

Un algorithme de seuillage itératif non-supervisé pour la décomposition parcimonieuse de signaux dans des unions de dictionnaires.

Matthieu KOWALSKI, Thomas RODET

Univ. Paris-Sud, Laboratoire des Signaux et Systèmes (CNRS-SUPELEC-Univ. Paris-Sud)
SUPELEC, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France
kowalski@lss.supelec.fr, rodet@lss.supelec.fr

Résumé – Les approches classiques de régressions parcimonieuses par seuillage itératif ont comme limitation le choix d’hyper-paramètres servant à régler le degré de parcimonie. Le réglage de ces paramètres pouvant être fastidieux, on propose donc une méthode non-supervisée. Pour des raisons algorithmiques, les méthodes de décomposition parcimonieuse utilisent généralement une pénalité ℓ_1 . Cette pénalité correspond à un *a priori* Laplacien qui code moins fidèlement l’information de parcimonie par rapport à un *a priori* Bernouilli Gaussien que nous utiliserons par la suite. Malgré ce choix, nous obtenons un algorithme de seuillage itératif de forme semblable à l’algorithme ISTA très populaire dans l’approche par optimisation convexe. Le principal avantage est que notre algorithme permet une décomposition automatique de notre signal en plusieurs couches, tout en pouvant s’affranchir du réglage des hyper-paramètres. On obtient un algorithme très simple à mettre en œuvre, qui reste dans la famille des algorithmes de seuillage itératifs. Cet algorithme se généralise directement dans le cas de modèle de parcimonie par groupe. On l’applique à la décomposition d’un signal audio en couches transitoire et tonale, chacune étant représentée dans un dictionnaire de Gabor.

Abstract – Sparse hypothesis is now widely used for ill posed inverse problem such as coding, denoising etc. We are interested in signal supposed to admit an expansion as the sum of two independant layers (such as edges+textures in images, transients+tonals in audio signals). Classical approaches of sparse regression by iterative thresholding involved the tune of hyperparameters. The tunage of these parameters can be fastidious, and we propose then an unsupervised algorithm. For algorithmical reasons, sparse expansion methods often use an ℓ_1 penalty. This penalty corresponds to a Laplacian prior which is not as suitable as a Bernoulli-Gaussian prior, used in the following, to take sparsity into account. With this choice for the prior, we still obtain a simple algorithm in the spirit of iterative thresholding algorithms. This algorithm can be extended straightforward to group-sparsity. It is applied on audio signal expansion in two layers (transients+tonals), where each are encoded in a Gabor dictionary.

1 Introduction

L’hypothèse de parcimonie est aujourd’hui classique pour la résolution de problèmes inverses, de codage, de débruitage etc. Un signal est supposé parcimonieux lorsqu’on peut l’écrire comme une combinaison linéaire d’un «petit nombre» d’atomes élémentaires. Par exemple, une image peut se représenter efficacement dans une base d’ondelette, et les signaux audio se prêtent bien aux repères de Gabor. Cependant, un signal fait souvent intervenir des composantes de nature différentes, et il est alors souhaitable de construire un dictionnaire comme l’union de plusieurs dictionnaires chacun judicieusement choisi pour représenter au mieux chacune de ces composantes. Par exemple, une image peut se voir comme la somme d’une image contenant les bords et d’une image contenant les textures ; un signal audio peut se voir comme la somme d’une partie tonale et d’une partie transitoire. Une telle approche se retrouve sous les noms de décompositions hybrides [2] ou morphologiques [9].

Afin d’obtenir une telle décomposition, une première ap-

proche consiste à utiliser le cadre de l’optimisation convexe. Starck *et al.* [9] utilisent dans ce cadre une attache aux données ℓ_2 et une pénalité ℓ_1 sur les coefficients morphologiques. L’avantage de cette modélisation est l’existence d’algorithmes efficaces comme FISTA (Fast Iterative Shrinkage/Thresholding Algorithm) [1]. Un inconvénient est le réglage d’hyperparamètres correspondant aux termes de régularisations. Pour éviter le réglage de ces hyperparamètres, des modèles aléatoires ont été proposés [6, 5, 3] dans le cadre des signaux audios, qui ont l’avantage d’estimer conjointement les couches tonales et transitoires, sans ce réglage fastidieux des hyperparamètres. Cependant, les algorithmes obtenus ne sont pas aussi simple à mettre en œuvre que les algorithmes de seuillage itératifs.

2 Modèle

On s’intéresse aux signaux, éventuellement complexes, de dimensions N qui peuvent s’écrire comme la somme de deux composantes et d’un résidu. Plus précisément, soient $\mathbf{U} = \{\mathbf{u}_n \in$

$\mathbb{C}^N\}_{n=1}^{N_1}$ et $\mathbf{V} = \{\mathbf{v}_m \in \mathbb{C}^N\}_{m=1}^{N_2}$ deux repères de \mathbb{C}^N (et donc $N_1, N_2 \geq N$). Soient Λ et Δ deux sous-ensembles finis de \mathbb{N} . On suppose qu'un signal $\mathbf{y} \in \mathbb{C}^N$ admet une décomposition Hybride [2] ou Morphologique [9] sous la forme

$$\mathbf{y} = \sum_{\lambda \in \Lambda} \alpha_\lambda \mathbf{u}_\lambda + \sum_{\delta \in \Delta} \beta_\delta \mathbf{v}_\delta + \mathbf{n}, \quad (1)$$

où $\mathbf{n} \in \mathbb{C}^N$ est du bruit, et où $\alpha_\lambda, \beta_\delta \in \mathbb{C}$ sont appelés les coefficients de synthèse. Les ensembles Λ et Δ sont appelés les cartes de significances : si $\lambda \in \Lambda$, alors $\alpha_\lambda \neq 0$, et si $\lambda \notin \Lambda$, alors $\alpha_\lambda = 0$.

On choisit un modèle Bernoulli-Gaussien complexe, qui prend naturellement en compte la parcimonie. On définit les variables indicatrices correspondant aux cartes Λ et Δ :

$$X_n = \begin{cases} 1 & \text{si } n \in \Lambda \\ 0 & \text{sinon,} \end{cases} \quad \tilde{X}_m = \begin{cases} 1 & \text{si } m \in \Delta \\ 0 & \text{sinon,} \end{cases} \quad (2)$$

avec p et \tilde{p} les probabilités d'appartenance : $\forall n, p = p(X_n = 1)$, $\forall m, \tilde{p} = p(\tilde{X}_m = 1)$. Le modèle de signal correspondant s'écrit alors

$$\mathbf{y} = \sum_{n=1}^{N_1} X_n \alpha_n \mathbf{u}_n + \sum_{m=1}^{N_2} \tilde{X}_m \beta_m \mathbf{v}_m + \mathbf{n}. \quad (3)$$

L'a priori Bernoulli-Gaussien s'écrit

$$p(\alpha_\lambda | X_\lambda) = (1 - X_\lambda) \delta_0 + X_\lambda \mathcal{CN}(0, \sigma^2), \quad (4)$$

$$p(\beta_\delta | \tilde{X}_\delta) = (1 - \tilde{X}_\delta) \delta_0 + \tilde{X}_\delta \mathcal{CN}(0, \tilde{\sigma}^2), \quad (5)$$

où les distributions gaussienne complexe sont univariées de densité [7] : $f(z) = \frac{1}{\pi \sigma^2} e^{-\frac{|z|^2}{\sigma^2}}$.

Les distributions gaussiennes complexes $\mathcal{CN}(\mu, \Sigma, C)$ sont définies par trois paramètres : la moyenne $\mu = \mathbb{E}\{\mathbf{z}\}$, la matrice de variance $\Sigma = \mathbb{E}\{(\mathbf{z} - \mu)(\mathbf{z} - \mu)^T\}$, et la matrice de relation $C = \mathbb{E}\{(\mathbf{z} - \mu)(\mathbf{z} - \mu)^T\}$. Lorsque la matrice de relation est nulle, on note simplement $\mathcal{CN}(\mu, \Sigma)$. Enfin, on suppose $\mathbf{n} \sim \mathcal{CN}(0, \sigma_0^2 \mathbf{I})$.

3 Dérivation de l'algorithme

On réécrit le problème de façon à faire apparaître une variable cachée, qui nous permettra d'exploiter le cadre des algorithmes d'Espérance-Maximisation (EM). On pourra alors donner une estimation des cartes de significances et des coefficients de synthèse.

3.1 GEM

Afin de simplifier les notations, on réécrit (1) sous la forme :

$$\mathbf{y} = \mathbf{U}\boldsymbol{\alpha} + \mathbf{V}\boldsymbol{\beta} + \mathbf{n} = \boldsymbol{\Phi}\boldsymbol{\theta} + \mathbf{n}, \quad (6)$$

avec $\boldsymbol{\Phi} = [\mathbf{U}\mathbf{V}]$ et $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$.

Le but est de trouver une estimation de $(\boldsymbol{\theta}, \Lambda, \Delta, p, \sigma^2, \tilde{p}, \tilde{\sigma}^2, \sigma_0^2)$ qui maximise l'a posteriori $p(\boldsymbol{\theta}, \Lambda, \Delta, p, \sigma^2, \tilde{p}, \tilde{\sigma}^2, \sigma_0^2 | \mathbf{y})$. En

notant l'ensemble des paramètres $T = \{p, \sigma^2, \tilde{p}, \tilde{\sigma}^2, \sigma_0^2\}$, on peut écrire

$$p(\boldsymbol{\theta}, \Lambda, \Delta, T | \mathbf{y}) = \frac{1}{p(\mathbf{y})} p(\mathbf{y} | \boldsymbol{\theta}, \Lambda, \Delta) p(\boldsymbol{\theta}, \Lambda, \Delta | T) p(T),$$

où $p(\mathbf{y} | \boldsymbol{\theta}, \Lambda, \Delta, \sigma_0^2) \sim \mathcal{N}(\boldsymbol{\theta}, \sigma_0^2 \mathbf{I})$ et $p(\boldsymbol{\theta}, \Lambda, \Delta | p, \sigma^2, \tilde{p}, \tilde{\sigma}^2) \sim \mathcal{BN}(p, \sigma^2; \tilde{p}, \tilde{\sigma}^2)$. En se basant sur [4], on peut utiliser une stratégie Espérance - Maximisation généralisée (GEM) afin de maximiser la vraisemblance pénalisée, en écrivant (6) comme

$$\mathbf{z} = \boldsymbol{\theta} + \mu \mathbf{n}_1; \quad \mathbf{y} = \boldsymbol{\Phi} \mathbf{z} + \mathbf{n}_2, \quad (7)$$

où \mathbf{z} sont les variables cachées, $\mathbf{n}_1 \sim \mathcal{CN}(0, \mathbf{I})$ et $\mathbf{n}_2 \sim \mathcal{CN}(0, \sigma_0^2 \mathbf{I} - \mu^2 \boldsymbol{\Phi} \boldsymbol{\Phi}^T)$ avec $\mu^2 < \frac{\sigma_0^2}{\|\boldsymbol{\Phi} \boldsymbol{\Phi}^*\|}$.

Le dictionnaire $\boldsymbol{\Phi}$ introduit un couplage entre les coefficients de synthèse, et la distribution des coefficients $(\mathbf{y}, \boldsymbol{\theta})$ devient corrélées et donc difficile à manipuler en pratique. L'introduction d'une variable cachée \mathbf{z} permet de manipuler ce couplage simplement, grâce à un algorithme GEM permettant d'estimer $\boldsymbol{\theta}, \Lambda, \Delta$ en maximisant la vraisemblance complète pénalisée $\log(p(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}, \Lambda, \Delta, T) + \log(p(\boldsymbol{\theta}, \Lambda, \Delta | T) p(T))$.

L'étape E En reprenant les calculs de [4], on a

$$\log p(\mathbf{y}, \mathbf{z} | \boldsymbol{\theta}, \Lambda, \Delta, T) = -\frac{\|\boldsymbol{\theta} - \mathbf{z}\|^2}{2\mu^2} + K(\mathbf{y}, \mathbf{z}),$$

où $K(\mathbf{y}, \mathbf{z})$ est une constante qui ne dépend pas de $\boldsymbol{\theta}$. On a aussi $\mathbf{z}^t = \mathbb{E}\{\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^t, \Lambda^t, \Delta^t, T\}$ où $p(\mathbf{z} | \mathbf{y}, \boldsymbol{\theta}^t, \Lambda^t, \Delta^t)$ est une loi normale dont la moyenne donne :

$$\hat{\mathbf{z}}^t = \boldsymbol{\theta}^t + \frac{\mu^2}{\sigma_0^2} \boldsymbol{\Phi}^* (\mathbf{y} - \boldsymbol{\Phi} \boldsymbol{\theta}^t).$$

L'étape M ne peut se faire directement car l'estimation jointe de $(\boldsymbol{\theta}, \Lambda, \Delta)$ est difficile. On propose alors une estimation séquentielle. On estime d'abord les cartes Λ, Δ , par marginalisation puis maximisation :

$$\hat{\Lambda}^{t+1}, \hat{\Delta}^{t+1} = \arg \max_{\Lambda, \Delta} \mathbb{E}\{\log(p(\Lambda, \Delta | \mathbf{y}, \mathbf{z}, T)) | \mathbf{y}, \boldsymbol{\theta}^t, \Lambda^t, \Delta^t, T\},$$

puis les coefficients $\boldsymbol{\theta}$ par

$$\hat{\boldsymbol{\theta}}^{t+1} = \arg \max_{\boldsymbol{\theta}} \mathbb{E}\{\log(p(\boldsymbol{\theta} | \hat{\Lambda}^{t+1}, \hat{\Delta}^{t+1}, \mathbf{y}, \mathbf{z}, T)) | \mathbf{y}, \boldsymbol{\theta}^t, \Lambda^t, \Delta^t, T\}$$

3.2 Estimation des cartes

La marginalisation donne

$$p(\Lambda, \Delta | \mathbf{y}, \mathbf{z}, T) = \int p(\boldsymbol{\theta}, \Lambda, \Delta | \mathbf{z}, \mathbf{y}, T) d\boldsymbol{\theta},$$

avec

$$p(\boldsymbol{\theta}, \Lambda, \Delta | \mathbf{z}, \mathbf{y}, T) \propto \prod_{\lambda=1}^{N_1} \frac{e^{-\frac{|\alpha_\lambda - z_\lambda|^2}{\mu^2}}}{\pi \mu^2} \left(\frac{p e^{-\frac{|\alpha_\lambda|^2}{\sigma^2}}}{\pi \sigma^2} X_\lambda + (1-p)(1-X_\lambda) \right) \prod_{\delta=2}^{N_2} \frac{e^{-\frac{|\beta_\delta - z_\delta|^2}{\mu^2}}}{\pi \mu^2} \left(\frac{\tilde{p} e^{-\frac{|\beta_\delta|^2}{\tilde{\sigma}^2}}}{\pi \tilde{\sigma}^2} \tilde{X}_\delta + (1-\tilde{p})(1-\tilde{X}_\delta) \right).$$

ce qui donne comme estimation pour la carte Λ (avec une estimation similaire pour Δ) :

$$p(X_\lambda = 1 | \mathbf{z}, \mathbf{y}, T) = \frac{p}{\pi(\mu^2 + \sigma^2)} e^{-\frac{z_\lambda^2}{(\mu^2 + \sigma^2)}}$$

$$p(X_\lambda = 0 | \mathbf{z}, \mathbf{y}, T) = \frac{1-p}{\pi\mu^2} e^{-\frac{z_\lambda^2}{\mu^2}}.$$

On obtient alors l'estimateur du MAP pour X_λ , par un simple seuillage sur z_λ^t :

$$\hat{X}_\lambda^{t+1} = \begin{cases} 1 & \text{si } |z_\lambda^t| > \sqrt{\frac{(\mu^2 + \sigma^2)\mu^2}{\sigma^2} \ln \left[\frac{1-p}{p} \frac{\mu^2 + \sigma^2}{\mu^2} \right]} \\ 0 & \text{sinon.} \end{cases}$$

Remarque 1 Dans le cas d'un modèle de carte qui regroupe des coefficients, on obtient une règle similaire de seuillage sur la norme des groupes. On suppose alors que la carte de signifiante peut s'exprimer à l'aide de deux indices : $\Delta = \Delta_t \times \Delta_f$. En introduisant une variable indicatrice à deux indices $X_{t,f}$ telle que pour tout f , $p(X_{t,f}) = p$ si $t \in \Delta_t$ et $p(X_{t,f}) = 1-p$ sinon, la règle s'écrit :

$$\hat{X}_{t,f}^{t+1} = \begin{cases} 1 & \text{si } \|\hat{z}_t^t\|^2 > \frac{|\Delta_f|(\mu^2 + \sigma^2)\mu^2}{\sigma^2} \ln \left[\frac{1-p}{p} \frac{\mu^2 + \sigma^2}{\mu^2} \right] \\ 0 & \text{sinon.} \end{cases}$$

3.3 Estimation de θ

Une fois les cartes estimées, l'estimation des θ^{t+1} devient une simple pondération des coefficients non nuls : si $\hat{X}_\lambda^{t+1} = 1$, alors

$$\hat{\theta}_\lambda^{t+1} = \arg \max_{\theta_\lambda} \frac{\|\hat{z}_\lambda^t - \theta_\lambda\|^2}{2\mu^2} + \frac{\|\theta_\lambda\|^2}{2\sigma^2} = \frac{\hat{z}_\lambda^t}{1 + \mu^2/\sigma^2}, \quad (8)$$

et $\hat{\theta}_\lambda^{t+1} = 0$ sinon, avec une règle similaire pour \hat{X}_δ^{t+1} .

3.4 Estimation des paramètres

Maintenant que les cartes de significances et que les coefficients sont connus, on peut estimer les paramètres du modèle $T = \{p, \sigma^2, \tilde{p}, \tilde{\sigma}^2, \sigma_0^2\}$ en utilisant un *a priori* non informatif de Jeffreys (voir par exemple [8]) sur les paramètres d'échelle $\sigma^2, \tilde{\sigma}^2, \sigma_0^2$, et une loi uniforme sur les probabilités d'appartenance p et \tilde{p} . On trouve alors les estimateurs statistiques classiques :

$$\hat{\sigma}^{2^{t+1}} = \frac{\sum_{\lambda \in \Lambda} \hat{X}_\lambda^{t+1} |\hat{\alpha}_\lambda^{t+1}|^2}{|\hat{\Lambda}^{t+1}|}, \quad \hat{\tilde{\sigma}}^{2^{t+1}} = \frac{\sum_{\delta \in \Delta} \hat{X}_\delta^{t+1} |\hat{\beta}_\delta^{t+1}|^2}{|\hat{\Delta}^{t+1}|}$$

$$\hat{\sigma}_0^{2^{t+1}} = \frac{\|\mathbf{y} - \Phi \hat{\theta}^{t+1}\|^2}{N}, \quad \hat{p}^{t+1} = \frac{|\hat{\Lambda}^{t+1}|}{N_1}, \quad \hat{\tilde{p}}^{t+1} = \frac{|\hat{\Delta}^{t+1}|}{N_2}$$

où $|\Lambda^{t+1}|$ et $|\Delta^{t+1}|$ représentent le cardinal des cartes de signifiante estimées.

3.5 Algorithme

L'algorithme 1, que l'on appellera Ad-ISTA-BG pour Adaptive Iterative Shrinkage/Thresholding Algorithm with a Bernoulli-Gaussian prior, résume les étapes précédentes.

Algorithme 1 : AdISTA-BG

répéter

Étape E : $\mathbf{z} = \boldsymbol{\theta} + \frac{\mu}{\sigma_0^2} \Phi^*(\mathbf{y} - \Phi \boldsymbol{\theta})$

Étape M : estimation des cartes de signifiante :

$$\forall \lambda, X_\lambda = \begin{cases} 1 & \text{si } |z_\lambda| > \sqrt{\frac{(\mu^2 + \sigma^2)\mu^2}{\sigma^2} \ln \left[\frac{1-p}{p} \frac{\mu^2 + \sigma^2}{\mu^2} \right]} \\ 0 & \text{sinon} \end{cases}$$

$$\forall \delta, \tilde{X}_\delta = \begin{cases} 1 & \text{si } |z_\delta| > \sqrt{\frac{(\mu^2 + \tilde{\sigma}^2)\mu^2}{\tilde{\sigma}^2} \ln \left[\frac{1-\tilde{p}}{\tilde{p}} \frac{\mu^2 + \tilde{\sigma}^2}{\mu^2} \right]} \\ 0 & \text{sinon} \end{cases}$$

estimation des coefficients :

$$\forall \lambda, \alpha_\lambda = \frac{X_\lambda z_\lambda}{1 + \frac{\mu^2}{\sigma^2}}; \quad \forall \delta, \beta_\delta = \frac{\tilde{X}_\delta z_\delta}{1 + \frac{\mu^2}{\tilde{\sigma}^2}}$$

estimation des paramètres :

$$\sigma^2 = \frac{\sum_{\lambda \in \Lambda} X_\lambda |\alpha_\lambda|^2}{|\Lambda|} \quad \tilde{\sigma}^2 = \frac{\sum_{\delta \in \Delta} \tilde{X}_\delta |\beta_\delta|^2}{|\Delta|}$$

$$\sigma_0^2 = \frac{\|\mathbf{y} - \Phi \boldsymbol{\theta}\|^2}{N} \quad p = \frac{|\Lambda|}{N_1} \quad \tilde{p} = \frac{|\Delta|}{N_2}$$

jusqu'à convergence ;

4 Résultats

On décompose un signal jouet de xylophone, bruité artificiellement, dans une union de deux dictionnaires de Gabor qui se différencient par la taille de leur fenêtre d'analyse. On choisit respectivement une taille de fenêtre de 2048 échantillons ($\simeq 46$ ms) pour la partie tonale et de 128 échantillons ($\simeq 3$ ms) pour la partie transitoire. On choisit de plus un *a priori* Bernoulli-Gaussien en groupe pour la partie transitoire, afin de retrouver des transitoires bien localisés en temps. La décomposition obtenue est donnée sur la figure 4. On retrouve en sortie de l'algorithme une estimation de $\hat{\sigma}_0^2 = 0.001$ pour $\sigma_0^2 = 0.0009$. En terme de débruitage, le rapport signal à bruit obtenu en sortie est de 16 dB, pour un signal d'entrée bruité à 12 dB.

5 Conclusion

L'algorithme proposé semble une bonne alternative aux MCMC pour ce modèle Bernoulli-Gaussien. Il permet d'éviter les problèmes de combinatoire qui apparaissent rapidement avec ce type de modèle. De plus son implémentation simple reste dans

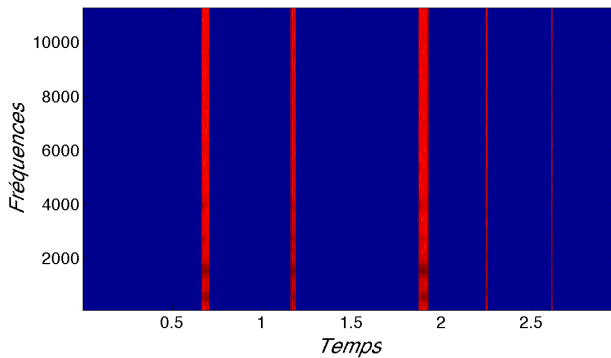
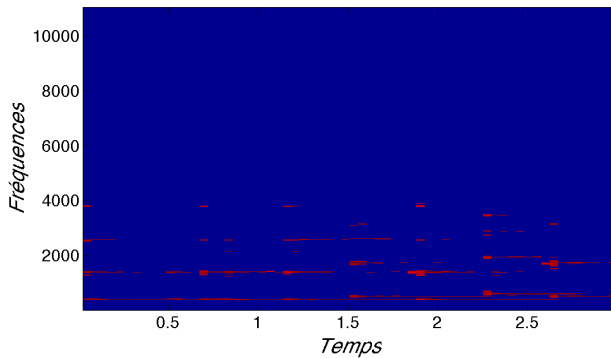
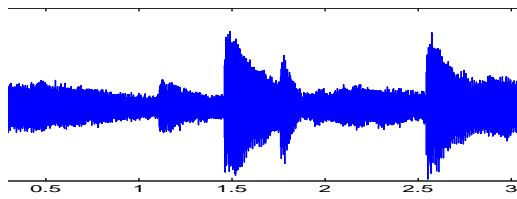


FIGURE 1 – Image temps-fréquence des couches estimées. Haut : signal de xylophone. Milieu : tonal. Droite : Transitoire avec une structure groupée.

la famille des algorithmes de type « seuillages itératifs ». Son caractère non supervisé permet d’obtenir une décomposition en couche acceptable, sans avoir à intervenir sur le réglage souvent fastidieux des hyper-paramètres.

Cependant, plusieurs remarques sont à faire. Tout d’abord, bien qu’exploitant le cadre GEM, seul la convergence vers un maximum local de la vraisemblance est assurée ; rien ne peut être conclu, *a priori*, sur la convergence des itérés en raison de la discontinuité introduite par le modèle de Bernoulli. Une autre remarque pratique importante est à faire sur le paramètre μ intervenant dans la loi de \mathbf{n}_2 (voir l’équation (7)). Ce paramètre, qui joue le rôle de la constante de Lipschitz dans les algorithmes de seuillage itératifs classiques [1], influe beaucoup sur le comportement de l’algorithme lorsque des repères de Gabor sont utilisés. Le réglage simple $\mu = \frac{\sigma_0}{\|\Phi\Phi^*\|}$, qui donne des résultats satisfaisant avec une union de bases MDCT, fait

converger l’algorithme vers un résultat trivial avec une union de repère de Gabor. Ce comportement peut s’expliquer par le fait que l’union de deux bases MDCT donne une matrice de dictionnaire peut corrélé et bien conditionnée, propriété que l’on perd alors avec les repères de Gabor. Il faut alors fixer ce μ à une valeur faible pour obtenir une décomposition satisfaisante, ce qui influe fortement sur la rapidité de l’algorithme.

Références

- [1] A. Beck and M. Teboulle. Fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2 :183 – 202, 2009.
- [2] L. Daudet and B. Torrèsani. Hybrid representations for audiophonic signal encoding. *Signal Processing*, 82(11) :1595–1617, 2002. Special issue on Image and Video Coding Beyond Standards.
- [3] C. Févotte, L. Daudet, S. J. Godsill, and B. Torrèsani. Sparse linear regression with structured priors and application to denoising of musical audio. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1) :174–185, Jan. 2008.
- [4] M. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8) :906–916, 2003.
- [5] M. Kowalski and B. Torrèsani. Random models for sparse signals expansion on unions of basis with application to audio signals. *IEEE Transactions On Signal Processing*, 56(8) :3468–3481, Aug. 2008.
- [6] S. Molla and B. Torrèsani. An hybrid audio scheme using hidden Markov models of waveforms. *Applied and Computational Harmonic Analysis*, 18(2) :137–166, 2005.
- [7] B. Picinbono. Second-order complex random vectors and normal distributions. *IEEE Transactions on Signal Processing*, 4(10) :2637—2640, 1996.
- [8] C. Robert. *The Bayesian Choice : A Decision Theoretic Motivation*. New York : Springer-Verlag, 1994.
- [9] J.-L. Starck, M. Elad, and D. L. Donoho. Image decomposition via the combination of sparse representation and a variational approach. *IEEE Transactions on Image Processing*, 14(10) :1570–1582, 2005.