#### **BLIND SOURCE SEPARATION FOR OPERATOR SELF-SIMILAR PROCESSES**

Hui Li (Advisor: Gustavo Didier)

Department of Mathematics, Tulane University

## 1. MOTIVATION

**Self-similarity and operator self-similarity.** An  $\mathbb{R}^n$ -valued signal  $\{\underline{Y}(t)\}$  is said to be operator self-similar (o.s.s) when it satisfies the scaling relation  $\{\underline{Y}(at)\}_{t\in\mathbb{R}} \stackrel{\text{fdd}}{=} \{a^H \underline{Y}(t)\}_{t\in\mathbb{R}}, a > 0$ , where fdd denotes the finite dimensional distributions, H is called the Hurst matrix, and  $a^H$  is given by the matrix exponential  $\sum_{k=0}^{\infty} \frac{\log^k(a) H^k}{k!}$ . O.s.s processes naturally generalize the univariate self-similar (s.s) processes, which have been used to model a wide range of phenomena in hydrodynamic turbulence [1], geophysics [2] and Internet traffic [3]. **Gaussian case.** The celebrated fractional Brownian motion (fBm) is the only Gaussian self-similar process with station-

(fBm) is the only Gaussian, self-similar process with stationary increments [4]. The natural multivariate generalization of fBm, called operator fBm (ofBm), was studied in [5, 6], and several papers are now devoted to inferential methods for of-Bm (e.g., [7, 8]).

**Non-Gaussian case.** Hermite processes are typically non Gaussian, self-similar, stationary increment processes. They appear as a consequence of non-central limit theorems. The Rosenblatt process (or fractional Rosenblatt motion, fRm) corresponds to the Hermite process of rank 2 [9]. Statistical inference for Hermite-type processes is particularly challenging: it was shown that wavelet-based estimators may display nonstandard convergence rates and asymptotic distributions in [10]. The modeling and statistical inference of multivariate Hermite-type fractional signals, while of great importance in applications, is an essentially unexplored research topic.

**Outline.** This thesis is dedicated to the problem of demixing (multivariate) fractional signals, i.e., extracting a source of independent s.s signals from a set of mixed signals. This blind source problem has been little studied under the framework of fractional stochastic processes. The work considers mainly two types of signals: Gaussian and non-Gaussian (Rosenblatt-type). The basic philosophy is to apply the same wavelet-based demixing procedure to these two types of signals (Section 2). Large size Monte Carlo simulations are used to illustrate that the finite-sample performance of the demixing method is very satisfactory (Section 4). Moreover, the asymptotic properties of the estimators will be studied for Gaussian and non-Gaussian cases separately (Section 3).

#### 2. METHODOLOGY

**Blind source separation.** For p = 1, ..., P, let  $X_{H_p}$  be a stationary-increment, s.s process with Hurst parameter  $H_p \in (0, 1)$ . Let W be a  $P \times P$  invertible matrix,  $\underline{H} = (H_1, ..., H_P)^T$  and  $^T$  denotes the vector transpose. Then, the process  $\{\underline{Y}(t)\}$  obtained by  $\underline{Y}(t) = W(X_{H_1}(t), ..., X_{H_P}(t))^T$  is o.s. with Hurst matrix  $H_W = W \text{diag}(\underline{H})W^{-1}$ . In practice, we only observe the mixed process  $\underline{Y}$  and need to estimate both W and H.

Wavelet-based demixing procedure. The wavelet transform of an o.s.s process  $\underline{Y}(t)$  is defined by

$$\mathbb{R}^P \ni D(2^j, k) = 2^{-j/2} \int_{\mathbb{R}} 2^{-j/2} \psi(2^{-j}t - k) \underline{Y}(t) dt, \quad (1)$$

 $j \in \mathbb{N} \cup \{0\}, k \in \mathbb{Z}$ . The mother wavelet function  $\psi(t)$  is characterized by its number of vanishing moments  $N_{\psi}$ , i.e., the integer  $N_{\psi}$  such that  $\forall n = 0, \ldots, N_{\psi} - 1, \int_{\mathbb{R}} t^k \psi(t) dt = 0$  and  $\int_{\mathbb{R}} t^{N_{\psi}} \psi(t) dt \neq 0$  [11]. The wavelet spectrum at scale  $2^j$  is the positive definite matrix  $\mathbb{E}D(2^j, 0)D(2^j, 0)^* =: \mathbb{E}W(2^j)$ , and its natural estimator, the sample wavelet transform, is the matrix statistic

$$W(2^j) = \frac{1}{n_j} \sum_{k=1}^{n_j} D(2^j, k) D(2^j, k)^*, \quad n_j = \frac{\nu}{2^j}, \quad (2)$$

where  $\nu$  is the number of sample points. By operator selfsimilarity, we can write that

$$\mathbb{E}W(2^{j}) = W\mathcal{E}^{1/2} \operatorname{diag}(2^{2jH_{1}}, \dots, 2^{2jH_{P}})\mathcal{E}^{1/2}W^{*} \quad (3)$$

where  $\mathcal{E} = \text{diag}(\eta(H_1), \dots, \eta(H_P))$ , and for  $i = 1, \dots, P$ ,

$$\eta(h_i) = -\sigma_i^2 \int_{\mathbb{R}^2} \psi(t)\psi(t') \frac{|t-t'|^{2H_i}}{2} dt dt'.$$

Equation (3) illustrates the fact that the matrices  $\mathbb{E}W(2^j)$  can be jointly diagonalized. In [12], we applied an exact joint diagonalization method to arrive at an estimator of the demixing matrix  $\widehat{W^{-1}}$ . The algorithm can be cast in the form of pseudocode as follows:

**Step 0:** From one observation  $\underline{Y}$ , sample wavelet variances  $W_Y(2^{J_1})$  and  $W_Y(2^{J_2})$  are computed as in Eq. 2 for suitable  $J_1$  and  $J_2$ ;

Step 1: Set  $\Theta = W_Y(2^{J_1})^{-1/2}$ ;

Step 2: Find eigenvectors Q of the matrix  $\Theta W_Y(2^{J_2})\Theta^*$ ; Step 3: Compute the demixing matrix  $B := Q\Theta$ . The estimator of the demixing matrix  $\widehat{W^{-1}}$  is defined to be the output of the demixing algorithm. Then, the demixed process  $\{\underline{Z}(t)\}$  is obtained by  $\underline{Z}(t) = \widehat{W^{-1}}\underline{Y}(t)$ , where (entrywise) univariate wavelet regression can be applied to estimate the Hurst parameters  $H_1, \ldots, H_P$  [13].

# 3. ASYMPTOTIC THEORY

**Gaussian case.** For p = 1, ..., P, let  $X_{H_p}$  be a fBm with Hurst parameter  $H_p \in (0, 1)$ . Then, the mixed signal  $\underline{Y}(t) = W(X_{H_1}(t), ..., X_{H_P}(t))^T$  is an ofBm. The sample wavelet variance (2) was shown to be asymptotically Gaussian [7]. By building upon this result, we proved that the asymptotic distributions of the eigenvalues/vectors of the sample wavelet variance  $W_Y(2^j)$  of  $\underline{Y}(t)$ , as well as that of the demixing matrix estimator  $\widehat{W^{-1}}$  are Gaussian ([14]).

**Non-Gaussian case (work in progress).** Suppose every  $X_{H_p}$  is a fRm with Hurst parameter  $H_p \in (\frac{1}{2}, 1)$ , we define the multivariate operator fractional Rosenblatt motion (ofRm) as the  $\mathbb{R}^{P}$ - valued process  $\underline{Y}(t) = W(X_{H_1}(t), \ldots, X_{H_P}(t))^T$ . We will study the asymptotic distributions of the wavelet variance and their eigenvalues/vectors, as well as that of the demixing matrix estimator. In particular, the convergence rate expected to depend on the underlying Hurst eigenvalues. Moreover, we will also study the case where the unmixed, original signals are independent fBms and fRms, i.e., it combines Gaussian and non-Gaussian components.

### 4. SIMULATION PERFORMANCE

Broad Monte Carlo studies showed that the demixing method works very well, over finite-samples, for both Gaussian and non-Gaussian instances, or combinations thereof. This can be illustrated based on the case where P = 4, the sample path size is  $\nu = 2^{16}$ , and the original signal is made up of 4 independent s.s processes (2 fBms with Hurst parameters 0.2 and 0.4 and 2 fRms with Hurst parameters 0.6 and 0.8), the mixed process was given by  $\underline{Y} = W\underline{X}$  (X is then a mixture of fBms and fRms). Figure 1 shows boxplots for the Monte Carlo distributions for each of the  $4 \times 4$  entries of  $\widehat{W^{-1}} - W^{-1}$ . This indicates that  $W^{-1}$  is remarkably well estimated with negligible biases. This results are reported for one arbitrarily chosen matrix W, since the performance for all instances of W was comparable. Monte Carlo studies not included further indicate that the demixing method also works well for pure ofBm and ofRm, and remains efficient even for relatively small sample size (down to  $\nu = 2^{10}$ ).

## 5. REFERENCES

- B.B. Mandelbrot, "Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier," J. Fluid Mech., vol. 62, pp. 331–358, 1974.
- [2] E. Foufoula-Georgiou and P. Kumar, Eds., *Wavelets in Geophysics*, Academic Press, San Diego, Calif., 1994.



Fig. 1. Boxplots for the  $4 \times 4$  entries of  $\widehat{W^{-1}} - W^{-1}$ .

- [3] M.S. Taqqu, W. Willinger, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," *SIGCOMM Comput. Commun. Rev.*, vol. 27, no. 2, pp. 5–23, 1997.
- [4] M. S. Taqqu, "Fractional Brownian motion and long range dependence," in *Theory and Applications of Long-Range Dependence (P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.)*, pp. 5–38. Birkhäuser, Boston, 2003.
- [5] M. Maejima and J. Mason, "Operator-self-similar stable processes," *Stochastic Processes and their Applications*, vol. 54, pp. 139–163, 1994.
- [6] G. Didier and V. Pipiras, "Integral representations and properties of operator fractional Brownian motions," *Bernoulli*, vol. 17, no. 1, pp. 1–33, 2011.
- [7] P. Abry and G. Didier, "Wavelet estimation for operator fractional Brownian motion," To appear in *Bernoulli*, pp. 1–30, 2016.
- [8] J.-F. Coeurjolly, P.-O. Amblard, and S. Achard, "Wavelet analysis of the multivariate fractional Brownian motion," *ESAIM: Probability and Statistics*, vol. 17, pp. 592–604, 2013.
- [9] M. S. Taqqu, "Convergence of integrated processes of arbitrary Hermite rank," Z. Wahrscheinlinchkertstheor. Verwandte Geb., vol. 50, no. 1, pp. 53–83, 1979.
- [10] J.-M. Bardet and C. Tudor, "A wavelet analysis of the Rosenblatt process: chaos expansion and estimation of the selfsimilarity parameter," *Stochastic Processes and their Applications*, vol. 120, no. 12, pp. 2331–2362, 2010.
- [11] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1998.
- [12] R.A. Horn and C.A. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [13] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," *IEEE Trans. Inform. Theory*, vol. 45, no. 3, pp. 878–897, 1999.
- [14] P. Abry, G. Didier, and H. Li, "Demixing operator fractioanl Brownian motion," http://arxiv.org/pdf/1607.05167v1.pdf, pp. 1-25, 2016.