PRINCIPAL COMPONENT ANALYSIS OF THE FRACTIONAL BROWNIAN MOTION FOR 0 < *H* < 0.5

Tolga Esat Özkurt[1], *Tayfun Akgül*[2], *Süleyman Baykut*[3]

[1]Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA
tolga@neuronet.pitt.edu

[2, 3]Department of Electronics and Communication Engineering, Istanbul Technical University, Istanbul, Turkey
tayfun@ehb.itu.edu.tr, baykut@ehb.itu.edu.tr

[2] TUBITAK - Marmara Research Center, Earth and Marine Sciences Research Institute, Kocaeli, Turkey

ABSTRACT

Principal component analysis (PCA) has been proposed for the estimation of the self-similarity parameter H, namely the Hurst parameter of $1/f$ processes, and an analytical proof is provided only for H=0.5 in a recent study [1]. In our paper, we extend this study by deriving explicit expressions and presenting an analytical proof for the range of $0 < H < 0.5$ (the anti-persistent part of the fractional Brownian motion). We also show via simulations that the accuracy of the estimated H values may decrease considerably as the theoretical H value increases towards the persistent part ($0.5< H < 1$).

1. INTRODUCTION

1/f processes are a family of statistically self-similar (SSS) processes. An SSS process $x(t)$ has the scaling property given below:

$$x(t) \overset{d}{=} a^{-H} x(at) \tag{1}$$

where H is the Hurst parameter, a is a positive real number and $\overset{d}{=}$ denotes statistical equivalence [2]. Self-similarity is a common property observed in many man-made and natural phenomena such as physiological heart-rate records [3], economical time-series [4], geophysical processes [5], teletraffic data [6], speech signal residuals [7], electromagnetic fluctuations [2] and many more. Such self-similar processes obey the following well-known power law relationship:

$$S_x(f) \approx \frac{\sigma_x^2}{|2\pi f|^{2H+1}} \tag{2}$$

where $S_x(f)$ is the power spectrum of the time series $x(t)$, σ_x^2 is the variance and f is the frequency.

Fractional Brownian motion (fBm) processes are one of the widely used models of the *1/f* processes. fBms are zero-mean, normally distributed, nonstationary random processes and their complete statistical characterization can be achieved by a single parameter, H, in the range of (0, 1). Since H is a significant parameter for *1/f* processes a number of techniques are developed to estimate it [8]. In a recent study, the Principal Component Analysis (PCA) method is suggested for estimating the H parameter of fBm processes and an analytical proof is provided for $H = 0.5$ only [1]. Here, we extend this study by providing an analytical proof for the range of $0 < H < 0.5$. Furthermore, we show through simulations that PCA method may not always yield accurate estimations when $0.5 < H < 1$.

This paper is organized as follows: In the next section, we summarize our derivations for $0 < H < 0.5$. Simulation results and performance evaluations are given in Section 3. Main conclusion is given in Section 4.

2. DERIVATIONS

In this section, we start with the standard definitions of fBm process $x(t)$, where the autocorrelation function $R(t_1,t_2)$ is given as [2]:

$$R(t_1,t_2) = \frac{1}{2}\sigma_H^2 \left\{ |t_1|^{2H} - |t_1 - t_2|^{2H} + |t_2|^{2H} \right\} \tag{3}$$

where

$$\sigma_H^2 = E\{x(1)^2\} = \Gamma(1-2H)\frac{\cos(\pi H)}{\pi H} \tag{4}$$

Here $\Gamma(.)$ is the gamma function and $E\{.\}$ denotes the expectation operator. Traditionally, the fBm processes are divided in three parts, i.e., $0 < H < 0.5$, antipersistent range where the process shows tendency to change its trend, sometimes referred as “negatively correlated”; $H = 0.5,$ classicial Brownian motion (Bm); and $0.5 < H < 1$, the process has the behavior to persist in the same way, sometimes referred as “positively correlated” [9].

The PCA method proposed in [1] relies on the eigenanalysis of the autocorrelation function in Eq. (3):

$$\mathbf{R}\varphi_i = \lambda_i \varphi_i \ , \ i=1,\dots,N \tag{5}$$

[1] T. E. Ozkurt’s work was done in Istanbul Technical University.
[3] S. Baykut’s work is supported by TUBITAK-BAYG.

where λ_i's are eigenvalues and $\boldsymbol{\varphi}_i$'s are the corresponding eigenvectors of the autocorrelation matrix $\mathbf{R}$ of the discretized process of $x(t)$ with the truncation length N. The Karhunen-Loéve expansion is known to be the continuous version of PCA whose fundamental equation can be given as [11]:

$$\int_{-T/2}^{T/2} R(t_1,t_2)\varphi(t_2)dt_2 = \lambda\varphi(t_1) \tag{6}$$

where $-T/2<t_1<T/2$, $-T/2<t_2<T/2$ and $\varphi(t)$'s are the orthonormal eigenfunctions corresponding to the eigenvectors. In [1], it is shown only for $H=0.5$ that by substituting the autocorrelation function of classical Brownian motion ($H=0.5$) in Eq. (6), the relationship between the eigenvalues, λ_i, and their indices i sorted in decreasing order is proportional:

$$\lambda_i \sim i^{-(2H+1)} \tag{7}$$

Following the same procedure, we consider the case $H \neq 0.5$ to explore whether the similar relationship holds. We start with substituting Eq. (3) into Eq. (6) and then apply the Karhunen-Loéve expansion for $H \neq 0.5$ which yields

$$\begin{aligned}\frac{1}{2}\sigma_H^2\Bigg\{&\int_{-T/2}^{T/2}|t_1|^{2H}\varphi(t_2)dt_2 - \int_{-T/2}^{t_1}(t_1-t_2)^{2H}\varphi(t_2)dt_2 \\ &- \int_{t_1}^{T/2}(t_2-t_1)^{2H}\varphi(t_2)dt_2 + \int_{-T/2}^{T/2}|t_2|^{2H}\varphi(t_2)dt_2\Bigg\} = \lambda\varphi(t_1)\end{aligned} \tag{8}$$

Using the Leibniz's formula [13], we differentiate both sides of the above expression with respect to the variable t_1 and obtain:

$$\begin{aligned}\sigma_H^2 H\Bigg\{&\int_{t_1}^{T/2}(t_2-t_1)^{2H-1}\varphi(t_2)dt_2 \\ &- \int_{-T/2}^{t_1}(t_1-t_2)^{2H-1}\varphi(t_2)dt_2\Bigg\} = \lambda\varphi'(t_1)\end{aligned} \tag{9}$$

Notice that this differentiation causes discontinuity at $H = 0.5$, which is out of the interval we are interested in.

Similar to [11], we try a solution for eigenfunctions as complex exponentials:

$$\varphi_i(t) = exp(j\omega_i t),\ \omega_i = \frac{2\pi}{T}i \tag{10}$$

whose derivative yields to:

$$\varphi_i'(t) = -j\omega_i\, exp(j\omega_i t) \tag{11}$$

By substituting Eq. (11) into Eq. (9) and after some algebra and change of variables, we reach to the following equation:

$$\lambda_i j\omega_i = \sigma_H^2 H\left\{\int_0^{T/2-t_1} 2j\sin(\omega_i t)t^{2H-1}dt - \int_{T/2-t_1}^{T/2+t_1}\exp(-j\omega_i t)t^{2H-1}dt\right\} \tag{12}$$

Assuming a large interval T, we seek for a solution when $T\to\infty$. Then Eq. (12) reduces to:

$$\lambda_i\omega_i = \sigma_H^2 H\left\{\int_0^{\infty}2\sin(\omega_i t)t^{2H-1}dt\right\} \tag{13}$$

Using the following property [14],

$$\int_0^{\infty}\frac{sin(ax)}{x^p}dx = \frac{a^{p-1}\pi}{2\Gamma(p)sin(p\pi/2)}\ ,\ 0<p<1 \tag{14}$$

the integral in Eq. (13) can be rewritten in terms of the gamma function as:

$$\lambda_i\omega_i = \frac{\omega_i^{-2H}H\pi}{\Gamma(1-2H)\cos(\pi H)}\sigma_H^2 \tag{15}$$

which is convergent only for $0 < H < 0.5$. By substituting Eq. (4) into Eq. (15), a relatively simple expression between λ_i and ω_i is obtained:

$$\lambda_i = \frac{1}{\omega_i^{2H+1}} = \left(\frac{2\pi}{T}i\right)^{-2H-1} \sim i^{-(2H+1)} \tag{16}$$

Hence, we are able to show that Eq. (7) also holds for $0 < H < 0.5$.

If we take the logarithm of both sides, and apply a linear fit algorithm, the slope of the straight line gives $2H + 1$, and then the H parameter can be easily estimated.

3. EXPERIMENTAL RESULTS

In this section we present the results of our experiments to illustrate the accuracy of the PCA based H parameter estimation.

3.1 Synthetic fBm data generation

We generated synthetic fBm data by using three different synthesis techniques:

- *Fourier Based Spectral (FBS) Method*: This method relies on constructing the power-law relationship for the signal in the spectral domain and taking the inverse Fourier transform [10].
- *Random Midpoint Displacement (RMD) Method*: This method is based on adding new midpoints to the recursively divided subintervals [15]
- *Wavelet Based Generation (WBG) Method*: This method examines the progression of the variances of the wavelet coefficients along scales. The wavelet coefficients at different scales are constructed and the time-domain process is obtained through an inverse transform using these coefficients.

3.2 fBm data examples and their PCA analyses

In Fig.1-(a), a synthetic fBm trace (length of 4096) with $H = 0.4$, and in Fig. 1-(b) the eigenvalue progression plot corresponding this trace are shown. The estimated H parameter, $\hat{H}$, is calculated as: 0.4114.

In Fig. 2-(a), a synthetic fBm trace (length of 4096) with $H = 0.8$ and in Fig. 2-(b) its eigenvalue progression are plotted respectively. In this example the H parameter is estimated as 0.5322 which is not accurate. It is observed in Fig. 2-(b) that a few dominant eigenvalues remain out of the linear progression of the rest.

It is relevant to mention here that, since the lag of the autocorrelation function corresponds to the number of eigenvalues, it is a critical parameter on the performance of the estimator. In practice, the lag can be selected from 1% to 10% of the data length [16]. In our experiments, where the data length is 4096, we have chosen it as 300 points to ease the computational burden.

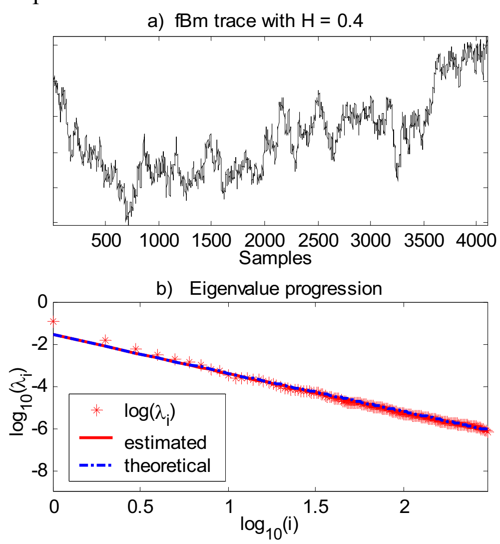

Figure 1: a) fBm trace (H = 0.4), b) Eigenvalue progression plot ($\hat{H}$ = 0.4114).

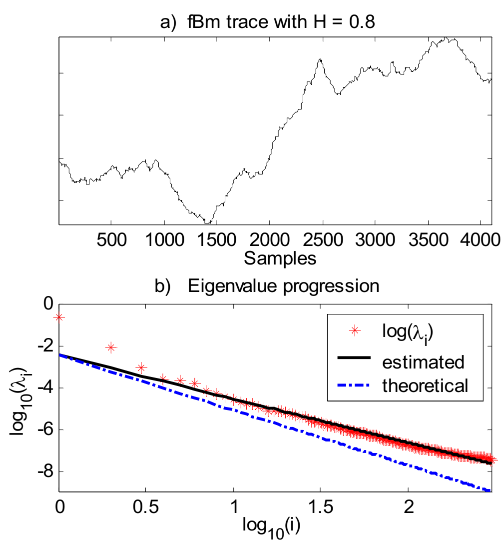

Figure 2: a) fBm trace (H = 0.8), b) Eigenvalue progression plot ($\hat{H}$ = 0.5322).

3.3 Analysis of fBms with different H

We generate a set of fBm traces having H between 0.1 and 0.9 with an increment of 0.1 using data length of 4096. For each H, 100 data sets are generated by FBS, RMD, and WBG algorithms. Then we estimated the H parameters of the fBm sample paths by the PCA based method.

To provide information on the bias of the estimation method we define the mean-square error (MSE) as:

$$MSE = \frac{1}{K}\sum_{i=1}^{K}(H_i - \hat{H}_i)^2 \tag{17}$$

Here, H is the theoretical Hurst parameter, and $\hat{H}$ is the estimated value. Square roots of MSE versus the theoretical H values are shown in Fig. 3. When H > 0.5 increase in MSE is observed, which indicates that the accuracy of the estimator drastically decreases for that region. This result can also be seen in Figure 4, where we provide the box plot of the estimated H values of the fBm data sets. When H > 0.5, the method may underestimate the H parameters with relatively higher variances.

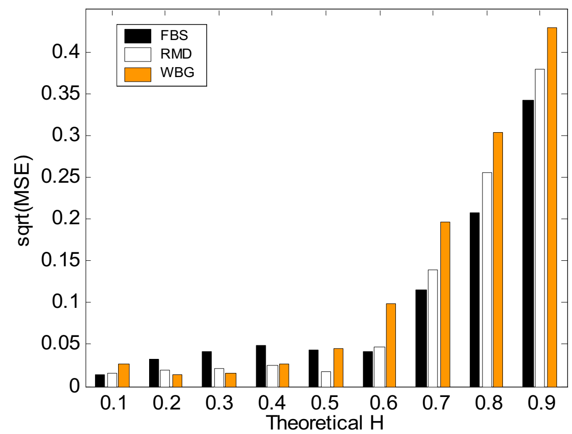

Figure 3: sqrt(MSE) versus H.

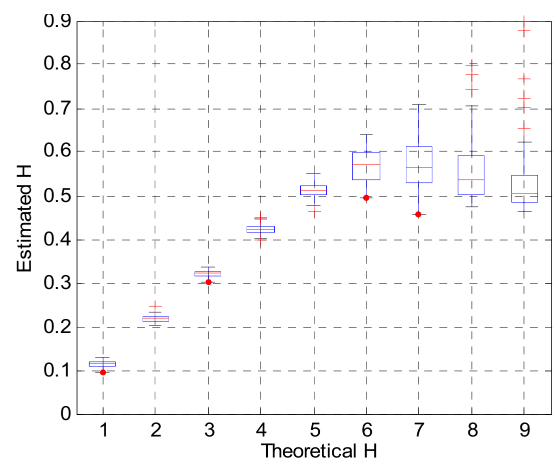

Figure 4: Box plot of the estimated H.

3.4 Effect of data length

In order to investigate the dependency of the PCA method on the data length, a set of Bm traces having lengths

N=1024, 2048, 4096, 8192 and 16384 are generated and tested. In Fig. 5, the square roots of MSE versus the lengths N are shown which reveals that the estimates do not overly sensitive to the data length.

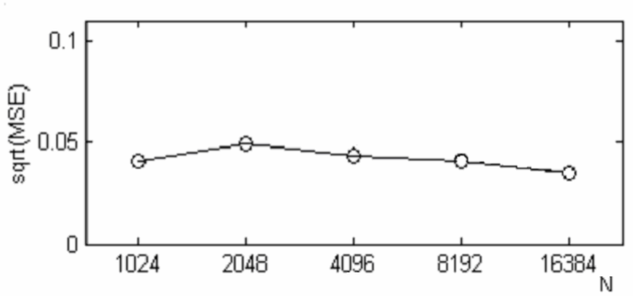

Figure 5: sqrt(MSE) versus the length of data N.

4. CONCLUSION

In this work, we provide an explicit expression for the antipersistent part ($H < 0.5$) of an fBm process which shows, for PCA, a linear eigenvalue progression in the logarithmic scale. However, because of divergence of the derived expression, the proof is not valid for $H > 0.5$. We include simulation results to show that PCA is an appropriate estimator when $0 < H < 0.5$ and $H = 0.5$. Nevertheless, the accuracy of the PCA method is questionable for $0.5 < H < 1$. We conclude that one should be careful while analyzing $1/f$ processes for $H > 0.5$ with PCA. As a matter of fact, using PCA to estimate the H parameter is not fully recommended in this paper. However, there are indications that the power-law eigenstructure can be related to some information retrieval applications with power-law networks, i.e., from Internet to biological networks of genes, and protein interactions, etc.

Acknowledgements

The authors would like to thank Melike Erol, Ali Yapar and Özkan Karabacak for their help, suggestions and comments.

REFERENCES

[1] J. B. Gao, Y. H. Cao, J. M. Lee, "Principal Component Analysis of *1/f* Noise," *Phys. Lett. A* 314, pp. 392-400, 2003.

[2] G. W. Wornell, "Wavelet-based Representations for the *1/f* Family of Fractal Processes," *Proc. of IEEE*, vol. 81, no. 10, pp. 1428-1450, Oct. 1993.

[3] C. K. Peng , J. Mietus, J. M. Hausdorff, S. Havlin, H. E. Stanley, A. L. Goldberger, "Long-range Anti-correlations and Non-Gaussian Behavior of the Heartbeat," *Phys. Rev. Lett.*70, pp. 1343-1346, 1993.

[4] J. Beran, *Statistics for Long-Memory Processes,* Chapman & Hall, 1994.

[5] M.C. Breslin, J. A. Belward, "Fractal Dimensions for Rainfall Time Series," *Mathematics and Computers in Simulation,* Vol. 48, no. 4-6, pp. 437-446, 1999.

[6] O. Cappe, E. Moulines, J. C. Pesquet, A. P. Petropulu, X. Yang, "Long-Range Dependence and Heavy-tail Modeling for Teletraffic Data," *IEEE Signal Processing Magazine*, vol. 19, no. 3, pp. 14-27, May 2002.

[7] Y. Li, T. Akgül, B. Onaral, "A Speech Quality Enhancement Method Using the *1/f* Nature of Speech Residual Signal," *Proc. of the 33rd Annual Conference on Information Science and Systems*, pp. 534-539, 1999.

[8] M. S. Taqqu, V. Teverovsky, W. Willinger, "Estimators for Long-range Dependence: An Empirical Study," *Fractals*, vol. 3, no. 4, pp 785-798, 1995.

[9] B. B. Mandelbrot, J. W. Van Ness, "Fractional Brownian Motions, Fractional Noise and Applications," *SIAM Review*, vol. 10, no. 4, pp. 422-437, 1968.

[10] D. G. Manolakis, V. K. Ingle, S. M. Kogon, *Statistical and Adaptive Signal Processing,* McGraw Hill, 2000.

[11] H. L. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I,* John Wiley & Sons, 1968.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* McGraw Hill, 1991.

[13] F. B. Hilderbrand, *Advanced Calculus for Applications*, Prentice-Hall, 1962.

[14] M. Abramowitz, I. A. Stegun, *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables,* Dover Publications, 1974.

[15] W. C. Lau, A. Erramilli, J. L. Wang, W. Willinger, "Self-similar Traffic Generation: The Random Midpoint Displacement Algorithm and Its Properties," *Proc. of the IEEE ICC*, pp. 466-472, 1995.

[16] T. E. Özkurt, T. Akgül, "Is PCA Reliable for the Analysis of Fractional Brownian Motion?" *Proc. of the 13th European Signal Processing Conference (EUSIPCO)*, Sept. 2005.