Multiscale demixing for multivariate self-similarity estimation

Herwig WENDT[1] Gustavo DIDIER[2] Marcus CARLSSON[3] Erik TROEDSSON[3] Patrice ABRY[4]

[1]CNRS, IRIT, Université de Toulouse, Toulouse, France

[2]Mathematics Department, Tulane University, New Orleans, LA, USA

[3]Centre for Mathematical Sciences, Lund University, Lund, Sweden

[4]CNRS, ENS de Lyon, Laboratoire de Physique, Lyon, France

WORK PARTIALLY SUPPORTED BY SWEDISH RESEARCH COUNCIL GRANT 2022-04917. G.D.'S LONG TERM VISITS TO ENS DE LYON WERE SUPPORTED BY THE SCHOOL, CNRS, AND SIMONS FOUNDATION COLLABORATION GRANT #714014.

Résumé – De nombreux systèmes réels génèrent des séries temporelles multivariées présentant des propriétés d'auto-similarité et d'invariance d'échelle. L'estimation des exposants de Hurst est cruciale pour analyser les dépendances à long terme, mais les méthodes traditionnelles basées sur l'analyse des valeurs propres induisent des distorsions, notamment en présence de variations d'amplitude d'échelle. Nous proposons une approche améliorant la précision de l'estimation via la diagonalisation conjointe de matrices aléatoires issues de la transformée en ondelettes. En approximant une matrice de mélange instantanée commune à plusieurs échelles, notre méthode surmonte les limites des régressions des valeurs propres, assurant des estimations plus fiables. Des simulations de Monte Carlo confirment sa supériorité sur les techniques conventionnelles en mélange orthogonal et non orthogonal.

Abstract – Estimating Hurst exponents is essential for analyzing long-range dependencies in multivariate time series, which often exhibit self-similarity and scale invariance. However, traditional eigenanalysis-based methods can introduce distortions, particularly when scaling amplitudes vary. To address this, we propose a novel approach that improves estimation accuracy by jointly diagonalizing wavelet random matrices. By approximating a common instantaneous mixing matrix across multiple scales, our method circumvents the limitations of scale-wise eigenvalue regressions, leading to more reliable self-similarity parameter estimates. Extensive Monte Carlo simulations demonstrate its advantages over conventional techniques in both orthogonal and non-orthogonal mixing scenarios.

1 Introduction

Scale invariance. Scale invariance is a key property in many domains, including physics and engineering. A signal X is *scale-invariant* (or fractal) when it exhibits self-similarity across scales, lacking a characteristic temporal scale. Unlike conventional statistical models, defined for fixed scales, such signals are characterized by *scaling exponents*, which quantify their behavior across multiple resolutions. A cornerstone model is *fractional Brownian motion* (fBm) [9], the only Gaussian, self-similar process with stationary increments. The Hurst exponent H governs its scaling law : $\{B_H(t)\}_{t\in\mathbb{R}} \overset{\text{f.d.d.}}{=} \{a^H B_H(t/a)\}_{t\in\mathbb{R}}, \quad 0 < H < 1$. Estimating H is essential in signal processing applications such as classification, anomaly detection, and system diagnosis. The *wavelet transform* provides a powerful framework for extracting scaling exponents [13].

Multivariate self-similarity. While univariate fBm is well understood, the increasing availability of multivariate data necessitates extensions to self-similar processes in higher dimensions. For instance, brain activity time series in neuroscience span from hundreds (MEG) to thousands (fMRI) of channels. In econometrics, detecting scaling laws in fractional time series is key to understanding long-run dependencies, such as *cointegration* [10]. *Operator fractional Brownian motion* (ofBm) is a canonical model for multidimensional scale-invariant structures in real data [8]. A key difficulty in such datasets is that multiple distinct Hurst exponents can define large-scale behavior along non-canonical axes, and failing to account for them can lead to severe estimation biases [2].

Eigenanalysis of Wavelet Random Matrices. *Random matrices* are fundamental in mathematical physics [5] and high-dimensional statistics [14]. *Wavelet random matrices* (WRMs) extend this framework to multivariate fractal systems, revealing multiple scaling laws as power laws in their eigenspectra, governed by Hurst exponents ; the authors first introduced WRMs in [1]. Despite strong asymptotic properties, WRM eigenanalysis struggles with finite samples, particularly due to variations in scaling *amplitudes*. After log-linearization, affine scaling laws may *cross over*, biasing Hurst exponent (slope) estimates. In the case of instantaneously correlated systems, blind source separation [6] provides inspiration for a solution. Namely, one can exactly joint diagonalize WRMs at two scales [3] and, hence, estimate the mixing matrix that distorts power laws. While effective in specific settings, this method is limited to two scales and lacks robustness. Furthermore, it is affected by significant crossover effects over small scales, impacting Hurst exponent estimation performance.

Goals and Contributions. This work introduces an efficient method for estimating Hurst exponents in instantaneously correlated, multivariate self-similar systems while addressing crossover distortions. Section 2 reviews the operator fractional Brownian motion (ofBm) model and wavelet eigenvalue-based estimation. Our first main contribution, detailed in Section 3, consists in a novel WRM diagonalization method across *multiple* scales, leveraging a recently proposed algorithm [12].

This approach advances WRM-based estimation by mitigating crossover effects. Section 4 presents an extensive numerical study, showing improved finite-sample performance over standard eigenanalysis and exact joint diagonalization [2]. Finally, Section 5 discusses implications and future research directions.

2 Multivariate self-similarity

2.1 Operator fractional Brownian motion

We recall the definition and key properties of ofBm [8]. Let $\underline{B}_{\underline{H},\Sigma}(t) = \big(B_{H_1}(t),\dots,B_{H_M}(t)\big)_{t\in\mathbb{R}}$ denote an M-dimensional fBm with Hurst exponents $\underline{H} = (H_1,\dots,H_M)$, where $0 < H_1 \le \cdots \le H_M < 1$. The covariance matrix Σ has entries $(\Sigma)_{\ell,\ell'} = \sigma_\ell\sigma_{\ell'}\rho_{\ell,\ell'}$, with variances σ_ℓ^2 and correlation coefficients $\rho_{\ell,\ell'}$. OfBm is the Gaussian, stationary-increment process $\underline{B}_{P,\underline{H},\Sigma}(t) = P\underline{B}_{\underline{H},\Sigma}(t)$, where P is an invertible $M\times M$ matrix that mixes the fBm components, thus altering the scaling coordinates. It satisfies the operator self-similarity relation

$$\{\underline{B}_{P,\underline{H},\Sigma}(t)\}_{t\in\mathbb{R}} \overset{\text{f.d.d.}}{=} \{a^{\mathbf{H}}\underline{B}_{P,\underline{H},\Sigma}(t/a)\}_{t\in\mathbb{R}}, \quad \forall a>0, \tag{1}$$

where the matrix Hurst exponent is $\mathbf{H} = P\,\mathrm{diag}(\underline{H})P^{-1}$, and $a^{\mathbf{H}} = \sum_{k=0}^{\infty}(\log^k a)\mathbf{H}^k/k!$. When the demixed process $P^{-1}\underline{B}_{P,\underline{H},\Sigma}(t) = \underline{B}_{\underline{H},\Sigma}(t)$ has uncorrelated entries, the process is *instantaneously correlated*. Hereinafter, we focus exclusively on such cases, setting $\Sigma = I$.

2.2 Wavelet analysis of ofBm

Multivariate Wavelet Transform. Let ψ represent a mother wavelet, which is a real-valued function that satisfies the condition $\int_{\mathbb{R}}\psi^2(t)\,dt = 1$. For each $k,j\in\mathbb{N}$, the multivariate discrete wavelet transform of the sequence $\{\underline{B}_{P,\underline{H},\Sigma}(t)\}_{t\in\mathbb{R}}$ is given by $D(2^j,k) = (D_1(2^j,k),\dots,D_M(2^j,k))$, where $D_\ell(2^j,k) = \langle 2^{-j/2}\psi(2^{-j}t-k)|\underline{B}_{P,\underline{H},\Sigma,\ell}(t)\rangle \in \mathbb{R}$ for $\ell\in\{1,\dots,M\}$. For a detailed overview of wavelet transforms, see [11]. The wavelet coefficients $\{D(2^j,k)\}_{k\in\mathbb{Z}}$ can be proven to satisfy, for every fixed octave j, the operator self-similarity relation [2]

$$\{D(2^j,k)\}_{k\in\mathbb{N}} \overset{\text{f.d.d.}}{=} \{2^{j(\mathbf{H}+\frac{1}{2}I)}D(1,k)\}_{k\in\mathbb{N}}. \tag{2}$$

Eigenanalysis. Starting from the measurements of $\underline{B}_{P,\underline{H},\Sigma}$, the *wavelet random matrix* (WRM) at octave $j = j_1,\dots,j_2$ is represented by the symmetric $M\times M$ matrices

$$S(2^j) = \frac{1}{n_j}\sum_{k=1}^{n_j} D(2^j,k)D(2^j,k)^T, \tag{3}$$

where n denotes the sample size and $n_j \simeq n/2^j$ corresponds to the number of wavelet coefficients available at scale 2^j. In [2] it was shown that, except for the trivial case $P = I$, estimation based on the *entrywise* multiscale behavior of $S(2^j)$ generally introduces arbitrarily large bias.

Alternatively, consider the eigen-decompositions of the random matrices in (3), $S(2^j) = U_j\Lambda(2^j)U_j^T, \quad j = j_1,\dots,j_2$, where $\Lambda(2^j) = \mathrm{diag}(\lambda_1(2^j),\dots,\lambda_M(2^j))$ is the diagonal matrix of eigenvalues of $S(2^j)$, and the orthogonal matrix $U_j\in\mathbb{R}^{M\times M}$ contains the eigenvectors of $S(2^j)$ as columns. It was established in [2] that

$$\lambda_m(2^j) \approx \zeta_m\cdot 2^{j(2H_m+1)}, \tag{4}$$

in the large-sample and large-scale limits, as $n,j\to\infty$, where $\zeta_m > 0$ are the *scaling amplitude* constants. Consequently, the Hurst exponents $(H_1,\dots,H_M)$ can be efficiently estimated with the wavelet estimator $(\widehat{H}_1,\dots,\widehat{H}_M)$ defined by the weighted log-eigenvalue regression :

$$\widehat{H}_m = \frac{1}{2}\left(\sum_{j=j_1}^{j_2} w_j\log_2\lambda_m(2^j) - 1\right),\ m=1,\dots,M, \tag{5}$$

where the weights w_j satisfy the conditions $\sum_j jw_j = 1$ and $\sum_j w_j = 0$ (cf. [13]). Under mild assumptions, it was demonstrated in [2] that $(\widehat{H}_1,\dots,\widehat{H}_M) \overset{\mathbb{P}}{\to} (H_1,\dots,H_M)$ as $n,j\to\infty$, with asymptotically Gaussian fluctuations.

3 Estimation by multiscale demixing

3.1 Approximate demixing of WRMs

Although the scale-wise eigenanalysis-based estimation is very robust, a key conceptual difficulty arises from the fact that it orders the eigenvalues at each scale independently by magnitude. Due to discrepancies in scaling amplitudes (linear coefficients in a log-log plot), this can lead to the *crossover effect*, i.e., the incorrect permutation of log-scaling curves stemming from eigenvalue ordering. This introduces slope estimation biases when used in (5). This situation is illustrated in Fig. 1 (left column). Here, we propose a new estimator for $(H_1,\dots,H_M)$ that leverages the fact that, by linearity, if we knew the true mixing matrix P, we could *demix* and recover the scaling relation of the fBm components entry-wise, $\underline{B}_{\underline{H},\Sigma}(t) = P^{-1}\underline{B}_{P,\underline{H},\Sigma}(t)$. In a finite-sample setting, our key idea is to *jointly diagonalize* the matrices $S(2^j)$ *approximately*, using a single nonsingular estimation $\hat{P}$ for the unknown matrix P. This yields the factorization

$$S(2^j) = \hat{P}\hat{\Lambda}(2^j)\hat{P}^T. \tag{6}$$

In (6), $\hat{\Lambda}(2^j)$ is an approximately diagonal matrix whose diagonal entries approximately reproduce the scaling (4),

$$\hat{\Lambda}_{mm}(2^j) \approx \check{\zeta}_m 2^{j(2H_m+1)}, \quad ,m=1,\dots,M. \tag{7}$$

To construct the wavelet estimator for $(H_1,\dots,H_M)$, we thus replace the scale-wise eigenvalues $\lambda_m(2^j)$ in (5) with $\hat{\Lambda}_{mm}(2^j)$. As a key advantage, if $P^{-1}\hat{P}\approx I$, the approach (6) preserves the *ordering of scaling laws* by approximately restoring the scaling relations of the individual fBm components.

3.2 Multiscale estimation of mixing matrix

The key step in the proposed approach is to construct an accurate estimate of the mixing matrix P, or its inverse $V = P^{-1}$, up to nonidentifiability by scaling and permutation factors [6]. In [3], this was attempted via the exact diagonalization of two WRMs $S(2^j)$ at different scales $j = J_1, J_2$. While this method

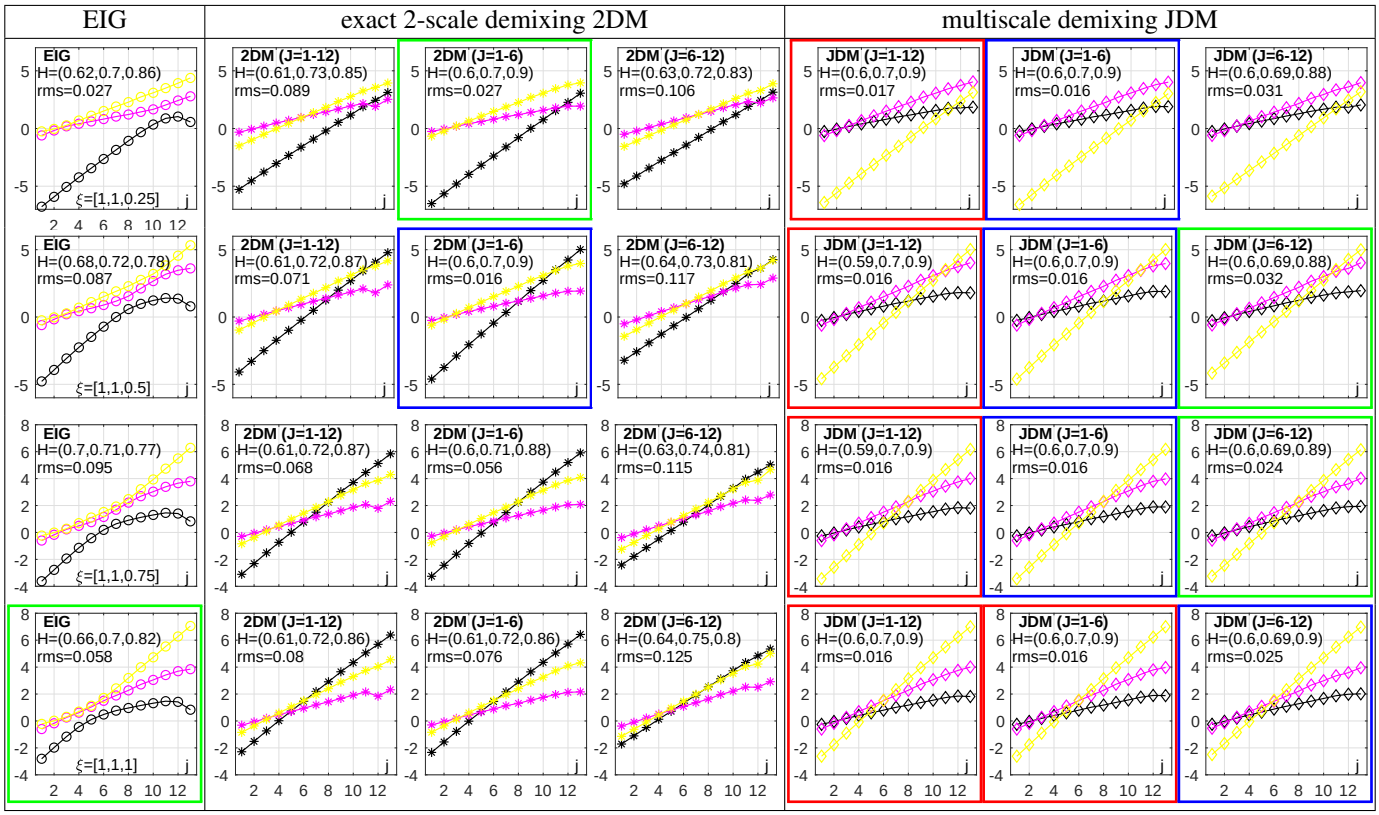

FIGURE 1 : **Log-scaling plots** for eigenvalues $\lambda_m(2^j)$ (EIG, left column), and $\hat{\Lambda}_{mm}(2^j)$ for 2DM (column 2 to 4) and JDM (column 5 to 7) computed for $(J_1, J_2) = (1, 12), (1, 6), (6, 12)$ (sub-column 1 to 3), respectively. From top to bottom : $\xi_3 = (0.25, 0.5, 0.75, 1)$, respectively for amplitude of 3rd fBm component (before mixing). Rms values for two scaling ranges are given in each subplot. Best, second best and third best results for each row are marked blue, red, green, respectively.

has shown good performance for several instances of scaling systems, it is limited by relying on only two scales of analysis, which may result in poor Hurst exponent estimates, especially when scaling laws cross over. Here we propose a more robust approach estimate P as an *approximate* diagonalizer for a *set* of WRMs $\mathbf{S} = \{S(2^{J_1}), \ldots, S(2^{J_2})\}$ for a *range of scales*. The problem of jointly diagonalizing a set of matrices has been extensively studied, and several algorithms have been proposed [6]. One possibility is to define the demixing matrix $\hat{V} = \hat{P}^{-1}$ as the minimizer of a functional

$$f_{\mathbf{S}}(V) = \frac{1}{2} \sum_{j=J_1}^{J_2} \sum_{i \neq j} \left| (V S(2^j) V^T)_{ij} \right|^2 \qquad (8)$$

that penalizes the magnitude of off-diagonal matrix, as $\hat{V} = \operatorname{argmin}_V f_{\mathbf{S}}(V)$. To avoid the trivial solution $V = 0$, and to fix the arbitrary scaling, V is further constrained to have unit column norm, i.e., V is a matrix on the oblique manifold. This approach was, for example, considered in [4], where expressions for the gradient, $\nabla f_{\mathbf{S}}|_V$, and Hessian operator, $H_{\mathbf{S}}|_V(W)$, were obtained and used in a trust region algorithm for solving (8). Here, we propose using these expressions in a conjugate gradient (CG) algorithm with a multiplicative change of basis and Newton step size recently put forth in a similar context in [12]. To this end, we adopt it for non-northogonal matrices V on the oblique manifold as follows, making use of the projection operator into the tangent space $\Pi_V(X) = X - V\mathrm{diag}(\mathrm{diag}(V^T X))$:

Initialization ($i = 0$) : Set $\mathbf{S}_0 = \mathbf{S}$ and $V_0 = I$.
Compute initial search direction $R_0 = -\Pi_V(\nabla f_{\mathbf{S}_0}|_I)$ and perform steps 4 to 7 (gradient descent step).

1. Set $\tilde{R}_{i-1} = (I + \lambda_{i-1} R_{i-1})^{-1} R_{i-1}$ given $R_{i-1}, \mathbf{S}_i$.
2. Compute gradient $\nabla f_{\mathbf{S}_i}|_I$ and $\tilde{\nabla} f_{\mathbf{S}_i}|_I = \Pi_V(\nabla f_{\mathbf{S}_i}|_I)$
3. Update search direction using Daniel's rule for nonlinear conjugate gradient, $R_i = \Pi_V(-\tilde{\nabla} f_{\mathbf{S}_i}|_I + \beta_{i-1} \tilde{R}_{i-1})$, where $\beta_{i-1} = \frac{\langle \tilde{\nabla} f_{\mathbf{S}_i}|_I, H_{\mathbf{S}_i}|_I(\tilde{R}_{i-1}) \rangle}{\langle \tilde{R}_{i-1}, H_{\mathbf{S}_i}|_I(\tilde{R}_{i-1}) \rangle}$ [7].
4. Compute Newton step-size $\lambda_i = -\frac{\langle \tilde{\nabla} f_{\mathbf{S}_i}|_I, R_i \rangle}{\langle R_i, H_{\mathbf{S}}|_I(R_i) \rangle}$.
5. Update matrix set $\mathbf{S}_{i+1} = ((I + \lambda_i R_i)\mathbf{S}_i (I + \lambda_i R_i)^T$.
6. Update estimate $V_{i+1} = (I + \lambda_i R_i) V_i$ and column-normalize.
7. If not converged, set $i = i + 1$ and go back to 1).

4 Numerical results

Monte Carlo Simulation. We apply the log-regression estimator in (5) through three different methods : the standard scale-wise eigenanalysis (*EIG*), demixing based on exact diagonalization of two scales as described in [3] (*2DM*), and the proposed multiscale demixing approach (*JDM*). These methods are evaluated over $N_{MC} = 100$ realizations of an ofBm process with $n = 2^{16}$, $M = 3$ and $H = (0.6, 0.7, 0.9)$. To enhance clarity in the log-log plots, and without affecting the generality of the analysis, we analyze the increment process of the ofBm. The mixing matrix is constructed as $P = P_\alpha A$, where P_α is a random orthonormal matrix that is modified

such that the last column forms an angle $\alpha = 66°$ with the second-to-last column, remaining orthogonal to all other columns, and $A = \mathrm{diag}(1, 1, \xi)$ modifies the scaling amplitude ζ_3 in (4) and hence the crossover of eigenvalues $\lambda_m(2^j)$; we study the cases $\xi_3 = (0.25, 0.5, 0.75, 1)$.

Estimation. For the analysis, we use a Daubechies wavelet with two vanishing moments and scaling range $(j_1, j_2) = (2, 12)$ for the regressions (5). For demixing (2DM and JDM), three different choices $(J_1, J_2) = (1, 12),\ (1, 6),\ (6, 12)$ are compared. The conjugate gradient (CG) algorithm is run for 1000 iterations on normalized matrices, $\tilde{S}(2^j) = S(2^j)/\|S(2^j)\|$. Performance is evaluated based on the mean, standard deviation (std), and root mean squared error (rms) calculated across realizations.

Performance for eigenvalue based estimation. Fig. 1 (left column) displays log-log plots of $\lambda_m(2^j)$ for four different scaling amplitudes (from top to bottom), the average estimations for each $H_m, m = 1, \ldots, M$, and the rms values (averaged over m). The scaling behavior is clearly affected by the inconsistent ordering of eigenvalues across scales (referred to as crossover), occurring both at finer scales ($j \approx 3$) and at coarser scales ($j \approx 11$ to $j \approx 5$ from top to bottom). This results in significant bias in the estimation, as indicated by large rms values up to approximately 0.1, with the exception of the first row, where the crossover happens at the largest scales.

Performance comparison for demixing. Fig. 1 presents log-log plots of $\hat{\Lambda}_{mm}(2^j)$ after demixing using 2DM (columns 2 to 4) and JDM (columns 5 to 7) for three different scale pairs, (J_1, J_2). The best, second best, and third best results, based on rms values, are highlighted with blue, red, and green frames, respectively. The results show that both 2DM and JDM effectively produce straight scaling laws for $\hat{\Lambda}_{mm}(2^j)$, free from crossovers, across all scaling amplitudes. However, a closer examination reveals that the estimates for H and rms values obtained with 2DM are generally worse than those from JDM in nearly all scenarios. In contrast, the JDM method consistently yields nearly unbiased estimates for H, with rms values significantly smaller than those from EIG and 2DM. Specifically, rms values are approximately 0.016 for all scenarios when $(J_1, J_2) = (1, 12)$ or $(1, 6)$, which is up to six times smaller compared to EIG and 2DM. When only the large scales are used for demixing ($(J_1, J_2) = (6, 12)$), the rms values are slightly larger but still outperform those of the other methods.

Overall, these findings demonstrate that the proposed JDM method showcases significant advantages over EIG and over 2DM, ensuring robust and improved performance for challenging data scenarios.

5 Conclusions

We introduced a multiscale demixing approach (JDM) for estimating Hurst exponents in multivariate self-similar processes, addressing issues arising from eigenvalue crossover and inconsistent ordering across scales. By applying approximate joint diagonalization of multiple wavelet random matrices, through a conjugate gradient algorithm, the proposed method effectively reconstructs the underlying scaling laws while preserving their relative ordering. Monte Carlo simulations demonstrated that JDM consistently outperforms both the standard eigenanalysis method and the two-scale demixing method, yielding nearly unbiased Hurst exponent estimates with significantly lower root mean squared errors. This highlights the robustness of the approach, particularly when a broad range of scales is utilized for demixing. These findings suggest that JDM provides a powerful alternative for multiscale analysis of self-similar processes. Future work includes extending this approach to more complicated instances of multivariate self-similarity – e.g., non-instantaneous correlations –, and studying the performance of the methodology in neuroscientific modeling.

Références

[1] P. ABRY, B. C. BONIECE, G. DIDIER et H. WENDT : On high-dimensional wavelet eigenanalysis. *Annals of Applied Probability*, 34(6):5287–5350, 2024.

[2] P. ABRY et G. DIDIER : Wavelet estimation for operator fractional Brownian motion. *Bernoulli*, 24(2):895, 2018.

[3] P. ABRY, G. DIDIER et H. LI : Two-step wavelet-based estimation for gaussian mixed fractional processes. *Stat. Inference Stochastic Process.*, 22:157–185, 2019.

[4] P.-A. ABSIL et K. A. GALLIVAN : Joint diagonalization on the oblique manifold for independent component analysis. *In IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, 2006.

[5] G. W. ANDERSON, A. GUIONNET et O. ZEITOUNI : *An Introduction to Random Matrices*. Numéro 118. Cambridge University Press, 2010.

[6] P. COMON et C. JUTTEN : *Handbook of Blind Source Separation : Independent Component Analysis and Applications*. Academic Press, 2010.

[7] J. W. DANIEL : The conjugate gradient method for linear and nonlinear operator equations. *SIAM J. Num. Anal.*, 4(1):10–26, 1967.

[8] G. DIDIER et V. PIPIRAS : Integral representations and properties of operator fractional Brownian motions. *Bernoulli*, 17(1):1–33, 2011.

[9] P. FLANDRIN : Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Info. Theory*, 38:910 – 917, 1992.

[10] J. HUALDE et P.M. ROBINSON : Semiparametric inference in multivariate fractionally cointegrated systems. *J. Econometrics*, 157(2):492–511, 2010.

[11] S. MALLAT : *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA, 1998.

[12] E. TROEDSSON, M. CARLSSON et H. WENDT : On gradient based descent algorithms for joint diagonalization of matrices. *In European Signal Processing Conference (EUSIPCO)*, Lyon, France, August 2024.

[13] D. VEITCH et P. ABRY : A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Info. Theory*, 45(3):878–897, 1999.

[14] M. J. WAINWRIGHT : *High-Dimensional Statistics : a Non-Asymptotic Viewpoint*. Cambridge U. Press, 2019.