

Approche bayésienne variationnelle non-supervisée basée sur les ondelettes

Yuling ZHENG¹, Aurélie FRAYSSE², Thomas RODET³

¹LIGM, Université Paris-Est, CNRS, Champs-sur-Marne, 77454 Marne-la-Vallée, France

²L2S, Université Paris-Sud, CNRS, Supelec, R. Joliot-Curie, 91190 Gif-sur-Yvette, France

³SATIE, ENS Cachan, CNAM, Av. du Prés. Wilson, 94230 Cachan, France

yuling.zheng@univ-mlv.fr, fraysse@lss.supelec.fr, trodet@satie.ens-cachan.fr

Résumé – Dans cet article, nous introduisons une approche bayésienne variationnelle pour résoudre les problèmes inverses linéaires en utilisant une décomposition sur un dictionnaire. Cette approche s’appuie sur un *a priori* de mélange infini de gaussiennes et une méthode d’approximation bayésienne variationnelle accélérée. Notre approche est non-supervisée et peut être utilisée pour résoudre n’importe quel problème inverse linéaire. Les résultats de reconstruction sont comparés aux approches de l’état de l’art sur un problème de déconvolution.

Abstract – In this paper we present a variational Bayesian approach using a sparse decomposition over a dictionary for solving linear inverse problems. This approach is based on a Gaussian Scale Mixture prior and an accelerated variational Bayesian approximation method. Our approach is unsupervised and can be used to solve any linear inverse problem. We show the good performance of our approach by comparing with state of the art approaches on a deconvolution problem.

1 Introduction

L’objectif de ce travail est de développer des approches de résolution des problèmes inverses mal posés de façon non-supervisée. De plus, nous voulons utiliser une information *a priori* de régularité par morceau. Les deux principales familles d’approches permettant d’introduire cette information sont les approches utilisant la variation totale [1] ou les approches utilisant un domaine transformé et un *a priori* de parcimonie [12].

L’introduction de la parcimonie peut se faire à l’aide d’un terme de régularisation défini par la norme L^1 par exemple [3]. Dans ce type d’approches on doit déterminer un hyperparamètre qui contrôle le compromis entre la fidélité aux données et la confiance aux informations *a priori*. Il est avantageux d’estimer automatiquement cet hyperparamètre, afin d’avoir des méthodes facilement utilisables par des non spécialistes. Suite aux premières approches de type courbe en L et validation croisée, des méthodes sont apparues utilisant l’estimateur de Stein de l’erreur quadratique moyenne (*Stein’s unbiased risk estimate (SURE)*) [11]. Cette approche a tout d’abord été introduite dans le cas du débruitage puis a été généralisée pour d’autres problèmes inverses linéaires [4, 9]. Ces approches sont intéressantes mais elles ont un degré de généralité moins important que les approches bayésiennes. Ces approches consistent à attribuer une loi *a priori* sur cet hyperparamètre et à l’estimer conjointement aux paramètres d’intérêt [6].

Nous développons une approche entièrement bayésienne conduisant à une estimation non-supervisée. De plus, nous introduisons une information de régularité par morceaux dans les

images en utilisant une projection sur un dictionnaire redondant ainsi qu’une information de parcimonie sur les coefficients [7].

L’autre avantage d’une approche bayésienne est qu’elle permet plus de flexibilité sur l’information *a priori*. Ainsi dans le cas présent nous utilisons aussi le fait que les images ont un degré de parcimonie variable en fonction de l’échelle des coefficients d’ondelette. Ainsi les coefficients de détails grossiers sont moins parcimonieux que les coefficients de détails plus fins. Cette propriété correspond à la majorité des images réelles. Afin de s’adapter au contenu géométrique des images, nous allons déterminer des hyperparamètres différents par type de coefficients suivant qu’ils soient fins ou grossiers, horizontaux, verticaux ou diagonaux. Pour satisfaire les conditions ci-dessus et avoir une loi *a priori* conjuguée avec une vraisemblance, nous avons choisi d’utiliser une loi appartenant à la classe des lois de mélange continue de gaussiennes (*Gaussian Scale Mixture (GSM)*) [12] qui comprend des lois à queue lourde favorisant les solutions parcimonieuses. Nous utilisons ici la loi gaussienne généralisée qui est un modèle bien adapté aux coefficients d’ondelette des images [2, 7].

Néanmoins, la loi *a posteriori* dans le cas non-supervisé est généralement complexe et les estimateurs classiques tels que le Maximum *a posteriori* ou l’Espérance *a posteriori* ne peuvent pas être calculés directement. Pour résoudre ce problème de manière efficace, plutôt que d’utiliser les approches Monte Carlo par chaînes de Markov (MCMC) [2], nous recourons aux approches d’approximation bayésienne variationnelle (BV) qui donnent une approximation analytique de forme plus simple

que la distribution originale. Par ailleurs, pour obtenir une approche plus efficace, nous utilisons ici une méthode BV accélérée que nous avons récemment proposée dans [13].

2 Modèle utilisé

Supposons que l'image inconnue $\mathbf{x} \in \mathbb{R}^P$ peut être représentée par une décomposition sur un dictionnaire $\mathbf{D} \in \mathbb{R}^{P \times N}$ sous la forme $\mathbf{x} = \mathbf{D}\mathbf{u}$ où $\mathbf{u} \in \mathbb{R}^N$ représente les coefficients associés. Nous considérons ici un modèle linéaire :

$$\mathbf{y} = \mathbf{A}\mathbf{D}\mathbf{u} + \mathbf{n}, \quad (1)$$

où $\mathbf{y} \in \mathbb{R}^M$ représente les données, l'opérateur $\mathbf{A} \in \mathbb{R}^{M \times P}$ est supposé connu et \mathbf{n} est un bruit blanc gaussien, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \gamma_n^{-1}\mathbf{I})$, avec γ_n l'inverse de la variance du bruit. Par conséquent, la vraisemblance est donnée par : $p(\mathbf{y}|\mathbf{u}, \gamma_n) = \mathcal{N}(\mathbf{A}\mathbf{D}\mathbf{u}, \gamma_n^{-1}\mathbf{I})$.

Concernant les coefficients \mathbf{u} , nous considérons une loi *a priori* séparable de la famille GSM. Néanmoins, les coefficients n'ont pas toujours les mêmes caractéristiques statistiques. Pour prendre cela en compte, nous divisons les coefficients \mathbf{u} en L sous-bandes et supposons que les paramètres des lois *a priori* des coefficients dans les sous-bandes différentes sont différents. Notons l'ensemble des indices des coefficients dans la l^e sous-bande par $(I_l)_{l=1, \dots, L}$. On écrit la loi *a priori* de \mathbf{u} comme

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

où $\mathbf{z} = (z_1, \dots, z_N)$ est le vecteur des variables cachées qui suivent des distributions $p(z_i|\boldsymbol{\tau}^l)$ de paramètres $\boldsymbol{\tau}^l$. Notons que $p(z_i|\boldsymbol{\tau}^l)$ dépend de $p(u_i|\gamma_p^l, \boldsymbol{\tau}^l)$. Ici, $\gamma_p = (\gamma_p^1, \dots, \gamma_p^L)$ est le vecteur des paramètres d'échelle de notre *a priori* GSM.

Au final, on introduit un hyperparamètre γ_n pour la loi du bruit, un vecteur d'hyperparamètres γ_p associé à l'échelle des lois *a priori* des coefficients et un vecteur d'hyperparamètres $\boldsymbol{\tau}$ associé aux paramètres de forme de ces mêmes lois. Les paramètres de forme permettent de régler le degré de parcimonie introduit *a priori*. Afin d'introduire un degré de parcimonie qui varie avec l'échelle des coefficients en ondelettes nous choisissons de fixer $\boldsymbol{\tau}$. Néanmoins, nous estimons γ_n et γ_p car ils règlent le compromis entre la fidélité aux données et la confiance à l'information *a priori*. Pour ce faire, des lois de Jeffreys non-informatives sont affectées à γ_n et γ_p .

En utilisant la règle de Bayes, nous pouvons en déduire la loi *a posteriori* jointe :

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

Cette distribution est de forme compliquée. Pour s'affranchir de la difficulté de calcul, nous pouvons recourir aux approches d'approximation bayésienne variationnelle (BV).

3 Les approches d'approximation BV

Notons l'ensemble des variables à estimer $\Theta = \{\mathbf{u}, \mathbf{z}, \gamma_n, \gamma_p\}$. Les approches d'approximation BV nous offrent une approximation séparable q_Θ de la loi *a posteriori* $p(\Theta|\mathbf{y}, \boldsymbol{\tau})$ en minimisant la divergence de Kullback-Leibler (\mathcal{KL}).

BV classique Supposons que $q_\Theta(\Theta) = \prod_i q_i(\Theta_i)$. L'approche BV classique [10] utilise la solution analytique donnée par :

$$q_i(\Theta_i) \propto \exp \left(\langle \log p(\mathbf{y}, \Theta) \rangle_{\prod_{j \neq i} q_j(\Theta_j)} \right), \quad (2)$$

où $p(\mathbf{y}, \Theta)$ est la distribution jointe qui est explicitement connue. Néanmoins, nous pouvons voir dans l'éq. (2) que chaque composante q_i dépend de toutes les autres composantes $(q_j)_{j \neq i}$. Par conséquent, on ne peut pas obtenir une expression explicite pour q_Θ sauf dans des cas extrêmement simples. En pratique, ce problème est généralement abordé en utilisant un algorithme alterné, qui met à jour une composante en fixant les autres à chaque étape, pour approcher la solution de façon itérative. Néanmoins, ce type d'approches n'est pas très efficace.

BV-MG Récemment, Fraysse *et al.* [5] ont proposé une méthode BV de type gradient, qui est beaucoup plus efficace que l'approche BV classique. Cette méthode a été améliorée dans [13], donnant lieu à une méthode d'approximation BV de sous-espace de mémoire de gradient (BV-MG) encore plus efficace. La méthode BV-MG utilise l'équation de mise à jour suivante :

$$\begin{aligned} q_i^{k+1}(\Theta_i) &= K^k(\mathbf{s}^k) q_i^k(\Theta_i) \left[\frac{q_i^r(\Theta_i)}{q_i^k(\Theta_i)} \right]^{s_1^k} \left[\frac{q_i^k(\Theta_i)}{q_i^{k-1}(\Theta_i)} \right]^{s_2^k}, \quad (3) \\ \text{avec } q_i^r(\Theta_i) &\propto \exp \left[\langle \log p(\mathbf{y}, \Theta) \rangle_{\prod_{j \neq i} q_j^k(\Theta_j)} \right] \end{aligned}$$

où $\mathbf{s}^k = [s_1^k, s_2^k]$ est le pas déterminé grâce au développement de Taylor d'ordre deux du critère et $K^k(\mathbf{s}^k)$ représente la constante de normalisation. L'éq. (3) montre que q_i^{k+1} ne dépend pas des $(q_j^{k+1})_{j \neq i}$, mais des $(q_j^k)_{j \neq i}$ connus dès la k^e itération, ce qui permet une mise à jour en parallèle et donc une accélération significative par rapport aux approches BV classiques.

4 Application des approches BV

En ce qui concerne l'hypothèse de séparabilité, nous considérons ici une séparabilité totale :

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

Comme $p(\mathbf{z}, \gamma_n, \gamma_p | \mathbf{u}, \mathbf{y}, \boldsymbol{\tau})$ est elle-même séparable, l'approximation BV classique donne directement des solutions explicites pour q_{z_i} , q_{γ_n} et $q_{\gamma_p^l}$. Néanmoins, ce n'est pas le cas pour q_{u_i} . Par conséquent, la méthode BV-MG est utilisée pour approcher q_{u_i} afin d'améliorer la vitesse de convergence.

Détermination de q_{u_i} Lorsqu'un *a priori* GSM est utilisé, la loi conditionnelle $p(u_i|z_i, \gamma_p^l)$ est une loi gaussienne qui est conjuguée avec la vraisemblance $p(\mathbf{y}|\mathbf{u}, \gamma_n)$. Par conséquent,

on sait que la loi approchante optimale $(q_{u_i})_{i=1,\dots,N}$ est gaussienne. Nous prenons donc

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

Ainsi, la mise à jour des lois q_{u_i} revient à la réactualisation de leurs paramètres : la moyenne \mathbf{m}_k et la variance $\boldsymbol{\sigma}_k^2$. En utilisant (3), nous avons les équations de mise à jour suivantes :

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

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

Dans les équations ci-dessus, nous avons omis les indices $(\cdot)_i$ pour alléger l'écriture. Par ailleurs, $\boldsymbol{\sigma}_r^2$ et \mathbf{m}_r sont deux paramètres de la fonction intermédiaire q_i^r dont les équations de mise à jour sont données par

$$(\boldsymbol{\sigma}_r^2)_i = \left[\langle \gamma_n \rangle^k (\mathbf{D}^T \mathbf{A}^T \mathbf{A} \mathbf{D})_{(i,i)} + \langle \gamma_p^l \rangle^k \langle z_i \rangle^k \right]^{-1}, \quad (4)$$

$$(\mathbf{m}_r)_i = (\boldsymbol{\sigma}_r^2)_i \langle \gamma_n \rangle^k \left[\mathbf{D}^T \mathbf{A}^T \mathbf{y} - \mathbf{D}^T \mathbf{A}^T \mathbf{A} \mathbf{D} \mathbf{m}_k + \text{diag}(\mathbf{D}^T \mathbf{A}^T \mathbf{A} \mathbf{D}) \circ \mathbf{m}_k \right]_i \quad (5)$$

où $\langle w \rangle^k = \mathbb{E}_{q_w^k}(w)$, $\text{diag}(\mathbf{M})$ est un vecteur incluant les éléments diagonaux de \mathbf{M} et \circ représente le produit d'Hadarnard.

Détermination de q_{z_i} Pour les variables cachées $(z_i)_{i=1,\dots,N}$, en utilisant (2), on a

$$q_{z_i}^{k+1}(z_i) \propto \exp \left(\frac{1}{2} \ln(z_i) - \frac{1}{2} \langle \gamma_p^l \rangle^k z_i \langle u_i^2 \rangle^k + \ln p(z_i | \boldsymbol{\tau}^l) \right)$$

$$\propto \sqrt{z_i} p(z_i | \boldsymbol{\tau}^l) \exp \left(-\frac{1}{2} \langle \gamma_p^l \rangle^k \langle u_i^2 \rangle^k z_i \right)$$

$$= p(z_i | \sqrt{\langle u_i^2 \rangle^k}, \langle \gamma_p^l \rangle^k, \boldsymbol{\tau}^l). \quad (6)$$

Nous pouvons voir que q_{z_i} dépend alors de $p(z_i | \boldsymbol{\tau}^l)$. Néanmoins, pour la plupart des lois de la famille GSM, la densité $p(z_i | \boldsymbol{\tau}^l)$ est généralement inconnue, ce qui ne permet pas de déterminer q_{z_i} [8]. Cependant, notre principale préoccupation est d'obtenir une expression explicite de q_{u_i} et non de q_{z_i} . Comme remarqué dans [8] et dans l'éq. (4), la mise à jour de q_{u_i} nécessite seulement la connaissance de l'espérance de z_i sous q_{z_i} .

Dans ce cas, l'objectif est de déterminer l'espérance de z_i sans connaître l'expression explicite de q_{z_i} . Comme $p(u_i | \gamma_p^l, \boldsymbol{\tau}^l)$ est dans la famille GSM, on peut utiliser les résultats de [8]

$$p'(u_i | \gamma_p^l, \boldsymbol{\tau}^l) = \frac{\partial}{\partial u_i} \int_0^\infty p(u_i | z_i, \gamma_p^l) p(z_i | \boldsymbol{\tau}^l) dz_i$$

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

qui nous permet d'obtenir

$$\mathbb{E}_{p(z_i | u_i, \gamma_p^l, \boldsymbol{\tau}^l)} \{z_i\} = -\frac{p'(u_i | \gamma_p^l, \boldsymbol{\tau}^l)}{\gamma_p^l u_i p(u_i | \gamma_p^l, \boldsymbol{\tau}^l)}. \quad (7)$$

En combinant (6) et (7), nous obtenons

$$\langle z_i \rangle^{k+1} = -\frac{p'(u_i | \gamma_p^l, \boldsymbol{\tau}^l)}{\gamma_p^l u_i p(u_i | \gamma_p^l, \boldsymbol{\tau}^l)} \Big|_{u_i = \sqrt{\langle u_i^2 \rangle^k}, \gamma_p = \langle \gamma_p^l \rangle^k}. \quad (8)$$

Dans ce travail, nous considérons aussi un cas particulier de la famille GSM : une loi gaussienne généralisée (GG) [12] dont la densité de probabilité s'écrit comme :

$$GG(u_i | \gamma_p^l, \boldsymbol{\tau}^l) = \frac{\sqrt{\gamma_p^l \boldsymbol{\tau}^l}}{2\Gamma(1/\boldsymbol{\tau}^l)} e^{-\sqrt{\gamma_p^l} u_i |\boldsymbol{\tau}^l|.} \quad (9)$$

Avec un *a priori* GG, l'espérance donnée par (8) devient

$$\langle z_i \rangle^{k+1} = \boldsymbol{\tau}^l \left[\langle \gamma_p^l \rangle^k ((\mathbf{m}_{k+1})_i^2 + (\boldsymbol{\sigma}_{k+1}^2)_i) \right]^{\frac{\boldsymbol{\tau}^l}{2} - 1}. \quad (10)$$

Détermination de q_{γ_n} et $q_{\gamma_p^l}$ Grâce aux lois *a priori* conjuguées pour γ_n et γ_p^l , les approximations optimales q_{γ_n} et $q_{\gamma_p^l}$ sont des lois Gamma. L'optimisation des lois revient à optimiser leurs paramètres. Nous ne montrons pas ici les équations de mise à jour de ces paramètres qui sont dans [5].

Comme nous avons reconstruit les coefficients \mathbf{u} , une étape de synthèse est appliquée à la fin pour obtenir une estimation de \mathbf{x} : $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{m}}$.

5 Résultats

Pour évaluer notre approche, nous montrons son application à un problème de déconvolution et comparons les résultats avec deux approches existantes : l'approche SURE-LET [9] dans le domaine des ondelettes et une approche non-supervisée basée sur la variation totale (TV) dans le domaine de l'image [13].

Dans toutes les simulations, nous utilisons une trame redondante constituée de neuf bases d'ondelettes orthogonales. Cette trame possède la propriété d'invariance que la transformée en ondelettes orthogonale (TO) n'a pas. Nous utilisons pour cela les ondelettes Daubechies 4 et nous faisons une décomposition sur 3 niveaux, ce qui conduit à $L = 10$ sous-bandes différentes.

Nous testons les trois approches avec les données générées à partir de deux images standards : *Lena* et *Cameraman*. Pour générer des données, un noyau de convolution uniforme de taille 9×9 a été utilisé. Ensuite, nous avons ajouté un bruit gaussien aux images convoluées à un SNR = 40 dB.

Notre approche est mise en œuvre avec les initialisations suivantes : la TO des données comme moyenne, 100 comme variance des coefficients à estimer, une valeur calculée à partir de la TO des données en utilisant l'équation de mise à jour de γ_n comme la valeur initiale de γ_n et γ_p^l . Concernant les paramètres de forme de l'*a priori* GG, nous avons pris 2, qui conduit à une loi gaussienne, pour les coefficients d'approximation, et [1, 0.9, 0.8], qui conduisent à de plus en plus de parcimonie, pour les coefficients de détails de l'échelle la plus grossière à l'échelle la plus fine.

Nous montrons dans la figure 1 (a) une image de *Lena*. Les reconstructions obtenues par les trois approches sont montrées dans la figure 1 (b)-(d). Nous pouvons voir que le résultat de notre approche (figure 1 (d)) est meilleur que celui obtenu par SURE-LET (figure 1 (b)). En comparant la figure 1 (c) et (d), nous pouvons voir que les détails de *Lena* sont mieux reconstruits par notre approche que par l'approche basée sur TV. Par

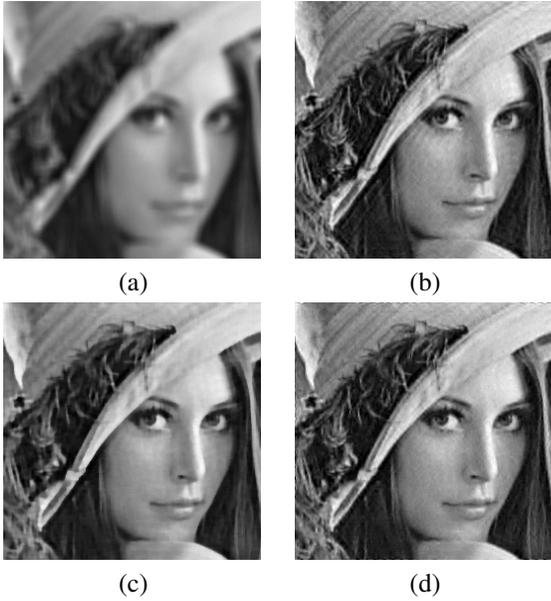


FIGURE 1 – (a) Image floue bruitée, images reconstruites par (b) SURE-LET [9] (c) TV [13] (d) l’approche proposée.

exemple, le bord du chapeau et la plume sur le chapeau sont plus nets que ceux dans la figure 1 (c).

TABLE 1 – PSNR (dB) OBTENU PAR SURE-LET, L’APPROCHE BASÉE SUR TV ET L’APPROCHE PROPOSÉE

	SURE-LET	TV	proposée
Lena	29.06	29.22	29.75
Cameraman	26.56	29.61	28.86

Nous montrons dans le tableau 1 les PSNR des reconstructions obtenues par les trois approches. Nous pouvons voir que dans les deux cas, l’approche proposée donne un meilleur PSNR que l’approche SURE-LET : pour *Lena*, notre approche donne 29.75 dB qui est 0.69 dB de plus que celui obtenu par SURE-LET, et pour *Cameraman*, notre approche donne 28.86 dB qui est 2.3 dB de plus que celui obtenu par SURE-LET. La bonne performance de notre approche par rapport à SURE-LET peut être expliquée par le fait qu’elle estime dix paramètres différents pour les coefficients des dix sous-bandes.

En comparant les PSNR de notre approche et de l’approche TV, nous pouvons voir que pour *Lena*, notre approche donne un PSNR 0.53 dB de plus que celui de l’approche TV, mais pour *Cameraman*, c’est l’approche TV qui donne un PSNR 0.75 dB de plus que celui de notre approche. En fait, l’image *Camera-man* est plutôt régulière par morceau et l’a priori TV marche très bien pour ce type d’images. Par contre, dans *Lena*, il y a plus de détails. Dans ce cas, c’est l’approche basée sur un a priori dans le domaine des ondelettes qui est la plus adaptée.

6 Conclusion

Dans cet article, nous avons développé une approche non-supervisée basée sur un a priori de la famille GSM dans un do-

main transformé à l’aide de l’approximation bayésienne variationnelle. L’avantage principal est que tous les hyperparamètres sont estimés automatiquement. Par ailleurs, cette approche est assez générale : elle est directement applicable pour tous types de problèmes linéaires (tomographie, inpainting, etc...), ce qui n’est pas forcément le cas des approche de type SURE. Les comparaisons avec les approches classiques montrent la bonne performance de notre approche.

Références

- [1] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Variational Bayesian super resolution. *IEEE Trans. Image Process.*, 20(4) :984–999, 2011.
- [2] 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.*, 58(11) :5560–5571, 2010.
- [3] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.*, 90(432) :1200–1224, 1995.
- [4] Y. C. Eldar. Generalized SURE for exponential families : Applications to regularization. *IEEE Trans. Signal Process.*, 57(2) :471–481, 2009.
- [5] A. Fraysse and T. Rodet. A measure-theoretic variational Bayesian algorithm for large dimensional problems. *SIAM J. Imaging Sci.*, 7(4) :2591–2622, 2014.
- [6] D. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
- [7] S. G. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7) :674–693, 1989.
- [8] 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.*, pages 1059–1066, 2005.
- [9] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux. A SURE approach for digital signal/image deconvolution problems. *IEEE Trans. Signal Process.*, 57(12) :4616–4632, 2009.
- [10] V. Šmídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer, 2006.
- [11] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, pages 1135–1151, 1981.
- [12] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky. Random cascades on wavelet trees and their use in analyzing and modeling natural images. *Appl. Comput. Harmon. Anal.*, 11(1) :89–123, July 2001.
- [13] Y. Zheng, A. Fraysse, and T. Rodet. Efficient variational Bayesian approximation method based on subspace optimization. *IEEE Trans. Image Process.*, 24(2) :681–693, 2015.