

Ondelettes analytiques, application à l'analyse des processus multivariés à longue mémoire

Irène GANNAZ¹, Sophie ACHARD², Marianne CLAUSEL³, François ROUEFF⁴

¹Université de Lyon, CNRS UMR 5208, INSA de Lyon, Institut Camille Jordan, France

²CNRS, Univ. Grenoble Alpes, GIPSA-lab, Grenoble, France

³Univ. Grenoble Alpes, LJK, Grenoble, France

⁴LTCI, Télécom-Paristech, Université Paris-Saclay

irene.gannaz@insa-lyon.fr, sophie.achard@gipsa-lab.grenoble-inp.fr,
marianne.clausel@univ-grenoble-alpes.fr, francois.roueff@telecom-paristech.fr

Résumé – Dans de nombreux domaines tels que la finance, la géophysique ou les neurosciences, les données se présentent sous la forme de séries temporelles multivariées. Un enjeu de l'analyse statistique est de prendre en compte cet aspect multivarié, notamment en raison de phénomènes de phase pouvant être induits par la présence de longue mémoire. Les ondelettes analytiques sont un outil adapté à ce cadre d'étude. Nous nous intéressons ainsi aux ondelettes quasi-analytiques introduites par Selesnick. Nous montrons tout d'abord leur existence et nous explicitons ensuite leur qualité analytique. Nous illustrons enfin l'avantage de l'utilisation de ces ondelettes dans un cas simple de processus multivariés à longue mémoire.

Abstract – Multivariate processes with long-range dependent properties are found in a large number of applications including finance, geophysics and neuroscience. Statistical analysis of such data is challenging because multivariate time series present phase phenomena. Analytic wavelets are well suited to deal with these characteristics. Our starting point is a paper of Selesnick which introduces quasi-analytic wavelets. We first establish the existence of these wavelets. We also give an exact formula quantifying their analytic quality. We then illustrate on simulations the relevance of quasi-analytic wavelets for multivariate time series analysis.

1 Introduction et contexte

Une ondelette analytique est une ondelette complexe dont la transformée de Fourier est nulle aux fréquences négatives. Par le théorème de Paley-Wiener, cette propriété de support en Fourier est incompatible avec une propriété de support compact de l'ondelette en temps. Néanmoins l'analyticité de l'ondelette présente de nombreux intérêts pour l'analyse de signaux, dès lors que la phase du contenu fréquentiel joue un rôle important, et pas uniquement son module [9]. Comme la propriété de support compact en temps facilite grandement l'utilisation des ondelettes en pratique, une approche essentielle pour ce type d'analyse consiste à construire des ondelettes complexes *approximativement* analytiques [9]. Bien que couramment utilisées en pratique, les ondelettes par facteur commun [8, 11] ont été peu étudiées et leur existence même n'était, à notre connaissance, pas démontrée. Ces ondelettes dépendent de deux paramètres entiers L et M qui déterminent respectivement l'ordre de l'approximation analytique et le nombre de moments nuls.

Notre contribution est de reprendre la construction de ces ondelettes en démontrant qu'elles peuvent en effet

être construites pour tout $L, M \geq 1$. Nous déterminons de plus une formule exacte et explicite permettant d'évaluer facilement leur *qualité analytique*, c'est-à-dire de contrôler la présence de fréquences négatives en fonction des paramètres de l'ondelette. L'ensemble de ces résultats sont présentés ici sans démonstration (voir [4]).

Dans un deuxième temps nous illustrons l'intérêt de ce type d'ondelette pour l'analyse de séries temporelles multivariées à longue mémoire qui trouve de nombreuses applications récentes [6, 12]. La phase joue un rôle essentiel pour modéliser l'interdépendance des composantes d'un processus multivarié. Elle apparaît en effet dans la densité spectrale croisée entre deux composantes du signal. Par ailleurs, l'analyse fréquentielle des processus stationnaires au second ordre, par périodogramme ou ondelettes réelles, présente un intérêt majeur dans le contexte de la longue mémoire [1, 3]. Pourtant du fait que ces analyses se concentrent en général aux basses fréquences, la question de la phase a été peu abordée. Il a été pourtant remarqué par [5, 10] que dans le cas de présence de longue mémoire, une phase peut néanmoins apparaître aux basses fréquences. Il nous a semblé intéressant d'étudier l'intérêt des ondelettes (approximativement) analy-

tiques pour l'analyse de ce type de processus. Nous nous contenterons ici de l'illustrer dans le cadre d'un exemple particulier déjà identifié par [10].

Cet article est organisé comme suit. Dans une première partie, nous reprenons la construction d'ondelettes analytiques par facteur commun proposée dans [7]. Nous montrons la validité de cette construction en explicitant le terme présent dans le facteur commun. Nous quantifions également la qualité de l'approximation analytique. Dans une deuxième partie, nous montrons les bonnes propriétés de la décomposition dans le cas du modèle multi-varié à longue mémoire $M(\mathbf{d})$ [2, 10].

2 Construction et existence de la base d'ondelettes

Il est impossible de construire une paire de Hilbert en imposant d'avoir des filtres à réponse impulsionnelle finie (RIF). Dans cette partie, nous détaillons la construction de [7] d'une paire d'ondelettes qui sont approximativement transformée de Hilbert l'une de l'autre.

2.1 Définition

Soient ψ_G et ψ_H deux fonctions réelles d'ondelettes. Elles forment une paire de fonctions de Hilbert si leurs transformées de Fourier, notées respectivement $\hat{\psi}_G$ et $\hat{\psi}_H$, vérifient

$$\hat{\psi}_G(\omega) = -i \operatorname{sign}(\omega) \hat{\psi}_H(\omega). \quad (1)$$

Les fonctions ψ_G et ψ_H sont associées à des filtres passe-bas, respectivement g_0 et h_0 . Nous notons G_0 et H_0 leurs fonctions de transfert, obtenues par une transformée en z ; par exemple, $G_0(z) = \sum_n g_{0n} z^{-n}$. La propriété (1) ne peut être obtenue avec des filtres RIF ([7]). Nous cherchons donc à trouver des filtres RIF qui vérifient l'approximation

$$G_0(e^{i\omega}) \sim H_0(e^{i\omega}) e^{-i\omega/2} \quad \text{quand } \omega \rightarrow 0. \quad (2)$$

Nous reprenons la construction des filtres finis de [7] qui satisfont cette condition et nous étudions son existence et la qualité de l'approximation (2).

2.2 Construction avec facteur commun

[7] propose de chercher des solutions sous la forme

$$H_0(z) = (1 + 1/z)^M Q_{L,M}(z) D_L(z), \quad (3)$$

$$G_0(z) = (1 + 1/z)^M Q_{L,M}(z) D_L(1/z) z^{-L}. \quad (4)$$

La présence de $(1 + 1/z)^M$ permet d'assurer une propriété de M moments nuls aux filtres. Le facteur D_L est la transformée en z d'un filtre RIF causal de taille L , $D_L(z) = 1 + \sum_{\ell=1}^L d_L(\ell) z^{-\ell}$ qui va permettre d'obtenir la propriété

de quasi-analyticité. Afin que les ondelettes vérifient l'approximation (2), il est proposé dans [7] de définir D_L par

$$\frac{e^{-i\omega L} D_L(e^{-i\omega})}{D_L(e^{i\omega})} = e^{-i\omega/2} + o_{\omega \rightarrow 0}(|\omega|^{2L+1}). \quad (5)$$

Une expression des coefficients $d_L(\ell)$, $\ell = 1, \dots, L$ est alors facilement obtenue dans [7]. Nous montrons tout d'abord que D_L peut être explicité sous la forme suivante.

Proposition 1. *Soit L un entier positif et soit $D_L(z) = 1 + \sum_{\ell=1}^L d_L(\ell) z^{-\ell}$ vérifiant (5). Alors, pour tout $z \in \mathbb{C}^*$,*

$$D_L(z) = \frac{z^{-L}}{4L+2} \left[(1 + \sqrt{z})^{2L+1} + (1 - \sqrt{z})^{2L+1} \right], \quad (6)$$

où \sqrt{z} désigne indifféremment l'une des deux racines complexes de z .

2.3 Existence

Nous nous intéressons plus spécifiquement au cadre de filtres vérifiant la propriété de reconstruction parfaite (voir [9] et les références associées). Pour les filtres H_0 et G_0 définis par (3) et (4), la reconstruction parfaite est obtenue si et seulement si

$$R_{L,M}(z) S_{L,M}(z) + R_{L,M}(-z) S_{L,M}(-z) = 2, \quad (7)$$

$$\text{avec } R_{L,M}(z) = Q_{L,M}(z) Q_{L,M}(1/z)$$

$$\text{et } S_{L,M}(z) = (2 + z + 1/z)^M D_L(z) D_L(1/z).$$

Notre première contribution est de montrer que (7) admet une solution, factorisable sous la forme souhaitée.

Théorème 2. *Soient L et M deux entiers positifs. Alors il existe $Q_{L,M}$ tel que $Q_{L,M}(1/z)$ soit un polynôme réel de z et tel que $R_{L,M}(z) = Q_{L,M}(z) Q_{L,M}(1/z)$ soit solution de (7).*

Ce théorème garantit l'existence d'une solution des équations (3)-(4)-(7). Aucune formule explicite de $R_{L,M}$ et $Q_{L,M}$ n'est disponible mais la preuve est constructive et la résolution numérique est possible. Un algorithme est proposé dans [8], sans garantie de succès.

2.4 Qualité de l'approximation analytique

À l'aide de l'expression donnée en Proposition 1, nous sommes en mesure d'expliciter la qualité de l'approximation de la propriété d'analyticité (1) recherchée pour le couple (ψ_G, ψ_H) .

Proposition 3. *Pour tout $\omega \in \mathbb{R}$,*

$$\hat{\psi}_G(\omega) = e^{-i\omega/2} e^{i\beta_L(\omega)} \hat{\psi}_H(\omega)$$

$$\hat{\psi}_G(\omega) = i e^{-i\eta_L(\omega)} \hat{\psi}_H(\omega).$$

avec

$$\beta_L(\omega) = 2(-1)^L \sum_{j=1}^{\infty} \operatorname{atan} \left(\tan^{2L+1}(\omega 2^{-j-2}) \right),$$

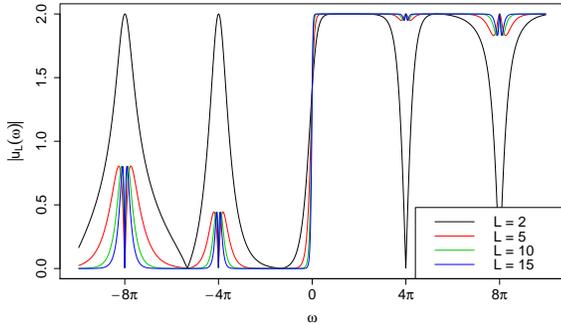
$$\eta_L(\omega) = 2(-1)^L \operatorname{atan} \left(\tan^{2L+1}(\omega/8 + \pi/4) \right) + \beta_L(\omega/2).$$

Ainsi,

$$\hat{\psi}_H(\omega) + i\hat{\psi}_G(\omega) = u_L(\omega)\hat{\psi}_H(\omega),$$

avec $u_L(\omega) = 1 - e^{-i\eta_L(\omega)}$. La figure 1 illustre le comportement de $u_L(\cdot)$. Sur les intervalles $(-4\pi, 0)$ et $(0, 4\pi)$, on peut aisément montrer que $u_L(\omega)$ tend respectivement vers 0 et vers 2 lorsque L tend vers l'infini. Les bonnes propriétés analytiques du couple $(\hat{\psi}_G, \hat{\psi}_H)$ sont alors assurées par le fait que $\hat{\psi}_H(\omega) \rightarrow 0$ lorsque $\omega \rightarrow 0$ grâce à la propriété de M moments nuls imposée par (3)-(4). Pour $|\omega| > 4\pi$, une décroissance de $\hat{\psi}_H(\omega)$ permet d'obtenir des résultats satisfaisants. Ce dernier point est vérifié numériquement.

FIGURE 1 – Approximation analytique. La figure représente $\omega \mapsto |u_L(\omega)|$ pour $L=2, 5, 10$.



3 Application aux processus $M(\mathbf{d})$

Afin d'illustrer l'utilité des ondelettes approximativement analytiques par facteur commun, nous proposons de les utiliser pour analyser des processus $M(\mathbf{d})$ bivariés avec différentes valeurs de \mathbf{d} . Nous montrerons que ces ondelettes permettent d'exhiber des corrélations ignorées par les ondelettes réelles. Les paramètres K et L sont fixés $K = 4$ et $L = 2$.

3.1 Définition du modèle simulé

À titre d'exemple, nous considérons ici un processus bivarié ARFIMA(0, \mathbf{d} , 0) défini par

$$X_\ell(k) = (1 - \mathbb{L})^{-d_\ell} u_\ell(k), \quad \ell = 1, 2, k \in \mathbb{Z},$$

avec \mathbb{L} opérateur de décalage et $\begin{pmatrix} u_1(k) \\ u_2(k) \end{pmatrix}$ i.i.d. de loi

$$\mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Omega\right), \text{ avec } \Omega = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}.$$

Alors la densité spectrale matricielle de (X_1, X_2) vérifie

$$f(\lambda) = \Lambda(\mathbf{d})\Omega\Lambda(\mathbf{d})^* \text{ où } \Lambda(\mathbf{d}) = \text{diag}((1 - e^{-i\lambda})^{-\mathbf{d}}).$$

Nous pouvons en déduire une approximation au premier ordre, quand $\lambda \rightarrow 0^+$, $f(\lambda) \sim \Sigma(\lambda)$ avec

$$\Sigma(\lambda) = \tilde{\Lambda}(\mathbf{d})^*\Omega\tilde{\Lambda}(\mathbf{d}), \text{ avec } \tilde{\Lambda}(\mathbf{d}) = \text{diag}(\lambda^{-\mathbf{d}}e^{-i\pi\mathbf{d}/2}).$$

Et quand $\lambda \rightarrow 0^-$, nous avons $f(\lambda) \sim \Sigma^*(\lambda)$. Nous introduisons ρ , coefficient de corrélation de $\Sigma(\lambda)$, $\rho = \Sigma_{1,2}(\lambda) / \sqrt{\Sigma_{1,1}(\lambda)\Sigma_{2,2}(\lambda)} = \Omega_{1,2} e^{i\pi(d_1 - d_2)/2}$ (ne dépend pas de λ). Dans ce cas particulier, la phase de la corrélation en 0^+ est donc égale à $\pi(d_1 - d_2)/2$. La discontinuité de la densité spectrale en 0 justifie l'utilisation de transformées analytiques afin d'identifier la phase. En effet il est nécessaire de ne considérer que les fréquences positives (ou négatives) si on ne veut pas perdre l'information de la phase.

Soient $\{W_{j,k}(\ell), j \geq 0, k \in \mathbb{Z}\}$ pour $\ell = 1, 2$ les coefficients d'ondelettes du processus X_ℓ obtenus à partir d'une fonction d'ondelette ψ . Notons θ_j la corrélation empirique entre $\{W_{j,k}(1), k \in \mathbb{Z}\}$ et $\{W_{j,k}(2), k \in \mathbb{Z}\}$ pour une échelle $j \geq 0$ donnée. On peut alors montrer que la valeur de θ_j quand $j \rightarrow \infty$ est liée à la valeur de ρ .

Plus précisément, lorsque ψ est réelle, la proposition 2 de [2] montre qu'asymptotiquement θ_j tend vers la partie réelle de ρ lorsque j tend vers l'infini. Lorsque la phase vaut $\pi/2$, c'est-à-dire ici lorsque $d_1 - d_2$ est proche de 1 modulo 2, la corrélation des coefficients d'ondelette réelle s'annule. Lorsque ψ est analytique, nous montrons sur simulations que ρ peut bien être identifiée par θ_j , en prenant en compte sa partie imaginaire. Nous notons dans la suite Re la partie réelle et Im la partie imaginaire.

3.2 Résultats numériques

Pour la première simulation, le paramètre longue mémoire \mathbf{d} est égal à $(0.2, 0.2)$. La phase exhibée ci-dessus est alors nulle. La figure 2 représente $\{\theta_j, j \geq 0\}$ avec des ondelettes réelles calculés à partir de signaux de longueur 2^{12} et 1000 simulations. Cette figure montre que ρ est bien identifiable dans ce cadre. Pour la deuxième simulation, le paramètre longue mémoire \mathbf{d} est égal à $(0.2, 1.2)$. La différence de phase est donc égale à $\pi/2$, et l'approximation de la densité spectrale est imaginaire pure. Dans ce contexte, la figure 2 illustre l'incapacité d'estimer ρ , dans la mesure où seule la partie imaginaire de f contient l'information relative à son comportement.

Dans la première simulation, avec $d_1 = d_2 = 0.2$, lorsque nous considérons une ondelette analytique, la figure 3 montre que nous retrouvons des résultats similaires à ceux obtenus avec une ondelette réelle avec l'information supplémentaire que la partie imaginaire des corrélations $\{\theta_j, j \geq 0\}$ reste nulle. Dans la deuxième simulation, avec $d_1 = 0.2$ et $d_2 = 1.2$, comme illustré par la figure 3, la partie réelle des corrélations $\{\theta_j, j \geq 0\}$ s'annule à partir d'une certaine échelle, comme observé pour les ondelettes réelles. Mais dans ce cadre, la partie imaginaire permet d'identifier la présence de corrélation. Plus précisément, pour les grandes échelles, les corrélations $\{\theta_j, j \geq 0\}$ sont proches en moyenne de ρ .

D'autres simulations, non décrites ici, montrent que la donnée de $\{\theta_j, j \geq 0\}$ obtenue avec des ondelettes analy-

FIGURE 2 – Boxplot des corrélations empiriques des coefficients d’ondelettes $\{\theta_j, j \geq 0\}$ pour les ondelettes réelles en fonction des échelles. En rouge, la partie réelle de la valeur théorique ρ du coefficient de corrélation du modèle ARFIMA.

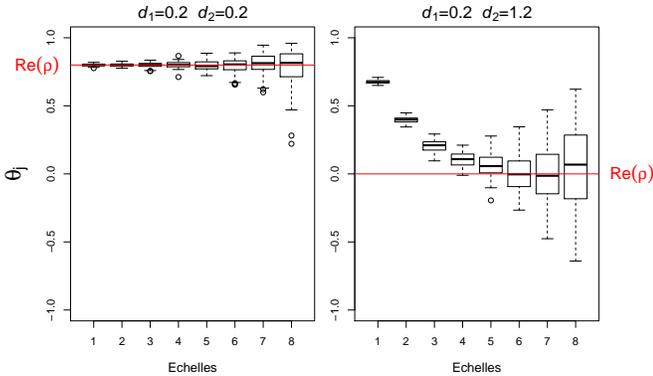
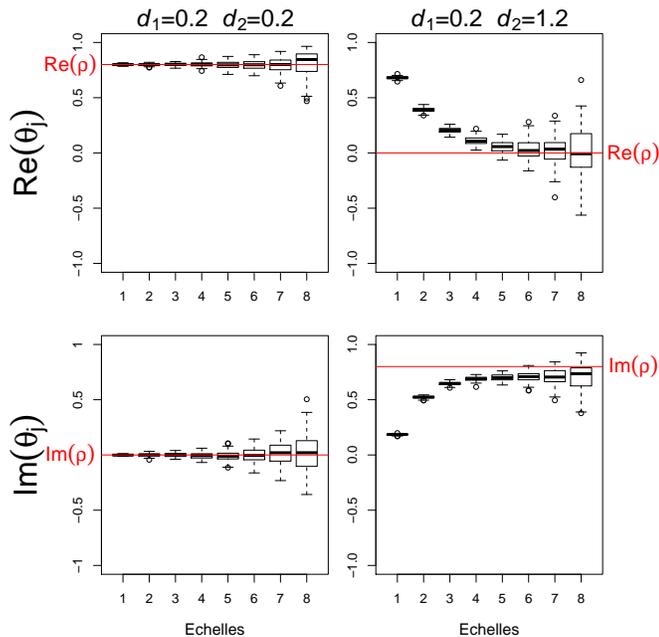


FIGURE 3 – Boxplot des corrélations empiriques des coefficients d’ondelettes $\{\theta_j, j \geq 0\}$ pour les ondelettes réelles en fonction des échelles. En rouge, la partie réelle ou imaginaire de la valeur théorique ρ du coefficient de corrélation du modèle ARFIMA.



tiques permet de retrouver module et phase de la densité spectrale en 0^+ .

4 Conclusion et perspectives

La première contribution de ce travail est d’avoir prouvé l’existence des ondelettes approximativement analy-

tique par facteur commun développées par [8]. De plus, nous avons établi une formule exacte permettant d’expliquer la qualité de l’approximation analytique.

Les simulations montrent des perspectives intéressantes pour l’utilisation de ces ondelettes en statistique. Une première étape est d’étudier du point de vue théorique le comportement des coefficients d’ondelettes, afin d’obtenir un résultat similaire à la proposition 2 de [2]. Dans un deuxième temps, nous souhaitons développer des procédures reposant sur ces ondelettes. Notre objectif est d’estimer les paramètres de longue mémoire et les propriétés de phase entre des signaux, dans le contexte de différentes classes de processus multivariés (tels que ceux décrits dans [5]).

Références

- [1] P. Abry and D. Veitch. Wavelet analysis of long-range-dependent traffic. *Information Theory, IEEE Transactions on*, 44(1) :2–15, 1998.
- [2] S. Achard and I. Gannaz. Multivariate wavelet whittle estimation in long-range dependence. *Journal of Time Series Analysis*, 37(4) :476–512, 2016.
- [3] G. Faÿ, E. Moulines, F. Roueff, and M. S. Taqqu. Estimators of long-memory : Fourier versus wavelets. *Journal of Econometrics*, 151(2) :159–177, 2009.
- [4] I. Gannaz, S. Achard, M. Clausel, and F. Roueff. New results on approximate hilbert pairs of wavelet filters with common factor. En préparation.
- [5] S. Kechagias and V. Pipiras. Definitions and representations of multivariate long-range dependent time series. *Journal of Time Series Analysis*, 2014.
- [6] V. Maxim, L. Şendur, M. J. Fadili, J. Suckling, R. Gould, R. Howard, and E. T. Bullmore. Fractional Gaussian noise, functional MRI and Alzheimer’s disease. *NeuroImage*, 25 :141–158, 2005.
- [7] I. W. Selesnick. Hilbert transform pairs of wavelet bases. *Signal Processing Letters, IEEE*, 8(6) :170–173, 2001.
- [8] I. W. Selesnick. The design of approximate hilbert transform pairs of wavelet bases. *Signal Processing, IEEE Transactions on*, 50(5) :1144–1152, 2002.
- [9] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury. The dual-tree complex wavelet transform. *Signal Processing Magazine, IEEE*, 22(6) :123–151, 2005.
- [10] K. Shimotsu. Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics*, 137(2) :277–310, 2007.
- [11] J.-P. Thiran. Recursive digital filters with maximally flat group delay. *Circuit Theory, IEEE Transactions on*, 18(6) :659–664, 1971.
- [12] B. Whitcher and M. J. Jensen. Wavelet estimation of a local long memory parameter. *Exploration Geophysics*, 31(1/2) :94–103, 2000.