HAL
open science

Extraction d'un noyau non stationnaire de processus gaussien à l'aide des ondelettes

Romain Ravaille, Olivier Alata, Rémi Emonet

To cite this version:

Romain Ravaille, Olivier Alata, Rémi Emonet. Extraction d'un noyau non stationnaire de processus gaussien à l'aide des ondelettes. GRETSI 2019, Aug 2019, Lille, France. hal-02308042

HAL Id: hal-02308042

https://hal.science/hal-02308042v1

Submitted on 8 Oct 2019

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.

Extraction d'un noyau non stationnaire de processus gaussien à l'aide des ondelettes

Romain RAVAILLE[1], Olivier ALATA[2], Rémi EMONET[2]

[1]Univ Lyon, Université Jean Monnet, CNRS UMR 5208, Institut Camille Jordan,
Maison de l'Université, 10 rue Tréfilerie, CS 82301, 42023 Saint-Etienne Cedex 2, France

[2]Lab. Hubert Curien CNRS UMR 5516,
Université Jean Monnet Saint-etienne, IOGS Univ. de Lyon, 42023 Saint-Étienne, France
r.ravaille@univ-st-etienne.fr, olivier.alata@univ-st-etienne.fr
remi.emonet@univ-st-etienne.fr

Résumé – Dans [1] Durrande et al. décomposent un noyau classique de processus gaussien (un Matern) en deux sous-noyaux : un noyau permettant de décrire une composante aléatoire périodique et une composante résiduelle. Cette décomposition s'inspire de la décomposition en série de Fourier et permet, via une comparaison des deux sous-noyaux, d'évaluer à quel point les données interpolées sont périodiques. Notre but est de présenter une extension de cette méthode en utilisant les ondelettes, nous permettant ainsi d'obtenir un noyau non stationnaire. Notre objectif final est de pouvoir à la fois extraire et modéliser des composantes non stationnaires dans des signaux bruités afin de les caractériser et les simuler.

Abstract – In [1] Durrande et al. separate a classical gaussian process kernel (a Matern) into two sub kernels : one containing a random periodic component and the other being a residual kernel. This decomposition was inspired by Fourier serie decomposition and enable, by comparison of the two sub kernels, an assessment of how much periodicity our data have. Our goal is to show an extension of this work by using wavelets, providing us with a non stationary kernel. The end goal is to both extract and model non stationnary component of a noisy signal to analyse or to do simulations.

1 Introduction

Les processus gaussiens ont déjà fait leurs preuves dans le cadre de la régression [5],[6] et de la classification (par exemple [4]). Notre approche va plus spécialement se concentrer sur le cadre de la régression, en s'inspirant tout particulièrement de l'approche de [1] utilisant les Espaces de Hilbert à Noyau reproduisant (« RKHS » : Reproducing Kernel Hilbert Space). Un RKHS est un espace de fonctions dans lequel les fonctions d'évaluation en un point sont des formes linéaires continues. Un RKHS H est défini par son noyau K_x tel que $\forall f \in H, f(x) =< f, K_x >$. Ce noyaux possède les mêmes caractéristiques qu'un noyau de processus gaussien (symétrique, définit positif). En fait, par le théorème de Moore–Aronszajn [8], on sait qu'il y a une équivalence entre un noyau de processus gaussien et un noyau reproduisant pour un RKHS (on peut construire l'un à partir de l'autre et vice versa).

Lorsqu'il s'agit d'interpolation, un cas classique consiste à utiliser un noyau dépendant uniquement de la distance entre deux points des données. Cela vient de l'idée naturelle qu'un point proche du point à interpoler intervient plus sur le résultat qu'un point éloigné. En particulier la famille des noyaux Matern$_{\frac{r}{2}}$ permet de générer des courbes qui sont dérivables $\lceil \frac{r}{2} \rceil - 1$ fois et donc de contrôler les courbes obtenues en simulation [3].

Une approche particulière est celle de [1] qui consiste à séparer un modèle de processus gaussien en deux parties, l'une périodique et l'autre résiduelle, afin d'analyser les données et de déterminer si la composante périodique est majoritaire. Un autre intérêt que nous voyons à cette méthode est qu'elle permet de simuler une trajectoire à partir d'un noyau dont des caractéristiques ont été extraites d'un signal d'origine. En ce sens, améliorer cette extraction d'information présente un intérêt pour simuler des courbes dont nous maîtrisons mieux les caractéristiques.

C'est pour cela que notre but est d'étendre le travail de [1] en créant un modèle non stationnaire d'analyse à l'aide des ondelettes. A cet effet, nous commencerons par des rappels méthodologiques, nous présenterons l'adaptation de ce modèle existant aux ondelettes avant de montrer un cas pratique. En partie 2, nous introduirons la régression par processus gaussiens et la séparation de noyau de [1] réalisée à l'aide d'un sous-ensemble de fonctions issu de la décomposition en série de Fourier. Nous étendrons également la décomposition de noyau au cas des ondelettes. La partie 3 sera dédiée à la présentation de l'algorithme puis des résultats actuels. Après une brève conclusion nous proposerons quelques perspectives.

2 Rappels méthodologiques

2.1 Le modèle de régression avec Processus Gaussiens

Lorsque l'on décide d'interpoler des données avec un processus gaussien, cela revient à dire que chaque sortie y liée à une entrée x d'une fonction f est donnée par la formule :

$$y = f(x) + \epsilon \tag{1}$$

avec $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ et on suppose que $f(.)$ est une réalisation d'un Processus Gaussien (GP) de moyenne $m(.)$ et de noyau $k(.,.)$, appelé aussi fonction d'autocovariance du processus.

En considérant les observations $\vec{Y} = [y_1, .., y_N]^T$ aux positions $\vec{X} = [x_1, .., x_N]^T$, le modèle d'interpolation par processus gaussien est déterminé par la moyenne et la variance de la distribution conditionnelle.

A partir de la matrice $K(\vec{X}, \vec{X}) = [k(x_i, x_j)]_{x_i, x_j \in \vec{X}}$, on a :

$$m(x) = K(x, \vec{X})[K(\vec{X}, \vec{X}) + \sigma_\epsilon^2 I]^{-1}\vec{Y} \tag{2}$$

$$v(x) = K(x, x) - K(x, \vec{X})[K(\vec{X}, \vec{X}) + \sigma_\epsilon^2 I]^{-1}K(\vec{X}, x). \tag{3}$$

Le noyaux k est donc un élément essentiel du modèle. Dans notre cas on s'intéresse plus particulièrement aux noyaux de la forme « Matern $\frac{r}{2}$ » noté $K_{\frac{r}{2}}$. En effet ces noyaux ont une expression relativement simple lorsque n est impair (les formules pour $r = 1, 3$ et 5 sont données dans [1] par exemple) et permettent de maîtriser les courbes obtenues en simulations. Le RKHS que l'on obtient à partir d'un noyau Matern$_{\frac{r}{2}}$ est bien connu : il s'agit d'un espace de Sobolev $W^{r,2}$ souvent noté $\mathcal{H}^r$.

2.2 Décomposition dans une base de Fourier

Dans un premier temps, présentons la décomposition d'un noyau de processus gaussien dans une base appartenant à son RKHS telle qu'elle est effectuée dans [1].

Une décomposition en série de Fourier est la projection d'une fonction périodique sur une base de Fourier. On note

$$B(x) = (sin(\frac{2\pi}{\lambda}x), ..., cos(\frac{2\pi}{\lambda}qx))^T \tag{4}$$

une base de Fourier tronquée ($q = 20$ par exemple) des fonctions périodiques dont la fondamentale est $\frac{1}{\lambda}$. L'espace issu de cette base de Fourier $\mathcal{H}_p$ est un sous-espace de $\mathcal{H}^r$ car les fonctions composant B sont toutes $\mathcal{C}^\infty$ et que $\mathcal{C}^\infty \subset \mathcal{H}^r \ \forall r$.

On peut utiliser cet ensemble B ainsi que G la matrice de Gram de B dans $\mathcal{H}^r$ afin de créer un sous-noyau k_p contenant les informations périodiques de l'ensemble B :

$$k_p(x_1, x_2) = B(x_1)^T G^{-1} B(x_2) \tag{5}$$

On définit la matrice de Gram par $G_{i,j} = < B_i, B_j >_{\mathcal{H}^r}$ où $G_{i,j}$ est l'élément de la ligne i et la colonne j de G et B_i désigne l'élément i de la base B. Notons ici que ce produit scalaire dépend de l'espace dans lequel nous travaillons et donc de r. On donne ci-dessous la formule pour $r = 1$ en restreignant le support à l'intervalle $[a, b]$ (les formules pour $r = 3$ et $r = 5$ sont données dans [1]).

$$< f, g >_{\mathcal{H}^1} = \frac{l}{2\sigma^2}\int_a^b (\frac{1}{l}g + g')(\frac{1}{l}h + h')dt + \frac{1}{\sigma^2}g(a)h(a) \tag{6}$$

L'espace complémentaire à $\mathcal{H}_p$ dans $\mathcal{H}^r$ que l'on note $\mathcal{H}_a$ est tel que [7] :

$$\mathcal{H}^r = \mathcal{H}_p \oplus \mathcal{H}_a \tag{7}$$

$$k(x_1, x_2) = k_p(x_1, x_2) + k_a(x_1, x_2). \tag{8}$$

Ainsi notre noyau est décomposé en deux parties : k_p qui permet de modéliser les fonctions de l'espace $\mathcal{H}_p$ et k_a correspondant à un résidu (calculé dans la pratique avec $k_a = k - k_p$). On appelle ce résidu la partie apériodique même si les fréquences qui n'ont pas été mises dans la base B peuvent se retrouver dans ce noyau.

Au final les deux noyaux donnent deux modèles qui vérifient chacun l'équation (2) (avec m_p, v_p et m_a, v_a respectivement). Ces deux modèles sont tels que :

$$m(x) = m_p(x) + m_a(x), \tag{9}$$

$$v(x) \neq v_p(x) + v_a(x), \tag{10}$$

à cause de la linéarité de l'espérance. Remarquons ici que suite à la décomposition, les moyennes gardent la propriété d'additivité, ce qui n'est pas le cas pour les variances.

2.3 Application avec des ondelettes

L'idée de départ de notre travail est qu'il est possible d'effectuer le même type de décomposition avec une base d'ondelettes plutôt qu'une base de Fourier. Cela passe par le choix d'une ondelette mère, d'un ensemble de translations et d'échelles (dilatations) que nous voulons utiliser. Ces choix sont similaires aux choix de λ et q dans [1] dans le sens où nous allons utiliser la base définie par ces paramètres pour décrire le signal. On notera le sous-espace généré par notre famille d'ondelettes $\mathcal{H}_w$.

Dans un premier temps, il convient de vérifier que les ondelettes avec lesquelles nous travaillons appartiennent à l'espace de Hilbert à noyaux reproduisant que nous étudions (ici un espace de Sobolev $\mathcal{H}^r$). Ce n'est pas le cas pour toutes les ondelettes : par exemple les ondelettes de Haar n'appartiennent pas à $\mathcal{H}^r$ quel que soit le r choisi.

Si l'ondelette mère est $\mathcal{C}^\infty$, les translations et dilatations de cette ondelette seront alors toutes $\mathcal{C}^\infty \subset \mathcal{H}^r$. Parmi les ondelettes continues, nos exemples se concentreront sur les dérivées de la gaussienne qui sont $\mathcal{C}^\infty$ et dont le représentant le plus connu est le « mexican hat »(dérivée d'ordre 2).

On note $\psi \in \mathcal{C}^\infty$ notre ondelette mère et on choisit l'ensemble des dilatations (échelles) $a = \{a_n\}_{n \leq N_a}$ avec N_a le nombre d'échelles utilisées. Ensuite on choisit un ensemble de

translations $b_n = \{b_{m,n}\}_{m \leq M_n}$ pour chaque échelle n. La base de fonctions peut alors être définie :

$$B^w = \{\psi(\frac{x - b_{m,n}}{a_n})\}_{a_n \in a, b_{m,n} \in b_n}. \tag{11}$$

On peut donc calculer $k_w(x_1, x_2) = B^w(x_1)^T G_w^{-1} B^w(x_2)$ et obtenir le noyau résiduel k_r par soustraction avec le noyau utilisé : $k_r = k - k_w$. L'information contenue dans le noyau k_w est donc celle de la base d'ondelettes choisie et k_r est un résidu. Remarquons ici que la formule pour calculer G_w est la même que précédemment : $G_{w,i,j} = < B_i^w, B_j^w >_{\mathcal{H}^r}$ avec B_i^w l'élément i de la famille de fonctions.

3 Implémentation et résultats

Présentons les étapes d'un code appliquant la séparation d'un noyau de Matern dans le cadre de l'interpolation. On utilisera pour ce faire un cas simple :

On échantillonne avec 5000 échantillons sur l'intervalle $[0, 100]$ la fonction

$$f(x) = \sin(\beta x^3) + \alpha x. \tag{12}$$

Nous prenons 100 points dans l'intervalle $[0, 100]$ avec lesquels on interpolera f via un processus gaussien. On veut ensuite pouvoir analyser le noyau obtenu avec les ondelettes en utilisant la méthode de [1].

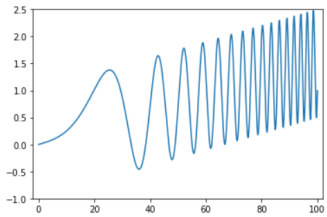

FIGURE 1 – Fonction réelle $\sin(\frac{x^3}{100^2}) + \frac{1.5x}{100}$.

Nous avons utilisé pour des raisons pratiques (calcul de G simple et rapide) une fonction proportionnelle à la dérivée d'une gaussienne comme ondelette et le noyau $K_{\frac{1}{2}}$. Ce noyau est de la forme : $K_{\frac{1}{2}}(x_1, x_2) = \sigma^2 \exp(\frac{|x_1 - x_2|}{l})$ avec σ et l ses hyperparamètres.

3.1 Algorithme

L'algorithme se déroule en cinq étapes :

1. Choisir un ensemble d'ondelettes B_Ψ (cf (11)).
2. Calculer $k_w = B_\Psi^T(X) G^{-1} B_\Psi(X)$.
3. Calculer $k_r = k_2 - k_{2,w}$ où k_2 est un noyau de Matern$_{\frac{1}{2}}$ et $k_{2,w}$ est un noyau défini comme k_w mais avec les mêmes hyperparamètres que k_2. *En sachant que les hyperparamètres d'un noyau n'ont aucune incidence sur les RKHS concernés, l'équation* (7) *est vraie pour tout choix d'hyperparamètres de nos noyaux. On peut donc en particulier choisir de calculer k_r à partir de noyaux ayant des hyperparamètres totalement différents du k_w calculé précédemment. k_r a donc son propre lot d'hyperparamètres dans l'étape suivante.*
4. Définir $k_f = k_w + k_r$, le noyau qui sera utilisé pour l'interpolation finale. *Chaque noyau composant k_f a son propre lot d'hyperparamètres. Par conséquent, k_f a 4 hyperparamètres (au lieu de 2) : σ_w, σ_r, l_w et l_r qui correspondent respectivement aux hyperparamètres de k_w et k_r. Avoir plus d'hyperparamètres permet notamment au modèle d'être plus flexible.*
5. Optimiser les hyper-paramètres de k_f : *On utilise une descente de gradient pour trouver le maximum de vraisemblance.*

   La log-vraisemblance de notre processus gaussien f sachant les points X est notée $\mathcal{L}_K = \mathcal{L}(\sigma_w, \sigma_r, l_w, l_r, X)$, et est définie par :

$$\begin{aligned}
\mathcal{L}_K &= \log(p(f(x)|x, \sigma_w, \sigma_r, l_w, l_r)) \\
&= -\frac{1}{2} f(X)^T K^{-1} f(X) \\
&- \frac{1}{2} \log(\det(K)) - \frac{|X|}{2} \log(2\pi) \\
&= -\frac{1}{2} f(X)^T (B_\Psi^T G^{-1} B_\Psi + K_r)^{-1} f(X) \\
&- \frac{1}{2} \log(\det(B_\Psi^T G^{-1} B_\Psi + K_r)) - \frac{|X|}{2} \log(2\pi).
\end{aligned} \tag{13}$$

3.2 Résultats

Dans un premier temps, si on applique l'algorithme présenté dans [1], avec $q = 20$ et $\lambda = 1$ à notre fonction, nous obtenons :

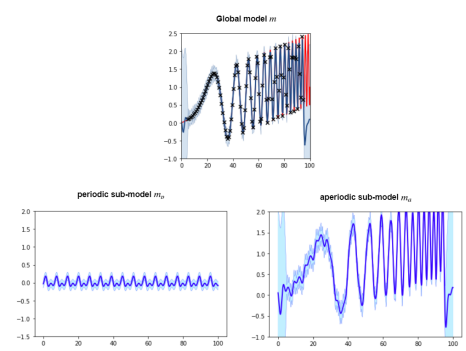

FIGURE 2 – Décomposition de la fonction f de l'équation (12) dans les conditions énoncées dans l'introduction de la partie 3, en utilisant l'algorithme proposé dans [1]. À gauche la moyenne et l'intervalle de confiance à 95% (obtenu avec la variance) sont représentés pour le noyau périodique. À droite la même représentation mais pour le noyau non périodique.

L'observation de la figure 2 nous permet de dire que le signal obtenu à partir du noyau périodique est peu informatif quant au signal d'origine. Le fait que le noyau k_p soit nécessairement périodique est ici un problème si on veut analyser correctement le signal.

Le choix le plus important pour notre algorithme est celui des dilatations a et translations b pour créer la base définie par l'équation (11). Pour mettre en valeur ce que permet de faire notre algorithme nous explorons deux cas : premièrement, nous utilisons $a = \{4\}$ pour explorer les basses fréquences du signal (voir figure 3). Ensuite nous utilisons $a = \{1\}$ pour extraire la partie haute fréquence et isoler les oscillations de fréquences croissantes de la fonction linéaire (voir figure 4). En utilisant les étalements temporels des fonctions (cf [2] partie 4.3.2 pour une explication plus détaillée), on peut obtenir une approximation du bon nombre de translations à prendre pour une échelle donnée. On notera N_t le nombre de translations régulières considérées. En général il est plus prudent de prendre plus de translations (en sachant cependant que cela revient à augmenter la taille de notre base et que cela ralentit donc les calculs).

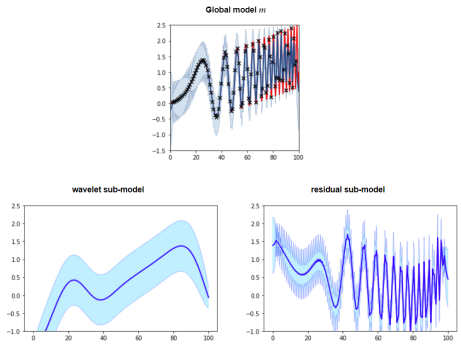

FIGURE 3 – Décomposition de la fonction f de l'équation (12) par notre algorithme avec $a = \{4\}$ et $N_t = 40$. À gauche le noyau obtenu avec les ondelettes à droite le résidu.

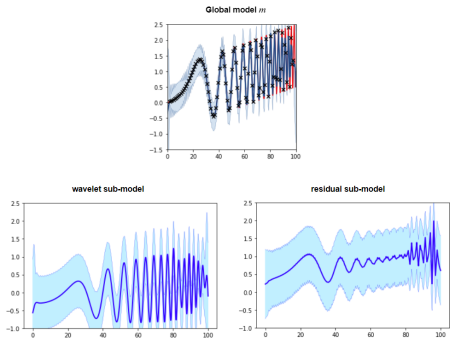

FIGURE 4 – Décomposition de la fonction f de l'équation (12) par notre algorithme avec $a = \{1\}$ et $N_t = 80$. À gauche le noyau obtenu avec les ondelettes à droite le résidu.

On remarquera ici que l'on peut combiner plusieurs échelles au besoin pour décrire plus d'information dans le signal. Nos ondelettes décrivent cependant une grande gamme de fréquences : par exemple pour $a = \{2\}$ on est déjà capable de repérer la composante linéaire ; Pour $a = \{4\}$ qu'une partie des basses fréquences a été extraite.

4 Conclusion

Dans ce papier nous avons présenté comment il est possible d'adapter le travail de Nicolas Durrande aux ondelettes.

Notre méthode comparée à celle exploitant une décomposition sur des fonctions sinusoïdales devrait permettre d'analyser des signaux qui ne sont pas nécessairement périodiques mais aussi ceux qui possèdent une composante non-stationnaire.

Dans de futurs travaux, réussir à optimiser l'algorithme (en utilisant par exemple les méthodes variationnelles) pourrait permettre d'avoir des résultats intéressants en termes d'analyse sur des données plus larges (plus de points). De plus, étendre ce travail à une ondelette dont la bande passante est plus restreinte permettrait une analyse plus fine.

Remerciements Nous tenons à remercier Olivier Roustant et Julian Tugaut pour les discussions très enrichissantes que nous avons eues concernant ce travail.

Références

[1] N. Durrande, J. Hensman, Ma. Rattray et N. D. Lawrence *Detecting periodicities with Gaussian processes* PeerJ Computer Science, 2016.

[2] S. Mallat. *A wavelet tour of signal processing*. Academic press, 1998.

[3] Santner, T. J., Williams, B. J., et Notz, W. I. . *The design and analysis of computer experiments.*. Springer Science et Business Media, 2013.

[4] Urtasun, Raquel, Darrell et Trevor, *Discriminative Gaussian Process Latent Variable Model for Classification.* 2007.

[5] T. Kailath, *RKHS approach to detection and estimation problems–I : Deterministic signals in Gaussian noise* IEEE Trans. on Inf. Theory, vol. 17, no. 5, p. 530-549, 1971.

[6] L. Clifton, D. A. Clifton, M. A. F. Pimentel, P. J. Watkinson et L. Tarassenko, *Gaussian process regression in vital-sign early warning systems* Annual International Conference of the IEEE int. EMBC, San Diego, CA, pp. 6161-6164, 2012.

[7] , Berlinet, Alain, *Reproducing Kernel Hilbert Spaces In Probability and Statistics*, Springer US, 2004

[8] N. Aronszajn, *Theory of Reproducing Kernels*, Transactions of the American Mathematical Society, Vol. 68, No. 3, p.337-404, 1950.