

Méthodes MCMC pour la Décomposition en Paquets d'Ondelettes de Signaux Transitoires

D. Leporini

Laboratoire des Signaux et Systèmes,
CNRS/UPS et GDR-PRC ISIS, ESE,
91192 Gif-sur-Yvette Cédex, France.

RÉSUMÉ

Les bases de paquets d'ondelettes fournissent un cadre approprié à la recherche d'une représentation optimale des processus stochastiques. Dans cet article, nous abordons le problème du choix de la "meilleure base" sous un angle bayésien, lorsque l'observation est un signal transitoire bruité, en considérant cette base optimale comme l'un des paramètres du modèle. Des distributions *a priori* non-homogènes sont introduites sur les coefficients de décomposition du signal, nécessitant la mise en œuvre d'algorithmes de Monte Carlo par chaînes de Markov à sauts réversibles. Diverses méthodes d'estimation, pouvant dépendre de plusieurs représentations, sont évaluées au travers de simulations.

ABSTRACT

Wavelet packets provide a natural framework to answer the problem of finding a "best representation" of stochastic signals. In this paper, we present a Bayesian approach to this problem when dealing with noisy transient signals by considering the basis itself as a model parameter. Prior non-homogeneous models for the underlying signal and noise coefficients are introduced and make it compelling to resort to recently introduced reversible jump Markov chain Monte Carlo algorithms. Various nonlinear signal estimates, which may depend on several representations, are subsequently obtained, and presented through several simulations.

1 Introduction

L'optimisation des représentations multi-échelles de signaux bruités, basée sur des approches statistiques, a suscité au cours de ces dernières années de nombreux travaux [8, 6, 9]. Cette optimisation s'effectue généralement sur un dictionnaire de bases orthonormées, et vise à discriminer de manière statistique les composantes du signal (déterministe) d'intérêt $x(t)$ de celles du bruit. Dans un certain nombre d'applications, il devient cependant essentiel de considérer le signal $x(t)$ comme un processus aléatoire, et de prendre en considération toute l'information *a priori* disponible. Il convient de noter que cette approche bayésienne a été précédemment proposée dans le cadre des décompositions en ondelettes [1, 2], et étendue dans [9] aux décompositions en paquets d'ondelettes. Cependant, la base optimale n'a, à notre connaissance, jamais été considérée comme l'un des objets de l'inférence bayésienne. Dans cet article, nous présentons une méthode permettant d'obtenir, dans le cas de signaux transitoires bruités, une description statistique de la base optimale au travers d'une estimation de sa distribution *a posteriori*. Cette approche générale présente deux avantages :

- Elle caractérise tout d'abord de manière statistique la représentation du signal observé.
- Elle permet ensuite la construction d'estimateurs basés sur l'ensemble du dictionnaire.

2 Position du problème

On considère un dictionnaire fini $\mathcal{D} = \{\mathcal{B}_1, \dots, \mathcal{B}_N\}$ de bases de décompositions possibles et l'on note $\mathbf{y}^{\mathcal{B}}$ le vecteur de dimension K représentant les échantillons du signal observé

$y(t)$ dans la base \mathcal{B} . Les décompositions considérées peuvent alors être schématisées sous la forme d'arbres binaires dont chaque nœud (j, m) (avec $j \in \{0, \dots, J\}$ et $m \in \{0, \dots, 2^j - 1\}$), correspond à une base $\mathcal{B}_{j,m}$ d'un sous-espace vectoriel des séquences à support compact dans $\{1, \dots, K\}$. On considère pour l'observation le modèle additif suivant

$$y(t) = x(t) + b(t), \quad t \in \{1, \dots, K\},$$

dans lequel le bruit $b(t)$ n'est pas supposé *a priori* i.i.d. gaussien. Ce modèle se démarque du cadre classique, même si des algorithmes d'estimation non-linéaire, étendant les méthodes introduites dans [3] au cas de bruits non nécessairement i.i.d. gaussiens, ont récemment été développés dans le cadre des décompositions en ondelettes [5]. On impose en réalité que le rapport du logarithme des polyspectres du bruit sur leur ordre soit uniformément borné, ce qui revient à limiter l'écart à la gaussianité et introduit des contraintes de mélange. Sous ces hypothèses, on montre que les coefficients de décomposition $\mathbf{b}^{\mathcal{B}_{j,m}}$ de $b(t)$ deviennent asymptotiquement i.i.d. gaussiens lorsque le niveau de résolution j tend vers l'infini [7]. Cette propriété est à la base de notre modèle.

3 Représentation optimale

La distribution de $\mathbf{y}^{\mathcal{B}^*}$, représentant le signal $y(t)$ dans la "meilleure base", est donnée par un modèle non-homogène de mélanges de gaussiennes

$$p(\mathbf{y} | \mathcal{B}_{j,m}^*, \boldsymbol{\theta}_{j,m}^*) = \prod_{k=1}^{K2^{-j}} \left[(1 - \varepsilon_{j,m}^*) g(\mathbf{y}^{\mathcal{B}_{j,m}^*}[k] | \sigma_{j,m}^{*2}) + \varepsilon_{j,m}^* g(\mathbf{y}^{\mathcal{B}_{j,m}^*}[k] | \tilde{\sigma}_{j,m}^{*2}) \right], \quad (1)$$

où $g(\cdot|s^2)$ représente la loi normale $\mathcal{N}(0, s^2)$, et $\boldsymbol{\theta}_{j,m} = [\varepsilon, \sigma^2, \tilde{\sigma}^2]_{j,m}$ désigne le vecteur de paramètres au nœud (j, m) . Ce modèle traduit une parsimonie locale des coefficients de $x(t)$ dans \mathcal{B}^* , qui s'avère bien adaptée aux phénomènes transitoires. Il convient par contre de rappeler que lorsque le bruit est i.i.d. gaussien et centré, ses coefficients $\mathbf{b}^{\mathcal{B}_{j,m}}$ demeurent i.i.d. gaussiens de même variance. Dans ce cas, une régularisation plus efficace peut être effectuée en imposant une variance de bruit commune à tous les nœuds de la décomposition. Un vecteur de variables indicatrices cachées $\mathcal{Q}_{j,m}$ est également introduit, définissant les probabilités conditionnelles

$$\begin{aligned} P_{Y_{j,m}[k]|Q_{j,m}[k]=0}(y_{j,m}[k]) &= g(y_{j,m}[k] | \sigma_{j,m}^2) \\ P_{Y_{j,m}[k]|Q_{j,m}[k]=1}(y_{j,m}[k]) &= g(y_{j,m}[k] | \tilde{\sigma}_{j,m}^2), \end{aligned}$$

et vérifiant $P(Q_{j,m}[k] = 1) = \varepsilon_{j,m} \in [0, 1]$. Le modèle (1) revient alors à considérer les variables binaires $(Q_{j,m}[k])_{1 \leq k \leq 2^j K}$ comme indépendantes. D'autres hypothèses, permettant par exemple de prendre en considération la corrélation entre les coefficients, ainsi que d'autres distributions paramétriques, pourraient néanmoins être choisies.

4 Distributions *a priori*

Dans une approche bayésienne, la base optimale \mathcal{B}^* et le vecteur de paramètres $\boldsymbol{\theta}_{j,m}$ sont considérés comme des variables aléatoires, et des distributions *a priori* doivent être précisées. Sans information supplémentaire, la loi sur l'arborescence est considérée comme uniforme sur le niveau de résolution. Des *a priori* favorisant certaines décompositions peuvent toutefois être introduits dans le cas de bruits corrélés. Nous utilisons, pour les paramètres, des distributions *a priori* conjuguées à partir de la vraisemblance (1), afin de mettre en œuvre un échantillonneur de Gibbs basé sur l'algorithme d'augmentation de données [10]. Une loi *a priori* beta $\mathcal{Be}(1, 3)$ est placée sur $\varepsilon_{j,m}$ afin de traduire la parsimonie de la représentation, et des lois inverse gamma $\mathcal{IG}(1, 1/2V_{j,m})$ sont choisies comme distributions *a priori* pour $\tilde{\sigma}_{j,m}^2$ et $\sigma_{j,m}^2$ sous la contrainte $\sigma_{j,m}^2 \leq \tilde{\sigma}_{j,m}^2$, où $V_{j,m}$ désigne une estimée robuste de la variance du bruit $\mathbf{b}^{\mathcal{B}_{j,m}}$. La dimension de l'espace des paramètres n'étant pas fixée – en raison du caractère non-homogène du modèle –, il apparaît nécessaire de recourir aux algorithmes de Monte Carlo par chaînes de Markov (MCMC) à sauts réversibles [4].

5 Présentation de l'algorithme

Afin d'obtenir des représentations arborescentes et des vecteurs de paramètres qui admettent $p(\mathcal{B}, \boldsymbol{\theta}|\mathbf{y})$ comme loi jointe, la base temporelle, ainsi qu'un vecteur de paramètres $\boldsymbol{\theta}^{(0)}$, définissent la représentation initiale. Pour accélérer la convergence, des algorithmes rapides de recherche de base optimale peuvent toutefois fournir une initialisation pertinente. Nous définissons ensuite les mouvements possibles de la chaîne de Markov : (a) une modification du vecteur de paramètres $\boldsymbol{\theta}_{j,m}$ où (j, m) désigne un nœud terminal, (b) la naissance de nœuds terminaux (*i.e.* une nouvelle décomposition) et (c) la suppression de nœuds terminaux (*i.e.* une recomposition). La transition (a) peut être développée à l'aide d'algorithmes

MCMC standards (échantillonneur de Gibbs, algorithmes de Metropolis-Hastings ou algorithmes hybrides), mais les transitions (b) et (c) induisent un changement dans la dimension de l'espace des paramètres. Dans le modèle (1), à titre d'exemple, la dimension est augmentée de 3 ($\boldsymbol{\theta}_{j,m} \in [0, 1] \times \mathbb{R}_+^2$) dans le cas d'une nouvelle décomposition. La bijection entre ces espaces est établie à l'aide d'un vecteur aléatoire \mathbf{u} i.i.d. $\mathcal{Be}(3, 3)$ de dimension 3. Plus précisément, les nouveaux vecteurs de paramètres $\boldsymbol{\theta}_{j+1,2m}$ et $\boldsymbol{\theta}_{j+1,2m+1}$ sont proposés selon

$$\begin{aligned} \varepsilon_{j+1,2m} &= 2u_1\varepsilon_{j,m}, \\ \varepsilon_{j+1,2m+1} &= 2(1-u_1)\varepsilon_{j,m}, \\ (1-\varepsilon_{j+1,2m})\sigma_{j+1,2m}^2 &= 2u_2(1-\varepsilon_{j,m})\sigma_{j,m}^2, \\ (1-\varepsilon_{j+1,2m+1})\sigma_{j+1,2m+1}^2 &= 2(1-u_2)(1-\varepsilon_{j,m})\sigma_{j,m}^2, \\ \varepsilon_{j+1,2m}\tilde{\sigma}_{j+1,2m}^2 &= 2u_3\varepsilon_{j,m}\tilde{\sigma}_{j,m}^2, \\ \varepsilon_{j+1,2m+1}\tilde{\sigma}_{j+1,2m+1}^2 &= 2(1-u_3)\varepsilon_{j,m}\tilde{\sigma}_{j,m}^2. \end{aligned}$$

Ces équations garantissent la conservation du nombre moyen de singularités ainsi que de l'énergie pour les deux composantes du mélange, d'après

$$\begin{aligned} \varepsilon_{j,m} &= \frac{\varepsilon_{j+1,2m} + \varepsilon_{j+1,2m+1}}{2}, \\ \varepsilon_{j,m}\tilde{\sigma}_{j,m}^2 &= \frac{\varepsilon_{j+1,2m}\tilde{\sigma}_{j+1,2m}^2}{2} \\ &+ \frac{\varepsilon_{j+1,2m+1}\tilde{\sigma}_{j+1,2m+1}^2}{2}, \\ (1-\varepsilon_{j,m})\sigma_{j,m}^2 &= \frac{(1-\varepsilon_{j+1,2m})\sigma_{j+1,2m}^2}{2} \\ &+ \frac{(1-\varepsilon_{j+1,2m+1})\sigma_{j+1,2m+1}^2}{2}. \end{aligned}$$

La probabilité d'acceptation de cette transition est alors donnée par

$$\alpha(s, s') = \min \left\{ 1, \frac{p(s' | \mathbf{y}) j_c(s')}{p(s | \mathbf{y}) j_b(s) p_U(\mathbf{u})} \left| \frac{\partial s'}{\partial(s, \mathbf{u})} \right| \right\}, \quad (2)$$

où s désigne l'état courant $(\mathcal{B}_{j,m}, \boldsymbol{\theta}_{j,m}, \mathbf{q}_{j,m})$, s' représente l'état cible $(\mathcal{B}_{j+1,2m}, \mathcal{B}_{j+1,2m+1}, (\boldsymbol{\theta}_{j+1,2m}, \boldsymbol{\theta}_{j+1,2m+1}), (\mathbf{q}_{j+1,2m}, \mathbf{q}_{j+1,2m+1}))$, et $j_m(s)$ correspond à la probabilité de choisir le mouvement m dans l'état s . La probabilité d'acceptation pour la recomposition correspondante est simplement obtenue en inversant les rapports dans (2). On vérifie que la chaîne ainsi générée est apériodique puisque tout voisinage de l'état courant est atteint avec une probabilité positive après une transition. Elle est également irréductible puisque le vecteur $\boldsymbol{\theta}_{j,m}$ est échantillonné à l'aide de distributions continues sur l'espace des paramètres, et que la chaîne peut passer d'une base quelconque à une autre, en modifiant un nœud à la fois.

6 Estimation bayésienne

L'échantillonneur présenté précédemment fournit une description statistique de la base optimale sous la forme d'une suite $(\mathcal{B}_k, \boldsymbol{\theta}_k)_{k \in \{1, \dots, N_p\}}$ de paramètres du modèle tirés selon leur loi *a posteriori* jointe, après convergence de la chaîne vers sa distribution stationnaire. Diverses méthodes d'estimation peuvent alors être proposées.

6.1 Maximum a posteriori

Les estimés des paramètres peuvent être obtenus par maximum *a posteriori* sur la distribution empirique

$$(\widehat{\mathcal{B}}^*, \widehat{\boldsymbol{\theta}}^*) = \arg \max_{\mathcal{B}, \boldsymbol{\theta}} p_{\mathcal{B}^*, \boldsymbol{\theta}^* | Y=y}(\mathcal{B}, \boldsymbol{\theta}).$$

Le signal $\mathbf{x}^{\mathcal{B}_{j,m}}$, pour $\mathcal{B}_{j,m} \subset \widehat{\mathcal{B}}^*$, est alors estimé par maximum *a posteriori* marginal, qui s'écrit

$$\widehat{x}_{j,m}[k](y_{j,m}[k], \widehat{\boldsymbol{\theta}}_{j,m}^*) = \begin{cases} \frac{\widehat{\sigma}_{j,m}^2 - \widehat{\sigma}_{j,m}^2}{\widehat{\sigma}_{j,m}^2} y_{j,m}[k] & \text{si } |y_{j,m}[k]| \geq \chi_{j,m}, \\ 0 & \text{si } |y_{j,m}[k]| < \chi_{j,m}, \end{cases}$$

où la valeur du seuil $\chi_{j,m} \geq 0$ est donnée par

$$\chi_{j,m}^2 = \max \left\{ \frac{2\widehat{\sigma}_{j,m}^2 \widehat{\sigma}_{j,m}^2}{\widehat{\sigma}_{j,m}^2 - \widehat{\sigma}_{j,m}^2} \ln \left(\frac{\widehat{\sigma}_{j,m}^2 (1 - \widehat{\varepsilon}_{j,m})}{\widehat{\sigma}_{j,m}^2 \widehat{\varepsilon}_{j,m}} \right), 0 \right\}.$$

6.2 Moyenne a posteriori

Il est possible d'estimer la moyenne *a posteriori* du signal $\mathbb{E}[\mathbf{x} | Y = \mathbf{y}]$, en moyennant les réalisations des paramètres du modèle selon

$$\mathbb{E}[\mathbf{x} | Y = \mathbf{y}] \approx \frac{1}{N_p} \sum_{k=1}^{N_p} \mathbb{E}[\mathbf{x} | \mathcal{B}^* = \mathcal{B}_k, \boldsymbol{\theta}^* = \boldsymbol{\theta}_k, Y = \mathbf{y}],$$

et ainsi de prendre en compte l'ensemble des représentations possibles. Cependant, cette estimation peut s'avérer extrêmement coûteuse dans le cas de distributions *a posteriori* complexes, et l'on préférera utiliser l'estimation

$$\widehat{\mathbf{x}}(Y) = \frac{1}{N_p} \sum_{k=1}^{N_p} \mathbb{E}[\mathbf{x} | \mathcal{B}^* = \mathcal{B}_k, \boldsymbol{\theta}^* = \widehat{\boldsymbol{\theta}}^*(\mathcal{B}_k), Y = \mathbf{y}],$$

dans laquelle

$$\{\widehat{\boldsymbol{\theta}}^*(\mathcal{B}_k)\} = \cup_{(j,m) \in \mathcal{B}_{j,m} \subset \mathcal{B}_k} \{\widehat{\boldsymbol{\theta}}_{j,m}^*\},$$

représente l'ensemble des estimées $\widehat{\boldsymbol{\theta}}_{j,m}^*$ au sens du maximum *a posteriori*. Dans le cas du modèle (1), l'expression de la moyenne *a posteriori*, connaissant les paramètres, est donnée par

$$\widehat{x}_{j,m}[k](y_{j,m}[k], \widehat{\boldsymbol{\theta}}_{j,m}^*) = \frac{\widehat{\varepsilon}_{j,m} (\widehat{\sigma}_{j,m}^2 - \widehat{\sigma}_{j,m}^2) g(y_{j,m}[k] | \widehat{\sigma}_{j,m}^2)}{\widehat{\sigma}_{j,m}^2 p_{Y_{j,m}[k] | \boldsymbol{\theta}_{j,m}^* = \widehat{\boldsymbol{\theta}}_{j,m}^*}(y_{j,m}[k])} y_{j,m}[k].$$

7 Simulations

7.1 Signaux synthétiques

Pour présenter les performances de l'algorithme proposé, on considère, dans un premier temps, un signal synthétique de $K = 1024$ échantillons correspondant à la somme des réalisations de deux processus Bernoulli-gaussien non-homogènes sur des bases distinctes, $\mathcal{B}_1^* = \mathcal{B}_{2,0} \oplus \mathcal{B}_{2,1} \oplus \mathcal{B}_{2,2} \oplus \mathcal{B}_{2,3}$ et

$\mathcal{B}_2^* = \mathcal{B}_{2,0} \oplus \mathcal{B}_{2,1} \oplus \mathcal{B}_{2,2} \oplus \mathcal{B}_{3,6} \oplus \mathcal{B}_{3,7}$. Le bruit additif est i.i.d. gaussien de variance unité et conduit à un rapport signal à bruit de 6.5 dB, mais nous rappelons que les propriétés statistiques du bruit ne sont pas supposées connues. La distribution empirique de la représentation optimale est présentée sur la Fig. 1 et reflète l'ambiguïté entre les bases \mathcal{B}_1^* et \mathcal{B}_2^* .

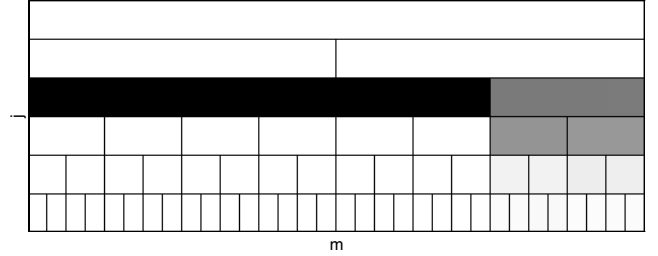


FIG. 1 — Distribution *a posteriori* empirique de la base optimale.

Une recherche de base optimale de type Minimum Description Length (MDL) [6], où la variance du bruit est connue, a également été développée. Les différents signaux sont présentés sur la Fig. 2, ainsi que l'erreur quadratique moyenne normalisée (EQMN) des signaux estimés, définie par

$$\text{EQMN} = \frac{\sum_{k=1}^K |x[k] - \widehat{x}[k]|^2}{\|\mathbf{x}\|_2^2}.$$

Nous considérons ensuite un signal doppler transitoire auquel est ajouté un bruit i.i.d. uniforme $\mathcal{U}[-0.5, 0.5]$. La méthode proposée est comparée à l'approche basée sur le seuillage dépendant du niveau de résolution des coefficients d'ondelettes [5]. L'ensemble des signaux est présenté sur la Fig. 4.

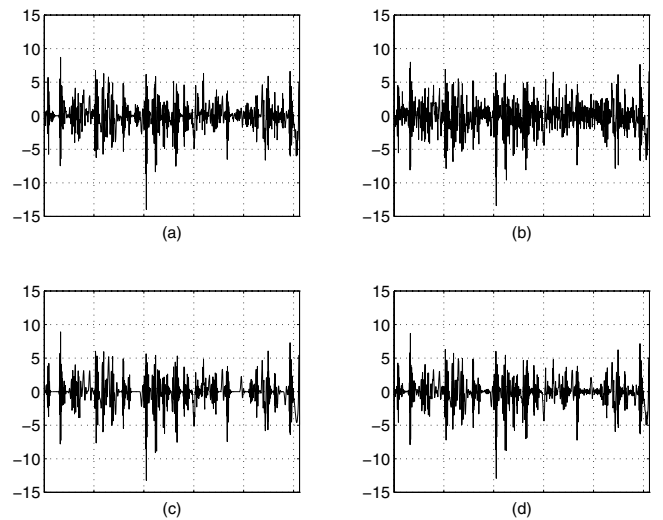


FIG. 2 — (a) signal original, (b) signal bruité, (c) signal estimé par MDL (EQMN = 0.14), et (d) moyenne *a posteriori* (EQMN = 0.07).

7.2 Signaux réels

Nous présentons dans cette partie les résultats obtenus avec un signal d'acoustique sous-marine de $K = 2048$ échantillons correspondant à des clics de crevettes dans un environnement gaussien corrélé. Le signal observé, l'estimé au sens du maximum *a posteriori*, ainsi que des détails de ces séquences sont présentés sur la Fig. 3.

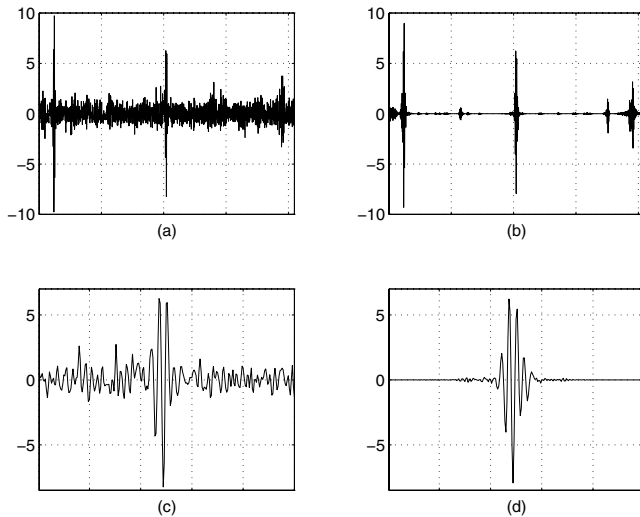


FIG. 3 — (a) signal observé, (b) signal estimé au sens du maximum *a posteriori*, (c) et (d) détails des signaux (a) et (b) respectivement.

8 Conclusion

Dans cet article, le problème de la représentation optimale des signaux stochastiques a été entièrement abordé sous un angle bayésien. Cette approche permet d'obtenir une description statistique du dictionnaire de bases considéré, et autorise la construction d'estimateurs basés sur l'ensemble des représentations.

Remerciements

L'auteur tient à exprimer ses remerciements à M. Di Martino (DCN/CERDSM) pour la mise à disposition des signaux biologiques d'acoustique sous-marine.

Références

- [1] H.A. Chipman, E.D. Kolaczyk, and R.E. McCulloch. Adaptive Bayesian Wavelet Shrinkage. Preprint, University of Chicago, 1996.
- [2] M. Clyde, G. Parmigiani, and B. Vidakovic. Multiple Shrinkage and Subset Selection in Wavelets. Preprint, Duke University, 1996.
- [3] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81 :425–455, Sept. 1994.

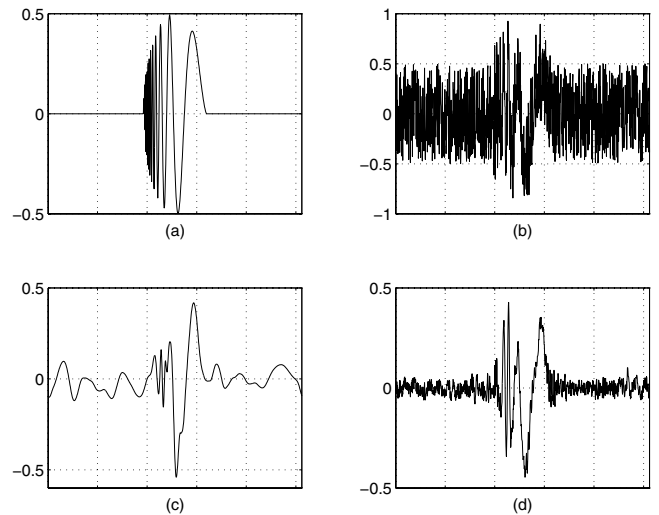


FIG. 4 — (a) signal original, (b) signal bruité, (c) signal estimé par seuillage des coefficients d'ondelettes (EQMN = 0.48), et (d) moyenne *a posteriori* (EQMN = 0.35).

- [4] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82 :711–732, 1995.
- [5] I.M. Johnstone and B.W. Silverman. Wavelet threshold estimators for data with correlated noise. Technical Report, Department of Statistics, Stanford University, 1994.
- [6] H. Krim and J.-C. Pesquet. On the statistics of best bases criteria. In A. Antoniadis, editor, *Wavelets and statistics*. Lecture Notes in Statistics, Springer Verlag, 1995.
- [7] D. Leporini and J.-C. Pesquet. High-Order Properties of M-Band Wavelet Packet Decompositions. In *IEEE Workshop on High-Order Statistics*, Banff, Canada, July 1997.
- [8] P. Moulin. A wavelet regularization method for diffuse radar target imaging and speckle noise reduction. *Journ. Math. Imaging and Vision*, 3 :123–134, 1993.
- [9] J.-C. Pesquet, H. Krim, D. Leporini, and E. Hamman. Bayesian Approach to Best Basis Selection. In *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, pages 2634–2638, Atlanta, USA, May 7-9 1996.
- [10] C.P. Robert. *Mixtures of distributions : inference and estimation*. Chapman and Hall, in Practical Markov chain Monte Carlo, pp. 441-464, London, 1996.