

Identification de signaux parcimonieux dans un dictionnaire hybride

Matthieu KOWALSKI, Bruno TORRESANI,

Laboratoire d'Analyse, Topologie et Probabilités, Université de Provence,
39 rue F. Joliot-Curie, 13453 Marseille cedex 13, France

Matthieu.Kowalski@cmi.univ-mrs.fr, Bruno.Torrésani@cmi.univ-mrs.fr

Résumé – Nous proposons et étudions une nouvelle famille d'algorithmes de décomposition de signaux sur des systèmes hybrides (par exemple, une unions de bases), basés sur une modélisation probabiliste. La modélisation repose sur deux ingrédients : un modèle pour les coefficients de la décomposition, et un modèle de *carte de signifiante*, décrivant les positions des coefficients significatifs dans l'espace des indices. Deux types de modélisation des cartes de signifiante sont proposées. La première, suivant un modèle de Bernoulli et appelée *non structurée*, ne privilégie aucune relation de dépendance entre coefficients significatifs. La seconde, qui repose sur un modèle de Bernoulli-hiérarchique, favorise certaines *structures* dans le domaine des indices.

Les algorithmes sont comparés et illustrés par des applications de débruitage

Abstract – We propose a new family of algorithm for expanding signals on a hybrid dictionary (for example, a union of bases), based on a probabilistic modeling. The models mainly rest on two ingredients: a model for the coefficients of the expansion, and a model for the index set of significant coefficients, termed *significance map*. Two types of significance map models are proposed. The first one, based upon a Bernoulli model, and called *unstructured*, does not favor any dependency between significant coefficients. The second one, based on a hierarchical Bernoulli model, favors certain types of *structures* in the index set.

The algorithms are compared and illustrated by denoising applications

1 Introduction

Un signal admet une décomposition parcimonieuse par rapport à un dictionnaire (base, ou repère redondant) de formes d'ondes $\{\varphi_n, n \in I\}$, I étant un index fixé, lorsqu'il peut s'écrire comme combinaison linéaire

$$x = \sum_{n=1}^N \alpha_{i_n} \varphi_{i_n},$$

N étant un entier très petit par rapport à la taille de l'index I . Dans ce travail, on se limite à des dictionnaires formés comme union de deux base orthonormées (ici des bases MDCT). L'existence de décompositions parcimonieuses permet de simplifier considérablement un certain nombre de tâches telles que la compression, le débruitage ou encore la séparation de sources.

Il existe plusieurs approches possibles pour l'identification d'une décomposition parcimonieuse d'un signal, parmi lesquelles on peut notamment citer les algorithmes de poursuite et les approches variationnelles. Une approche classique est de minimiser l'erreur quadratique entre le signal estimé et le signal de référence (souvent bruité), tout en minimisant la norme L_1 de la décomposition du signal dans un dictionnaire. Cette approche ainsi que plusieurs algorithmes sont décrits en détail dans [1, 2, 3].

Nous décrivons ici une approche probabiliste, permettant de formuler l'identification d'une décomposition parcimonieuse comme un problème d'estimation. L'estimation porte essentiellement sur les paramètres du modèle et les formes d'onde du dictionnaire présentes dans le dé-

veloppement, et est suivie par une régression du signal sur les formes d'onde sélectionnées.

L'estimation peut également être effectuée dans un cadre Bayésien, via des algorithmes de type MCMC [4]. Ces algorithmes sont *a priori* plus puissants, et fournissent des estimateurs MAP et MMSE des modèles étudiés. L'approche que nous proposons est toutefois beaucoup plus efficace algorithmiquement, et fournit des résultats de qualité sensiblement équivalente.

2 Signaux aléatoires hybrides

Soit \mathcal{H} un espace de Hilbert, qu'on suppose de dimension finie pour simplifier. Soient \mathbf{V} et \mathbf{U} deux bases orthonormées de \mathcal{H} , et soit $\mathcal{D} = \mathbf{V} \cup \mathbf{U}$ le dictionnaire constitué par l'union de ces deux bases. Nous considérons des modèles de signaux de la forme

$$x = \sum_{n \in I} X_n \alpha_n v_n + \sum_{m \in I} \tilde{X}_m \beta_m u_m + r, \quad (1)$$

où

1. Les X_n et \tilde{X}_m sont des variables aléatoires binaires, identiquement distribuées :

$$X_n \sim \mathcal{B}(p); \quad \tilde{X}_m \sim \mathcal{B}(\tilde{p}).$$

Elles génèrent des ensembles aléatoires $\Delta = \{n \in I, X_n = 1\}$ et $\Lambda = \{m \in I, \tilde{X}_m = 1\}$, appelés **cartes de signifiante**.

2. Les α_n (resp. β_m), appelés **coefficients de synthèse**, sont des variables aléatoires normales centrées indépendantes

$$\alpha_n \sim X_n \mathcal{N}(0, \sigma_n^2) + (1 - X_n) \delta_0 \quad (2)$$

$$\beta_m \sim \tilde{X}_m \mathcal{N}(0, \tilde{\sigma}_m^2) + (1 - \tilde{X}_m) \delta_0 \quad (3)$$

3. Les variances σ_n^2 et $\tilde{\sigma}_m^2$ sont fixées, et peuvent dépendre de l'indice n (n étant généralement un indice temps-fréquence, les variances dépendent alors de la partie fréquentielle de n). Pour simplifier, on se limite dans cette contribution au cas constant $\sigma_n^2 = \sigma^2$, et $\tilde{\sigma}_m^2 = \tilde{\sigma}^2$ (les signaux à temps continu nécessitent de considérer des variances dépendant de la fréquence).
4. r est une couche résiduelle, modélisée comme un bruit blanc Gaussien de variance s^2 .

Un tel modèle permet de reproduire avec succès le comportement des densités de probabilités des coefficients MDCT obtenus après décomposition du signal dans la base [5].

Le problème posé est le suivant. A partir d'une réalisation d'un tel signal, on cherche à identifier la réalisation correspondante des cartes Λ et Δ , et les valeurs des paramètres. Dans un second temps, les coefficients α et β sont identifiés par régression.

Ce modèle peut être rendu plus riche et complexe de plusieurs façons. Nous considérons notamment le cas d'un modèle « structuré », dans lequel la carte de signifiante Λ est modélisée via un modèle de Bernoulli hiérarchique, exploitant la nature « temps-fréquence » des bases considérées. L'hypothèse 1. plus haut est alors remplacée par

- 1'. Les \tilde{X}_m sont des variables aléatoires binaires iid, et engendrent la carte de signifiante Δ . Les variables aléatoires $X_{n=k,\nu}$ sont distribuées suivant une loi de Bernoulli $\mathcal{B}(p_1)$, conditionnellement à des indicatrices temporelles T_k , elles aussi distribuées suivant une loi de Bernoulli $\mathcal{B}(p_2)$. En d'autres termes,

$$\tilde{X}_m \sim \mathcal{B}(\tilde{p})$$

$$X_{k,\nu} \sim T_k \mathcal{B}(p_2) + (1 - T_k) \delta_0 \text{ avec } T_k \sim \mathcal{B}(p_1) .$$

Cette modélisation est particulièrement adaptée aux structures transitoires des signaux audios. Lors de l'attaque d'une note, et encore plus particulièrement les attaques percussives, l'image temps-fréquence associée fait clairement apparaître des lignes verticales localisées en temps. Le modèle de Bernoulli-hiérarchique permet de tenir compte de cette particularité en sélectionnant en premier les lignes temporelles où se situe l'attaque.

Un modèle similaire peut être étendu pour les parties tonales, qui font apparaître des structures rectilignes en fréquence. Cependant, la forte variation des fréquences au cours d'un même signal audio, qui comporte en général plusieurs notes, fait que le modèle de Bernoulli simple est mieux adapté.

Remarque 1 Une autre extension de ce modèle, étudiée en détails dans [6] consiste à rendre les variances des coefficients dépendante de la fréquence. On montre alors que les estimateurs que nous décrivons ci-dessous peuvent être

modifiés, en introduisant une pondération des coefficients d'analyse, pour prendre en compte cette nouvelle situation, plus réaliste en pratique.

3 Algorithmes

3.1 Estimation des cartes de signifiante

3.1.1 Modèle de Bernoulli

Dans le cadre du modèle de Bernoulli (i.e. sans structure), l'algorithme proposé est basé sur l'étude des **coefficients d'analyse** $a_n = \langle x, v_n \rangle$, et $b_m = \langle x, u_m \rangle$, et les résultats suivants :

1. Conditionnellement aux cartes Λ et Δ , les coefficients d'analyse a_n suivent une loi normale centrée, de variance

$$w_n := \mathbb{E} \{ a_n^2 \} = X_n \sigma^2 + p_n(\Delta) \tilde{\sigma}^2 + s^2, \quad (4)$$

où $p_n(\Delta) = \sum_{\delta \in \Delta} |\langle v_n, u_\delta \rangle|^2$.

Ainsi, suivant que $X_n = 0$ ou 1 , le coefficient d'analyse a_n a un comportement différent.

2. Cette différence de comportement est caractérisée par la distribution du poids $p_n(\Delta)$. Celle-ci est centrée sur

$$\mathbb{E}_\Delta \{ p_n(\Delta) \} = p := \mathbb{P} \{ X_n = 1 \},$$

et d'autant plus concentrée autour de cette valeur que les deux bases sont différentes, cette dernière notion s'exprimant via l'égalité

$$\text{Var} \{ p_n(\Delta) \} = p(1-p) \tilde{\sigma}^4 \sum_m |\langle v_n, u_\delta \rangle|^4 .$$

Ainsi, plus les produits scalaires mixtes des vecteurs des deux bases sont petits, plus la distribution des poids est « piquée » sur sa valeur moyenne.

3. Des résultats similaires sont obtenus pour les coefficients b_m .

Remarque 2 Conditionnellement à la carte Δ , les coefficients a_n sont distribués suivant un mélange de deux types de gaussiennes, de variances respectives $p_n(\Delta) \tilde{\sigma}^2 + s^2$ et $\sigma^2 + p_n(\Delta) \tilde{\sigma}^2 + s^2$. En moyenne ces deux types de gaussiennes peuvent être approchés par deux gaussiennes de variances $p \tilde{\sigma}^2 + s^2$ et $\sigma^2 + p \tilde{\sigma}^2 + s^2$. Une troisième gaussienne due au bruit peut se distinguer dans le mélange.

Cette remarque est utilisée pour l'estimation des cartes à partir des coefficients d'analyse dans l'algorithme suivant.

Algorithme 1 (Estimation EM d'une carte)

1. Calcul des coefficients d'analyses a_n (res. b_n) pour l'estimation de la carte de signifiante transitoire (res. tonale).
2. Estimation des paramètres du modèle par un algorithme EM. Suivant la parcimonie désirée, on pourra estimer un mélange de deux ou trois gaussiennes.

3. Estimation de la carte de signifiante via la classification des coefficients par Maximum A Posteriori (MAP).

Cet algorithme peut aussi être utilisé comme l'initialisation d'un algorithme type CEM pour l'estimation simultanée des cartes de signifiante.

Algorithme 2 (Estimation CEM des cartes) *On itère les trois étapes suivantes, jusqu'à convergence.*

1. Estimation des paramètres du modèle : σ , $\tilde{\sigma}$, p et \tilde{p} .
2. Calcul des poids $p_n(\Delta)$ et $\tilde{p}_n(\Lambda)$.
3. Réestimation des cartes de signifiante Δ et Λ par classification selon le MAP de chacun des coefficients d'après l'équation (4).

Ici, chaque coefficient est classé selon le MAP de sa propre distribution, au contraire de l'algorithme précédent où la distribution de l'ensemble des coefficients était approximée par un mélange de quelques gaussiennes.

3.1.2 Modèle de Bernoulli-hiérarchique

Dans le cadre du modèle de Bernoulli hiérarchique, l'estimation de la carte de signifiante Λ doit être effectuée différemment. On introduit les coefficients

$$c_k = \sum_{\nu=1}^{|\Lambda_f|} a_{k,\nu}^2 \quad (5)$$

Suivant que l'indice (k, ν) appartient à la carte temporelle $\Lambda_t = \{k : T_k = 1\}$ ou non, le coefficient d'analyse $a_{k,\nu}$ est distribué suivant l'une ou l'autre des deux types de gaussiennes (voir remarque 2). La somme des carrés de variables aléatoire normale étant distribuée selon une loi du χ^2 , on peut exploiter cette remarque pour estimer dans un premier temps la carte de signifiante temporelle, via un test statistique. Ceci conduit à l'algorithme suivant.

Algorithme 3 (Estimation structurée d'une carte)

1. Estimation de la carte temporelle par un test d'adéquation des c_k à une loi du χ^2
2. Estimation de la carte complète par classification des coefficients précédemment sélectionnés, par l'algorithme 1.

3.2 Estimation des coefficients

Une fois les cartes de signifiante estimées, les coefficients de synthèse peuvent être estimés à leur tour, par simple régression.

Soit $x \in \mathcal{H}$, on note $\hat{\Lambda}$ et $\hat{\Delta}$ les estimées des cartes Λ et Δ et $\mathcal{H}_{\hat{\mathcal{D}}}$ le sous-espace de \mathcal{H} engendré par le dictionnaire $\hat{\mathcal{D}} = \{u_\delta, \delta \in \hat{\Delta}\} \cup \{v_\lambda, \lambda \in \hat{\Lambda}\}$.

L'estimée \hat{x} du signal x par régression L^2 est sa projection orthogonale sur $\mathcal{H}_{\hat{\mathcal{D}}}$. Les estimations α_λ et β_δ des coefficients sont données par inversion de la matrice de Gram du dictionnaire $\hat{\mathcal{D}}$, que l'on suppose inversible. Des reconstructions partielles n'utilisant que les coefficients $\hat{\alpha}$ (resp. $\hat{\beta}$) forment les deux **couches** de la décomposition.

Dans le cas considéré ici, les deux bases étant deux bases MDCT de différentes résolutions temps-fréquence, on parlera de **couche tonale** (bonne résolution fréquentielle) et de **couche transitoire** (bonne résolution temporelle).

Notons qu'il est possible de renforcer la parcimonie de la décomposition du signal en utilisant une régression L^1 type « basis-pursuit denoising » [1] (qui est l'estimateur du lasso [2]).

4 Résultats

Les diverses variantes de l'algorithme proposé ont été mises en oeuvre sous MATLAB. L'application choisie pour tester les algorithmes sur des signaux audiophoniques est le débruitage, pour lequel la décomposition en couches « tonale + transitoire + bruit » est particulièrement adaptée.

Les résultats obtenus sont de qualité comparable à l'état de l'art. Les temps de calcul sont tout à fait encourageants (de l'ordre de 1 à 3 minutes pour traiter 1 sec de son, dans une implémentation MATLAB, sur un PC Linux Pentium 4, 1.2 GhZ). Ils peuvent être écoutés sur [7].

Le modèle de Bernoulli-hiérarchique est comparé au modèle de Bernoulli simple sur le signal de glockenspiel bruité afin d'obtenir un rapport signal à bruit (SNR) de 6 dB. Ce signal se prête particulièrement bien à la décomposition en couches, en raison de l'attaque percussive nette présente au début de chaque note jouée. Le signal ainsi que les différentes couches obtenues après décomposition par chacun des algorithmes sont représentés en figure 1. Le modèle étant différent pour la partie transitoire, c'est sur cette partie que la différence ressort. Le modèle de Bernoulli-hiérarchique permet d'obtenir des attaques bien marquées en temps, avec très peu d'artefacts entre deux notes. Le modèle de Bernoulli récupère de l'information basse-fréquence, et laisse entendre un bruit tout le long du signal transitoire. Ce bruit se retrouve lors de la reconstruction, qui donne un débruitage moins convaincant. En terme de SNR, les résultats sont équivalents : 17.1 dB pour le débruitage par Bernoulli simple contre 17.5 dB pour le débruitage par Bernoulli-hiérarchique.

Les résultats obtenus en débruitage par l'algorithme 3 ont été comparés à ceux donnés par un algorithme MCMC, sur le signal de piano utilisé dans [4]. Le signal original est bruité avec un bruit blanc gaussien, afin d'obtenir un SNR d'entrée de 10 dB. Après débruitage par le modèle de Bernoulli-hiérarchique, le SNR obtenu est de 18 dB. Si le SNR paraît bien inférieur aux 20 dB obtenus par certains algorithmes MCMC de [4], ils sont tout à fait comparables en terme de qualités d'écoute : notre modèle présente moins d'artefact musicaux que la plupart des sons débruités par MCMC.

L'algorithme a aussi été testé sur deux extraits de signaux issus des archives radiophoniques françaises, gracieusement fournis par l'INA¹.

¹Institut National de l'Audiovisuel. Merci à Vincent Fromont pour les extraits audio. Les résultats et les fichiers son se trouvent sur le site [7].

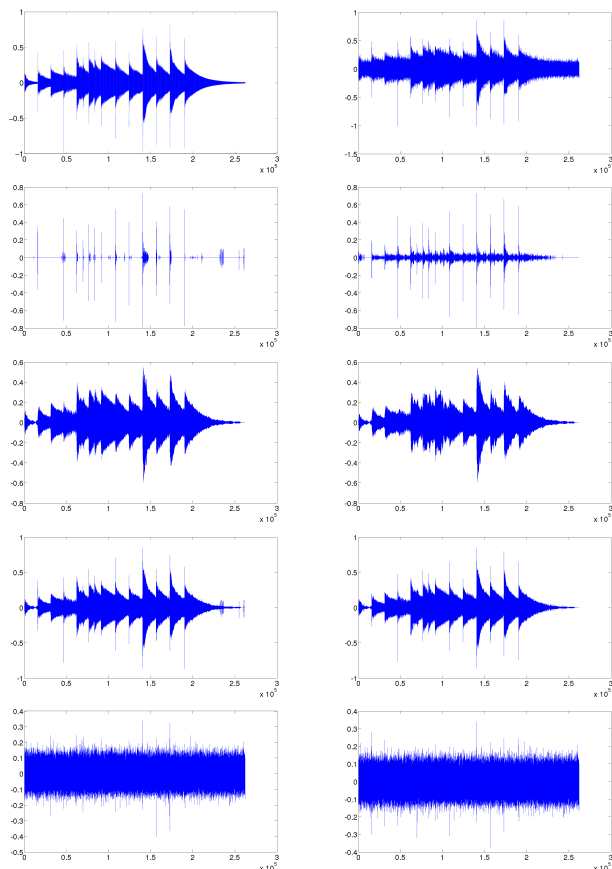


FIG. 1 – Débruitage du signal de glockenspiel. Comparaison entre le modèle de Bernoulli-hiérarchique et le modèle de Bernoulli simple. En haut : signal original et signal bruité. De haut en bas, à gauche modèle de Bernoulli-hiérarchique, à droite modèle de Bernoulli simple : couche transitoire, couche tonale, signal débruité (couche tonale + couche transitoire), résidu.

Le premier extrait présente un bruit de fond à basse fréquence et des clics. Le deuxième extrait présente un bruit de fond ainsi que de nombreux clics et craquements. On appelle ici « bruit de fond » le bruit qui n'est pas du aux clics. C'est le bruit qui s'apparente au bruit blanc gaussien du modèle.

De manière général, les clics restent présents lors de la reconstruction *tonal + transitoire* à la sortie de l'algorithme. Dans l'extrait 1, le bruit basse fréquence donne de gros coefficients lors de la transformation MDCT. Malgré la pondération (voir remarque 1), il reste de nombreux artefacts dus à ces coefficients MDCT importants. Le « bruit de fond » est cependant atténué. Pour l'extrait numéro 2, le meilleur résultat semble être celui obtenu lorsque l'algorithme est itéré une seconde fois sur le résidu : la partie tonale récupère en effet de l'information en haute fréquence. En revanche, la partie transitoire ajoute des clics lors de la deuxième itération, plutôt que d'améliorer le résultat. En ajoutant la partie tonale de la deuxième itération, à la partie transitoire de la 1ère, on obtient un résultat correct, mais présentant un bruit « musical » à la place des clics (les clics recouvrant ce bruit). Sur cet extrait, le « bruit de fond » est en majorité retiré.

Comme l'on peut s'y attendre, les clics sont capturés

par la partie transitoire du signal. On pourrait alors imaginer une discrimination entre les différents transitoires pour effectuer un décliquage.

5 Conclusion

Nous avons vu une approche originale, fondée sur une modélisation aléatoire, d'analyse parcimonieuse des signaux audio. Elle permet de rendre compte des différentes composantes présentes dans un signal audio et a été validée par une application au débruitage.

Cette approche, motivée par les observations faites sur des signaux réels, donne un modèle simple dont les résultats sont encourageants.

Toutefois, ce modèle est réservé à l'analyse des signaux audio, et ne permet pas de synthétiser des sons satisfaisant à l'écoute. Il demande de plus plusieurs améliorations. En particulier, une dépendance entre les couches tonales et transitoires (une attaque de note précède sa tenue) ainsi que des structures entre les coefficients d'une même couche devraient compléter le modèle. Ces deux points permettraient de le rendre plus réaliste et devraient améliorer sensiblement les estimations.

Le problème majeur de ce type d'approche, est qu'on ne dispose d'aucune définition d'un transitoire ou d'un tonal. Elle est ici implicite, un transitoire et un tonal étant définis par le modèle de carte et, surtout, par les bases choisies pour les représenter.

Références

- [1] S. S. Chen, D. L. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [2] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society*, vol. 58, pp. 267–288, 1996.
- [3] B. Efrin, T. Hastie, T. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–451, 2004.
- [4] C. Févotte, L. Daudet, S. J. Godsill, and B. Torrèsani, "Sparse regression with structured priors : Application to audio denoising," in *IEEE International Conference on Acoustics, Speech, and Audio Signal*, Toulouse, France, May 2006.
- [5] M. Kowalski and B. Torrèsani, "A study of bernoulli and structured random audio models," in *Proceedings of the conference on Signal Processing with Adaptive and Sparse Structured Representations (SPARS'05)*, R. Gribonval, Ed., Rennes, France, November 2005, pp. 59–62.
- [6] —, "Random models for audio signals expansion on hybrid mdct dictionaries," 2007, *Preprint*.
- [7] [Online]. Available : <http://www.cmi.univ-mrs.fr/~kowalski/GRETSI07.html>