

Une petite introduction aux représentations parcimonieuses de signaux et d'images

PAR JÉRÔME LANDRÉ

2008-07-01



Table des matières

1	Introduction	2
2	Représentation « classique » des signaux	2
3	Représentations parcimonieuses	3
3.1	Représentation exacte	3
3.2	Représentation approchée	3
4	Etat de l'art	4
4.1	Problématique scientifique	4
4.2	Critère de parcimonie	4
4.3	Algorithmes de décomposition	5
4.3.1	Meilleure base	5
4.3.2	Poursuite de base	5
4.3.3	Poursuite adaptative	5
4.4	Construction de dictionnaire	5
4.5	Applications	6
5	Perspectives de recherche	7
5.1	Indexation d'images	7
5.2	Création d'une bibliothèque de fonctions	8
	Bibliographie	8

1 Introduction

Les représentations parcimonieuses connaissent une évolution croissante depuis quelques années pour la communauté du traitement du signal et des images. Cet article tente de résumer l'état actuel des connaissances à propos de ce sujet qui intéresse de plus en plus d'équipes de recherche.

Voici tout d'abord une définition d'un signal parcimonieux [21] : « Un signal est dit parcimonieux lorsque la plupart de ses coefficients sont (approximativement) nuls ». Les figures 1 et 2 montrent des exemples de signaux parcimonieux.

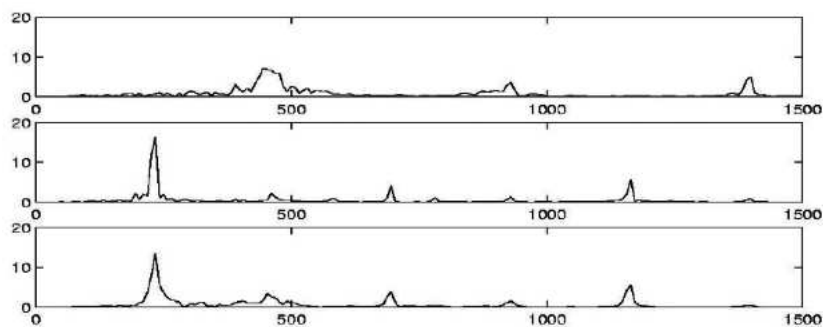


Figure 1. Trois exemples de signaux parcimonieux.



Figure 2. Transformée DCT-2D de l'image cameraman, les valeurs non-nulles sont concentrées en haut à gauche sur l'image transformée, la plupart des valeurs de niveaux de gris dans l'image sont nulles.

Les signaux sont classiquement représentés sous la forme d'une transformation linéaire de leurs coefficients. La section suivante présente cette représentation « classique ».

2 Représentation « classique » des signaux

Les techniques classiques de représentation des signaux utilisent en général une description des composantes sur une base sur laquelle la représentation du signal est unique (base de Fourier, base d'ondelettes orthogonales par exemple).

Soit y un signal réel de taille N ($y \in \mathbb{R}^N$), la représentation du signal sur la base B s'écrit matriciellement :

$$y = Bx \quad (1)$$

B est une base de \mathbb{R}^N , les vecteurs $b_i \in \mathbb{R}^N$ sont des vecteurs orthogonaux, $\forall i \neq j, b_i \perp b_j$, de norme unitaire, $\|b_i\| = 1$. Il y a unicité de x , décomposition de y sur la base B . La transformation est inversible (si B est inversible) et, dans ce cas, on a directement :

$$\begin{aligned} B^{-1}y &= B^{-1}Bx \\ B^{-1}y &= x \end{aligned} \quad (2)$$

Depuis quelques années, de nombreux chercheurs s'intéressent à une autre technique de représentation des signaux et des images : les représentations parcimonieuses.

3 Représentations parcimonieuses

À l'inverse, les représentations parcimonieuses (*sparse representations*) consistent en la décomposition du signal sur un dictionnaire comprenant un nombre d'éléments (ou atomes) très supérieur à la dimension du signal. Cette décomposition va introduire dans la nouvelle représentation du signal un grand nombre de valeurs nulles, on parle de représentation parcimonieuse pour cette raison. On trouve aussi les termes représentation « éparsée » et représentation « creuse ».

Soit y un signal réel de taille N ($y \in \mathbb{R}^N$), on cherche à représenter ce signal par sa décomposition sur une famille génératrice liée de K vecteurs de taille N avec $K > N$.

3.1 Représentation exacte

On se place dans le cas où on souhaite une représentation du signal sans perte d'information, c'est-à-dire qu'il y a la même information dans le signal représenté x et dans le signal de départ y .

On note :

$$y = Dx \quad (3)$$

En notation matricielle :

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} d_{1,1} & \dots & d_{1,K} \\ \vdots & \ddots & \vdots \\ d_{N,1} & \dots & d_{N,K} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix}$$

Où les vecteurs $d_i = \begin{pmatrix} d_{1,i} \\ \vdots \\ d_{N,i} \end{pmatrix}$ sont des vecteurs générateurs liés de l'espace \mathbb{R}^N .

Le dictionnaire étant redondant, il existe une infinité de solutions pour l'équation (3). Il faut donc trouver des méthodes pour choisir parmi cette infinité de décompositions celle qui va répondre, selon un critère à définir, au problème posé. Par exemple pour une application de compression sans perte d'information (*lossless compression*), on va chercher la décomposition la plus creuse possible (celle contenant le plus de termes nuls) afin de compresser efficacement le signal.

3.2 Représentation approchée

Lorsqu'on souhaite une représentation approchée du signal (par exemple pour la compression avec perte d'information, *lossy compression*), on utilise un critère de différence entre le signal de départ et sa représentation.

Soit $\epsilon \in \mathbb{R}^{*+}$, $\epsilon > 0$, l'équation (4) donne la condition pour avoir une bonne représentation du signal en fixant ϵ aussi petit que l'on veut, selon une norme L^p .

$$\|y - Dx\|_p < \epsilon \quad (4)$$

La norme L^p étant définie par : $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$.

L'idée est de se rapprocher aussi près que l'on souhaite du signal y de départ en réglant le paramètre ϵ .

4 Etat de l'art

Depuis quelques années, les représentations parcimonieuses de signaux et d'images font l'objet de développements importants [20], [9], [10]. Le développement de ces techniques est issu des travaux de Stéphane Mallat [14], [3]. Une étude bibliographique assez complète sur le sujet [5] a été proposée en 2007 et est en cours de publication.

4.1 Problématique scientifique

Les problématiques scientifiques que posent ces représentations sont au nombre de trois :

1. La définition du **critère de parcimonie** est une étape importante de toutes les méthodes utilisant ces représentations,
2. Il est nécessaire de trouver des **méthodes de calcul de x** , connaissant le signal y et le dictionnaire D ,
3. D'autre part, il faut aussi développer des techniques de **construction du dictionnaire D** en fonction de l'application à réaliser.

Ces trois aspects ont donné lieu à de nombreuses publications qui sont au coeur de la thématique des représentations parcimonieuses de signaux et d'images.

4.2 Critère de parcimonie

La résolution des équations 3 et 4 nécessite de définir des contraintes sur les coefficients α_i . Ainsi, la décomposition obtenue sera la meilleure selon ces contraintes. En général, ce critère de parcimonie s'écrit :

$$\hat{\alpha} = \arg \min_{\alpha} F(\alpha) \quad (5)$$

$F(\alpha)$ est une fonction à définir, mais en général, on prend la norme L^p des vecteurs dans l'espace \mathbb{R}^N . La norme L^p étant définie par : $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$.

Sachant que la norme L^0 mesure le nombre de termes non-nuls dans la décomposition x . En minimisant L^0 , on obtient donc la décomposition optimale comprenant un nombre minimal de termes non-nuls dans x . Toutefois, la minimisation exacte selon L^0 fait partie de la classe des problèmes NP-complets (*NP-hard*).

On se contente donc de travailler avec la norme L^1 , c'est-à-dire la somme des valeurs absolues des coefficients. Ce critère de norme L^1 offre une solution moins optimale que la norme L^0 , mais qui a l'avantage d'être calculable par des techniques de programmation linéaire.

4.3 Algorithmes de décomposition

Il existe trois approches principales pour obtenir la décomposition d'un signal y sur un dictionnaire redondant D :

4.3.1 Meilleure base

L'approche « meilleure base » (*Best basis*) [2] considère le dictionnaire D comme la réunion de B bases distinctes, puis à rechercher (exhaustivement ou non) parmi ces bases celle qui offre la décomposition optimale selon le critère de parcimonie retenu.

4.3.2 Poursuite de base

L'approche « Poursuite de base » (*Basis pursuit*) [6] utilise des techniques de programmation linéaire pour minimiser la norme L_1 de la décomposition. La complexité de l'algorithme est importante, la solution est en général satisfaisante, sans toutefois atteindre le résultat qui aurait été obtenu avec la norme L_0 . Mais l'utilisation de la norme L_0 est un problème d'optimisation NP-complet, non calculable sur un ordinateur.

4.3.3 Poursuite adaptive

L'algorithme « Poursuite adaptative » (*Matching pursuit*) [14] et son extension « Poursuite adaptative orthogonale » (*Orthogonal matching pursuit*) [18] consistent à trouver itérativement dans le dictionnaire l'élément (vecteur) le plus corrélé avec le signal, d'ôter la contribution de ce vecteur au signal et de recommencer le processus jusqu'à obtention d'un nombre d'éléments égal à la dimension du signal. Il suffit ensuite de décomposer le signal sur la famille génératrice obtenue pour trouver les coefficients α_i . Cet algorithme de type « glouton » (*greedy*) n'est pas optimal, mais il offre de bonnes propriétés de décroissance de l'erreur entre le signal original et le signal parcimonieux.

Il existe d'autres méthodes hybrides basées sur les trois méthodes ci-dessus.

4.4 Construction de dictionnaire

Le choix du dictionnaire D a une très grande influence sur la qualité de la décomposition x du signal y . En effet, si le dictionnaire ne contient pas du tout ou très peu d'atomes adaptés aux structures présentes dans le signal y , la décomposition x sera, quelle que soit la méthode choisie, mauvaise.

Le choix du dictionnaire peut être orienté selon plusieurs critères :

1. Connaissances a priori :
 1. Simplicité calculatoire : exemple des atomes de Gabor, des paquets d'ondelettes et des cosinus locaux qui possèdent des propriétés intéressantes,
 2. Répartition et variation fréquentielle des signaux à analyser : atomes harmoniques, chirplets, ...
2. Apprentissage à partir des données :
 1. Signaux de même nature : l'analyse en composantes indépendantes permet de construire des dictionnaires adaptés à une famille similaire de signaux,
 2. Approche probabiliste : algorithme approximation-maximisation (Expectation-Maximization EM)

Une technique récente (k-SVD) [1] donne un algorithme qui trouve un dictionnaire optimal selon le critère de parcimonie retenu pour un signal ou une image.

4.5 Applications

Les utilisations principales des représentations parcimonieuses sont le débruitage de signaux et d'images [?], la séparation de sources [4], la représentation de fonctions [11] et la compression [15] en raison de la nature très compressible de cette représentation (beaucoup de coefficients nuls dans les vecteurs x et codage simple du dictionnaire D).

L'équipe METISS [16] de l'INRIA effectue des recherches en séparation de sources qui utilisent les représentations parcimonieuses de signaux sonores. Une librairie de développement basée sur les représentations parcimonieuses MPTK (Matching pursuit ToolKit) [17] est présente sur le site de cette équipe.

La génération d'un dictionnaire pour les applications de traitement d'images peut-être réalisée soit à partir d'un dictionnaire pré-calculé correspond à une famille de fonctions connues : ondelettes, cosinus discret, soit à partir des images elles-mêmes, ce qui permet une adaptation très forte de la représentation parcimonieuse aux images considérées.

Ainsi le dictionnaire de signature d'images (*Image-Signature-Dictionary*) [2] permet de construire un dictionnaire adapté à la famille d'images à traiter par apprentissage.

Plus récemment, l'algorithme k-SVD a été appliqué à la construction de dictionnaire pour la segmentation et la génération de textures [19] avec d'excellents résultats. La figure 3 donne un exemple de résultat d'apprentissage d'un dictionnaire à partir d'une texture. La figure 4 donne un exemple de synthèse de texture à partir du dictionnaire appris précédemment.

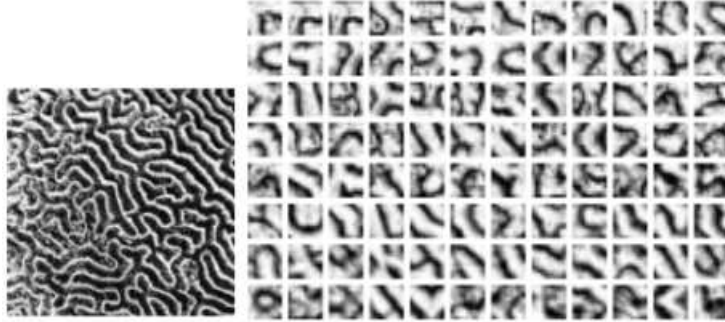


Figure 3. Un exemple de construction de dictionnaire à partir d'une image de texture.
(Image Gabriel Peyré, CEREMADE)

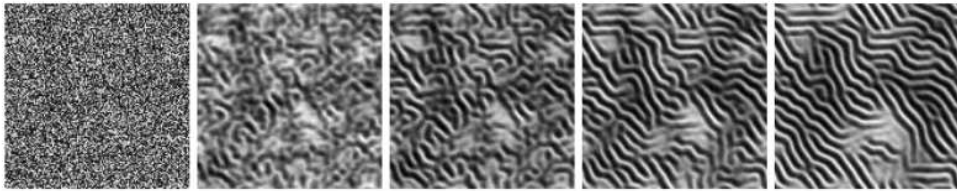


Figure 4. Exemple de synthèse de texture par itérations et contrôle de la parcimonie (ici $s=2$).
(Image Gabriel Peyré, CEREMADE)

L'idée de combiner les représentations parcimonieuses à l'analyse multi-échelle [12] apporte un intérêt très important pour le codage arborescent des informations contenues dans les images. L'idée est d'analyser de façon multi-échelle les informations contenues dans les images à partir d'apprentissage multi-échelles de dictionnaires. La figure 5 donne la structure arborescente de la décomposition. L'idée de la décomposition multi-échelle utilisant plusieurs dictionnaires est donnée en figure 6. La figure 7 représente un dictionnaire multi-échelle appris sur l'image « bateau ». Un exemple d'application pour le débruitage d'image est donné et propose d'excellents résultats.

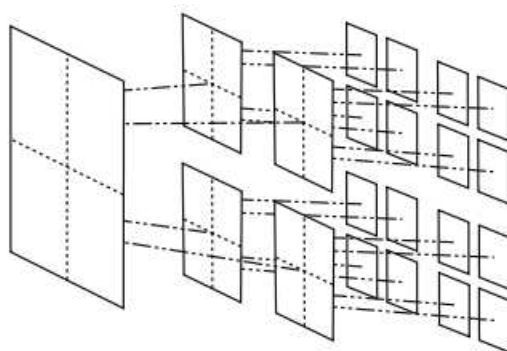


Figure 5. Exemple de décomposition multi-échelle sur une hiérarchie de dictionnaires.
(Image Julien Mairal, INRIA)

$$\begin{aligned}
 & \text{Image} = \alpha_0 \begin{matrix} \blacksquare \\ \blacksquare \end{matrix} + \alpha_1 \begin{matrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{matrix} + \alpha_2 \begin{matrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{matrix} + \alpha_3 \begin{matrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{matrix} + \alpha_4 \begin{matrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{matrix} + \alpha_5 \begin{matrix} \blacksquare & \blacksquare \\ \blacksquare & \blacksquare \end{matrix} + \\
 & \alpha_6 \begin{matrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{matrix} + \alpha_7 \begin{matrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{matrix} + \alpha_8 \begin{matrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{matrix} + \alpha_9 \begin{matrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{matrix} + \alpha_{10} \begin{matrix} \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare & \blacksquare \end{matrix} + \dots
 \end{aligned}$$

Figure 6. Exemple de décomposition multi-échelle d'un bloc d'image.
(Image Julien Mairal, INRIA)

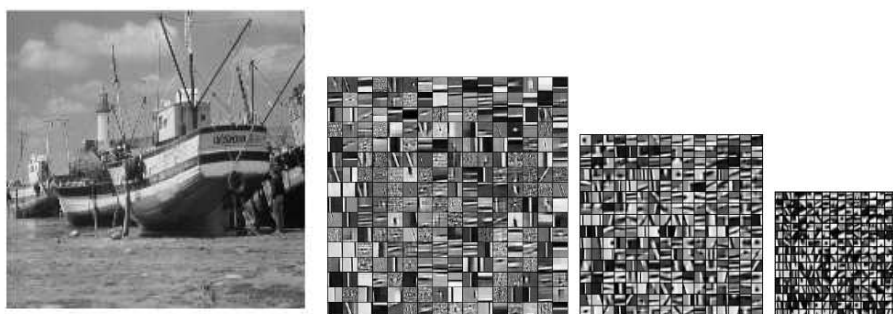


Figure 7. Une image et son dictionnaire multi-échelles (trois dans cet exemple) appris.
(Image Julien Mairal, INRIA)

Les travaux récents indiquent que les représentations parcimonieuses sont parfaitement adaptées au débruitage (*denoising*) et à la reconstruction d'images dégradées (*inpainting*) en niveaux de gris et en couleurs [13]. Il y a peu (voire pas du tout) d'application en indexation d'images à ce jour. Ce manque d'application peut ouvrir des perspectives de recherche dans cette voie.

5 Perspectives de recherche

Cette partie présente mon projet de recherche à moyen (et/ou long) terme sur les représentations parcimonieuses de signaux et d'images.

5.1 Indexation d'images

Les représentations parcimonieuses semblent très bien adaptées à la représentation très concise d'information. En indexation d'images, on cherche des représentants des images sous la forme de vecteurs, on utilise ensuite des techniques de classification sur ces vecteurs afin de classer les images.

L'utilisation de dictionnaires calculés sur des exemples d'images de familles connues pourrait permettre de classer facilement les images d'une base de données en familles visuellement similaires à l'aide de distances calculées sur des vecteurs éparses.

L'utilisation de dictionnaires multi-échelles pourrait offrir la possibilité de travailler avec des vecteurs à plusieurs niveaux pour accélérer les temps de recherche d'images (technique proposée dans ma thèse).

La principale question à résoudre est la stabilité de l'algorithme de construction de dictionnaire sur différentes images. En d'autres termes, le dictionnaire calculé sur une image d'une famille donnée sera-t-il proche du dictionnaire calculé sur une autre image de la même famille ?

Je propose donc d'étudier les représentations parcimonieuses appliquées au domaine l'indexation et de la recherche d'images. C'est mon projet de recherche pour les années à venir. Ces travaux, en raison de leur originalité, seront sans doute source de publications dans des conférences et dans des revues internationales.

5.2 Création d'une bibliothèque de fonctions

Il existe peu de bibliothèques permettant de manipuler les représentations parcimonieuses de signaux et d'images. En parallèle avec les recherches en indexation d'images, le développement d'une bibliothèque de fonctions disponible pour les chercheurs du domaine pourrait permettre des échanges et des avancées dans le domaine.

L'idée de base est de programmer une bibliothèque sous la forme d'un plugin ImageJ (en Java) et selon les temps de calcul et les performances obtenues, proposer une version C/C++ de la bibliothèque sous OpenCV avec appui éventuel sur des librairies bas-niveaux propriétaires (Intel IPP pour les processeurs Intel et AMD APL pour les processeurs AMD).

La création de cette bibliothèque permettrait de produire de nombreuses publications dans des conférences et des revues internationales et offrirait des possibilités d'échange et de collaboration avec de nombreux acteurs internationaux.

Bibliographie

- [1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, November 2006.
- [2] Michal Aharon and Michael Elad. Sparse and redundant modeling of image content using an image-signature-dictionary (isd). *To appear in SIAM*, 2008.
- [3] F. Bergeaud and S. Mallat. Matching pursuit of images. In *ICIP*, pages 53–56, 1995.
- [4] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden. Sparsity and morphological diversity in blind source separation. *IEEE Transactions on Image Processing*, 16(11):2662–2674, Nov. 2007.
- [5] A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *to appear in SIAM Review*, 2008.
- [6] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal of Sci. Comput.*, 20(1):33–61, 1999.
- [7] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15(12):3736–3745, December 2006.
- [8] M.J. Fadili, J.-L. Starck, and F. Murtagh. Inpainting and zooming using sparse representations. *The Computer Journal, Oxford Journals*, July 2007.
- [9] Rosa Maria Figueras i Ventura. *Sparse Image Approximation with Application To Flexible Image Coding*. PhD thesis, EPFL, September 2005.
- [10] Rémi Gribonval. *Sur quelques problèmes mathématiques de modélisation parcimonieuse*. PhD thesis, HDR, Université de Rennes 1, October 2007.
- [11] P. Honein , C. Richard, and J.-C. Bermudez. Online nonlinear sparse approximation of functions. In *IEEE ISIT'07*, june 2007. Nice - France.
- [12] G. Sapiro J. Mairal and M. Elad. Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Modeling and Simulation*, 7(1):214–241, April 2008.

- [13] J. Mairal, M. Elad, and G. Sapiro. Sparse representation for color image restoration. *IEEE Transactions on Image Processing*, 17(1):53–69, January 2008.
- [14] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.
- [15] Davide Mattered, Francesco Palmieri, and Michele Di Monte. A comparison of signal compression methods by sparse solution of linear systems. In *WIRN VIETRI 2002: Proceedings of the 13th Italian Workshop on Neural Nets-Revised Papers*, pages 146–151, London, UK, 2002. Springer-Verlag.
- [16] METISS. 2007 research project activity report. Technical report, INRIA, 2007.
- [17] MPTK. Matching pursuit toolkit. IRISA, 2008.
- [18] Y.C. Pati, R. Rezaifar, and P.S. Krishnaprasad. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1:40–44, November 1993.
- [19] Gabriel Peyré. Sparse modeling of textures. CEREMADE, université Paris Dauphine, 2007.
- [20] Joel Aaron Trop. *Topics in Sparse Approximation*. PhD thesis, University of Texas at Austin, August 2004.
- [21] Allen Y. Yang. Compressed sensing meets machine learning - classification of mixture subspace models via sparse representation. Mini Lectures in Image Processing, TRUST Center Seminar, University of California, Berkeley, USA, 2008.