

Estimation adaptative de signaux localement harmoniques

Stéphanie DUBOST, Olivier CAPPÉ

ENST- TSI / CNRS URA 820
46, rue Barrault, 75634 Paris Cedex 13 France
dubost@tsi.enst.fr, cappe@tsi.enst.fr

Résumé – Nous nous intéressons dans cet article à une nouvelle modélisation à temps discret de signaux localement harmoniques. Nous étudions ensuite une méthode d'estimation des paramètres du modèle reposant sur l'approximation locale polynomiale. Après avoir présenté les performances asymptotiques de ces estimateurs, nous proposons une application au débruitage d'un signal de parole voisé.

Abstract – This paper deals with a new discrete time modelisation for locally harmonic signals. We first define the model and present an estimation approach based on local polynomial approximation. We then consider the asymptotic properties of this type of estimators and describe an application to voiced speech enhancement.

1 Introduction

L'analyse et l'estimation de signaux quasi sinusoïdaux lentement modulés en fréquence et en amplitude, ou de superpositions de tels signaux, est une problématique classique du traitement du signal qui possède de nombreuses applications. Une approche bien établie consiste par exemple à utiliser des modèles dits "à phase polynomiale" [3]. Récemment, V. Katkovnic a jeté les bases d'un traitement non paramétrique du problème (voir notamment [4], [8]). Pour certaines applications comme le traitement de la parole, cette approche semble particulièrement appropriée dans la mesure où l'évolution temporelle de certains paramètres (en l'occurrence surtout des amplitudes) ne se prête pas facilement à une modélisation paramétrique [2].

On propose dans cette contribution, un cadre permettant une analyse asymptotique simplifiée pour les modèles localement sinusoïdaux, considérés dans le cadre paramétrique ou non paramétrique. On s'intéresse ensuite au cas de signaux localement harmoniques, en considérant plus spécifiquement l'application de tels modèles au cas de signaux de parole voisés.

2 Modélisation des signaux localement harmoniques

On définit tout d'abord une classe de signaux réels à temps discret localement sinusoïdaux (ou quasi-sinusoïdaux) se prêtant à l'analyse asymptotique, c'est à dire pour laquelle l'augmentation du nombre d'observations permet effectivement de se rapprocher d'une situation analysable de manière simple. Plus précisément, soit :

$$s_{t,T} = a\left(\frac{t}{T}\right) \cos\left(T\phi\left(\frac{t}{T}\right)\right) + b\left(\frac{t}{T}\right) \sin\left(T\phi\left(\frac{t}{T}\right)\right) \quad \text{pour } t = 1, \dots, T \quad (1)$$

où T désigne le nombre d'observations et

- Les amplitudes en phase et quadrature $a(u)$ et $b(u)$ sont des fonctions définies sur l'intervalle $[0, 1]$ que l'on suppose deux fois continuellement différentiables (C^2).
- La phase $\phi(u)$ est une fonction croissante et dérivable sur $[0, 1]$. De plus, sa dérivée notée $\omega(u) \triangleq \dot{\phi}(u)$ est supposée être C^2 et vérifier la condition suivante : $\pi > \omega(u) \geq \omega_{\min} > 0$.

Cette écriture ne correspond ni à l'observation d'une portion croissante d'un signal discret fixé (en ce sens que $s_{1,\dots,T+1,T+1}$ n'est pas obtenu par concaténation de $s_{1,\dots,T,T}$ et de $s_{T+1,T+1}$) ni à l'échantillonnage de plus en plus fin d'une fonction à temps continu fixée comme il est d'usage en estimation fonctionnelle.

Un examen plus détaillée de (1) montre que lorsque T augmente, le signal observé autour du point u_0T ($s_{t,T}$ pour t proche de u_0T avec $u_0 \in [0, 1]$ fixé) correspond bien à un signal localement sinusoïdal avec des variations d'amplitude et de fréquence instantanée de plus en plus faibles. On peut ainsi montrer que

Lemme 1.

$$\left| s_t - 2 \cos\left(\omega\left(\frac{t}{T}\right)\right) s_{t-1} + s_{t-2} \right| \leq \frac{C}{T} \quad (2)$$

la constante C ne dépendant que des fonctions a, b et ω , (mais pas de t).

(2) justifie que $\omega\left(\frac{t}{T}\right)$ puisse être considéré comme la fréquence instantanée du signal s_t lorsque le nombre d'observations devient grand. De même, on peut montrer que $\sqrt{a^2(u_0) + b^2(u_0)}$ correspond asymptotiquement à l'amplitude instantanée du signal $s_{E(u_0T)}$ pour $u_0 \in [0, 1]$.

Comparé à d'autres écritures comme celle utilisée dans [4], (1) présente plusieurs avantages :

- Il s'agit intrinsèquement d'une écriture à temps discret dans laquelle la fréquence instantanée est définie de manière non-ambigüe via (2),

- $a(u)$, $b(u)$ et $\omega(u)$ peuvent être considérées comme des fonctions fixées indépendamment de T (ainsi, dans un contexte paramétrique où, par exemple, $\omega(u) = \omega_0 + \omega_1 u$ le domaine de variation des paramètres ω_0 et ω_1 est fixée une fois pour toutes indépendamment de T - à comparer avec la convention utilisée dans [3])
- Cette écriture permet de fixer le comportement asymptotique des sommes de cosinus (cf. lemme 2 et corollaire 1) ce qui est indispensable pour obtenir une évaluation asymptotique valide des performances d'estimation.

Pour alléger les notations, nous noterons dans la suite s_t au lieu de $s_{t,T}$.

L'extension naturelle de ce modèle aux cas de signaux localement harmoniques est donnée par :

$$s_t = \sum_{k=1}^K \left[a_k \left(\frac{t}{T} \right) \cos(kT\phi\left(\frac{t}{T}\right)) + b_k \left(\frac{t}{T} \right) \sin(kT\phi\left(\frac{t}{T}\right)) \right] \quad (3)$$

On suppose de plus que l'on dispose d'observations bruitées du signal

$$X_t = s_t + B_t \quad (4)$$

où le bruit B_t est un bruit blanc centré de variance σ^2 .

3 Estimation des paramètres

On s'intéresse ici à l'estimation des amplitudes et de la fréquence du signal en un point $u_0 \in [0, 1]$ (c'est à dire des paramètres qui caractérisent le signal observé autour de l'indice temporel $u_0 T$).

Il s'agit d'un problème d'estimation non-paramétrique, que l'on propose d'aborder sous l'angle de la régression polynomiale locale [1]. L'idée est d'approximer le comportement des fonctions inconnues autour du point u_0 par des développements de Taylor. Par exemple, on supposera que pour $u_0 - h \leq \frac{t}{T} \leq u_0 + h$

$$\phi\left(\frac{t}{T}\right) \approx \phi(u_0) + \left(\frac{t}{T} - u_0\right) \omega(u_0) \quad (5)$$

La qualité de l'approximation dépend alors de l'ordre de grandeur des termes négligés, c'est-à-dire des dérivées d'ordre supérieur de ω au point u_0 ainsi que de la "largeur de bande" h sur laquelle l'approximation est utilisée. De même, les fonctions $a_k(u)$ et $b_k(u)$ peuvent être approximées au voisinage du point u_0 par des constantes ($a(u_0)$ et $b(u_0)$ respectivement). L'estimation des paramètres de l'approximation est obtenue par minimisation d'un critère des moindres carrés fenêtrés (de manière à réduire l'influence des points situés au bord de l'intervalle considéré).

Le choix de la "largeur de bande" h de la fenêtre d'analyse influe notablement sur les estimations obtenues. On cherche généralement à fixer h de manière "adaptative" (c'est à dire indépendamment pour chaque valeur de u_0 , de manière à s'adapter aux variations de régularité des fonctions inconnues) en optimisant le compromis biais-variance: quand h diminue, le biais dû à l'approximation

diminue mais la variance augmente, et l'effet inverse est observé lorsque h augmente.

Dans le cas du modèle donné par (3), en utilisant une approximation polynomiale (de degré 1 pour ϕ et de degré 0 pour les amplitudes a_k, b_k) et en notant

$$\begin{cases} \alpha_k(u) = a_k(u) \cos(T\phi(u_0)) + b_k(u) \sin(T\phi(u_0)) \\ \beta_k(u) = -a_k(u) \sin(T\phi(u_0)) + b_k(u) \cos(T\phi(u_0)) \end{cases}$$

le critère à minimiser s'écrit

$$\sum_{t=1}^T K_h\left(\frac{t}{T} - u_0\right) \left[Y_t - \sum_{k=1}^K (\alpha_k \cos(k(t - u_0 T)\Omega) + \beta_k \sin(k(t - u_0 T)\Omega)) \right] \quad (6)$$

où

$$\begin{cases} \Omega = \omega(u_0) \\ \alpha_k = \alpha_k(u_0) \\ \beta_k = \beta_k(u_0) \end{cases}$$

K_h est la fenêtre (appelée "noyau" dans la littérature sur l'estimation non-paramétrique), obtenue par dilatation d'un noyau prototype $K(t)$ ($K_h(t) = K(t/h)$) de support $[-1, 1]$.

On note que si la fonction ϕ est inconnue, les paramètres α_k et β_k obtenus par minimisation de (6) ne donneront les amplitudes a_k et b_k qu'à une rotation près: dans le modèle (3), considéré en un point donné u_0 seules la fréquence fondamentale $\omega(u_0)$ et ses dérivées ainsi que les amplitudes $\sqrt{a_k(u_0)^2 + b_k(u_0)^2}$ et leurs dérivées peuvent être estimées.

4 Analyse des performances

Nous étudions ici les propriétés asymptotiques des estimateurs basés sur l'approximation polynomiale locale sous les conditions classiques

$$\begin{aligned} h &\rightarrow 0 \\ hT &\rightarrow +\infty \end{aligned} \quad (7)$$

La première condition stipule que la largeur (absolue) de la fenêtre d'analyse tend vers 0 ce qui garantit que le biais décroît asymptotiquement, tandis que la seconde condition (nombre d'observations dans la fenêtre tendant vers l'infini) implique de même une diminution asymptotique de la variance.

On notera dans ce qui suit μ_i (respectivement ν_i) les moments d'ordre i de la fenêtre de pondération (resp. du carré de la fenêtre de pondération)

$$\mu_i = \int_{-1}^1 K(u) u^i du \quad \text{et} \quad \nu_i = \int_{-1}^1 K^2(u) u^i du$$

4.1 Résultats préliminaires

Les résultats qui suivent reposent essentiellement sur le comportement des sommes de cosinus sous les hypothèses (7) dont on montre que:

Lemme 2. *Soit ϕ une fonction dérivable dont la dérivée notée ω est lipschitzienne, et telle que pour tout $u \in [0, 1]$,*

$\omega(u) > \omega_{min} > 0$.

Soit ν une constante positive : $\nu > 0$.

Alors

$$\frac{1}{T^\nu} \sum_{t=1}^T \cos(T\phi(\frac{t}{T})) \xrightarrow{T \rightarrow \infty} 0$$

Corollaire 1. On suppose les conditions du lemme 2 vérifiées.

Soit w une fonction lipschitzienne sur $[0, 1]$

Alors pour tout entier $r \geq 1$ et ν constante positive :

$$\frac{1}{T^{r+\nu}} \sum_{t=1}^T w(\frac{t}{T}) t^r \cos(T\phi(\frac{t}{T})) \xrightarrow{T \rightarrow \infty} 0$$

Il est à noter que ces résultats permettent de borner supérieurement les sommes de cosinus de manière beaucoup plus fine (et beaucoup plus simplement) que ceux portant sur les *sommes de Weyl* [3], et ce grâce à l'écriture adoptée en (1). Par ailleurs, l'hypothèse $\omega(u) > \omega_{min} > 0$ du lemme 2 est indispensable (pour $\omega(u) = \gamma u$ par exemple les sommes de cosinus ne tendent vers 0 que si $\nu > 1/2$).

4.2 Cas paramétrique

On suppose ici dans un premier temps que les amplitudes et la fréquence sont constantes sur toute la fenêtre d'analyse. Le vecteur de paramètres à estimer est alors $\theta_g = (a_1, b_1, \dots, a_K, b_K, \omega)^T$ que l'on notera $\theta_g = ((\theta_a^T, \omega)^T)^T$ avec $\theta_a = (a_1, b_1, \dots, a_K, b_K)^T$. L'estimateur $\hat{\theta}_g$ de θ_g obtenu par minimisation du critère (6) vérifie alors :

- $\hat{\theta}_g$ est fortement consistant, i.e. $\hat{\theta}_g \xrightarrow{p.s.} \theta_g$.
- $H(\hat{\theta}_g - \theta_g) \rightarrow \mathcal{N}(0, 2\sigma^2 \Sigma)$
où les matrices H et Σ sont les matrices diagonales de taille $(2K+1, 2K+1)$ suivantes :

$$H = \text{diag}((hT)^{1/2}, (hT)^{1/2}, \dots, (hT)^{1/2}, (hT)^{3/2})$$

$$\Sigma = \text{diag}(\frac{\nu_0}{\mu_0^2}, \frac{\nu_0}{\mu_0^2}, \dots, \frac{\nu_0}{\mu_0^2}, \frac{\nu_2}{\mu_0^2 \sum_k k^2 (a_k^2 + b_k^2)})$$

On constate que l'ordre de convergence des estimateurs des amplitudes est beaucoup plus faible que celui de l'estimateur de la fréquence ($(hT)^{-1/2}$ contre $(hT)^{-3/2}$) et que de plus la variance asymptotique de l'estimateur des amplitudes est exactement la même que dans le cas où la fréquence est supposée connue [7].

En considérant des expansions d'ordre supérieur pour les variations d'amplitude et de fréquence fondamentale, on obtiendrait des résultats analogues grâce au lemme 2.

4.3 Cas non paramétrique

Les travaux de Katkovnik [8] suggèrent que dans le cas non-paramétrique la conclusion concernant le comportement asymptotique des estimateurs d'amplitudes reste valide : il est le même, que la fréquence fondamentale soit ou non supposée connue. Il resterait toutefois à vérifier ce résultat dans le contexte du modèle défini en (1).

Dans la suite on considère l'évolution de phase $\phi(u)$ connue exactement au voisinage de u_0 . Notons que dans ce cas les amplitudes sont alors estimées exactement, et

non plus "à une rotation près". Le critère à minimiser se réécrit donc simplement

$$\sum_{t=1}^T K_h(\frac{t}{T} - u_0) [Y_t - \sum_{k=1}^K (a_k \cos(kT\phi(\frac{t}{T})) + b_k \sin(kT\phi(\frac{t}{T})))] \quad (8)$$

L'estimation non paramétrique des amplitudes $\theta_a = (a_1(u_0), b_1(u_0), \dots, a_K(u_0), b_K(u_0))^T$ obtenue par minimisation du critère (6) vérifie :

- La variance des estimateurs des paramètres $a_k(u_0)$ et $b_k(u_0)$ est asymptotiquement équivalente à

$$\frac{2\sigma^2 \nu_0}{hT \mu_0^2}$$

- Le biais de l'estimateur $\hat{a}_k(u_0)$ (respectivement $\hat{b}_k(u_0)$) est asymptotiquement équivalent à

$$\frac{\mu_2}{\mu_0} \ddot{a}_k(u_0) h^2 \quad (\text{respectivement } \frac{\mu_2}{\mu_0} \ddot{b}_k(u_0) h^2)$$

en supposant qu'en plus des hypothèses énoncées en (7), on a $Th^2 \rightarrow +\infty$.

Une remarque importante est qu'à un facteur 2 près sur la variance asymptotique, ces résultats sont les mêmes que ceux que l'on obtiendrait si l'on observait directement les fonctions amplitudes isolément dans un bruit de même variance σ^2 [1]. Outre de souligner que dans le modèle (1) la modulation par la quasi-sinusoïde est quasiment "transparente" d'un point de vue statistique, ce résultat suggère que dans le modèle quasi-harmonique décrit par (3), il est légitime de traiter les harmoniques indépendamment les unes des autres (comme s'il s'agissait exactement de régresseurs orthogonaux).

5 Application

Les modèles harmoniques (ou plus généralement, sinusoïdaux) constituent désormais un outil standard pour le traitement de signaux de parole. Cette approche a notamment permis des avancées significatives dans les domaines du codage et de la synthèse de parole [6, 2].

Afin d'illustrer la pertinence de l'approche non paramétrique détaillée précédemment, on considère dans cette partie l'analyse d'un segment de parole. Une remarque importante est que les harmoniques du signal de parole étant de puissances très différentes (en particuliers les premières harmoniques sont de puissance beaucoup plus grande que les dernières), l'ajout de bruit n'a pas le même effet sur toutes les harmoniques.

La fréquence fondamentale est estimée au préalable, de manière usuelle (estimateur du maximum de vraisemblance en supposant un modèle harmonique exact) sur des fenêtres d'une trentaine de millisecondes. L'évolution de phase est alors approximée par la somme cumulée des fréquences estimées

$$T\phi(\frac{t}{T}) \simeq \sum_{i=1}^t \hat{\omega}_{\frac{i}{T}}$$

($\phi(0)$ étant arbitrairement fixé à 0).

Les amplitudes des harmoniques sont alors estimées par minimisation du critère de régression (8). Dans une optique de simplification, on pourrait également estimer ces amplitudes individuellement (par transformée de Fourier) dans la mesure où les résultats exposés au paragraphe 4.3 suggèrent que les régresseurs restent quasi-orthogonaux bien que n'étant plus exactement des sinusoides.

Comme nous l'avons signalé dans la partie 3, le choix de la largeur h de la fenêtre d'analyse est primordial pour une bonne estimation des amplitudes. Plus la largeur de la fenêtre est grande, plus le lissage de l'évolution temporelle est important, inversement, lorsqu'on est en présence d'un bruit de niveau élevé, le bruit sera d'autant plus atténué que la largeur de la fenêtre est grande. Pour sélectionner h de manière adaptative (c'est à dire optimalement pour chaque point d'estimation), nous avons appliqué la procédure décrite dans [5] (voir aussi [4]). Cette approche permet de comparer les estimations obtenues pour différentes valeurs de h en ne faisant appel qu'aux résultats concernant la variance des estimateurs (ce qui la distingue notablement des procédures plus classiques exposées dans [1] qui implique également une estimation du biais).

Compte tenu des remarques précédentes sur l'effet du bruit au niveau de chaque harmonique et de leur quasi orthogonalité, cette procédure est appliquée indépendamment pour chacune des harmoniques.

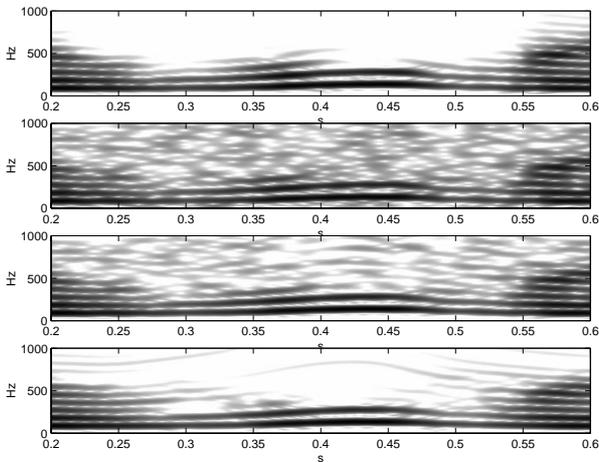


Figure 1: **A** Signal original; **B** Signal bruité (RSB 10 dB dans la bande $[0,1$ kHz]); **C** Estimation sinusoidale avec $2h$ (largeur de la fenêtre d'analyse) correspondant à 25 ms; **D** Estimation sinusoidale adaptative.

La figure 1 représente le spectrogramme à bande étroite des différents signaux dans la bande $[0,1$ kHz]. L'option **C** correspond à la pratique usuelle en traitement de la parole qui consiste à analyser le signal par trames de 25 ms. Il est clair que dans ce cas, la modélisation est fortement perturbée par la présence du bruit, les harmoniques d'ordre supérieurs étant, pour une fenêtre de 25 ms, "noyées" dans le bruit de fond. Par comparaison, le résultat obtenu sur la figure 1-D avec la procédure adaptative est bien différent. On observe une très forte diminution du niveau des composantes erratiques dues au bruit. Cette diminution de

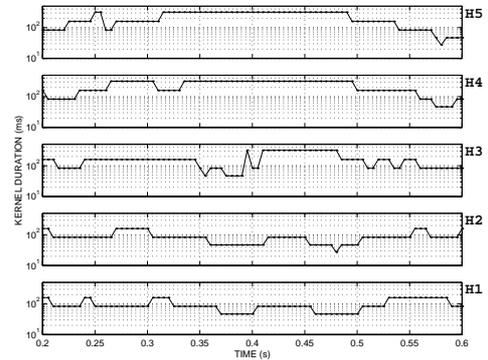


Figure 2: Durées des fenêtres d'analyse sélectionnée en fonctions du temps pour les cinq premiers harmoniques.

la variance d'estimation dans les zones fortement bruitées s'accompagne d'un accroissement de l'effet de lissage temporel dans ces mêmes zones. La figure 2 montre que les durées de fenêtre choisies sont effectivement plus grandes pour les harmoniques plus élevées et ont tendance à augmenter lorsque la puissance du signal diminue (dans la zone centrale de la figure).

Références

- [1] J. Fan et I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall, 1996.
- [2] G. Fay, E. Moulines, O. Cappé, et F. Bimbot. Polynomial quasi-harmonic models for speech analysis and synthesis. Dans *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, pages II-865-II-868, Seattle, May 1998.
- [3] G.Zhou. On the use of high order ambiguity function for multicomponent polynomial phase signals. Dans *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, volume 5, pages 3629-3632, 1997.
- [4] V. Katkovnik. On adaptive local polynomial approximation with varying bandwidth. Dans *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, 1998.
- [5] O. V. Lepski et V. G. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, 25(6): 2512-2546, 1997.
- [6] T. F. Quatieri et R. J. McAulay. Audio signal processing based on sinusoidal analysis/synthesis. Dans M. Kahrs and K. Brandenburg, editors, *Applications of Signal Processing to Audio and Acoustics*. Kluwer Academic Publishers, 1998.
- [7] B. G. Quinn et P. J. Thomson. Estimating the frequency of a periodic function. *Biometrika*, 78(1): 65-74, 1991.
- [8] V.Katkovnik. Non parametric local polynomial approximation of the time varying frequency and amplitude. *Commun.statist.-Theory meth*, 24(12): 3001-3025, 1995.