WAVELET BASED UNSUPERVISED VARIATIONAL BAYESIAN IMAGE RECONSTRUCTION APPROACH

Yuling Zheng[*], *Aurélia Fraysse*[†], *Thomas Rodet*[‡]

[*] LIGM, Université Paris-Est, CNRS, Champs-sur-Marne, 77454 Marne-la-Vallée, France
[†] L2S, Université Paris-Sud, CNRS, Supelec, R. Joliot-Curie, 91190 Gif-sur-Yvette, France
[‡] SATIE, ENS Cachan, CNAM, Av. du Prés. Wilson, 94230 Cachan, France
yuling.zheng@univ-mlv.fr, fraysse@lss.supelec.fr, trodet@satie.ens-cachan.fr

ABSTRACT

In this paper, we present a variational Bayesian approach in the wavelet domain for linear image reconstruction problems. This approach is based on a Gaussian Scale Mixture prior and an improved variational Bayesian approximation method. Its main advantages are that it is unsupervised and can be used to solve various linear inverse problems. We show the good performance of our approach through comparisons with state of the art approaches on a deconvolution problem.

***Index Terms*—** unsupervised approach, wavelet transform, variational Bayesian, GSM, Generalized Gaussian

1. INTRODUCTION

Most image reconstruction tasks can be considered as ill-posed linear inverse problems. The resolution of such problems generally relies on the introduction of additional information thanks to regularization terms or prior distributions either in the spatial or in a transform domain (e.g. Fourier, wavelet, ...). It is popular to treat image reconstruction problems in the wavelet domain as it provides sparse representations for a large class of images.

As a result, sparsity prior information on wavelet coefficients can be introduced to regularize ill-posed inverse problems. To do this, L^1 norm regularization has been used extensively, e.g. [1]. However, a main difficulty encountered is a proper choice of hyperparameters which control the trade-off between data fidelity and regularization terms. Thanks to the work in [2], for denoising problems, hyperparameters can be determined by minimizing the Stein's unbiased risk estimate (SURE) [1]. SURE has also been generalized (GSURE) [3–6] for general linear inverse problems. Another way to solve this problem is to work in the Bayesian framework where we jointly estimate hyperparameters and unknown coefficients by assigning prior distributions to both of them [7].

In this paper, we choose to work in the Bayesian framework. To obtain a more accurate estimation, rather than using orthogonal wavelet transforms, we consider a dictionary decomposition over an union of wavelet bases [8]. Concerning the prior for unknown coefficients, we consider a Gaussian Scale Mixture (GSM) model [9,10] which encompasses many heavy tailed priors, e.g. Generalized Gaussian (GG), which has shown considerable success in wavelet domain image reconstructions [11, 12]. Moreover, in order to adapt sparsity degrees and relative importance of prior information to different subbands, we take different hyperparameters (shape and scale parameters) for GSM priors in different subbands.

Nevertheless, due to the introduction of hyperpriors, the involved posterior distribution is too complicated for the computation of classical estimators such as the Maximum *A Posterior* (MAP) and the Posterior Mean (PM). To tackle this problem efficiently, rather than using MCMC approaches [12], we resort to variational Bayesian methods [13–15] which provide an approximate posterior distribution of simpler form than the original one. The PM estimator can be more easily derived from this approximate distribution. Furthermore, to get a more efficient method, we adopt here an improved variational Bayesian algorithm recently proposed in [15] which can be well adapted to large dimensional problems.

The rest of this paper is organized as follows: we present the involved Bayesian model in Section 2; Section 3 is devoted to the introduction of our variational Bayesian approach whereas algorithm evaluation through simulation results on a deconvolution problem is given in Section 4; Finally, we draw our conclusions in Section 5.

2. BAYESIAN MODELING

The unknown image $\mathbf{x} \in \mathbb{R}^P$ can be represented through a dictionary expansion as $\mathbf{x} = \mathbf{D}\mathbf{u}$ where $\mathbf{D} \in \mathbb{R}^{P\times N}$ and $\mathbf{u} \in \mathbb{R}^N$ denotes the associated coefficients. In the case of an overcomplete dictionary, e.g. the union of several orthonormal bases, we have $N > P$. We consider in the following a linear forward model in the transformed domain:

$$\mathbf{y} = \mathbf{A}\mathbf{D}\mathbf{u} + \mathbf{n}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^M$ denotes the data, the operator $\mathbf{A} \in \mathbb{R}^{M\times P}$ is assumed to be known and $\mathbf{n}$ is a Gaussian white noise,

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \gamma_n^{-1}\mathbf{I})$, with γ_n as the inverse of the noise variance. Therefore $p(\mathbf{y}|\mathbf{u}, \gamma_n) = \mathcal{N}(\mathbf{ADu}, \gamma_n^{-1}\mathbf{I})$.

Concerning the prior for the coefficients $\mathbf{u}$, we consider a GSM. In fact, all the coefficients do not exhibit the same statistical characteristics. To take this fact into account, we divide the coefficients into L subbands and we assign GSM priors with different parameters to coefficients in different subbands. By using $(I_l)_{l=1,\dots,L}$ to denote sets of indices of coefficients in the l^{th} subband, the prior distribution of $\mathbf{u}$ can be written as

$$\begin{aligned} p(\mathbf{u}|\boldsymbol{\gamma}_p, \boldsymbol{\tau}) &= \prod_{l=1}^{L} \prod_{i \in I_l} p(u_i|\gamma_p^l, \boldsymbol{\tau}^l) \\ &= \prod_{l=1}^{L} \prod_{i \in I_l} \int_{\mathbb{R}} \mathcal{N}(u_i|0, (z_i\gamma_p^l)^{-1}) p(z_i|\boldsymbol{\tau}^l) \mathrm{d}z_i \end{aligned}$$

where $\mathbf{z} = (z_1, \dots, z_N)$ is the vector of hidden variables whose distributions are given by $p(z_i|\boldsymbol{\tau}^l)$ with parameters denoted by $\boldsymbol{\tau}^l$. We note that $p(z_i|\boldsymbol{\tau}^l)$ depends on the expression of $p(u_i|\gamma_p^l, \boldsymbol{\tau}^l)$. Moreover, in the above equation, $\boldsymbol{\gamma}_p = [\gamma_p^1, \dots, \gamma_p^L]$ are scale parameters of our GSM prior.

In total, three types of hyperparameters are involved in the above Bayesian formulation: γ_n, $\boldsymbol{\gamma}_p$ and $\boldsymbol{\tau} = (\boldsymbol{\tau}^1, \dots, \boldsymbol{\tau}^L)$. Generally, $\boldsymbol{\tau}$ are shape parameters of the GSM prior which determine the type of prior information introduced. As a result, we choose to fix $\boldsymbol{\tau}$ according to our prior knowledge. However, we estimate γ_n and $\boldsymbol{\gamma}_p$ since they determine a compromise between data fidelity and fidelity to the prior information.

For γ_n and $(\gamma_p^l)_{l=1,\dots,L}$, Jeffreys' non-informative priors are assigned. Using the Bayes' rule, we can derive the posterior distribution of the unknown parameters given data

$$\begin{aligned} p(\mathbf{u}, \mathbf{z}, \gamma_n, \boldsymbol{\gamma}_p|\mathbf{y}, \boldsymbol{\tau}) \propto {\gamma_n}^{M/2} \exp\left[-\frac{\gamma_n \|\mathbf{y} - \mathbf{ADu}\|^2}{2}\right] \\ \times \prod_{l=1}^{L} \prod_{i \in I_l} \sqrt{z_i \gamma_p^l} \exp\left[-\frac{\gamma_p^l}{2} z_i u_i^2\right] p(z_i|\boldsymbol{\tau}^l) \\ \times \gamma_n^{-1} \prod_{l=1}^{L} (\gamma_p^l)^{-1}. \end{aligned} \quad (2)$$

3. VARIATIONAL BAYESIAN APPROACHES

In the following, we introduce a variable $\Theta = \{\mathbf{u}, \mathbf{z}, \gamma_n, \boldsymbol{\gamma}_p\}$ which includes all the parameters to be estimated. The estimation of these parameters is based on the joint posterior distribution given by (2). However, this distribution is intractable since its partition function is difficult to calculate in practice. To tackle this problem, we resort to variational Bayesian approximations (VBA) which generate a separable approximation q_Θ of the true posterior distribution $p(\Theta|\mathbf{y}, \boldsymbol{\tau})$ by minimizing the Kullback-Leibler divergence between them. Assuming that $q_\Theta(\Theta) = \prod_i q_i(\Theta_i)$, classical VBA gives the following analytic solution (see [13] for details)

$$q_i(\Theta_i) \propto \exp\left(\langle \log p(\mathbf{y}, \Theta) \rangle_{\prod_{j \neq i} q_j(\Theta_j)}\right), \quad (3)$$

where $p(\mathbf{y}, \Theta)$ is the joint distribution which is explicitly known. As shown by (3), each distribution q_i depends on all the other distributions q_j with j different from i. In practice, this dependence implies the use of iterative methods such as the Gauss-Seidel one, which are not very efficient to iteratively approximate q_Θ. Recently in a prior work [14], an efficient exponentiated gradient based VBA method has been proposed. This method has been further developed in [15], leading to a more efficient Memory Gradient subspace based variational Bayesian approximation (MG-VBA) method. The MG-VBA integrates the subspace optimization principle and adopts the following updating equation:

$$\begin{aligned} q_i^{k+1}(\Theta_i) = & K^k(\mathbf{s}^k) q_i^k(\Theta_i) \left(\frac{\langle \log p(\mathbf{y}, \Theta) \rangle_{\prod_{j \neq i} q_j^k(\Theta_j)}}{q_i^k(\Theta_i)}\right)^{s_1^k} \\ & \times \left(\frac{q_i^k(\Theta_i)}{q_i^{k-1}(\Theta_i)}\right)^{s_2^k}, \end{aligned} \quad (4)$$

where $\mathbf{s}^k = [s_1^k, s_2^k]$ is the two-dimensional algorithm step size. In this work, we adopt the approximate optimal step size proposed in [15] thanks to the second order Taylor expansion of the objective criterion. We can see from (4) that q_i^{k+1} does not depend on q_j^{k+1} with j different from i, but depends on $\prod_j q_j^k$ which is known from the kth iteration. As a result, all the $(q_i^{k+1})_{i=1,\dots,N}$ are updated in parallel.

Concerning the separability assumption, we consider here a total separability given as follows

$$q_\Theta(\Theta) = \left(\prod_{l=1}^{L} \prod_{i \in I_l} q_{u_i}(u_i) q_{z_i}(z_i)\right) q_{\gamma_n}(\gamma_n) \prod_{l=1}^{L} q_{\gamma_p^l}(\gamma_p^l).$$

In fact, since $p(\mathbf{z}, \gamma_n, \boldsymbol{\gamma}_p|\mathbf{u}, \mathbf{y}, \boldsymbol{\tau})$ is separable, the classical VBA yields directly explicit solutions for $(q_{z_j})_{j=1,\dots,N}$, q_{γ_n} and $(q_{\gamma_p^l})_{l=1,\dots,L}$. Nevertheless, this is not the case for $(q_{u_i})_{i=1,\dots,N}$. Therefore, we adopt the MG-VBA for the optimization of $(q_{u_i})_{i=1,\dots,N}$.

3.1. Determination of q_{u_i}

Since a GSM prior is used, the conditional distribution $p(u_i|z_i, \gamma_p^l)$ is a Gaussian one, which is conjugate with the Gaussian likelihood $p(\mathbf{y}|\mathbf{u}, \gamma_n)$. Therefore, the optimal approximate distributions $(q_{u_i})_{i=1,\dots,N}$ belong to a Gaussian family. As a result, we take

$$q_{u_i}^k(u_i) = \mathcal{N}\left((\mathbf{m}_k)_i, (\boldsymbol{\sigma}_k^2)_i\right).$$

In this case, the optimization of $(q_{u_i})_{i=1,\dots,N}$ is performed by optimizing their parameters: mean $\mathbf{m}_k$ and variance $\boldsymbol{\sigma}_k^2$.

Using (4), the following update equations have been obtained (details can be found in [15].):

$$\boldsymbol{\sigma}_{k+1}^2 = \left[\frac{1}{\boldsymbol{\sigma}_k^2} + s_1\left(\frac{1}{\boldsymbol{\sigma}_r^2} - \frac{1}{\boldsymbol{\sigma}_k^2}\right) + s_2\left(\frac{1}{\boldsymbol{\sigma}_k^2} - \frac{1}{\boldsymbol{\sigma}_{k-1}^2}\right)\right]^{-1},$$

$$\mathbf{m}_{k+1} = \boldsymbol{\sigma}_{k+1}^2\left[\frac{\mathbf{m}_k}{\boldsymbol{\sigma}_k^2} + s_1\left(\frac{\mathbf{m}_r}{\boldsymbol{\sigma}_r^2} - \frac{\mathbf{m}_k}{\boldsymbol{\sigma}_k^2}\right) + s_2\left(\frac{\mathbf{m}_k}{\boldsymbol{\sigma}_k^2} - \frac{\mathbf{m}_{k-1}}{\boldsymbol{\sigma}_{k-1}^2}\right)\right].$$

In the above equations, we omit all the indications of vector component $(\cdot)_i$ for the sake of clarity. Moreover, $\boldsymbol{\sigma}_r^2$ and $\mathbf{m}_r$ are two intermediate variables updated using the two following equations:

$$\left(\boldsymbol{\sigma}_r^2\right)_i = \left[\langle\gamma_n\rangle^k \left(\mathbf{D}^T\mathbf{A}^T\mathbf{A}\mathbf{D}\right)_{(i,i)} + \left\langle\gamma_p^l\right\rangle^k \langle z_i\rangle^k\right]^{-1}, \quad (5)$$

$$\left(\mathbf{m}_r\right)_i = \left(\boldsymbol{\sigma}_r^2\right)_i \langle\gamma_n\rangle^k \left[\mathbf{D}^T\mathbf{A}^T\mathbf{y} - \mathbf{D}^T\mathbf{A}^T\mathbf{A}\mathbf{D}\mathbf{m}_k + \operatorname{diag}(\mathbf{D}^T\mathbf{A}^T\mathbf{A}\mathbf{D}) \circ \mathbf{m}_k\right]_i \quad (6)$$

where $\langle w\rangle^k = \mathbb{E}_{q_w^k}(w)$, $\operatorname{diag}(\mathbf{M})$ is a vector containing the diagonal elements of $\mathbf{M}$ and $\circ$ denotes the Hadamard product between two vectors. Actually, the variables $(\boldsymbol{\sigma}_r^2)_i$ and $(\mathbf{m}_r)_i$ are determined by the auxiliary function $q_i^r(\Theta_i) \propto \exp\left(\langle\log p(\mathbf{y},\Theta)\rangle_{\prod_{j\neq i} q_j^k(\Theta_j)}\right)$.

3.2. Determination of q_{z_i}

For the hidden variables $(z_i)_{i=1,\ldots,N}$, using (3), we can obtain

$$\begin{aligned} q_{z_i}^{k+1}(z_i) &\propto \exp\left(\frac{1}{2}\ln(z_i) - \frac{\langle\gamma_p^l\rangle^k}{2} z_i \langle u_i^2\rangle^k + \ln p(z_i|\boldsymbol{\tau}^l)\right) \\ &\propto \sqrt{z_i}\, p(z_i|\boldsymbol{\tau}^l) \exp\left(-\frac{\langle\gamma_p^l\rangle^k}{2}\langle u_i^2\rangle^k z_i\right) \\ &= p(z_i|\sqrt{\langle u_i^2\rangle^k}, \langle\gamma_p^l\rangle^k, \boldsymbol{\tau}^l). \end{aligned} \quad (7)$$

We can see that q_{z_i} depends on $p(z_i|\boldsymbol{\tau}^l)$. However, for most distributions in the GSM family, we do not know the explicit expression of $p(z_i|\boldsymbol{\tau}^l)$. As a result, the explicit expression of q_{z_i} is not known either. Nevertheless, our objective is not to obtain q_{z_i} but q_{u_i}. As noted in [16] and also shown by (5), to determine q_{u_i}, it is enough to know the expectation of q_{z_i}.

In this case, the main challenge is to determine the expectation of q_{z_i} without knowing its explicit expression. Since $p(u_i|\gamma_p^l, \boldsymbol{\tau}^l)$ belongs to the GSM family, we can obtain (see [16] for details)

$$\begin{aligned} p'(u_i|\gamma_p^l, \boldsymbol{\tau}^l) &= \frac{\partial}{\partial u_i}\int_0^\infty p(u_i|z_i,\gamma_p^l) p(z_i|\boldsymbol{\tau}^l)\mathrm{d}z_i \\ &= -\gamma_p^l u_i p(u_i|\gamma_p^l, \boldsymbol{\tau}^l)\mathbb{E}_{p(z_i|u_i,\gamma_p^l,\boldsymbol{\tau}^l)}\{z_i\}, \end{aligned}$$

which allows us to get

$$\mathbb{E}_{p(z_i|u_i,\gamma_p^l,\boldsymbol{\tau}^l)}\{z_i\} = -\frac{p'(u_i|\gamma_p^l,\boldsymbol{\tau}^l)}{\gamma_p^l u_i p(u_i|\gamma_p^l,\boldsymbol{\tau}^l)}. \quad (8)$$

Combining (7) and (8), we obtain the expectation

$$\langle z_i\rangle^{k+1} = -\left.\frac{p'(u_i|\gamma_p^l,\boldsymbol{\tau}^l)}{\gamma_p^l u_i p(u_i|\gamma_p^l,\boldsymbol{\tau}^l)}\right|_{u_i=\sqrt{\langle u_i^2\rangle^k}, \gamma_p=\langle\gamma_p^l\rangle^k}. \quad (9)$$

In this work, we consider also one special case of the GSM family: the Generalized Gaussian (GG) distribution whose density is given by:

$$GG(u_i|\gamma_p^l,\tau^l) = \frac{\sqrt{\gamma_p^l}\tau^l}{2\Gamma(1/\tau^l)} e^{-|\sqrt{\gamma_p^l}u_i|^{\tau^l}} \quad (10)$$

where $\tau^l > 0$ is the shape parameter of the GG distribution.

With a GG prior, the expectation given by (9) becomes

$$\langle z_i\rangle^{k+1} = \tau^l\left[\langle\gamma_p^l\rangle^k\left((\mathbf{m}_{k+1})_i^2 + (\boldsymbol{\sigma}_{k+1}^2)_i\right)\right]^{\frac{\tau^l}{2}-1}. \quad (11)$$

3.3. Determination of q_{γ_n} and $q_{\gamma_p^l}$

Thanks to the conjugate priors for the hyperparameters, the optimal approximate distributions q_{γ_n} and $(q_{\gamma_p^l})_{l=1,\ldots,L}$ are Gamma ones. As a result, we take

$$\begin{aligned} q_{\gamma_n}^k(\gamma_n) &= \mathcal{G}(\beta^k, \xi^k), \\ q_{\gamma_p^l}^k(\gamma_p^l) &= \mathcal{G}(\eta_l^k, \zeta_l^k). \end{aligned}$$

Therefore, the optimization of q_{γ_n} and $q_{\gamma_p^l}$ can be performed by updating their parameters. Using (3), we can obtain the following update equations for parameters of q_{γ_n} and $(q_{\gamma_p^l})_{l=1,\ldots,L}$:

$$\begin{aligned} \beta^{k+1} &= \frac{M}{2} = \beta \\ \xi^{k+1} &= \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{m}_{k+1}\|^2 + \frac{1}{2}\sum_{i=1}^N (\mathbf{D}^T\mathbf{A}^T\mathbf{A}\mathbf{D})_{(i,i)}(\boldsymbol{\sigma}_{k+1}^2)_i \\ \eta_l^{k+1} &= \frac{\operatorname{card}\{I_l\}}{2} = \eta_l \\ \zeta_l^{k+1} &= \frac{1}{2}\sum_{i\in I_l}\langle z_i\rangle^{k+1}\left[(\mathbf{m}_{k+1})_i^2 + (\boldsymbol{\sigma}_{k+1}^2)_i\right], \end{aligned}$$

where $\operatorname{card}\{I_l\}$ is the number of elements in the set I_l.

The PM estimator is used for each parameter. However, we reconstruct the dictionary coefficients $\mathbf{u}$ instead of the unknown image $\mathbf{x}$. As a result, we perform the following reconstruction operation to get an estimation of the unknown image: $\widehat{\mathbf{x}} = \mathbf{D}\widehat{\mathbf{m}}$.

4. EXPERIMENTAL RESULTS

The proposed approach is evaluated through an application to a deconvolution problem which is covered by the linear forward model (1). In deconvolution problems, $\mathbf{A}$ corresponds

to a convolution operator. In the following, we present simulation results obtained by the proposed approach and compare the results with two existing approaches: a SURE-LET approach in wavelet domain [4], and a supervised total variation (TV) regularized least-squares deconvolution approach in the image domain which computes the MAP estimate thanks to a primal-dual algorithm [17]. For the TV based approach, the hyperparameter is manually tuned to obtain the best result which has the highest PSNR.

In our simulations, we used Symlet-8 wavelets over three decomposition levels, leading to 10 subbands. Moreover, we used a frame constructed by the union of nine translated wavelet bases which allows reducing blocky artifacts caused by dyadic shifts underlying the orthogonal wavelet transform.

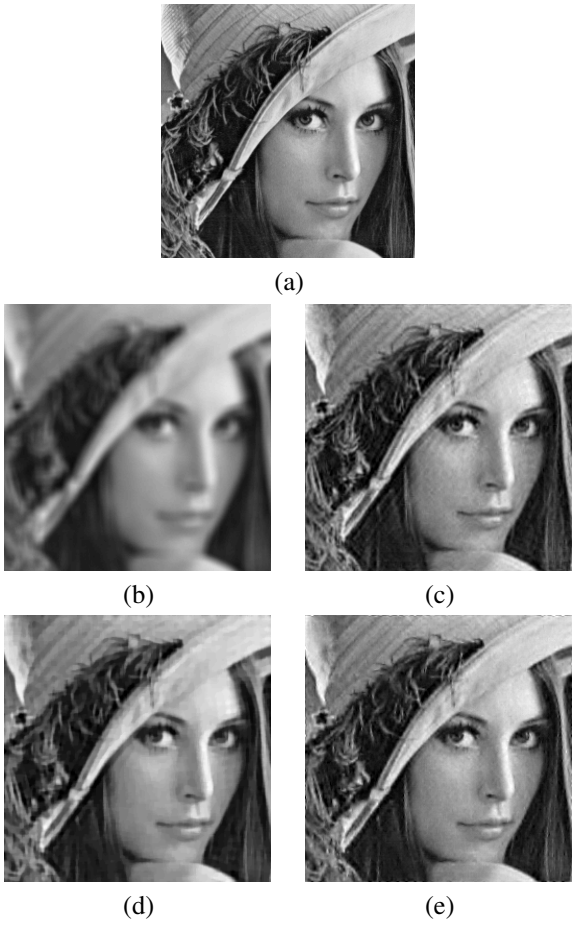

Fig. 1. (a) Original image, (b) blurred noisy image, restored ones with (c) SURE-LET [4] (d) TV (e) proposed approach.

The application was based on synthetic data generated from two images, *Lena* and *Cameraman*. A 9×9 uniform blur was applied to original images and Gaussian noises were added to the blurred ones resulting in a SNR equal to 40 dB.

The proposed approach was implemented with the following initializations: the wavelet transform of the observed data as the mean and 100 as the variance of dictionary coefficients. From these initial values, we compute the initialization of γ_n using updating equations given in Section 3.3 and $(\gamma_p^l)_{l=1,\ldots,10}$ are initialized by the same value. Concerning the shape parameters of the GG priors, 2, which gives a Gaussian distribution, was assigned to the GG prior for coarse approximation coefficients since they are generally not sparse and $[1 \ 0.9 \ 0.8]$ were assigned to the GG priors for detail coefficients from the coarsest to the finest level to enforce higher degrees of sparsity for coefficients at finer scales.

We show in Fig. 1 (a) the original *Lena* image and in Fig. 1 (b) its blurred version. The reconstructions obtained by SURE-LET [4], the TV based approach and the proposed one are shown in Fig. 1 (c), (d) and (e), respectively. We can see that the result of the proposed approach (Fig. 1 (e)) is of slightly better quality than that of SURE-LET (Fig. 1 (c)). Moreover, by comparing Fig. 1 (d) and (e), we can see that details of *Lena* are better reconstructed by the proposed approach than the TV based one, e.g. the brim of the hat, textures and the feather on the hat (upper left corner of *Lena* image) of Fig. 1 (e) are sharper than those in Fig. 1 (d).

Table 1. PSNR (dB) OBTAINED BY SURE-LET, TV BASED APPROACH AND THE PROPOSED APPROACH.

	SURE-LET	TV	proposed
Lena	29.06	28.19	**29.75**
Cameraman	26.56	27.41	**28.86**

We show in Table 1 PSNR of reconstructions obtained by the three approaches. The highest PSNR in each case is highlighted in bold. We can see that in both cases, it is the proposed approach that gives the best PSNR. For *Lena*, the proposed approach gives 29.75 dB which is 0.69 dB higher than SURE-LET and 1.56 dB higher than TV. For *Cameraman*, our approach gives 28.86 dB which is 2.3 dB larger than SURE-LET and 1.45 dB higher than TV. However, concerning the execution time, the proposed approach is slower than SURE-LET. For *Lena*, SURE-LET takes 95 seconds whereas the proposed approach takes 273 seconds. Moreover, for *Cameraman*, SURE-LET takes 44 seconds whereas the proposed approach takes 136 seconds. Our approach jointly estimate parameters and hyperparameters in an iterative way which leads to better results but requires more computation time.

In our approach, hyperparameters were determined automatically. For the inverse noise variance γ_n, our approach gives 2.14×10^5 for *Lena* which is close to the true value: 2.26×10^5, and gives 1.08×10^5 for *Cameraman*, which is also close to the true value: 1.09×10^5. We show also in Table 2 converged values of the scale parameters $(\gamma_p^l)_{l=1,\ldots,10}$ of the GG prior for *Lena*. From (2) we can see that larger values of γ_p^l lead to greater importance of the sparse prior information. In Table 2, we can see that the value of γ_p^l is larger for coefficients at finer scales: the γ_p^l of scale 1 is larger than that of

Table 2. VALUES OF γ_p^l ESTIMATED BY OUR APPROACH. FROM THE COARSEST TO THE FINEST SCALE: $3 \rightarrow 1$.

	Approxi -mation	Scale	Horizontal	Vertical	Diagonal
γ_p^l	0.03	3	21.21	44.70	65.33
		2	252.19	642.47	342.41
		1	2.25×10^3	4.46×10^3	530.57

scale 2 which is larger than that of scale 3, which means that we give much more importance on the sparsity prior information for finer scale coefficients. This result is coherent with the fact that coefficients of finer scales are sparser than those of coarser scales. Furthermore, γ_p^l for coarse approximation coefficients is small, which means that the coarse approximation coefficients are mainly determined by the data. For *Cameraman*, we do not show the estimation of γ_p^l since it exhibits a similar performance to *Lena*.

We need to note that another advantage of the proposed approach is that it can be easily used to treat other linear inverse problems, e.g. tomographic reconstruction problems.

5. CONCLUSION

In this paper, by using variational Bayesian approximations, we proposed an unsupervised Bayesian approach based on a prior distribution of the GSM family in a transform domain for linear inverse problems. The first main advantage of this approach is that a large number of hyperparameters can be estimated automatically. The second advantage is that it can be used to solve various linear inverse problems. Experimental results showed that the proposed approach can well estimate the hyperparameters and gives better reconstructions than classical approaches. Furthermore, the proposed approach can be easily applied to large dimensional problems.

REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, 1995.

[2] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, pp. 1135–1151, 1981.

[3] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, 2009.

[4] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616–4632, 2009.

[5] S. Vaiter, C. A. Deledalle, G. Peyré, C. Dossal, and J. Fadili, "Local behavior of sparse analysis regularization: Applications to risk estimation," *Appl. Comput. Harmon. Anal.*, vol. 35, no. 3, pp. 433–451, 2013.

[6] F. Xue, F. Luisier, and T. Blu, "Multi-Wiener SURE-LET deconvolution.," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1954–1968, 2013.

[7] G. Zhang and N. Kingsbury, "Fast L0-based image deconvolution with variational Bayesian inference and Majorization-Minimization," in *GlobalSIP 2013 Symposium on Optimization in Machine Learning and Signal Processing*, 2013, pp. 1081–1084.

[8] J. L. Starck, J. Fadili, and F. Murtagh, "The undecimated wavelet decomposition and its reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 297–309, 2007.

[9] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, 2003.

[10] J. M. Bioucas-Dias, "Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors," *IEEE Trans. Image Process.*, vol. 15, no. 4, pp. 937–951, 2006.

[11] 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.

[12] L. Chaâri, J.-C. Pesquet, J. Y. Tourneret, P. Ciuciu, and A. Benazza-Benyahia, "A hierarchical Bayesian model for frame representation," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5560–5571, 2010.

[13] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*, Springer-Verlag, New York, USA, 2005.

[14] A. Fraysse and T. Rodet, "A measure-theoretic variational Bayesian algorithm for large dimensional problems.," *SIAM J. Imaging Sci.*, vol. 7, no. 4, pp. 2591–2622, 2014.

[15] Y. Zheng, A. Fraysse, and T. Rodet, "Efficient variational Bayesian approximation method based on subspace optimization," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 681–693, 2015.

[16] J. Palmer, K. Kreutz-Delgado, B. D. Rao, and D. P. Wipf, "Variational EM algorithms for non-Gaussian latent variable models," in *Adv. Neural. Inf. Process. Syst. 18*, pp. 1059–1066. MIT Press, 2006.

[17] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators," *Set-Valued Var. Anal.*, vol. 20, no. 2, pp. 307–330, 2012.