Parcimonie Sociale et application à l'audio inpainting

Kai SIEDENBURG[1], Monika DÖRFLER[2], Matthieu KOWALSKI[3]

[1]Schulich School of Music – McGill University Montreal, Canada
550 Sherbrooke Street West Montreal, Quebec, Canada H4A 2N3

[2]NuHAG – Faculty of Mathematics, Univ. of Vienna
Alserbachstraße 23, A-1090 Wien, Austria

[3]Laboratoire des Signaux et Systèmes – Supelec–CNRS–Univ Paris-Sud
3, rue Joliot-Curie, 91192 Gif-sur-Yvette, France
kai.siedenburg@mail.mcgill.ca, monika.doerfler@univie.ac.at,
matthieu.kowalski@lss.supelec.fr

Résumé – On considère le problème d'audio ≪ inpainting ≫ en utilisant des algorithmes de seuillage itératif basés sur le principe de ≪ parcimonie sociale ≫ (ou ≪ social sparsity ≫). Ces seuils permettent de prendre en compte le voisinage de coefficients dans un dictionnaire temps-fréquence. On propose une formulation convexe non contrainte pour le problème d'audio inpainting. On utilise ensuite des opérateurs de seuillage structurés dans les algorithmes itératifs, tels que le *group-Lasso fenêtré*, ou le seuillage de *Wiener empirique persistant*. Ces nouveaux opérateurs permettent d'améliorer la qualité de reconstruction par rapport à un simple opérateur de seuillage doux. L'algorithme obtenu est rapide, simple à mettre en oeuvre (car appartenant à la famille des algorithmes de seuillage itératifs) et permet d'améliorer la qualité de reconstruction.

Abstract – The audio inpating problem is under consideration, using iterative thresholding algorithms and the "social sparsity" principle. Indeed, these recently introduced thresholding/shrinkage operators allows one to take into account the neighborhood of the synthesis coefficient inside time-frequency dictionnary.

We first start from a new unconstrained convex formulation for the audio inpainting problem. The chosen structured thresholding operator are the so called windowed group-Lasso and the persistent empirical Wiener. The use of such operator allows one to improve the quality of the reconstruction, compared to a simple soft-thresholding operator.

The resulted algorithm is fast, simple to implement, and outperform the state of the art constrained matching pursuit algorithm in term of SNR for the reoncstruction, and of speed of convergence.

1 Introduction

L'écrêtage d'un signal, ou phénomène de saturation, est un problème de restauration des signaux audio lorsque que le gain maximum d'un système est saturé, et résulte en un seuillage des échantillons de trop forte amplitude. Cela se traduit par le phénomène connu de saturation qui, lorsqu'il n'est pas intentionnellement recherché, donne une distortion de signal peu plaisante à l'écoute.

On s'intéresse ici à l'estimation des échantillons manquants ou corrompus d'un signal audio, tâche dans laquelle l'estimation des échantillons écrêtés entre parfaitement. Si les approches classiques de ce problème utilisent des modèles autorégressif [10] ou des estimations bayesiennes [8], les auteurs de [1] ont proposé un cadre général d'*audio inpainting* (en référence au problème classique d'inpainting rencontré en image [5]) reposant sur la parcimonie. En pratique, la reconstruction des échantillons manquant ou corrompu se fait à l'aide des échantillons restés intacte. Le modèle proposé par [1] s'écrit comme

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\alpha\|_0 \quad \text{s.t.} \quad \|\mathbf{y}^r - \mathbf{M}^r \mathbf{\Phi} \alpha\|_2^2 \leq \epsilon \,, \qquad (1)$$

avec $\mathbf{\Phi}$ l'opérateur de synthèse correspondant au dictionnaire choisi (en général de type ≪ Gabor ≫), $\mathbf{y}^r$ les échantillons du signal observé $\mathbf{y}$, qui correspondent à la partie fiable du signal, c'est-à-dire les échantillons non clippés ou bien ceux disponibles dans le cas de paquets de données perdus. De la même façon, $\mathbf{M}^r$ est une matrice formée à partir des lignes de la matrice identité choisies telles que les entrées correspondent aux échantillons fiables. L'estimation du signal restauré se faisant classiquement par la synthèse $\hat{\mathbf{s}} = \mathbf{\Phi}\hat{\boldsymbol{\alpha}}$.

Pour le problème plus spécifique du *desécrêtage*, qu'on appellera aussi par l'anglicisme *déclippage* ou *déclipping*, les auteurs proposent d'ajouter une contrainte complémentaire au problème (1). En effet, les échantillons reconstruits sont supposés prendre une valeur plus grande que celle donnée par le seuil d'écrêtage, en valeur absolue. Par analogie avec la définition de $\mathbf{M}^r$, on note $\mathbf{M}^c$ la matrice correspondante aux échantillons clippés. On note aussi $\boldsymbol{\theta}^{clip}$ le vecteur des échantillons clippés, qui ne

prennent que la valeur $\pm\theta^{clip}$, dépendant du signe de la vraie valeur de $\mathbf{y}$ en k. Le problème devient alors

$$\hat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_0 \quad (2)$$
$$\text{s.t. } \|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 \leq \epsilon \quad \text{and} \quad |\mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}| \geq |\boldsymbol{\theta}^{clip}|$$

Dans [1], le problème de minimization (1) est résolu en utilisant le matching pursuit orthogonal (OMP), ou, afin d'utiliser de l'information supplémentaire dans le cas du déclippage, un OMP contraint. Cependant, l'approche ne prend pas en compte de corrélation temporelle sur les coefficients.

On expose dans la section 2 une formulation convexe non contrainte pour résoudre le problème de déclippage, ainsi que de nouveaux opérateurs de seuillage prenant en compte une certaine corrélation temporelle des coefficients de synthèse d'un dictionnaire de Gabor. La section 3 illustre les performances de l'algorithme proposé, comparées à l'état de l'art donné dans [1].

2 Cadre proposé

On commence par exposer la formulation convexe non contrainte pour les problème de déclippage, avant d'introduire les opérateurs de seuillage utilisés en pratique.

2.1 Une formulation convexe non contrainte

La relaxation convexe sous contrainte linéaire la plus naturelle du problème (2) donne :

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_1 \quad (3)$$
$$\text{s.t.} \quad |\mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}| \geq |\boldsymbol{\theta}^{clip}|$$

Cependant, un tel problème est difficile à résoudre. En effet, l'opérateur de proximitié de la norme ℓ_1 sous contrainte linéaire ne peut pas se calculer sous forme analytique. On est alors obligé de faire appelle à des algorithmes itératifs pour le calculer, ce qui peut donner des algorithmes très lourd en temps de calcul. On pourra se référer à [4] où un algorithme de type Douglas-Rachford est utilisé à l'intérieur de l'algorithme explicite-implicite.

Grâce à la fonction « charnière au carré », on propose une formulation convexe non contrainte, qui permet de forcer les échantillons restaurés à prendre une valeur supérieure au seuil d'écrêtage. Le problème résultant donne

$$\underset{\alpha}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y}^r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \frac{\mu}{2}[\boldsymbol{\theta}^{clip} - \mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}]_+^2 + \lambda\|\boldsymbol{\alpha}\|_1 \quad (4)$$

avec

$$[\boldsymbol{\theta}^{clip} - \mathbf{x}]_+^2 = \sum_{k:\theta_k^{clip}>0} (\theta_k^{clip} - x_k)_+^2 + \sum_{k:\theta_k^{clip}<0} (-\theta_k^{clip} + x_k)_+^2 .$$

Ainsi, la pénalité $\boldsymbol{\alpha} \mapsto \frac{\mu}{2}[\boldsymbol{\theta}_{clip} - \mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}]_+^2$ a bien pour but de respecter la contrainte introduite dans [1], qui force les échantillons restaurés à prendre une valeur supérieure au seuil de clippage.

Puisque la fonction $[.]_+^2$, et donc la fonction

$$\boldsymbol{\alpha} \mapsto \frac{1}{2}\|\mathbf{y}_r - \mathbf{M}^r\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \frac{\mu}{2}[\boldsymbol{\theta}^{clip} - \mathbf{M}^c\boldsymbol{\Phi}\boldsymbol{\alpha}]_+^2 ,$$

est différentiable, de dérivée Lipschitz continue, n'importe quel algorithme de la famille ISTA (Iterative Shrinkage/Thresholding Algorithm ou algorithme explicite-implicite) peut être utilisé afin de résoudre le problème d'optimisation correspondant [3].

2.2 Opérateurs de seuillage structurés

On explore de plus les bénéfices des operateurs de « social sparsity » pour ce problème. Étant donnée la structure du problème de déclippage, il semble assez naturel de prendre en compte une corrélation sur les coefficients de synthèse : les composantes du signal qui se déroule au cours du temps introduisent une persistance temporelle sur les coefficients. D'un autre coté, les coefficients de haute énergie « isolés » peuvent être attribués à une corruption des données, et ne devrait donc ne pas être pris en compte dans le processus de restauration.

Par extension de l'opérateur de seuillage doux usuel correspondant à l'opérateur de proximité de la pénalité ℓ^1, qui apparait dans (4), il devient alors possible d'exploiter des propriétés de persistence du signal par un système de voisinage temps-fréquence [11]. Par conséquent, on présente les résultats correspondant au Lasso, mais aussi au group-Lasso fenêtré (windowed group Lasso – WGL), Wiener Empirique (EW) et son compagnon social, le persistent EW (PEW). Ces deux derniers opérateurs permettent une exponentiation au carré de l'amplitude des coefficients (voir [12]). Plus formellement, les opérateurs de seuillages ont les expressions suivantes :

– Lasso :

$$\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^L(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda}{|\alpha_{tf}|}\right)^+$$

– EW :

$$\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^{EW}(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda^2}{|\alpha_{tf}|^2}\right)^+$$

– WGL :

$$\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^{WGL}(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda}{\sqrt{\sum_{t'\in\mathcal{N}(t)}|\alpha_{t'f}|^2}}\right)^+$$

où $\mathcal{N}(t)$ désigne le voisinage de l'indice t.

– PEW :

$$\tilde{\alpha}_{tf} = \mathbb{S}_\lambda^{PEW}(\alpha_{tf}) = \alpha_{tf}\left(1 - \frac{\lambda^2}{\sum_{t'\in\mathcal{N}(t)}|\alpha_{t'f}|^2}\right)^+$$

où $\mathcal{N}(t)$ désigne le voisinage de l'indice t.

L'opérateur EW [7] est aussi connu comme le non-négative garrote [2]. Cet opérateur a été introduit pour limiter l'effet de « shrinkage » des coefficients induit par le lasso. Ainsi, les coefficients d'amplitude élevée sont laissés quasiment inchangés, et les résultats obtenus avec cet opérateur sont à rapprocher de ceux obtenus après une étape de débiaisage [6].

L'opérateur WGL permet de prendre une décision « coefficients par coefficients » en prenant en compte son voisinage. En effet, l'énergie est calculée sur tout un voisinage, puis comparée à un seuil. Ainsi, un coefficient d'énergie relativement élevé mais isolé pourra être mis à zéro, tandis qu'un coefficient de faible énergie au milieu de coefficients « fort » pourra être conservé. Le but sera par la suite de favoriser des structures temporelles dans les dictionnaires temps-fréquence.

Enfin, l'opérateur PEW a été construit sur l'opérateur WGL pour les mêmes raisons que le EW : limiter le seuillage des forts coefficients.

Équipé de ces opérateurs de seuillage généralisé, on utilise une version sur-relaxé de l'algorithme de seuillage itératif. Cette version sur-relaxée, bien que moins connu que la version accélérée FISTA, est en pratique aussi rapide, et apparait même plus robuste. Il est étudié plus particulièrement dans [3].

Algorithm 1: ISTA relaxé

Initialization : $\boldsymbol{\alpha}^{(0)} \in \mathbb{C}^N$, $\mathbf{z}^{(0)} = \boldsymbol{\alpha}^{(0)} k = 1$,
$\gamma = \|\boldsymbol{\Phi}\boldsymbol{\Phi}^*\|$ **repeat**
  $\mathbf{g}1 = -\boldsymbol{\Phi}^*\mathbf{M}^{c^T}(\mathbf{y}^r - \mathbf{M}^c\boldsymbol{\Phi}\mathbf{z}^{(k-1)})$;
  $\mathbf{g}2 = -\mu\boldsymbol{\Phi}^*\mathbf{M}^T[\boldsymbol{\theta}^{clip} - \mathbf{M}\boldsymbol{\Phi}\boldsymbol{\alpha}]_+$;
  $\boldsymbol{\alpha}^{(k)} = \mathbb{S}_{\lambda/\gamma}\left(\mathbf{z}^{(k-1)} - \frac{1}{\gamma}(\mathbf{g}1 + \mathbf{g}2)\right)$;
  $\mathbf{z}^{(k)} = \boldsymbol{\alpha}^{(k)} + \gamma(\boldsymbol{\alpha}^{(k)} - \boldsymbol{\alpha}^{(k-1)})$;
  $k = k + 1$;
until *convergence*;

où γ est le coefficient de relaxation. Pour le lasso, la convergence des itérés $\alpha^{(k)}$ vers un minimiseur de (4) est montré pour $-1 < \gamma < 1/2$, et la convergence de la fonction vers son minimum pour $1/2 \leq \gamma < 1$.

3 Résultats numériques

Pour les expériences, on utilise les données disponibles sur la page `http://small-project.eu/` et la toolbox correspondante pour l'audio inpainting. Plus particulièrement, les algorithmes sont comparés sur la base de donnée des signaux de musique échantillonnés à 16 kHz, après une normalisation des signaux afin que la valeur maximale des échantillons n'excède pas 1. Les expériences ont été menées avec deux valeurs d'écrêtage : 0.2 et 0.6.

Pour l'algorithme de seuillage itératif 1, on a choisi un dictionnaire de Gabor avec une fenêtre de longueur 512 échantillons (soit une durée de 32 ms), et un recouvrement de 50% entre deux fenêtres. Les expériences ne font pas de débruitage en plus de la restauration des échantillons écrêtés, ainsi, la valeur du paramètre λ est choisie proche de 0. On utilise ici la stratégie classique de redémarrage à chaud avec des valeurs de λ décroissantes [9]. Le voisinage des opérateurs WGL et PEW est choisi de manière à prendre en compte les quatre coefficients précédents et les quatre coefficients suivant le coefficient considéré sur l'axe temporel. Le voisinage considéré est donc exclusivement temporel et comporte neuf coefficients au total.

Une première série d'expériences montre que la valeur $\mu = 1$ suffit à respecter la contrainte de reconstruction des échantillons écrêtés. Cette valeur vient confirmer une intuition naturelle : la square hinge peut être vue comme une attache aux données ℓ_2, mais ne prenant pas en compte les erreurs de « sur-estimation ». Si l'on remplace le terme de squared hinge par une simple norme ℓ_2 (avec $\mu = 1$), cela revient a reconstruire le signal observé sans faire d'inpainting.

On résume dans le tableau 3 les performances obtenues par les quatre opérateurs de seuillage, ainsi que le résultat de référence donné dans [1] avec le matching pursuit orthogonal contraint (et un dictionnaire de Gabor).

TABLE 1 – Résultats (en dB) des différents algorithmes de déclippage

	Clipped	C-OMP-Gabor	L	WG-L	EW	PEW
0.2	6.4	10.2	7.0	6.9	12.7	14.6
0.6	15	22.9	17.4	17.5	22.7	24.3

On remarque que le Lasso et le group-Lasso fenêtré obtiennent de piètre performance pour ce problème. Vu les très bons résultats obtenus par les opérateurs EW et PEW, construit pour préserver un maximum d'énergie, cela s'explique directement par le phénomène de seuillage, et confirme les observations faites sur d'autres problèmes inverses [6]. L'apport de l'information de voisinage avec l'opérateur PEW permet d'améliorer la qualité de $2dB$.

On montre sur la figure 1 l'évolution du SNR en fonction de λ pour les différents opérateurs de seuillage, qui confirme le choix $\lambda \to 0$.

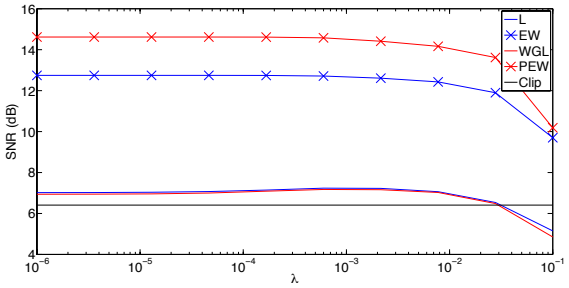

FIGURE 1 – Évolution du SNR en fonction de λ

Enfin, on montre sur la figure 2 le résultat de la restauration obtenue avec l'opérateur PEW sur un signal de musique particulier.

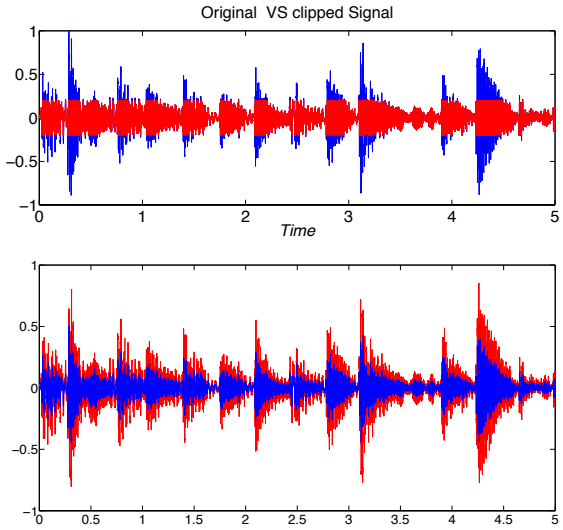

FIGURE 2 – Haut : signal original (bleu) et signal clippé (rouge). Bas : signal original (bleu) et signal restauré (rouge).

4 Conclusion

Les principales contributions présentées dans cet article sont la formulation convexe non contrainte du problème de restauration des échantillons écrêtés, et les operateurs de seuillage structurés.

La formulation non contrainte permet d'utiliser les algorithmes d'optimisation classiques de la littérature. Il reste à régler de manière optimale le nombre d'itération et la décroissance du paramètre λ afin d'obtenir un algorithme vraiment rapide en pratique.

Les opérateurs de seuillage structurés, qui font référence à la ¡¡social sparsity¿¿ ont d'abord été construit de manière heuristique. Ils permettent très simplement de prendre compte l'information de voisinage d'un coefficient afin de favoriser les structures voulues, de manière plus souple que des groupes de coefficients.

Par la suite, une étude expérimentale plus poussé du problème d'inpainting sera intéressante à mener, sans se limiter au cas particulier des signaux écrêtés.

Références

[1] Amir Adler, Valentin Emiya, Maria Jafari, Michael Elad, Rémi Gribonval, and Mark D. Plumbley. Audio inpainting. *IEEE Transactions on Audio, Speech and Language Processing*, 20(3) :922–932, 2012.

[2] A. Antoniadis. Wavelet methods in statistics : Some recent developments and their applications. *Statistics Surveys*, 1 :16–55, 2007.

[3] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. New York, Springer edition, 2011.

[4] Caroline Chaux, Jean-Christophe Pesquet, and Nelly Pustelnik. Nested iterative algorithms for convex constrained image recovery problems. *SIAM Journal on Imaging Sciences*, 2(2) :730–762, 2009.

[5] Michael Elad, Jean-Luc Starck, David L. Donoho, and P Querre. Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Journal on Applied and Computational Harmonic Analysis*, 19 :340–358, November 2005.

[6] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction : application to compressed sensing and other inverse problems. *IEEE Journal on Selected Topics in Signal Processing*, 1 :586–598, 2007.

[7] S. Ghael, A. Sayeed, and R. Baraniuk. Improved wavelet denoising via empirical wiener filtering. In *Proceedings of SPIE*, pages 389–399, 1997.

[8] Simon J Godsill and Peter JW Rayner. A bayesian approach to the restoration of degraded audio signals. *Speech and Audio Processing, IEEE Transactions on*, 3(4) :267–278, 1995.

[9] A. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1-minimization : Methodology and convergence. *SIAM Journal on Optimisation, 19(3),* ., 19(3) :1107–1130, 2008.

[10] AJEM Janssen, R Veldhuis, and L Vries. Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(2) :317–330, 1986.

[11] M. Kowalski, K. Siedenburg, and M. Dörfler. Social sparsity ! neighborhood systems enrich structured shrinkage operators. *IEEE transactions on signal processing*, 61(10) :2498–2511, 2013.

[12] K. Siedenburg and M. Dörfler. Persistent time-frequency shrinkage for audio denoising. *Journal of the Audio Engineering Society (AES)*, 61(1), 2013.