



HAL
open science

Parcimonie Adaptative Structurée par Blocs Dyadiques

Gabriel Peyré, Jalal M. Fadili, Christophe Chesneau

► **To cite this version:**

Gabriel Peyré, Jalal M. Fadili, Christophe Chesneau. Parcimonie Adaptative Structurée par Blocs Dyadiques. GRETSI'11, 2011, Bordeaux, France. 4 pp. hal-00744811

HAL Id: hal-00744811

<https://hal.science/hal-00744811v1>

Submitted on 24 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Parcimonie Adaptative Structurée par Blocs Dyadiques

Gabriel PEYRE¹, Jalal FADILI², Christophe CHESNEAU³

¹Ceremade CNRS-Univ. Dauphine ²GREYC CNRS-ENSICAEN-Univ. Caen ³LMNO CNRS-Univ. Caen

Gabriel.Peyre@ceremade.dauphine.fr, Jalal.Fadili@greyc.ensicaen.fr

Christophe.Chesneau@math.unicaen.fr

Résumé – Cet article traite de l’estimation adaptative de la structuration de la parcimonie en blocs dyadiques, en minimisant des mesures appropriées du risque d’estimation. Ce travail propose une nouvelle manière d’inférer adaptativement à partir des observations, la structuration des blocs pour des signaux/images parcimonieux. Pour ce faire, la théorie du risque de Stein est mise en oeuvre. La meilleure structuration est celle minimisant ce risque. Nous proposons (i) des expressions analytiques du risque; (ii) des estimateurs judicieux du risque en vue de l’implémentation numérique; (iii) un algorithme rapide de programmation dynamique aboutissant à la partition optimale. Une illustration numérique en débruitage d’images synthétiques et naturelles est donnée.

Abstract – This paper proposes a novel method to adapt the block-sparsity structure to the observed noisy data. Towards this goal, the Stein risk estimator framework is exploited, and the block-sparsity is dyadically organized in a tree. The adaptation of the sparsity structure is obtained by finding the best recursive dyadic partition, whose terminal nodes (leaves) are the blocks, that minimizes a data-driven estimator of the risk. Our main contributions are (i) analytical expression of the risk; (ii) a novel estimator of the risk; (iii) a fast algorithm that yields the best partition. Numerical results illustrated on wavelet-domain denoising of synthetic and natural images show the improvements brought by our adaptive approach.

1 Introduction

La parcimonie est devenue indéniablement un concept majeur en traitement du signal et des images, et son utilisation comme source d’a priori a connu un essor considérable ces dernières années avec des implications majeures en compression, restauration, et plus généralement les problèmes inverses, etc. L’intérêt de la parcimonie a sans doute culminé grâce à l’émergence plus récente de l’échantillonnage compressé (compressed sensing) qui renouvelle la vision de Shannon de l’échantillonnage.

Il est cependant bien connu que dans une décomposition multi-échelle d’une image naturelle, dans les sous-bandes de détails, les coefficients de forte valeur tendent à se regrouper autour des singularités et ce quelle que soit l’orientation. Ces agglomérats de coefficients décrivent un chaînage géométrique local. Cette dépendance incite à introduire des modèles joints plus élaborés prenant en compte cette structuration de la parcimonie, typiquement par *blocs*, lors de la modélisation, mais aussi pour les traitements ultérieurs notamment en inférence Bayésienne pour la résolution de nombreux problèmes inverses.

Parcimonie et estimation par bloc

L’approche désormais classique en débruitage [14, 13] opère par seuillage individuel (terme-à-terme) des coefficients dans une base appropriée, où le signal/image est supposé parcimonieux, e.g. typiquement une base d’ondelettes. Toutefois, ce seuillage individuel atteint un certain compromis entre biais et variance dans le risque quadratique. Ce compromis est toutefois non-optimal et peut être amélioré en exploitant l’information provenant des coefficients voisins. En clair, ces coefficients doivent être seuillés par blocs (ou groupes). Il a été ainsi montré qu’une telle procédure, avec une taille de blocs fixe appropriée, atteint des vitesses minimax optimales sur plusieurs boules fonctionnelles (typiquement Besov) [4, 5, 8]. D’un point de vue pra-

tique, les résultats de [8, 7, 21] montrent clairement l’apport du seuillage par blocs. Le seuillage par blocs a été utilisé pour des images multicanales dans [7], et dans [21] pour le débruitage audio. Dans ces travaux, et en s’inspirant de [6], les hyperparamètres du débruiteur (seuil, taille des blocs), sont choisis comme ceux minimisant l’estimateur de Stein sans biais du risque (SURE).

Les estimateurs de seuillage par blocs de taille fixe ont été étendus au cas des problèmes inverses dans [9]. Par ailleurs, depuis peu, la régularisation parcimonieuse convexe par blocs (structuration connue a priori) fait l’objet d’une activité de recherche importante aussi bien en théorie statistique et d’apprentissage (group Lasso [22, 1]) qu’en échantillonnage compressé [2].

Partitionnement dyadique La méthodologie CART [3] repose sur l’idée de partitionnement récursif, avec des applications en analyse statistique des données. En commençant par la partition triviale contenant tout le domaine, une sous-partition est construite en décomposant en deux parties horizontalement et verticalement. La procédure est répétée récursivement à chaque prédécesseur pour en produire les partitions filles. La version originale du CART autorise des partitions dans des proportions arbitraires de chaque père, alors que le CART dyadique, telle que nous l’entendons ici, découpera chaque prédécesseur en parts égales (i.e. carrés, voir Fig. 1). Bien qu’il autorise moins de partitions possibles, le CART dyadique permet de trouver la partition optimale (à définir dans quel sens) en temps linéaire. Il existe des liens forts entre le CART et l’algorithme de meilleure base dans un dictionnaire structuré en arbre 2^d -aire [12, 10]. Dans ces deux méthodes, les arbres et leur élagage sont des structures fondamentales. De plus, ils sont tous deux basés sur le résultat fondamental qui montre que pour une fonction additive, qui assigne un coût à un arbre 2^d -aire et ses sous-arbres, le sous-arbre optimal la minimisant est obtenu par programmation

dynamique procédant par élagage récursif bottom-up de l'arbre. Ce lien a aussi été exploité pour le débruitage [16], ainsi que pour l'approximation non-linéaire d'images géométriques dans le domaine spatial [11, 19], ou celui des ondelettes [15, 20, 17]. L'adaptativité offerte par ce partitionnement dyadique donne accès à des performances bien meilleures que celles de l'approximation non-adaptative.

Apprentissage de la structuration des blocs Dans un travail récent, les auteurs dans [18] proposent d'étendre le cadre de l'apprentissage parcimonieux pour estimer conjointement le dictionnaire des atomes et la structuration par blocs de la parcimonie, à partir d'une base de signaux/images d'apprentissage. Toutefois, la solution du problème de minimisation marginale par rapport à la structuration étant éminemment NP-difficile, ils l'approchent, sans garantie théorique, par un algorithme heuristique de clustering agglomératif.

2 Estimateurs Non-Linéaires par Blocs et Risques

Nous considérons un scénario simple de débruitage, où l'on observe $y = x + w \in \mathbb{R}^N$, avec x le signal original à restaurer, et w un bruit additif Gaussien blanc de variance σ^2 . L'a priori sur le signal x est qu'il exhibe une parcimonie structurée par blocs dans un espace transformé donné.

2.1 Estimateurs par blocs

Une segmentation B par blocs est une partition disjointe de l'ensemble des indices

$$\{0, \dots, N-1\} = \bigcup_{b \in B} b, \quad b \cap b' = \emptyset, \forall b \neq b'.$$

En chaque bloc $b \in B$, nous définissons un estimateur par seuillage $\hat{x} = S_{\lambda, B}(y)$ de x

$$\forall i \in b, \quad \hat{x}[i] = \rho_\lambda(\|y_b\|)y[i], \quad \text{où } \|y_b\|^2 = \frac{1}{|b|} \sum_{i \in b} |y[i]|^2, \quad (1)$$

où $|b|$ est la taille du bloc. En clair, un tel estimateur seuille conjointement tous les coefficients d'un bloc si le rapport signal sur bruit intra-bloc est en dessous d'un certain seuil dépendant de $\lambda \geq 0$, sinon une atténuation isotrope ρ_λ est appliquée à tout le bloc.

Dans ce travail, nous considérons, sans perte de généralité¹, deux règles de seuillage par bloc communément utilisées : le seuillage doux et le seuillage de James-Stein (JS) définis respectivement par

$$\rho_\lambda^{\text{Soft}}(a) = \max\left(0, 1 - \frac{\lambda}{\|a\|}\right), \quad \rho_\lambda^{\text{JS}}(a) = \max\left(0, 1 - \frac{\lambda^2}{\|a\|^2}\right). \quad (2)$$

On peut montrer que chacun de ces deux estimateurs correspond à la solution unique d'un estimateur Bayésien MAP pour une pénalité qui favorise en effet la parcimonie par bloc. Ce résultat est omis ici pour des raisons évidentes de place. Notons que contrairement au seuillage doux, qui souffre d'un biais constant, le seuillage JS engendre un biais bien plus petit qui s'atténue d'autant plus que l'énergie du bloc augmente.

1. D'autres estimateurs par seuillage peuvent être utilisés, pourvu que $a \mapsto \rho_\lambda(a)$ corresponde à une application faiblement différentiable permettant d'appliquer le lemme de Stein discuté ci-après.

2.2 Risques des estimateurs

Risque SURE Le risque quadratique moyen d'un estimateur S est noté $\mathbb{E}_w(\|S(y) - x\|^2)$, et l'espérance est par rapport à la loi du bruit. Le meilleur estimateur est celui minimisant cette quantité. Toutefois, sans connaissance d'un oracle donnant accès au signal original x , minimiser le risque est illusoire.

Sous hypothèse de différentiabilité faible de $y \in \mathbb{R}^n \mapsto S(y)$, le lemme de Stein permet d'obtenir une estimée sans biais du risque, appelée SURE, qui ne dépend que de l'observation y et aucunement du signal x . Son expression générale s'écrit

$$J(y, S) = n\sigma^2 + \|y - S(y)\|^2 + 2\sigma^2 \text{div}(S - \text{Id})(y) \quad (3)$$

où $\text{div}(f)(x) = \sum_i \frac{\partial f_i}{\partial x_i}(x)$ est la divergence de $f = (f_i)_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$. Le risque SURE est sans biais puisqu'on peut montrer que

$$\mathbb{E}_w(\|S(y) - x\|^2) = \mathbb{E}_w(J(y, S)).$$

Il est ainsi naturel d'utiliser (3), calculé à partir d'une (ou plusieurs) réalisation y , comme une estimée du risque, et de la minimiser pour adapter les hyperparamètres de l'estimateur S (seuil, taille des blocs) comme nous l'avons vu en introduction. Dans ce travail, nous proposons le cadre du risque de Stein pour estimer à la fois le seuil et la structuration optimale des blocs.

Risque d'une partition par bloc Par application de la formule (3), le risque SURE de chaque bloc $b \in B$ associé à chacun des estimateurs $y_b \mapsto \rho_\lambda(\|y_b\|)y_b$ définis dans (1)-(2) est :

$$J^{\text{Soft}}(y_b, \lambda, \sigma) = |b|\sigma^2 + (|b|\|y_b\|^2 - 2|b|\sigma^2) I(\|y_b\| < \lambda) + \left(|b|\lambda^2 - 2\sigma^2(|b| - 1) \frac{\lambda}{\|y_b\|}\right) I(\|y_b\| \geq \lambda) \quad (4)$$

$$J^{\text{JS}}(y_b, \lambda, \sigma) = |b|\sigma^2 + (|b|\|y_b\|^2 - 2|b|\sigma^2) I(\|y_b\| < \lambda) + \frac{|b|\lambda^2 - 2\sigma^2(|b| - 2)}{\|y_b\|^2/\lambda^2} I(\|y_b\| \geq \lambda), \quad (5)$$

où $I(\omega)$ est l'indicatrice de l'événement ω . Remarquons au passage que $J(y_b, \lambda, \sigma) = \sigma^2 J(y_b/\sigma, \lambda/\sigma, 1)$, et que $|b|\|y_b\|^2/\sigma^2$ suit une loi du $\chi_{|b|}^2(|b|\|x_b\|^2/\sigma^2)$ non-centrale de paramètre de non-centralité $|b|\|x_b\|^2/\sigma^2$ et à $|b|$ degrés de liberté. Par cette simple remarque et quelques arrangement algébriques, on peut montrer le résultat suivant.

Proposition 1. *Considérons $\sigma = 1$. Posons $L = |b|$, $\mu = L\|x_b\|^2/2$ et $\lambda_b = \lambda\sqrt{L}$. Pour tout $L > 2$,*

$$\mathbb{E}_w(J^{\text{Soft}}(y_b, \lambda, 1)) = 2\mu + \sum_{k \in \mathbb{N}} \frac{e^{-\mu} \mu^k}{k!} \left(-\sqrt{2}\lambda_b(L-1) \frac{\Gamma(\frac{L-1}{2} + k)}{\Gamma(\frac{L}{2} + k)} P_{L+2k-2} - (L+2k)P_{L+2k+2} + L(\lambda_b^2 + 2)P_{L+2k} \right) \quad (6)$$

$$\mathbb{E}_w(J^{\text{JS}}(y_b, \lambda, 1)) = 2\mu + \sum_{k \in \mathbb{N}} \frac{e^{-\mu} \mu^k}{k!} \left(\frac{\lambda_b^2(4-L(\lambda_b^2-2))}{L+2k-2} P_{L+2k-2} - (L+2k)P_{L+2k+2} + 2LP_{L+2k} \right) \quad (7)$$

$$\text{où } P_\nu = \Pr(\chi_\nu^2 > \lambda_b^2) = \frac{\Gamma(\nu/2, \lambda_b^2)}{\Gamma(\nu/2)}.$$

Ces expressions du risque ont l'avantage de ne dépendre que de μ , et non pas de x_b , ce qui est une différence de taille.

Etant donné une structuration des blocs B , le risque qui lui est associé est alors

$$J_B(y, \lambda, \sigma) = \sum_{b \in B} J(y_b, \lambda, \sigma), \quad (8)$$

où $J(y_b, \lambda, \sigma)$ est (4) où (5).

Estimateur empirique Bien que le SURE (avec une seule réplification de y) soit un estimateur sans biais du risque, il peut présenter une variance non-négligeable. On peut montrer que c'est typiquement le cas pour nos estimateurs Soft et JS. Ainsi, estimer la partition optimale en minimisant le risque SURE peut s'avérer assez imprécis, particulièrement en présence de bruit. Pour contourner ce problème, nous proposons de recourir à de l'intégration Monte-Carlo, à partir des expressions (4)-(5)². En effet, une estimée potentiellement biaisée mais à variance inférieure de $\mathbb{E}_w(J^{\text{JS}}(y_b, \lambda, \sigma))$ est donnée par

$$\hat{J}^{\text{JS}}(y_b, \lambda, \sigma) = \frac{1}{K} \sum_{i=1}^K \left(|b|\sigma^2 + (z_i - 2|b|\sigma^2)I(z_i < \lambda^2|b|) + \frac{\lambda^2|b|(\lambda^2|b| - 2\sigma^2(|b| - 2))}{z_i} I(z_i \geq \lambda^2|b|) \right), \quad (9)$$

où z_i sont des réalisations d'une variable $\sigma^2 \chi_{|b|}^2(\hat{\mu})$, $\hat{\mu} = |b| \max(0, \|y_b\|^2/\sigma^2 - 1)$. Le biais peut provenir de l'estimateur du paramètre de non-centralité utilisé. Une formule similaire peut s'écrire pour \tilde{J}^{Soft} à partir de (4). Dans nos expérimentations, nous avons utilisé typiquement $K = 100$ réalisations. Pour lever toute ambiguïté par la suite, nous éviterons la terminologie SURE, car notre estimateur n'en ait pas stricto sensu, et nous dénommerons le risque ainsi estimé le risque de Stein.

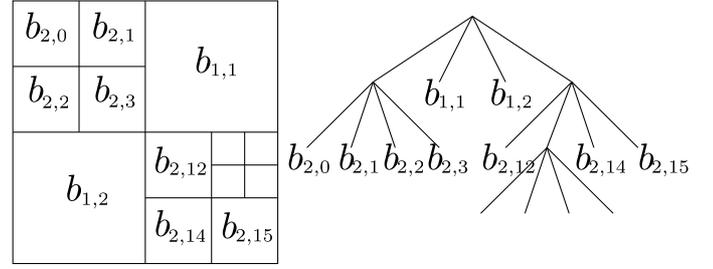
3 Structuration de la Parcimonie par Partitionnement Dyadique

Rappelons que nous visons le couple (λ^*, B^*) minimiseur global de $\min_{\lambda, B} \mathbb{E}_w(\|S_{B, \lambda}(y) - x\|^2)$. Sans accès à x , ce couple est approché par le minimiseur global du risque de Stein (9). Sans structuration additionnelle sur la nature des partitions, ce problème de d'optimisation est évidemment combinatoire. Pour obtenir un algorithme rapide exploitant la méthodologie CART, nous exploitons le fait que les partitions des blocs ont une structure hiérarchique en un arbre obtenu par subdivision récursive dyadique des blocs de la racine vers les feuilles.

3.1 Partitionnement par quadtree

Pour des raisons de simplicité, nous détaillons le cas 2D des subdivisions en quadtree. Notre procédure reste toutefois valide en toute dimension et pour des schémas de subdivision arbitraires possiblement non-stationnaires. Ainsi, dans ce qui reste, l'ensemble $\{0, \dots, N - 1\}$ indexe les pixels en 2D, les blocs b correspondent à des carrés dyadiques regroupant des sous-ensembles de ces pixels.

Une partition B est associée à un quadtree T qui est sous-arbre de la forêt d'arbres contenant toutes les partitions dyadiques possibles. Les blocs d'une telle partition sont les feuilles $\mathcal{L}(T)$ de ce sous-arbre T . T liste un ensemble de blocs $b_{j,i}$ indexés par leur profondeur $0 \leq j \leq J = \log_2(N)/2$ dans l'arbre, et leur position $0 \leq i < 4^j$. La Fig. 1 montre un exemple de quadtree T (droite) et la partition dyadique correspondante (gauche), où chaque bloc $b_{j,i}$ correspond à un carré de $N/4^j$ pixels et ayant pour fils quatre blocs $(b_{j+1,4i}, b_{j+1,4i+1}, b_{j+1,4i+2}, b_{j+1,4i+3})$ de taille $N/4^{j+1}$.



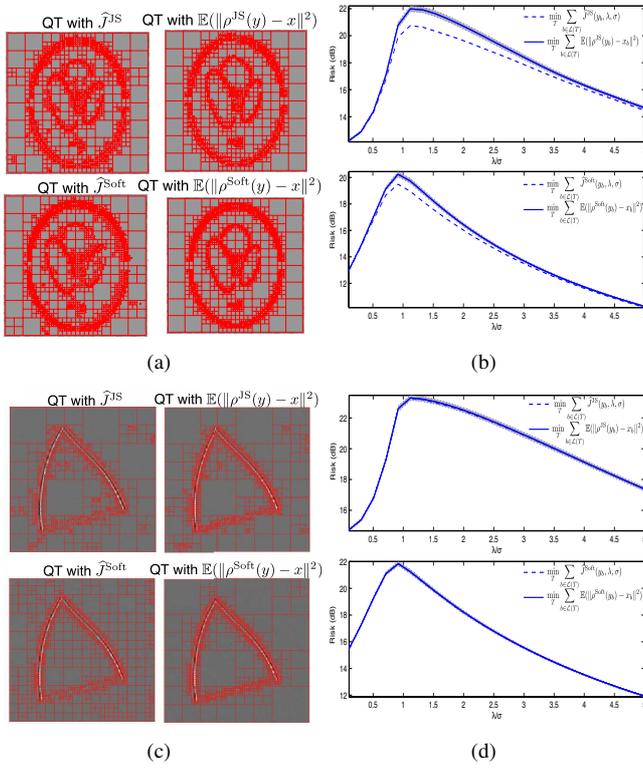


FIGURE 2 – (a) : Exemple de partition dyadique (QT) d’une sous-bande ondelettes de deux images en minimisant le risque exact (avec oracle), et notre estimateur du risque de Stein. (b) : Le risque (en dB) exact et de Stein en fonction de λ/σ . La zone grisée correspond à l’intervalle de confiance puisque l’espérance dans le risque exact est estimée par une moyenne sur 1E3 réalisations.

Débruitage dans le domaine des ondelettes La Fig. 3 des résultats de débruitage utilisant des ondelettes invariantes par translation. La valeur de λ a été aussi optimisée pour le seuillage par blocs de taille fixe. L’estimateur JS étant sujet à moins de biais, il est systématiquement meilleur que le seuillage doux. L’adaptation à la structuration apporte une amélioration notable visuelle et quantitative du résultat de débruitage, comparée à utiliser une taille de bloc fixe.

5 Conclusion

Nous avons proposé un nouveau cadre pour l’estimation adaptative de la structuration de la parcimonie en blocs dyadiques. Son potentiel d’application a été illustré sur le problème de débruitage, mais nous pensons pouvoir l’étendre judicieusement aux problèmes inverses. Cet aspect, en plus de quelques questions théoriques ouvertes sont autant de perspectives de ce travail.

Références

- [1] F. R. Bach. Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.*, 9 :1179–1225, June 2008.
- [2] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inf. Theo.*, 56 :1982–2001, 2010.
- [3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*. Wadsworth, Belmont, CA, 1984.
- [4] T. Cai. Adaptive wavelet estimation : a block thresholding and oracle inequality approach. *Ann. Statist.*, 27(3) :898–924, 1999.

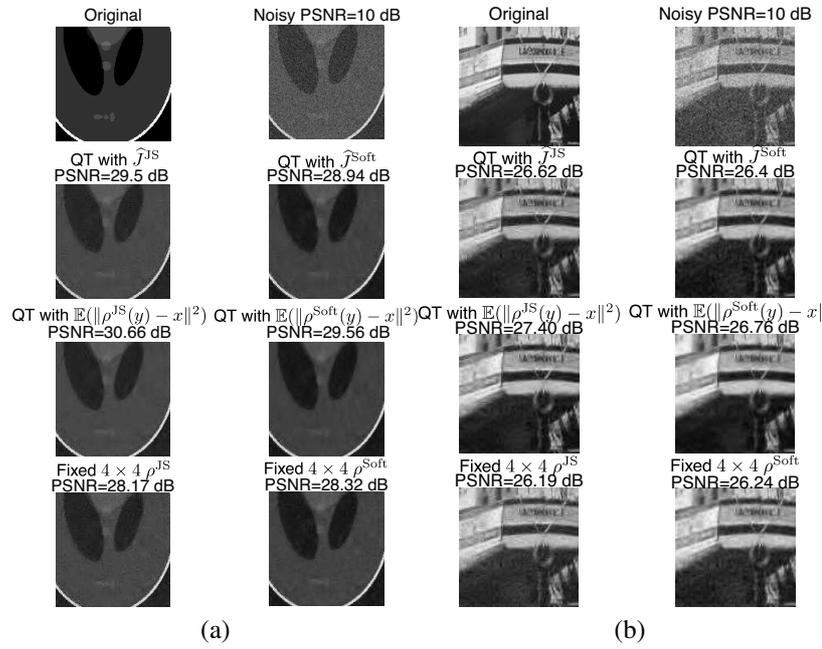


FIGURE 3 – Zoom on denoising results PSNR=10dB. (a) : Synthetic image. (b) : Natural image.

- [5] T. Cai and B.W. Silverman. Incorporate Information on Neighboring Coefficients into Wavelet Estimation. *Sankhya*, 63 :127–148, 2001.
- [6] T. Cai and H. Zhou. A data-driven block thresholding approach to wavelet estimation. *Annals of Statistics*, 37(2) :569–595, 2009.
- [7] C. Chau, L. Duval, A. Benazza-Benyahia, and J.-C. Pesquet. A nonlinear stein based estimator for multichannel image denoising. *IEEE Trans. on Sig.*, 56(8) :3855–3870, 2008.
- [8] C. Chesneau, M.J. Fadili, and J.-L. Starck. Stein block thresholding for image denoising. *App. Comp. Harm. Anal.*, 28 :67–88, 2009.
- [9] C. Chesneau, M.J. Fadili, and J.-L. Starck. Stein block thresholding for wavelet-based image deconvolution. *Elec. J. Statistics*, 4 :415–435, 2010.
- [10] R. Coifman and V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Inf. Theo.*, IT-38(2) :713–718, Mar. 1992.
- [11] D. Donoho. Wedgelets : Nearly minimax estimation of edges. *Annals of Stat.*, 27(3) :859–897, 1999.
- [12] D. L. Donoho. Cart and best-ortho-basis : A connection. *Annals of Stat.*, 25(5) :1870–1911, 1997.
- [13] D. L. Donoho and I. Johnstone. Minimax estimation by wavelet shrinkage. *Annals of Stat.*, 26 :879–921, 1998.
- [14] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81 :425–455, 1994.
- [15] P.L. Dragotti and M. Vetterli. Wavelet footprints : Theory, algorithms and applications. *IEEE Trans. Signal Proc.*, 51(5) :1306–1323, Mat 2003.
- [16] H. Krim, D. Tucker, S. Mallat, and D. Donoho. On denoising and best signal representation. *IEEE Trans. Info. Theory*, 45(7) :319–335, Nov 1999.
- [17] S. Mallat and G. Peyré. Orthogonal bandlet bases for geometric images approximation. *Commun. on Pure and Appl. Math.*, 61(9) :1173–1212, 2008.
- [18] K. Rosenblum, L. Zelnik-Manor, and Y. Eldar. Dictionary optimization for block-sparse representations. In *AAAI Fall 2010 Symposium on Manifold Learning*, 2010.
- [19] R. Shukla, P.L. Dragotti, M. Do, and M. Vetterli. Rate distortion optimized tree structured compression algorithms for piecewise smooth images. *IEEE Trans. Image Proc.*, 14(3), 2005.
- [20] M. Wakin, J. Romberg, H. Choi, and R. Baraniuk. Wavelet-domain approximation and compression of piecewise smooth images. *IEEE Trans. Image Proc.*, 15(5) :1071–1087, May 2006.
- [21] G. Yu, S. Mallat, and E. Bacry. Audio denoising by time-frequency block thresholding. *IEEE Trans. on Sig.*, 56(5) :1830–1839, 2008.
- [22] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. of The Roy. Stat. Soc. B*, 68(1) :49–67, 2006.