Débruitage de signaux à l'aide d'ondelettes

par Béatrice Pesquet-Popescu,

Maître de Conférences à l'Ecole Nationale Supérieure des Télécommunications de Paris.

Jean-Christophe Pesquet,

Professeur à l'Université de Marne la Vallée.

et Eric Moulines

Professeur à l'Ecole Nationale Supérieure des Télécommunications de Paris.

Plan

| 1 | Gér | néralités | 3 | |
|----------|------------------|--|----|--|
| 2 | Approche minimax | | | |
| | 2.1 | Notion d'oracle | 4 | |
| | 2.2 | Estimateurs par seuillage | 6 | |
| | 2.3 | L'optimalité du seuillage | 9 | |
| | 2.4 | Bruit de variance inconnue | 10 | |
| 3 | La méthode SURE | | | |
| | 3.1 | Estimer le risque sans biais | 12 | |
| | 3.2 | Application aux coefficients d'ondelettes | 13 | |
| | 3.3 | Algorithme de recherche du seuil optimal | 14 | |
| | 3.4 | Une assurance complémentaire (tout risque!) | 15 | |
| | 3.5 | Exemple de débruitage d'un signal Doppler | 16 | |
| | 3.6 | Comparaison des méthodes en débruitage d'images | 18 | |
| 4 | Ave | ec un peu d' <i>a priori</i> | 19 | |
| | 4.1 | Le point de vue bayésien | 19 | |
| | 4.2 | Quelques bons a priori sur les coefficients d'ondelettes | 23 | |
| | 4.3 | Détermination des hyperparamètres | 27 | |

| 5 | Extensions | 27 |
|--------------|--|----|
| \mathbf{A} | Preuve de la Proposition 1 | 28 |
| В | Simplification de l'équation intégrale (6) | 36 |
| \mathbf{C} | Preuve de la Proposition 2 | 37 |
| D | Preuve du Lemme 1 | 41 |
| ${f E}$ | Preuve de la Proposition 4 | 41 |
| \mathbf{F} | Preuve de la Proposition 5 | 42 |

1 Généralités

Les ondelettes et paquets d'ondelettes fournissent des moyens simples et efficaces d'estimer de façon non linéaire des signaux, en présence de bruit [11].

Partons de la situation assez commune où l'on observe un signal r(t) qui est la somme d'un signal utile inconnu s(t) et d'un bruit aléatoire b(t). Ces deux signaux seront supposés réels. Après décomposition sur une base d'ondelettes¹, on conserve un modèle additif :

$$c_j^r[k] = c_j^s[k] + c_j^b[k],$$

où les lettres r, s et b servent à distinguer les coefficients d'ondelettes des différents signaux mis en jeu. Par souci de simplicité, nous supposerons qu'une décomposition en ondelettes orthonormales périodiques (cf. paragraphe ??) est utilisée, de telle sorte que $1 \le j \le j_{\text{max}}$ et $0 \le k < K2^{-j}$, le signal (à la résolution j = 0) étant de longueur K, multiple de $2^{j_{\text{max}}}$. Pour les coefficients d'approximation au niveau de résolution j_{max} , le rapport signal sur bruit est généralement assez élévé pour qu'on puisse considérer que, pour $k \in \{0, \ldots, 2^{j_{\text{max}}} - 1\}$,

$$a_{j_{\max}}^s[k] \simeq a_{j_{\max}}^r[k]. \tag{1}$$

Le postulat sous-jacent dans les méthodes de débruitage à l'aide d'ondelettes est qu'un choix convenable de la base de décomposition facilite la discrimination du signal et du bruit. En effet, le signal est généralement assez "régulier" pour pouvoir être bien décrit par un petit nombre de coefficients d'ondelettes de grande amplitude. Au contraire, le bruit, souvent blanc et stationnaire, a tendance à s'équi-répartir sur toutes les composantes. Dans le paragraphe suivant, nous ferons l'hypothèse que b(t) est un bruit blanc, centré et de densité spectrale de puissance σ^2 . Sa décomposition dans toute base orthonormale conduit donc à une suite de coefficients centrés et décorrélés, de variance σ^2 . Si de plus, b(t) est gaussien, ces coefficients sont également gaussiens.

¹Certaines conditions techniques [3] sont nécessaires pour pouvoir définir les coefficients d'ondelettes d'un signal aléatoire.

2 Approche minimax

Les méthodes présentées dans cette partie sont le résultat d'une série de travaux récents de Donoho et Johnstone [8] qui ont été étendus par Averkamp et Houdré []. Les principes sur lesquels reposent ces approches minimax s'inscrivent cependant dans une tradition statistique beaucoup plus ancienne.

2.1 Notion d'oracle

Tout d'abord, considérons le signal s(t) comme déterministe, d'énergie finie. Une méthode simple pour estimer s(t), ou de façon équivalente ses coefficients d'ondelettes, consiste à produire le signal débruité $\widehat{s}(t)$ de coefficients d'ondelettes : $c_j^{\widehat{s}}[k] = \theta_j[k] \, c_j^r[k], \, \theta_j[k]$ étant un facteur d'atténuation à appliquer à chaque composante. L'idée est de préserver les coefficients peu bruités et de pouvoir atténuer, voire éliminer, les coefficients trop fortement bruités. En règle générale, on cherche à minimiser l'erreur quadratique moyenne (EQM) d'estimation, qui, compte tenu de l'hypothèse (1), s'écrit :

$$\mathcal{E}^{2}(s) = \mathbb{E}\{\|s - \hat{s}\|^{2}\} = \sum_{j=1}^{j_{\text{max}}} \sum_{k=0}^{K2^{-j}-1} \mathbb{E}\{(c_{j}^{s}[k] - c_{j}^{\hat{s}}[k])^{2}\}, \qquad (2)$$

où le symbole $E\{\cdot\}$ désigne l'espérance de la quantité aléatoire en argument. Les coefficients d'atténuation optimaux sont alors :

$$\theta_j[k] = \frac{(c_j^s[k])^2}{(c_j^s[k])^2 + \sigma^2} ,$$

et l'EQM de l'erreur vaut :

$$\mathcal{E}_a^2(s) = \sigma^2 \sum_{j=1}^{j_{\text{max}}} \sum_{k=0}^{K2^{-j}-1} \frac{(c_j^s[k])^2}{(c_j^s[k])^2 + \sigma^2} .$$

Démonstration. Intéressons-nous à l'un des termes apparaissant dans la sommation (2). Pour simplifier les notations, nous allons omettre les indices j et k. Cette simplification dans les écritures des formules pourra être reprise dans des démonstrations ultérieures.

Nous avons donc:

$$E\{(c^{s} - c^{\hat{s}})^{2}\} = E\{(c^{s} - \theta c^{r})^{2}\} = E\{(c^{s} - \theta (c^{s} + c^{b}))^{2}\}$$

$$= (1 - \theta)^{2}(c^{s})^{2} - 2(1 - \theta)\theta c^{s}E\{c^{b}\} + \theta^{2}E\{(c^{b})^{2}\}$$

$$= (1 - \theta)^{2}(c^{s})^{2} + \theta^{2}\sigma^{2}.$$
(3)

On voit que l'erreur est minimale pour $-2(1-\theta)|c^s|^2+2\theta\sigma^2=0,$ ce qui conduit à

$$\theta = \frac{(c^s)^2}{(c^s)^2 + \sigma^2} \ .$$

On a alors bien:

$$E\{(c^s - c^{\hat{s}})^2\} = \frac{\sigma^2(c^s)^2}{\sigma^2 + (c^s)^2}.$$

En pratique, cet estimateur idéal n'est pas exploitable car il nécessite de connaître les coefficients $c_j^s[k]$ qu'on cherche justement à déterminer! On dit qu'il s'agit d'un oracle.

Notons qu'un oracle plus rudimentaire est obtenu en contraignant les coefficients $\theta_j[k]$ à être binaires ($\in \{0,1\}$). Le problème d'estimation se ramène alors, en quelque sorte, à un problème de détection où l'on décide de conserver ou, au contraire, de rejeter les coefficients selon leurs valeurs relativement à l'écart-type du bruit. La relation (3) montre, en effet, qu'on doit prendre :

$$\theta_j[k] = \begin{cases} 0 & \text{si } |c_j^s[k]| < \sigma \\ 1 & \text{sinon,} \end{cases}$$

ce qui conduit à l'EQM :

$$\mathcal{E}_o^2(s) = \sum_{j=1}^{j_{\text{max}}} \sum_{k=0}^{K2^{-j}-1} \min\{(c_j^s[k])^2, \sigma^2\} \ . \tag{4}$$

On a évidemment $\mathcal{E}_a^2(s) \leq \mathcal{E}_o^2(s)$. De plus, en exploitant le fait que,

$$\forall (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+^*, \quad \frac{1}{2}\min(x,y) \le \frac{xy}{x+y}$$

on conclut que

$$\frac{\mathcal{E}_o^2(s)}{2} \le \mathcal{E}_o^2(s) \le \mathcal{E}_o^2(s) .$$

Ceci montre qu'en terme d'EQM, le second oracle n'est au pire que deux fois moins bon que le premier.

Pour mieux cerner l'intérêt du choix d'une "bonne" base pour la représentation du signal, imaginons que seuls Q coefficients d'ondelettes de s soient non nuls et que leurs valeurs absolues soient supérieures à l'écart-type du bruit (ce qui n'est possible que si l'énergie $||s||^2$ du signal est supérieure à σ^2). La relation (4) montre alors que $\mathcal{E}_o^2(s) = Q\sigma^2$. Il faut se souvenir que l'erreur quadratique introduite par le bruit sur chaque composante du signal est initialement égale à σ^2 et qu'il y a en tout $K_m = K(1 - 2^{-j_{\text{max}}})$ coefficients d'ondelettes. L'oracle binaire permet donc de réduire l'EQM d'un facteur K_m/Q . Par conséquent, le débruitage est d'autant plus performant que la décomposition sur la base choisie permet de bien concentrer l'énergie du signal (i.e. que Q est petit).

2.2 Estimateurs par seuillage

Par des techniques élémentaires de seuillage des coefficients d'ondelettes, on peut approcher les performances idéales de ces oracles. Deux autres types d'estimateurs sont couramment utilisés pour cela :

- le seuillage dur,

$$c_j^{\hat{s}}[k] = \begin{cases} c_j^r[k] & \text{si } |c_j^r[k]| > \chi, \\ 0 & \text{sinon }, \end{cases}$$

- le seuillage doux,

$$c_j^{\hat{s}}[k] = \begin{cases} c_j^r[k] - \chi & \text{si } c_j^r[k] > \chi, \\ c_j^r[k] + \chi & \text{si } c_j^r[k] < -\chi, \\ 0 & \text{sinon,} \end{cases}$$
$$= \text{sign}(c_j^r[k])(|c_j^r[k]| - \chi)_+,$$

où la valeur du seuil χ est un réel positif.² Les graphes des figures ?? font ressortir les points faibles de ces deux estimateurs :

- le seuillage dur à un comportement discontinu au voisinage de $\mp \chi$;
- le seuillage doux introduit un biais de $\mp \chi$ dans l'estimation des coefficients $c_i^s[k]$ de grande amplitude.

²Pour tout réel x, sign(x) vaut 1 si x > 0, -1 si x < 0 et 0 sinon, $(x)_+ = 1$ si x > 0 et 0 sinon.

Pour déterminer la valeur du seuil optimal, on peut adopter une stratégie minimax, c'est à dire examiner quelle est l'erreur minimale atteignable pour le signal se révélant le plus défavorable. Notre objectif va être de comparer l'EQM $\mathcal{E}^2_{\chi}(s)$ résultant de l'usage d'un estimateur de seuillage à celle de l'oracle binaire. En fait, comme nous allons le voir, la comparaison ne va pas exactement se faire avec $\mathcal{E}^2_o(s)$ mais avec une quantité qui lui est fortement apparentée : $\sigma^2 + \mathcal{E}^2_o(s)$. Les calculs étant plus simples à mener avec le seuillage doux, nous allons nous limiter à l'étude de cet estimateur.

Proposition 1. Si $K_m = K(1-2^{-j_{\text{max}}}) \ge 4$ et si les coefficients d'ondelettes du bruit $c_j^b[k]$ sont de même loi marginale $\tilde{\mu}$ symétrique (de variance σ^2), alors en employant un seuillage doux,

$$\inf_{\chi \ge 0} \sup_{s} \frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{o}^{2}(s)} = \sup_{s} \frac{\mathcal{E}_{\chi_{m}}^{2}(s)}{\sigma^{2} + \mathcal{E}_{o}^{2}(s)}$$
$$= \Lambda_{\chi_{m}} = \frac{K_{m}(\chi_{m}^{2} + \sigma^{2})}{(K_{m} + 1)\sigma^{2}} ,$$

où χ_m est l'unique solution dans \mathbb{R}_+ de l'équation

$$2(K_m + 1) \int_{\gamma}^{\infty} (z - \chi)^2 \tilde{\mu}(dz) = \chi^2 + \sigma^2 .$$
 (5)

Démonstration. Voir l'annexe A.

Remarques.

- Le résultat précédent ainsi que ceux que nous allons démontrer dans la suite de ce paragraphe sont très généraux et s'appliquent, en fait, à une décomposition sur une base orthonormale quelconque.
- Il est judicieux de réécrire l'équation (5) en utilisant la loi $\bar{\mu}$ du bruit normalisé (de variance ramenée à 1) :

$$2(K_m+1)\int_{\bar{\chi}_m}^{\infty} (z-\bar{\chi}_m)^2 \bar{\mu}(dz) = \bar{\chi}_m^2 + 1.$$
 (6)

où $\bar{\chi}_m = \chi_m/\sigma$ désigne le seuil normalisé. Cette équation montre que $\bar{\chi}_m$ est indépendant de σ et peut donc être tabulé pour une forme de loi donnée, en fonction de K_m . A titre d'exemple, dans le tableau 1, les valeurs du seuil normalisé ont été calculées numériquement dans le cas gaussien. Dans ce cas,

on peut montrer que la résolution de l'équation intégrale (6) se réduit à la résolution numérique de l'équation plus simple suivante (v. annexe B) :

$$\frac{1}{2}\operatorname{erf}\left(\frac{\bar{\chi}_m}{\sqrt{2}}\right) + \frac{\bar{\chi}_m}{\bar{\chi}_m + 1} \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\chi}_m^2}{2}} = \frac{K_m}{2(1 + K_m)}.$$
 (7)

– Dans le cas d'une décomposition en ondelettes, il est peu réaliste de supposer que les coefficients du bruit suivent la même loi $\tilde{\mu}$ (ou même $\bar{\mu}$, après normalisation) à tous les niveaux de résolution. Une situation importante où ce cas se produit néanmoins est celui d'un bruit b(t) gaussien, la loi $\tilde{\mu}$ étant alors une gaussienne centrée d'écart-type σ . Dans les autres cas, un modèle de bruit plus satisfaisant peut consister à choisir une loi $\tilde{\mu}_j$ dépendant du niveau de résolution j considéré. La proposition 1 doit alors être adaptée pour tenir compte de cette nouvelle hypothèse. On peut, par exemple, songer à appliquer un seuil différent à chaque niveau de résolution en le déterminant à l'aide de la formule (6). On peut aussi retenir la valeur maximale parmi ces seuils.

Il peut être instructif d'examiner le comportement asymptotique des estimateurs de seuillage préconisés par la proposition 1. Le résultat suivant fournit la valeur asymptotique du seuil optimal.

Proposition 2. Supposons que les coefficients d'ondelettes du bruit soient de densité marginale :

$$\forall z \in \mathbb{R}, \qquad \mu(z) = Ce^{-h(z)}$$

où $C \in \mathbb{R}_+^*$ et h est une fonction paire, continue, strictement croissante sur \mathbb{R}_+ telle que

$$\lim_{z \to \infty} z^{-\beta} h(z) = \gamma \in \mathbb{R}_+^* , \qquad \beta \ge 1$$

et

$$\forall (z_1, z_2) \in \mathbb{R}^2_+, \quad h(z_1 + z_2) \ge h(z_1) + h(z_2) .$$

Quand $K_m \to \infty$, on a

$$\chi_m \sim \chi_{\rm U} = h^{-1}(\ln K_m) \tag{8}$$

$$\Lambda_{\chi_m} \sim \Lambda_{\rm U} = \frac{\chi_{\rm U}^2}{\sigma^2} + 1 \ . \tag{9}$$

Démonstration. Voir l'annexe C.

Remarques.

– Les hypothèses de cette proposition sont vérifiées pour les lois gaussiennes généralisées correspondant au cas où $h(z) = \gamma |z|^{\beta}$ avec $\beta \geq 1$. Nous aurons l'occasion de revenir sur ces lois au paragraphe 4.2. Les lois gaussiennes appartiennent à cette famille ($\gamma = (2\sigma^2)^{-1}$ et $\beta = 2$). On trouve alors

$$\chi_{\rm U} = \sigma \sqrt{2 \ln K_m}.\tag{10}$$

- Dans la pratique des estimateurs de seuillage, il est assez tentant d'employer $\chi_{\rm U}$ plutôt que χ_m , en raison de la simplicité de l'expression (8). Comme l'illustre le tableau 1 dans le cas gaussien, l'écart avec le seuil optimal est cependant significatif pour un nombre modéré de valeurs.
- La valeur $\chi_{\rm U}$ est parfois qualifiée de seuil universel.³

2.3 L'optimalité du seuillage

Le choix d'un opérateur de seuillage pour le débruitage des coefficients d'ondelettes est séduisant en raison de sa simplicité, mais il peut paraître assez abitraire. N'existe-t-il pas d'autres formes d'estimateurs qui conduiraient à de meilleurs résultats? Nous allons voir que la réponse à cette question est négative si l'on se place dans le cas gaussien et dans des conditions asymptotiques.

En reprenant les notations des deux paragraphes précédents, on a :

Proposition 3. Si les coefficients du bruit sont i.i.d. $\mathcal{N}(0, \sigma^2)$ alors

$$\lim_{K_m \to \infty} \inf_{\hat{s}} \sup_{s} \left(\frac{\mathcal{E}^2(s)}{\left(\sigma^2 + \mathcal{E}_o^2(s)\right) \Lambda_{\chi_m}} \right) = 1.$$

où $\inf_{\hat{s}}$ désigne l'inf. sur l'ensemble de tous les estimateurs possibles du signal s(t).

$$D\acute{e}monstration.$$

On se réfère souvent à cette propriété en disant que les estimateurs de seuillage sont asymptotiquement minimax.

 $^{^3}$ Comme le souligne la remarque précédente, cette dénomination peut laisser un peu perplexe.

| K_m | χ_m | Λ_{χ_m} | χ_U | Λ_U |
|--------|----------|--------------------|----------|-------------|
| 4 | 0.683524 | 1.173764 | 1.665109 | 3.772589 |
| 8 | 0.877372 | 1.573139 | 2.039334 | 5.158883 |
| 16 | 1.076456 | 2.031772 | 2.354820 | 6.545177 |
| 32 | 1.276276 | 2.549217 | 2.632769 | 7.931472 |
| 64 | 1.474135 | 3.124256 | 2.884054 | 9.317766 |
| 128 | 1.668605 | 3.754906 | 3.115134 | 10.704061 |
| 256 | 1.859020 | 4.438616 | 3.330218 | 12.090355 |
| 512 | 2.044916 | 5.171582 | 3.532230 | 13.476649 |
| 1024 | 2.226161 | 5.949982 | 3.723297 | 14.862944 |
| 2048 | 2.402888 | 6.770567 | 3.905027 | 16.249238 |
| 4096 | 2.575057 | 7.629058 | 4.078668 | 17.635532 |
| 8192 | 2.742753 | 8.521655 | 4.245212 | 19.021827 |
| 16384 | 2.906252 | 9.445722 | 4.405465 | 20.408121 |
| 32768 | 3.065703 | 10.398216 | 4.560089 | 21.794415 |
| 65536 | 3.221205 | 11.375990 | 4.709640 | 23.180710 |
| 131072 | 3.373025 | 12.377200 | 4.854586 | 24.567004 |
| 262144 | 3.521304 | 13.399528 | 4.995328 | 25.953299 |

TAB. 1 – Comparaison entre le seuil universel et seuil donné par la formule (7) dans le cas gaussien.

2.4 Bruit de variance inconnue

Pour mettre en œuvre un seuillage, il est nécessaire de disposer de la variance du bruit. Lorsque celle-ci est inconnue, on peut l'estimer à partir des coefficients d'ondelettes à la résolution la plus fine (j=1), en supposant que le signal varie suffisamment lentement pour que la plupart de ses coefficients $(c_1^s[k])_{0 \le k < K/2}$ soient nuls.

Intéressons-nous au cas d'un bruit gaussien. Soit Z une variable aléatoire $\mathcal{N}(0, \sigma^2)$. La valeur médiane z_{med} de la distribution de |Z| est définie par

$$\begin{split} & \mathrm{P}\left(|Z| < z_{\mathrm{med}}\right) = \mathrm{P}\left(|Z| > z_{\mathrm{med}}\right) \\ \Leftrightarrow & \frac{1}{\sqrt{2\pi}\sigma} \int_0^{z_{\mathrm{med}}} \exp\!\left(-\frac{z^2}{2\sigma^2}\right) dz = \frac{1}{\sqrt{2\pi}\sigma} \int_{z_{\mathrm{med}}}^{\infty} \exp\!\left(-\frac{z^2}{2\sigma^2}\right) dz \;. \end{split}$$

Cette valeur est donc telle que

$$\int_{-\infty}^{z_{\text{med}}/\sigma} \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \, dz = \frac{3}{4}$$

et $z_{\rm med}/\sigma$ correspond ainsi au dernier quartile de la loi normale. En inversant la fonction de répartition, on trouve que

$$z_{\rm med} \simeq 0.6745 \,\sigma \,\,. \tag{11}$$

Appliquons ce résultat aux coefficients $(c_1^r[k])_{0 \le k < K/2}$ en supposant qu'idéalement tous les coefficients $c_1^s[k]$ sont nuls, ce qui impliquerait que $c_1^r[k]$ seraient $\mathcal{N}(0, \sigma^2)$. Un estimateur empirique de z_{med} est alors donné par le point médian $\text{Med}[c_1^r]$ des coefficients $\{|c_1^r[k]|, 0 \le k < K/2\}$ rangés par ordre croissant (ou décroissant). Un estimateur de l'écart-type se déduit de la relation (11):

$$\widehat{\sigma} = \frac{1}{0.6745} \mathrm{Med}|c_1^r| \ .$$

L'intérêt de cet estimateur par rapport à l'estimateur usuel de la variance est qu'il est plus robuste aux données aberrantes. De telles données correspondent typiquement aux valeurs pour lesquelles les coefficients d'ondelettes $c_1^s[k]$ sont non nuls (et les coefficients $c_1^r[k]$ ne sont donc pas $\mathcal{N}(0, \sigma^2)$). Nous avons fait l'hypothèse que ces coefficients sont en nombre réduit mais ils peuvent très bien être de grande amplitude.

3 La méthode SURE

Souvent, les approches minimax fournissent des estimations trop pessimistes car elles sont optimales dans la pire des situations possibles qui, fort heureusement, est rarement celle à laquelle on se trouve confronté. Une meilleure stratégie peut consister à estimer empiriquement la valeur de l'erreur quadratique moyenne et à chercher à minimiser cette dernière pour une forme particulière de l'estimateur des données. Cette stratégie est résumée par la méthode SURE.⁵

⁴On peut, par exemple, prendre la valeur de rang $\lfloor K/4 \rfloor$ après classement par ordre croissant où $\lfloor \cdot \rfloor$ désigne la partie entière inférieure.

⁵Les inconditionnels de l'académie française seront enclins à prononcer "sssuure" mais, pour faire pro, il faut dire "choure" (Stein Unbiased Risk Estimate). Evidemment, cela reste avec l'accent roumain que cela a le plus de charme!

3.1 Estimer le risque sans biais

Supposons qu'on dispose d'une observation Y = x + Z d'un réel x, en présence d'un bruit aléatoire Z de variance σ^2 . Considérons T(Y), un estimateur de x, et réécrivons le, pour notre convenance, sous la forme $Y + \gamma(Y)$. Imaginons – comme c'est souvent le cas en statistique – qu'on cherche à réaliser un choix "optimal" de T (ou de façon équivalente, de γ), au sein d'une classe donnée d'estimateurs (par exemple, les estimateurs linéaires ou les estimateurs polynomiaux de degré maximum fixé). La façon usuelle de procéder est de se donner un critère statistique qu'on appelle un risque moyen, correspondant à l'espérance d'une certaine "distance" entre x et T(Y). Si l'on s'intéresse plus particulièrement à la distance euclidienne, le risque s'écrit simplement :

$$\epsilon^2(x) = \mathbb{E}\{\left(x - T(Y)\right)^2\} \ .$$

Comme nous avons déjà eu l'occasion de le voir au paragraphe 2.1, ce risque est une fonction de x, la quantité à estimer, et ne constitue donc pas un critère viable pour optimiser T.

Les choses en vont autrement, si par chance, il existe une fonction J telle que $\epsilon^2(x) = E\{J(Y)\}$. La variable aléatoire J(Y) fournit un estimateur sans biais du risque. Comme elle ne dépend que de l'observation, on peut légitimement espérer l'utiliser au sein de procédures d'estimation empirique afin d'aboutir à un choix optimisé de T.

Pour mieux comprendre comment concevoir un estimateur sans biais de $\epsilon^2(x)$, développons l'expression de ce critère :

$$\epsilon^{2}(x) = E\{(x - Y - \gamma(Y))^{2}\}\$$

= $\sigma^{2} - 2(xE\{\gamma(Y)\} - E\{Y\gamma(Y)\}) + E\{\gamma(Y)^{2}\}$.

On observe ainsi qu'à l'exception du terme $x \to \{\gamma(Y)\}$, le risque quadratique est déjà pratiquement sous la forme désirée (espérance d'une fonction de Y). Le verrou qui subsiste peut être levé, dans le cas gaussien, par la formule de Stein [4]:

Lemme 1. Si Z est $\mathcal{N}(0, \sigma^2)$ et γ est une fonction continue, dérivable par morceaux et telle que, pour tout $x \in \mathbb{R}$,

$$\lim_{|z| \to \infty} \gamma(z) \exp\left(-\frac{(z-x)^2}{2\sigma^2}\right) = 0$$

$$\mathrm{E}\{\gamma(x+Z)^2\} < \infty, \quad \mathrm{E}\{|\gamma'(x+Z)|\} < \infty$$

alors

$$x \mathbb{E}\{\gamma(Y)\} = \mathbb{E}\{Y\gamma(Y)\} - \sigma^2 \mathbb{E}\{\gamma'(Y)\} . \tag{12}$$

Démonstration. V. Annexe D.

Notons que les conditions imposées à la fonction γ sont assez peu restrictives. En particulier, elles sont vérifiées pour des fonctions non linéaires continues et dérivables par morceaux, à croissance au plus polynomiale.⁶

Compte tenu de la formule de Stein et, en nous plaçant sous les hypothèses qui nous ont permis de l'établir, nous pouvons affirmer que

$$J(Y) = \sigma^2 + 2\sigma^2 \gamma'(Y) + \gamma(Y)^2 \tag{13}$$

constitue un estimateur sans biais du risque quadratique.

Ce résultat reste valable quand Y = X + Z où X est une variable aléatoire de variance finie, indépendante de $Z \sim \mathcal{N}(0, \sigma^2)$. En effet, le risque devient alors

$$\epsilon^2 = E\{(X - T(Y))^2\} = E_X\{E_Z\{(X - T(Y))^2\}\}.$$

Or, d'après les calculs précédents, la fonction J définie par (13), est telle que $\mathbb{E}_Z\{J(Y)\} = \mathbb{E}_Z\{\left(X - T(Y)\right)^2\}$, ce qui permet de conclure que $E\{J(Y)\} = \epsilon^2$.

3.2 Application aux coefficients d'ondelettes

Dans la suite, nous considérons que les coefficients d'ondelettes du signal sont des variables aléatoires de variance finie et, qu'à une même résolution, ils sont i.i.d. Les coefficients $(c_j^b[k])_{0 \le k < 2^{-j}K}$ sont, de plus, supposés i.i.d. $\mathcal{N}(0, \sigma_j^2)$ et indépendants des coefficients du signal.

Il est facile de vérifier que le seuillage doux vérifie les hypothèses du lemme $1\,\mathrm{avec^7}$

$$\gamma(z) = \begin{cases} -\chi & \text{si } z \ge \chi, \\ \chi & \text{si } z \le -\chi, \\ -z & \text{sinon.} \end{cases}$$

⁶Ceci signifie qu'il existe $m \in \mathbb{N}$ et $A \in \mathbb{R}_+$ tel que $\forall z \in \mathbb{R}, |\gamma(z)| \leq A|t|^m$.

⁷Par contre, ce lemme n'est pas directement applicable au seuillage dur qui correspond à une fonction γ discontinue.

L'EQM d'estimation, à un niveau de résolution j, $\epsilon_j^2(\chi) = \mathbb{E}\{(c_j^s[k] - c_j^{\hat{s}}[k])^2\}$, peut donc être ré-écrite à l'aide de la formule de Stein, sous la forme : $\epsilon_j^2(\chi) = \mathbb{E}\{J_\chi(c_j^r[k])\}$. La fonction J_χ est paramétrée par la valeur du seuil χ ainsi que par la variance du bruit. Elle se déduit de la relation (13) :

$$J_{\chi}(z) = \begin{cases} z^2 - \sigma_j^2 & \text{si } |z| \le \chi, \\ \chi^2 + \sigma_j^2 & \text{sinon.} \end{cases}$$

L'expression de $\epsilon_j^2(\chi)$ obtenue est uniquement fonction des coefficients d'ondelettes du signal observé. Ceci fournit un moyen d'estimer empiriquement $\epsilon_j^2(\chi)$ en calculant :

$$\hat{\epsilon}_j^2(\chi) = \frac{1}{2^{-j}K} \sum_{k=0}^{K2^{-j}-1} J_{\chi}(c_j^r[k]).$$

Compte tenu des hypothèses statistiques que nous avons adoptées, $(c_j^r[k])_{0 \le k < 2^{-j}K}$ est une suite de variables aléatoires i.i.d. et la loi des grands nombres nous assure que $\hat{\epsilon}_j^2(\chi)$ fournit une estimation consistante de $\hat{\epsilon}_j^2(\chi)$.

Il ne nous reste plus qu'à déterminer, à chaque niveau de résolution j, une valeur $\chi_{j,S}$ du seuil qui permet de minimiser $\hat{\epsilon}_j^2(\chi)$.

3.3 Algorithme de recherche du seuil optimal

A chaque niveau de résolution j, un algorithme d'optimisation simple permet de déterminer $\chi_{j,S}$.

Supposons que les coefficients $(c_j^r[k])_{0 \le k < 2^{-j}K}$ soient ré-ordonnés⁹ par ordre décroissant de leurs valeurs absolues :

$$|c_j^r[0]| \ge |c_j^r[1]| \ge \ldots \ge |c_j^r[2^{-j}K - 1]|$$
.

Si $k_0 \in \{1, \dots, 2^{-j}K - 1\}$ et $\chi \in \mathbb{R}_+$ sont tels que :

$$|c_j^r[k_0-1]| > \chi \ge |c_j^r[k_0]|$$
,

⁸Ceci signifie que $\hat{\epsilon}_j^2(\chi)$ converge presque sûrement vers $\epsilon_j^2(\chi)$ quand $2^{-j}K \to \infty$.

⁹Par abus de notation, nous ne distinguerons pas les valeurs ré-ordonnées des coefficients d'ondelettes originaux.

on a

$$2^{-j}K\hat{\epsilon}_{j}^{2}(\chi) = \sum_{k=0}^{k_{0}-1} (\chi^{2} + \sigma_{j}^{2}) + \sum_{k=k_{0}}^{K2^{-j}-1} ((c_{j}^{r}[k])^{2} - \sigma_{j}^{2})$$
$$= k_{0}\chi^{2} + (2k_{0} - K2^{-j})\sigma_{j}^{2} + \sum_{k=k_{0}}^{K2^{-j}-1} (c_{j}^{r}[k])^{2}.$$
(14)

On voit alors que $\hat{\epsilon}_j^2(\chi)$ est une fonction croissante de χ sur l'intervalle $[|c_j^r[k_0]|, |c_j^r[k_0-1]|]$. Son minimum sur cet intervalle est donc atteint en $\chi = |c_j^r[k_0]|$.

De façon similaire, si $\chi < |c_j^r[2^{-j}K - 1]|$,

$$\hat{\epsilon}_j^2(\chi) = \chi^2 + \sigma_j^2$$

et le minimum est obtenu pour $\chi = 0$. Par ailleurs, si $\chi \geq |c_j^r[0]|$,

$$2^{-j}K\hat{\epsilon}_j^2(\chi) = \sum_{k=0}^{K2^{-j}-1} (c_j^r[k])^2 - K2^{-j}\sigma_j^2$$
 (15)

et $\hat{\epsilon}_i^2(\chi)$ devient donc une fonction constante.

En résumé, on est ramené à la recherche du seuil optimal $\chi_{j,S}$, au sein d'un ensemble fini de valeurs : $\{|c_j^r[0]|, \dots, |c_j^r[2^{-j}K-1]|, 0\}$. Cette recherche peut être effectuée de manière exhaustive, en évaluant récursivement les valeurs prises par l'EQM estimée. L'initialisation se fait en déterminant $2^{-j}K\hat{\epsilon}_j^2(|c_j^r[0]|)$ à l'aide de la formule (15). L'équation récursive permettant de calculer les autres valeurs $2^{-j}K\hat{\epsilon}_j^2(|c_j^r[k_0]|)$, pour $k_0 \in \{1, \dots, 2^{-j}K-1\}$, se déduit de la relation (14) :

$$2^{-j}K\hat{\epsilon}_j^2(|c_j^r[k_0]|) - 2^{-j}K\hat{\epsilon}_j^2(|c_j^r[k_0-1]|) = k_0((c_j^r(k_0))^2 - (c_j^r(k_0-1))^2) + 2\sigma_j^2.$$

Au coût en calculs de ce traitement $(O(2^{-j}K)$ opérations), il faut ajouter celui correspondant au ré-ordonnancement des coefficients, ce qui correspond à $O(2^{-j}K\log_2(2^{-j}K))$ opérations.

L'avantage de cette méthode (appelée *sureshrink* dans la littérature anglo-saxone) est qu'elle s'ajuste automatiquement aux données à traiter.

3.4 Une assurance complémentaire (tout risque!)

Même si l'estimateur de Stein est non biaisé, il introduit inéluctablement une certaine erreur dans l'estimation du risque. Ce phénomène est surtout sensible lorsque la puissance des coefficients d'ondelettes du signal, à un niveau de résolution donné, est faible par rapport à la variance du bruit. Cette estimation erronée peut induire une détermination incorrecte du seuil optimal.

Au niveau de résolution j, un estimateur empirique de la puissance des coefficients d'ondelettes du signal est

$$\overline{(c_j^s)^2} = \frac{1}{K2^{-j}} \sum_{k=0}^{K2^{-j}-1} (c_j^r[k])^2 - \sigma_j^2$$

Lorsque cette puissance est inférieure à un certain niveau critique $\lambda_{j,K}$, on peut considérer que l'estimation du risque est trop mauvaise pour être exploitable et recourir à la valeur de seuil "universelle" $\sigma\sqrt{2\ln(K2^{-j})}$. En d'autres termes, on utilise le seuil "hybride" :

$$\chi_{j,H} = \begin{cases} \chi_{j,S} & \text{si } \overline{(c_j^s)^2} > \lambda_{j,K} \\ \sigma_j \sqrt{2 \ln(K2^{-j})} & \text{sinon.} \end{cases}$$

Il reste maintenant à définir le paramètre $\lambda_{j,K}$. Donoho et Johnstone [9] ont montré qu'en choisissant :

$$\lambda_{j,K} = \sigma_j^2 (K2^{-j})^{-1/2} \ln(K2^{-j})^{3/2}$$
,

l'estimateur a un comportement pratiquement identique à celui qui résulterait de l'emploi du seuil idéal minimisant le "vrai" risque $\epsilon_j^2(\chi)$. ¹⁰

3.5 Exemple de débruitage d'un signal Doppler

La figure 1 compare les performances de *sureshrink* à celles du seuillage universel pour un signal de type Doppler, bruité par adjonction d'un bruit blanc gaussien, centré. Les rapports signal sur bruit (RSB) sont donnés à titre indicatif. Bien que le résultat fourni par le seuillage universel puisse apparaître meilleur visuellement, ce qui lui a valu le nom de *visushrink*, cette méthode a tendance à trop fortement lisser le signal.

¹⁰En fait, Donoho et Johnstone considèrent le cas où le signal à estimer est déterministe et travaillent à partir de majorations uniformes de l'erreur d'estimation des coefficients d'ondelettes.

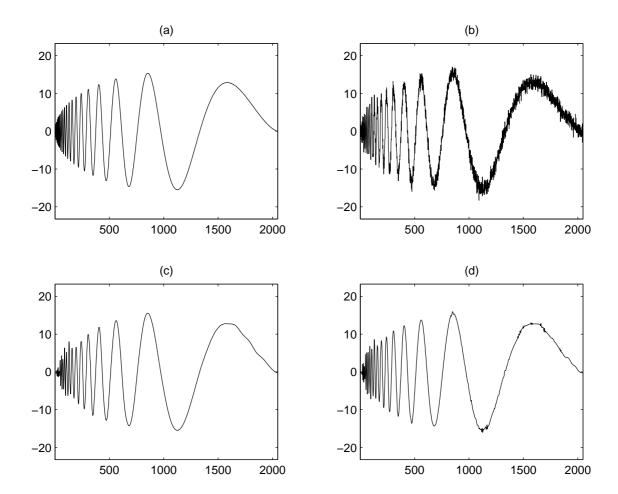


FIG. 1 – Exemple de débruitage : signaux original (a), bruité, RSB = 18.86 dB (b), après seuillage universel, RSB = 23.80 dB (c), débruité à l'aide de sureshrink, RSB = 27.45 dB (d).

3.6 Comparaison des méthodes en débruitage d'images

La figure 4 compare les performances d'un seuillage dur et d'un seuillage doux sur l'image *lena* (cf. figure 2) affectée par un bruit additif, blanc, gaussien, centré, d'écart-type 30. L'image bruitée est présentée sur la figure 3. Le seuil est donné par la formule (10).



Fig. 2 – Image *lena* originale.

Par comparaison, le débruitage par filtrage de Wiener est présenté figure 5.

La méthode *sureshrink* est illustrée sur la figure 6, de même que l'influence du choix de la base d'ondelettes. Ici, des symlets de longueur 8 et une base de Haar sont comparées, et on peut remarquer les effets de blocs produits par les ondelettes de Haar.

Finalement, une méthode dite "invariante par translation", qui utilise une trame d'ondelettes plutôt qu'une base, a été mise en œuvre, en utilisant toujours des symlets 8. Ses performances sont illustrées par la figure 7.



FIG. 3 – Image originale bruitée par un bruit blanc gaussien d'écart-type 30, RSB = 12,93 dB.

4 Avec un peu d'a priori

Les deux stratégies précédentes n'utilisent aucune connaissance statistique a priori sur le signal à estimer. Pourtant, la littérature abondante sur les approches bayésiennes montre que l'exploitation de telles informations (quand elles sont fiables) permet d'améliorer sensiblement les performances d'estimation. Aussi n'est-il pas surprenant qu'un ensemble de travaux plus récents [18, 5, 1, 6] aient étudié le couplage d'approches bayésiennes à des représentations multirésolution.

4.1 Le point de vue bayésien

Reprenons le problème d'estimation en présence d'un bruit additif, vu au paragraphe 3.1. L'approche bayésienne pour aborder ce problème consiste à probabiliser l'ensemble des données du problème : le bruit Z évidemment, mais aussi la quantité à estimer X. Nous supposerons que Z et X sont des variables aléatoires indépendantes, que la densité du bruit est μ et que celle de X est p_X . La loi p_X caractérisant X est appelée loi a priori car elle permet, dans le problème considéré, de modéli-



Fig. 4 – Comparaison entre le seuillage dur en haut (RSB = 19,72 dB) et seuillage doux en bas (RSB = 18,24 dB).



Fig. 5 – Débruitage par filtrage de Wiener, RSB = 17,44 dB.

ser toute les connaissances a priori dont on peut disposer (par exemple, positivité, valeur moyenne, dispersion,...) sur la nature de la quantité à estimer. Les formes paramétriques de μ et p_X doivent être connues.

Disposant d'une réalisation (observation) y de la variable aléatoire Y = X + Z, la loi de X conditionnellement à Y, $p_X(x \mid Y = y)$, résume en termes probabilistes toute l'information apportée sur X par la connaissance de Y. Cette densité est appelée loi a posteriori. Plusieurs estimateurs bayésiens de X peuvent être construits à partir de cette loi a posteriori selon le choix de la forme du risque : moyenne a posteriori, médiane a posteriori, . . . Dans ce paragraphe, nous nous limiterons à l'estimateur du Maximum A Posteriori, qui est généralement le plus commode à calculer. Il consiste a prendre comme estimation de X à partir de y

$$\hat{x} = \arg\max_{x} p_X(x \mid Y = y).$$

En utilisant la règle de Bayes, on a, pour tout y tel que $p_Y(y) \neq 0$,

$$p_X(x \mid Y = y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_Y(y \mid X = x)p_X(x)}{p_Y(y)} = \frac{\mu(y-x)p_X(x)}{p_Y(y)}$$



Fig. 6 – Débruitage par sureshrink: symlets 8, RSB = 22,06 dB et Haar, RSB = 20,60 dB.



FIG. 7 – Débruitage par une méthode sureshrink invariante par translation, RSB = 22,91 dB.

En remarquant que le dénominateur ne dépend pas de x et en exploitant le caractére strictement croissant de la fonction log, on voit que l'estimateur MAP est donné par

$$\hat{x} = \arg\min_{x} \left[-\ln(\mu(y-x)) - \ln(p_X(x))\right]. \tag{16}$$

4.2 Quelques bons a priori sur les coefficients d'ondelettes

Appliquons ces principes bayésiens au débruitage à l'aide d'ondelettes. Nous adoptons les hypothèses simplificatrices suivantes :

- le bruit et le signal sont mutuellement indépendants;
- leurs coefficients d'ondelettes respectifs sont des variables aléatoires indépendantes;
- au niveau de résolution j, les coefficients d'ondelettes du bruit sont $\mathcal{N}(0, \sigma_j^2)$, avec $\sigma_j > 0$, et ceux du signal sont i.i.d. de densité de probabilité p_j .

Il nous reste à effectuer le choix de la loi a priori p_j . En l'absence d'informations expérimentales spécifiques sur le signal recherché, de nombreuses possibilités sont envisageables. Si la base de décomposition est bien adaptée au signal, on s'attend

néanmoins à obtenir une représentation compacte : les coefficients $c_j^s[k]$ sont soit nuls (ou de très faibles valeurs) avec une forte probabilité, soit de grande amplitude (avec une probabilité beaucoup plus faible). Un fait intéressant que nous allons observer au travers des exemples suivants est que la plupart des *a priori* raisonnables conduisent à des estimateurs de type seuillage. 11

Lois laplacienne et gaussienne généralisée

En codage de la parole ou des images, on modélise souvent les erreurs de prédiction à l'aide d'une loi laplacienne. Les coefficients d'ondelettes présentant un certain air de ressemblance avec des erreurs de prédiction, on peut songer à faire ce choix pour la loi *a priori* :

$$p_j(u) = \frac{1}{\sqrt{2}\eta_j} \exp\left(-\frac{\sqrt{2}|u|}{\eta_j}\right)$$

Le paramètre $\eta_j > 0$ correspond à l'écart-type de la loi laplacienne. Remarquons, qu'à variance égale, cette loi a bien une valeur en 0 supérieure à celle d'une loi gaussienne alors que sa queue de distribution décroît plus lentement. On montre que :

Proposition 4. L'estimateur MAP correspondant à une distribution a priori laplacienne centrée, d'écart-type $\eta_j > 0$ et à une loi de bruit $\mathcal{N}(0, \sigma_j^2)$ est le seuillage doux de seuil $\chi_{j,B} = \sqrt{2}\sigma_j^2/\eta_j$.

Démonstration. Voir l'annexe E.

La distribution laplacienne se révèle souvent apropriée à la modélisation des coefficients d'ondelettes d'échelle fine (j petit). Dans le même temps, un théorème centrale limite multirésolution [] nous assure que, lorsqu'on décompose en ondelettes un processus stationnaire du second ordre, ses coefficients d'ondelettes tendent à devenir gaussiens pour j grand. Il peut donc être intéressant de rechercher une famille de distributions a priori permettant d'assurer une transition graduelle entre la loi laplacienne et la gaussienne. Les distributions gaussiennes généralisées (appelées aussi distributions en exponentielle de puissance) offrent cette possibilité. Un

¹¹On pourra se référer à ?? pour connaître la raison profonde de ce mystère (régularisation non différentiable).

¹²Il est cependant assez peu vraisemblable de supposer le signal à estimer stationnaire.

certain nombre d'études experimentales, en traitement d'images notamment [16], ont confirmé la pertinence de ces modèles.

La loi gaussienne généralisée $\mathcal{GG}(\alpha_j, \beta_j), (\alpha_j, \beta_j) \in (\mathbb{R}_+^*)^2$, est définie par :

$$p_j(u) = \frac{\beta_j}{2\alpha_j \Gamma(1/\beta_j)} \exp\left(-\frac{|u|^{\beta_j}}{\alpha_j^{\beta_j}}\right)$$

où Γ est la fonction gamma. La loi exponentielle symétrique (Laplacienne) étant un cas particulier correspondant à $\beta_j = 1$, on s'attend à pouvoir généraliser la proposition 4. Cette extension n'est cependant possible que pour une plage réduite de valeurs de β_j .

Proposition 5. Si $\beta_j \leq 1$, l'estimateur MAP correspondant à un a priori $\mathcal{GG}(\alpha_j, \beta_j)$ et à une loi de bruit $\mathcal{N}(0, \sigma_j^2)$ est un estimateur de seuillage dans le sens où

$$c_j^{\hat{s}}[k] = 0 \iff |c_j^{\hat{s}}[k]| \le \chi_{j,B}$$

où

$$\chi_{j,B} = \frac{2 - \beta_j}{2(1 - \beta_j)} \left(\frac{2\sigma_j^2 (1 - \beta_j)}{\alpha_j^{\beta_j}} \right)^{1/(2 - \beta_j)}.$$

Démonstration. Voir l'annexe F.

Remarques.

– On montre dans l'annexe F que, si $\beta_j > 1$ ou si $\beta_j < 1$ et $|c_j^{\hat{s}}[k]| > \chi_{j,B}$, alors $c_j^{\hat{s}}[k]$ est solution de l'équation

$$c_j^{\hat{s}}[k] + \frac{\sigma_j^2 \beta_j}{\alpha_j^{\beta_j}} \operatorname{sign}(c_j^{\hat{s}}[k]) |c_j^{\hat{s}}[k]|^{\beta_j - 1} = c_j^r[k].$$

On voit donc qu'on obtient une fonction non linéaire de $c_j^r[k]$ telle que

$$|c_i^{\hat{s}}[k]| \leq |c_i^r[k]|.$$

Lorsque cette dernière inégalité est vérifiée on dit qu'on a un *estimateur de* rétrécissement (shrinkage, en anglais). Notons que tous les estimateurs qui sont étudiés dans ce cours satisfont cette propriété de rétrécissement.

– Quand $\beta_j < 1$, le seuillage obtenu s'apparente plus à un seuillage dur qu'à un seuillage doux puisqu'il présente des discontinuités en $\mp \chi_{j,B}$ (cf. annexe F).

Loi Bernoulli-gaussienne

Une autre forme d'a priori qui paraît naturelle dans le contexte considéré est la loi Bernoulli-gaussienne :

$$p_j(u) = (1 - \gamma_j)\delta + \gamma_j \frac{1}{\sqrt{2\pi}\eta_j} \exp\left(-\frac{u^2}{2\eta_j^2}\right), \quad \eta_j > 0, 0 < \gamma_j < 1,$$

où δ est la distribution de Dirac. Avec une probabilité $1 - \gamma_j$ les coefficients d'ondelettes sont supposés nuls et, avec une probabilité γ_j (généralement petite), ils sont considérés porteurs d'information. Dans le second cas de figure, les coefficients $c_j^s[k]$ sont pris gaussiens, centrés, de variance η_j^2 .

L'inconvénient de ce modèle est qu'il conduit à un estimateur MAP dégénéré (cf. relation ??). Pour surmonter cette difficulté, on peut recourir à un artifice consistant à introduire des variables aléatoires "cachées" $q_j[k]$ qui sont binaires, indépendantes et telles que chaque composante $c_i^s[k]$ de s(t) est :

- $\mathcal{N}(0, \eta_i^2)$ quand $q_j[k] = 1$;
- nulle, quand $q_i[k] = 0$.

On a ainsi $P(q_j[k] = 1) = \gamma_j$.

Le bruit étant gaussien, la loi conditionnelle de $c_i^r[k]$, connaissant $q_j[k]$, est :

- $\mathcal{N}(0, \sigma_i^2) \text{ si } q_j[k] = 0;$
- $\mathcal{N}(0, \eta_j^2 + \sigma_j^2)$ si $q_j[k] = 1$.

Par conséquent, l'estimateur MAP de $q_j[k]$ est donné par $\hat{q}_j[k]=1$ si

$$\frac{1}{\sigma_j} \exp\left(-\frac{(c_j^r[k])^2}{2\sigma_j^2}\right) P(q_j[k] = 0) < \frac{1}{\sqrt{\eta_j^2 + \sigma_j^2}} \exp\left(-\frac{(c_j^r[k])^2}{2(\eta_j^2 + \sigma_j^2)}\right) P(q_j[k] = 1)$$

et $\hat{q}_i[k] = 0$, dans le cas contraire. On obtient alors :

$$\hat{q}_j[k] = \begin{cases} 1 & \text{si } |c_j^r[k]| > \chi_{j,B}, \\ 0 & \text{sinon,} \end{cases}$$

où le seuil bayésien $\chi_{j,B} \geq 0$ est tel que :

$$\chi_{j,\mathrm{B}}^2 = \frac{2\sigma_j^2(\eta_j^2 + \sigma_j^2)}{\eta_j^2} \ln \left(\frac{(1 - \gamma_j)\sqrt{\eta_j^2 + \sigma_j^2}}{\gamma_j \sigma_j} \right)$$

Ce seuil, de même que celui de la proposition 4 est indépendant de la durée du signal [15]. Rappelons qu'au contraire, le seuil minimax croît en fonction du nombre de données, ce qui n'est pas toujours très satisfaisant pour des signaux de longue durée. Finalement, il reste à estimer les composantes du signal. Quand $\hat{q}_j[k] = 0$, on a naturellement $c_j^{\hat{s}}[k] = 0$ et, quand $\hat{q}_j[k] = 1$, on peut procéder à une estimation de type MAP en déterminant :

$$c_j^{\hat{s}}[k] = \arg\max_{u} p_{c_j^{s}[k]}(u \mid c_j^{r}[k], q_j[k] = 1)$$
.

Sachant que $q_j[k] = 1$, $(c_j^s[k], c_j^r[k])$ est un vecteur gaussien et la loi conditionnelle précédente est donc gaussienne d'espérance

$$\mathbb{E}\{c_j^s[k] \mid c_j^r[k], q_j[k] = 1\} = \frac{\mathbb{E}\{c_j^s[k]c_j^r[k] \mid q_j[k] = 1\}}{\mathbb{E}\{(c_j^r[k])^2 \mid q_j[k] = 1\}} c_j^r[k] = \frac{\eta_j^2}{\sigma_j^2 + \eta_j^2} c_j^r[k] .$$

On aboutit de cette manière à une règle de seuillage qui réalise une forme de compromis entre une estimation linéaire en moyenne quadratique et un seuillage dur :

$$c_{j}^{\hat{s}}[k] = \begin{cases} \frac{\eta_{j}^{2}}{\sigma_{j}^{2} + \eta_{j}^{2}} c_{j}^{r}[k] & \text{si } \hat{q}_{j}[k] = 1, \\ 0 & \text{sinon.} \end{cases}$$

4.3 Détermination des hyperparamètres

Un point délicat dans ces méthodes bayésiennes est la détermination des paramètres du modèle statistique (par exemple, γ_j et η_j^2 , dans le cas Bernoulli-gaussien). Pour estimer ces paramètres, on doit généralement employer des méthodes itératives (maximum de vraisemblance généralisé, algorithme EM [7], méthodes MCMC [17],...). Les diverses solutions envisageables correspondent à des compromis performance/complexité différents.

5 Extensions

Les principes des méthodes de débruitages que nous venons de présenter peuvent être appliqués à des décompositions en paquets d'ondelettes, qui conduisent généralement à une amélioration de la qualité des estimations réalisées. Pour le choix de la meilleure base, il est alors préférable d'utiliser un critère propre au débruitage, défini, par exemple, à l'aide de la fonction $J_{\chi_{i,S}}$ introduite ci-dessus (voir aussi [14]).

Il faut aussi souligner que l'utilisation de représentations invariantes par translation (cf. paragraphe ??) permet presque systématiquement de réduire le niveau de bruit de manière plus efficace.

Pour conclure, signalons que des problèmes similaires à ceux de débruitage se posent pour l'estimation de densités de probabilité [12] ou de densités spectrales de puissance [10], pour la construction de tests statistiques [2] ainsi qu'en déconvolution [13].

Annexes

A Preuve de la Proposition 1

Lemme 2. Soit μ une densité de probabilité paire, définie sur \mathbb{R} , de variance finie σ^2 . La fonction définie sur \mathbb{R}_+ par :

$$g_0: \quad \chi \mapsto 2 \int_{\chi}^{\infty} (z - \chi)^2 \mu(z) dz$$
 (17)

est continue, décroissante et telle que :

$$g_0(0) = \sigma^2$$

et

$$\lim_{\chi \to \infty} g_0(\chi) = 0.$$

Démonstration. Montrons que g_0 est décroissante. Soit $0 \le \chi_1 \le \chi_2$. Donc, $\forall z \ge \chi_2$, $z - \chi_1 \ge z - \chi_2 \ge 0$.

$$g_0(\chi_1) = 2 \int_{\chi_1}^{\infty} (z - \chi_1)^2 \mu(z) dz \ge 2 \int_{\chi_1}^{\infty} (z - \chi_2)^2 \mu(z) dz.$$

D'où:

$$g(\chi_1) \ge 2 \int_{\chi_2}^{\infty} (z - \chi_2)^2 \mu(z) dz = g_0(\chi_2).$$

Montrons que g_0 est continue. On a :

$$g_0(\chi) = 2\left(\int_{\chi}^{\infty} z^2 \mu(z) dz - 2\chi \int_{\chi}^{\infty} z \mu(z) dz + \chi^2 \int_{\chi}^{\infty} \mu(z) dz\right).$$

Les fonctions $\chi \mapsto \int_{\chi}^{\infty} z^2 \mu(z) dz$, $\chi \mapsto \int_{\chi}^{\infty} z \mu(z) dz$ et $\chi \mapsto \int_{\chi}^{\infty} \mu(z) dz$ étant continues, le résultat s'en déduit immédiatement.

Remarquons, en outre, que

$$g_0(0) = 2 \int_0^\infty z^2 \mu(z) dz = \int_{-\infty}^\infty z^2 \mu(z) dz = \sigma^2.$$

(la moyenne est nulle, la densité de probabilité étant paire).

Enfin

$$0 \le g_0(\chi) = 2 \int_{\chi}^{\infty} (z - \chi)^2 \mu(z) dz \le 2 \int_{\chi}^{\infty} z^2 \mu(z) dz$$

puisque $\forall z \geq \chi, \ 0 \leq z - \chi \leq z$. Donc

$$\lim_{\chi \to \infty} \int_{\chi}^{\infty} z^2 \mu(z) dz = 0 \quad \Rightarrow \quad \lim_{\chi \to \infty} g_0(\chi) = 0.$$

La preuve de la Proposition 1 est assez longue et procéde en plusieurs étapes. Dans la suite, nous noterons T la fonction d'estimation (seuillage doux, dans le cas présent) $(c_j^{\hat{s}}[k] = T[c_j^r(k)])$. Par ailleurs, pour simplifier la présentation, nous supposerons que la loi des coefficients du bruit admet une densité $\mu(\cdot)$ (le cas d'une mesure de probabilité quelconque se traite de manière identique).

– Majoration de $\mathcal{E}_{\chi}^{2}(s)$

Notons:

$$E\{(c_j^s[k] - c_j^{\widehat{s}}[k])^2\} = \int_{-\infty}^{\infty} \left(c_j^s[k] - T\left[c_j^s[k] + z\right]\right)^2 \mu(z) dz = g(\chi, c_j^s(k)). \quad (18)$$

On a donc

$$\mathcal{E}_{\chi}^{2}(s) = \sum_{j,k} g(\chi, c_{j}^{s}(k)).$$

Posons

$$\Lambda_{\chi} = \sup_{x \in \mathbb{R}} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)}.$$

On a:

$$\begin{split} \mathcal{E}_{\chi}^{2}(s) &\leq \Lambda_{\chi} \sum_{j,k} \left[\frac{\sigma^{2}}{K_{m}} + \min((c_{j}^{s}[k])^{2}, \sigma^{2}) \right] \\ &= \Lambda_{\chi} \left[\sigma^{2} + \sum_{j,k} \min((c_{j}^{s}[k])^{2}, \sigma^{2}) \right] \\ &= \Lambda_{\chi}(\sigma^{2} + \mathcal{E}_{0}^{2}(s)). \end{split}$$

On trouve donc, à l'issue de cette première étape, que :

$$\frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)} \leq \Lambda_{\chi} \quad \Rightarrow \quad \sup_{s} \frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)} \leq \Lambda_{\chi}.$$

Le but que nous allons maintenant poursuivre va être d'établir l'existence et l'unicité de χ_m et de montrer que

$$\Lambda_{\chi_m} = \inf_{\chi \ge 0} \sup_{x \in \{0, \infty\}} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)}$$

- Etude de
$$\sup_{x \in \{0,\infty\}} \frac{g(\chi,x)}{\frac{\sigma^2}{K_m} + \min(x^2,\sigma^2)}$$

On a :

$$g(\chi, x) = \int_{-\infty}^{\infty} [x - T(x+z)]\mu(z)dz.$$

De plus,

$$T(x+z) = \begin{cases} x+z+\chi & \text{si } z < -x-\chi \\ 0 & \text{si } -x-\chi \le z \le -x+\chi \\ x+z-\chi & \text{si } z > -x+\chi, \end{cases}$$

ce qui permet de développer l'expression de $g(\chi, x)$:

$$g(\chi, x) = \int_{-\infty}^{-x - \chi} (z + \chi)^2 \mu(z) dz + x^2 \int_{-x - \chi}^{-x + \chi} \mu(z) dz + \int_{-x + \chi}^{\infty} (z - \chi)^2 \mu(z) dz$$

Si $x \to \infty$, alors $g(\chi, x) = \int_{-\infty}^{+\infty} (z - \chi)^2 \mu(z) dz$. On trouve donc que $g(\chi, \infty) =$ $\sigma^2 + \chi^2$. Il s'agit clairement d'une fonction continue strictement croissante sur \mathbb{R}_{+} .

Par ailleurs, on a:

$$g(\chi,0) = \int_{-\infty}^{-\chi} (z+\chi)^2 \mu(z) dz + \int_{\chi}^{\infty} (z-\chi)^2 \mu(z) dz$$

ce qui, en exploitant la symétrie de μ , conduit à

$$g(\chi, 0) = 2 \int_{\chi}^{\infty} (z - \chi)^2 \mu(z) dz = g_0(\chi).$$

Les propriétés de la fonction g_0 ont été données dans le Lemme 2.

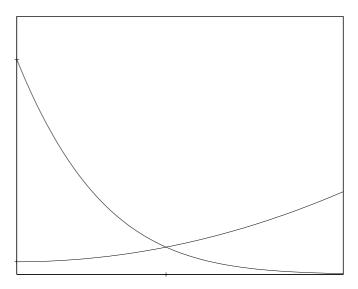


Fig. 8 – Graphes de $\frac{g(\chi,\infty)}{\frac{\sigma^2}{K_n}+\sigma^2}$ et de $\frac{g(\chi,0)}{\frac{\sigma^2}{K_m}}$

A la lumière de ces résultats, traçons les graphes respectifs de $\frac{g(\chi,\infty)}{\frac{\sigma^2}{\chi^2} + \sigma^2}$ et de $\frac{g(\chi,0)}{\frac{\sigma^2}{K_m}}$ (Fig. 8).

On voit donc qu'il existe un unique point $\chi_m \in \mathbb{R}_+$ tel que

$$\frac{g(\chi_m, 0)}{\frac{\sigma^2}{K_m}} = \frac{g(\chi_m, \infty)}{\frac{\sigma^2}{K_m} + \sigma^2} = \frac{\sigma^2 + \chi_m^2}{(1 + \frac{1}{K_m})\sigma^2}$$

et que

$$\frac{\sigma^2 + \chi_m^2}{(1 + \frac{1}{K_m})\sigma^2} = \inf_{\chi \ge 0} \sup_{x \in \{0, \infty\}} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)}$$

Notons que

$$g(\chi_m, 0) = \frac{\sigma^2 + \chi_m^2}{(1 + K_m)} \Leftrightarrow 2 \int_{\chi}^{\infty} (z - \chi_m)^2 \mu(z) dz = \frac{\sigma^2 + \chi_m^2}{(1 + K_m)},$$

ce qui montre que le χ_m ainsi trouvé vérifie l'équation (5). Pour χ fixé, $\frac{\sigma^2 + \chi^2}{(1 + K_m)}$ décroît avec K_m , ce qui montre que χ_m est une fonction

croissante non bornée de
$$K_m$$
.

– Détermination de $\sup_{|x| \ge \sigma} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)}$

Montrons que $\forall \chi \in \mathbb{R}_+$, $x \mapsto g(x, \chi)$ est croissante sur \mathbb{R}_+ . En effet, $\forall k \in \mathbb{R}_+$, $\forall x \in \mathbb{R}_+$, on a :

$$g(x+k,\chi) = \int_{-\infty}^{\infty} [x+k-T(x+k+z)]^2 \mu(z) dz = \int_{-\infty}^{\infty} [x+k-T(u+k)]^2 \mu(u-x) du,$$

où u = z + x. On a aussi :

$$g(x,\chi) = \int_{-\infty}^{\infty} [x - T(u)]^2 \mu(u - x) du$$

Comme

$$[x+k-T(u+k)]^{2} = \begin{cases} (u-x+\chi)^{2} & \text{si } u < -\chi - k \\ (x+k)^{2} & \text{si } -\chi - k \le u \le \chi - k \\ (u-x-\chi)^{2} & \text{si } u > \chi - k. \end{cases}$$

et

$$[x - T(u)]^{2} = \begin{cases} (u - x + \chi)^{2} & \text{si } u < -\chi \\ x^{2} & \text{si } -\chi \le u \le \chi \\ (u - x - \chi)^{2} & \text{si } u > \chi, \end{cases}$$

on trouve donc $\forall (k, x) \in (\mathbb{R}_+)^2$, $[x + k - T(u + k)]^2 \ge [x - T(u)]^2$, donc $g(x + k, \chi) \ge g(x, \chi)$ (v. Fig. 9).

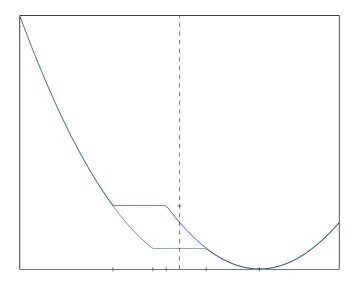


FIG. 9 – Graphes de $g(x + k, \chi)$ et $g(x, \chi)$.

En raison de l'hypothèse de symétrie de μ et du caractère impair de T, on vérifie que $x \mapsto g(x, \chi)$ est paire : $\forall x \in \mathbb{R}$,

$$g(-x,\chi) = \int_{-\infty}^{\infty} \left[-x - T(-x+z) \right]^2 \mu(z) dz = \int_{-\infty}^{\infty} \left[x - T(x-z) \right]^2 \mu(z) dz$$
$$= \int_{-\infty}^{\infty} \left[x - T(x-z) \right]^2 \mu(-z) dz = \int_{-\infty}^{\infty} \left[x - T(x+z) \right]^2 \mu(z) dz = g(x,\chi).$$

On déduit des 2 propriétés démontrées ci-dessus que : $\forall x \in \mathbb{R}_+$,

$$\sup_{|x| \ge \sigma} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)} = \frac{1}{\sigma^2 (1 + \frac{1}{K_m})} \sup_{|x| \ge \sigma} g(\chi, x) = \frac{g(\chi, \infty)}{\sigma^2 (\frac{1}{K_m} + 1)}$$

- Détermination de sup $\frac{g(x,\chi)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)}$ Montrons que $\forall x \in \mathbb{R}, \ x^2 + g(\chi, 0) \ge g(\chi, x)$

$$x^{2} + T(z)^{2} = \begin{cases} (z + \chi)^{2} + x^{2} & \text{si } z < -\chi \\ x^{2} & \text{si } -\chi \le z \le \chi \\ (z - \chi)^{2} + x^{2} & \text{si } z > \chi \end{cases}$$

et

$$[x - T(x+z)]^{2} = \begin{cases} (z+\chi)^{2} & \text{si } z < -\chi - x \\ x^{2} & \text{si } -\chi - x \le z \le \chi - x \\ (z-\chi)^{2} & \text{si } z > \chi - x \end{cases}$$

Comparons graphiquement les deux fonctions (v. Fig. 10). On en déduit que $\forall x \in \mathbb{R}_+, \forall z \in \mathbb{R}_+$

$$x^{2} + T(z)^{2} \ge [x - T(x+z)]^{2}$$

$$\Rightarrow \int_{-\infty}^{\infty} [x^{2} + T(z)^{2}] \mu(z) dz \ge \int_{-\infty}^{\infty} [x - T(z)]^{2} \mu(z) dz$$

$$\Leftrightarrow x^{2} + q(\chi, 0) \ge q(\chi, x).$$

L'inégalité reste vraie dans le cas où x < 0, du fait de la parité de $x \mapsto g(\chi, x)$. On en déduit que, $\forall x \in]-\sigma, \sigma[$,

$$\frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(\sigma^2, x^2)} = \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + x^2} \\
\leq \frac{x^2 + g(\chi, 0)}{\frac{\sigma^2}{K_m} + x^2} = 1 + \frac{g(\chi, 0) - \frac{\sigma^2}{K_m}}{\frac{\sigma^2}{K_m} + x^2}.$$

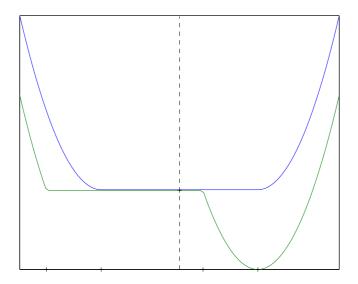


FIG. 10 – Graphes de $x^2 + g(\chi, 0)$ et $g(\chi, x)$.

Si on fait l'hypothèse supplémentaire que $g(\chi,0) \ge \frac{\sigma^2}{K_m}$, alors on a :

$$\frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(\sigma^2, x^2)} \le 1 + \frac{g(\chi, 0) - \frac{\sigma^2}{K_m}}{\frac{\sigma^2}{K_m}} = \frac{g(\chi, 0)}{\frac{\sigma^2}{K_m}}.$$

En résumé, si $g(\chi,0) \ge \frac{\sigma^2}{K_m}$, on trouve que

$$\sup_{|x|<\sigma} \frac{g(\chi,x)}{\frac{\sigma^2}{K_m} + \min(\sigma^2, x^2)} = \frac{g(\chi,0)}{\frac{\sigma^2}{K_m}},$$

ce qui, ajouté au résultat de l'étape précédente, montre que :

$$\Lambda_{\chi} = \sup_{x \in \mathbb{R}} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(\sigma^2, x^2)} = \sup_{x \in \{0, \infty\}} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(\sigma^2, x^2)}.$$

- Vérification de la relation $g(\chi_m, 0) \ge \frac{\sigma^2}{K_m}$. Rappelons que nous avons obtenu la relation

$$g(\chi_m, 0) = \frac{\sigma^2 + \chi_m^2}{1 + K_m}$$

ce qui signifie que l'inégalité qu'on cherche à prouver équivaut à

$$K_m(\sigma^2 + \chi_m^2) \ge (1 + K_m)\sigma^2 \quad \Leftrightarrow \quad \chi_m^2 \ge \frac{\sigma^2}{K_m}$$

Par ailleurs, $g(\chi,0) = E\{T(Z)^2\}$, où Z est une v.a. de loi μ et $T(Z)^2 = (|Z| - \chi)_+^2 = t(Z)^2$, avec $t(z) = (\sqrt{|z|} - \chi)_+^2$.

La fonction t est convexe et, d'après l'inégalité de Jensen,

$$E\{t(Z)^2\} \ge t\left(E\{Z^2\}\right)$$

On a aussi

$$g(\chi, 0) \ge t(\mathbb{E}\{Z^2\}) = t(\sigma^2) = \mathbb{E}\{t(Y^2)\} = g_Y(\chi, 0),$$

où Y est une v.a. de loi

$$p_Y(y) = \frac{1}{2} [\delta_{\sigma} + \delta_{-\sigma}]$$

et $g_Y(\chi, x)$ est l'équivalent de la fonction $g(\chi, x)$, où l'on a remplacé la loi μ par la loi $p_Y(y)$, de variance σ^2 .

En reprenant le graphe ?? et en notant que $g(\chi, \infty) = \sigma^2 + \chi^2 = g_Y(\chi, \infty)$ ne dépend pas de la loi, on voit que $\chi_m \geq \chi_m^Y$, où χ_m^Y est l'unique solution de l'équation

$$g_{Y}(\chi_{m}^{Y}, 0) = \frac{\sigma^{2} + (\chi_{m}^{Y})^{2}}{1 + K_{m}}$$

$$\Leftrightarrow 2 \int_{\chi_{m}^{Y}}^{\infty} (y - \chi_{m}^{Y})^{2} p_{Y}(y) dy = \frac{\sigma^{2} + (\chi_{m}^{Y})^{2}}{1 + K_{m}}$$

$$\Rightarrow \chi_{m}^{Y} < \sigma \quad \text{et} \quad (\sigma - \chi_{m}^{Y})^{2} = \frac{\sigma^{2} + (\chi_{m}^{Y})^{2}}{1 + K_{m}}.$$

Considérons le cas particulier où $K_m = 4$:

$$5(\sigma - \chi_m^Y)^2 = \sigma^2 + (\chi_m^Y)^2 \quad \Leftrightarrow \quad 4(\chi_m^Y)^2 - 10\sigma\chi_m^Y + 4\sigma^2 = 0$$

On a $\Delta' = 9\sigma^2$, d'où $\chi_m^Y = 2\sigma$ ou $\sigma/2$.

On retient la seconde solution, qui est inférieure à σ . De plus, on sait que χ_m^Y est une fonction croissante de K_m . Par conséquent, si $K_m \geq 4$, alors $\chi_m^Y \geq \frac{\sigma}{2}$ ou $(\chi_m^Y)^2 \geq \frac{\sigma^2}{4} \geq \frac{\sigma^2}{K_m}$. Donc, $\chi_m^2 \geq \frac{\sigma^2}{K_m}$.

En conclusion, ayant démontré la relation $g(\chi_m,0) \geq \frac{\sigma^2}{K_m}$, on peut affirmer que

$$\begin{split} \Lambda_{\chi_m} &= \sup_{x \in \{0, \infty\}} \frac{g(\chi_m, x)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)} \\ &= \frac{g(\chi_m, 0)}{\frac{\sigma^2}{K_m}} = \frac{g(\chi_m, \infty)}{\sigma^2(\frac{1}{K_m} + 1)} \\ &= \frac{\sigma^2 + \chi_m^2}{\sigma^2(\frac{1}{K_m} + 1)} \end{split}$$

- Minoration de sup $\frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)}$

 $\sup_{s} \frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{b}^{2}(s)} \text{ est minorée par les valeurs de } \frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)} \text{ obtenues quand } \forall (j, k), c_{j}^{s}(k) = 0 \text{ ou } \forall (j, k), c_{j}^{s}(k) \to \infty, \text{ soit}$

$$\sup_{s} \frac{\mathcal{E}_{\chi}^2(s)}{\sigma^2 + \mathcal{E}_0^2(s)} \geq \max\left\{\frac{K_m g(\chi, 0)}{\sigma^2}, \frac{K_m g(\chi, \infty)}{\sigma^2 + K_m \sigma^2}\right\} = \sup_{x \in \{0, \infty\}} \frac{g(\chi, x)}{\frac{\sigma^2}{K_m} + \min(x^2, \sigma^2)}$$

et.

$$\inf_{\chi \ge 0} \sup_{s} \frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)} \ge \inf_{\chi \ge 0} \sup_{x \in \{0, \infty\}} \frac{g(\chi, x)}{\frac{\sigma^{2}}{K_{m}} + \min(x^{2}, \sigma^{2})}$$

$$= \Lambda_{\chi_{m}} \ge \sup_{s} \frac{\mathcal{E}_{\chi_{m}}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)}$$

Or, par définition,

$$\sup_{s} \frac{\mathcal{E}_{\chi_m}^2(s)}{\sigma^2 + \mathcal{E}_0^2(s)} \ge \inf_{\chi \ge 0} \sup_{s} \frac{\mathcal{E}_{\chi}^2(s)}{\sigma^2 + \mathcal{E}_0^2(s)}$$

On en déduit que

$$\inf_{\chi \ge 0} \sup_{s} \frac{\mathcal{E}_{\chi}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)} = \sup_{s} \frac{\mathcal{E}_{\chi_{m}}^{2}(s)}{\sigma^{2} + \mathcal{E}_{0}^{2}(s)} = \Lambda_{\chi_{m}}.$$

B Simplification de l'équation intégrale (6)

En utilisant χ_m , l'équation (6) s'écrit :

$$\int_{\bar{\chi_m}}^{\infty} (z - \bar{\chi_m})^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{\bar{\chi_m}^2 + 1}{2(\bar{\chi_m} + 1)},$$

d'où

$$\int_{\bar{\chi_m}}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - 2\chi_m \int_{\bar{\chi_m}}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \chi_m^2 \int_{\bar{\chi_m}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{\chi_m^2 + 1}{2(\chi_m + 1)}.$$

Par ailleurs,

$$\int_{\bar{\chi_m}}^{\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \left[-e^{-\frac{z^2}{2}} \right]_{\chi_m}^{\infty} = \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\chi_m}^2}{2}}$$

et, en intégrant par parties la première intégrale, on a :

$$\int_{\bar{\chi_m}}^{\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \left[-z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \right]_{\bar{\chi_m}}^{\infty} + \int_{\bar{\chi_m}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$= \bar{\chi_m} \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\chi_m}^2}{2}} + \int_{\bar{\chi_m}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

On a donc

$$(\bar{\chi_m}^2 + 1) \int_{\bar{\chi_m}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - \bar{\chi_m} \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\chi_m}^2}{2}} = \frac{(\bar{\chi_m}^2 + 1)}{2(K_m + 1)}$$
$$\frac{1}{2} - \int_0^{\bar{\chi_m}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - \frac{\bar{\chi_m}}{\bar{\chi_m}^2 + 1} \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\chi_m}^2}{2}} = \frac{1}{2(K_m + 1)}$$

On obtient finalement l'équation

$$\int_0^{\bar{\chi_m}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \frac{\bar{\chi_m}}{\bar{\chi_m}^2 + 1} \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\chi_m}^2}{2}} = \frac{K_m}{2(K_m + 1)}$$

En introduisant la fonction

$$\operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt$$

on est ramené à l'équation :

$$\frac{1}{2}\operatorname{erf}\left(\frac{\bar{\chi_m}}{\sqrt{2}}\right) + \frac{\bar{\chi_m}}{\bar{\chi_m}^2 + 1} \frac{1}{\sqrt{2\pi}} e^{-\frac{\bar{\chi_m}^2}{2}} = \frac{K_m}{2(K_m + 1)}.$$

C Preuve de la Proposition 2

 χ_m est la solution de l'équation $q(\chi) = K_m$, où

$$q(\chi) = \frac{\sigma^2 + \chi^2}{2C \int_{\chi}^{\infty} (z - \chi)^2 e^{-h(z)} dz} - 1.$$

Soit $\delta \in]0,1[$. Etudions le comportement asymptotique (quand $\chi \to \infty$) de

$$\begin{split} \frac{\int_{\chi}^{\infty} (z-\chi)^2 e^{-h(z)} dz}{(\sigma^2 + \chi^2) e^{-h[(1+\delta)\chi]}} &= \frac{\int_{0}^{\infty} z^2 e^{-h(z+\chi)} dz}{(\sigma^2 + \chi^2) e^{-h[(1+\delta)\chi]}} \\ &= \int_{0}^{\infty} z^2 e^{h[(1+\delta)\chi] - h(z+\chi) - \ln(\sigma^2 + \chi^2)} dz \\ &\geq \int_{0}^{A} z^2 e^{h[(1+\delta)\chi] - h(z+\chi) - \ln(\sigma^2 + \chi^2)} dz, \quad A > 0. \end{split}$$

Or, $\forall z \in [0, A]$,

$$h[(1+\delta)\chi] - h(z+\chi) - \ln(\sigma^2 + \chi^2) \ge h[(1+\delta)\chi] - h(A+\chi) - \ln(\sigma^2 + \chi^2),$$

puisque h a été supposée croissante. On a donc

$$\frac{\int_{\chi}^{\infty} (z - \chi)^{2} e^{-h(z)} dz}{(\sigma^{2} + \chi^{2}) e^{-h[(1+\delta)\chi]}} \ge e^{h[(1+\delta)\chi] - h(A+\chi) - \ln(\sigma^{2} + \chi^{2})} \int_{0}^{A} z^{2} dz$$

$$= \frac{A^{3}}{3} e^{h[(1+\delta)\chi] - h(A+\chi) - \ln(\sigma^{2} + \chi^{2})}.$$

Comme $\lim_{\chi \to \infty} h(A + \chi) = \infty$, on a :

$$\lim_{\chi \to \infty} \frac{h[(1+\delta)\chi]}{h(A+\chi)} = \lim_{\chi \to \infty} \frac{(1+\delta)^{\beta}\chi^{\beta}}{(A+\chi)^{\beta}} = (1+\delta)^{\beta}$$

et

$$\lim_{\chi \to \infty} \frac{\ln(\sigma^2 + \chi^2)}{h(A + \chi)} = \lim_{\chi \to \infty} \frac{\ln(\sigma^2 + \chi^2)}{\gamma \chi^\beta} = 0,$$

on a donc:

$$h[(1+\delta)\chi] - h(A+\chi) - \ln(\sigma^2 + \chi^2) \equiv h(A+\chi)[(1+\delta)^{\beta} - 1]$$

quand $\chi \to \infty$, ce qui montre que

$$\lim_{\chi \to \infty} h[(1+\delta)\chi] - h(A+\chi) - \ln(\sigma^2 + \chi^2) = \infty$$

et qui permet de conclûre que

$$\lim_{\chi \to \infty} \frac{\int_{\chi}^{\infty} (z - \chi)^2 e^{-h(z)} dz}{(\sigma^2 + \chi^2) e^{-h[(1 + \delta)\chi]}} = \infty$$

$$\Rightarrow \lim_{\chi \to \infty} q(\chi) e^{-h[(1+\delta)\chi]} = 0.$$

De façon similaire, étudions

$$\frac{\int_{\chi}^{\infty} (z-\chi)^2 e^{-h(z)} dz}{(\sigma^2 + \chi^2) e^{-h[(1+\delta)\chi]}} = \int_{0}^{\infty} z^2 e^{h[(1-\delta)\chi] - h(z+\chi) - \ln(\sigma^2 + \chi^2)} dz.$$

De plus, on a $\forall z \in \mathbb{R}_+$,

$$h[(1-\delta)\chi] - h(z+\chi) - \ln(\sigma^2 + \chi^2) \le h[(1-\delta)\chi] - h(z) - h(\chi) - \ln(\sigma^2 + \chi^2),$$

d'où

$$0 \le \frac{\int_{\chi}^{\infty} (z - \chi)^2 e^{-h(z)} dz}{(\sigma^2 + \chi^2) e^{-h[(1+\delta)\chi]}}$$

$$= \le e^{h[(1-\delta)\chi] - h(\chi) - \ln(\sigma^2 + \chi^2)} \int_0^{\infty} z^2 e^{-h(z)} dz$$

$$= \frac{\sigma^2}{2C} e^{h[(1-\delta)\chi] - h(\chi) - \ln(\sigma^2 + \chi^2)}$$

et

$$\lim_{\chi \to \infty} h[(1 - \delta)\chi] - h(\chi) - \ln(\sigma^2 + \chi^2) = \lim_{\chi \to \infty} h(\chi)[(1 - \delta^{\beta}) - 1] = -\infty.$$

On en déduit que

$$\lim_{\chi \to \infty} \frac{\int_{\chi}^{\infty} (z - \chi)^2 e^{-h(z)} dz}{(\sigma^2 + \chi^2) e^{-h[(1+\delta)\chi]}} = 0$$

et donc

$$\lim_{\chi \to \infty} q(\chi) e^{-h[(1+\delta)\chi]} = \infty.$$

Compte tenu des deux limites calculées précédemment, on peut affirmer qu'il existe $\eta_{\delta} > 0$ tel que, pour tout $\chi > \eta_{\delta}$, on ait :

$$q(\chi)e^{-h[(1+\delta)\chi]} \le 1$$

et

$$q(\chi)e^{-h[(1-\delta)\chi]} \ge 1$$

Mais χ_m est tel que $q(\chi_m) = K_m$ et nous avons déjà mentionné (cf. Preuve précédente) que $K_m \to \infty \quad \Rightarrow \quad \chi_m \to \infty$. Il existe donc $N_\delta \in \mathbb{N}$ tel que, pour tout $k_m > N_\delta$, on ait :

$$K_m e^{-h[(1+\delta)\chi_m]} \le 1$$

et

$$K_m e^{-h[(1-\delta)\chi_m]} \ge 1,$$

ce qui nous conduit à

$$h[(1+\delta)\chi_m] \ge \ln K_m$$

 $h[(1-\delta)\chi_m] \le \ln K_m$

ou encore

$$(1+\delta)\chi_m \ge h^{-1}(\ln K_m)$$
$$(1-\delta)\chi_m \le h^{-1}(\ln K_m)$$

(h étant strictement croissante et continue, on peut bien définir h^{-1} , qui est aussi strictement croissante).

On a finalement :

$$\frac{1}{1+\delta}h^{-1}(\ln K_m) \le \chi_m \le \frac{1}{1-\delta}h^{-1}(\ln K_m)$$

et on en déduit que

$$\frac{1}{1+\delta} \le \lim_{K_m \to \infty} \frac{\chi_m}{h^{-1}(\ln K_m)} \le \frac{1}{1-\delta}.$$

La propriété étant vraie pour $\delta > 0$ arbitrarement petit, ceci implique que :

$$\lim_{K_m \to \infty} \frac{K_m}{h^{-1}(\ln K_m)} = 1$$

De la relation

$$\Lambda_{\chi_m} = \frac{\sigma^2 + \chi_m^2}{\sigma^2 (1 + \frac{1}{K_m})}$$

on déduit immédiatement que

$$\lim_{K_m \to \infty} \frac{\sigma^2 \Lambda_{\chi_m}}{\sigma^2 + [h^{-1}(\ln K_m)]^2} = 1.$$

D Preuve du Lemme 1

On a:

$$xE\{\gamma(Y)\} = E\{(Y - Z)\gamma(Y)\} = E\{Y\gamma(Y)\} - E\{Z\gamma(Y)\}.$$

Or

$$E\{Z\gamma(Y)\} = \int_{-\infty}^{\infty} z\gamma(x+z) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} dz$$

(l'existence de cette intégrale est garantie par le fait que $E\{\gamma(Y)^2\} < \infty$). Par intégration par parties, on a :

$$E\{Z\gamma(Y)\} = \frac{1}{\sqrt{2\pi}\sigma} \left\{ \left[-\gamma(x+z)\sigma^2 e^{-\frac{z^2}{2\sigma^2}} \right]_{-\infty}^{\infty} + \sigma^2 \int_{-\infty}^{\infty} \gamma'(x+z)e^{-\frac{z^2}{2\sigma^2}} dz \right\}$$

(l'existence de cette intégrale est déduite de l'hypothèse $\mathrm{E}\{|\gamma(Y)'|\}<\infty$) et puisque

$$\lim_{|z| \to \infty} \gamma(z) e^{-\frac{(z-x)^2}{2\sigma^2}} = 0,$$

on a:

$$E\{Z\gamma(Y)\} = \sigma^2 \int_{-\infty}^{\infty} \gamma'(x+z) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} dz = \sigma^2 E\{\gamma'(Y)\}.$$

E Preuve de la Proposition 4

Ommetons les indices j et k, comme nous l'avons déjà fait, de façon à alléger les notations.

D'après la relation (16),

$$c^{\hat{s}} = \arg\min_{c^s} M(c^s),$$

οù

$$M(c^s) = \frac{1}{2\sigma^2}(c^r - c^s)^2 + \frac{\sqrt{2}|c^s|}{n}.$$

La fonction étant continue et tendant vers l'infini en $\pm \infty$, elle posséde nécessairement un minimum global. Elle est, de plus, strictement convexe, ce qui garantit l'unicité de ce minimum.

Si $c^{\hat{s}} > 0$, alors

$$M'(c^s) = 0 \iff -\frac{1}{\sigma^2}(c^r - c^{\widehat{s}}) + \frac{2}{\eta} = 0 \iff c^{\widehat{s}} = c^r - \frac{\sqrt{2}\sigma^2}{\eta}$$

On a donc nécessairement $c^r > \frac{\sqrt{2}\sigma^2}{\eta}$.

De façon similaire, si $c^{\hat{s}} < 0$, alors $c^{\hat{s}} = c^r + \frac{\sqrt{2}\sigma^2}{n}$

On peut donc en déduire que, si $|c^r| \leq \frac{\sqrt{2}\sigma^2}{\eta}$, alors $c^{\hat{s}} = 0$. Si $|c^r| > \frac{\sqrt{2}\sigma^2}{\eta}$, soit $c^{\hat{s}} = c^r - \text{sign}(c^r) \frac{\sqrt{2}\sigma^2}{\eta}$, soit $c^{\hat{s}} = 0$. Comparons les valeurs de M en ces deux points. On a :

$$M(c^{r} - \operatorname{sign}(c^{r}) \frac{\sqrt{2}\sigma^{2}}{\eta}) = \frac{1}{2\sigma^{2}} \frac{2\sigma^{4}}{\eta^{2}} + \frac{\sqrt{2}}{\eta} \left(|c^{r}| - \frac{\sqrt{2}\sigma^{2}}{\eta} \right) = \frac{2}{\eta} |c^{r}| - \frac{\sigma^{2}}{\eta^{2}}$$

De plus,

$$M(0) = \frac{1}{2\sigma^2} (c^r)^2.$$

Le problème se réduit donc à comparer les fonctions $u\mapsto \frac{\sqrt{2}}{\eta}u-\frac{\sigma^2}{\eta^2}$ et $u\mapsto \frac{1}{2\sigma^2}u^2$ pour $u > \frac{\sqrt{2}\sigma^2}{\eta}$. Les deux fonctions prennent la même valeur en $\frac{\sqrt{2}\sigma^2}{\eta}$, égale à $\frac{\sigma^2}{\eta^2}$. La première fonction étant affine de "pente" $\frac{\sqrt{2}}{\eta}$ et la seconde étant quadratique de dérivée égale à $\frac{\sqrt{2}}{\eta}$, on a nécessairement

$$M(c^r - \operatorname{sign}(c^r) \frac{\sqrt{2}\sigma^2}{\eta}) < M(0).$$

On peut conclure que, pour $|c^r| > \frac{\sqrt{2}\sigma^2}{\eta}$, on a $c^{\hat{s}} = c^r - \text{sign}(c^r) \frac{\sqrt{2}\sigma^2}{\eta}$.

Preuve de la Proposition 5 \mathbf{F}

Calculons

$$c^{\hat{s}} = \arg\min_{c^s} M(c^s),$$

οù

$$M(c^{s}) = \frac{1}{2\sigma^{2}}(c^{r} - c^{s})^{2} + \frac{|c^{s}|^{\beta}}{\alpha^{\beta}}.$$

La fonction M est continue et tend vers l'infini en ∞ , elle posséde nécessairement un minimum global.

Intéressons-nous au cas où $c^r \geq 0$ (le cas $c^r < 0$ se traite symétriquement). On a pour $c^s \neq 0$:

$$M'(c^s) = \frac{c^s - c^r}{\sigma^2} + \frac{\beta \operatorname{sign}(c^s) |c^s|^{\beta - 1}}{\alpha^{\beta}}$$

et, par conséquent, si $c^s < 0$, alors $M'(c^s) < 0$ et, si $c^s > c^r$, alors $M'(c^s) > 0$. Ceci montre que $c^{\hat{s}} \in [0, c^r]$

Si
$$c^{\hat{s}} \neq 0$$
, alors $M'(c^s) = 0 \iff g(c^{\hat{s}}) = c^r$, avec

$$\forall u \in \mathbb{R}_+, \quad g(u) = u + \frac{\sigma^2 \beta}{\alpha^{\beta}} u^{\beta - 1}.$$

Etudions la fonction g. Sa dérivée est

$$\forall u \in \mathbb{R}_+, \quad g'(u) = 1 + \frac{\sigma^2 \beta(\beta - 1)}{\alpha^{\beta}} u^{\beta - 2}.$$

On voit que cette dérivée s'annule en

$$u_0 = \left[\frac{\sigma^2 \beta (1 - \beta)}{\alpha^{\beta}} \right]^{\frac{1}{2 - \beta}}$$

et est strictement positive (resp. négative) si $u > u_0$ (resp. $u < u_0$). La fonction g est donc minimale pour $u = u_0$.

- Si $c^r < g(u_0)$, on ne peut trouver $c^{\hat{s}} \neq 0$ tel que $g(c^{\hat{s}}) = c^r$. On en déduit que $c^{\hat{s}} = 0$. Comme nous allons le voir, $g(u_0)$ ne constitue qu'un minorant du seuil.
- Si $c^r = g(u_0)$, $M'(c^s) = \frac{1}{\sigma^2}(g(c^s) c^r) \ge 0$ et la valeur minimale de $M(c^s)$ est atteinte en $c^{\hat{s}} = 0$.
- Si $c^r > g(u_0)$, il existe $(u_1, u_2) \in]0, u_0[\times]u_0, \infty[$ (dépendant de c^r) tel que $g(u_1) = g(u_2) = c^r$. Par ailleurs,

$$M'(c^s) > 0$$
 si $c^s \in]0, u_1[\cup]u_2, \infty[$
 $M'(c^s) < 0$ si $c^s \in]u_1, u_2[$.

Ceci indique que M posséde un unique minimum local sur $]0, \infty[$, qui est localisé en $u_2 > u_0$. Pour savoir si ce minimum est global, il faut comparer M(0) à $M(u_2)$:

$$M(0) < M(u_2)$$

$$\Leftrightarrow \frac{(c^r)^2}{2\sigma^2} < \frac{(c^r - u_2)^2}{2\sigma^2} + \frac{u_2^{\beta}}{\alpha^{\beta}}$$

$$\Leftrightarrow -2c^r u_2 + u_2^2 + \frac{2\sigma^2}{\alpha^{\beta}} u_2^{\beta} > 0$$

En exploitant la relation $g(u_2) = c^r$, on remarque que cette inégalité équivaut à :

$$-u_2^2 + \frac{2\sigma^2}{\alpha^\beta} (1-\beta) u_2^\beta > 0 \quad \Leftrightarrow \quad u_2 < \left(\frac{2\sigma^2 (1-\beta)}{\alpha^\beta}\right)^{\frac{1}{2-\beta}} \tag{19}$$

La fonction g étant strictement croissante sur $[u_0, \infty[$, ceci est aussi équivalent à

$$g(u_2) < g\left(\frac{2\sigma^2(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}} \Leftrightarrow c^r < \chi_{\rm B}$$

avec

$$\chi_{\mathrm{B}} = \left(\frac{2\sigma^{2}(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}} + \frac{\sigma^{2}\beta}{\alpha^{\beta}} \left(\frac{2\sigma^{2}(1-\beta)}{\alpha^{\beta}}\right)^{\frac{\beta-1}{2-\beta}}$$

$$= \left(\frac{2\sigma^{2}(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}} + \beta \left(\frac{\sigma^{2}}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}} (2(1-\beta))^{\frac{\beta-1}{2-\beta}}$$

$$= \left(\frac{2\sigma^{2}(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}} \left(1+\beta (2(1-\beta))^{-1}\right)$$

$$= \frac{2-\beta}{2(1-\beta)} \left(\frac{2\sigma^{2}(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}}.$$

En remarquant que $g(u_0) < \chi_B$ (puisque $u_0 < \left(\frac{2\sigma^2(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}}$), on a, en résumant les étapes précédentes :

- soit $0 \le c^r \le \chi_{\mathrm{B}}$ et $c^{\widehat{s}} = 0$
- soit $c^r > \chi_{\rm B}$ et $c^{\hat{s}}$ est l'unique solution u_2 supérieure à u_0 de l'équation

$$g(u) = c^r (20)$$

Compte tenu de l'équation (19), on a $M(0) = M(u_2)$ quand $u_2 = \left(\frac{2\sigma^2(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}}$, ce qui montre que

$$\lim_{c^r \to \chi_{\rm B}} c^{\hat{s}} = \left(\frac{2\sigma^2(1-\beta)}{\alpha^{\beta}}\right)^{\frac{1}{2-\beta}}.$$

$$c^r > \chi_{\rm B}$$
(21)

Toute l'étude que nous venons de réaliser a utilisé l'hypothèse $\beta < 1$. Le cas $\beta = 1$ a déjà été traité (cf. Proposition 4). Quand $\beta > 1$, M est différentiable sur \mathbb{R} et la fonction g est strictement croissante de 0 à ∞ , sur \mathbb{R}_+ . L'estimateur $c^{\hat{s}}$ est alors toujours l'unique solution de l'équation (20).

Références

- [1] F. Abramovich, T. Sapatinas, and B. W. Silverman. Wavelet thresholding via a Bayesian approach. J. R. Statist. Soc. B, 60:725–749, 1998.
- [2] A. Antoniadis, I. Gijbels, and G. Grégoire. Model selection using wavelet decomposition and applications. Technical Report Discussion paper 9508, Institute of Statistics, Université Catholique de Louvain, Belgium, 1996. To appear in Biometrika.
- [3] R. Averkamp and C. Houdré. Some distributional properties of the continuous wavelet transform of random processes. *IEEE Trans. Informat. Theory*, IT-44(3):1111–1124, 1998.
- [4] J. O. Berger. Statistical decision theory and Bayesian analysis. Springer-Verlag, New York, USA, 1985.
- [5] H. A. Chipman, E. D. Kolaczyck, and R. E. McCulloch. Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, 92:1413– 1421, 1997.
- [6] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based statistical signal processing using hidden Markov models. *IEEE Trans. Signal Processing*, SP-46:886–902, Apr. 1998.
- [7] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Ann. Roy. Stat. Soc.*, 11:1–38, Dec. 1977.
- [8] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrin-kage. *Biometrika*, 81:425–455, Sept. 1994.
- [9] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224, Dec. 1995.
- [10] H.-Y. Gao. Choice of thresholds for wavelet shrinkage estimate of the spectrum. Journal of Time Series Analysis, 18(3), 1997.
- [11] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. Wavelets, approximation and statistical applications. Lecture notes in statistics, Springer-Verlag, Berlin, 1998.

- [12] I. Johnstone, G. Kerkyacharian, and D. Picard. Estimation d'une densité de probabilité par méthode d'ondelettes. *C. R. Acad. Sci. Paris*, 315, Série I :211–216, 1991.
- [13] J. Kalifa and S. Mallat. Minimax restoration and deconvolution. In P. Müller and B. Vidakovic, editors, *Bayesian inference in wavelet-based models*, pages 115–138. Lecture Notes in Statistics, Springer Verlag, 1999.
- [14] H. Krim, D. Tucker, S. Mallat, and D. Donoho. On denoising and best signal representation. *IEEE Trans. Informat. Theory*, 45:2225–2238, Nov. 1999.
- [15] D. Leporini, J.-C. Pesquet, and H. Krim. Best basis representations with prior statistical models. In P. Müller and B. Vidakovic, editors, *Bayesian inference* in wavelet-based models, pages 155–172. Lecture Notes in Statistics, Springer Verlag, 1999.
- [16] E. P. Simoncelli. Bayesian denoising of visual images in the wavelet domain. In P. Müller and B. Vidakovic, editors, *Bayesian inference in wavelet-based models*, pages 291–308. Lecture Notes in Statistics, Springer Verlag, 1999.
- [17] A. Smith and G. Roberts. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *J. R. Statist. Soc. B*, 55:3–23, 1993.
- [18] B. Vidakovic. Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. Journal of the American Statistical Association, 93:173–179, 1998.