Correspondence

PCA Based Hurst Exponent Estimator for fBm Signals Under Disturbances

Li Li, Jianming Hu, Yudong Chen, and Yi Zhang

***Abstract*—In this paper, the validity of PCA eigenspectrum based Hurst exponent estimator proposed in[J. B. Gao, Y. Cao, and J.-M. Lee, "Principal Component Analysis of $1/f^{\alpha}$ noise," *Phys. Lett. A*, vol. 314, no. 5–6, pp. 392–400, 2003] for single fBm signal is proved. Moreover, how to apply this estimator for fBm signals corrupted with some other signals are discussed. Theoretical analysis and experiments show that it can also be used for 1) mixed fBm signals with different Hurst exponents, 2) fBm signals corrupted with additive Gaussian white noise when the signal-to-noise ratio (SNR) is not too small, and 3) fBm signals corrupted with additive deterministic sine/cosine signals. However, the estimation accuracy depends on the SNR value for the first two situations.**

***Index Terms*—Fractal Brownian motion (fBm), Hurst exponent, principal component analysis (PCA).**

I. Introduction

Fractional Brownian motion (fBm) is a simple and useful model that demonstrates long-range dependencies (LRD) and $1/f^{\alpha}$-type spectral behaviors. It has attracted increasing interests during the last four decades [1], because it is found to be prevalent in natural and man-made systems.

In many practical applications, we need to estimate the only parameter that describes the complexity of the data, the Hurst exponent H, in the context of signal processing [2]–[5]. Various methods have been proposed to estimate H, including those based upon power spectral density (PSD) and discrete wavelet transform (DWT). However, the estimation may still be inaccurate sometimes, especially when the analyzed signals are mixed with other signals.

Recently, a new Hurst estimator was proposed in [6]. It was conjectured that the eigenvalues from principal component analysis (PCA) of fBm processes decay as a power-law, with the exponent being $-(2H+1)$. Since the PCA eigenspectrum is usually smoother than the power spectrum, we may get more accurate and robust estimation results of H than those PSD-based methods.

Manuscript received May 29, 2008; accepted January 28, 2009. First published March 10, 2009; current version published June 17, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Lucas Parra. This work is supported in part by the National Basic Research Program of China (973 Project) 2006CB705506, the Hi-Tech Research and Development Program of China (863 Project) 2007AA11Z222, the National Science & Technology Pillar Program (11.5 Program) 2006BAJ18B02, and the National Natural Science Foundation of China 60774034, 50708055.

L. Li, J. Hu, and Y. Zhang are with the Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Automation, Tsinghua University, Beijing 100084, China (e-mail: li-li@mail.tsinghua.edu.cn; hujm@mail.tsinghua.edu.cn;) zhyi@mail.tsinghua.edu.cn).

Y. Chen is with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: ydchen@mail.utexas.edu).

Digital Object Identifier 10.1109/TSP.2009.2016877

However in [6], this conjecture is only analytically proved for regular Brownian motion ($H = 1/2$, that is, Wiener process) by using the famous continuous Karhunen–Loève (K-L) expansion of Wiener process. In the following studies [7]–[9], this conjecture is analytically proved for $0 < H < 1/2$, too. But the accuracy of PCA based method was still questioned for $1/2 < H < 1$.

To resolve the interesting problem, we first discuss the relations between different representation types of fBm, continuous/discrete K-L transformation and PCA in a rigorous way. Based on these backgrounds, a proof of the conjecture for $0 < H < 1$ is then provided. Furthermore, We study the performance of this estimator for some corrupted fBm signals, e.g., mixed fBm signals with different Hurst exponents, fBm signals corrupted with additive Gaussian white noise, and fBm signals corrupted with additive deterministic sine/cosine signals. Experiments indicate that it may still be used when the fBm signals/components are dominant. However, the estimation results needs to be carefully examined.

II. The PCA Spectrum of Single fBm Signals

A. Brief Overview of the Backgrounds

For simplicity, we focus on the standard fBm process in this section but the conclusions can be easily extended to general fBm processes. Suppose $B = \{B_t, t \geq 0\}$ is a standard fBm process with Hurst exponent H ($0 < H < 1$). Thus, it has zero-mean, continuous sample paths with autocorrelation function [1], [10], [11]

$$R_b(t_1, t_2) = E\left[B_{t_1} B_{t_2}\right] = \frac{1}{2}\left({t_1}^{2H} + {t_2}^{2H} - |t_1 - t_2|^{2H}\right). \quad (1)$$

Notice that the standard fBm is a centered process ($E[B_t] = 0$), the auto-covariance K_b is identical to the autocorrelation R_b here. From this definition, we may have different representation types of B_t, e.g., series expansions, spectral representations, and integral representations [11]. For example, according to the theory of reproducing kernel space and continuous Karhunen-Loève expansion, we can expand B_t within a finite time interval $[0, T]$ as a linear combination of orthogonal basis functions [12]

$$B_t = \sum_{n=1}^{\infty} c_n \phi_n(t) \quad (2)$$

where $\phi_n(t)$ belongs to a set of orthonormal functions in the interval $[0, T]$,

$$\int_0^T \phi_n(t)\phi_m^*(t)dt = \delta[n-m] \quad (3)$$

where $*$ denotes complex conjugate. The coefficients c_n are pairwise uncorrelated random variables. For any sample path $\bar{B}_t$, the corresponding $\bar{c}_n$ can be determined by[1]

$$\bar{c}_n = \int_0^T \bar{B}_t \phi_n^*(t)dt. \quad (4)$$

[1]Throughout this paper, we avoid using Riemann–Stieltjes integral for conceptual convenience.

Since the covariance function $R_b(t_1, t_2)$ is jointly continuous and symmetric within $[0, t_1 \wedge t_2]$, it can be regarded as a positive definite kernel. Based on spectral theory, we can apply the following Mercer's theorem [13].

Lemma 1 (Mercer's theorem): Suppose K is a continuous symmetric nonnegative definite kernel, then there is a set of orthonormal basis functions (eigenfunctions) $\psi_n(t)$ of $L^2[0,T]$ consisting of eigenfunctions of a certain operator Q_K such that the corresponding sequence of eigenvalues λ_n is nonnegative. The eigenfunctions corresponding to nonzero eigenvalues are continuous on $L^2[0,T]$ and K has the following representation, where the convergence is absolute and uniform:

$$K(s,t) = \sum_{n=1}^{\infty} \lambda_n \psi_n(s) \psi_n^*(t), \quad 0 < s,t < T. \tag{5}$$

According to Mercer's theorem, we have the following relations from (3)–(5) (see also [14]–[17]):

$$R_b(t_1,t_2) = \sum_{n=1}^{\infty} \lambda_n \phi_n(t_1) \phi_n^*(t_2) \tag{6}$$

$$E\{c_n c_m^*\} = \lambda_n \delta[n-m] \tag{7}$$

where λ_n are the corresponding eigenvalues of the nth orthonormal functions.

Equivalently, the above continuous K-L expansion can be characterized by the following integral equation:

$$\int_0^T R_b(t_1,t_2) \phi_n^*(t_2) dt_2 = \lambda_n \phi_n(t_1). \tag{8}$$

The functions ϕ_n in (6) equal to those in (2)–(4). According to (7), if we have a series expansion of B_t like (2), then the variance of its coefficients equals the eigenvalues in (6).

In applications, we often need to carry out the discrete Karhunen-Loève transformation (decompositions) from empirical sampled data instead of the continuous K-L transformations. Particularly, we will focus on the real-valued fBm processes, since PCA based H estimator are usually applied to real-valued fBm processes.

The estimation of the eigenfunctions of the continuous integral (8) from the associated discrete version (9) generated by a finite sample of M data points was discussed in [18]–[22]

$$\frac{1}{M} \sum_{l=1}^{M} \tilde{R}_b(e_k, e_l) \tilde{\phi}_m(e_l) = \tilde{\lambda}_m \tilde{\phi}_m(e_k) \tag{9}$$

where M is the truncation length. $\tilde{\phi}_m \approx \phi_m$ can be taken as the discretized eigenfunction (also called eigenvectors), which span a M-dimensional inner product space.

There are several methods to obtain the approximate numerical solution of (8) based on the discretized form of (5). Especially, we have the following lemma [22].

Lemma 2: Suppose we perform an eigenvalue decomposition on the $M \times M$ dimensional auto-covariance matrix (Gram matrix) $\tilde{R}_b(e_k, e_l)$, where the eigendecomposition satisfies $\tilde{R}_b(e_k, e_l) U = US$ (the columns of the $M \times M$ matrix U are the eigenvectors of the Gram matrix and the diagonal matrix S contains the corresponding eigenvalues $\tilde{\lambda}_m$). When M is large enough, we can then have an approximate estimation of the actual eigenvalues as

$$\tilde{\lambda}_m \approx M \lambda_m. \tag{10}$$

Finally, we will discuss the relations between discrete K-L expansion and PCA. Mathematically, PCA can be viewed as an orthogonal linear transformation that transforms the sampled data to a new coordinate such that the largest variance by any projection of the data comes to lie on the first principal direction, then the second greatest variance on the second principal direction, and so on. This technique is also known as Hotelling transformation. In practice, PCA can be executed via singular value decomposition (SVD) and is fast enough when handling large M. Noticing discrete K-L and PCA are two orthogonal decompositions operating in the same M-dimensional inner product space, we have the following lemma [18], [22], [23].

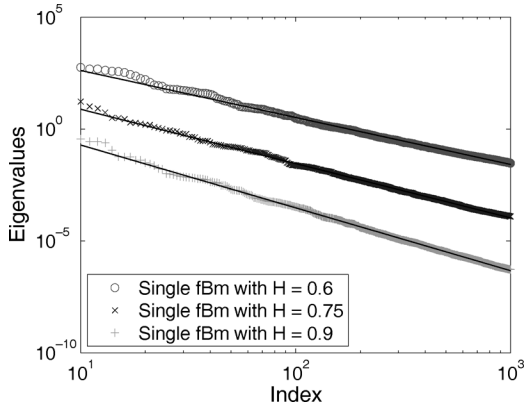

Fig. 1. Examples of PCA eigenspectrum for single fBm signals with $1/2 < H < 1$: (a) true value $H = 0.6$, estimated value $\hat{H} = 0.5725$; (b) true value $H = 0.75$, estimated value $\hat{H} = 0.7267$; and (c) true value $H = 0.9$, estimated value $\hat{H} = 0.8877$.

Lemma 3: The eigenvalues obtained from the discrete K-L expansions is equivalent to the eigenvalues obtained from PCA of the same auto-covariance matrix.

B. The Hurst Exponent Estimator

Based on the aforementioned backgrounds, we have the following relation between the Hurst exponent H of a fBm process and its PCA spectrum.

Theorem 1: When the sampling data length M is a sufficiently large constant, the PCA eigenvalue spectrum of a fBm process with Hurst exponent H decays as a power-law

$$\tilde{\lambda}_m \approx M \lambda_m \sim m^{-(2H+1)}, \quad m = 1, \ldots, M. \tag{11}$$

A detailed proof of this Theorem is presented in the Appendix.

Theorem 1 indicates that when plotted in a log–log scale, the variation of the eigenvalue spectrum with the index is approximately a straight line with slope $\alpha = -(2H+1)$. Thus, we can estimate H based on this property, since the PCA eigenspectrum is usually smoother than the power spectrum. The test fBm signals used below are generated by using the method proposed in [24]. Particularly, Fig. 1 shows some numerical verifications for single fBm signals with $1/2 < H < 1$. Table I shows the means and variances of the estimated $\hat{H}$ from 50 samples of each single fBm processes, respectively. In the next section, we will focus on the performance of this estimator for some fBm signals corrupted with other signals.

III. PCA Based Hurst Estimator for Corrupted fBm Signals

A. Mixed fBm Signals With Different Hurst Exponents

Mixed fBm signals had been discussed in several literatures, e.g., [25]–[28]. Here, we only discuss the mixture of two standard fBm signals for simplicity. Let's consider $\breve{B}_t$ which consists of two additive independent standard fBm signals $B_{1,t}$, $B_{2,t}$ with Hurst exponents H_1 and H_2, respectively

$$\breve{B}_t = \kappa_1 B_{1,t} + \kappa_2 B_{2,t} \tag{12}$$

TABLE I
MEANS AND VARIANCES OF THE ESTIMATED $\hat{H}$ FROM50 SAMPLES OF EACH SINGLE FBM PROCESSES, RESPECTIVELY

H	Mean of $\hat{H}$	Variance of $\hat{H}$
0.1	0.0658	4.6378×10^{-4}
0.2	0.1587	3.6230×10^{-4}
0.3	0.2836	3.5273×10^{-4}
0.4	0.3910	3.8746×10^{-4}
0.5	0.5072	3.7745×10^{-4}
0.6	0.5570	9.8238×10^{-5}
0.7	0.6690	1.7242×10^{-4}
0.8	0.7798	1.4205×10^{-4}
0.9	0.8869	1.0183×10^{-4}

where κ_1 and κ_2 are two positive scaling coefficients that define the mixture ratio/proportion of the signals. Without losing any generality, we assume $H_1 < H_2$ here.

Generally, the autocorrelation of the sum of two completely uncorrelated functions (the cross-correlation is zero for all time lag) is the sum of the autocorrelations of each function separately. Thus, we can have the corresponding series expansion of $\breve{B}_t$ as

$$\breve{B}_t = \sum_{n=1}^{\infty} (\kappa_1 c_{1,n} + \kappa_2 c_{2,n}) \phi_n(t) \tag{13}$$

where $c_{1,n}$ and $c_{2,n}$ are pairwise uncorrelated random variables satisfying $E\{c_{1,n}^2\} = \lambda_{1,n}$, $E\{c_{2,n}^2\} = \lambda_{2,n}$. Here $\lambda_{1,n}$ and $\lambda_{2,n}$ denote the corresponding eigenvalues of $B_{1,t}$ and $B_{2,t}$ respectively.

Because of independence, we can have

$$\begin{aligned} \lambda_{1+2,n} &= E\{(\kappa_1 c_{1,n} + \kappa_2 c_{2,n})^2\} = \kappa_1^2 \lambda_{1,n} + \kappa_2^2 \lambda_{2,n} \\ &\sim \kappa_1^2 n^{-(2H_1+1)} + \kappa_2^2 n^{-(2H_2+1)}. \end{aligned} \tag{14}$$

If $\kappa_1 \gg \kappa_2$, we have $\kappa_1^2 n^{-(2H_1+1)} \gg \kappa_2^2 n^{-(2H_2+1)}$, and thus

$$\lambda_{1+2,n} \sim \kappa_1^2 n^{-(2H_1+1)}. \tag{15}$$

If $\kappa_1 \ll \kappa_2$, there exists a positive solution (may be not an integer) for the following equation of variable n

$$\kappa_1^2 n^{-(2H_1+1)} = \kappa_2^2 n^{-(2H_2+1)}. \tag{16}$$

Let's denote the solution as n^*. It is easy to find

$$\lambda_{1+2,n} \sim \begin{cases} \kappa_2^2 n^{-(2H_2+1)}, & \text{for } n \ll n^* \\ \kappa_1^2 n^{-(2H_1+1)}, & \text{for } n \gg n^*. \end{cases} \tag{17}$$

This indicates that the PCA eigenspectrum of the mixed signal are the linear addition of the PCA eigenspectrum of two fBm signals. In a log-log scale, if the $\kappa_1 \gg \kappa_2$, the eigenvalue spectrum of the mixed signal would be one straight line whose slope is $\alpha = -(2H_1 + 1)$; otherwise the spectrum crossovers from $-(2H_2 + 1)$ to $-(2H_1 + 1)$. More precisely, when $\kappa_1 \ll \kappa_2$, the eigenvalue spectrum of the mixed signal consists of three parts: one line segment with $\alpha = -(2H_2 + 1)$ when n is small, the intermediate transition part, and another line segments with $\alpha = -(2H_1 + 1)$ when n is large. The above conclusions can be easily extended to the mixture of more than 2 fBm signals. Some typical examples can be found in Fig. 2.

Table II shows some typical estimated means and variances of $\hat{H}_1$, when $\kappa_1 > \kappa_2$. As illustrated in Fig. 2(b), the PCA eigenspectrum of the mixed signal is dominated by the fBm signal with H_1 under such conditions and we can hardly detect the existence of the fBm signal with H_2 from the PCA eigenspectrum of the mixed signal. Results show that the estimation error of the dominant Hurst exponent $\hat{H}_1$ caused by such disturbance is relatively small.

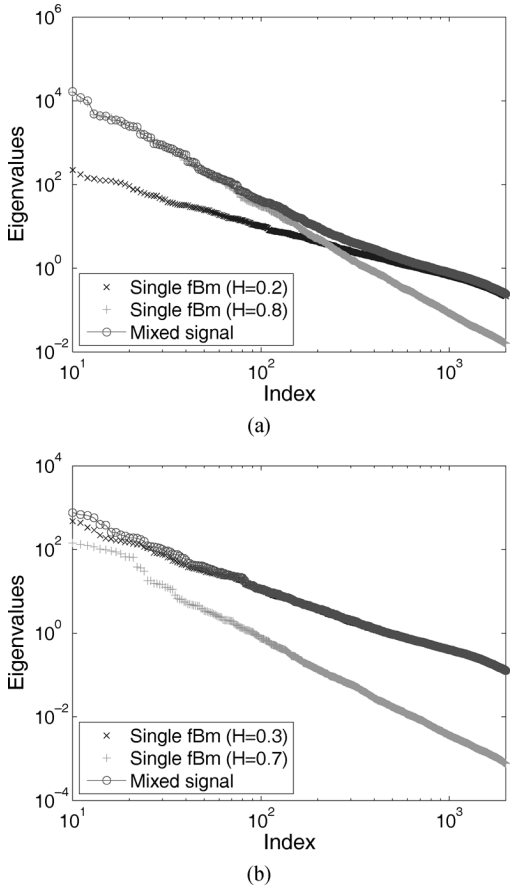

Fig. 2. PCA eigenspectrum of the mixed fBm signal, where (a) $H_1 = 0.2$, $H_2 = 0.8$, $\kappa_1/\kappa_2 = 2/3$, (b) $H_1 = 0.3$, $H_2 = 0.7$, $\kappa_1/\kappa_2 = 4$.

TABLE II
MEANS AND VARIANCES OF THE ESTIMATED $\hat{H}_1$ FROM50 SAMPLES OF TWO MIXED FBM SIGNALS RESPECTIVELY, WHERE $\kappa_1 > \kappa_2$

H_1	H_2	κ_1/κ_2	Mean of $\hat{H}_1$	Variance of $\hat{H}_1$
0.3	0.4	2	0.2660	5.0713×10^{-4}
0.3	0.5	2	0.2990	6.0350×10^{-4}
0.3	0.6	2	0.2879	6.2217×10^{-4}
0.3	0.7	2	0.3202	5.4289×10^{-4}
0.3	0.8	2	0.3071	6.4341×10^{-4}
0.3	0.9	2	0.3088	9.3707×10^{-4}
0.5	0.6	2	0.5035	1.0414×10^{-3}
0.5	0.7	2	0.5176	9.0487×10^{-4}
0.5	0.8	2	0.5282	9.9433×10^{-4}
0.5	0.9	2	0.5349	9.4357×10^{-4}
0.7	0.8	2	0.7190	8.4905×10^{-4}
0.7	0.9	2	0.7255	9.3461×10^{-4}

Simulations show that when $H_1 \approx H_2$, the existence of the fBm signal with H_2 would be hard to detect, since it may be recognized as disturbance from noise or sampling. Table III shows some typical estimated means and variances of $\hat{H}_1$ and $\hat{H}_2$, when $\kappa_1 < \kappa_2$ and especially the crossover phenomena is notable to be identified (H_1 is

TABLE III
MEANS AND VARIANCES OF THE ESTIMATED $\hat{H}_1$ FROM50 SAMPLES OF TWO MIXED FBM SIGNALS RESPECTIVELY, WHERE $\kappa_1 < \kappa_2$

H_1	H_2	κ_1/κ_2	Mean of $\hat{H}_1$ and $\hat{H}_2$	Variance of $\hat{H}_1$ and $\hat{H}_2$
0.2	0.5	1/1.1	0.1866, 0.4608	$2.3384, 6.9200 \times 10^{-3}$
0.2	0.6	1/1.1	0.1904, 0.5804	$1.7368, 6.8930 \times 10^{-3}$
0.2	0.7	1/1.1	0.1987, 0.7170	$1.8411, 9.6574 \times 10^{-3}$
0.2	0.8	1/1.1	0.1944, 0.8182	$1.6766, 6.4438 \times 10^{-3}$
0.2	0.9	1/1.1	0.1947, 0.8726	$2.0258, 5.9505 \times 10^{-3}$

notably different from H_2). As illustrated in Fig. 2(a), the estimation accuracy of $\hat{H}_2$ increases with the value of κ_2/κ_1. Tests also show that computing the eigenvalue spectrum at large n usually requires larger memory space and more computational cost to perform PCA. Thus, it would be actually difficult to estimate $\hat{H}_1$ when $\kappa_1 \ll \kappa_2$. Anyway, it should be pointed out that we are usually more interested in the existence of the crossover phenomena itself, since this indicates the coexistences/mixture of two different fBm signals with different Hurst exponents. And this might be a reasonable explanation for some important empirical data series (i.e., traffic flows in urban road networks) yielding complex crossover phenomena.

B. fBm Signals Corrupted With Additive Gaussian White Noise

The fBm signals corrupted by additive Gaussian white noise has also been discussed in several reports, e.g., [29]. For simplicity, let us consider $\breve{B}_t$ which consists of a standard fBm signal $B_{1,t}$ with Hurst exponent H_1 and another independent standard Gaussian white noise G_t (zero mean and unit variance)

$$\breve{B}_t = \kappa_3 B_{1,t} + \kappa_4 G_t \tag{18}$$

where κ_3 and κ_4 are two positive scaling coefficients that define the mixture ratio/proportion of the signals (the SNR here). Usually, we have $\kappa_3 > \kappa_4$.

By definition, we know that the autocorrelation function of white noise is a unitary matrix which satisfies

$$R_G(t_1, t_2) = E\{G_{t_1} G_{t_2}\} = \frac{1}{2}\delta(t_1 - t_2). \tag{19}$$

Thus, the numerical eigenspectrum of the Gaussian white noise should be a straight line with slope $\alpha_0 \approx 0$ in the log-log scale (the slope is not strictly 0 because of the finite sampling length effect). And the corresponding eigenspectrum is $\lambda_{4,n} = 1/2$.

Because of independence, we can have

$$\lambda_{3+4,n} = \kappa_3^2 \lambda_{3,n} + \kappa_4^2 \lambda_{4,n} \sim \kappa_3^2 n^{-(2H_1+1)} + \frac{1}{2}\kappa_4^2 K_1 \tag{20}$$

where $\lambda_{3,n}$ and $\lambda_{4,n}$ denote the corresponding eigenvalues of $B_{1,t}$ and G_t respectively. K_1 is a constant.

Similar to the above discussions, we can see if the $\kappa_3 \gg \kappa_4$ (and also $\kappa_3 \gg \kappa_4 K_1$), the eigenvalue spectrum of the mixed signal crossovers from $\alpha_1 = -(2H_1 + 1)$ to α_0; otherwise when $\kappa_3 \ll \kappa_4$ (and also $\kappa_3 \ll \kappa_4 K_1$), the eigenvalue spectrum of the signal would be one straight line whose slope is $\alpha = \alpha_0$ and we can hardly estimate H_1 accurately in such cases via this method. Two typical examples can be found in Fig. 3. More simulation results are given in Table IV, which indicates that the estimator error (partly reflected as the estimation variance) increases when the SNR value decreases. Further simulations show that the estimation error will be generally unacceptable when $\kappa_3/\kappa_4 < 1$.

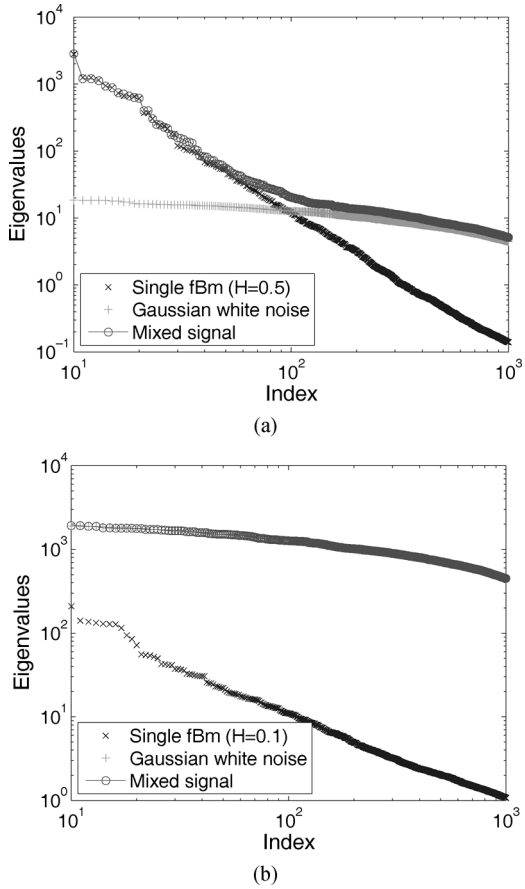

Fig. 3. PCA eigenspectrum of the fBm signal corrupted with Gaussian white noise, where (a) $H = 0.5$, $\kappa_3/\kappa_4 = 1/2$, (b) $H = 0.1$, $\kappa_3/\kappa_4 = 1/20$.

TABLE IV
MEANS AND VARIANCES OF THE ESTIMATED $\hat{H}$ FROM50 SAMPLES OF EACH SIGNAL FBM SIGNALS CORRUPTED BY WHITE NOISE, RESPECTIVELY

H	κ_3/κ_4	Mean of $\bar{H}$	Variance of $\bar{H}$
0.1	4	0.0958	3.3409×10^{-3}
0.2	4	0.2045	4.6963×10^{-3}
0.3	4	0.2967	4.7829×10^{-3}
0.4	4	0.4292	3.8613×10^{-3}
0.5	4	0.5101	4.8498×10^{-3}
0.6	4	0.6191	4.9724×10^{-3}
0.7	4	0.7374	9.1588×10^{-3}
0.8	4	0.8381	6.1777×10^{-3}
0.9	4	0.9611	6.7738×10^{-3}
0.1	20	0.0716	6.3355×10^{-3}
0.2	20	0.1560	1.5059×10^{-3}
0.3	20	0.2900	9.9115×10^{-4}
0.4	20	0.3911	1.0694×10^{-3}
0.5	20	0.5071	1.4522×10^{-3}
0.6	20	0.5995	1.6151×10^{-3}
0.7	20	0.7022	1.7367×10^{-3}
0.8	20	0.8129	1.7400×10^{-3}
0.9	20	0.9053	2.2848×10^{-3}

C. fBm Signals Corrupted With Additive Deterministic Signals

Sometimes we need to analyze the fBm signals corrupted with additive deterministic signals, especially periodical signals like sine/cosine waves [30]. Particularly, let's consider a simple example here in which $\breve{B}_t$ consists of a standard fBm signals $B_{1,t}$ with Hurst exponents H_1 and a sine signal $S_t = \sin(2\pi f_0 t)$ (f_0 is the eigenfrequency of this sine wave, satisfying $f_0 \nrightarrow 0$ and $f_0 \nrightarrow \infty$)

$$\breve{B}_t = \kappa_5 B_{1,t} + \kappa_6 S_t \tag{21}$$

where κ_5 and κ_6 are two positive scaling coefficients that define the mixture ratio/proportion of the signals.

Here we also assume the mixed signals are uniformly sampled at a sufficient rate f_1. In other words, based on *Nyquist–Shannon* sampling theorem, we have $f_1 > 2f_0$. Moreover, the sampling data length M is long enough to cover several periods of the sine wave, that is, $M/f1 \gg 1/f_0$.

If we choose the orthonormal function set as the widely applied $\hat{\phi}_n(t) = \sqrt{2/T}\sin[((n-1/2)\pi/T)t]$ (see [6] and [17]), we have

$$S_t = \sum_{n=1}^{\infty}\left(\sqrt{\frac{2}{T}}\frac{2\pi f_0 \cos[2\pi f_0 T](-1)^{n+1}}{\frac{(n-\frac{1}{2})^2\pi^2}{T^2} - 4\pi^2 f_0^2}\right)\hat{\phi}_n(t) \tag{22}$$

which indicates that when $n \to +\infty$, we have

$$\lambda_{6,n} \sim (c_{6,n})^2 \sim \left(n^{-2}\right)^2 = n^{-4}. \tag{23}$$

Because of independence, we can have

$$\lambda_{5+6,n} = \kappa_5^2\lambda_{5,n} + \kappa_6^2\lambda_{6,n} \sim \kappa_5^2 n^{-(2H_1+1)} + \kappa_6^2 K_2 n^{-4} \tag{24}$$

where $\lambda_{5,n}$ and $\lambda_{6,n}$ denote the corresponding eigenvalues of $B_{1,t}$ and S_t respectively. K_2 is a constant.

Similar to the above discussions, we can see that only the first few eigenvalues will deviate from the line with slope $-(2H_1+1)$ (the number of influenced eigenvalues are partly determined by κ_5/κ_6), because $2H_1 + 1 < 4$. Thus, we can estimate H by checking the slope of the eigenvalue spectrum segment at large enough n.

In such situations, to check the power spectrum of the mixed signal might be a better choice, because the influence of the sine signal is isolated as a single peak and can be easily detected/removed. A typical example is shown in Fig. 4. More simulating results are given in Table V, which indicates that the proposed Hurst estimator is quite robust for fBm signals corrupted by sine signals. Noticing any periodical signals can be decomposed into the combinations of sine/cosine signals, the above conclusions can be easily extended.

IV. Conclusion

The argument on the validity of PCA eigenspectrum based Hurst exponent estimator for single fBm signals is analytically solved in this paper. Moreover, theoretical analysis and experiments show the following:

1) it can estimate some or even all the H parameters for mixed fBm signals with different Hurst exponents, depending on the mixing ratios of the fBm signals. But it may require larger sampling data length and more PCA computation cost;
2) it can estimate the H parameter for fBm signals corrupted with additive Gaussian white noise, if the SNR value is not too small; and
3) it can estimate the H parameter for fBm signals corrupted with additive deterministic sine/cosine signals.

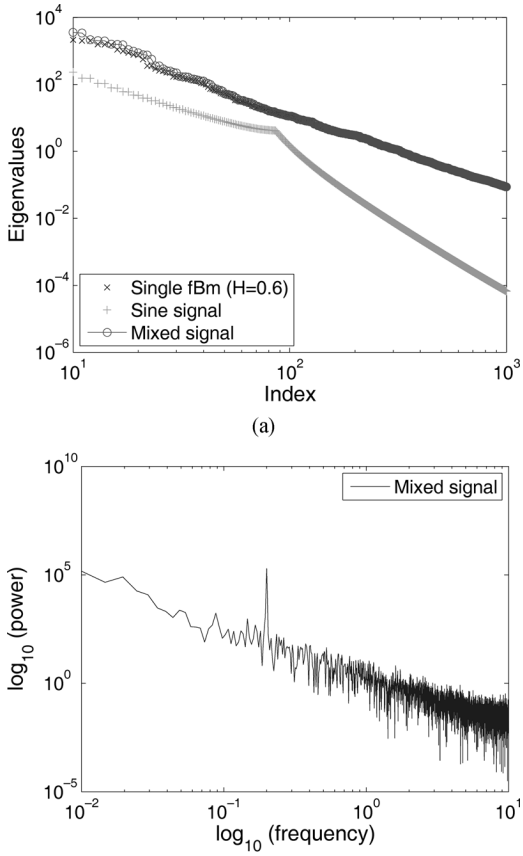

Fig. 4. (a) PCA eigenspectrum of fBm signal corrupted with sine signal, where $H = 0.6$, $f_0 = 0.2$ Hz, $f_1 = 20$ Hz, $\kappa_5/\kappa_6 = 1/10$. $M = 2000$ and (b) the corresponding power-spectrum of the mixed fBm signal, in which the peak caused by sine signal can be easily located around 0.2 Hz.

TABLE V
Means and Variances of the Estimated $\hat{H}$ From50 Samples of Each Signal fBm Signals Mixed With Sine Signals, Respectively. the Eigenfrequencies of Sine Signals Vary Within [0.2, 10] Hz. the Sampling Frequency Is Constant 20 Hz

H	κ_5/κ_6	Mean of $\bar{H}$	Variance of $\bar{H}$
0.1	$1/50 \sim 1$	0.0574	3.8237×10^{-4}
0.2	$1/50 \sim 1$	0.1827	8.6726×10^{-4}
0.3	$1/50 \sim 1$	0.3088	6.1619×10^{-4}
0.4	$1/50 \sim 1$	0.4238	6.6795×10^{-4}
0.5	$1/50 \sim 1$	0.5330	9.1839×10^{-4}
0.6	$1/50 \sim 1$	0.6066	9.5720×10^{-4}
0.7	$1/50 \sim 1$	0.7345	5.8028×10^{-4}
0.8	$1/50 \sim 1$	0.8420	5.3046×10^{-4}
0.9	$1/50 \sim 1$	0.9446	5.6849×10^{-4}

Appendix
Proof of Theorem 1

Based on Lemma 1–3, what is left to do is to prove the asymptotics of the eigenvalues for the continuous K-L decomposition. A direct but very complicated proof of the asymptotic properties of the K-L expansion eigenvalues can be found in [31]. In this paper, we provide a much simpler proof using the following conclusion obtained recently [12], [32]–[34].

Lemma 4 [33]: Let $\cdots < \omega_{-1} < \omega_0 = 0 < \omega_1 < \cdots$ be the real zeros of the Bessel function J_{1-H} of the first kind, we can expand a standard fBm process B_t, $0 < H < 1$, within a finite time interval $[0, T]$ as

$$B_t = \sum_{n\in\mathbb{Z}} z_n \frac{e^{2i\omega_n t/T} - 1}{\frac{2i\omega_n}{T}} \tag{25}$$

where z_n, $n \in \mathbb{Z}$, are independent Gaussian random variables with mean zero and variance shown in (26). $V_T = S_T(0,0)$, and the reproducing kernel S_T can be represented as (27).

$$\begin{aligned} &E\left[z_n^2\right] \\ &= \left[S_T\left(\frac{2\omega_n}{T}, \frac{2\omega_n}{T}\right)\right]^{-1} \\ &= \begin{cases} \left[(2-2H)\Gamma^2(1-H)\left(\frac{\omega_n}{2}\right)^{2H} J_{-H}^2(\omega_n) V_T\right]^{-1}, & \omega_n \neq 0 \\ V_T^{-1}, & \omega_n = 0 \end{cases} \end{aligned} \tag{26}$$

$$\begin{aligned} &\frac{S_T(2\omega, 2\omega)}{S_T(0,0)} \\ &= (2-2H)\Gamma^2(1-H)\left(\frac{T\omega}{2}\right)^{2H} \\ &\quad \times \left(J_{1-H}^2(T\omega) + \frac{2H-1}{T\omega} J_{-H}(T\omega) J_{1-H}(T\omega) + J_{-H}^2(T\omega)\right). \end{aligned} \tag{27}$$

Moreover, the expansion is rate-optimal, which holds

$$\mathbb{E} \sup_{t\in[0,T]} \left| \sum_{|n|>N} z_n \frac{e^{2i\omega_n t/T} - 1}{\frac{2i\omega_n}{T}} \right| \lesssim N^{-H}\sqrt{\log N}. \tag{28}$$

Since *Lemma 4* gives the exact expansion representation, we can prove the decaying law of the eigenvalues by examining the coefficients of the above K-L expansion directly.

Proof: From (26), noticing $\omega_n \neq 0$, we have

$$\lambda_n = E\left[c_n^2\right] = E\left[\left(\frac{z_n}{\omega_n}\right)^2\right] = K\omega_n^{-2}\left[\omega_n^{2H} J_{-H}^2(\omega_n)\right]^{-1} \tag{29}$$

where K is a constant.

Let's consider $J_{-H}^2(\omega_n)$ in (29) first. Generally, we have the following Neumann function $Y_\alpha(x)$, which satisfies [35]

$$J_{-\alpha}(x) = J_\alpha(x)\cos(\alpha\pi) - Y_\alpha(x)\sin(\alpha\pi). \tag{30}$$

Notice that the curves of Bessel functions $J_\alpha(x)$ $(\alpha > 0)$ converges to oscillating sine/cosine functions that decay proportionally to $1/\sqrt{x}$ for large x. Especially, for $x \gg \left|\alpha^2 - 1/4\right|$, we have [35]

$$J_\alpha(x) = \sqrt{\frac{2}{\pi x}}\cos\left(x - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right) + O(|x|^{-3/2}) \tag{31}$$

$$Y_\alpha(x) = \sqrt{\frac{2}{\pi x}}\sin\left(x - \frac{\alpha\pi}{2} - \frac{\pi}{4}\right) + O(|x|^{-3/2}). \tag{32}$$

Since ω_n are the zeros of $J_{1-H}(x)$, we have

$$\cos\left(\omega_n - \frac{(1-H)\pi}{2} - \frac{\pi}{4}\right) = O(|\omega_n|^{-1}) \tag{33}$$

$$\sin\left(\omega_n - \frac{(1-H)\pi}{2} - \frac{\pi}{4}\right) = \pm 1 + O(|\omega_n|^{-1}). \tag{34}$$

Moreover, we have

$$J_H(\omega_n) = \sqrt{\frac{2}{\pi\omega_m}}\cos\left(\omega_n - \frac{H\pi}{2} - \frac{\pi}{4}\right) + O(|\omega_n|^{-3/2}) \tag{35}$$

$$Y_H(\omega_n) = \sqrt{\frac{2}{\pi\omega_n}}\sin\left(\omega_n - \frac{H\pi}{2} - \frac{\pi}{4}\right) + O(|\omega_n|^{-3/2}). \tag{36}$$

Using the triangle formula $\sin(A \pm B) = \sin(A)\cos(B) \pm \cos(A)\sin(B)$, $\cos(A \pm B) = \cos(A)\cos(B) \mp \sin(A)\sin(B)$, we can further have

$$J_H(\omega_n) = \pm\sqrt{\frac{2}{\pi\omega_n}}\sin\left(\frac{(1-2H)\pi}{2}\right) + O(|\omega_n|^{-3/2}) \tag{37}$$

$$Y_H(\omega_n) = \pm\sqrt{\frac{2}{\pi\omega_n}}\cos\left(\frac{(1-2H)\pi}{2}\right) + O(|\omega_n|^{-3/2}). \tag{38}$$

Substituting (37), (38) into (30), we have

$$\begin{aligned} &J_{-H}(\omega_n) \\ &= \pm\sqrt{\frac{2}{\pi\omega_n}}\sin\left(\frac{(1-2H)\pi}{2}\right)\cos(H\pi) \\ &\quad \mp\sqrt{\frac{2}{\pi\omega_n}}\cos\left(\frac{(1-2H)\pi}{2}\right)\sin(H\pi) + O(|\omega_n|^{-3/2}) \\ &= \pm\sqrt{\frac{2}{\pi\omega_n}}\sin\left(\frac{(1-4H)\pi}{2}\right) + O(|\omega_n|^{-3/2}) \sim \omega_n^{-1/2}. \end{aligned} \tag{39}$$

Based on (29) and (39), we have

$$\lambda_n \sim \omega_n^{-2}\left[\omega_n^{2H}\left(\omega_n^{-1/2}\right)^2\right]^{-1} \sim \omega_n^{-(2H+1)}. \tag{40}$$

Since $\alpha = (1-H)$ is a real number here, all the zeros of $J_\alpha(s)$ are real. These real zeros ω_n are asymptotically periodic for large n as

$$\omega_n = a_1 n + a_0 + O(n^{-1}) \sim n \tag{41}$$

where a_1 and a_0 are certain real constants.

Thus we reach the conclusion by combining (40) and (41). ■

REFERENCES

[1] B. Mandelbrot and J. W. van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Rev.*, vol. 10, no. 4, pp. 422–437, 1968.

[2] R. Fischer and M. Akay, "A comparison of analytical methods for the study of fractional Brownian motion," *Ann. Biomed. Eng.*, vol. 24, no. 4, pp. 537–543, 1996.

[3] G. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*. Upper Saddle River, NJ: Prentice-Hall PTR, 1996.

[4] *Theory and Application of Long-Range Dependence*, P. Doukhan, G. Oppenheim, and M. S. Taqqu, Eds. Boston, MA: Birkhäuser, 2003.

[5] T. Karagiannis, M. Molle, and M. Faloutsos, "Understanding the limitations of estimation methods for long-range dependence," UCR-CS-2006–10245, 2006.

[6] J. B. Gao, Y. Cao, and J.-M. Lee, "Principal component analysis of $1/f^\alpha$ noise," *Phys. Lett. A*, vol. 314, no. 5-6, pp. 392–400, 2003.

[7] T. E. Özkurt and T. Akgül, "Is PCA reliable for the analysis of fractional Brownian motion?," presented at the Eur. Signal Processing Conf., Antalya, Turkey, Sep. 4–8, 2005.

[8] T. E. Özkurt, T. Akgül, and S. Baykut, "Principle component analysis of the fractional Brownian motion for $0 < H < 0.5$," presented at the IEEE Conf. Acoustics, Speech, Signal Processing (ICASSP), Toulouse, France, May 14–19, 2006.

[9] M. Erol, T. Akgül, S. Oktug, and S. Baykut, "On the use of principle component analysis for the Hurst parameter estimation of long-range dependent network traffic," *Lecture Notes in Computer Science*, vol. 4263, pp. 464–473, 2006.

[10] Y. S. Mishura, *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. New York: Springer-Verlag, 2008.
[11] F. Biagini, Y. Hu, B. Oksendal, and T. Zhang, *Stochastic Calculus for Fractional Brownian Motion and Applications*. London, U.K.: Springer, 2008.
[12] T. Kuhn and W. Linde, "Optimal series representation of fractional Brownian sheets," *Bernoulli*, vol. 8, no. 5, pp. 669–696, 2002.
[13] P. D. Hislop and I. M. Sigal, *Introduction to Spectral Theory: With Applications to Schrödinger Operators*. New York: Springer, 1995.
[14] E. Masry, B. Liu, and K. Steiglitz, "Series expansion of wide-sense stationary random processes," *IEEE Trans. Inf. Theory*, vol. 14, no. 6, pp. 792–796, Nov. 1968.
[15] A. M. Yaglom, *Correlation Theory of Stationary and Related Random Functions*. New York: Springer-Verlag, 1987, vol. I.
[16] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*. New York: McGraw-Hill, 1999.
[17] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York: McGraw-Hill, 2002.
[18] H. Ogawa and E. Oja, "Can we solve the continuous Karhunen–Loéve eigenproblem from discrete data?," *IEICE Trans.*, vol. E69-E, no. 9, pp. 1020–1029, 1986.
[19] C. K. I. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G. Diettrich, and V. Tresp, Eds. Cambridge, MA: MIT Press, 2001.
[20] V. Popovici and J.-P. Thiran, "PCA in autocorrelation space," in *Proc. IEEE Conf. Pattern Recogn.*, 2002, vol. 2, pp. 132–135.
[21] M. Girolami, "Orthogonal series density estimation and the kernel eigenvalue problem," *Neural Comput.*, vol. 14, no. 3, pp. 669–688, 2002.
[22] C. G. Wu, Y. C. Liang, W. Z. Lin, H. P. Lee, and S. P. Lim, "A note on equivalence of proper orthogonal decomposition methods," *J. Sound Vibrat.*, vol. 265, no. 5, pp. 1103–1110, 2003.
[23] V. Popovici and J.-P. Thiran, "Pattern recognition using higher-order local autocorrelation coefficients," *Pattern Recogn. Lett.*, vol. 25, no. 10, pp. 1107–1113, Jul. 2004.
[24] P. Abry and F. Sellan, "The wavelet-based synthesis for the fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation," *Appl. Comput. Harmon. Anal.*, vol. 3, no. 4, pp. 377–383, 1996.
[25] P. Cheridito, "Mixed fractional Brownian motion," *Bernoulli*, vol. 7, no. 6, pp. 913–934, 2001.
[26] C. El-Nouty, "The fractional mixed fractional Brownian motion," *Stat. Probab. Lett.*, vol. 65, no. 2, pp. 111–120, 2003.
[27] M. Zili, "On the mixed fractional Brownian motion," *J. Appl. Math. Stoch. Anal.*, vol. 2006, no. Article id. 32435, 2006.
[28] H. van Zanten, "When is a linear combination of independent fBm's equivalent to a single fBm?," *Stoch. Process. Appl.*, vol. 117, no. 1, pp. 57–70, 2007.
[29] S. Baykut, T. Akgül, and S. Ergintav, "Estimation of spectral exponent parameter of $1/f$ process in additive white background noise," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. Id. 63129, 2007.
[30] S. Baykut, T. E. Özkurt, M. Erol, and T. Akgül, "The influence of a single tone sinusoid over Hurst estimators," presented at the Eur. Signal Processing Conf., Antalya, Turkey, Sep. 4–8, 2005.
[31] J. C. Bronski, "Small ball constants and tight eigenvalue asymptotics for fractional Brownian motions," *J. Theoretic. Probab.*, vol. 16, no. 1, pp. 87–100, 2003.
[32] K. Dzhaparidzea and J. H. van Zantenb, "A series expansion of fractional Brownian motion," *Probab. Theory Related Fields*, vol. 130, no. 1, pp. 39–55, 2004.
[33] K. Dzhaparidzea and J. H. van Zantenb, "Krein's spectral theory and the Paley–Wiener expansion for fractional Brownian motion," *Ann. Probab.*, vol. 33, no. 2, pp. 620–644, 2005.
[34] K. Dzhaparidzea and J. H. van Zantenb, "Optimality of an explicit series expansion of the fractional Brownian sheet," *Stat. Probab. Lett.*, vol. 71, no. 4, pp. 295–301, 2005.
[35] G. N. Watson, *A Treatise on the Theory of Bessel Functions*. London, U.K.: Cambridge Univ. Press, 1995.

Generalized Identifiability Conditions for Blind Convolutive MIMO Separation

Marc Castella and Eric Moreau

***Abstract*—This correspondence deals with the problem of source separation in the case where the output of a multivariate convolutive mixture is observed: we propose novel and generalized conditions for the blind identifiability of a separating system. The results are based on higher-order statistics and are valid in the case of stationary but not necessarily i.i.d. signals. In particular, we extend recent results based on second-order statistics only. The approach relies on the use of so called reference signals. Our new results also show that only weak conditions are required on the reference signals: this is illustrated by simulations and opens up the possibility of developing new methods.**

***Index Terms*—Blind source separation, contrast functions, higher order statistics, independent component analysis, MIMO convolutive mixtures, MIMO identification, reference system, semi-blind methods.**

I. Introduction

The problem of blind source separation has a very large scope of potential applications such as telecommunications, array processing, audio processing and biology. This explains a high interest in the field over the last decades. Numerous achievements have already been accomplished through the now well-recognized concept of independent component analysis (ICA), see, e.g., [8]. The first extensively studied mixture model was the instantaneous one. However, more recently, the more general convolutive model has been considered [1], [9], [14] and successful algorithms have been proposed [16], [18]. In particular, this has been done through the derivation of so called contrast functions. Such functions are very useful since they provide both identifiability conditions and criteria to be optimized. Classically they are based on statistics of order higher than or equal to three, which may lead to complicated optimization schemes. To simplify the optimization step, further improvements, which are based on the idea of reference signals, have been recently introduced [4], [6], [7], [12], [13]. In a semi-blind context, the reference signals can be interpreted as a partial knowledge on the sources: this has not been fully exploited up to now and it is done in our correspondence.

Two practical ways to consider the problem of blind identifiability can be found. The first one uses second order statistics through matrix algebra decompositions, see, e.g., [10], [11], and [17]. The second one uses higher order statistics (higher than three) through contrast functions, see, e.g., [9] and [14]–[16]. In all cases, questions about identifiability conditions remain unclear and particularly the link between the above two points of view. To a certain extent, the use of reference signals in conjunction with higher order statistics makes a bridge between different statistics. This is what we extensively use in the following.

Manuscript received October 07, 2008; accepted January 27, 2008. First published February 24, 2009; current version published June 17, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alfred Hanssen.

M. Castella is with the Institut Télécom; Télécom & Management SudParis, UMR-CNRS 5157, 91011 Évry Cedex, France (e-mail: marc.castella@it-sudparis.eu).

E. Moreau is with the University of Sud Toulon Var, ISITV, LSEET UMR-CNRS 6017, BP56, F-83162 La Valette du Var Cedex, France (e-mail: moreau@univ-tln.fr).

Digital Object Identifier 10.1109/TSP.2009.2016259