|\boldsymbol{Y}$, by considering some parametric probabilistic model and by estimating the associated hyperparameters. A useful example where this characterization may be of great interest is frame-based signal/image denoising under a Bayesian framework. Actually, denoising in the wavelet domain using wavelet frame decompositions has already been investigated since the seminal work in [40] as this kind of representation provides sparse description of regular signals. The related hyperparameters have then to be estimated. 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 II-A, for redundant frames, F^* is not bijective, which makes the hyperparameter estimation problem more difficult. This paper presents hierarchical Bayesian algorithms to address this issue.

III. 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. In the following, $f(\boldsymbol{x}|\boldsymbol{\theta})$ denotes the pdf of the frame coefficients that depends on an unknown hyperparameter vector $\boldsymbol{\theta}$ and $f(\boldsymbol{\theta})$ is the *a priori* pdf for the hyperparameter vector $\boldsymbol{\theta}$. In compliance with the observation model (3) and the constraint (4), $\boldsymbol{n}$ is assumed to be uniformly distributed on the ball

$$B_{\boldsymbol{0},\delta} = \{\boldsymbol{a} \in \mathbb{R}^L \mid N(\boldsymbol{a}) \leq \delta\}. \tag{5}$$

From (3), it can be deduced that $f(\boldsymbol{y}|\boldsymbol{x})$ is the uniform pdf on the closed convex ball $B_{F^*\boldsymbol{x},\delta}$ defined as

$$B_{F^*\boldsymbol{x},\delta} = \{\boldsymbol{y} \in \mathbb{R}^L \mid N(\boldsymbol{y} - F^*\boldsymbol{x}) \leq \delta\}. \tag{6}$$

Denoting by $\boldsymbol{\Theta}$ the random vector associated with the hyperparameter vector $\boldsymbol{\theta}$ and using the hierarchical structure between $\boldsymbol{Y}$, $\boldsymbol{X}$ and $\boldsymbol{\Theta}$, the conditional pdf of $(\boldsymbol{X}, \boldsymbol{\Theta})$ given $\boldsymbol{Y}$ can be written as

$$f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{x}) f(\boldsymbol{x}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \tag{7}$$

where $\propto$ means *proportional to*.

In this work, we assume that frame coefficients are *a priori* independent with marginal GG distributions. This assumption has been successfully used in many studies [41]–[45] and leads to the following frame coefficient prior

$$f(x_k|\alpha_k,\beta_k) = \frac{\beta_k}{2\alpha_k\Gamma\left(\frac{1}{\beta_k}\right)}\exp\left(-\frac{|x_k|^{\beta_k}}{\alpha_k^{\beta_k}}\right) \tag{8}$$

where $\alpha_k > 0, \beta_k > 0$ (with $k \in \mathbb{K}$) are the scale and shape parameters associated with x_k, which is the kth component of the frame coefficient vector $\boldsymbol{x}$ and $\Gamma(.)$ is the Gamma function. Note that small values of the shape parameters are appropriate for modeling sparse signals. For instance, when $\beta_k = 1$, for $k \in \mathbb{K}$, (8) reduces to the Laplace prior which plays a central role in sparse signal recovery [46] and compressed sensing [47]. By introducing $\gamma_k = \alpha_k^{\beta_k}$, the frame prior can be rewritten as[2]

$$f(x_k|\gamma_k,\beta_k) = \frac{\beta_k}{2\gamma_k^{1/\beta_k}\Gamma\left(\frac{1}{\beta_k}\right)}\exp\left(-\frac{|x_k|^{\beta_k}}{\gamma_k}\right). \tag{9}$$

The distribution of a frame coefficient generally differs from one coefficient to another. However, some frame coefficients can have very similar distributions (that can be defined by the same hyperparameters β_k and γ_k). Consequently, we propose to split the frame coefficients into G different groups. The gth group will be parameterized by a unique hyperparameter vector denoted as $\boldsymbol{\theta}_g = (\beta_g, \gamma_g)$ (after the reparameterization aforementioned). In this case, the frame prior can be expressed as

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{g=1}^{G}\left[\left(\frac{\beta_g}{2\gamma_g^{1/\beta_g}\Gamma\left(\frac{1}{\beta_g}\right)}\right)^{n_g}\exp\left(-\frac{1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\right)\right] \tag{10}$$

where the summation covers the index set S_g of the elements of the gth group containing n_g elements and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_G)$. Note that in our simulations, each group g will correspond to a given wavelet subband. A coarser classification may be made when using multiscale frame representations by considering that all the frame coefficients at a given resolution level belong to a same group.

The hierarchical Bayesian model for the frame decomposition is completed by the following improper hyperprior:

$$f(\boldsymbol{\theta}) = \prod_{g=1}^{G} f(\boldsymbol{\theta}_g) = \prod_{g=1}^{G}[f(\gamma_g)f(\beta_g)]$$

$$\propto \prod_{g=1}^{G}\left[\frac{1}{\gamma_g}1_{\mathbb{R}^+}(\gamma_g)1_{[0,3]}(\beta_g)\right] \tag{11}$$

where $1_A(\xi)$ is the function defined on $A \subset \mathbb{R}$ by $1_A(\xi) = 1$ if $\xi \in A$ and $1_A(\xi) = 0$ otherwise.

The motivations for using this kind of prior are summarized:

- the interval $[0, 3]$ covers all possible values of β_g encountered in practical applications. Moreover, there is no additional information about the parameter β_g.
- The prior for γ_g is a Jeffrey's distribution that reflects the absence of knowledge about this parameter. This kind of prior is often used for scale parameters [48].

The resulting posterior distribution is given by (12) as shown at the bottom of the page. The Bayesian estimators [e.g., the maximum *a posteriori* (MAP) or minimum mean square error (MMSE) estimators] associated with the posterior distribution (12) have no simple closed-form expression. The next section studies different sampling strategies that allow one to generate samples asymptotically distributed according to the posterior distribution (12). The generated samples will be used to estimate the unknown parameter and hyperparameter vectors $\boldsymbol{x}$ and $\boldsymbol{\theta}$.

IV. Sampling Strategies

This section proposes different MCMC methods to generate samples asymptotically distributed according to the posterior $f(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})$ defined in (12).

A. Hybrid Gibbs Sampler

A very standard strategy to sample according to (12) is provided by the Gibbs sampler (GS). GS iteratively generates samples distributed according to conditional distributions associated with the target distribution. Precisely, the GS iteratively generates samples distributed according to $f(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ and $f(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{y})$.

1) Sampling the Frame Coefficients: Straightforward calculations yield the following conditional distribution

$$f(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \propto 1_{C_\delta}(\boldsymbol{x})\prod_{g=1}^{G}\exp\left(-\frac{1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\right) \tag{13}$$

where C_δ is defined in (4). This conditional distribution is a product of GG distributions truncated on C_δ. Actually, sampling according to this truncated distribution is not always easy to perform since the adjoint frame operator F^* is usually of large dimension. However, two alternative sampling strategies are detailed in what follows.

[2]The interest of this new parameterization will be clarified in Section IV.

$$f(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) \propto 1_{C_\delta}(\boldsymbol{x})\prod_{g=1}^{G}\left[\left(\frac{\beta_g}{2\gamma_g^{1/\beta_g}\Gamma\left(\frac{1}{\beta_g}\right)}\right)^{n_g}\exp\left(-\frac{1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\right)\left(\frac{1}{\gamma_g}1_{\mathbb{R}^+}(\gamma_g)1_{[0,3]}(\beta_g)\right)\right]. \tag{12}$$

a) Naive Sampling: This sampling method proceeds by sampling according to independent GG distributions

$$\prod_{g=1}^{G} \exp\left(-\frac{1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\right) \tag{14}$$

and then accepting the proposed candidate $\boldsymbol{x}$ only if $N(\boldsymbol{y} - F^*\boldsymbol{x}) \leq \delta$. This method can be used for any frame decomposition and any norm. However, it can be inefficient because of a very low acceptance ratio when δ takes small values.

b) Gibbs Sampler: This sampling method is designed to sample more efficiently from the conditional distribution in (13) when the considered frame is the union of M orthonormal bases and $N(\cdot)$ is the Euclidean norm. In this case, the analysis frame operator and the corresponding adjoint can be written as

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_M \end{bmatrix}$$

and

$$F^* = [F_1^* \dots F_M^*],$$

respectively, where $\forall m \in \{1,\dots,M\}$, F_m is the decomposition operator onto the mth orthonormal basis such as $F_m^*F_m = F_mF_m^* = \mathrm{I}$. In what follows, we will decompose every $\boldsymbol{x} \in \mathbb{R}^K$ with $K = ML$ as $\boldsymbol{x} = \left[\boldsymbol{x}_1^\top,\dots,\boldsymbol{x}_M^\top\right]^\top$ where $\boldsymbol{x}_m \in \mathbb{R}^L$, for every $m \in \{1,\dots,M\}$. The GS for the generation of frame coefficients draws vectors according to the conditional distribution $f(\boldsymbol{x}_n|\boldsymbol{x}_{-n},\boldsymbol{y},\boldsymbol{\theta})$ under the constraint $N(\boldsymbol{y} - F^*\boldsymbol{x}) \leq \delta$, where $\boldsymbol{x}_{-n}$ is the reduced size vector of dimension $\mathbb{R}^{K-L}$ built from $\boldsymbol{x}$ by removing the nth vector $\boldsymbol{x}_n$. If $N(\cdot)$ is the Euclidean norm, we have for every $n \in \{1,\dots,M\}$

$$\begin{aligned} N(\boldsymbol{y} - \sum_{m=1}^{M} F_m^*\boldsymbol{x}_m) \leq \delta &\\ &\Leftrightarrow \left\| F_n^*(F_n\boldsymbol{y} - \sum_{m=1}^{M} F_nF_m^*\boldsymbol{x}_m)\right\| \leq \delta \\ &\Leftrightarrow \left\| F_n\boldsymbol{y} - \sum_{m\neq n} F_nF_m^*\boldsymbol{x}_m - \boldsymbol{x}_n\right\| \leq \delta \\ &\Leftrightarrow N(\boldsymbol{x}_n - \boldsymbol{c}_n) \leq \delta, \\ &\quad \text{where } \boldsymbol{c}_n = F_n\left(\boldsymbol{y} - \sum_{m\neq n} F_m^*\boldsymbol{x}_m\right). \end{aligned}$$

To sample each $\boldsymbol{x}_n$, we propose to use an MH step whose proposal distribution is supported on the ball $B_{\boldsymbol{c}_n,\delta}$ defined by

$$B_{\boldsymbol{c}_n,\delta} = \{\boldsymbol{a} \in \mathbb{R}^L \mid N(\boldsymbol{a} - \boldsymbol{c}_n) \leq \delta\}. \tag{15}$$

Random generation from a pdf q_δ defined on $B_{\boldsymbol{0},\delta}$ is described in Appendix A. Having a closed form expression of this pdf is important to be able to calculate the acceptance ratio of the MH move. To take into account the value of $\boldsymbol{x}_n^{(i-1)}$ obtained at the previous iteration $(i-1)$, it may however be preferable to choose a proposal distribution supported on a restricted ball of radius $\eta \in]0,\delta[$ containing $\boldsymbol{x}_n^{(i-1)}$. This strategy similar to the random walk MH algorithm [21, p. 287] results in a better exploration of regions associated with large values of the conditional distribution $f(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$. More precisely, we propose to choose a proposal distribution defined on $B_{\hat{\boldsymbol{x}}_n^{(i-1)},\eta}$, where $\hat{\boldsymbol{x}}_n^{(i-1)} = P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n) + \boldsymbol{c}_n$ and P is the projection onto the ball $B_{\boldsymbol{0},\delta-\eta}$ defined as

$$\forall \boldsymbol{a} \in \mathbb{R}^L, \quad P(\boldsymbol{a}) = \begin{cases} \boldsymbol{a}, & \text{if } N(\boldsymbol{a}) \leq \delta - \eta \\ \frac{\delta-\eta}{N(\boldsymbol{a})}\boldsymbol{a}, & \text{otherwise.} \end{cases} \tag{16}$$

This choice of the center of the ball guarantees that $B_{\hat{\boldsymbol{x}}_n^{(i-1)},\eta} \subset B_{\boldsymbol{c}_n,\delta}$. Moreover, any point of $B_{\boldsymbol{c}_n,\delta}$ can be reached after consecutive draws in $B_{\hat{\boldsymbol{x}}_n^{(i-1)},\eta}$. Note that the radius η has to be adjusted to ensure a good exploration of $B_{\boldsymbol{c}_n,\delta}$. In practice, it may also be interesting to fix a small enough value of η (compared with δ) so as to improve the acceptance ratio.

Remark: Alternatively, a GS can be used to draw successively the L elements $(x_{n,l})_{1\leq l\leq L}$ of $\boldsymbol{x}_n$ under the following constraint for every $\forall l \in \{1,\dots,L\}$ by setting $s_{n,l} = \sum_{k\neq l}(x_{n,l} - c_{n,k})^2$:

$$\| \boldsymbol{x}_n - \boldsymbol{c}_n \| \leq \delta \Leftrightarrow -\sqrt{\delta^2 - s_{n,l}} \leq x_{n,l} - c_{n,l} \leq \sqrt{\delta^2 - s_{n,l}}$$

where $c_{n,k}$ is the kth element of the vector $\boldsymbol{c}_n$. However, this method is very time-consuming since it proceeds sequentially for each component of the high dimensional vector $\boldsymbol{x}$.

2) Sampling the Hyperparameter Vector: 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 $f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y})$. Straightforward calculations allow us to obtain the following results:

$$\begin{aligned} f(\gamma_g|\beta_g,\boldsymbol{x},\boldsymbol{y}) &\propto \gamma_g^{-n_g/\beta_g-1} \exp\left(-\frac{1}{\gamma_g}\sum_{k\in S_g}|x_k|^{\beta_g}\right) 1_{\mathbb{R}^+}(\gamma_g) \\ f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y}) &\propto \frac{\beta_g^{n_g} 1_{[0,3]}(\beta_g)}{\gamma_g^{n_g/\beta_g}\left[\Gamma\left(\frac{1}{\beta_g}\right)\right]^{n_g}} \exp\left(\sum_{k\in S_g}\frac{-|x_k|^{\beta_g}}{\gamma_g}\right). \end{aligned} \tag{17}$$

Consequently, due to the new parameterization introduced in (9), $f(\gamma_g|\beta_g,\boldsymbol{x},\boldsymbol{y})$ is the pdf of the inverse gamma distribution $\mathcal{IG}\left(n_g/\beta_g, \sum_{k\in S_g}|x_k|^{\beta_g}\right)$ that is easy to sample. Conversely, it is more difficult to sample according to the truncated pdf $f(\beta_g|\gamma_g,\boldsymbol{x},\boldsymbol{y})$. This is achieved by using an MH move whose proposal $q(\beta_g \mid \beta_g^{(i-1)})$ is a Gaussian distribution truncated on the interval $[0,3]$ with standard deviation $\sigma_{\beta_g} = 0.05$ [49]. Note that the mode of this distribution is the value of the parameter $\beta_g^{(i-1)}$ at the previous iteration $(i-1)$.

The resulting method is the hybrid GS summarized in Algorithm 1. Although this algorithm is intuitive and simple to implement, it must be pointed out that it was derived under the restrictive assumption that the considered frame is the union of M orthonormal bases. When these assumptions do not hold, another algorithm proposed in the next section allows us to sample frame coefficients and the related hyperparameters by exploiting algebraic properties of frames.

Algorithm 1 Hybrid GS to simulate according to $f(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})$ (superscript $\cdot^{(i)}$ indicates values computed at ith iteration)

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

② Sampling $\boldsymbol{x}$

for $n=1$ to M **do**

—Compute $\boldsymbol{c}_n^{(i)} = F_n\left(\boldsymbol{y} - \sum_{m<n} F_m^* \boldsymbol{x}_m^{(i)} - \sum_{m>n} F_m^* \boldsymbol{x}_m^{(i-1)}\right)$ and $\hat{\boldsymbol{x}}_n^{(i-1)} = P(\boldsymbol{x}_n^{(i-1)} - \boldsymbol{c}_n^{(i)}) + \boldsymbol{c}_n^{(i)}$.

—Simulate $\boldsymbol{x}_n^{(i)}$ as follows:

• Generate $\tilde{\boldsymbol{x}}_n^{(i)} \sim q_\eta(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n^{(i-1)})$ where q_η is defined on $B_{\boldsymbol{0},\eta}$ (see Appendix A).

• Compute the ratio $r(\tilde{\boldsymbol{x}}_n^{(i)},\boldsymbol{x}_n^{(i-1)})$ (see the equation at the bottom of the page) and accept the proposed candidate with the probability $\min\{1, r(\tilde{\boldsymbol{x}}_n^{(i)},\boldsymbol{x}_n^{(i-1)})\}$.

end for

③ Sampling $\boldsymbol{\theta}$

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

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

—Simulate $\beta_g^{(i)}$ as follows:

• Generate $\tilde{\beta}_g^{(i)} \backsim q(\beta_g \mid \beta_g^{(i-1)})$

• Compute the ratio

$$r(\tilde{\beta}_g^{(i)},\beta_g^{(i-1)}) = \frac{f(\tilde{\beta}_g^{(i)}|\gamma_g^{(i)},\boldsymbol{x}^{(i)},\boldsymbol{y})q(\beta_g^{(i-1)} \mid \tilde{\beta}_g^{(i)})}{f(\beta_g^{(i-1)}|\gamma_g^{(i)},\boldsymbol{x}^{(i)},\boldsymbol{y})q(\tilde{\beta}_g^{(i)} \mid \beta_g^{(i-1)})}$$

and accept the proposed candidate with the probability $\min\{1, r(\tilde{\beta}_g^{(i)},\beta_g^{(i-1)})\}$.

end for

④ Set $i \leftarrow i+1$ and goto ② until convergence.

B. Hybrid MH Sampler Using Algebraic Properties of Frame Representations

As a direct generation of samples according to $f(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ is generally impossible, we propose here an alternative that replaces the Gibbs move by an MH move. This MH move aims at sampling globally a candidate $\boldsymbol{x}$ according to a proposal distribution. This candidate is accepted or rejected with the standard MH acceptance ratio. The efficiency of the MH move strongly depends on the choice of the proposal distribution for $\boldsymbol{x}$. We denote as $\boldsymbol{x}^{(i)}$ the ith 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 II-B) 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, any frame coefficient vector can be decomposed as $\boldsymbol{x} = \boldsymbol{x}_H + \boldsymbol{x}_{H^\perp}$, 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.[3] 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, the following separable form of the proposal pdf is considered:

$$q(\boldsymbol{x} \mid \boldsymbol{x}^{(i)}) = q\left(\boldsymbol{x}_H \mid \boldsymbol{x}_H^{(i-1)}\right) q\left(\boldsymbol{x}_{H^\perp} \mid \boldsymbol{x}_{H^\perp}^{(i-1)}\right) \tag{18}$$

where $\boldsymbol{x}_H^{(i-1)} \in H, \boldsymbol{x}_{H^\perp}^{(i-1)} \in H^\perp$ and $\boldsymbol{x}^{(i-1)} = \boldsymbol{x}_H^{(i-1)} + \boldsymbol{x}_{H^\perp}^{(i-1)}$, i.e., $\boldsymbol{x}_H$ and $\boldsymbol{x}_{H^\perp}$ are sampled independently.

If we consider the decomposition $\boldsymbol{x} = \boldsymbol{x}_H + \boldsymbol{x}_{H^\perp}$, sampling $\boldsymbol{x}$ in C_δ is equivalent to sampling $\boldsymbol{\lambda} \in \bar{C}_\delta$, where $\bar{C}_\delta = \{\boldsymbol{\lambda} \in \mathbb{R}^L | N(\boldsymbol{y} - F^*F\boldsymbol{\lambda}) \le \delta\}$. 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 $\bar{C}_\delta$ can be easily achieved, e.g., by generating $\boldsymbol{u}$ from a 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, taking into account the sampled value at the previous iteration $\boldsymbol{x}_H^{(i-1)} = F\boldsymbol{\lambda}^{(i-1)} = F(F^*F)^{-1}\boldsymbol{u}^{(i-1)}$ may be interesting. Similarly to Section IV-A.1b, and to random walk generation techniques, we proceed by generating $\boldsymbol{u}$ in $B_{\hat{\boldsymbol{u}}^{(i-1)},\eta}$ where $\eta \in]0,\delta[$ and $\hat{\boldsymbol{u}}^{(i-1)} = P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) + \boldsymbol{y}$. This allows us to draw a vector $\boldsymbol{u}$ such that $\boldsymbol{x}_H = F(F^*F)^{-1}\boldsymbol{u} \in C_\delta$ and $N(\boldsymbol{u} - \boldsymbol{u}^{(i-1)}) \le 2\eta$. The generation of $\boldsymbol{u}$ can then be performed as explained in Appendix A provided that $N(\cdot)$ is an ℓ_p norm with $p \in [1,+\infty]$.

Once we have simulated $\boldsymbol{x}_H = F\boldsymbol{\lambda} \in H \cap C_\delta$ (which ensures that $\boldsymbol{x}$ is in C_δ), $\boldsymbol{x}_{H^\perp}$ has to be sampled as an element of $H^\perp$. Since $\boldsymbol{y} = F^*\boldsymbol{x} + \boldsymbol{n} = F^*\boldsymbol{x}_H + \boldsymbol{n}$, there is no information in $\boldsymbol{y}$ about $\boldsymbol{x}_{H^\perp}$. As a consequence, we propose to sample $\boldsymbol{x}_H$ by drawing $\boldsymbol{z}$ according to the Gaussian distribution $\mathcal{N}(\boldsymbol{x}^{(i-1)}, \sigma_{\boldsymbol{x}}^2 \boldsymbol{I})$ and by projecting $\boldsymbol{z}$ onto $H^\perp$, i.e.,

$$\boldsymbol{x}_{H^\perp} = \Pi_{H^\perp}\boldsymbol{z} \tag{19}$$

where $\Pi_{H^\perp} = \mathrm{I} - F(F^*F)^{-1}F^*$ is the orthogonal projection operator onto $H^\perp$.[4]

[3]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 $\mathrm{Null}(F^*) = \{\boldsymbol{x} \in \mathbb{R}^K | F^*\boldsymbol{x} = \boldsymbol{0}\}$.

[4]Note here that using a tight frame makes the computation of both $\boldsymbol{x}_H$ and $\boldsymbol{x}_{H^\perp}$ much easier due to the relation $F^*F = \mu\mathrm{I}$.

$$r(\tilde{\boldsymbol{x}}_n^{(i)},\boldsymbol{x}_n^{(i-1)}) = \frac{f(\tilde{\boldsymbol{x}}_n^{(i)}|\boldsymbol{\theta}^{(i-1)},(\boldsymbol{x}_m^{(i)})_{m<n},(\boldsymbol{x}_m^{(i-1)})_{m>n},\boldsymbol{y})\; q_\eta\left(\boldsymbol{x}_n^{(i-1)} - P(\tilde{\boldsymbol{x}}_n^{(i)} - \boldsymbol{c}_n^{(i)}) - \boldsymbol{c}_n^{(i)}\right)}{f(\boldsymbol{x}_n^{(i-1)}|\boldsymbol{\theta}^{(i-1)},(\boldsymbol{x}_m^{(i)})_{m<n},(\boldsymbol{x}_m^{(i-1)})_{m>n},\boldsymbol{y}) q_\eta\left(\tilde{\boldsymbol{x}}_n^{(i)} - \hat{\boldsymbol{x}}_n^{(i-1)}\right)}$$

Let us now derive the expression of the proposal pdf. It can be noticed that, if $K > L$, there exists a linear operator $F_\perp$ from $\mathbb{R}^{K-L}$ to $\mathbb{R}^K$ which is semi-orthogonal (i.e., $F_\perp^* F_\perp = \mathrm{I}$) and orthogonal to F (i.e., $F_\perp^* F = 0$), such that

$$\boldsymbol{x} = \underbrace{F\boldsymbol{\lambda}}_{\boldsymbol{x}_H} + \underbrace{F_\perp \boldsymbol{\lambda}_\perp}_{\boldsymbol{x}_{H^\perp}} \tag{20}$$

and $\boldsymbol{\lambda}_\perp = F_\perp^* \boldsymbol{x} \in \mathbb{R}^{K-L}$. Standard rules on bijective linear transforms of random vectors lead to

$$q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)}) = |\det([F \quad F_\perp])|^{-1} q(\boldsymbol{\lambda} \mid \boldsymbol{x}^{(i-1)}) q(\boldsymbol{\lambda}_\perp \mid \boldsymbol{x}^{(i-1)}) \tag{21}$$

where, due to the bijective mapping between $\boldsymbol{\lambda}$ and $\boldsymbol{u} = F^* F \boldsymbol{\lambda}$

$$q(\boldsymbol{\lambda} \mid \boldsymbol{x}^{(i-1)}) = \det(FF^*) q_\eta(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)}) \tag{22}$$

and $q(\boldsymbol{\lambda}_\perp \mid \boldsymbol{x}^{(i-1)})$ is the pdf of the Gaussian distribution $\mathcal{N}(\boldsymbol{\lambda}_\perp^{(i-1)}, \sigma_{\boldsymbol{x}}^2 \boldsymbol{I})$ with mean $\boldsymbol{\lambda}_\perp^{(i-1)} = F_\perp^* \boldsymbol{x}^{(i-1)}$. Recall that q_η denotes a pdf defined on the ball $B_{\mathbf{0},\eta}$ as expressed in Appendix A. Due to the symmetry of the Gaussian distribution, it can be deduced that

$$\frac{q(\boldsymbol{x}^{(i-1)} \mid \boldsymbol{x})}{q(\boldsymbol{x} \mid \boldsymbol{x}^{(i-1)})} = \frac{q_\eta(\boldsymbol{u}^{(i-1)} - P(\boldsymbol{u} - \boldsymbol{y}) - \boldsymbol{y})}{q_\eta(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)})}. \tag{23}$$

This expression remains valid in the degenerate case when $K = L$ (yielding $\boldsymbol{x}_{H^\perp} = \mathbf{0}$). Finally, it is important to note that, if q_η is the uniform distribution on the ball $B_{\mathbf{0},\,\eta}$, the above ratio reduces to 1, which simplifies the computation of the MH acceptance ratio. The final algorithm is summarized in Algorithm 2. Note that the sampling of the hyperparameter vector is performed as for the hybrid GS in Section IV-A.2.

Algorithm 2 Hybrid MH sampler using algebraic properties of frame representations to simulate according to $f(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$

① Initialize with some $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\theta}_g^{(0)})_{1\le g\le G} = (\gamma_g^{(0)}, \beta_g^{(0)})_{1\le g\le G}$ and $\boldsymbol{u}^{(0)} \in B_{\boldsymbol{y},\delta}$. Set $\boldsymbol{x}^{(0)} = F(F^*F)^{-1}\boldsymbol{u}^{(0)}$ and $i = 1$.

② Sampling $\boldsymbol{x}$

- Compute $\hat{\boldsymbol{u}}^{(i-1)} = P(\boldsymbol{u}^{(i-1)} - \boldsymbol{y}) + \boldsymbol{y}$.
- Generate $\tilde{\boldsymbol{u}}^{(i)} \backsim q_\eta(\boldsymbol{u} - \hat{\boldsymbol{u}}^{(i-1)})$ where q_η is defined on $B_{\mathbf{0},\eta}$ (see Appendix A).
- Compute $\tilde{\boldsymbol{x}}_H^{(i)} = F(F^*F)^{-1}\tilde{\boldsymbol{u}}^{(i)}$.
- Generate $z^{(i)} \backsim \mathcal{N}(\boldsymbol{x}^{(i-1)}, \sigma_{\boldsymbol{x}}^2 \boldsymbol{I})$.
- Compute $\tilde{\boldsymbol{x}}_{H^\perp}^{(i)} = \Pi_{H^\perp} z^{(i)}$ and $\tilde{\boldsymbol{x}}^{(i)} = \tilde{\boldsymbol{x}}_H^{(i)} + \tilde{\boldsymbol{x}}_{H^\perp}^{(i)}$.
- Compute the ratio

$$r(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)}) = \frac{f(\tilde{\boldsymbol{x}}^{(i)}|\boldsymbol{\theta}^{(i-1)}, \boldsymbol{y}) q_\eta\left(\boldsymbol{u}^{(i-1)} - P(\tilde{\boldsymbol{u}}^{(i)} - \boldsymbol{y}) - \boldsymbol{y}\right)}{f(\boldsymbol{x}^{(i-1)}|\boldsymbol{\theta}^{(i-1)}, \boldsymbol{y}) q_\eta\left(\tilde{\boldsymbol{u}}^{(i)} - \hat{\boldsymbol{u}}^{(i-1)}\right)}$$

and accept the proposed candidates $\tilde{\boldsymbol{u}}^{(i)}$ and $\tilde{\boldsymbol{x}}^{(i)}$ with probability $\min\{1, r(\tilde{\boldsymbol{x}}^{(i)}, \boldsymbol{x}^{(i-1)})\}$.

③ Sampling $\boldsymbol{\theta}$

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

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

—Simulate $\beta_g^{(i)}$ as follows:

- Generate $\tilde{\beta}_g^{(i)} \backsim q(\beta_g \mid \beta_g^{(i-1)})$
- Compute the ratio

$$r(\tilde{\beta}_g^{(i)}, \beta_g^{(i-1)}) = \frac{f(\tilde{\beta}_g^{(i)}|\gamma_g^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y}) q(\beta_g^{(i-1)} \mid \tilde{\beta}_g^{(i)})}{f(\beta_g^{(i-1)}|\gamma_g^{(i)}, \boldsymbol{x}^{(i)}, \boldsymbol{y}) q(\tilde{\beta}_g^{(i)} \mid \beta_g^{(i-1)})}$$

and accept the proposed candidate with the probability $\min\{1, r(\tilde{\beta}_g^{(i)}, \beta_g^{(i-1)})\}$.

end for

④ Set $i \leftarrow i + 1$ and goto ② until convergence.

V. Simulation Results

A. Validation Experiments

1) Example 1: To show the effectiveness of our algorithm, a first set of experiments is carried out on synthetic images. As a FR, we use 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. The ℓ_2 norm is used for $N(\cdot)$ in (3) with $\delta = 10^{-4}$. To generate a synthetic image (of size 128×128), we synthesize wavelet frame coefficients $\boldsymbol{x}$ from known prior distributions.

Let $\boldsymbol{x}_1 = (a_1, (h_{1,j}, v_{1,j}, d_{1,j})_{1\le j\le 2})$ and $\boldsymbol{x}_2 = (a_2, (h_{2,j}, v_{2,j}, d_{2,j})_{1\le j\le 2})$ be the sequences of wavelet basis coefficients generated in $\mathcal{B}_1$ and $\mathcal{B}_2$, where a, h, v, d stand for approximation, horizontal, vertical and diagonal coefficients and the index j is the resolution level. Wavelet frame coefficients are generated from a GG distribution in accordance with the chosen priors. The coefficients in each subband are modeled with the same values of the hyperparameters α_g and β_g, which means that each subband forms a group of index g. The number of groups (i.e., the number of subbands) G is therefore equal to 14. A uniform prior distribution over $[0, 3]$ is chosen for parameter β_g whereas a Jeffrey's prior is assigned to each parameter γ_g. For each group, the hyperparameters β_g and γ_g are first generated from a uniform prior distribution over $[0, 3]$ and a beta distribution, respectively. Drawing the hyperparameters from different distributions than the priors allows us to evaluate the robustness of our approach to modeling errors. A set of frame coefficients is then randomly generated to synthesize the observed data. The hyperparameters are then supposed unknown, sampled using the proposed algorithm, and estimated based on the generated samples by:

(i) computing the mean according to the MMSE principle;
(ii) computing the MAP estimate.

Having reference values, the normalized mean square erors (NMSEs) related to the estimation of each hyperparameter belonging to a given group (here a given subband) are computed

TABLE I
NMSEs for the Estimated Hyperparameters Using the MMSE and MAP Estimators

	MMSE				MAP			
	Sampler 1		Sampler 2		Sampler 1		Sampler 2	
	β	α	β	α	β	α	β	α
$h_{1,1}$	0.019	0.016	0.012	0.030	0.025	0.021	0.013	0.039
$v_{1,1}$	0.022	0.021	0.022	0.026	0.029	0.032	0.034	0.051
$d_{1,1}$	0.007	0.030	0.011	0.044	0.013	0.037	0.025	0.051
$h_{1,2}$	0.042	0.044	0.021	0.026	0.055	0.051	0.033	0.037
$v_{1,2}$	0.011	0.018	0.020	0.019	0.021	0.027	0.031	0.022
$d_{1,2}$	0.009	0.012	0.023	0.041	0.017	0.020	0.024	0.038
a_1	0.040	0.043	0.039	0.023	0.046	0.050	0.052	0.034
$h_{2,1}$	0.036	0.043	0.015	0.025	0.045	0.051	0.019	0.038
$v_{2,1}$	0.041	0.057	0.025	0.031	0.049	0.056	0.034	0.042
$d_{2,1}$	0.008	0.017	0.029	0.023	0.021	0.026	0.037	0.035
$h_{2,2}$	0.019	0.021	0.016	0.034	0.025	0.029	0.024	0.041
$v_{2,2}$	0.011	0.009	0.013	0.022	0.020	0.015	0.019	0.030
$d_{2,2}$	0.018	0.019	0.011	0.040	0.023	0.027	0.019	0.041
a_2	0.025	0.031	0.010	0.028	0.033	0.038	0.017	0.032

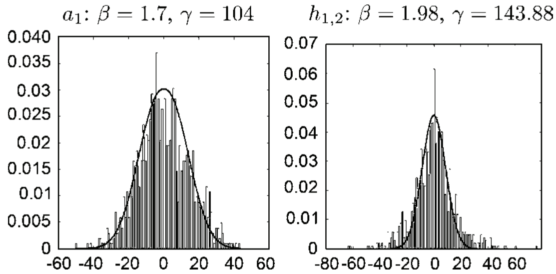

Fig. 1. Examples of empirical approximation (left) and detail (right) histograms and pdfs of frame coefficients corresponding to a synthetic image.

from 30 Monte Carlo runs. The NMSEs computed for the estimators associated with the two samplers of Sections IV-A and IV-B are reported in Table I. Table I shows that the proposed algorithms (using Sampler 1 of Section IV-A and Sampler 2 of Section IV-B) provide accurate estimates of the hyperparameters using the MMSE or the MAP estimator (with a slightly better performance for the MMSE estimator). The two samplers perform similarly for this experiment. However, one advantage of Sampler 2 is that it can be applied to different kinds of redundant frames, unlike Sampler 1. Indeed, as reported in Section IV-A, the conditional distribution (13) is generally difficult to sample when the FR is not a union of orthonormal bases.

To further illustrate the good performance of the proposed estimator, Fig. 1 shows two examples of empirical histograms of wavelet frame coefficients (corresponding to $\mathcal{B}_1$) that are in good agreement with the corresponding pdfs obtained after replacing the hyperparameters by their estimates.

2) Example 2: In this experiment, another FR is considered, namely a tight frame version of the translation invariant wavelet transform [50] with Daubechies filters of length 8. The ℓ_2 norm is also used for $N(\cdot)$ in (3) with $\delta = 10^{-4}$. We use the same process to generate frame coefficients as for Example 1. The coefficients in each subband (i.e., each group) are modeled with the same values of the hyperparameters γ_g and β_g, the number of groups being equal to 7. The same priors for the hyperparameters γ_g and β_g as for Example 1 are used.

After generating the hyperparameters and frame coefficients, the hyperparameters are then sampled using the proposed algorithm, and estimated using the MMSE estimator. Table II shows NMSEs based on reference values of each hyperparameter, where the frame coefficient vector is denoted by $\boldsymbol{x} = (a, (h_j, v_j, d_j)_{1 \le j \le 2})$. Note that Sampler 1 is difficult to be implemented in this case because of the used frame properties. Consequently, only NMSE values for Sampler 2 have been reported in Table II.

TABLE II
NMSEs for the Estimated Hyperparameters Using the MMSE and MAP Estimators With Sampler 2

	MMSE		MAP	
	β	α	β	α
h_1	0.050	0.027	0.056	0.035
v_1	0.024	0.007	0.029	0.011
d_1	0.050	0.014	0.051	0.021
h_2	0.037	0.028	0.044	0.033
v_2	0.051	0.044	0.057	0.050
d_2	0.040	0.012	0.043	0.021
a	0.040	0.050	0.046	0.055

TABLE III
NMSEs for the Estimated Hyperparameters Using the MMSE and MAP Estimates With Sampler 2

	MMSE		MAP	
	β	α	β	α
SB_1	0.007	0.027	0.0120	0.071
SB_2	0.002	0.032	0.011	0.056
SB_3	0.004	0.011	0.009	0.023
SB_4	0.001	0.018	0.008	0.022
SB_5	0.001	0.006	0.006	0.012
SB_6	0.010	0.040	0.028	0.048
SB_7	0.009	0.020	0.018	0.07
SB_8	0.002	0.021	0.009	0.033

3) Example 3: A third frame is considered in this experiment to show the versatility of our approach with respect to the choice of the FR: the contourlet transform [12] with Ladder filters over two resolution levels. The ℓ_∞ norm is used for $N(\cdot)$ in (3) with $\delta = 10^{-4}$. We use the same procedure to generate frame coefficients as for Examples 1 and 2. The coefficients in each of the eight groups are modeled with the same values of the hyperparameters γ_g and β_g and the same hyperparameter priors. After generating the hyperparameters and frame coefficients, the hyperparameters are then supposed unknown and estimated using the MMSE estimator based on samples drawn with Sampler 2. Table III shows NMSEs based on reference values of each hyperparameter.

B. Convergence Results

To be able to automatically stop the simulated chain and ensure that the last simulated samples are appropriately distributed according to the posterior distribution of interest, a convergence monitoring technique based on the potential scale reduction factor (PSRF) is used by simulating several chains in parallel (see [51] for more details). This convergence monitoring technique indicates that sample convergence arises as soon as PSRF < 1.2. Using the union of two orthonormal bases as a FR, Figs. 2 and 3 illustrate the variations w.r.t. the iteration number of the NMSE between the MMSE estimator and a reference estimator (computed by using a large number of burn-in and computation iterations, so as to guarantee that convergence has been achieved). The NMSE plots show that

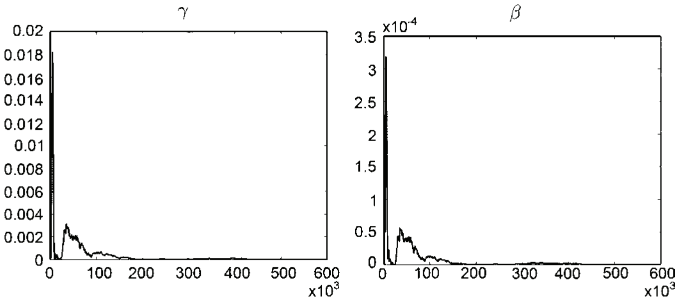

Fig. 2. NMSE between the reference and current MMSE estimators w.r.t. iteration number corresponding to $v_{1,1}$ in $\mathcal{B}_1$.

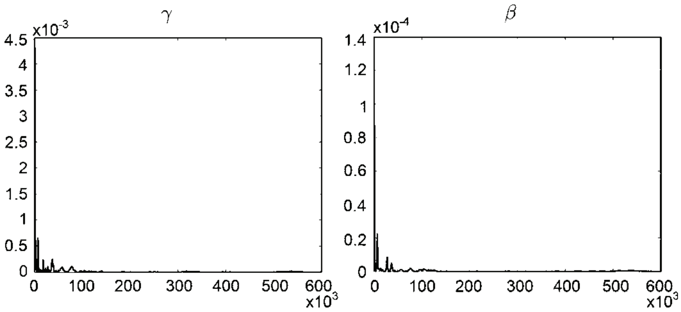

Fig. 3. NMSE between the reference and current MMSE estimators w.r.t. iteration number corresponding to $v_{2,2}$ in $\mathcal{B}_2$.

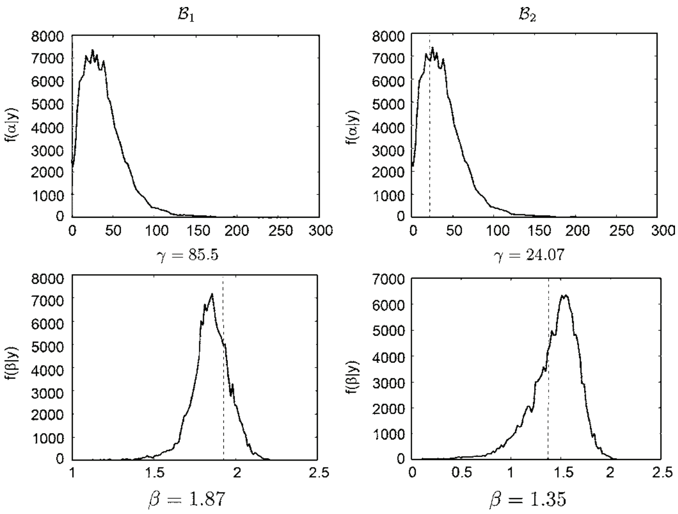

Fig. 4. Ground truth values (dashed line) and posterior distributions (solid line) of the sampled hyperparameters γ and β, for the subbands $h_{1,2}$ and $h_{2,2}$ in $\mathcal{B}_1$ and $\mathcal{B}_2$, repectively.

convergence is reached after about 150,000 iterations (burn-in period of 100,000 iterations), which corresponds to about 4 hours of computational time using Matlab 7.7 on an Intel Core 4–3 GHz architecture. When comparing the two proposed samplers in terms of convergence speed, it turns out from our simulations that Sampler 1 shows faster convergence than Sampler 2. Indeed, Sampler 1 needs about 110,000 iterations to converge, which reduces the global computational time to about 3 hours.

The posterior distributions of the hyperparameters β and γ related to the subbands $h_{1,2}$ and $h_{2,2}$ in $\mathcal{B}_1$ and $\mathcal{B}_2$ are shown in Fig. 4, as well as the known original values. It is clear that the modes of the posterior distributions are around the ground truth value, which confirms the good estimation performance of the proposed approach.

Note that when the resolution level increases, the number of subbands also increases, which leads to a higher number of hyperparameters to be estimated and a potential increase of the required computational time to reach convergence. For example, when using the union of two orthonormal wavelet bases with two resolution levels, the number of hyperparameters to estimate is $G = 28$.

C. Application to Image Denoising

1) Example 1: In this experiment, we are interested in recovering an image (the *Boat* image of size 256 × 256 coded at 8 bpp) from its noisy observation affected by a noise $\boldsymbol{n}$ uniformly distributed over the ball $[-\delta, \delta]^{256\times 256}$ with $\delta = 30$. We recall that the observation model for this image denoising problem is given by (3). The noisy image in Fig. 5(b) is simulated using the available reference image $\boldsymbol{y}_{\mathrm{ref}}$ in Fig. 5(a) and the noise properties described above.

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 (as for validation experiments in Section V-A) is used as a tight FR. Denoising is performed using the MMSE estimator denoted as $\hat{\boldsymbol{x}}$ computed from sampled wavelet frame coefficients. The adjoint frame operator is then applied to recover the denoised image from its denoised estimated wavelet frame coefficients ($\hat{\boldsymbol{y}} = F^*\hat{\boldsymbol{x}}$). The obtained denoised image is depicted in Fig. 5(d). For comparison purpose, the denoised image using a variational approach [52], [53] based on a MAP criterion using the estimated values of the hyperparameters with our approach is illustrated in Fig. 5(c). This comparison shows that, for denoising purposes, the proposed method gives better visual quality than the other reported methods. Signal-to-noise ratio ($\mathrm{SNR} = 20\log_{10}\left(\|\boldsymbol{y}_{\mathrm{ref}}\|/\|\boldsymbol{y}_{\mathrm{ref}} - \hat{\boldsymbol{y}}\|\right)$) and structural similarity (SSIM) [54] values are also given in Table IV to quantitatively evaluate denoising performance. Additional comparisons with respect to Wiener filtering and the algorithm developed in [36] (denoted here by SLR) are given in this table. Note that SLR can be applied only when the employed frame is the union of orthonormal bases, while our approach remains valid for any FR. Note also that SLR and Wiener filtering are basically designed to deal with Gaussian noise. This comparison shows that assuming the right noise model is essential to achieve good denoising performance. On the other hand, comparisons with the variational approach, which accounts for the right uniform noise model and uses the same FR and coefficient groups, show that the improvement achieved by our algorithm is not only due to the model choice. The SNR and SSIM values are given for two additional test images (*Sebal* and *Tree*) with different textures and contents to better illustrate the good performance of the proposed approach and its robustness to model mismatch. The corresponding original, noisy and denoised images are displayed in Figs. 6 and 7.

It is worth noticing that the visual quality and quantitative results show that the denoised image based on the MMSE estimator of the wavelet frame coefficients is better than the one obtained with the Wiener filtering or the variational approach.

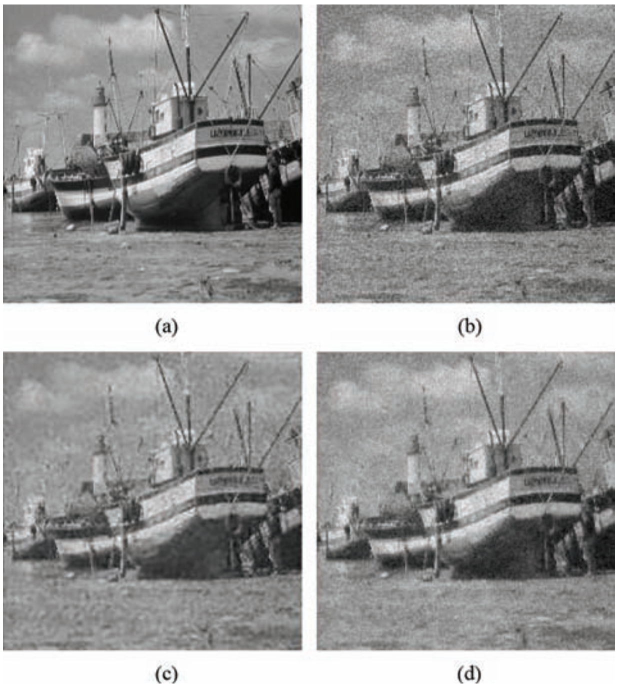

Fig. 5. Original *Boat* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

TABLE IV
SNR AND SSIM VALUES FOR THE NOISY AND DENOISED IMAGES

		Noisy	Wiener	Variational	SLR	MCMC
	SNR	16.67	18.02	18.41	18.40	19.20
Boat	SSIM	0.521	0.553	0.570	0.563	0.614
	SNR	13.85	14.40	15.04	14.98	15.69
Sebal	SSIM	0.642	0.695	0.704	0.697	0.701
	SNR	17.19	19.27	19.29	19.38	19.82
Tree	SSIM	0.662	0.768	0.765	0.776	0.785

For the latter approach, it must be emphasized that the choice of the hyperparameters always constitute a delicate problem, for which our algorithm brings a numerical solution. Note also that compared with the variational approach, our algorithm recovers sharper and better denoised edges. However, our approach seems to be less performant in smooth regions, even if it does not introduce blurring effects like the variational approach.

In contrast with Wiener filtering and the variational approach which are very fast, SLR and our approach are more time-consuming. Table V gives the iteration numbers and computational times for the used methods on an Intel Core 4–3 GHz architecture using a Matlab implementation. However, a high gain in computational time can be expected through code optimization and parallel implementation using multiple CPU cores. In fact, since the frame coefficients are split into G groups with a couple of hyperparameters for each of them, a high number of loops is required, which is detrimental to the computational time in a Matlab implementation.

2) Example 2: In this experiment, we are interested in recovering an image (the *Straw* image of size 128×128 coded at 8 bpp) from its noisy observation affected by a noise $\boldsymbol{n}$ uniformly distributed over the centered ℓ_p ball of radius δ

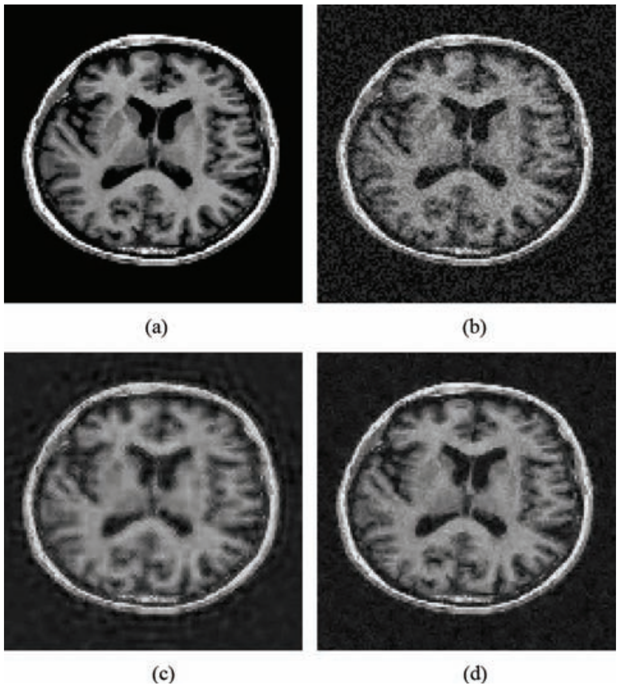

Fig. 6. Original 128×128 *Sebal* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

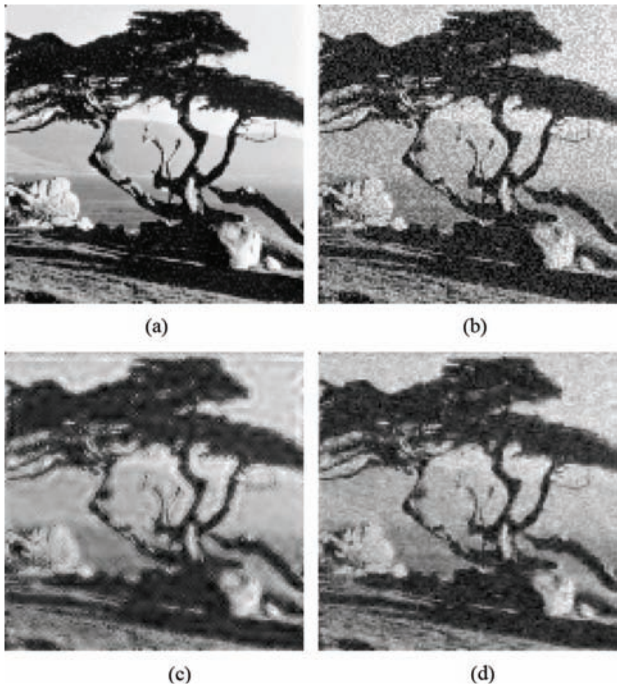

Fig. 7. Original 128×128 *Tree* image (a), noisy image (b), denoised images using a variational approach (c) and the proposed MMSE estimator (d).

TABLE V
COMPUTATIONAL TIME (IN MINUTES) FOR THE USED METHODS

	Wiener	Variational	SLR	MCMC
Iterations	1	100	100,000	150,000
Computational time	0.002	3	60	130

when $p \in \{1, 2, 3\}$. Experiments are conducted using two different FRs: the translation invariant wavelet transform with a

TABLE VI
SNR AND SSIM VALUES FOR THE NOISY AND DENOISED *STRAW* IMAGES

				translation invariant wavelet		contourlet	
		Noisy	Wiener	Variational	MCMC	Variational	MCMC
$\delta = 300000$	SNR (dB)	15.56	16.42	16.67	18.11	17.76	18.79
$p = 1$	SSIM	0.719	0.705	0.730	0.755	0.678	0.803
$\delta = 3000$	SNR (dB)	16.46	17.03	17.84	19.02	18.61	19.21
$p = 2$	SSIM	0.749	0.720	0.758	0.796	0.719	0.808
$\delta = 700$	SNR (dB)	16.14	17.05	17.65	19.29	18.28	19.44
$p = 3$	SSIM	0.734	0.720	0.671	0.771	0.698	0.788

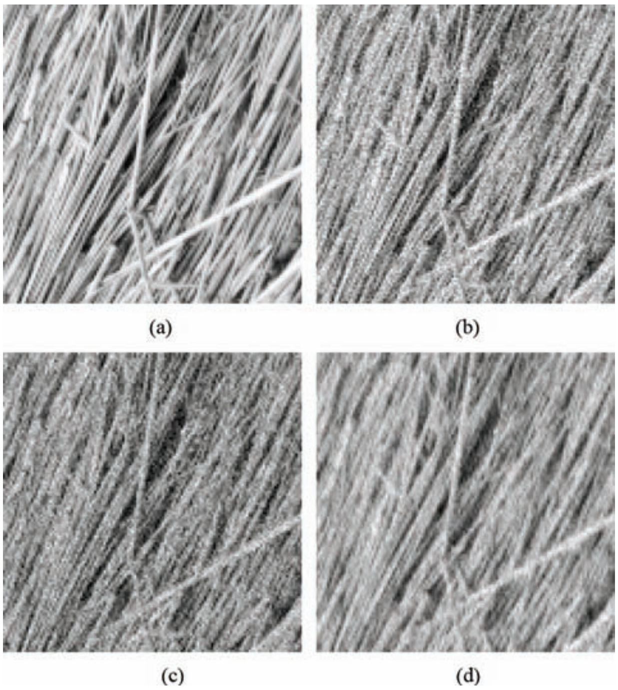

Fig. 8. Original *Straw* image (a), noisy image (b) and denoised images using the variational approach (c) and the proposed MMSE estimator (d).

Symmlet filter of length 8 and the contourlet transform with Ladder filters, both over 3 resolution levels. The ℓ_p norm $(p \in \{1,2,3\})$ was used for $N(\cdot)$ in (3). Figs. 8(a) and 8(b) show the original and noisy images using a uniform noise over the ℓ_2 ball of radius 3000. When using the translation invariant wavelet transform, Figs. 8(c) and 8(d) illustrate the results generated by the denoising strategies based on the variational approach and the MMSE estimator using frame coefficients sampled with our algorithm. Table VI shows the SNR and SSIM values for noisy and denoised images using the proposed MMSE estimator for different values of p and δ. This second set of image denoising experiments shows that the proposed approach performs well when using different kinds of FRs and various noise properties, which emphasizes its robustness to modelling errors.

VI. CONCLUSION

This paper proposed a hierarchical Bayesian model for frame coefficient from a noisy observation of a signal or image of interest. The signal perturbation was modeled by introducing a bound on a distance between the signal and its observation. A hierarchical model based on this maximum distance property was then defined. This model assumed flexible GG priors for the frame coefficients. Vague priors were assigned to the hyperparameters associated with the frame coefficient priors. Different sampling strategies were proposed to generate samples distributed according to the joint distribution of the parameters and hyperparameters of the resulting Bayesian model. The generated samples were finally used for estimation purposes. Our validation experiments indicated that the proposed algorithms provide an accurate estimation of the frame coefficients and hyperparameters. The good quality of the estimates was confirmed on statistical processing problems in image denoising. Clearly, the proposed Bayesian approach outperforms the other methods because it allows us to use the right noise model and an appropriate frame coefficient prior. The numerous experiments which were conducted also showed that the proposed algorithm is robust to model mismatch. The hierarchical model studied in this paper assumed GG priors for the frame coefficients. However, the proposed algorithm might be generalized to other classes of prior models. Another direction of research for future work would be to extend the proposed framework to situations where the observed signal is degraded by a linear operator.

APPENDIX
SAMPLING ON THE UNIT ℓ_p BALL

This appendix explains how to sample vectors in the unit ℓ_p ball ($p \in]0,+\infty]$) of $\mathbb{R}^L$. First, it is interesting to note that sampling on the unit ball can be easily performed in the particular case $p = +\infty$, by sampling independently along each space coordinate according to a distribution on the interval $[-1,1]$. Thus, this appendix focuses on the more difficult problem associated with a finite value of p. In the following, $\|\cdot\|_p$ denotes the ℓ_p norm. We recall the following theorem:

Theorem A.1 [55]: Let $\boldsymbol{A} = [A_1,\ldots,A_{L'}]^\top$ be the random vector of i.i.d. components which have the following $\mathrm{GG}(p^{1/p},p)$ pdf

$$\forall a_i \in \mathbb{R}, \quad f(a_i) = \frac{p^{1-1/p}}{2\Gamma\left(\frac{1}{p}\right)} \exp\left(-\frac{|a_i|^p}{p}\right). \tag{24}$$

Let $\boldsymbol{U} = [U_1,\ldots,U_{L'}]^\top = \boldsymbol{A}/\|\boldsymbol{A}\|_p$. Then, the random vector $\boldsymbol{U}$ is uniformly distributed on the surface of the ℓ_p unit sphere of $\mathbb{R}^{L'}$ and the joint pdf of $U_1,\ldots,U_{L'-1}$ is

$$f(\boldsymbol{u}) = \frac{p^{L'-1}\Gamma\left(\frac{L'}{p}\right)}{2^{L'-1}\left(\Gamma\left(\frac{1}{p}\right)\right)^{L'}} \times \left(1 - \sum_{k=1}^{L'-1} |u_k|^p\right)^{(1-p)/p} 1_{D_{p,L'}}(\boldsymbol{u}) \tag{25}$$

where $\boldsymbol{u} = (u_1, \ldots, u_{L'-1})$ and $D_{p,L'} = \{\boldsymbol{u} \in \mathbb{R}^{L'-1} \mid \sum_{k=1}^{L'-1} |u_k|^p < 1\}$.

The uniform distribution on the unit ℓ_p sphere of $\mathbb{R}^{L'}$ will be denoted by $\mathcal{U}(L', p)$. The construction of a random vector distributed within the ℓ_p ball of $\mathbb{R}^L$ with $L < L'$ can be derived from Theorem A.1 as expressed.

Theorem A.2 [55]: Let $\boldsymbol{U} = [U_1, \ldots, U_{L'}]^\top \sim \mathcal{U}(L', p)$. For every $L \in \{1, \ldots, L'-1\}$, the pdf of $\boldsymbol{V} = [U_1, \ldots, U_L]^\top$ is given by

$$q_1(\boldsymbol{v}) = \frac{p^L \Gamma\left(\frac{L'}{p}\right)\left(1 - \sum_{k=1}^{L} |u_k|^p\right)^{(L'-L)/p-1}}{2^L \left(\Gamma\left(\frac{1}{p}\right)\right)^L \Gamma\left(\frac{(L'-L)}{p}\right)} 1_{D_{p,L+1}}(\boldsymbol{v}). \tag{26}$$

In particular, if $p \in \mathbb{N}^*$ and $L' = L + p$, we obtain the uniform distribution on the unit ℓ_p ball of $\mathbb{R}^L$.

Sampling from a distribution q_η on an ℓ_p ball of radius $\eta > 0$ is straightforwardly deduced by scaling $\boldsymbol{V}$.

ACKNOWLEDGMENT

One of the authors (J.-C. Pesquet) would like to thank Prof. C. Vignat for some fruitful discussions.

REFERENCES

[1] H. Kawahara, "Signal reconstruction from modified auditory wavelet transform," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3549–3554, Dec. 1993.

[2] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu, "Autocalibrated parallel MRI reconstruction in the wavelet domain," in *IEEE Int. Symp. on Biomed. Imag. (ISBI)*, Paris, France, May 14–17, 2008, pp. 756–759.

[3] E. L. Miller, "Efficient computational methods for wavelet domain signal restoration problems," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 1184–1188, Apr. 1999.

[4] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. Wajs, "A forward-backward algorithm for image restoration with sparse representations," in *Signal Process. Adaptat. Sparse Structured Represent.*, Rennes, France, Nov. 2005, pp. 49–52.

[5] M. B. Martin and A. E. Bell, "New image compression techniques using multiwavelets and multiwavelet packets," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 500–510, Apr. 2001.

[6] M. De Meuleneire, "Algebraic quantization of transform coefficients for embedded audio coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal (ICASSP)*, Las Vegas, NV, Apr. 4, 2008, pp. 4789–4792.

[7] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1999.

[8] R. Coifman, Y. Meyer, and V. Wickerhauser, "Wavelet analysis and signal processing," in *Wavelets and Their Applications*. Boston, MA: Jones & Bartlett, 1992, pp. 153–178.

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

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

[11] 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, Oct. 2006.

[12] M. N. Do and M. Vetterli, "The contourlet transform: An efficient directional multiresolution image representation," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2091–2106, Dec. 2005.

[13] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. Image Process.*, vol. 14, no. 4, pp. 423–438, Apr. 2005.

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

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

[16] A. B. Watson, G. Y. Yang, J. A. Solomon, and J. Villasenor, "Visibility of wavelet quantization noise," *IEEE Trans. Image Process.*, vol. 6, no. 8, pp. 1164–1175, Aug. 1997.

[17] Y. Zhang, A. O. Hero, and W. L. Rogers, "A bounded error estimation approach to image reconstruction," in *Proc. Nucl. Sci. Symp. Med. Imag. Conf.*, Orlando, FL, Oct. 1992, pp. 966–968.

[18] S. Rangan and V. K. Goyal, "Recursive consistent estimation with bounded noise," *IEEE Trans. Inf. Theory*, vol. 47, no. 1, pp. 457–464, Jan. 2001.

[19] T. M. Breuel, "Finding lines under bounded error," *Pattern Recogn.*, vol. 29, no. 1, pp. 167–178, Jan. 1996.

[20] M. Alghoniemy and A. H. Tewfik, "A sparse solution to the bounded subset selection problem: A network flow model approach," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal (ICASSP)*, Montreal, Canada, May 17–21, 2004, pp. 89–92.

[21] C. Robert and G. Castella, *Monte Carlo Statistical Methods*. New York: Springer, 2004.

[22] O. Cappé, "A Bayesian approach for simultaneous segmentation and classification of count data," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 400–410, Feb. 2002.

[23] C. Andrieu, P. M. Djuric, and A. Doucet, "Model selection by MCMC computation," *Signal Process.*, vol. 81, pp. 19–37, Jan. 2001.

[24] M. Ichir and A. Mohammad-Djafari, "Wavelet domain blind image separation," in *Proc. SPIE Tech. Conf. Wavelet Appl. Signal and Image Process. X*, San Diego, CA, Aug. 2003.

[25] A. Jalobeanu, L. Blanc-Féraud, and J. Zerubia, "Hyperparameter estimation for satellite image restoration using a MCMC maximum likelihood method," *Pattern Recogn.*, vol. 35, no. 2, pp. 341–352, Nov. 2002.

[26] S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline, "Joint detection-estimation of brain activity in functional MRI: A multichannel deconvolution solution," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3488–3502, Sep. 2005.

[27] 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, Sep. 2009.

[28] 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, Sep. 2007.

[29] S. Zeger and R. Karim, "Generalized linear models with random effects a Gibbs sampling approach," *J. Amer. Statist. Assoc.*, vol. 86, no. 413, pp. 79–86, Mar. 1991.

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

[31] W. K. Hastings, "Monte Carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[32] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, 1984.

[33] D. Leporini and J.-C. Pesquet, "Bayesian wavelet denoising: Besov priors and non-Gaussian noises," *Signal Process.*, vol. 81, no. 1, pp. 55–67, Jan. 2001.

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

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

[36] C. Févotte and S. J. Godsill, "Sparse linear regression in unions of bases via Bayesian variable selection," *IEEE Signal Process. Lett.*, vol. 13, no. 7, pp. 441–444, Jul. 2006.

[37] P. J. Wolfe and S. J. Godsill, "Bayesian estimation of time-frequency coefficients for audio signal enhancement," in *Adv. Neural Inf. Process. Syst.*, 2003, vol. 15, pp. 1197–1204.

[38] M. Kowalski and B. Torrésani, "Random models for sparse signals expansion on unions of bases with application to audio signals," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3468–3481, Aug. 2008.

[39] S. Molla and B. Torrésani, "A hybrid scheme for encoding audio signal using hidden Markov models of waveforms," *Appl. Comput. Harmon. Anal.*, vol. 18, no. 2, pp. 137–166, Mar. 2005.

[40] D. L. Donoho, "Denoising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[41] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[42] R. L. Joshi, H. Jafarkhani, J. H. Kasner, T. R. Fischer, N. Farvardin, M. W. Marcellin, and R. H. Bamberger, "Comparison of different methods of classification in subband coding of images," *IEEE Trans. Image Process.*, vol. 6, no. 11, pp. 1473–1486, Nov. 1997.

[43] E. P. Simoncelli and E. H. Adelson, "Noise removal via Bayesian wavelet coring," in *IEEE Int. Conf. on Image Process. (ICIP)*, Lausanne, Switzerland, Sep. 16–19, 1996, pp. 379–382.

[44] P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes using generalized-Gaussian priors," in *Proc. IEEE-SP Int. Symp. Time-Frequency Time-Scale Analysis*, Pittsburgh, PA, Oct. 1998, pp. 633–636.

[45] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans. Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[46] M. W. Seeger and H. Nickisch, "Compressed sensing and bayesian experimental design," in *Int. Conf. Mach. Learn.*, Helsinki, Finland, Jul. 5–9, 2008, pp. 912–919.

[47] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Fast Bayesian compressive sensing using Laplace priors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal (ICASSP)*, Taipei, Taiwan, Apr. 19–24, 2009, pp. 2873–2876.

[48] H. Jeffreys, "An invariant form for the prior probability in estimation problems," *Proc. Royal Soc. London. Series A*, vol. 186, no. 1007, pp. 453–461, 1946.

[49] N. Dobigeon and J.-Y. Tourneret, "Truncated multivariate Gaussian distribution on a simplex," Univ. Toulouse, France, Tech. rep., Jan. 2007.

[50] R. Coifman and D. Donoho, "Translation-invariant de-noising," in *Wavelets and Statistics*. New York: Springer, 1995, vol. 103, Lecture Notes in Statistics, pp. 125–150.

[51] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statist. Sci.*, vol. 7, no. 4, pp. 457–472, Nov. 1992.

[52] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, "A variational formulation for frame-based inverse problems," *Inv. Problems*, vol. 23, no. 4, pp. 1495–1518, Aug. 2007.

[53] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inv. Problems*, vol. 24, no. 4, p. 27, Nov. 2008.

[54] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[55] A. K. Gupta and D. Song, "Lp-norm uniform distribution," *J. Statist. Plann. Inference*, vol. 125, no. 2, pp. 595–601, Feb. 1997.

Lotfi Chaâri (S'08) was born in Sfax, Tunisia, in 1983. He received the engineering and master degrees in telecommunication from the École Supérieure des Communications (SUP'COM), Tunis, Tunisia in 2007.

He is currently pursuing the Ph.D. degree in 2008 with Professor J.-C. Pesquet at the Laboratoire d'Informatique (UMR-CNRS 8049) of the Université Paris-Est Marne-la-Vallée. His research is focused on Bayesian and variational approaches for image processing and their application to parallel MRI reconstruction.

Jean-Christophe Pesquet (S'89–M'91–SM'99) received the engineering degree from Supélec, Gif-sur-Yvette, France, in 1987, the Ph.D. degree from the Université Paris-Sud (XI), Paris, France, in 1990, and the Habilitation à Diriger des Recherches from the Université Paris-Sud in 1999.

From 1991 to 1999, he was a Maître de Conférences at the Université Paris-Sud, and a Research Scientist at the Laboratoire des Signaux et Systèmes, Centre National de la Recherche Scientifique (CNRS), Gif sur Yvette. He is currently a Professor with Université Paris-Est Marne-la-Vallée, France, and a Research Scientist with the Laboratoire d'Informatique of the university (UMR-CNRS 8049).

Jean-Yves Tourneret (SM'08) received the ingénieur degree in electrical engineering from the École Nationale Supérieure d'Électronique, d'Électrotechnique, d'Informatique et d'Hydraulique of Toulouse (ENSEEIHT), France, in 1989 and the Ph.D. degree from the National Polytechnic Institute, Toulouse, in 1992.

He is currently a professor with the University of Toulouse (ENSEEIHT), France and a member of the IRIT laboratory (UMR 5505 of the CNRS). His research activities are centered around statistical signal processing with a particular interest to Markov chain Monte Carlo methods.

Dr. Tourneret was the Program Chair of the European Conference on Signal Processing (EUSIPCO), held in Toulouse in 2002. He was also member of the Organizing Committee for the International Conference ICASSP'06 held in Toulouse in 2006. He has been a member of different technical committees including the Signal Processing Theory and Methods (SPTM) Committee of the IEEE Signal Processing Society (2001–2007 and 2010-present). He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Philippe Ciuciu (M'02) was born in France in 1973. He graduated from the École Supérieure d'Informatique Électronique Automatique, Paris, France, in 1996. He received the DEA and Ph.D. degrees in signal processing from the Université de Paris-sud, Orsay, France, in 1996 and 2000, respectively.

In November 2000, he joined the fMRI Signal Processing Group as a Postdoctoral Fellow with Service Hospitalier Frédéric Joliot (CEA, Life Science Division, Medical Research Department) in Orsay, France. Since November 2001, he has been a permanent CEA Research Scientist. In 2007, he moved to NeuroSpin, an ultrahigh magnetic field center in the LNAO lab. Currently, he is Principal Investigator of the Neurodynamics Program in collaboration with A. Kleinschmidt (DR INSERM) at the Cognitive Neuroimaging Unit (INSERM/CEA U562). His research is focused on the application of statistical methods (e.g., Bayesian inference, model selection), stochastic agorithms (MCMC), and wavelet-based regularized approaches to functional brain imaging (fMRI) and to parallel MRI reconstruction.

Amel Benazza-Benyahia received the engineering degree from the National Institute of Telecommunications, Evry, France, in 1988, the Ph.D. degree from the Université Paris-Sud (XI), Paris, France, in 1993, and the Habilitation Universitaire from the Ecole Supérieure des Communications (SUP'COM), Tunis, Tunisia, in 2003.

She is currently a Professor with the Department of Applied Mathematics, Signal Processing and Communications, SUP'COM. She is also a Research Scientist with the Unité de Recherche en Imagerie Satellitaire et ses Applications (URISA), SUP'COM. Her research interests include multispectral image compression, signal denoising, and medical image analysis.