Solution parcimonieuse pour des systèmes linéaires sous déterminés.

Jean-Jacques Fuchs
IRISA/Université de Rennes I
Campus de Beaulieu - 35042 Rennes Cedex
fuchs@irisa.fr

Résumé – Le problème traité est le suivant : étant donnés une matrice A de dimension $(n,\ m)$ avec $m > n$ et un vecteur $b = Ax_o$ avec x_o ayant $q < n$ composantes non nulles, sous quelles conditions sur A et q est il possible de retrouver x_o en résolvant un programme linéaire? Nous considérons les cas où A est une matrice de Vandermonde ou une matrice de Fourier réelle et les composantes de x_o sont positives ou nulles et obtenons des conditions suffisantes moins fortes que celles connues pour des matrices A quelconques.

Abstract – We consider the following problem : given a $n \times m$ matrix A with $m > n$ and a vector $b = Ax_o$ with x_o having $q < n$ nonzero components, under which conditions on A and q is it possible to recover x_o by solving a linear program ? We consider the cases where A is a Vandermonde matrix or a real Fourier matrix and the components of x_o are known to be ≥ 0 and get sufficient conditions that are weaker than those known for abitrary A matrices.

1 Introduction

Dans de nombreuses applications, on est amené à rechercher la décomposition la plus parcimonieuse d'un signal sur les éléments d'un jeu largement redondant de vecteurs de *bases*. On peut alors distinguer principalement trois types de problèmes ou définir la solution optimale de trois façons différentes : ⋄ trouver la décomposition exacte utilisant le plus petit nombre de vecteurs, ⋄ trouver la décomposition de complexité donnée qui minimise l'erreur de reconstruction, ou ⋄ trouver la décomposition la plus parcimonieuse ayant une erreur de reconstruction inférieure à un certain seuil [1]. Nous avions considéré ce type de problème dès le Gretsi 1997 [6] mais sa complexité dans le cas général ne permet d'obtenir que des résultats peu exploitables.

Ce n'est que très récemment en considérant un cas particulier et simple du premier de ces problèmes que des résultats concrets ont été obtenus [2], [3]. En utilisant les résultats plus généraux de [6] nous avons pu améliorer ces premiers résultats [5] et nous nous proposons ici de poursuivre dans le cas de deux types de *bases* ou dictionnaires fréquemment utilisés.

2 Formulation du problème

Soient m vecteurs a_j de dimension n et A la matrice de dimension (n,m) associée, toute combinaison linéaire b de ces m vecteur-colonnes s'écrit alors $b = Ax$ avec x un vecteur à m composantes. Si b est construit à partir d'un petit nombre de colonnes de A, il est possible qu'il n'existe pas d'autre décomposition aussi ou plus parcimonieuse. Mais comme retrouver la décomposition la plus parcimonieuse à partir de la seule connaissance de b est une tâche d'une complexité prohibitive, on s'intéresse dans [2]-[5] au cas où cela est possible en résolvant le problème d'optimisation suivant :

$$\min_x \|x\|_1 \quad \text{sous :} \quad Ax = b \qquad \text{(PL1)}$$

où $\|x\|_1 = \sum\ |x_i|$ désigne la norme ℓ_1 de x. Car on espère que choisir parmi toutes les solutions d'un système linéaire sous déterminés, celle de norme ℓ_1 minimale donne une solution ayant peu (le moins possible) de composantes non nulles. Ce problème d'optimisation se transforme sans difficulté en un programme linéaire dont l'optimum est simple à obtenir [7]. On associe à la variable x_i deux variables x_i^+ et x_i^- positives ou nulles représentant la partie positive et négative de x_i et on remplace alors x_i par $x_i^+ - x_i^-$ et $|x_i|$ par $x_i^+ + x_i^-$. En posant $z^T = [\ x_i^{+T}\ \ x_i^{-T}\ ]$, (PL1) s'écrit :

$$\min_z\ \mathbf{1}^T z \quad \text{sous :} \quad [A\ \ -A]\ z = b, \quad z \geq 0$$

On sait que pour ce programme linéaire l'optimum est toujours atteint en un point ayant au plus n composantes non nulles mais, en supposant qu'il existe au moins un point admissible bien plus parcimonieux, on veut que l'optimum soit en un tel point que l'on appelle alors dégénéré.

En notant $\|x\|_0$ le nombre de composantes non nulles de x et en supposant que les colonnes a_j de A sont de norme euclidienne égale à un, on peut montrer que si une solution x de $Ax = b$ vérifie :

$$\|x\|_0 \leq \frac{1}{2}(1 + \frac{1}{M}) \quad \text{avec :} \ M = \sup_{1 \leq i \neq j \leq m} |a_i^T a_j| \qquad (1)$$

alors c'est aussi l'unique point solution de (PL1). Comme (PL1) est un problème d'optimisation convexe, il n'admet pas d'autre optimum et il n'existe donc pas de solution plus parcimonieuse car elle vérifierait (1) et on aurait une contradiction. On a donc établi que si une solution x satisfait (1), c'est la solution la plus parcimonieuse, elle est unique et solution de (PL1).

Les articles publiés à ce jour [2], [3] établissent ce résultat dans le cas où $m = 2n$ et A est la concaténation de deux matrices orthogonales. Il est montré dans [4], [5] qu'il reste vrai pour une matrice A quelconque dont les colonnes sont de norme euclidienne égale à un. Bien entendu quand le nombre m de colonnes augmente, M dans (1) augmente et le seuil sur $\|x\|_0$ diminue.

Dans la suite nous allons obtenir un résultat plus fort dans le cas de deux matrices A bien particulières et de pondérations positives ou nulles. Le fait de supposer que les pondérations sont positives ou nulles change la nature de la question et, contrairement à (1), le résultat que nous obtenons ne fait plus intervenir le nombre de colonnes.

Nous considérons d'une part la matrice de Vandermonde dont les colonnes sont de la forme :

$$v(\alpha) = [1 \;\; \alpha \;\; \alpha^2 \;\; ... \;\; \alpha^k \;\; ... \;\; \alpha^{n-1}]^T \tag{2}$$

et d'autre part la matrice de $Fourier$ réelle dont les colonnes sont de la forme :

$$\begin{aligned} s(\omega) \quad &= \frac{1}{2}(v(e^{i\omega}) + v(e^{-i\omega})) \qquad (3)\\ = [1 \quad \cos\omega \qquad &\cos 2\omega \;\; ... \;\; \cos k\omega \;\; ... \;\; \cos(n-1)\omega]^T. \end{aligned}$$

Comme dans [5] nous exploitons les conditions nécessaires et suffisantes satisfaites en un optimum de (PL1) et celles garantissant l'unicité de cet optimum.

3 Conditions d'optimalité d'un PL

Soit le programme linéaire, noté (PL) :

$$\min_x \mathbf{1}^T x \quad \text{sous : } \quad Ax = b \quad \text{ et } \quad x \geq 0 \tag{PL}$$

où $\mathbf{1}$ est un vecteur colonne dont toutes les composantes valent un. (PL) correspond à (PL1) quand les composantes de x sont positives ou nulles, le dual de (PL) est [7] :

$$\max_d b^T d \quad \text{sous : } \quad A^T d \leq \mathbf{1} \tag{PLD}$$

Pour b donné, soit $x_o \geq 0$ un point admissible de (PL) nous notons $\bar{x}_o > 0$ le vecteur de dimension réduite des composantes non nulles de x_o et $\bar{A}_o$ la matrice des colonnes de A associées à $\bar{x}_o$, on a alors $Ax_o = \bar{A}_o\bar{x}_o = b$. Le point x_o est alors un optimum de (PL) si et seulement si on peut lui associer un point d_o vérifiant $A^T d_o \leq \mathbf{1}$ et $\mathbf{1}^T x_o = b^T d_o$, qui est alors l'optimum du dual. De plus x_o est l'unique optimum de (PL) si aucune des contraintes du dual n'est dégénérée en d_0 : $a_j^T d_o < 1 \;\; \forall a_j \notin \; \bar{A}_o$.
En résumé, l'optimum de (PL) est unique et égal à x_o si :

$$\exists d_o \;\ni \qquad \bar{A}_o^T d_o = \mathbf{1} \quad \text{et} \quad a_j^T d_o < 1, \qquad \forall a_j \notin \bar{A}_o \tag{4}$$

Ce n'est qu'une condition suffisante (d'unicité), mais comme on s'intéresse à un problème primal dont l'optimum x_0 est dégénéré, c.a.d. dont l'optimum satisfait $\|x_o\|_0 < n$, on ne peut s'attendre à mieux. Quand l'optimum du primal est dégénéré, celui du dual d_o est indéterminé et cela signifie aussi que dés que l'on perturbe b, le primal admet différents optimums non parcimonieux.

Dans la suite nous donnons une condition sur $\|x\|_o$ garantissant l'existence d'un vecteur d satisfaisant (4) dans deux cas bien particuliers de matrice A.

4 Le cas de la matrice de Vandermonde

Les vecteurs colonnes de A sont de la forme $v(\alpha)$ (2) avec α prenant par exemple ses valeurs α_i en m points sur une grille régulière dans (-1, +1). Nous cherchons sous quelles conditions, *a priori* sur A et $\|x_o\|_o$, la résolution de (PL) avec $b = Ax_o$ permet de retrouver x_o dont les composantes sont supposées positives ou nulles.

Nous allons montrer que c'est le cas si :

$$\|x_0\|_0 \leq \lfloor \frac{n-1}{2} \rfloor \tag{5}$$

où $\lfloor z \rfloor$ désigne la partie entière de $z \geq 0$, en indiquant comment obtenir un vecteur d_o vérifiant (4) dans ce cas.

On peut noter que contrairement à (1), cette condition suffisante ne dépend pas de m par le biais de M et ne requiert pas une normalisation à un des colonnes de A.

L'idée est de construire un polynôme qui s'annule au point α_i des colonnes de $\bar{A}_o$ à préserver et qui est strictement positif ailleurs. On déduit alors de ce polynôme un vecteur d satisfaisant (4).

A $v(\alpha)$, on associe le polynôme : $P_\alpha(z) = -\alpha + z$ dont le carré : $P_\alpha^2(z) = \alpha^2 - 2\alpha z + z^2$ vérifie $P_\alpha^2(\alpha) = 0$ et $P_\alpha^2(\beta) > 0, \quad \forall \beta \neq \alpha$. Au polynôme $P_\alpha^2(z)$, on associe ensuite le vecteur $f^T = [\; \alpha^2 \;\; -2\alpha \;\; 1 \;\; 0 \;\; 0 \;\; ... \;\; 0 \;]$ qui vérifie $f^T v(\alpha) = 0$ et $f^T v(\beta) > 0, \quad \forall \beta \neq \alpha$ et enfin le vecteur $d^T = [\; 1 \; 0 \; 0 \; 0 \;\; ... \;\; 0 \;] - f^T$:

$$d^T = [\; 1\text{-}\alpha^2 \;\; 2\alpha \;\; -1 \;\; 0 \;\; 0 \;\; ... \;\; 0 \;]$$

qui vérifie (4). S'il s'agit d'isoler deux colonnes de paramètres α et β, on procède de le même façon pour construire le vecteur associé au produit des polynômes : $P_\alpha^2(z)P_\beta^2(z)$. Comme pour isoler, de cette façon, q colonnes il faut un polynôme de degré $2q$, les vecteurs f et d associés ont $2q+1$ composantes non nulles d'où la borne (5) sur $q = \|x_o\|_o$, le nombre maximum de colonnes identifiables.

Ce résultat et cette borne n'ont rien de surprenant et peuvent se déduire également de la théorie de la réalisation développée en automatique [8] pour trouver un modèle d'un système linéaire dynamique invariant d'ordre fini à partir de sa réponse impulsionnelle. Un système discret d'ordre un, possédant un pôle en α, a une réponse impulsionnelle partielle proportionnelle à $v(\alpha)$ et la matrice à structure de Hankel carrée construite à l'aide des n composantes disponible de la réponse impulsionnelle partielle est d'ordre $\lfloor \frac{n+1}{2} \rfloor$ et de rang 1. De la même façon la matrice de Hankel construite à l'aide des composantes du vecteur colonne $b = Ax_o$ qui contient les n premiers échantillons de la réponse impulsionnelle d'un système linéaire constitué de la mise en parallèle de $\|x_0\|_0$ système d'ordre un est de rang : $\min(\|x_0\|_0, \lfloor \frac{n+1}{2} \rfloor)$. Si $\|x_0\|_0 < \lfloor \frac{n+1}{2} \rfloor$ ou encore

$\|x_0\|_0 \leq \lfloor \frac{n-1}{2} \rfloor$, elle est donc de rang déficiente et il est facile de montrer que les polynômes associés au vecteur de son noyau ont pour racines communes les α, β, .. des colonnes $v(\alpha)$ de $\bar{A}_o$ utilisées pour construire b.

Nous avons supposé : $\alpha_i \in (-1,\ +1)$, mais dans la démonstration cette hypothèse n'est pas utilisé et peut donc être levée. Du coté, théorie de la réalisation l'hypothèse sur le signe des composantes de x n'est pas requise, dans notre contexte elle est essentielle si on veut passer d'une condition du style (1) à la condition (5).

5 Le cas de la matrice de Fourier réelle

Les vecteurs colonnes de A sont de la forme $s(\omega)$ (3) avec ω prenant m valeurs dans $(0,\ \pi)$, par exemple sur une grille régulière. La condition suffisante que nous allons obtenir est la même (5) que dans le cas de la matrice de Vandermonde.

On procède d'ailleurs de façon similaire. Le polynôme $P_\omega(z)$ qui s'annule quand $z = e^{\pm i\omega}$ est $P_\omega(z) = (z - e^{i\omega})(z - e^{-i\omega}) = z^2 - 2\cos\omega z + 1$. Mais il faut maintenant prendre le module au carré de ce polynôme pour qu'il soit positif ou nul en $e^{\pm i\omega'}$ car si on l'élève au carré ce polynôme à coefficients réels cela n'entraine pas que $P^2_\omega(e^{i\omega'})$ est réel et positif. Sur le cercle unité on obtient alors le polynôme réciproque :

$$\begin{aligned} |P_\omega(z)|^2 &= P_\omega(z)P_\omega(z^{-1}) \\ &= 2 + 4\cos^2\omega - 4\cos\omega(z + z^{-1}) + (z^2 + z^{-2}) \end{aligned}$$

auquel il faut associer le vecteur :

$$f^T_\omega = [1{+}2\cos^2\omega \quad -4\cos\omega \quad 1 \quad 0 \quad 0 \ldots 0\ ]$$

pour avoir : $f^T_\omega s(\omega) = 0$ et $f^T_\omega s(\omega') > 0\ \ \forall\omega' \neq \omega$ et donc le vecteur :

$$d^T_\omega = [-2\cos^2\omega \ \ 4\cos\omega \quad -1 \quad 0 \quad 0 \quad \ldots 0]$$

qui vérifie (4) dans le cas où $\|x_o\|_0 = 1$. S'il s'agit d'isoler deux ou davantage de colonnes, on procède de le même façon pour construire d'abord le polynôme réciproque de degré double, puis le vecteur f_ω et enfin le vecteur d_ω associé. Le nombre maximal de colonnes que l'on peut ainsi isoler est à nouveau $\lfloor \frac{n-1}{2} \rfloor$.

Ce résultat également n'est pas surprenant et se déduit d'un théorème de Caratheodory [9] qui dit que si on se donne n-1 nombres complexes $\{b_1,\ b_2,\ \ldots, b_{n-1}\}$, non tous nuls, alors il existe une décomposition unique :

$$b_j = \sum_1^P \ \alpha_p\ e^{ij\omega_p}, \qquad j = 1, \ldots, n-1$$

avec $\alpha_p > 0$, $\omega_p \neq \omega_k \in]-\pi, \pi]$ et $P \leq n-1$.

En particularisant ce théorème au cas où les b_j sont réels, on observe qu'à $\omega_p \neq 0$ ou π, est alors associé un $\omega_{p'} = -\omega_p$ avec $\alpha_{p'} = \alpha_p$. Dans notre contexte, ce théorème établit donc que si : $b = \sum_1^q x_\ell s_{\omega_\ell}$ avec les ω_ℓ sur la grille de discrétisation et $q \leq \lfloor \frac{n-1}{2} \rfloor$ alors cette décomposition est unique. Nous revenons sur cette observation dans le paragraphe suivant.

La théorie de la réalisation s'applique aussi à ce cas mais on obtient un seuil sur $\|x_o\|_o$ deux fois plus bas car chaque sinusoïde induit une rang 2 dans la matrice de Hankel. Par contre le signe des pondérations x_i n'est pas contraint dans le cadre de cette théorie.

6 Remarque sur l'unicité de la solution

On peut observer que la première ligne de la matrice A dans (PL) vaut $\mathbf{1}^T$ aussi bien dans le cas des colonnes $v(\alpha)$ que des colonnes $s(\omega)$. Cela entraine que la première ligne de $Ax = b$ s'écrit $\mathbf{1}^T x = b(1)$, pour tout x admissible. Mais comme le critère de (PL) est précisément $\mathbf{1}^T x$ cela signifie que tous les points admissibles ont le même coût qui est donc aussi le coût minimal.

Cela signifie aussi que la condition d'unicité (5) de l'optimum est en fait une condition d'unicité des points admissibles : sous la condition (5) l'ensemble des points admissibles de (PL) se réduit à un seul point, le point x_o.

Si le point x_o est le seul point admissible de (PL) le critère d'optimisation ne joue aucun rôle, il est optimal non seulement s'il s'agit de minimiser la norme ℓ_1 mais également pour tout autre critère linéaire de la forme $c^T x$.

Comme normaliser les colonnes de la matrice A revient à transformer le critère $\mathbf{1}^T x$ en un critère linéaire de la forme $c^T x$, on a démontré que le même résultat est valide si on remplace (PL) par :

$$\min_x \mathbf{1}^T x \quad \text{sous :} \quad \Delta x = b \quad \text{et} \quad x \geq 0 \qquad \text{(PLn)}$$

où Δ est la matrice A dont les vecteurs colonne ont été normalisées.

En résumé, la démonstration de l'unicité de l'optimum de (PL) obtenue en construisant un vecteur d vérifiant (4) et l'observation que pour le programme linéaire initial considéré tous les points admissibles ont le même coût, impliquent l'unicité du point admissible ce qui reste valide si on passe de (PL) à (Pln).

On peut arriver à la même conclusion en observant que le vecteur f que nous avons introduit dans les paragraphes précédents vérifie le théorème suivant :

Théorème : La solution x_o du système : $Ax = b, \quad x \geq 0,$ est unique si :

$$\exists f \ \ni \qquad \bar{A}_0^T f = 0 \quad \text{et} \quad a_j^T f > 0 \ \ \forall a_j \notin \bar{A}_0 \qquad (6)$$

où $\bar{A}_0$ est la matrice à nombre de colonnes minimal vérifiant $Ax_0 = \bar{A}_0\bar{x}_0 = b$. $\qquad \triangle$

Ce théorème, qui ne s'applique que si la solution x_0 est dégénérée (car sinon $f = 0$), est une conséquence immédiate du lemme suivant que nous allons démontrer :

Lemme : La variable x_j est nulle en tout point de l'ensemble, supposé non vide, des solutions du système : $\{Ax =$

$b, \quad x \geq 0\}$ si et seulement si $\exists f \quad \ni \quad b^T f = 0, \quad A^T f \geq 0, \quad a_j^T f > 0$.

Démonstration du lemme: Pour démontrer que la condition est nécessaire, on introduit le programme linéaire: $\max x_j = \min -e_j^T x$ sous $Ax = b, \quad x \geq 0$ dont le dual s'écrit: $\max b^T d$ sous $A^T d \leq -e_j$.

Par hypothèse, le primal admet un optimum x_0 avec $x_{0,\ j} = 0$, le dual admet alors un optimum d_0 qui vérifie $b^T d_0 = 0$ et $A^T d_0 \leq 0$ et $a_j^T d_0 \leq -1$ et il suffit de prendre $f = -d_0$ pour établir le résultat.

La condition est suffisante car elle entraine que $f^T Ax = f^T b = 0$. Mais $f^T Ax = 0$ avec $f^T A \geq 0$, $x \geq 0$ implique que $x_j = 0$ dès que $f^T a_j > 0$. $\quad \triangle$.

On peut rapprocher ce résultat du lemme de Farkas qui dit que le système:
$Ax = b, \quad x \geq 0,$ a une solution si et seulement si $b^T d \leq 0$ pour tout d satisfaisant $A^T d \leq 0$.

Du coté du programme linéaire dual (PLD), on peut de façon similaire observer que $d^* = e_1$, la première colonne de la matrice identité, est toujours une solution optimale du dual (PLD): c'est un point admissible du dual qui réalise le même coût que le primal puisque $b^T e_1 = b(1) = \mathbf{1}^T x$.

Que $d^* = e_1$ soit un optimum du dual qui en a d'autres, signifie que seule une modification de $b(1)$ influence le coût optimal primal et que la modification d'une autre composante de b changera la solution mais pas son coût.

7 Commentaire

Il ne semble pas possible d'étendre ces résultats au cas où la matrice A est obtenue par concaténation des deux matrices considérées ici. Il n'existe pas de polynôme qui soit positif ou nul sur toutes les colonnes et ne s'annulent que sur la sélection désirée.

Cette limitation n'existe pas du tout du coté théorie de la réalisation car il est parfaitement possible de "réaliser" une réponse impulsionnelle associée à un système linéaire dynamique constitué par la mise en parallèle d'une part de systèmes du premier ordre (les colonnes de la matrice de Vandermonde) et d'autre part d'oscillateurs (les colonnes de la matrice de Fourier réelle). Le seuil sur le nombre maximal est cependant moins favorable car pour la matrice de Hankel un oscillateur est de rang deux.

8 Conclusions

Comme nous l'avons indiqué, retrouver la décomposition minimale dont on sait qu'elle existe dans les deux cas considérés est une opération que l'on savait faire soit en passant par la théorie de la réalisation et en construisant une matrice de Hankel dans le cas de la matrice de Vandermonde, soit en passant par un théorème de Carathéodory et en construisant une matrice de Toeplitz dans le cas de la matrice de Fourier réelle, l'originalité ici est d'avoir montré que cela était également possible en résolvant un programme linéaire. Les opérations réalisées dans les cas "Toeplitz" ou "Hankel" : décomposition en valeurs et vecteurs propres, recherche de racines d'un polynôme, sont évidemment de nature plus complexes que celles requises dans la résolution d'un programme linéaire. Cela explique la différence de performances: restriction des signes et impossibilité de concaténer les deux bases. Sans parler du fait que pour les autres algorithmes il n'est pas nécessaire de discrétiser le paramètre d'intérêt pour se ramener à un nombre fini m de vecteurs de base.

Références

[1] M. Nafie, A.H. Tewfik and M. Ali. Deterministic and iterative solutions to subset selection problems. *IEEE Trans. on S.P.*, 50, 7, 1591-1601, 2002.

[2] D.L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. on I.T.*, 47, 11, 2845-2862, 2001.

[3] M. Elad an A.M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. on I.T.*, 48, 9, 2558-2567, 2002.

[4] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *Internal Report INRIA*, 1499, nov. 2002.

[5] J.J. Fuchs. On sparse representations in arbitrary redundant bases. soumis à *IEEE Trans. on I.T.*, déc. 2002.

[6] J.J. Fuchs. Une approche à l'estimation et l'identification simultanées. *XVI Colloque GRETSI*, vol. 2, pp. 1273-1276, Grenoble, 1997.

[7] D. G. Luenberger. Introduction to linear and nonlinear programming. *Addison Wesley*, 1973.

[8] T. Kailath. Linear systems. *Prentice Hall*, 1980.

[9] U. Grenander and G. Szego. Toeplitz forms and their applications. *Berkeley, Univ. Calif. Press*, 1958.

[10] N. Moal and J.J. Fuchs. Sinusoids in white noise: a quadratic programming approach. *IEEE ICASSP*, vol. IV, 2221-2224, Seattle, 1998.