Principal component analysis of $1/f^{\alpha}$ noise

J.B. Gao [*], Yinhe Cao, Jae-Min Lee

Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

Abstract

Principal component analysis (PCA) is a popular data analysis method. One of the motivations for using PCA in practice is to reduce the dimension of the original data by projecting the raw data onto a few dominant eigenvectors with large variance (energy). Due to the ubiquity of $1/f^{\alpha}$ noise in science and engineering, in this Letter we study the prototypical stochastic model for $1/f^{\alpha}$ processes—the fractional Brownian motion (fBm) processes using PCA, and find that the eigenvalues from PCA of fBm processes follow a power-law, with the exponent being the key parameter defining the fBm processes. We also study random-walk-type processes constructed from DNA sequences, and find that the eigenvalue spectrum from PCA of those random-walk processes also follow power-law relations, with the exponent characterizing the correlation structures of the DNA sequence. In fact, it is observed that PCA can automatically remove linear trends induced by patchiness in the DNA sequence, hence, PCA has a similar capability to the detrended fluctuation analysis. Implications of the power-law distributed eigenvalue spectrum are discussed.
© 2003 Elsevier B.V. All rights reserved.

1. Introduction

Of the types of activity that characterize complex systems, the most ubiquitous and puzzling is perhaps the appearance of $1/f^{\alpha}$ noise, a form of temporal or spatial fluctuation characterized by a power-law decaying power spectral density. Some of the older literatures on this subject can be found, for example, in Press [1], Bak [2], and Wornell [3]. Some of the more recently discovered $1/f^{\alpha}$ processes are in traffic engineering [4–6], DNA sequence [7–9], human cognition [10], coordination [11], posture [12], dynamic images [13,14], the distribution of prime numbers [15], and multistable visual perception [16], among many others. In this Letter, we study $1/f^{\alpha}$ processes by principal component analysis (PCA).

PCA is closely related to singular value decomposition (SVD). In the continuous case, PCA is called Karhunen–Loève expansion. The latter is often called proper orthogonal decomposition (POD) in turbulence [17] and empirical orthogonal functions (EOFs) in meteorology [18]. Because of its conceptual simplicity and widely available codes based on well-studied numerical schemes, PCA is one of the most popular tools for data analysis. For example, PCA has been used to analyze DNA microarray data [19,20] and brain functional imaging data [21]. One of the motivations for using PCA is the expectation that the raw data may be projected onto a few dominant eigenvectors with large variance (or energy), thus the di-

mension of the raw data can be dramatically reduced. A fundamental question for us to ask is how often may an experimental dataset belong to this category? This question prompts us to consider PCA of the ubiquitous $1/f^\alpha$ noise to shed new light on when and how PCA may be used in practice.

The Letter is organized as follows. In Section 2 we briefly describe PCA, SVD, and Karhunen–Loève expansion, so that the Letter is self-contained. In Section 3 we analyze fractional Brownian motion (fBm) processes by PCA. We shall show that the distribution of the eigenvalues from PCA of fBm processes obeys a power-law, with the exponent being the key parameter defining the fBm processes. In Section 4 we study a random-walk-type process constructed from a DNA sequence, and show that the distribution of the eigenvalues from PCA of the constructed random-walk-type process is again a power-law, with the exponent correctly characterizing the correlation structure of the DNA sequence. Conclusions and discussions can be found in Section 5.

2. Brief overview of principal component analysis, singular value decomposition, and Karhunen–Loève expansion

First, we consider principal component analysis (PCA). PCA is the eigenanalysis of the autocorrelation (or auto-covariance) matrix R:

$$R = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ R_{21} & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nn} \end{pmatrix}.$$

Let λ_i and ϕ_i be the ith eigenvalue and corresponding eigenvector, then we have

$$R\phi_i = \lambda_i \phi_i, \quad i = 1, 2, \ldots, n.$$

Since the matrix R is symmetric and positive-definite, the eigenvalues are all positive, and the eigenvectors corresponding to different eigenvalues are orthogonal.

Next, we consider singular value decomposition (SVD). SVD of a matrix $A_{n \times m}$ is the following:

$$A = U \Sigma V^T,$$

where $U_{n \times n}$ and $V_{m \times m}$ are orthogonal matrices, $\Sigma_{n \times m}$ is a specific matrix which will be specified shortly, and

T denotes transpose of a matrix. The computation can be carried out by first forming AA^T or $A^T A$, then do eigenanalysis by finding all the positive eigenvalues and eigenvectors. The elements of Σ are all zero except $\Sigma_{ii} = \sigma_i$, for $i = 1, 2, \ldots, r$, where r is the number of positive eigenvalues from either AA^T or $A^T A$, and σ_i^2 is the ith eigenvalue of AA^T (or $A^T A$). The relation between PCA and SVD is clear if one forms the matrix A by taking delayed coordinates as its row vectors.

Finally, we consider the Karhunen–Loève expansion. Here, a signal $x(t)$ is expanded as

$$x(t) = \sum_{n=1}^{\infty} c_n \psi_n(t), \quad 0 < t < T,$$

where $\psi(t)$ is a set of orthonormal functions in the interval $(0, T)$:

$$\int_0^T \psi_n(t) \psi_m^*(t) \, dt = \delta[n - m]$$

and the coefficients c_n are random variables given by

$$c_n = \int_0^T x(t) \psi_n^*(t) \, dt,$$

where $*$ denotes complex conjugate. The basis functions $\psi(t)$ are the solutions to the following integral equation:

$$\int_0^T R(t_1, t_2) \psi(t_2) \, dt_2 = \lambda \psi(t_1), \quad 0 < t_1 < T,$$

where $R(t_1, t_2)$ is the autocorrelation function of the process $x(t)$. Note that the Karhunen–Loève expansion does not require the process $x(t)$ to be stationary.

With these background information, we now go to Section 3 to study PCA of the fractional Brownian motion (fBm) processes.

3. Principal component analysis of the fractional Brownian motion processes

A convenient framework for studying $1/f^\alpha$ processes is the self-affine stochastic processes $X = \{X(t),$

$t \geqslant 0\}$ defined by

$$X(\lambda t) =_d \lambda^H X(t), \quad t \geqslant 0 \tag{1}$$

for $\lambda > 0$, $0 < H < 1$, where $=_d$ denotes equality in distribution. The process is usually assumed to start at the origin. H is called the self-similarity parameter, or the Hurst parameter. The following properties can be easily derived from the above definition:

$$E[X(t)] = \frac{E[X(\lambda t)]}{\lambda^H} \qquad \text{mean}, \tag{2}$$

$$\text{Var}[X(t)] = \frac{\text{Var}[X(\lambda t)]}{\lambda^{2H}} \qquad \text{variance}, \tag{3}$$

$$R_x(t, s) = \frac{R_x(\lambda t, \lambda s)}{\lambda^{2H}} \qquad \text{autocorrelation}. \tag{4}$$

By arguing that $\lambda^{-H} X(\lambda t)$ and $X(t)$ should have the same power spectral density [22], one can obtain:

$$S(f) \sim f^{-\alpha} \tag{5}$$

with

$$\alpha = 2H + 1. \tag{6}$$

This explains why self-affine stochastic processes can be used to study $1/f^\alpha$ noise.

The prototypical model for self-affine stochastic processes is the fractional Brownian motion (fBm) process. FBm is a Gaussian process with mean 0, stationary increments, variance

$$E\left[\left(B_H(t)\right)^2\right] = t^{2H} \tag{7}$$

and covariance:

$$E\left[B_H(s)B_H(t)\right] = \frac{1}{2}\left\{s^{2H} + t^{2H} - |s - t|^{2H}\right\}, \tag{8}$$

where H is the Hurst parameter. When $H = 1/2$, fBm reduces to the standard Brownian motion (Bm) process. Bm is also called a Wiener process. When $0 < H < 1/2$, fBm is said to have negatively correlated increments, a property of anti-persistence [23]: a jump up is more likely to be followed by a jump down. When $1/2 < H < 1$, fBm is said to have persistent correlation [23]: a jump up is more likely to be followed by another jump up. Fig. 1 shows several fBm processes with different H. We observe that when H becomes larger, the process becomes less irregular and more trendy. That is what anti-persistence and persistence means.

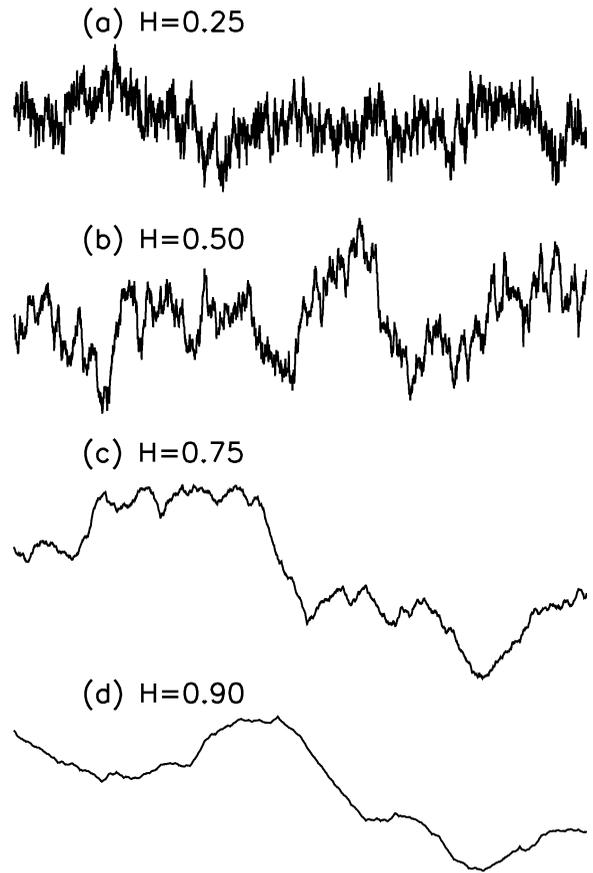Fig. 1. Several fBm processes with different H.

Since the power-spectral density (PSD) of a fBm process decays as

$$S(f) \sim \frac{1}{f^{2H+1}},$$

we propose the following conjecture:

Conjecture. *When n is large, the eigenvalue spectrum from PCA of a fBm process with parameter H decays as a power-law*: $\lambda_n \sim n^{-(2H+1)}$.

For a Bm process, the above conjecture can be analytically proven to be true. Denote a Bm defined in the interval $[0, T]$ by $w(t)$. Its Karhunen–Loève decomposition can be readily solved, giving [24] eigenfunctions

$$\psi(t) = \sqrt{\frac{2}{T}} \sin \omega_n t, \quad \omega_n = \frac{(2n+1)\pi}{2T}$$

and

$$w(t) = \sqrt{\frac{2}{T}} \sum_{n=1}^{\infty} c_n \sin \omega_n t,$$

$$c_n = \sqrt{\frac{2}{T}} \int_0^T w(t) \sin \omega_n t \, dt,$$

where the coefficients c_n are uncorrelated with variance being the eigenvalues, $E[c_n^2] = \lambda_n$. When the eigenvalues are sorted in a decreasing order, as is usually done, then the eigenvalue spectrum decays as a power-law:

$$\lambda_n \sim \frac{1}{\omega_n^2} \sim (n + 1/2)^{-2}. \tag{9}$$

When n is large, we can simply write

$$\lambda_n \sim n^{-2} = n^{-(2H+1)}, \quad H = 1/2. \tag{10}$$

We have numerically carried out PCA of fBm processes with many different H, and found the above conjecture to be always true. The fBm processes we analyzed are generated using fast Fourier transform filtering. That is, we start from a "white noise" sequence, whose spectral density, $S_W(f)$, is a constant, and filter the sequence with a transfer function $T(f) \propto f^{-(H+1/2)}$. The output sequence then has the desired spectral density, $S(f) \propto |T(f)|^2 S_W(f) \propto f^{-(2H+1)}$, and random phases. We then compute the auto-covariance matrix from the synthesized fBm processes with different H. The dimensions of the auto-covariance matrices range from 256×256 to 1024×1024. Eigenanalysis of those auto-covariance matrices always reveal a power-law decaying eigenvalue spectrum. Two examples for the auto-covariance matrices of size 1024×1024 are shown in Fig. 2 for $H = 0.25$ and $H = 0.75$. Clearly, we observe that when plotted in a log-log scale, the variation of the eigenvalue spectrum with the index is a straight line, with the slope of the line being $\alpha = 2H + 1$.

Since the eigenfunctions from the Karhunen–Loève expansion of the standard Bm processes are simple sine functions, to examine how similar the eigenvectors from the eigenanalysis of the auto-covariance matrices of fBm processes to the sine functions, we have plotted in Figs. 3 and 4 the first 10 leading eigenvectors for the fBm process with $H = 0.25$ and 0.75. We observe that when $H \neq 1/2$, the eigenvectors are no
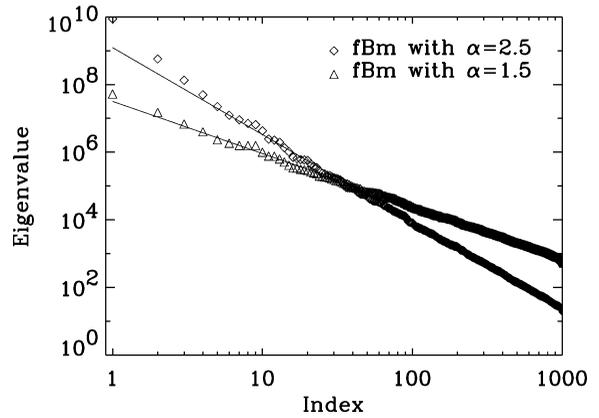

Fig. 2. Eigenvalue spectrum for the fBm processes with $H = 0.25$ and 0.75.

longer simple sine functions. This may indicate that analytically proving the above conjecture may be difficult. It is also of interest to note that the eigenvectors for the fBm process with $H = 0.25$ are less "smooth" than those of the fBm process with $H = 0.75$. This is consistent with Fig. 1 as well as the notion of anti-persistent and persistent correlations.

What are the relations between PCA and the wavelet decomposition of $1/f^\alpha$ processes? For simplicity, we consider dyadic orthonormal wavelet decomposition of fBm. Following the notations of [3], we write

$$x(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x_n^m \psi_n^m(t),$$

where m represents the time scale $t = 2^{-m}$, and the random variables x_n^m are wavelet coefficients. The wavelet coefficients across different time scales are only weakly correlated, hence, somewhat similar to the eigenvalues of PCA. A more remarkable similarity between PCA and wavelet decomposition of fBm is that the variance of x_n^m also decays with the time scale as a power-law, just as the decay of the eigenvalue spectrum of PCA with the index:

$$\text{var} \, x_n^m = \sigma^2 2^{-\alpha m} = \sigma^2 t^\alpha, \quad t = 2^{-m}.$$

However, if one wishes to index $\text{var} \, x_n^m$ by the scale index m so that $\text{var} \, x_n^m$ appear in a decreasing order, then $\text{var} \, x_n^m$ decreases with m exponentially fast. This contrasts with the power-law decay of the eigenvalue spectrum of PCA with the index. This can be considered a difference between PCA and the wavelet de-
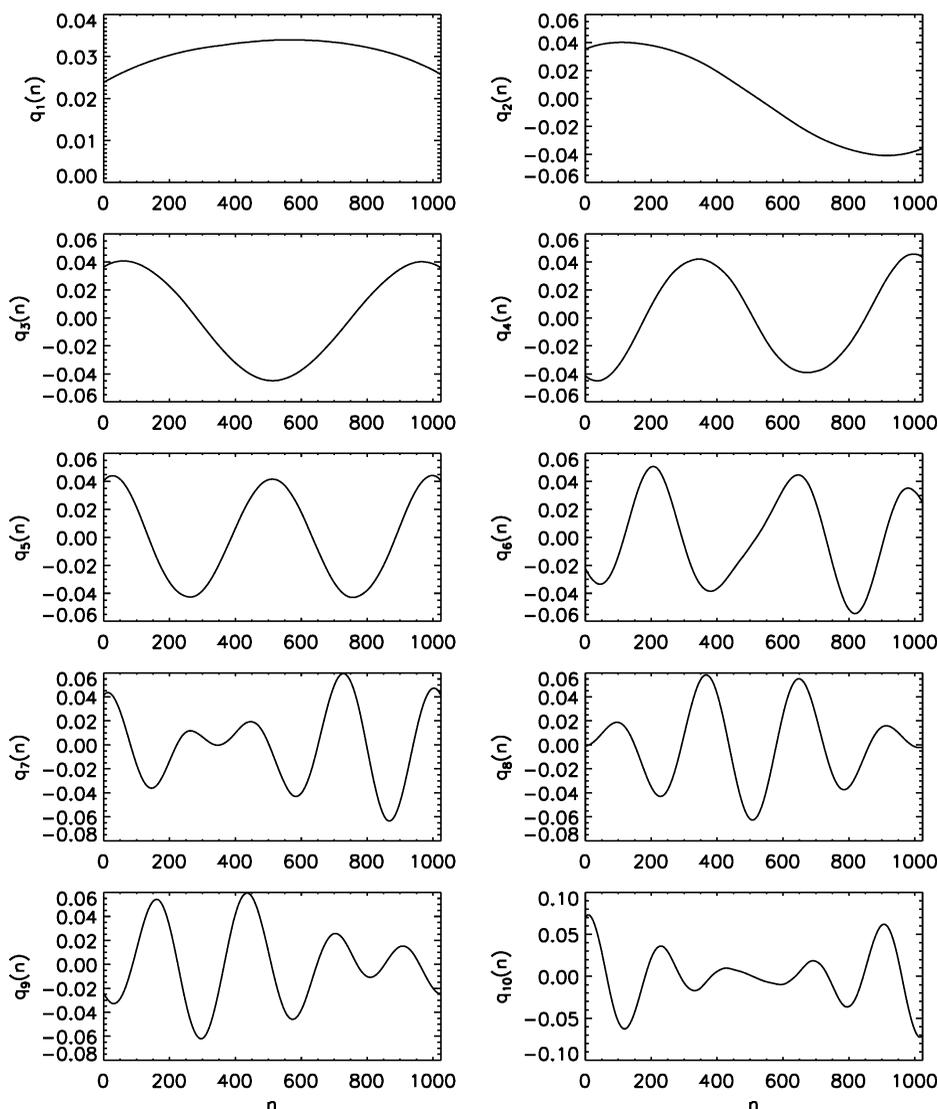
Fig. 3. The first 10 leading eigenvectors for the fBm process with $H = 0.25$.

composition of fBm. This difference may be due to the fact that wavelets are typically a lot more localized in time (and space) than the eigenfunctions shown in Figs. 3 and 4. At this point, it is of interest to note that the difference between PCA and wavelet decomposition has been considered by Field when applying them to describe the receptive fields of cells in the visual pathway (see [25] and references therein). One of the reasons that Field favors wavelets more than PCA is that PCA does not produce localized eigenfunctions to represent the receptive fields.

4. Principal component analysis of DNA sequences

A DNA sequence is made up of four nucleotides, adenine (A), guanine (G), thymine (T), and cytosine (C). A and G are purines, while T and C are pyrimidines. In recent years, the study of the correlation structures of a long DNA sequence has attracted much attention. For a recent review focusing on scaling features of DNA, see Stanley et al. [26]. It was discovered by Karlin et al. [27] that patchi-
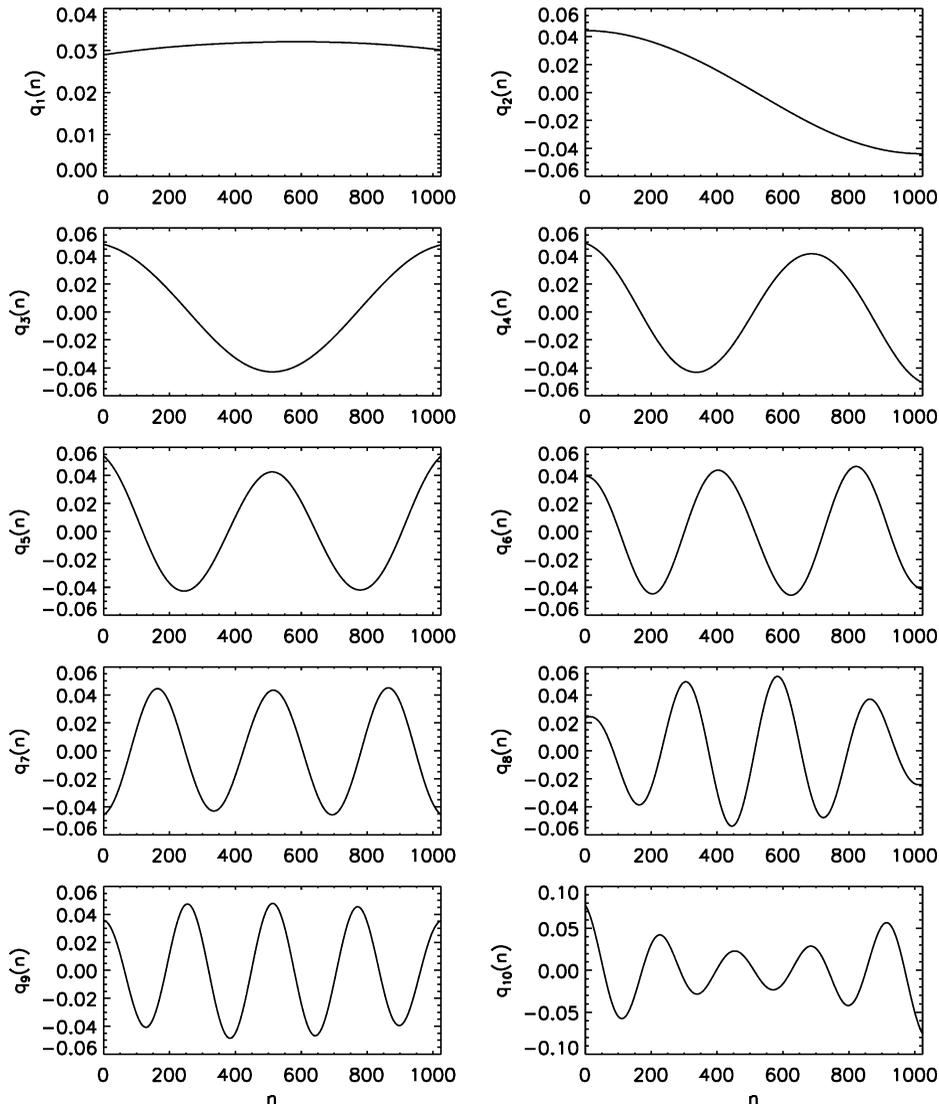
Fig. 4. The first 10 leading eigenvectors for the fBm process with $H = 0.75$.

ness, for example, a long patch of A or G in the bacteria lambda phage, may cause a DNA sequence to have long-range-correlations. To find genuine long-range-correlations not induced by linear trends including patchiness, Peng et al. [28] developed a method called detrended fluctuation analysis (DFA). Using DFA, the correlations in the lambda phage has indeed been eliminated. In this section, we analyze random-walk-type processes constructed from DNA sequences using PCA. We shall use the lambda phage to illustrate typical results.

Specifically, we have followed Peng et al. [28] by first mapping pyrimidines C and T to $u(i) = +1$, and purines A and G to $u(i) = -1$, then constructing a random-walk-type process

$$y(n) = \sum_{i=1}^{n} u(i).$$

Fig. 5(a) shows the random-walk process, $y(n)$ vs. n, for the bacteria lambda phage. Its eigenvalue spectrum is shown in Fig. 5(b), where we clearly observe a
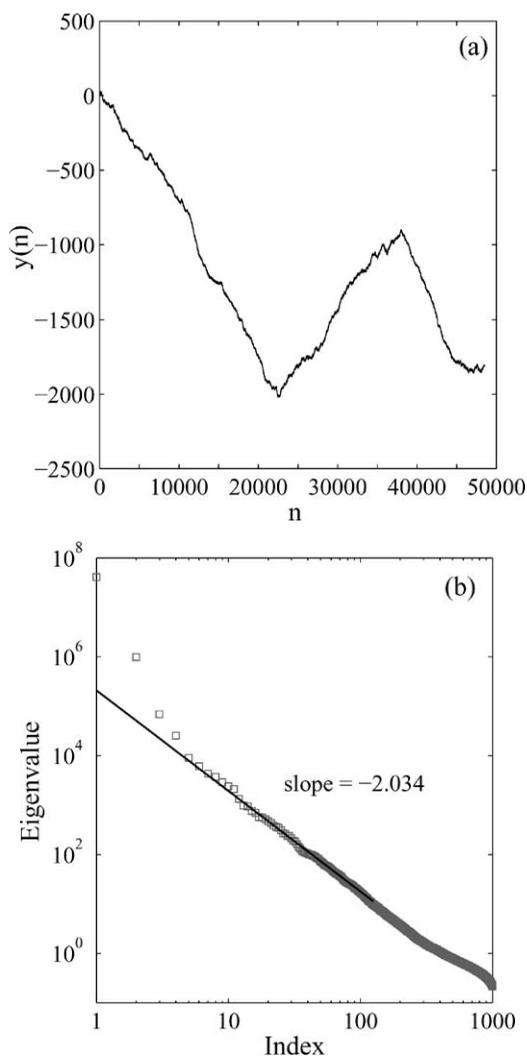
Fig. 5. The random-walk-type process constructed from the DNA sequence of the bacteria lambda phage (a) and its eigenvalue spectrum (b).

5. Conclusions and discussion

In this Letter, we have performed PCA of $1/f^\alpha$ noise modeled by fBm processes, and found the eigenvalues from PCA of fBm processes to follow a power-law, with the exponent being the key parameter defining the fBm processes. We have also studied random-walk-type processes constructed from DNA sequences, and found the eigenvalue spectrum from PCA of those random-walk processes to also follow power-law relations, with the exponent characterizing the correlation structures of the DNA sequence. In a matter of fact, we have found that PCA can automatically eliminate long-range-correlations due to patchiness in a DNA sequence, thus works similarly to DFA.

A nice example relevant to this study may be human vision. Various researchers have used various forms of factor analysis on human contrast sensitivity data to recover the spatial filtering characteristics of the individual mechanisms that underlie human contrast sensitivity. Typically, these analysis yield a number of bandpass filters with the lowest spatial frequency operator having a large eigenvalue and the progressively higher spatial frequency channels having smaller eigenvalues. It will be of interest to find out if the decay of eigenvalues follows a power-law, as in PCA, or an exponential distribution, as in wavelet decompositions, of $1/f$ processes.

Our finding has an immediate important implication that PCA can be conveniently used to study random fractal processes by finding the key parameter characterizing the process under study. Of course, eigenanalysis of a matrix may be computationally expensive, if the matrix is very big. When this is the case, other methods of analyzing $1/f^\alpha$ processes, including fluctuation analysis and detrended fluctuation analysis [26], may be preferred.

Our motivation for this study is to critically examine whether PCA can be used for reducing the dimension of an experimental dataset by projecting the raw data onto a few dominant eigenvectors with large variance (eigenvalues). The ubiquity of $1/f^\alpha$ processes imply that the eigenvalue distribution obtained using PCA is often a power-law. When this is true, then the dataset under study forms a random fractal. While the summation of the first few eigenvalues may contain most of the energy, projection of the raw dataset onto those few eigenvectors may however destroy the frac-

straight line in a log-log plot, indicating that the eigenvalue spectrum follows a power-law distribution. The slope of the straight line in Fig. 5(b) is 2.034, thus $H = 0.517$, very similar to the value estimated using DFA [28], which is 0.51. Hence, PCA has automatically removed the long-range-correlations induced by patchiness. Because H is very close to 0.5, it is considered [28] that the lambda phage DNA sequence does not have the long-range-correlations. We have analyzed many other DNA sequences and observed that PCA functions similarly to DFA.

tal features and correlation structures of the original dataset. This is evident if we consider the standard Bm. If we retain 90% of the energy, then about 10 eigenvalues and eigenfunctions will be chosen for projection. Superposition of 10 simple sine functions certainly will not constitute a true Bm process. Hence, before performing projection, we strongly recommend to examine whether the eigenvalues follow a power-law.

In practice, one may expect that power-law distributed eigenvalue spectrum is truncated. That is, the power-law only holds up to the Nth eigenvalue. If this is the case, then a natural way of reducing the dimension of the raw data would be to keep the first N eigenvectors and project the original data onto those eigenvectors.

Since PCA is only one of the popular methods for dimension reduction, our discussion should carry over to other methods for dimension reduction. These methods include clustering analysis and factor analysis. The latter, for example, has been widely used in the behavioral sciences. There, one of the most important questions that have to be addressed is how many factors to retain. In some analyses, it has been suggested that the number of factors retained should be a function of the grain of the data [29–31]. Those analyses were done by investigators not aware of fractals and $1/f$ noise, and thus, may need to be revisited. We emphasize that in this kind of studies, one should check whether factors may follow a power-law or an exponential distribution, as in the PCA and wavelet decomposition of $1/f$ noise, where the eigenvalues of PCA decay in a power-law manner while the wavelet coefficients decay with the scale index m exponentially fast. The exponential decay of factors is particularly interesting, since when the factors are plotted in linear scale and in a decreasing order, the first few factors will decrease sharply, while smaller factors may appear to remain more or less constant after the steep decline. This reminds us of the popular scree plot [32]. We recommend that before smaller factors are discarded, they should be checked if they follow either power-law or exponential distributions.

Acknowledgements

References

[1] W.H. Press, Comments Astrophys. 7 (1978) 103.
[2] P. Bak, How Nature Works: the Science of Self-Organized Criticality, Copernicus, 1996.
[3] G.M. Wornell, Signal Processing with Fractals: a Wavelet-Based Approach, Prentice Hall International, Englewood Cliffs, NJ, 1996.
[4] W.E. Leland, M.S. Taqqu, W. Willinger, D.V. Wilson, IEEE/ACM Trans. Netw. 2 (1994) 1.
[5] J. Beran, R. Sherman, M.S. Taqqu, W. Willinger, IEEE Trans. Commun. 43 (1995) 1566.
[6] V. Paxson, S. Floyd, IEEE/ACM Trans. Netw. 3 (1995) 226.
[7] W. Li, K. Kaneko, Europhys. Lett. 17 (1992) 655.
[8] R. Voss, Phys. Rev. Lett. 68 (1992) 3805.
[9] C.-K. Peng, S.V. Buldyrev, A.L. Goldberger, S. Havlin, F. Sciortino, M. Simons, H.E. Stanley, Nature 356 (1992) 168.
[10] D.L. Gilden, T. Thornton, M.W. Mallon, Science 267 (1995) 1837.
[11] Y. Chen, M. Ding, J.A.S. Kelso, Phys. Rev. Lett. 79 (1997) 4501.
[12] J.J. Collins, C.J. De Luca, Phys. Rev. Lett. 73 (1994) 764.
[13] V.A. Billock, Physica D 137 (2000) 379.
[14] V.A. Billock, G.C. de Guzman, J.A.S. Kelso, Physica D 148 (2001) 136.
[15] M. Wolf, Physica A 241 (1997) 493.
[16] J.B. Gao, I. Merk, W.W. Tung, V. Billock, K.D. White, J.G. Harris, V.P. Roychowdhury, Phys. Rev. Lett. (2003), submitted for publication.
[17] G. Berkooz, P. Holmes, J.L. Lumley, Annu. Rev. Fluid Mech. 25 (1993) 539.
[18] D. Dommenget, M. Latif, J. Climate 15 (2002) 216.
[19] N.S. Holter, M. Mitra, A. Maritan, M. Cieplak, J.R. Banavar, N.V. Fedoroff, Proc. Nat. Acad. Sci. USA 97 (2000) 8409.
[20] O. Alter, P.O. Brown, D. Botstein, Proc. Nat. Acad. Sci. USA 97 (2000) 10101.
[21] G. Zuendorf, N. Kerrouche, K. Herholz, J.C. Baron, Hum. Brain Mapp. 18 (2003) 13.
[22] D. Saupe, Algorithms for random fractals, in: H. Peitgen, D. Saupe (Eds.), The Science of Fractal Images, Springer-Verlag, Berlin, 1988, pp. 71–113.
[23] B.B. Mandelbrot, The Fractal Geometry of Nature, Freeman, San Francisco, 1982.
[24] A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw–Hill, New York, 1991, pp. 412–416.
[25] D.J. Field, Neural Comput. 6 (1994) 559.
[26] H.E. Stanley, S.V. Buldyrev, A.L. Goldberger, S. Havlin, C.K. Peng, M. Simons, Physica A 273 (1999) 1.
[27] S. Karlin, V. Brendel, Science 259 (1993) 677.
[28] C.K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Phys. Rev. E 49 (1994) 1685.

[29] L. Guttman, A new approach to factor analysis: the radix, in: P.F. Lazarfeld (Ed.), Mathematical Thinking in the Social Sciences, Free Press, New York, 1954, pp. 258–348.

[30] K.G. Joreskog, British J. Math. Stat. Psych. 23 (1970) 121.

[31] D.H. Peterzell, J.S. Werner, P.S. Kaplan, Vision Res. 33 (1993) 381.

[32] J.C. Nunnally, I.H. Bernstein, Psychometric Theory, McGraw–Hill, New York, 1994.