

Une approche à l'estimation et l'identification simultanées

Jean-Jacques FUCHS

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

RÉSUMÉ

On propose une nouvelle approche permettant simultanément d'estimer l'ordre et d'identifier les paramètres d'un modèle paramétrique. Elle s'applique notamment quand les observations peuvent être modélisées comme la somme d'un nombre inconnu de signaux appartenant à une famille paramétrée connue et d'un bruit blanc. L'idée est de trouver une décomposition parcimonieuse de l'observation sur la famille que l'on a discrétisée par rapport au(x) paramètre(s). La décomposition est non unique et la sélection d'une représentation utilisant un petit nombre de composantes est assurée par la minimisation de la norme ℓ_1 des pondérations. L'algorithme sous-jacent a été appliqué à différents contextes, on propose ici une analyse de ses résultats.

ABSTRACT

The problem of fitting a model composed of a number of superimposed signals to noisy observations is considered. We do not assume that the number of signals is known *a priori* and perform simultaneously the detection of the number of components and the estimation of their parameters. The computations requirements are those of a quadratic programming algorithm. The scheme can also be seen as a decomposition approach using a maximum likelihood criterion with an ℓ_1 additive regularization term. We present here a tentative analysis of the performances of the algorithm that has been successfully applied to different classes of problems [3], [5].

1 Introduction

Dans de nombreux domaines, on est amené à représenter un vecteur d'observation bruité comme la somme d'un nombre inconnu de signaux appartenant à une famille connue. Parmi les problèmes [1-6] pouvant être modélisés de cette façon, on peut citer : la décomposition d'une somme d'un petit nombre de sinusoides en ses composantes dont il s'agit d'estimer les paramètres [4], l'estimation de retards [5] associés aux différents trajets en sonar actif ou en radar mais également l'identification de processus ARMA par la méthode des moments [6].

Si le bruit peut être supposé additif et gaussien, on peut résoudre le problème au sens du maximum de vraisemblance (MV). Mais cette approche a deux inconvénients : il faut connaître *a priori* le nombre de signaux présent et, dans des cas difficiles, elle requiert la connaissance d'un point initial proche de l'optimum car le bassin d'attraction du maximum global est alors étroit. Les autres approches proposées utilisant des algorithmes itératifs du type gradient [1], ou du type EM ou relevant de la programmation dynamique [2] souffrent, en général, des deux mêmes désavantages.

Nous proposons une approche qui traite simultanément les deux problèmes de l'estimation du nombre de signaux et de l'identification de leurs paramètres. Les approches bayésiennes [4] développées dans le cadre des méthodes de déconvolution impulsionnelle abordent le même type de problème.

2 Modélisation

On considère des modèles de la forme :

$$\hat{b} = b_0 + e = \sum_{i=1}^{q_0} \alpha_i a(\theta_i) + e$$

où \hat{b} , b_0 , $a(\theta)$, e sont des vecteurs de dimension N , \hat{b} est le vecteur d'observations bruité, e est le vecteur du bruit additif supposé centré et de matrice de covariance $\sigma_e^2 I$. L'objectif est de retrouver la partie déterministe b_0 de \hat{b} : le nombre q_0 de composantes et les $\{\alpha_i, \theta_i\}$,

La dépendance de $a(\theta)$ en fonction de θ est connue et non-linéaire, θ peut être un scalaire (retard) ou un vecteur (fréquence et phase initiale). La difficulté du problème dépend notamment du rapport signal à bruit (RSB), de la *proximité* des composantes (fréquences ou retards *proches*) et de la dynamique des amplitudes α_i .

En supposant que la norme euclidienne de $a(\theta)$ est indépendante de θ et normalisée un, nous allons définir un RSB global comme étant égal à $a(\theta)^T a(\theta) / \sigma_e^2 = 1 / \sigma_e^2$. Pour définir le RSB de chaque composante il faut également tenir compte de son amplitude α_i propre.

Dans ce contexte, avec un nombre N de composantes du vecteur d'observation constant, dire que l'on atteint asymptotiquement la borne de Cramer-Rao signifie qu'on l'atteint quand le RSB global tend vers l'infini. Il suffit, pour s'en convaincre, de penser à \hat{b} comme étant un vecteur lui-même estimé à partir de T observations, e est alors le *bruit d'estimation*, dont la variance σ_e^2 tend vers zéro quand T tend vers l'infini et on retrouve bien l'asymptote usuelle.

Si le bruit e est gaussien, l'estimation des paramètres α_i, θ_i par une approche du maximum de vraisemblance (MV)

consiste à résoudre, pour un ordre q fixé :

$$\min_{\alpha, \theta} \|\hat{b} - \sum_{i=1}^q \alpha_i a(\theta_i)\|_2^2 \quad MV(q)$$

Cette approche nécessite donc la connaissance de q le nombre de composantes et, dans les cas difficiles, d'un point initial très proche de l'optimum. Pour $q > q_0$, de fausses composantes modélisant essentiellement le bruit apparaissent. On peut vérifier que l'amplitude α associée à une telle fausse composante est de l'ordre de σ_e et que la décroissance qu'elle induit sur f est de l'ordre de σ_e^2 .

3 Développement

L'objectif ultime de notre approche est de faire aussi bien que le MV mais sans connaître *a priori* ni l'ordre exact q_0 ni un point initial proche de l'optimum. Transformons peu à peu le problème d'optimisation précédent. Une approche voisine de celle du MV consiste à minimiser :

$$\min_{\alpha, \theta} \|\hat{b} - \sum_{i=1}^q \alpha_i a(\theta_i)\|_2^2 + \lambda \|\underline{\alpha}\|_1 \quad MV B(q, \lambda)$$

On a introduit un terme de régularisation dans la fonctionnelle du MV qui fait intervenir la norme ℓ_1 du vecteur $\underline{\alpha}$ de dimension q contenant les pondérations α_i . On peut donner diverses interprétations à ce terme. La pondération λ qui apparaît dans $MVB(q, \lambda)$ peut notamment s'interpréter comme un multiplicateur de Lagrange du problème d'optimisation suivant :

$$\min_{\alpha, \theta} \|\hat{b} - \sum_{i=1}^q \alpha_i a(\theta_i)\|_2^2 \text{ sous } \|\underline{\alpha}\|_1 \leq B$$

En effet, le lagrangien de ce problème est $MVB(q, \lambda)$. Son optimum est atteint au point selle du Lagrangien, pour une valeur $\lambda^*(B)$ du multiplicateur de Lagrange dépendant de B . À un λ dans $MVB(q, \lambda)$ correspond donc un B dans cette dernière formulation et *vice versa*. On peut noter que pour B suffisamment grand, la contrainte n'est plus active et on est ramené à $MV(q)$.

$MVB(q, \lambda)$ n'est pas plus facile à résoudre que $MV(q)$ et souffre des mêmes inconvénients indiqués plus haut. Mais il est possible de le remplacer par une version approchée, approximée dont la solution est facile à obtenir. Pour simplifier l'exposé, plaçons nous dans le cas où le paramètre θ est un scalaire (par exemple, un retard entre les différentes répliques d'un même signal). Nous discrétisons alors θ et considérons qu'il ne peut plus prendre que M valeurs θ_m . La valeur de M dépend du pas de discrétisation qui est à choisir en fonction des performances attendues et donc du RSB. Aucune dépendance n'existe entre M et N . Nous posons $a_m = a(\theta_m)$ et notons A la matrice de dimension (N, M) ayant pour colonnes les vecteurs a_m . Pour M suffisamment grand et donc un pas de discrétisation suffisamment petit, une composante $a(\theta_i)$ arbitraire sera très bien approximée par une interpolation linéaire de ses deux colonnes a_m voisines. La reconstruction de la partie utile b_0 de \hat{b} nécessite alors essentiellement $2q_0$ colonnes.

Dans un premier temps, on peut alors considérer le problème d'optimisation suivant :

$$\min_{\beta_j, m_j} \|\hat{b} - \sum_{j=1}^{2q} \beta_j a_{m_j}\|_2^2 + \lambda \|\underline{\beta}\|_1$$

où la recherche en θ est restreinte aux points d'une grille et dont la solution est proche de celle de $MVB(q, \lambda)$ pour un λ bien choisi. Ce problème est toujours difficile à résoudre et une recherche exhaustive sur la grille de valeur de θ_i peu réaliste. Il souffre donc des mêmes inconvénients. Finalement, nous le remplaçons par :

$$\min_X \|\hat{b} - AX\|_2^2 + \lambda \|X\|_1 \quad P1(\lambda)$$

où la minimisation ne porte plus que sur les M pondérations x_k du vecteur X . Le nombre de composantes requises, l'ordre du modèle, a disparu et est laissé libre. La différence entre $P1(\lambda)$ et les formulations précédentes est importante.

L'optimum X^* de ce problème est unique et facile à obtenir par des algorithmes standards de programmation quadratique. Il s'agit maintenant de montrer que pour un choix judicieux de λ cette solution X^* a un petit nombre de composantes non nulles (environ $2q$ avec $q > q_0$), que ces composantes sont groupés principalement en q paquets de deux. L'amplitude α à associer à un tel paquet est la somme des composantes et la valeur de θ le barycentre des θ_m correspondants.

En utilisant les ordres de grandeurs relatifs à l'amplitude et à la diminution de coût associées à une fausse composante, nous proposons de prendre $\lambda = O(\sigma_e)$. Ce choix basé sur l'interprétation de λ comme multiplicateur de Lagrange semble effectivement convenir dans la pratique [5].

4 Analyse

4.1 Généralités

Nous allons essayer d'évaluer les performances de la solution X^* de $P1(\lambda)$ pour un λ bien choisi. Il est prudent de commencer par étudier le problème en absence de bruit. Il n'est pas trivial puisque une première difficulté est déjà de retrouver l'ordre exact q_0 du modèle. En absence de bruit, la valeur suggérée plus haut pour λ est nulle et le vecteur \hat{b} devient b_0 . On peut montrer qu'alors, $P1(\lambda)$ est génériquement équivalent à :

$$\min_X \|X\|_1 \text{ sous } AX = b_0 \quad PL$$

PL est un programme linéaire, il admet, comme $P1(\lambda)$, un optimum global unique qui est atteint en un point, noté X_0 dans la suite, possédant au plus N composantes non nulles : on reconstruit b_0 à l'aide de N colonnes bien choisies de A . De façon plus précise, on peut établir qu'il existe λ_1 (petit) tel que pour $\lambda \in]0, \lambda_1[$, $P1(\lambda)$ et PL ont "même solution" si on convient de dire qu'ils ont "même solution" quand les colonnes de A utilisées sont les mêmes.

Comme pour un RSB non nul, notre approche va au mieux, fournir une estimée X^* au voisinage de X_0 la solution de PL , il est important d'analyser, dans chaque cas particulier (pour chaque classe de modèles), l'intérêt de la solution de

PL. Il semble même qu'évaluer ce qui se passe pour un RSB *raisonnable* relève ensuite de l'expérimentation. Le maximum de vraisemblance a, asymptotiquement de bonnes performances mais savoir *a priori* où commence l'asymptote est en général impossible et relève également de l'expérimentation.

4.2 PL retrouve la bonne solution

Nous commençons par étudier sous quelles conditions sur les colonnes de la matrice A et donc sur la famille paramétrée $a(\theta)$, PL retrouve la bonne solution [3]. On suppose que le second membre b_o est construit à l'aide de $q_o < N$ (et même $q_o \ll N$) colonnes de A . On pose $b_o = AX_o = \bar{A}\bar{X}_o$. De façon générale, on note avec des $\bar{\quad}$ les restrictions aux composantes non nulles, \bar{A} est ainsi la matrice (N, q_o) qui contient les bonnes colonnes de A . On cherche sous quelles conditions, PL retrouve cette solution appelée dégénérée car $q_o < N$.

Le dual de PL est : $\max_d d^T b_o$, sous $\|d^T A\|_\infty \leq 1$ sans contrainte de signe sur les variables duales d .

Pour que la solution exacte X_o , caractérisée par \bar{A} , soit l'optimum unique de PL, il faut et il suffit de savoir lui associer un vecteur dual d tel que le couple (X_o, d) soit optimal. La condition nécessaire et suffisante s'écrit :

$$(C) \quad \exists d \ni \bar{A}^T d = \text{sgn} \bar{X}_o \\ \text{et } |a_j^T d| < 1 \quad \forall a_j \notin \bar{A}.$$

On constate que dans le cas dégénéré qui est celui qui nous intéresse, d n'est pas défini de façon unique. Par ailleurs, l'inégalité stricte pour les colonnes n'appartenant pas à \bar{A} est là pour garantir que l'optimum X_o est unique.

La condition trouvée dit qu'il doit exister un hyperplan séparateur (strict) entre les bonnes colonnes et les fausses. On peut noter qu'elle ne fait pas intervenir les amplitudes associées aux différentes colonnes.

Pour un problème (une famille $a(\theta)$) donné et pour un scénario fixé, on peut alors vérifier si un vecteur d existe. En posant $A = [\bar{A}N]$, le "meilleur" d est celui qui réalise : $\min_d \|N^T d\|_\infty$ sous $\bar{A}^T d = \text{sgn} \bar{X}_o$. On peut, plus simplement, vérifier si $\min_d \|N^T d\|_2$ sous les mêmes contraintes, satisfait $\|N^T d\|_\infty < 1$.

Quand un tel d existe, c'est la différence $1 - \|N^T d\|_\infty$ qui joue un rôle crucial dans la suite. C'est elle qui indique si un scénario est facile ou difficile et jusqu'à quel point l'approche est robuste au bruit ou aux erreurs d'interpolation. Elle est fonction de la difficulté du scénario mais également du pas de discrétisation en θ , que nous notons h , utilisé lors de la construction de A . Globalement, elle décroît avec h .

Il est évidemment impossible de vérifier si la condition (C) est satisfaite pour un scénario donné puisque le scénario est inconnu. Mais il apparaît que dans la pratique pour les familles $a(\theta)$ évoquées dans l'introduction, elle est très généralement satisfaite même pour des scénarios difficiles. Dans le cas où b est une somme de sinusoides, le théorème de Carathéodory (celui utilisé pour justifier la méthode de Pisarenko) établit que la condition est toujours satisfaite. La question est ouverte pour d'autres familles. Nous supposons en général que cette condition est satisfaite.

4.3 L'optimum de P1(λ)

Nous indiquons maintenant les conditions que satisfait X^* l'optimum de P1(λ) :

$$\min_X \|\hat{b} - AX\|_2^2 + \lambda \|X\|_1 \quad P1(\lambda)$$

Il s'agit du minimum d'une fonction convexe mais non continûment différentiable. Une CNS pour que X^* soit l'optimum global est que le vecteur 0 soit un sous gradient de la fonction en X^* . Le vecteur y est un sous gradient de f en X^* si $f(X) \geq f(X^*) + y^T(X - X^*)$. La condition revient à écrire que la fonction ne diminue pas quand on s'écarte de X^* . Cette condition s'écrit ¹ :

$$\|A^T(\hat{b} - AX^*)\|_\infty \leq \lambda/2$$

De façon plus détaillée, pour les p composantes non nulles de X^* la fonction est continûment différentiable et le sous gradient se réduit au gradient qui doit donc être nul. En remplaçant $\|X\|_1$ par $X^T \text{sgn} X$ et en notant, comme plus haut, par des $\bar{\quad}$ les composantes non nulles, on trouve :

$$(C1) \quad -2\bar{A}^T(\hat{b} - \bar{A}\bar{X}^*) + \lambda \text{sgn} \bar{X}^* = 0$$

que l'on peut réécrire :

$$\Rightarrow (\bar{A}^T \bar{A})\bar{X}^* = \bar{A}^T \hat{b} - \lambda/2 \text{sgn} \bar{X}^*$$

\bar{X}^* est donc solution d'un système linéaire (implicite car $\text{sgn} \bar{X}^*$ apparaît au second membre) dont la matrice est $(\bar{A}^T \bar{A})$. Comme la fonction à optimiser est convexe, elle admet toujours un optimum et ce système a une solution même si cette matrice n'est pas inversible ($p > N$). Dans ce cas, on peut se ramener à un système inversible d'ordre N et donc réduire le nombre de composantes de \bar{X}^* à N au maximum. Les composantes non nulles de l'optimum notées \bar{X}^* sont données par :

$$\bar{X}^* = (\bar{A}^T \bar{A})^{-1} \bar{A}^T \hat{b} - \lambda/2 (\bar{A}^T \bar{A})^{-1} \text{sgn} \bar{X}^*$$

Le second terme peut être considéré comme un biais dû au terme de régularisation qui est facile à corriger. Ce biais fait bien diminuer la norme ℓ_1 de \bar{X}^* puisque $\text{sgn} \bar{X}^{*T} (\bar{A}^T \bar{A})^{-1} \text{sgn} \bar{X}^* > 0$.

Pour les composantes nulles de X^* , le vecteur 0 est un sous gradient si on y perd à les prendre différentes de zéro. La dérivée partielle du le terme quadratique doit donc être inférieure à λ en valeur absolue :

$$(C2) \quad |2a_j^T(\hat{b} - \bar{A}\bar{X}^*)| < \lambda \quad \forall a_j \notin \bar{A}$$

En regroupant l'ensemble de ces conditions, on a :

$$\|A^T(\hat{b} - \bar{A}\bar{X}^*)\|_\infty \leq \lambda/2$$

avec "égalité" pour les colonnes de A actives et "inégalité" pour les autres. Ces conditions ont une interprétation intéressante. Elles indiquent qu'après avoir soustrait $\bar{A}\bar{X}^*$ à l'observation \hat{b} , la sortie du (d'un nouveau) filtre adapté est saturée en $\lambda/2$ pour les colonnes qui contribuent à expliquer la solution et reste inférieure à $\lambda/2$ pour les autres.

Ce sont ces conditions (CNS) qui permettent d'établir que pour λ suffisamment petit PL et P1(λ) ont génériquement "même" solution (au terme de biais près). En fait, comme nous

¹P1(λ) est équivalent à $\min_X \|AX\|_2^2$ sous $\|A^T(AX - b)\|_\infty \leq \frac{\lambda}{2}$.

nous intéressons exclusivement à des cas dégénérés ($p < N$), PL et P1(λ) ont "même" solution sous une condition plus forte que (C), à savoir :

$$(C') \quad \begin{aligned} |a_j^T d^\parallel| < 1 \quad \forall a_j \notin \bar{A} \\ \text{ou } d^\parallel = \bar{A}(\bar{A}^T \bar{A})^{-1} \text{sgn} \bar{X}_o \end{aligned}$$

4.4 Cas d'une composante unique

On considère le cas où $q_o = 1$, en absence de bruit on a alors $\hat{b} = a_o$ une colonne de A . La condition (C') est alors toujours satisfaite en prenant $d = d^\parallel = a_o$ puisque les colonnes de A sont indépendantes et de norme euclidienne égale à un. L'optimum X^* de P1(λ) n'a alors qu'une composante non nulle (la bonne) qui vaut $1 - (\lambda/2)$ pour $\lambda \in]0, \lambda_{max}[$ avec $\lambda_{max} = 2$ et $X^* = 0$ pour $\lambda > 2$.

En présence de bruit e faible ($\sigma_e \ll 1$), le second membre devient $\hat{b} = a_o + e$, PL a génériquement une solution à N composantes non nulles, celle associée à a_o étant plus grande (de l'ordre de 1) que les autres (de l'ordre de σ_e). Pour λ très petit P1(λ) a cette même "solution". On a vu plus haut comment déduire l'optimum X^* de P1(λ) de l'optimum X_o de PL et la solution reste la "même" aussi longtemps qu'aucune nouvelle composante de X^* ne s'annule (en gros $\lambda \leq \sigma_e$). Pour λ plus grand, un certain nombre de composantes de X_* s'annulent et enfin dans une certaine plage de λ seule la composante associée à a_o sera non nulle (en gros pour $\lambda \in [2\sigma_e, 2]$). Pour λ encore plus grand ($\lambda \geq 2 \max_j |a_j^T \hat{b}|$), toutes les composantes de X^* sont nulles.

Si le bruit est plus important et/ou le pas de discrétisation suffisamment petit, deux composantes celle de a_o et d'une colonne voisine vont dominer les autres dans la solution X_o du PL. Il en est de même pour l'optimum de P1(λ) pour une certaine plage de λ , seules ces deux composantes seront non nulles. Cela correspond au fait qu'en présence de bruit, la solution $\hat{\theta}_o$ du MV est non strictement identique à θ_o (associé à a_o) mais est comprise entre θ_o et θ_1 , par exemple.

4.5 Cas de q composantes

On considère le cas où il y a q_o composantes. Si elles sont bien séparées, on est pratiquement ramené au cas précédent car, en général, $|a_j^T a_i| \rightarrow 0$ quand $|i - j|$ devient grand. On suppose pour simplifier que les q_o composantes sont des colonnes de A . On a donc $b_o = \bar{A} \bar{X}_o$. On fait l'hypothèse que la condition (C') est satisfaite. Comme indiqué plus haut, elle doit être satisfaite avec une inégalité stricte qui traduit que l'optimum est unique et robuste aux perturbations.

On va montrer que sous cette hypothèse, pour un bruit suffisamment faible il existe un λ (une plage de λ) pour lequel P1(λ) a exactement cette solution.

En présence de bruit, PL a une solution non dégénérée avec q_o composantes dominant les $N - q_o$ autres, pour λ très petit P1(λ) a cette même solution puis, en augmentant λ , les fausses colonnes disparaissent peu à peu.

En écrivant que l'on souhaite retrouver les bonnes colonnes avec les mêmes signes que dans \bar{X}_o , on obtient l'ex-

pression suivante pour la solution de P1(λ) souhaitée $\bar{X} = (\bar{A}^T \bar{A})^{-1} \{ \bar{A}^T (b_o + e) - (\lambda/2) \text{sgn} \bar{X}_o \}$,

Cette solution satisfait par construction (C1), il reste à montrer qu'elle satisfait (C2) pour λ bien choisi :

$$|a_j^T \{ b + e - \bar{A} \bar{X} \}| \leq \lambda/2 \quad \forall a_j \notin \bar{A}$$

En posant $e = e^\perp + e^\parallel$ avec e^\parallel la composante appartenant à l'image de \bar{A} , on a $\bar{A}(\bar{A}^T \bar{A})^{-1} \bar{A}^T (b + e) = b + e^\parallel$ et il reste à établir que :

$$|a_j^T \{ e^\perp + (\lambda/2) \bar{A}(\bar{A}^T \bar{A})^{-1} \text{sgn} \bar{X}_o \}| \leq \lambda/2$$

$$\Leftrightarrow |a_j^T e^\perp| \leq (\lambda/2)(1 - |a_j^T d^\parallel|)$$

Cette dernière relation est vérifiée pour un bruit suffisamment faible dès que (C') est satisfaite strictement.

5 Conclusions

Cette analyse est incomplète. Il reste notamment à étudier le cas où b_o est constitué de composantes ne figurant pas parmi les colonnes de A . Même en absence de bruit, pour reconstruire une telle composante, PL aura recours à N colonnes de A . Mais pour un pas de discrétisation h donné, mis à part les deux colonnes voisines de cette composante les poids affectés aux $N - 2$ autres colonnes seront petits, en $O(h^2)$ comme l'erreur d'approximation linéaire. On peut alors montrer que pour λ bien choisi, P1(λ) n'aura que deux composantes non nulles. Comme pour le bruit, les colonnes de A nécessaires pour modéliser les erreurs d'interpolation peuvent être gommées par un choix judicieux de λ . Pour compléter cette analyse, il reste également à indiquer comment choisir le pas de discrétisation h à appliquer à θ . Les résultats semblent extrêmement prometteurs notamment pour des estimations de retards [5].

Références

- [1] B.M. Bell and T.E. Ewart. Separating multipaths by global optimization of a multidimensional matched filter. *IEEE-T-ASSP*, 37 :378–391, mar. 1986.
- [2] Y. Bresler and A. Macovski. Exact ML parameter estimation of superimposed exponential signals in noise. *IEEE-T-ASSP*, 34 :1081–1089, oct. 1986.
- [3] J.J. Fuchs. Linear programming in spectral estimation. Application to array processing In *Proc. ICASSP*, VI, pp. 3161–3164, Atlanta, may 1996.
- [4] F. Dublanchet. Contribution de la méthodologie bayésienne à l'analyse spectrale de raies pures et à la goniométrie haute résolution. *Thèse de l'université Paris-Sud Orsay*, oct. 1996.
- [5] J.J. Fuchs. Multipath time-delay estimation. In *Proc. ICASSP*, I, pp. 527–530, Munich, April 1997.
- [6] N. Moal et J.J. Fuchs. Estimation de l'ordre et identification des paramètres d'un processus ARMA *GRETSI*, 1997.