Extension de Matching Pursuit au modèle des moindres carrés totaux

Sébastien Maria

▶ **To cite this version:**

Sébastien Maria. Extension de Matching Pursuit au modèle des moindres carrés totaux. MajecSTIC 2005 : Manifestation des Jeunes Chercheurs francophones dans les domaines des STIC, IRISA – IETR – LTSI, Nov 2005, Rennes, pp.188-195. inria-00000725

HAL Id: inria-00000725

https://inria.hal.science/inria-00000725

Submitted on 15 Nov 2005

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.

Extension de Matching Pursuit au modèle des moindres carrés totaux

Sébastien MARIA
IRISA/Université de Rennes I
Campus de Beaulieu, 35042 Rennes Cedex, France
sebastien.maria@irisa.fr

Résumé : La modélisation d'une observation est un problème récurrent dans de nombreux domaines. Une idée assez naturelle est d'imaginer un ensemble important de signaux parmi lesquels certains seraient susceptibles de la décrire de façon satisfaisante. Très souvent pour modéliser ce lien, on utilise une régression linéaire. Le vecteur d'observation peut donc être décrit comme une somme de vecteurs de signaux potentiels. Nous supposons que seul un nombre restreint de signaux suf£t à une modélisation convenable, il devient alors nécessaire de les identi£er (cela revient à chercher une représentation parcimonieuse de notre observation).

Le modèle considéré est sous-déterminé c'est-à-dire que le nombre de prédicteurs potentiels (n) est beaucoup plus important que la taille du vecteur d'observation (m) : $m << n$. Les algorithmes de sélection de variables utilisés en statistiques ne s'appliquent que dans le cas surdéterminé ($m > n$). Différentes procédures ont vu le jour pour pallier cette dif£culté. Celle qui est probablement la plus utilisée pour les modèles sous-déterminés est appelée Matching Pursuit (MP). Nous décrirons donc cet algorithme ainsi qu'une de ses variantes : Orthogonal Matching Pursuit (OMP). Puis nous proposerons un algorithme plus général qui prend en compte la possibilité que du bruit gaussien puisse perturber à la fois l'observation et les prédicteurs. Nous introduirons dans ce but le modèle des moindres carrés totaux.

Mots-clés : estimation, matching pursuit, orthogonal matching pursuit, moindres carrés, moindres carrés totaux, parcimonie, systèmes sous-determinés.

1 INTRODUCTION

Chercher à approcher une observation d'une façon parcimonieuse correspond à trouver une représentation convenable d'un signal observé en l'écrivant comme combinaison linéaire d'un nombre restreint de signaux élémentaires judicieusement choisis parmi un ensemble de signaux important. Ce problème est apparu dès les années cinquante dans le domaine des statistiques sous le nom de sélection de variables ou sélection de sous-ensembles. Pour réaliser cette sélection, plusieurs algorithmes ont déjà été proposés dans la littérature s'appuyant sur les propriétés statistiques de l'estimée de l'observation. Ils visaient à déterminer quels étaient les signaux qui participaient signi£cativement à la régression ainsi qu'éliminer ceux qui n'avaient que peu d'in¤uence [Draper, 1981]. Cependant, ce type d'algorithmes n'est réalisable que pour des modèles dits sur-déterminés (c'est-à-dire que la taille du vecteur d'observation (m) est plus important que le nombre de signaux potentiels (n) qui sont capables de le décrire, appelés prédicteurs : $m > n$).

Or, ces dix dernières années, la communauté du traitement du signal s'est intéressée au cas sous-déterminé. En effet chercher une approximation parcimonieuse d'un signal trouve des applications dans de nombreux domaines tels que la compression audio et l'analyse du son [Gribonval, 2003], l'image [Frossard, 2004], la vidéo [Nguyen, 2003], l'estimation de canaux [Karabulut, 2004], la modélisation de signaux [Rissanen, 1979], l'apprentissage machine [Girosi, 1998] ou encore la régularisation [Daubechies, 2004], etc. Or dans le cas sous-déterminé, les algorithmes de sélection de variables utilisés en statistiques ne sont pas applicables. Pour résoudre cette dif£culté, divers algorithmes furent proposés. Les principaux sont Matching pursuit [Mallat, 1993], Orthogonal Matching Pursuit [Pati, 1993] ainsi que d'autres s'appuyant sur les propriétés de la norme $\ell 1$ matricielle dont le principe apparaît dans [Tibshirani, 1996], [Chen, 1998] ou encore [Fuchs, 2001]. Ils répondent tous au critère de parcimonie de la représentation de l'observation. Ils vont même au-delà car ils permettent une bonne approximation du signal observé lorsque celui-ci se trouve contaminé par du bruit. Pour clari£er le modèle, nous considérons la régression linéaire suivante :

$$y = Ax + e = \sum_{i=1}^{n} a_i x_i + e \qquad (1)$$

Ici, y représente le vecteur d'observation de dimension m. La matrice de régression $A \in \mathbb{R}^{m \times n}$ est nommée dictionnaire. Elle contient tous les signaux prédicteurs potentiels à savoir $a_1, ..., a_n$ (colonnes de cette même matrice) qui sont appelés atomes, ils sont supposés être normalisés. Généralement deux possibilités se présentent concernant le dictionnaire : soit l'observateur le construit lui même de sorte qu'il soit constitué de suf£sament d'atomes pour décrire au mieux l'observation, soit l'observateur dispose d'un dictionnaire imposé dès le

départ à partir duquel il veut modéliser l'observation (cela peut être un certain type de signaux élémentaires ou des signaux mesurés par exemple).
Le vecteur e modélise l'erreur de mesure, elle sera représentée par du bruit gaussien. Finalement, x est un vecteur de dimension n contenant les coef£cients de la régression. Dans la recherche d'une solution parcimonieuse, ce vecteur x possède beaucoup de composantes nulles.

L'un des buts de cette communication est de présenter l'algorithme qui est probablement le plus utilisé pour résoudre les cas sous-déterminés, à savoir Matching pursuit (MP). Il est détaillé dans la partie 2. Nous nous intéressons dans la partie 3 à une version plus élaborée de MP, à savoir Orthogonal Matching Pursuit (OMP) proposé dans [Pati, 1993].
L'autre objectif de cette communication est d'introduire un algorithme plus général que les deux premiers. En effet, nous considérons la possibilité que des nuisances perturbent à la fois le vecteur y mais aussi le dictionnaire. Dans ce but, nous modi£ons (1) par le modèle des moindres carrés totaux (Total Least Squares, TLS) [Golub, 1980], [Van Huffel, 1991] à savoir

$$y = Ax + e \quad \text{avec} \quad Z = A + N, \tag{2}$$

où seuls le vecteur y et la matrice Z sont connus ou observés, e et N modélisant les erreurs de mesure. L'étude des moindres carrés totaux est faite dans la partie 4 et son lien avec une version plus élaborée que MP et OMP en partie 5. Finalement la partie 6 donne quelques résultats expérimentaux visant à comparer les différentes procédures.

2 L'ALGORITHME MATCHING PURSUIT (MP)

Trouver une solution parcimonieuse à partir du dictionnaire en utilisant une recherche exhaustive est infaisable en pratique (trop de possibilités combinatoires). Pour résoudre ce problème, des méthodes sous optimales furent proposées. L'une d'entre elle, appelée Matching Pursuit, apparaît très simple dans son fonctionnement et son implémentation. Matching Pursuit est un algorithme itératif qui construit une estimée de l'observation y en ajoutant, à chaque itération, un atome à l'ensemble des atomes déjà pré-selectionnés. Nous allons maintenant expliquer son fonctionnement.

Au début de l'algorithme, on détermine l'atome le plus corrélé avec l'observation notée y. Lorsque celui-ci est sélectionné, on choisit comme estimée le projeté orthogonal de y sur cet atome. Les itérations suivantes visent à estimer l'information restante (c'est ce qu'on nomme résidu). On procède alors de la même manière que précedemment. On cherche l'atome le plus corrélé avec le résidu, puis on approxime ce résidu en le projetant sur l'atome selectionné. On ajoute £nalement l'approximation du résidu à l'estimée courante de l'observation. On renouvelle ainsi de suite cette méthode jusqu'à ce qu'un critère d'arrêt soit atteint (norme du résidu inférieure à un seuil donné par exemple). Pour le détail de l'algorithme, on notera plus formellement $r^{(k)}$ le résidu entre y et son estimée $\hat{y}^{(k)}$ à l'itération k de l'algorithme.

Description de l'algorithme :

Initialisation à $\hat{y}^{(0)} = 0$, $r^{(0)} = y$ et $k = 1$. A l'itération k :

1. Calcul de toutes les corrélations $\{\langle a_i, r^{(k-1)}\rangle = a_i^T . r^{(k-1)}\}_{i=1..n}$
2. Choix de la plus forte de ces corrélations :
$$i_k = \arg\max_i \mid \langle a_i, r^{(k-1)}\rangle \mid$$
3. Mise à jour de l'estimée et de son résidu :
   - $\hat{y}^{(k)} = \hat{y}^{(k-1)} + \langle a_{i_k}, r^{(k-1)}\rangle a_{i_k}$
   - $r^{(k)} = y - \hat{y}^{(k)} = r^{(k-1)} - \langle a_{i_k}, r^{(k-1)}\rangle a_{i_k}$
4. Incrémentation de k, $(k \leftarrow k+1)$ et retour aux étapes 1, 2 et 3 jusqu'à ce qu'un critère d'arrêt soit atteint ($\|y - \hat{y}^{(k)}\|_2 \leq \epsilon$ où ϵ doit être spéci£é au départ).

MP apparaît donc comme une procédure à la mise en oeuvre très pratique et dont le coût en temps de calcul est le moins important pour toutes les procédures visant à résoudre les problèmes sous-déterminés. On peut cependant donner une limitation à cet algorithme qui va conduire à la procédure OMP. Prenons pour cela le cas concret des 2 premières itérations de l'algorithme.
A la première itération, supposons qu'on choisisse l'atome a_1 qui est l'atome le plus corrélé avec l'observation. L'estimée $\hat{y}^{(1)}$ va correspondre au projeté orthogonal de y sur l'espace engendré par l'atome a_1. Le résidu quant à lui va appartenir à son espace orthogonal c'est à dire que $\langle a_1, r^{(1)}\rangle = 0$. Lors de la deuxième itération, un deuxième atome va être selectionné, supposons que ce soit a_2. Le résidu $r^{(1)}$ va être projeté sur a_2, d'où l'obtention d'un $r^{(2)}$. A ce stade de l'algorithme voici ce que l'on peut conclure : $\langle a_1, r^{(1)}\rangle = 0$; $\langle a_2, r^{(2)}\rangle = 0$; mais par contre $\langle a_1, r^{(2)}\rangle \neq 0$. Ainsi, la sélection dans MP n'exclut pas de choisir une nouvelle fois l'atome a_1 à une autre itération de l'algorithme. Sur un exemple précis (tel que ceux vus dans la partie 5) où l'on construit un dictionnaire $A \in \mathbb{R}^{128\times 256}$ avec 5 atomes appartenant effectivement à la régression, 5 itérations de l'algorithme ne garantissent pas que l'on retrouvera les 5 atomes du départ. Ainsi, la théorie af£rme que, dans un cas non bruité, l'estimée converge mais cette convergence peut être très lente. La convergence est garantie pour un nombre in£ni d'itérations et c'est en cela qu'OMP va apparaître plus performant.

3 VERS UNE VERSION PLUS EFFICACE DE MP : ORTHOGONAL MATCHING PURSUIT

L'algorithme OMP (Orthogonal Matching Pursuit) peut fournir, selon l'application, un bon compromis pour

ses performances d'estimation et le temps de calcul nécessaire à son fonctionnement. OMP est appelé de cette façon car il correspond à une version modi£ée de MP. Il permet de résoudre le problème lié à la sélection d'atomes qui ont été déjà choisis. En effet, dans cette procédure, k itérations permettent d'obtenir k atomes distincts. Pour expliquer cela, nous allons étudier OMP plus en détail.

De la même manière que pour MP, à chaque étape de l'algorithme, on cherche à déterminer l'atome le plus corrélé avec le résidu courant. Cependant, contrairement à MP, on ne cherche pas à reconstruire le résidu puis à l'ajouter à l'estimée de l'itération précédente mais on recalcule l'estimée à chaque itération. A chaque fois que l'on détermine un nouvel atome, on choisit comme estimée $\hat{y}^{(k)}$, le projeté de y sur l'espace engendré par tous ces atomes. Ainsi, au bout de k itérations, $\forall i = 1..k$, $\langle a^{(i)}, r^{(k)} \rangle = 0$, ce qui fait qu'à chaque itération un nouvel atome sera selectionné.

Nous conservons les notations précédemment utilisées à savoir $r^{(k)}$ le résidu entre y et son estimée $\hat{y}^{(k)}$ pour l'étape k de l'algorithme. L'ensemble des atomes sélectionnés à l'itération k sera contenu dans la matrice $A_k = \left[a^{(1)} \mid \cdots \mid a^{(k)}\right]$. On notera pour £nir $\hat{x}^{(k)}$ le vecteur contenant les coef£cients de la régression linéaire. Nous rappelons donc que dans l'algorithme OMP, ils sont réévalués à chaque itération pour faire en sorte que le résidu n'appartienne pas à l'espace engendré par la matrice A_k.

Description de l'algorithme :

Initialisation à $\hat{y}^{(0)} = 0$, $r^{(0)} = y$, $A_0 = [\,]$ et $k = 1$. A l'itération k :

1. Calcul de toutes les corrélations $\{\langle a_i, r^{(k-1)} \rangle = a_i^T . r^{(k-1)}\}_{i=1..n}$
2. Choix de la plus forte de ces corrélations :

$$i_k = \arg\max_i \mid \langle a_i, r^{(k-1)} \rangle \mid$$

3. Mise à jour de l'estimée :
   - $a^{(k)} = a_{i_k}$ et donc $A_k = \left[A_{k-1} \mid a^{(k)}\right]$
   - $\hat{x}^{(k)} = A_k^+ y \Rightarrow \hat{y}^{(k)} = A_k \hat{x}^{(k)}$
   - alors $r^{(k)} = y - \hat{y}^{(k)}$
4. Incrémentation de k, $(k \leftarrow k + 1)$ et retour aux étapes 1, 2 et 3 jusqu'à ce que le critère d'arrêt soit atteint.

où $A_k^+ = (A_k^T A_k)^{-1} A_k^T$ est la pseudo-inverse de Moore-Penrose de A_k. La première chose à remarquer dans cet algorithme est que l'étape 3 correspond bien à la projection de Y sur l'espace engendré par les atomes $a^{(j)}$ pour $j = 1..k$. En effet la matrice $A_k A_k^+$ est par dé£nition un projecteur. Comme nous le verrons plus loin, ce type de résultat est à comparer à la solution des moindres carrés (la matrice A_k possède plus de lignes que de colonnes et c'est à ce titre que l'on peut parler de solution au problème des moindres carrés applicables seulement pour dans le cas sur-déterminé).

D'autre part, chaque étape de la procédure nécessite le calcul de la pseudo-inverse d'une matrice. Cela accroît grandement le coût en terme de temps de calcul pour la procédure. Cette dif£culté peut être contournée en utilisant les résultats de la pseudo-inverse d'une matrice partitionnée [Greville, 1960]. Ces résultats permettent de calculer A_k^+ à partir de A_{k-1}^+ ce qui réduit considérablement le coût. L'avantage d'OMP est que l'on peut décider très facilement de représenter l'observation avec un nombre £xé d'atomes sans avoir besoin de surveiller l'apparition de nouveaux atomes dans l'estimation. On peut aussi démontrer en théorie que contrairement à MP, OMP assure que, dans le cas non bruité, l'estimée converge pour un nombre £ni d'itérations.

4 LA NOUVELLE PROCÉDURE TLS-OMP

4.1 Introduction au modèle des moindres carré totaux (TLS)

Dans un premier temps, nous allons rappeler le résultat bien connu des moindres carrés classiques. Si on considère le modèle de régression linéaire (1) dans le cas surdéterminé, c'est-à-dire lorsque $m > n$,

$$y = Ax + e$$

avec e correspondant à un bruit gaussien $N(0, \sigma^2 I)$. Le problème des moindres carrés consiste à résoudre le problème d'optimisation suivant :

$$\begin{aligned} & \min_x \|y - Ax\|_2^2 \\ \Leftrightarrow \quad & \min_{r\ x} \|r\|_2^2 \quad \text{sous} \quad Ax = y + r, \end{aligned} \tag{3}$$

où $\|r\|_2^2 = \sum r_i^2$ correspond au carré de la norme ℓ_2. Cela correspond à déterminer l'estimée au maximum de vraisemblance de y. L'optimum d'un tel problème est atteint pour $\hat{x} = A^+ y$. Dans ce cas, A est supposé être de rang colonne plein c'est-à-dire que toutes les colonnes de la matrice sont linéairement indépendantes entre elles. Cette solution rappelle très clairement la solution proposée par l'algorithme OMP. En effet à la k-ième étape de l'algorithme, le vecteur x est estimé par $\hat{x}^{(k)} = A_k^+ \hat{y}^{(k)}$ où la matrice A_k possède bien plus de lignes que de colonnes. L'idée proposée ici est de remplacer le modèle (1) par un modèle plus général (2) soit :

$$y = Ax + e \quad \text{avec} \quad Z = A + N,$$

où seuls Z et y sont connus. On peut ré-écrire le modèle linéaire en isolant la i-ième ligne :

$$\begin{aligned} y_i &= a_i^T x + e_i \\ z_i^T &= a_i^T + n_i^T . \end{aligned}$$

On suppose que les colonnes de la matrice de bruit N (c'est à dire le bruit qui perturbe les atomes) sont indépendantes entre elles. Pour simpli£er la présentation, on suppose aussi que le bruit est de même nature sur l'observation et sur le dictionnaire. Ainsi, les erreurs e_i et n_i sont supposées être des bruits gaussiens indépendants avec respectivement une variance σ^2 et une matrice de covariance $\sigma^2 I$. Le problème d'optimisation correspondant à (3) pour le problème TLS est le suivant :

$$\min_{x,A} \|y - Ax\|_2^2 + \|(Z - A)\|_F^2, \tag{4}$$

où $\|A\|_F^2 = \sum_{i,j} a_{i,j}^2 = \text{ trace } (AA^T)$ correspond au carré de la norme de Frobenius de la matrice A. Il convient maintenant de réécrire ce problème pour faire apparaître TLS sous sa forme la plus classique. Notons $r = y - Ax$ et $R = Z - A$, alors nous avons :

$$\min_{r,R,x} \|r\|_2^2 + \|R\|_F^2 \qquad \text{sous} \qquad y - r = (Z - R)x$$

En introduisant les notations suivantes des matrices augmentées : $C = [A \mid Ax]$, $\hat{C} = [Z \mid y]$ et $\Delta = [R \mid r]$, (4) devient clairement la forme habituelle des moindres carrés totaux :

$$\min_{\Delta} \|\Delta\|_F^2 \qquad \text{sous} \qquad \hat{C} = C + \Delta \tag{5}$$

$A \in \mathbb{R}^{m \times n}$ et c'est un matrice de rang colonne plein, à savoir $rang(A) = n$. Lorsqu'on crée la matrice augmentée C, on ajoute à A une nouvelle colonne qui est combinaison linéaire de ces autres colonnes, alors $rang(C) = rang(A) = n$. La matrice $\hat{C}$, quant à elle, possède un rang égal à $n + 1$. Ainsi, résoudre (5), revient à déterminer la matrice Δ qui, si on la soustrait à la matrice $\hat{C}$, fait chuter le rang de $\hat{C}$ de 1. On peut voir que cela revient aussi à déterminer la matrice Δ qui est de norme minimale et qui possède un rang de 1. Dans [Golub, 1980] et [Van Huffel, 1991], on démontre que la solution est obtenue à partir de la décomposition en valeurs singulières (SVD) de $\hat{C}$: $\hat{C} = \hat{U}\hat{S}\hat{V}^T$. Si on note $\hat{s}_{min}$ la plus petite valeur singulière de cette SVD et si on note $\hat{u}_{min}$ et $\hat{v}_{min}$ les vecteurs singuliers associés à $\hat{s}_{min}$, on trouve que $\hat{\Delta} = \hat{s}_{min}\hat{u}_{min}\hat{v}_{min}^T$. De plus, l'optimum $\hat{x}$ est donné par

$$[\hat{x}^T \mid -1] = \hat{v}_{min}^T / \hat{v}_{min}(n+1), \tag{6}$$

Ce qui correspond à normaliser la dernière composante de $\hat{v}_{min}$ à -1. Nous pouvons maintenant proposer un algorithme de type OMP mais intégrant ce résultat.

4.2 Détail de l'algorithme

Nous supposons donc que nous n'avons accès qu'aux données Z et y et pas à la matrice A. L'algorithme présenté ici possède donc la même structure que les deux précédents. On le nomme TLS-OMP car il correspond à une version plus élaborée que OMP. Le point qui change par rapport à OMP est l'estimation des coef£cients de régression et donc de l'estimée de y à la k-ième itération de l'algorithme.

Dans la suite, nous choisissons les notations suivantes : à l'étape k, $\hat{y}^{(k)}$, $r^{(k)}$ et $x^{(k)}$ correspondent toujours aux mêmes valeurs estimées que pour OMP. La matrice $Z_k = \left[z^{(1)} \mid ... \mid z^{(k)}\right]$ contiendra l'ensemble des atomes (à priori perturbés) qui ont été selectionnés. On notera $\hat{C}_k$ la matrice augmentée $\left[Z_k \mid y\right]$ et $\hat{u}_{min}^{(k)}$, $\hat{s}_{min}^{(k)}$, $\hat{v}_{min}^{(k)}$ les différents éléments liés à la résolution de TLS que l'on a vu dans la section précédente. La matrice $\hat{N}_k$ correspondra à l'estimation de N et $\hat{A}_k$ à l'estimation de A_k. Finalement par commodité, nous considérerons que $v(1:p)$ correspond aux p premières composantes du vecteur v.

Description de l'algorithme :

Initialisation à $\hat{y}^{(0)} = 0$, $r^{(0)} = y$, $Z_0 = [\,]$ et $k = 1$. A l'itération k :

1. Calcul de toutes les corrélations $\{\langle z_i, r^{(k-1)}\rangle = z_i^T . r^{(k-1)}\}_{i=1..n}$
2. Choix de la plus forte de ces corrélations :
$$i_k = \arg\max_i \mid \langle z_i, r^{(k-1)}\rangle \mid$$
3. Mise à jour de l'estimée :
   - $z^{(k)} = z_{i_k}$ et donc $Z_k = \left[Z_{k-1} \mid z^{(k)}\right]$,
   - Calcul de la décomposition en valeurs singulières de $\hat{C}_k = \hat{U}_k\hat{S}_k\hat{V}_k^T$,
   - Sélection de $\hat{u}_{min}^{(k)}$, $\hat{s}_{min}^{(k)}$et $\hat{v}_{min}^{(k)}$,
   - Estimation $\hat{N}_k = \hat{s}_{min}^{(k)} . \left[\hat{u}_{min}^{(k)}\right] . \left[\hat{v}_{min}^{(k)}(1:n)\right]^T$ puis de $\hat{A}_k = Z_k - \hat{N}_k$,
   - $\hat{x}^{(k)} = -\frac{v_{min}^{(k)}(1:n)}{v_{min}^{(k)}(n+1)} \;\Rightarrow\; \hat{y}^{(k)} = \hat{A}_k\hat{x}^{(k)}$,
   - $r^{(k)} = y - \hat{y}^{(k)}$.
4. Incrémentation de k, $(k \leftarrow k + 1)$ et retour aux étapes 1, 2 et 3 jusqu'à ce que le critère d'arrêt soit atteint.

De par leur structure, les trois algorithmes sont tout à fait semblables. En effet, le critère de sélection de l'atome pour chaque itération est le choix de la corrélation maximale du résidu avec l'estimée quelle que soit la procédure utilisée. Cependant même si, au même titre que OMP, TLS-OMP sélectionne au bout de k itérations, k atomes distincts, le coût du calcul des SVD est assez important. Il n'y a d'ailleurs pas possibilité d'utiliser une astuce comme celle vue dans OMP pour réduire ce coût additionnel.

5 RÉSULTATS EXPÉRIMENTAUX

Dans cette section nous allons observer les performances des trois algorithmes et montrer que l'algorithme TLS-OMP prend en compte des cas plus généraux que ses deux prédécesseurs c'est à dire qu'en possédant les mêmes

capacités d'estimation pour un modèle non perturbé il s'avère plus performant sur des modèles où le bruit est présent partout. Pour ce faire nous allons les étudier sur des données simulées. Nous construisons un dictionnaire réel A de taille 128×256 construit aléatoirement avec des valeurs comprises entre -1 et 1. Nous normalisons tous les atomes du dictionnaire puis nous choisissons aléatoirement 5 atomes que nous pondérons pour construire notre observation. Dans un premier temps, nous étudions le cas non bruité.

5.1 Le modèle non bruité

Si on applique les trois algorithmes à des données non bruitées, les procédures ont les mêmes résultats en terme de détection. En effet, MP, OMP et TLS-OMP retrouvent bien les bons atomes qui ont été insérés dans la régression. C'est ce qu'on peut visualiser sur la £gure 1. Nous n'avons représenté qu'une seule courbe pour les trois algorithmes car visuellement le résultat est le même. En abscisse, nous avons les indices des atomes du dictionnaire et en ordonnée les coef£cients de la régression correspondants. En bleu, ce sont les coef£cents réels et les 'étoiles rouges' donnent la valeur des coef£cients estimés. La détection est correcte et dans l'a priori, on peut supposer que l'estimation est bonne pour les trois algorithmes.

Suit ensuite dans la table 1, la norme du résidu obtenu après application des trois algorithmes. Pour OMP et TLS-OMP, on a effectué 5 itérations de la procédure, pour MP, on en a effectué 10 fois plus pour obtenir une reconstruction correct de l'observation. Il apparaît que l'estimation est bien meilleure pour OMP et TLS-OMP que pour MP même si celui-ci fournit une estimation tout à fait convenable.

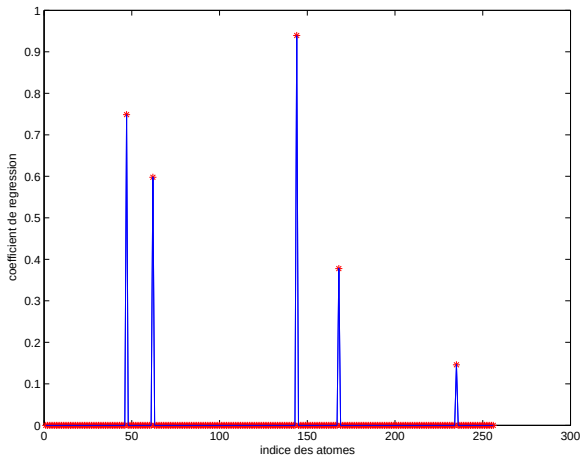

FIG. 1 – Estimation des coef£cients de régression dans le cas sans bruit pour MP, OMP, TLS-OMP

5.2 Le modèle bruité

Nous perturbons maintenant tout le modèle de régression linéaire avec du bruit gaussien de moyenne nulle et

	MP	OMP	TLS-OMP
$\|y - \hat{y}\|_2$	$6.11\ e^{-4}$	$5.97\ e^{-16}$	$5.28\ e^{-16}$

TAB. 1 – Norme du résidu dans le cas non bruité

d'écart-type $\sigma = 0,015$. Pour OMP et TLS-OMP, nous avons effectué 5 itérations correspondant aux 5 atomes insérés dans le modèle. Pour MP, nous avons effectué un nombre d'itération qui correspond à la sélection de 5 atomes. Nous obtenons trois résultats différents pour chaque algorithme, ils apparaissent dans les £gures ci-dessous.

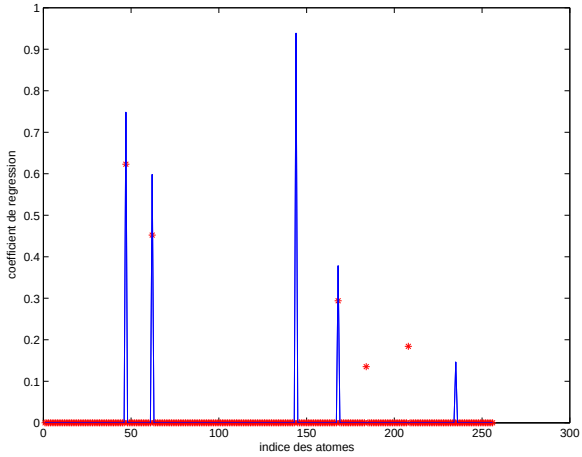

FIG. 2 – Estimation des coef£cients de régression en présence de bruit pour MP

Ici il est clair que MP (£g. 2) n'est plus ef£cace : il ne retrouve plus les bons atomes quelque soit le nombre d'itération qu'il fait. Nous pouvons donc passer à la comparaison entre OMP (£g. 3) et TLS-OMP (£g. 4).

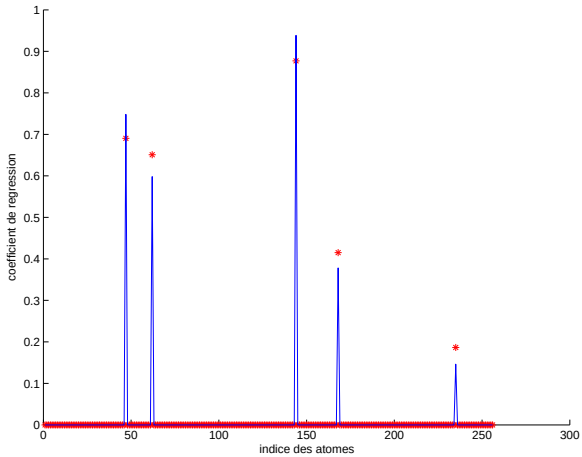

FIG. 3 – Estimation des coef£cients de régression en présence de bruit pour OMP

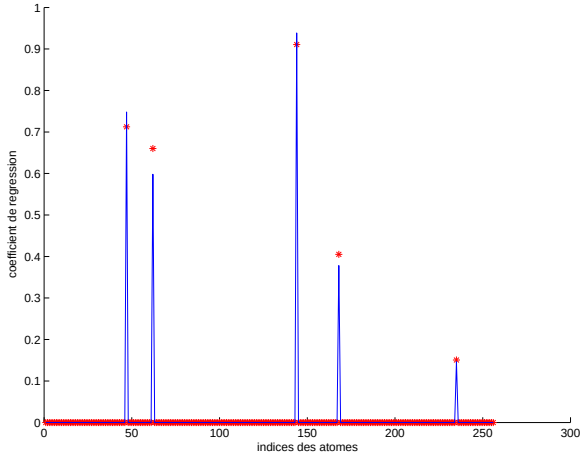

FIG. 4 – Estimation des coef£cients de régression en présence de bruit pour TLS-OMP

Les deux algorithmes possèdent les mêmes propriétés de détection dans le cas bruité. La différence se fait au niveau de la qualité de l'estimation. Visuellement, on peut voir que le dernier atome est mieux estimé par TLS-OMP que par OMP. Ceci se traduit aussi dans la table 2 qui donne la valeur du résidu pour chaque estimation.

	MP	OMP	TLS-OMP
$\|y - \hat{y}\|_2$	0.1833	0.2676	0.0911

TAB. 2 – Norme du résidu dans le cas bruité

La qualité d'une estimation va dépendre du modèle utilisé. Dès que l'on va choisir d'appliquer des algorithmes tels MP ou OMP, il convient de s'interroger si TLS-OMP ne serait pas plus approprié. Si le dictionnaire est parfaitement connu, MP ou OMP (selon le cas de £gure) peut convenir parfaitement. Cependant dès qu'apparaît une incertitude sur la connaissance du dictionnaire, il est préférable de supposer que ce dictionnaire est connu à une erreur près. C'est le modèle TLS qui intervient alors et, dans ce cas, les résultats vus précédemment démontrent d'une meilleure qualité de l'estimation.

6 CONCLUSION

Nous avons présenté dans cette communication deux algorithmes très utilisés par des domaines qui nécessitent soit l'approximation, soit la décomposition, soit l'estimation d'une observation par un certain type de signaux. Ces algorithmes ont de très bonnes performances dans le cas où l'observation et les régresseurs ne sont pas perturbés. Nous avons proposé ensuite un modèle plus général qui prend en compte la possibilité de nuisance sur l'observation mais aussi sur ses prédicteurs. Cela nous a conduit à décrire le modèle des moindres carrés totaux. Ce type d'algorithme apparaît comme une bonne alternative à OMP si la modélisation du problème admet la possibilité d'incertitude sur la connaissance des prédicteurs.

Par soucis de clarté et pour simpli£er notre présentation, nous avons supposé que le bruit qui perturbe notre observation était identique à celui qui perturbe les atomes. Ce n'est bien évidemment pas toujours le cas. Il serait tout à fait possible d'intégrer une hypothèse supplémentaire à notre modèle prenant en compte cet aspect. Une telle modi£cation ne changerait cependant pas la structure de l'algorithme.
Une autre perspective de ce travail serait de s'intéresser à d'autres algorithmes utilisés dans les cas sous-déterminés basés notamment sur l'utilisation de la norme $\ell 1$ ([Tibshirani, 1996], [Chen, 1998], [Fuchs, 2001]). Il semble tout à fait possible d'établir des extensions de ce type d'algorithmes au modèle des moindres carrés totaux.

Références

[Chen, 1998] Chen S. S., Donoho D. L. et Saunders M. A., Atomic decomposition by basis pursuit. *SIAM J. Scienti£c computing*, 20(1) :33–61.

[Daubechies, 2004] Daubechies I., Defrise M. et De mol C., An iterative thresolding algorithm for linear inverse problems with a sparsity contrainst. *Communication on Pure and Applied Mathematics*, 57(11) :1413–1457.

[Draper, 1981] Draper N. et Smith H., *Applied regression analysis*. Wiley.

[Frossard, 2004] Frossard P., Vandergheynst P., Figueras i Ventura R. M. et Kunt M., A posteriori quantization of progressive matching pursuit streams. *IEEE Trans. Signal Proc.*, 52(2) :525–535.

[Fuchs, 2001] Fuchs J. J., On the application of the global matched £lter to doa estimation with uniform circular arrays. *IEEE Trans. on Signal Proc.*, 49(4) :702–709.

[Girosi, 1998] Girosi F., An equivalence between sparse approximation and support vector machines. *Neural Comput.*, 10(6) :1455–1480.

[Golub, 1980] Golub G. H. et Loan C. F. V., An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17(6) :883–893.

[Greville, 1960] Greville T., Some applications of the pseudoinverse of a matrix. *SIAM Review*, 2 :15–22.

[Gribonval, 2003] Gribonval R. et Bacry E., Harmonic decomposition of audio signals with matching pursuit. *IEEE Trans. Signal Proc.*, 51(1) :101–111.

[Karabulut, 2004] Karabulut G., Kurt T. et Yongacoglu A., Channel estimation by matching pursuit algorithm. *In Proc. of the IEEE 60th Vehicular Technology Conference, Los Angeles*.

[Mallat, 1993] Mallat S. et Zhang Z., Matching pursuit with time-frequency dictionnaries. *IEEE Trans. on Signal Proc.*, 41(12) :3397–3415.

[Nguyen, 2003] Nguyen T. et Zakhor A., Matching pursuits based multiple description video coding for lossy environments. *In Proc. of the 2003 IEEE International Conference on Image Processing, Barcelona*.

[Pati, 1993] Pati Y. C., Rezaiifar R. et Krishnaprasad P. S., Orthogonal matching pursuit : Recursive function

approximation with applications to wavelet decomposition. *In Proc. of 27th Asimolar Conf. on Signals, Systems and Computers, Los Alamitos*.

[Rissanen, 1979] Rissanen J., Modeling by shortest data description. *Automatica*, 14 :465–471.

[Tibshirani, 1996] Tibshirani R., Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society, Series B*, 58 :267–288.

[Van Huffel, 1991] Van Huffel S. et Vanderwalle J., *The Total Least Squares problem*. SIAM.