Wavelet Estimation of Fractional Brownian Motion Embedded in a Noisy Environment

Lei Zhang, Paul Bao, *Senior Member, IEEE*, and Xiaolin Wu, *Senior Member, IEEE*

***Abstract*—This correspondence proposes a wavelet-based fractional Brownian motion (fBm) signal estimation scheme. Despite the fact that wavelet transform approximately whitens the fBm processes, it is observed that statistical dependencies still exist across adjacent wavelet scales and between neighboring wavelet coefficients. These dependencies can be exploited to improve the estimation of fBm signals embedded into noise. The idea is to reorganize the wavelet coefficients into a scale–time mixture model and then carry out the minimum mean-square-error estimation (MMSE) using the model. Experiments show that the proposed scheme obtains better estimates than Wornell and Oppenheim's algorithm, in which the wavelet dependencies are not utilized.**

***Index Terms*—Fractional Brownian motion(fBm), intra- and inter-scale dependency, signal estimation, wavelet transform.**

I. Introduction

Fractional Brownian motion (fBm) was introduced to model the processes with long-term dependencies and statistical self-similarity [1], [3]. Since the fBm processes are nonstationary while having stationary increments, the time–frequency techniques, especially wavelet transform (WT) [14]–[16], are suitable tools in analyzing the behaviors of the fBm processes. The self-similarity property of fBm implies that any portion of a continuous fBm process can be viewed as a scaled version of itself. Thus, WT, with its innate multiresolution structure, naturally endows a good description of fBm.

Orthogonal wavelet transform (OWT) is almost a Karhunen–Loève-like transform that well decorrelates a rich class of random processes [12]. As for fBm signals, several papers [4], [5], [7], [10], [11] have discussed their correlation structures in wavelet domain. They showed that due to the $1/f$-type spectral behavior of fBm, the auto- and cross-correlation functions of its OWT coefficients decay exponentially along scales. With this property, some wavelet-based approaches have been proposed to efficiently estimate the parameters of fBm processes from noise-free or noisy measurements [5], [6], [8], [9], [13]. Wornell and Oppenheim [6] and Kaplan and Kuo [9] developed two approximate maximum-likelihood estimators that, respectively, applied to fBm signals and discrete fractional *Gaussian* noise (DFGN). The estimator of Fieguth and Willsky [8] models an fBm signal as a multiscale tree by Haar wavelet. Wornell and Oppenheim [6] also recovered the fBm signals by the minimum mean-squared-error estimation (MMSE) technique with the corresponding estimated parameters. In [13], Hwang gave a Wiener filter of noisy fBm signals with nonorthognal wavelet transforms. Wornell and Oppenheim's method is a specific case of the Wiener filter when OWT is used.

Although an fBm signal is significantly whitened by OWT, its wavelet coefficients still have some dependencies inter-scale and intra-scale. This correspondence aims at developing a noisy fBm signal estimation scheme, in which the wavelet intra- and inter-scale dependencies are exploited to improve the signal estimates. The main idea is to form a random vector by combining wavelet coefficients at two adjacent scales which are associated to each other by descendant–ancestor relationship. By computing the statistics of the vector variable, the MMSE technique is then applied to the multiscale model to recover the fBm signals.

The rest of the correspondence is organized as follows. Section II introduces the wavelet decorrelation of fBm processes. Section III develops the proposed multiscale MMSE-based scheme of fBm signal estimation. In order to demonstrate the efficacy and advantages of the new estimation scheme, simulation results are presented in Section IV. Section V is the conclusion.

II. Representation of Fractional Brownian Motion by Wavelet Transform

Fractional Brownian motion (fBm) is a generalization of ordinary Brownian motion [1]. It is a zero-mean nonstationary stochastic process $B_H(t)$, whose covariance function is

$$r_{B_H}(t,s) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right) \tag{1}$$

where parameter $0 < H < 1$ controls the roughness of the process. The ordinary Brownian motion is the case when $H = 1/2$. In [3], Flandrin stated that the average spectrum of fBm could be written as

$$S_{B_H}(\omega) = \frac{\sigma^2}{|\omega|^{2H+1}}. \tag{2}$$

The above equation implies that fBm is competitive to model those stochastic processes with long-term dependencies such as $1/f$-like processes.

Although fBm is nonstationary, its increments are stationary and statistically self-similar. That is, the statistic properties of process $B_H(t+s) - B_H(t)$ depend only on S and there is [1]

$$B_H(t+s) - B_H(t) \cong b^{-H}\left[B_H(t+s\cdot b) - B_H(t)\right] \tag{3}$$

where symbol "$\cong$" means identical in distribution and b is a constant. In discrete computation, a continuous fBm process is sampled as $B_H[k] = B_H(k\cdot\Delta t)$, where Δt is the sampling period. The increment, defined as

$$X_H[k] = B_H[k+1] - B_H[k] \tag{4}$$

is called the discrete fractional *Gaussian* noise (DFGN) [2]. DFGN is a zero-mean stationary *Gaussian* process whose autocorrelation function is

$$r_{X_H}[k] = \frac{\sigma_X^2}{2}\left[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right] \tag{5}$$

where σ_X^2 is the variance of sequence $X_H[k]$.

Since fBm characterizes the nonstationary and self-similar features, the WT [14]–[16], as an inherent multiscale (multiresolution) representation, is naturally an appropriate and powerful tool in analyzing fBm signals. Suppose function ψ is a mother wavelet basis, WT represents a function f as the linear combination of a series of dilations and translations of ψ : $f = \sum d_{j,n}\psi_{j,n}$, where $\psi_{j,n}(t) = 2^{-j/2}\psi(2^{-j}t - n)$, $j, n \in \mathrm{Z}$. Suppose $a_{0,n}^B$ is the initial fBm sequence given at the finest scale $j = 0$, i.e., $a_{0,n}^B = \int \phi(t-n)B_H(t)dt$, where function ϕ is the associated scale function of wavelet function ψ. Then wavelet coefficients $d_{j,n}^B$ can be computed recursively as

$$a_{j,n}^B = \sum_l h[2n-l]a_{j-1,l}^B \qquad d_{j,n}^B = \sum_l g[2n-l]a_{j-1,l}^B \tag{6}$$

Manuscript received August 24, 2002; revised December 15, 2003.

L. Zhang and X. Wu with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4L8, Canada (e-mail: johnray@mail.ece.mcmaster.ca; xwu@mail.ece.mcmaster.ca).

P. Bao is with the School of Computer Engineering, Nanyang Technological University, Singapore, 639798 (e-mail: paulbao@ie.cuhk.edu.hk).

Communicated by A. Kavčić, Associate Editor for Detection and Estimation.
Digital Object Identifier 10.1109/TIT.2004.833357

where h and g are the quadrature mirror filters (QMF) associated with the wavelet ψ [14].

It has been shown by many authors [4], [7], [10], [11] that OWT well whitens the fBm processes, and the correlation function of wavelet coefficients $d_{j,n}^B$ is [4]

$$E\left(d_{j,n}^B d_{k,m}^B\right) = \frac{\sigma^2}{2}(2^k)^{2H+1}\left(-\int A_\psi\left(2^{j-k}, \tau-(2^{j-k}n-m)\right)|\tau|^{2H}d\tau\right) \tag{7}$$

where $A_\psi(\alpha,\tau)=\sqrt{\alpha}\int\psi(t)\psi(\alpha t-\tau)dt$ is called the *reproducing kernel* of wavelet ψ. Wavelet coefficients $d_{j,n}^B$ are weakly correlated in scale and time, and there is [11]

$$E\left(d_{j,n}^B d_{k,m}^B\right) \sim \mathrm{O}\left(\left|2^j n - 2^k m\right|^{2(H-R)}\right) \tag{8}$$

for all j, k such that

$$\left|2^j n - 2^k m\right| > \max\left(2^j K_1 - 2^k K_2, 2^j K_2 - 2^k K_1\right)$$

where R is the number of vanishing moments of wavelet ψ and $[-K_1, K_2]$ is the support of ψ, $K_1, K_2 > 0$.

It is easily derived from (7) that the variance of $d_{j,n}^B$ is

$$r_j = Var\left(d_{j,n}^B\right) = C_d(2^{2H+1})^j \tag{9}$$

where $C_d = -\frac{\sigma^2}{2}\int A_\psi(1,\tau)|\tau|^{2H}d\tau$ is a constant. Obviously, the variance r_j decays exponentially along scales j.

III. The Signal Estimation Algorithm

A. Problem Description

Suppose we have a sequence of observations of fBm signal $B_H[K]$

$$Y[k] = B_H[k] + \varepsilon[k] \tag{10}$$

where ε is additive *Gaussian* white noise with zero-mean and variance σ_ε^2. Since OWT is a linear transform, we have

$$d_{j,n}^Y = d_{j,n}^B + d_{j,n}^\varepsilon \tag{11}$$

where $d_{j,n}^Y$, $d_{j,n}^B$, and $d_{j,n}^\varepsilon$ are the OWTs of $Y[k]$, $B_H[k]$, and $\varepsilon[k]$, respectively. Suppose the wavelet is orthonormal, $d_{j,n}^\varepsilon$ is then white noise with zero mean and variance σ_ε^2, and $d_{j,n}^\varepsilon$ is also *Gaussian* due to the linearity of OWT.

An interesting problem is to estimate the parameter H and the signal $B_H[k]$ from noisy measurement. Abry *et al.* gave a survey of the fBm estimation problem in [5]. In [6], Wornell and Oppenheim developed a maximum-likelihood estimator of H and other parameters by assuming that $d_{j,n}^B$ is totally uncorrelated by OWT. They also recovered the fBm process $B_H[k]$ by the MMSE method with the estimated parameters. This signal estimation approach was extended to nonorthogonal wavelet transforms and generalized as a Wiener filter by Hwang [13]. Kaplan and Kuo [9] implemented OWT on the associated DFGN signal of an fBm process with Haar wavelet, and they reported better estimation results of H with a similar maximum-likelihood estimator to that in [6]. Fieguth and Willsky [8] proposed another maximum-likelihood estimator of H with a multiscale framework technique, which models the fBm as a multiscale tree by Haar wavelet.

Although the Haar wavelet is not a good candidate in many signal processing applications for that it has the least vanishing moments, interestingly it is rather a good choice in fBm estimation. The main reason is that the Haar wavelet has the shortest support among all wavelets, making it be advantageous in describing the self-similarity property (refer to (3)) of fBm. Furthermore, Haar wavelet does not need a periodic extension of fBm data, which is a common problem to other wavelets of longer supports. With Haar wavelet, there exists greater inter-scale correlation than with the other wavelets. This would have been a shortcoming of Haar wavelet transform in other applications. But fortunately to our estimation approach, as we will see in the next subsection that the additional information provided by the wavelet inter-scale dependencies can be exploited to improve the performance of fBm estimation in noisy environments. With these considerations, the Haar wavelet is also employed in this correspondence.

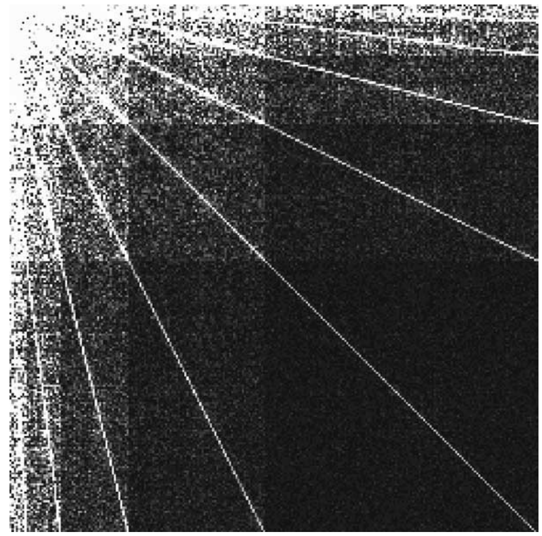

Fig. 1. The experimental covariance matrix of fBm processes $(H = 0.65)$. The correlation information is mainly existed across scales and between neighboring coefficients.

B. Intra- and Inter-Scale Dependencies Exploited Model

Although the OWT is an approximate whitening filter bank, there still exist some dependencies inter- and intra- wavelet scales. The dependency information could be employed to improve the signal processing performance [18]. In Fig. 1 we illustrate an experimental covariance matrix of the wavelet coefficients of fBm $(H = 0.65)$ with Haar basis. It is observed that the wavelet dependencies distribute mainly across scales and among neighboring coefficients within each band. (This is implied by the auxiliary diagonals in Fig. 1.) In this correspondence, the inter-scale dependencies between adjacent two scales and the intra-scale dependencies between nearby two wavelet coefficients are exploited for improving noise reduction results of fBm.

Using Haar basis, whose vanishing moments number is $R = 1$, it can be seen from (8) that $E\left(d_{j,n}^B d_{j,m}^B\right) \sim \mathrm{O}\left(\left|2^j(n-m)\right|^{2(H-1)}\right)$, and $E\left(d_{j,n}^B d_{j,m}^B\right)$ decays quickly for increasing $|n-m|$ at scale j. Considering the covariance of $d_{j,n}^B$ with $d_{j,m}^B$ such that $|n-m|=1$, one can derive from (7) that

$$r_{j,0} = E\left(d_{j,n}^B d_{j,m}^B\,\big|_{|n-m|=1}\right) = C_0(2^{2H+1})^j \tag{12}$$

where

$$C_0 = -\frac{\sigma^2}{2}\int A_\psi(1,\tau-n+m)|\tau|^{2H}d\tau$$

is a constant.

Suppose $k < j$, from (7) it follows that $d_{j,n}^B$ is almost uncorrelated with $d_{k,m}^B$, except for its descendants, i.e., those coefficients with $2^{j-k}n \le m \le 2^{j-k}(n+1)-1$ [4]. Let $k = j-1$, the covariances of parent $d_{j,n}^B$ with its two descendants $d_{j-1,2n}^B$ and $d_{j-1,2n+1}^B$ can be derived from (7) as

$$\begin{aligned} r_{j,1} &= E\left(d_{j,n}^B d_{j-1,2n}^B\right) = C_1(2^{2H+1})^j \\ r_{j,2} &= E\left(d_{j,n}^B d_{j-1,2n+1}^B\right) = C_2(2^{2H+1})^j \end{aligned} \tag{13}$$

where C_1 and C_2 are constants. Similar to the variance r_j in (9), the covariances $r_{j,0}$, $r_{j,1}$, and $r_{j,2}$ also decay exponentially along scales.

Combining the ancestor–descendant-associated coefficients $d_{j,n}^Y$, $d_{j-1,2n}^Y$, and $d_{j-1,2n+1}^Y$, we have

$$\vec{d}_{j,n}^Y = \vec{d}_{j,n}^B + \vec{d}_{j,n}^\varepsilon \tag{14}$$

with

$$\vec{d}_{j,n}^* = \mathrm{col}\left\{d_{j,n}^*, d_{j-1,2n}^*, d_{j-1,2n+1}^*\right\}, ^* \in \{Y, B, \varepsilon\} \tag{15}$$

where col means column vector. Since $\vec{d}_{j,n}^B$ and $\vec{d}_{j,n}^\varepsilon$ are uncorrelated and jointly *Gaussian* distributed with zero means, the MMSE of $\vec{d}_{j,n}^B$ is

$$\hat{\vec{d}}_{j,n}^B = P_j^B(P_j^B + P^\varepsilon)^{-1}\vec{d}_{j,n}^Y \tag{16}$$

with P_j^B and P^ε being the covariance matrices of vector variables $\vec{d}_{j,n}^B$ and $\vec{d}_{j,n}^\varepsilon$, respectively. Because $d_{j,n}^\varepsilon$ is white *Gaussian*, it follows that $P^\varepsilon = \mathrm{diag}\left[\sigma_\varepsilon^2, \sigma_\varepsilon^2, \sigma_\varepsilon^2\right]$ is a diagonal matrix. The noise level σ_ε can be estimated by the maximum-likelihood method of Wornell and Oppenheim [6] or the one of Kaplan and Kuo [9]. In this development, we employed the median estimator of Donoho and Johnstone [17]. σ_ε is estimated from the median absolute value of the wavelet coefficients at the finest scale, i.e., $\hat{\sigma}_\varepsilon = \mathrm{Median}(\left|d_{1,n}^Y\right|)/0.6745$.

In (16), the intra- and inter-scale dependency between wavelet coefficients offers additional information to estimate $d_{j,n}^B$, which is advantageous over estimating $d_{j,n}^B$ solely by its observation $d_{j,n}^Y$. Next we discuss the calculation of matrix P_j^B that is required by the implementation of (16).

C. Estimation of Covariance Matrix P_j^B

Covariance matrix P_j^B should be estimated to compute the MMSE of $\vec{d}_{j,n}^B$, the elements of P_j^B are

$$P_j^B = E\left[\vec{d}_{j,n}^B \cdot (\vec{d}_{j,n}^B)^T\right] = \begin{bmatrix} r_j & r_{j,1} & r_{j,2} \\ r_{j,1} & r_{j-1} & r_{j-1,0} \\ r_{j,2} & r_{j-1,0} & r_{j-1} \end{bmatrix}. \tag{17}$$

With the measurements $d_{j,n}^Y$ and the fact that $d_{j,n}^\varepsilon$ is white noise uncorrelated with $d_{j,n}^B$, there are

$$\begin{aligned} r_j &\approx \frac{1}{N_j}\sum_{n=0}^{N_j-1}(d_{j,n}^Y)^2 - \sigma_\varepsilon^2 \\ r_{j-1,0} &\approx \frac{1}{N_j}\sum_{n=0}^{N_j-1} d_{j-1,2n}^Y \cdot d_{j-1,2n+1}^Y \\ r_{j,1} &\approx \frac{1}{N_j}\sum_{n=0}^{N_j-1} d_{j,n}^Y \cdot d_{j-1,2n}^Y \\ r_{j,2} &\approx \frac{1}{N_j}\sum_{n=0}^{N_j-1} d_{j,n}^Y \cdot d_{j-1,2n+1}^Y \end{aligned} \tag{18}$$

where N_j is the number of wavelet coefficients at scale j.

The estimated values of r_j, $r_{j,0}$, $r_{j,1}$, and $r_{j,2}$ should be further processed because we have the prior information (9), (12), and (13) of them. Denote

$$\beta = 2^{2H+1}. \tag{19}$$

In this work, we estimate parameter β by the approximate maximum-likelihood method in [6] or [9]. Denote the estimation of β by $\hat{\beta}$, with (9) we have

$$\begin{bmatrix} r_1 \\ \vdots \\ r_J \end{bmatrix} \approx \begin{bmatrix} \hat{\beta}^1 \\ \vdots \\ \hat{\beta}^J \end{bmatrix} C_d \tag{20}$$

where J is the total number of wavelet decomposition scale. The least-squares estimation of C_d is then

$$\hat{C}_d = \left(\begin{bmatrix} \hat{\beta}^1 \\ \vdots \\ \hat{\beta}^J \end{bmatrix}^T \begin{bmatrix} \hat{\beta}^1 \\ \vdots \\ \hat{\beta}^J \end{bmatrix}\right)^{-1} \begin{bmatrix} \hat{\beta}^1 \\ \vdots \\ \hat{\beta}^J \end{bmatrix}^T \begin{bmatrix} r_1 \\ \vdots \\ r_J \end{bmatrix}. \tag{21}$$

Finally, the estimates of r_j, $j = 1, \ldots, J$, are as follows:

$$\begin{bmatrix} \hat{r}_1 \\ \vdots \\ \hat{r}_J \end{bmatrix} = \begin{bmatrix} \hat{\beta}^1 \\ \vdots \\ \hat{\beta}^J \end{bmatrix} \hat{C}_d. \tag{22}$$

Similarly, the least-squares estimates $\hat{C}_0$, $\hat{C}_1$ and $\hat{C}_2$ of C_0, C_1 and C_2, can be calculated, and the estimates of $r_{j,0}$, $r_{j,1}$, and $r_{j,2}$ are then obtained as in (22). Substituting them into (17) would bring forth $\hat{P}_j^B$, the estimate of P_j^B. The MMSE of $\vec{d}_{j,n}^B$ is then implemented as

$$\hat{\vec{d}}_{j,n}^B = \hat{P}_j^B(\hat{P}_j^B + P^\varepsilon)^{-1}\vec{d}_{j,n}^Y. \tag{23}$$

Since the noise component of $d_{j,n}^Y$ is relatively stronger at finer scales than at coarser scales, we only use the measurements at coarser scales to improve the estimations at finer scales. Once estimation $\hat{\vec{d}}_{j,n}^B$ is computed, the elements $\hat{d}_{j-1,2n}^B$ and $\hat{d}_{j-1,2n+1}^B$ inside it are extracted as the final estimation results of finer scale $j-1$. Similarly, the estimate $\hat{d}_{j,n}^B$ at scale j is extracted from the estimation vector $\hat{\vec{d}}_{j+1,n}^B$. At the coarsest scale J, both $\hat{d}_{J,n}^B$ and $\hat{d}_{J-1,n}^B$ are obtained from $\hat{\vec{d}}_{j,n}^B$.

IV. Simulation

We compare the proposed fBm denoising method that exploits wavelet intra- and inter-scale dependencies of fBm signals with the algorithm of Wornell and Oppenheim (WO) [6]. In the WO scheme, the parameters r_j, σ_ε^2, and β are estimated simultaneously by a maximum-likelihood-based algorithm, and then the MMSE process is carried out independently at each scale. The WO method assumes that the fBm wavelet coefficients are totally whitened and it does not exploit the wavelet dependencies.

In this section, both the WO algorithm and the proposed scheme were implemented with Haar wavelet. By our experiments, using other wavelets with higher vanishing moment R to implement the WO algorithm does not offer better results. (The reason may be that the shortest support of Haar wavelet makes it posses more advantages in describing the self-similarity property of fBm processes.) And higher vanishing moment R will make the interscale correlation very weak (referring to (8)), so for the proposed scheme that exploits the interscale dependencies in denoising, higher R does not improve the performance either. Actually, our experiments show that the denoising performance of the proposed scheme is deteriorated by more regular wavelets.

In our experiments, the fBm signals were simulated by the Cholesky decomposition approach [2], which is a very accurate realization of fBm process. The sample length of the realized fBm signal is 512. We ran simulations for different levels of signal roughness (different fractal dimensions), with the Hurst parameter H being $0.3, 0.5, 0.7$, and 0.9, respectively. And for each value of H, we imposed different levels of additive *Gaussian* white noise on the generated fBm signals.

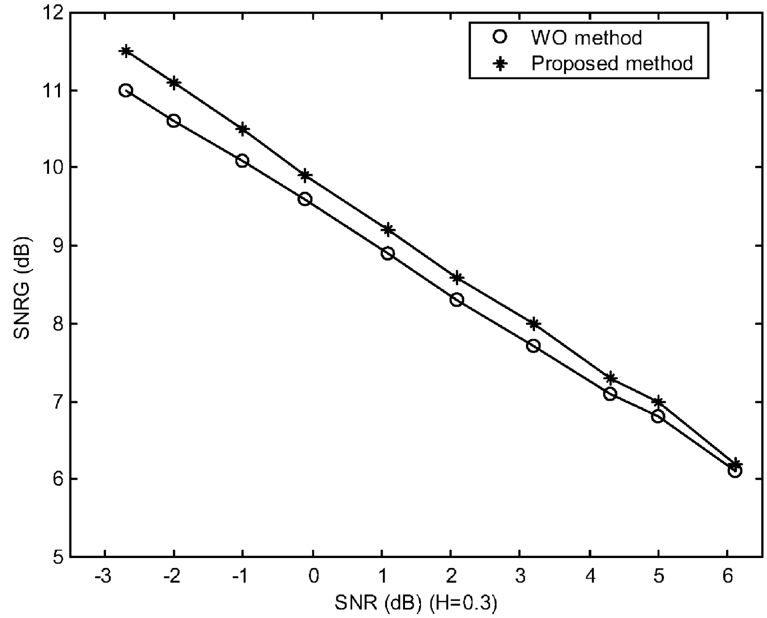

Fig. 2. The average SNRG results of the proposed method and the WO method with parameter H being 0.3. For each noise level, we ran simulations for 1000 times and calculated the average SNRG.

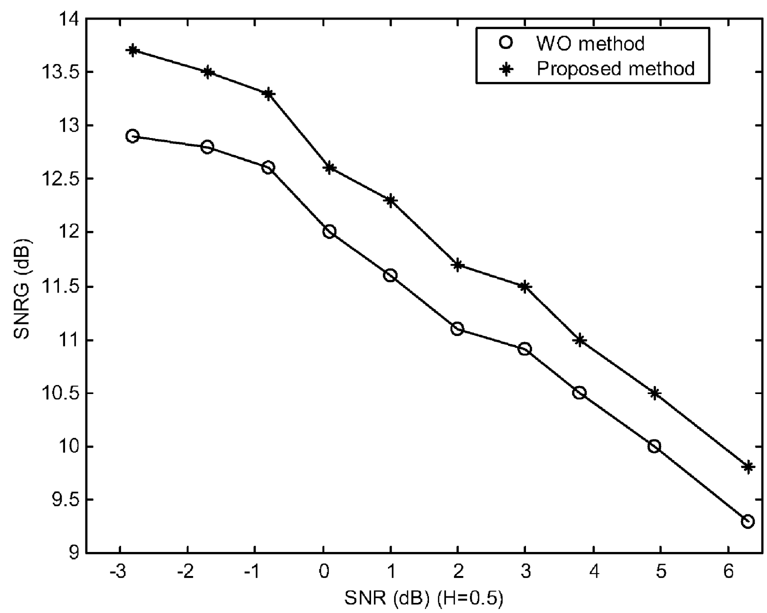

Fig. 3. The average SNRG results of the proposed method and the WO method with parameter H being 0.5. For each noise level, we ran simulations for 1000 times and calculated the average SNRG.

The proposed method and the WO method were then used to recover the true fBm signals from the noisy environments. For a fixed value of H and for a given noise level, the simulation was run for 1000 times to calculate the average signal-to-noise ratio gain (SNRG). Here, the signal-to-noise ratio (SNR) and SNRG are defined as follows. Referring to (10), the SNR of Y, the observation of B_H, is

$$\mathrm{SNR} = 10\log_{10}\frac{E[B_H^2]}{E[\varepsilon^2]}. \tag{24}$$

Denote by $\hat{B}_H$ the estimation of B_H from Y, we define the SNRG of $\hat{B}_H$ as

$$\mathrm{SNRG} = 10\log_{10}\frac{E[\varepsilon^2]}{E[(B_H - \hat{B}_H)^2]}. \tag{25}$$

In Figs. 2–5, we illustrate the SNRG results of the estimated fBm signal versus the SNR of the noisy fBm signal. It can be seen that the proposed scheme performs better than the WO algorithm. And as the parameter value H increases, the improvement of the proposed scheme over the WO scheme becomes more significant. When $H = 0.3$, the proposed method achieves about 0.3-dB improvements over the WO algorithm. The gap in estimation performance increases to about 0.6, 0.9, and 1.2 dB, respectively, for $H = 0.5$, $H = 0.7$, and $H = 0.9$. This should be well anticipated, because a larger value of H makes the fBm signal smoother, and creates greater dependencies inter- and intra-wavelet scales. These dependencies are well exploited by our scheme to the benefit of noise reduction.

Fig. 6 plots an fBm signal ($H = 0.7$) with 512 samples and its noisy version with SNR being -3.01 dB. The estimated signals are illustrated

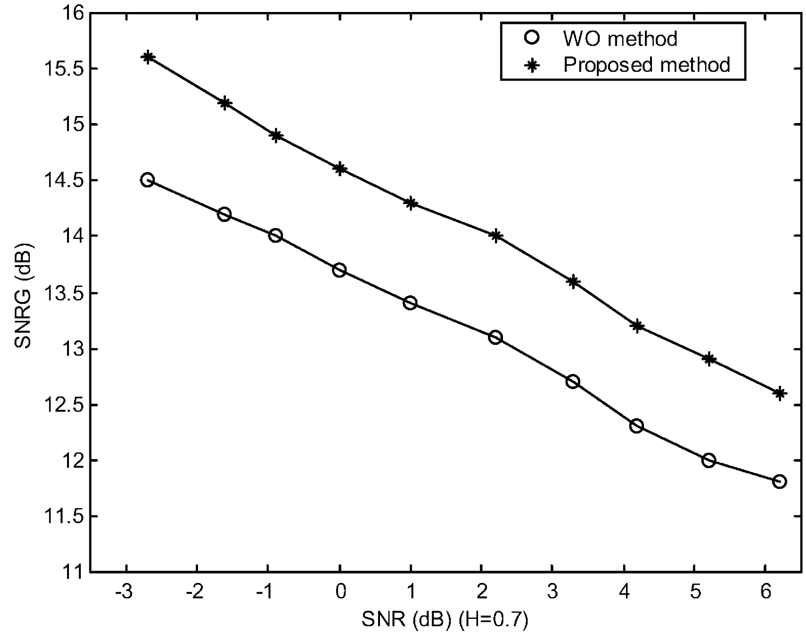

Fig. 4. The average SNRG results of the proposed method and the WO method with parameter H being 0.7. For each noise level, we ran simulations for 1000 times and calculated the average SNRG.

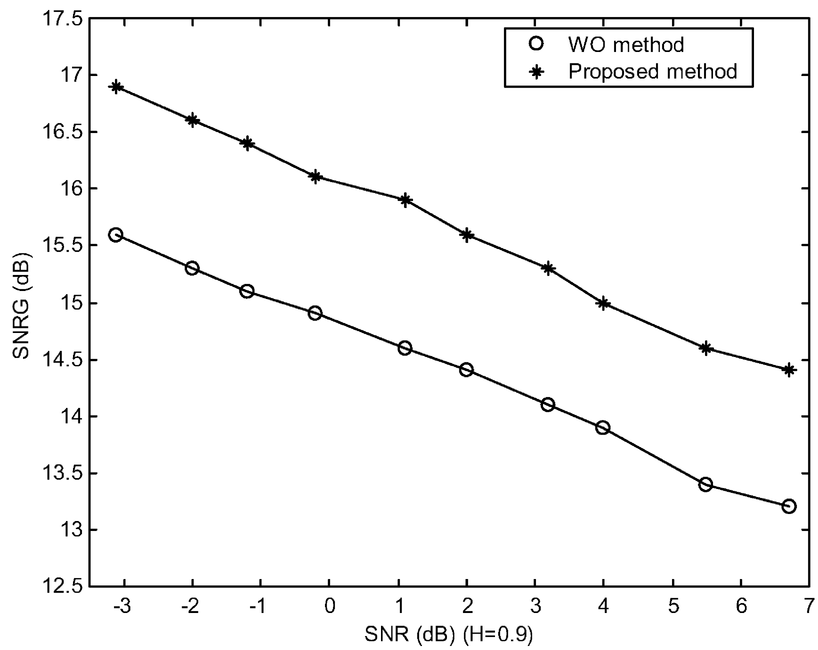

Fig. 5. The average SNRG results of the proposed method and the WO method with parameter H being 0.9. For each noise level, we ran simulations for 1000 times and calculated the average SNRG.

in Fig. 7, where the SNRG by the WO algorithm is 14.05 dB and that by our scheme is 15.03 dB. As anticipated, our model efficiently utilizes the scale–time dependency information to improve the estimation performance.

V. Conclusion

A new wavelet-based method for estimating fBm signals embedded into noises was developed in this correspondence. Although wavelets are approximate whitening filters for fBm processes, significant correlations are still prevalent across adjacent wavelet scales and between neighboring coefficients of a given scale. To exploit these intra- and inter-scale dependencies, we form a random vector of ancestor–descendant-associated wavelet coefficients, to which the MMSE technique is applied to estimate the fBm signals. Our simulations showed that the exploitation of the additional wavelet scale–time dependencies improves the estimation performance. It is also illustrated that the proposed estimation scheme works better for larger values of H, because the wavelet inter- and intra-scale dependencies of fBm signals increase

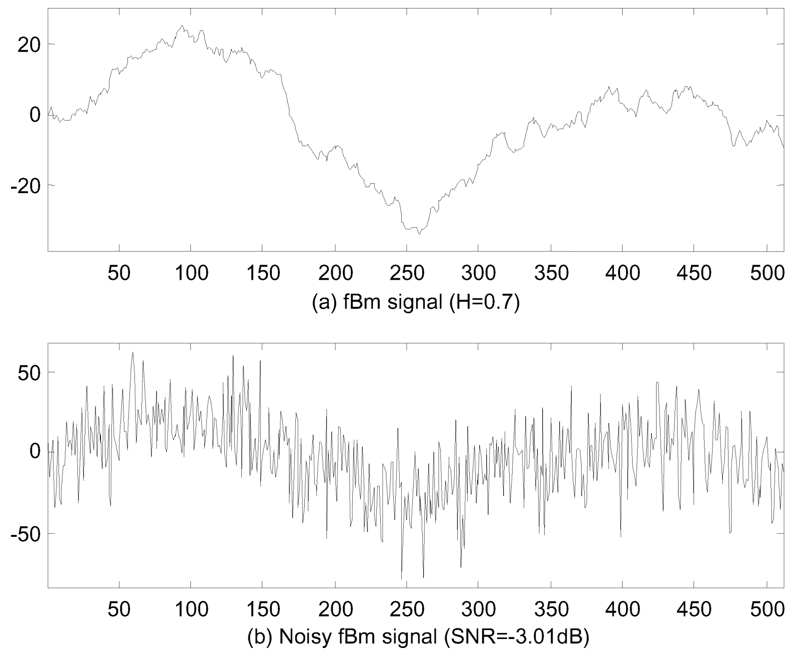

Fig. 6. (a) A simulated original fBm signal with $H = 0.7$. (b) The noisy version of the fBm signal (SNR $= -3.01$ dB).

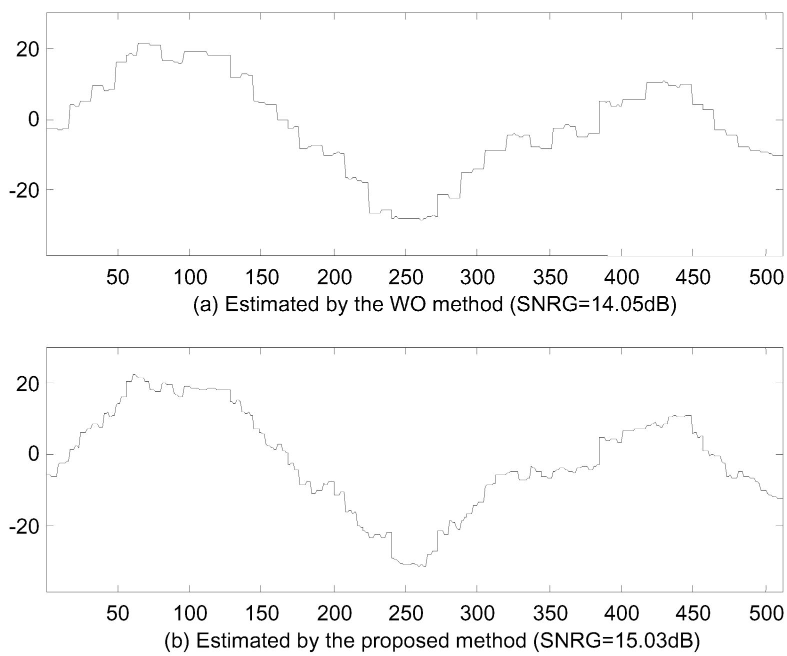

Fig. 7. (a) Estimated fBm signal by the WO algorithm (SNRG $=$14.05 dB). (b) Estimated fBm signal by the proposed intra- and inter-scale dependencies exploited scheme (SNR $=$15.03 dB).

in H, which are efficiently exploited by our scheme to improve the signal estimates.

References

[1] B. B. Mandelbrot and B. J. Van Ness, "Fractional Brownian motion, fractional noises and applications," *SIAM Rev.*, vol. 10, no. 4, pp. 422–438, 1968.

[2] T. Lundahl, W. J. Ohley, S. M. Kay, and R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. Med. Imag.*, vol. MI-5, pp. 152–161, Sept. 1986.

[3] P. Flandrin, "On the spectrum of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 35, pp. 197–199, Jan. 1989.

[4] ——, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 910–917, Mar. 1992.

[5] P. Abry, P. Flandrin, M. S. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation and synthesis of scaling data," in *Self-Similar Network Traffic and Performance Evaluation*, K. Park and W. Willinger, Eds. New York: Wiley, 2000, ch. 2, pp. 39–88.

[6] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noisy measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 40, pp. 611–623, Mar. 1992.

[7] G. W. Wornell, "A Karhunen–Loève-like expansion for $1/f$ process via wavelets," *IEEE Trans. Inform. Theory*, vol. 36, pp. 859–861, July 1990.

[8] P. W. Fieguth and A. S. Willsky, "Fractal estimation using models on multiscale trees," *IEEE Trans. Signal Processing*, vol. 44, pp. 1297–1300, May 1996.

[9] L. M. Kaplan and C.-C. J. Kuo, "Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis," *IEEE Trans. Signal Processing*, vol. 41, pp. 3554–3562, Dec. 1993.
[10] R. W. Dijkerman and R. R. Mazumdar, "On the correlation structure of the wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1609–1612, Sept. 1994.
[11] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inform. Theory*, vol. 38, pp. 904–909, Mar. 1992.
[12] R. W. Dijkerman and R. R. Mazumdar, "Wavelet representations of stochastic processes and multiresolution stochastic models," *IEEE Trans. Signal Processing*, vol. 42, pp. 1640–1652, July 1994.
[13] W.-L. Hwang, "Estimation of fractional Brownian motion embedded in a noisy environment using nonorthogonal wavelets," *IEEE Trans. Signal Processing*, vol. 47, pp. 2211–2219, Aug. 1999.
[14] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
[15] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.
[16] B. Jawerth and W. Sweldens, "An overview of wavelet based multiresolution analyzes," *SIAM Rev.*, vol. 36, pp. 377–412, 1994.
[17] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
[18] J. Liu and P. Moulin, "Information-theoretic analysis of inter-scale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Processing*, vol. 10, pp. 1647–1658, Nov. 2001.

Estimating Entropy on m Bins Given Fewer Than m Samples

Liam Paninski

***Abstract*—Consider a sequence p_N of discrete probability measures, supported on m_N points, and assume that we observe N independent and identically distributed (i.i.d.) samples from each p_N. We demonstrate the existence of an estimator of the entropy, $H(p_N)$, which is consistent even if the ratio N/m_N is bounded (and, as a corollary, even if this ratio tends to zero, albeit at a sufficiently slow rate).**

***Index Terms*—Approximation theory, bias, consistency, distribution-free bounds, entropy, estimation.**

Earlier work has examined the problem of estimating the entropy of a discrete distribution p, with support on $m < \infty$ "bins," given N independent and identically distributed (i.i.d.) samples from p. It has long been recognized [1] that the crucial quantity in this estimation problem is the ratio N/m: if the number of samples is much greater than the number of bins, the estimation problem is easy, and *vice versa*. This correspondence concentrates on the hard part of this problem: how do we estimate the entropy when N/m_N is bounded? (To allow the precise statement of asymptotic results, it is convenient here to let $m = m_N$ depend on N; see [2] for motivation, a brief review, and some recent results.) We show that a consistent estimator of the entropy exists in this regime (thus proving the main conjecture of [2]); the most surprising implication of this result is that it is possible to accurately estimate the entropy on m bins, given N samples, even when N/m is small (provided that both N and m are sufficiently large). We give an existence proof of this result here; see [2] for a more constructive demonstration of an estimator which numerically appears to have this interesting and useful consistency property.

Manuscript received March 11, 2003; revised March 10, 2004. This work was supported by predoctoral and postdoctoral fellowships from the Howard Hughes Medical Institute.
The author is with the Gatsby Computational Neuroscience Unit, University College London, London WC1N 3AR, U.K. (e-mail: liam@gatsby.ucl.ac.uk).
Communicated by A. B. Nobel, Associate Editor for Nonparametric Estimation, Classification, and Neural Networks.
Digital Object Identifier 10.1109/TIT.2004.833360

The entropy of a discrete distribution p is defined, as usual, as

$$H(p) = -\sum_{i=1}^{m} p_i \log p_i$$

where i indexes the support points of p, and the logarithm is taken to be natural. Our main result is as follows.

Theorem 1: Let $N/m_N \geq c > 0$, uniformly in N. Then there exists an estimator $\hat{H}_N$ for the entropy H which is uniformly consistent in mean square; that is,

$$E(\hat{H}_N - H)^2 < \epsilon(c, N)$$

with $\epsilon(c, N) \searrow 0$ as $N \to \infty$.

Note that the above statement is uniform over all distributions supported on m_N bins; the main practical implication, therefore, is that we can construct entropy estimators with surprisingly small "worst case" risk, given just m and N. We have as an easy corollary.

Corollary 2: There exists an estimator which is uniformly consistent even if $N/m_N \to 0$, sufficiently slowly.

More colloquially, we can estimate the entropy on m bins given fewer than m samples, as advertised. This is interesting in that it shows, in a sense, that the individual probabilities p need not be precisely estimated for the entropy estimate to be consistent.

On the other hand, in [2] we showed that N/m_N cannot decay faster than $N^{-\alpha}$, $\alpha > 0$, for consistency to hold, so the result is somewhat delicate. We present another partial converse here, indicating that not all functionals of the form $\sum_i f(p_i)$ can be estimated so easily, even for f smooth and vanishing at $p = 0$.

Proposition 3: Define the power sum

$$F(p) \equiv \sum_{i=1}^{m} p_i^{\alpha}, \qquad 0 < \alpha < 1.$$

If $\limsup N^{\alpha/(1-\alpha)}/m_N < \infty$, then

$$\liminf_{N} \inf_{\hat{F}_N} \max_{p} E(\hat{F}_N - F)^2 > 0$$

where the second infimum is taken over all possible estimators for F, and the maximum over all probability measures on m_N bins.

In particular, we need many more than m samples to estimate the power sum on m bins, whenever the exponent $\alpha \leq 1/2$. This result also quantifies the intuition that $F(p)$ becomes harder to estimate as α decreases (and, in fact, is impossible to estimate—in a "worst case" sense, at least—as $\alpha \to 0$, where we interpret $F(p)$ as counting the number of bins i for which $p_i > 0$).

The proof of the main theorem is built on ideas from [2]. Our estimator will be of the linear form

$$\hat{H}_{a,N} \equiv \sum_{j=0}^{N} a_{j,N} h_j$$

where the count statistics h_j are defined as

$$h_j \equiv \sum_{i=1}^{m} 1(n_i = j)$$