La technique du Matching Pursuit

Présentation de M2RI - Module ARD
Université de Rennes 1 - INSA de Rennes

Cédric Fleury

12 Novembre 2007

1 Introduction

Le Matching Pursuit [7] [1] est une technique de traitement du signal, dite de décomposition parcimonieuse, qui permet de représenter un signal par une combinaison linéaire de formes d'ondes appartenant à un ensemble, appelé dictionnaire. Ces formes d'ondes sont sélectionnées de telle façon qu'elles représentent le mieux possible la structure du signal. La combinaison linéaire des formes d'ondes est qualifiée de parcimonieuse car elle n'a que peu de coefficients significatifs. Cela pourra être utilisé pour extraire facilement de l'information du signal.

Par exemple, en utilisant un grand dictionnaire redondant de différentes formes d'ondes bien localisées soit en temps soit en fréquence, appelé dictionnaire « temps - fréquence », on peut ainsi avoir une bonne décomposition du signal quelque soit ses caractéristiques. Cette décomposition flexible est particulièrement intéressante pour représenter les signaux dont la localisation de leurs composantes varie largement en temps et en fréquence.

La partie 2 présentera la notion de dictionnaire en donnant l'exemple du dictionnaire « temps - fréquence ». L'algorithme du Matching Pursuit, ainsi que son utilisation avec un dictionnaire « temps - fréquence » seront présentés en partie 3. Enfin, nous verrons dans la partie 4, les différentes applications du Matching Pursuit dans le traitement du signal.

Notations Comme Mallat et Zhang [7], on notera $\mathbf{L}^2(\mathbb{R})$ l'espace d'Hilbert des fonctions à valeurs complexes tel que, pour $f \in \mathbf{L}^2(\mathbb{R})$

$$\|f\|^2 = \int_{-\infty}^{+\infty} |f(t)|^2 dt < +\infty.$$

Un espace d'Hilbert est une généralisation en dimension quelconque d'un espace Euclidien.
Le produit scalaire de $(f,g) \in \mathbf{L}^2(\mathbb{R})^2$ est défini par

$$\langle f, g\rangle = \int_{-\infty}^{+\infty} f(t)\bar{g}(t)dt,$$

où $\bar{g}(t)$ est le conjugué complexe de $g(t)$.
La transformée de Fourier de $f(t) \in \mathbf{L}^2(\mathbb{R})$ s'écrit $\hat{f}(\omega)$ et est défini par

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt.$$

Enfin, dans l'espace $\mathbf{L}^2(\mathbb{R})$, un signal est considéré comme une fonction $f(t) \in \mathbf{L}^2(\mathbb{R})$.

2 Décomposition en atomes « Temps - Fréquence »

Pour extraire de l'information d'un signal, il peut être intéressant de décomposer ce signal en une famille de fonctions bien localisées à la fois en temps et en fréquence. Ces fonctions, appelées atomes « temps - fréquence », sont regroupées dans un dictionnaire. Mallat et Zhang [7] proposent de générer un tel dictionnaire en modifiant l'échelle, en translatant et en modulant une fenêtre simple $g(t) \in \mathbf{L}^2(\mathbb{R})$. Soit l'échelle $s > 0$, la modulation en fréquence ξ et la translation u, on note $\gamma = (s, u, \xi) \in \Gamma = \mathbb{R}^+ \times \mathbb{R}^2$ et on définit un atome

« temps - fréquence » de la manière suivante :

$$g_\gamma(t) = \frac{1}{\sqrt{s}} g(\frac{t-u}{s}) e^{i\xi t}.$$

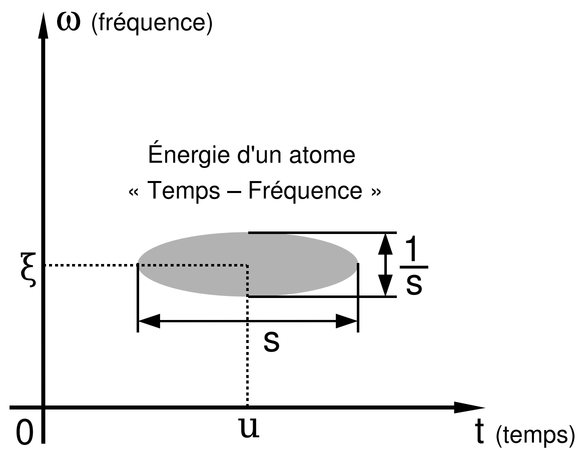

Fig. 1 – Représentation de l'énergie d'une atome « temps - fréquence » en fonction de l'échelle s, la modulation en fréquence ξ et la translation u.

Comme le montre la figure 1 :

- par rapport au temps, la fonction $g_\gamma(t)$ est centrée autour de u et son énergie est concentrée au voisinage de u avec une taille proportionnelle à s.
- par rapport à la fréquence, la transformée de Fourier $\hat{g}_\gamma(\omega)$ est centrée autour de ξ et son énergie concentrée au voisinage de ξ avec une taille proportionnelle à $\frac{1}{s}$.

Le dictionnaire obtenu est alors la famille de vecteurs $\mathcal{D} = (g_\gamma(t))_{\gamma\in\Gamma}$. On dit que le dictionnaire est complet uniquement si la combinaison linéaire des vecteurs de $\mathcal{D}$ est dense dans l'espace d'Hilbert, ici $\mathbf{L}^2(\mathbb{R})$. Le plus petit dictionnaire complet possible est une base de $\mathbf{L}^2(\mathbb{R})$. Cependant, dans la plupart des cas, le dictionnaire est largement redondant, ce qui permet d'avoir une plus grande liberté dans la décomposition des signaux.

Pour représenter efficacement un signal, c'est-à-dire une fonction $f(t)$, on doit sélectionner un sous-ensemble approprié d'atomes $(g_{\gamma_n}(t))_{n\in\mathbb{N}}$ avec $\gamma_n = (s_n, u_n, \xi_n)$ tel que

$$f(t) = \sum_{n=-\infty}^{+\infty} a_n(g_{\gamma_n}(t))$$

Les coefficients a_n dépendent de l'atome $g_{\gamma_n}(t)$ choisi. Ils portent des informations sur la structure du signal. Ce sont donc eux que l'on va chercher à calculer en plus de l'indice γ_n de l'atome choisi.

3 Matching Pursuit

3.1 Principes généraux

Le Matching Pursuit est un algorithme itératif qui décompose un signal grâce à un dictionnaire de vecteurs $\mathcal{D}$ appartenant à l'espace d'Hilbert $\mathbf{H}$ du signal, comme par exemple le dictionnaire « temps - fréquence » dans $\mathbf{L}^2(\mathbb{R})$ vu précédemment. A chaque itération, l'algorithme devra choisir l'atome $(g_\gamma)_{\gamma\in\Gamma}$ qui correspond le mieux au signal à reconnaître. Pour cela, il faudra choisir une fonction de corrélation qui permettra de mesurer la corrélation entre les différents atomes $(g_\gamma)_{\gamma\in\Gamma}$ et le signal. Cette fonction nous donnera donc le coefficient a_n de la décomposition parcimonieuse. Comme dans l'algorithme initial de Mallat et Zhang [7], nous utiliserons le produit scalaire comme fonction de corrélation. Cependant, il existe des versions plus récentes, comme celle de Gribonval et coll. [3], qui utilisent des fonctions de corrélation différentes afin d'améliorer l'algorithme.

On notera $R^n f$ le résidu de la fonction f après n itérations pour $n > 0$. $R^n f$ est ce qui n'a pas encore était traité dans la fonction f. On pose $R^0 f = f$.

On suppose que l'on est arrivé à la $n^{ième}$ itération et que l'on a calculé le résidu $R^n f$. On va alors choisir grâce à la fonction de corrélation (ici le produit scalaire) un atome $g_{\gamma_n} \in \mathcal{D}$ qui correspond le mieux à $R^n f$

$$|\langle R^n f, g_{\gamma_n}\rangle| = \sup_{\gamma\in\Gamma} |\langle R^n f, g_\gamma\rangle|$$

Le résidu $R^n f$ est de nouveau décomposé en

$$R^n f = \langle R^n f, g_{\gamma_n}\rangle\, g_{\gamma_n} + R^{n+1} f,$$

ce qui définit le résidu à l'ordre $n+1$. Etant donné que $R^{n+1} f$ est orthogonal à g_{γ_n}, l'énergie du résidu peut s'écrire

$$\|R^n f\|^2 = |\langle R^n f, g_{\gamma_n}\rangle|^2 + \|R^{n+1} f\|^2.$$

Après m itérations, le Matching Pursuit décompose le signal f en

$$f = \sum_{n=0}^{m} \langle R^n f, g_{\gamma_n}\rangle\, g_{\gamma_n} + R^m f.$$

De la même façon, on peut trouver une équation de conservation de l'énergie telle que

$$\|f\|^2 = \sum_{n=0}^{m} |\langle R^n f, g_{\gamma_n}\rangle|^2 + \|R^m f\|^2.$$

Malgré que cette décomposition ne soit pas linéaire, la conservation de l'énergie est respectée comme si c'était une décomposition linéaire orthogonale. Le signal f peut être caractérisé par une double séquence $(\langle R^n f, g_{\gamma_n}\rangle, \gamma_n)_{n\in\mathbb{N}}$, où γ_n est l'indice de l'atome sélectionné dans le dictionnaire et $\langle R^n f, g_{\gamma_n}\rangle$ est le produit scalaire correspondant.

Si on s'arrête à l'ordre m, on approxime alors f avec une erreur égale à $R^m f$. Si les vecteurs $(g_{\gamma_n})_{0\leqslant n<m}$ ne sont pas orthogonaux, cette approximation n'est pas celle qui approxime le mieux f. En effet, Mallat et Zhang [7] ont montré que, grâce à une technique de « back-projection » qui consiste à re-projeter le résidu $R^m f$ dans l'espace généré par les vecteurs $(g_{\gamma_n})_{0\leqslant n<m}$, cela permet de réduire l'énergie de l'erreur $\|R^m f\|^2$.

A l'inverse, si l'on continue à itérer, Mallat et Zhang [7] ont prouvé que l'erreur $\|R^m f\|$ converge exponentiellement vers zéro dans un espace de dimensions finies. On obtient donc

$$f = \sum_{n=0}^{+\infty} \langle R^n f, g_{\gamma_n}\rangle\, g_{\gamma_n}.$$

Dans la plupart des cas, on arrête l'itération lorsque l'erreur est devenue inférieure à une précision ϵ désirée. Le nombre d'itération nécessaire pour obtenir cette condition dépend du taux de décroissement de $\|R^m f\|$. Mais, étant donné que cette erreur $\|R^m f\|$ décroît exponentiellement jusqu'à zéro, on peut être sûr que l'algorithme converge vers une solution.

3.2 Matching Pursuit avec un dictionnaire « Temps - Fréquence »

Pour les dictionnaires d'atomes « temps - fréquence », le Matching Pursuit permet une décomposition qui s'adapte à la fois dans le temps et selon la fréquence du signal à traiter. Il décompose toute fonction $f \in \mathbf{L}^2(\mathbb{R})$ en une somme d'atomes « temps - fréquence » complexes qui sont corrélés à ces résidus.

Afin d'illustrer la décomposition d'un signal par Matching Pursuit dans le domaine « temps - fréquence », Mallat et Zhang [7] proposent une nouvelle distribution de l'énergie « temps - fréquence » obtenu en sommant les distributions de Wigner de chaque atome sélectionné par l'algorithme du Matching Pursuit. Cette nouvelle distribution est définie par

$$Ef(t,\omega) = \sum_{n=0}^{+\infty} \langle R^n f, g_{\gamma_n}\rangle\, Wg_{\gamma_n}(t,\omega),$$

où $Wg_{\gamma_n}(t,\omega)$ est la distribution de Wigner de g_{γ_n}. La distribution de Wigner d'une fonction $f(t)$ est

$$Wf(t,\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} f(t+\frac{\tau}{2})\bar{f}(t-\frac{\tau}{2})e^{-i\omega\tau}d\tau.$$

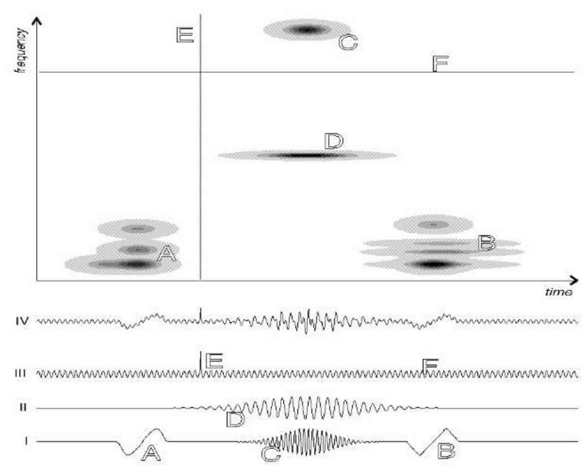

FIG. 2 – Distribution de l'énergie « temps - fréquence » du signal IV, qui est la somme des signaux I, II et III.

Cette nouvelle distribution permet de représenter l'énergie d'un signal dans le plan « temps - fréquence », comme le montre le figure 2. L'énergie des atomes du dictionnaire est représentée par des lignes horizontales pour les fonctions cosinus (F), des lignes verticales pour les Dirac (E) et des ellipses proportionnelles aux valeurs du temps et de la fréquence pour les fonctions de Gabor (C,D).

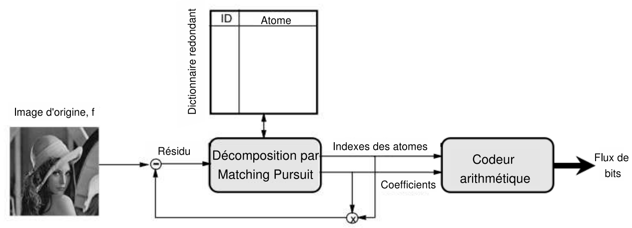

Fig. 3 – Compression d'un image grâce au Matching Pursuit

4 Applications

Le Matching Pursuit a de nombreuses applications dans le traitement du signal. Nous allons voir quels peuvent être les différents traitements que le Matching Pursuit permet d'appliquer au signal, ainsi que des exemples d'applications de ces traitements dans les différents domaines qui utilisent le traitement du signal, en particulier le traitement du son et celui des images. La décomposition en atomes « temps - fréquence » par Matching Pursuit comme nous l'avons vu précédemment est bien adaptée pour le traitement du son. Cependant, la technique du Matching Pursuit peut également être appliquée beaucoup d'autres domaines en changeant simplement le type de dictionnaire utilisé. Par exemple, pour traiter des images, on pourra utilisé un Matching Pursuit avec un dictionnaire de Gabor à 2 dimensions comme le proposent Mallat et Bergeaud [6].

4.1 Compression

Le Matching Pusrsuit peut permettre de compresser un signal en le codant par la double séquence $(\langle R^n f, g_{\gamma_n}\rangle, \gamma_n)_{n\in\mathbb{N}}$ où l'on garde que les n premiers atomes de la décomposition selon le taux de compression souhaité. Par exemple, Figueras i Ventura et coll. [4] proposent de coder une image par Matching Pursuit comme présenté sur la figure 3.

Les avantages que voient Figueras i Ventura et coll. à ce type de codage sont :

- une bonne performance de compression à bas taux (comparable, voire meilleure à celle du JPEG-2000),
- une très bonne robustesse de l'image recomposée après décodage au changement d'échelle, même si l'image recomposée est de taille supérieure à l'image d'origine,
- une possibilité de changer le taux de compression selon le besoin sans avoir à recoder toute l'image. En effet, il suffit de garder un nombre plus ou moins important des premiers atomes de la décomposition par Matching Pursuit pour changer le taux de compression. Moins on garde de coefficients, meilleur sera le taux de compression, mais la qualité de l'image sera bien sûr moins bonne.

4.2 Débruitage et extraction de structures cohérentes

Pour pouvoir retirer un bruit présent dans le signal, ce n'est pas l'information contenu dans le signal qui est importante, mais la cohérence de cette dernière par rapport au système d'interprétation du signal. Mallat et Zhang [7] ont étudié la notion de cohérence et décrit un algorithme qui isole les structures du signal qui sont cohérentes par rapport à un dictionnaire donné. Pour cela, à la chaque itération n de l'algorithme du Matching Pursuit, ils proposent de mesurer le taux de corrélation entre l'atome choisi dans le dictionnaire g_{γ_n} et le résidu $R^n f$

$$\tilde{\lambda}(R^n f) = \frac{|\langle R^n f, g_{\gamma_n}\rangle|}{\|R^n f\|},$$

et de le comparer à l'espérance du taux de corrélation $E(\tilde{\lambda}(R^n W))$ entre l'atome g_{γ_n} et le résidu du bruit $R^n W$. Ils définissent les structures cohérentes du signal comme les m premiers vecteurs $(g_{\gamma_n})_{0\leqslant n<m}$ qui ont un plus grand taux de corrélation avec $R^n f$ que avec $R^n W$. C'est-à-dire que f a m structures cohérentes si et seulement si

pour $0 \leqslant n < m$

$$\tilde{\lambda}(R^n f) > E(\tilde{\lambda}(R^n W)),$$

et

$$\tilde{\lambda}(R^m f) \leqslant E(\tilde{\lambda}(R^m W))$$

Mallat et Zhang [7] utilisent cette méthode pour extraire de la parole d'un signal contenant du bruit, afin de pouvoir reconstituer ensuite le signal avec la parole sans le bruit grâce au m premiers termes de la décomposition par Matching Pursuit.

4.3 Reconnaissance d'information

Grâce à ses propriétés, le Matching Pursuit peut également être utilisé pour reconnaître des informations dans le signal. En effet, si on utilise le Matching Pursuit avec un dictionnaire correspondant à différentes variantes des fragments d'informations à reconnaître, on pourra certainement retrouver quelles sont les informations présentes dans le signal en fonction des coefficients de sa décomposition parcimonieuse. Gribonval et Bacry [2] utilisent ainsi une version améliorée, un Matching Pursuit pour retrouver des notes dans un morceau de musique.

Ils définissent un nouveau dictionnaire, dit harmonique, puis ils l'utilisent pour reconnaître les différentes notes de musique présentes dans le signal grâce à un Matching Pursuit spécialisé à ce dictionnaire harmonique. Cette méthode permet de reconnaître les notes même dans des conditions difficiles, comme par exemple, des durées de note très différentes, la présence de réverbérations ou de bruit, etc.

5 Conclusion

Le Matching Pursuit permet de décomposer un signal en une combinaison linéaire de différents atomes présents dans un dictionnaire. Ces atomes sont choisis en fonction de leur corrélation avec le signal à traiter : on choisit en premier, les atomes qui ont la plus forte corrélation avec le signal. Le signal peut donc être caractérisé par les indices des atomes choisis, ainsi que par les coefficients de corrélation entre le signal et ces derniers.

Un des principaux avantages du Matching Pursuit est que le décomposition résultant de l'algorithme est parcimonieuse, c'est-à-dire qu'elle ne contient que peu de coefficients significatifs. En effet, l'information du signal est regroupée dans les coefficients des n premiers atomes. Cette caractéristique de la décomposition rend possible l'utilisation du Matching Pursuit dans de nombreux domaines du traitement du signal. Par exemple,

- pour compresser un signal, il suffit de le coder avec les indices des atomes et les coefficients de corrélation.
- pour débruiter un signal, il suffit de garder les premiers atomes qui correspondent à des structures cohérentes et de ne pas garder les autres qui correspondent au bruit.
- pour reconnaître de l'information dans un signal, il suffit de regarder les premiers atomes qui ressemblent le plus au signal et ainsi d'en déduire de l'information.

Néanmoins, le Matching Pursuit est souvent considéré comme un algorithme un peu trop lent, en particulier s'il utilise un dictionnaire redondant de très grande taille. En effet, malgré des améliorations proposées par Mallat et Zhang [7] pour améliorer la version originale du Matching Pursuit, la complexité d'une itération de l'algorithme est d'au moins $O(N)$, où N est le nombre d'échantillons du signal traité. Sachant que pour atteindre une précision acceptable pour la reconstruction du signal après traitement, il faut en moyenne un nombre d'itération égal à une fraction de N, d'où une complexité générale de l'algorithme d'au moins $O(N^2)$. Cependant, il existe maintenant des versions plus récentes de l'algorithme du Matching Pursuit qui permettant de réduire les temps de calcul. Par exemple, Gribonval et Krstulovic [5] proposent une méthode qui permet de descendre la complexité d'une itération à $O(\log N)$ et donc la complexité de l'algorithme générale à $O(N \log N)$

Références

[1] G. Davis, S. Mallat, and Z. Zhang. Adaptive time-frequency approximations with matching pursuits. Technical Report TR1994-657, 1994.

[2] R. Gribonval and E. Bacry. Harmonic decompositions of audio signals with matching pursuit. *IEEE Transactions on Signal Processing*, 51(1) :101–111, January 2003.

[3] R. Gribonval, X. Depalle, E. Bacry, and S. Mallat. Sound signal decomposition using a high resolution matching pursuit. In *Proceedings of International Computer Music Conference (ICMC'96)*, pages 293–296, August 1996.

[4] R. Figueras i Ventura, P. Vandergheynst, and P. Frossard. Low rate and scalable image coding with redundant representations. Technical Report TR-ITS-03.02, June 2003.

[5] S. Krstulovic and R. Gribonval. MPTK : Matching pursuit made tractable. In *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP'06)*, volume 3, pages 496–499, Toulouse, 2006.

[6] S. Mallat and F.Bergeaud. Matching pursuit : Adaptative representations of images. *Computational and Applied Mathematics*, 15(2), October 1996.

[7] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12) :3397–3415, 1993.