Journal of Computational and Applied Mathematics 236 (2011) 728–742

Contents lists available at SciVerse ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Hierarchical entropy analysis for biological signals

Ying Jiang [a], C.-K. Peng [b], Yuesheng Xu [c,a,*]

[a] Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou 510275, PR China

[b] Margret and H.A. Rey Institute for Nonlinear Dynamics in Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

[c] Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA

ARTICLE INFO

Keywords:
Sample entropy
Hierarchical decomposition
Heartbeat data

ABSTRACT

We develop a hierarchical entropy (HE) method to quantify the complexity of a time series based on hierarchical decomposition and entropy analysis. The proposed method is applied to the Gaussian white noise and the $1/f$ noise. We prove that the difference frequency components of the Gaussian white noise with the same scale factor have the same value of entropies, and the values decline as the scale factor increases. We also apply the HE method to the $1/f$ noise, and prove mathematically that a lower frequency component of a $1/f$ noise is also a $1/f$ noise and verify numerically that a higher frequency component of a $1/f$ random vector is approximately equal to a Gaussian random vector. The theoretical results are confirmed by numerical results. Moreover, we show that the HE method is an efficient method to analyze heartbeat signals by applying it to the cardiac interbeat interval time series of healthy young and elderly subjects, congestive heart failure (CHF) subjects and atrial fibrillation (AF) subjects.

1. Introduction

The complexity of a biological system is to measure its "meaningful structural richness", which exhibits relatively higher regularity than the output of random phenomena, where the word "regularity" means the predictability of the meaningful structure in the output. Understanding the complexity of a biological system is vitally important in both theory and practice. The complex fluctuations exhibited by an output signal of a physiologic system often contain information of underlying interacting mechanisms which regulate the system. Such mechanisms operate across multiple spatial and temporal scales.

Traditional entropy-based analysis methods, such as approximate entropy and sample entropy, evaluate the number of appearances of repetitive patterns of a time series to quantify the degree of its regularity. Uncorrelated random signals (white noise) are highly unpredictable but not structurally complex (see, [1]). However, since the traditional entropy methods only consider the unpredictability of signals, they assign to these signals the highest entropy values. As a result, the traditional entropy measures grow monotonically with the degree of randomness. When applied to the physiologic time series, traditional entropy-based methods may lead to misleading results. They may assign higher entropy values to certain pathologic cardiac rhythms that generate erratic outputs than to healthy cardiac rhythms that are exquisitely regulated by multiple interacting control mechanisms [1]. To overcome the shortcoming of the traditional entropy-based methods and give more precise descriptions of the complexity of physiologic signals, the multiscale entropy (MSE) method was recently introduced in [1] to analyze the complexity of the system. It considers not only the entropy of the original time series but also that entropy of different scales of the time series, where a scale is generated by averaging its previous scale. It incorporates the interrelationship of entropy and scale. Compared to traditional entropy methods, MSE has the advantages

* Corresponding author at: Department of Mathematics, Syracuse University, Syracuse, NY 13244, USA.
E-mail addresses: yjiang80@gmail.com (Y. Jiang), yxu06@syr.edu, xuyuesh@mail.sysu.edu.cn (Y. Xu).

0377-0427/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2011.06.007

of being applicable to both physiologic and physical signals of finite length. Results of using the MSE analysis are consistent with the consideration that both completely ordered and completely random signals are not complex [1–5]. In particular, the MSE methods show that correlated random signals (colored noise) are more complex than uncorrelated random signals (white noise).

A generic feature of pathologic dynamics should reappear in different scales of the output signal, in contrast to the outputs of random phenomena, which may disappear in lower frequency components of a large scale. Each scale of a time series has a lower frequency component which is generated by averaging the components in the previous scale and a higher frequency component which is generated by taking the difference of two consecutive scales. The generic feature may be stored in the lower frequency component or in the higher frequency component or in both. Hence, considering only low frequency components of the multiple scales of a time series is not adequate. The entropy of the higher frequency components of a time series should provide useful information in addition to that encoded in the lower frequency components. The hierarchical entropy analysis proposed in this paper takes into consideration the entropy of the higher frequency components of a time series.

This paper is organized in six sections. In Section 2, we recall the notions of sample entropy and multiscale entropy, and introduce the hierarchical entropy method. Section 3 is devoted to the hierarchical entropy analysis of the Gaussian white noise. We show that the hierarchical components of the real Gaussian random vector are all real Gaussian random vectors. Based on this observation, results regarding the hierarchical entropy analysis of the Gaussian random vector are obtained. These theoretical results are verified by a numerical example. We present in Section 4 the hierarchical entropy analysis of the $1/f$ noise. We prove mathematically that a lower frequency component of a $1/f$ random vector is also a $1/f$ random vector and verify numerically that a higher frequency component of a $1/f$ random vector is approximately equal to a Gaussian random vector. In Section 5, we apply the hierarchical entropy method to analyze heartbeat signals. We show that the hierarchical entropy analysis is an efficient method for analysis of physiological data. It provides additional useful information which leads to improvement of the existing multiscale entropy method. We close the paper with a conclusion section.

2. The hierarchical entropy method

We introduce in this section a hierarchical entropy method to quantify the complexity of a physiologic system. For this purpose, we first review the notion of the sample entropy and the multiscale entropy.

The sample entropy was introduced in [6] to measure the "complexity" of a time series data set sampled from a continuous process. It quantifies the degree of regularity of a system by evaluating the number of appearances of repetitive patterns in its output time series. It is based on a theoretical sample entropy model which estimates the conditional probability that two vectors that are close to each other for m components will remain close at the next component (see, [6]). To describe the concept of the sample entropy, for an $n \in \mathbb{N}$ we define notations $\mathbb{Z}_n := \{0, 1, \ldots, n-1\}$ and $\mathbb{Z}_n^+ := \{1, 2, \ldots, n\}$. For a given time series $\mathbf{x} := (x_j : j \in \mathbb{Z}_N)$ and $m \in \mathbb{Z}_N$, we let

$$u_m(j) := [x(j+k) : k \in \mathbb{Z}_m] \quad \text{for } j \in \mathbb{Z}_{N-m+1},$$

which are the pattern templates embedded in the time series $\mathbf{x}$. We then construct a sequence

$$\mathbf{u}_m := (u_m(j) : j \in \mathbb{Z}_{N-m+1}).$$

The positive integer m is the length of the pattern templates. The distance between two vectors $u_m(\ell)$ and $u_m(j)$ is defined by

$$d[u_m(\ell), u_m(j)] := \max\{|x(\ell+k) - x(j+k)| : k \in \mathbb{Z}_m\}.$$

We are interested in computing the probability of vectors in the sequence $\mathbf{u}_m$ within a given distance from a fixed vector. Specifically, for a given tolerance $r > 0$ and for a fixed $\ell \in \mathbb{Z}_N$, we let A_ℓ^m denote the number of vectors $u_m(j)$ with $j > \ell$ which satisfy the condition $d[u_m(j), u_m(\ell)] \le r$. We select $j > \ell$ to avoid double counting the matched vectors or counting $u_m(\ell)$ as a matched vector. We call the positive real number r the tolerance for accepting matches and the vector $u_m(\ell)$ the template. The probability of vectors $u_m(j) \in \mathbf{u}_m$ that are within tolerance r of the vector $u_m(\ell)$ is then given by

$$C_\ell^m(\mathbf{x}, r) := \frac{A_\ell^m}{N-m+1}.$$

We then sum these probabilities for all $\ell \in \mathbb{Z}_{N-m+1}$ to obtain

$$C^m(\mathbf{x}, r) := \sum_{\ell \in \mathbb{Z}_{N-m+1}} C_\ell^m(\mathbf{x}, r).$$

The sample entropy of the time series $\mathbf{x}$ is now defined by

$$S_m(\mathbf{x}, r) := -\ln \frac{C^{m+1}(\mathbf{x}, r)}{C^m(\mathbf{x}, r)}. \tag{2.1}$$

The sample entropy $S_m(\mathbf{x}, r)$ measures the degree of randomness of the time series $\mathbf{x}$ with the tolerance r. Fast algorithms for computing the sample entropy were presented in [7]. However, as noted in [8], there is no straightforward relationship between the entropy-based regularity and complexity.

The multiscale entropy method was introduced in [1] to measure the "complexity" of a time series aiming at overcoming the shortcoming of traditional entropy methods such as the sample entropy. According to [2], the complexity of a biological system should measure its ability to adapt and function in an ever-changing environment, a biological system operates across multiple spatial and temporal scales and thus its complexity is also multiscaled. Moreover, a wide class of disease states and aging reduce the adaptive capacity of the individual and degrade the information carried by output variables (see, [2]). Based on these consideration, the multiscale entropy method measures the complexity of the time series by using the entropy of its lower frequency components of different scales. Specifically, for a given time series $\mathbf{x} := (x_\ell : \ell \in \mathbb{Z}_N)$, we construct a sequence of time series $\mathbf{y}_n := (y_j^{(n)} : j \in \mathbb{Z}_{N_n})$, for $1 \le n \le N$, by

$$y_j^{(n)} := \frac{1}{n} \sum_{\ell=jn}^{(j+1)n-1} x_\ell, \quad j \in \mathbb{Z}_{N_n},$$

where $N_n := \left\lfloor \frac{N}{n} \right\rfloor$, with $\lfloor x \rfloor$ being the largest integer not greater that x. Note that $\mathbf{y}_1 = \mathbf{x}$ and for $n \ge 2$, these time series present lower frequency components of the original time series in different scales. The multiscale entropy is then obtained by calculating the sample entropy of each time series $\mathbf{y}_n$ and it is the vector whose components are the sample entropies $S_m(\mathbf{y}_n, r)$. The multiscale entropy method takes into account the information contained in multiple scales of the original time series and use a vector of scale-dependent entropies to describe the "complexity" of a time series. Since it illustrates the variation of information redundancies appeared in multiple scales, the multiscale entropy method can describe the "complexity" of a time series more meaningfully than traditional entropy-based methods. For example, traditional entropy-based methods assign the highest values to white noise which is highly unpredictable. From a global point of view, white noise has a simple description and does not contain any information. The multiscale entropy method can present very well such a feature of white noise by using a vector of scale-dependent entropies. See [2] for more discussion of this point.

While the multiscale entropy method focuses on the lower frequency components of different scales of a time series, it ignores the information in the higher frequency components which are the differences between two consecutive scales. Moreover, the original time series cannot be reconstructed by using only its lower frequency components. For a time series, information may be stored in a multiscale manner in its lower frequency components, or in its high frequency components or in both lower and higher frequency components. The multiscale entropy measures the complexity very well for those time series whose information is mainly stored in its lower frequency components. But it will miss the information stored in a high frequency component. This observation leads us to introducing the hierarchical entropy method.

We now introduce the hierarchical entropy (HE) method. We define an averaging operator Q_0 for the time series $\mathbf{s} := (s_j : j \in \mathbb{Z}_{2^n})$ of length 2^n by

$$Q_0(\mathbf{s}) := \left(\frac{s_{2j} + s_{2j+1}}{2} : j \in \mathbb{Z}_{2^{n-1}} \right).$$

The time series $Q_0(\mathbf{s})$ with length 2^{n-1} is the low frequency component of $\mathbf{s}$ at scale 2. We also define a difference operator Q_1 for time series $\mathbf{s}$ by

$$Q_1(\mathbf{s}) := \left(\frac{s_{2j} - s_{2j+1}}{2} : j \in \mathbb{Z}_{2^{n-1}} \right),$$

and we call $Q_1(\mathbf{s})$ a difference frequency component of $\mathbf{s}$. The time series $Q_1(\mathbf{s})$ with length 2^{n-1} is the high frequency component of $\mathbf{s}$ at scale 2. Note that the original time series $\mathbf{s}$ can be reconstructed from $Q_0(\mathbf{s})$ and $Q_1(\mathbf{s})$. In fact, we have that

$$\mathbf{s} = ((Q_0(\mathbf{s}))_j + (Q_1(\mathbf{s}))_j, (Q_0(\mathbf{s}))_j - (Q_1(\mathbf{s}))_j : j \in \mathbb{Z}_{2^{n-1}})^T.$$

As a result, the time series $Q_0(\mathbf{s})$ and $Q_1(\mathbf{s})$ constitute a two-scale analysis for the time series $\mathbf{s}$.

The operators Q_j for $j \in \mathbb{Z}_2$ have a matrix representation. Specifically, for $j \in \mathbb{Z}_2$, we observe that

$$Q_j := \begin{pmatrix} \frac{1}{2} & \frac{(-1)^j}{2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{(-1)^j}{2} & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \frac{1}{2} & \frac{(-1)^j}{2} \end{pmatrix}_{2^{n-1} \times 2^n}. \tag{2.2}$$

Note that the matrix form of these operators depends on the length of the time series to which they apply. For notational simplicity, we do not indicate their dependence on n since it can be understood from the context. The operator Q_0 and Q_1 correspond to the low and high pass filters of the Haar wavelet (cf., [9,10]), respectively.

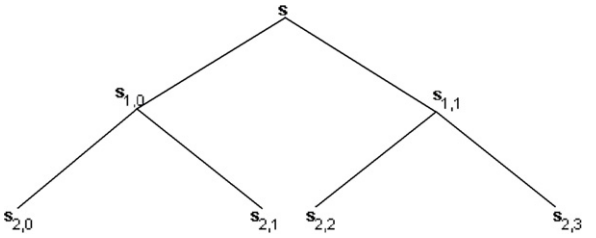

Fig. 1. Hierarchical decomposition of signal **s** with three scales.

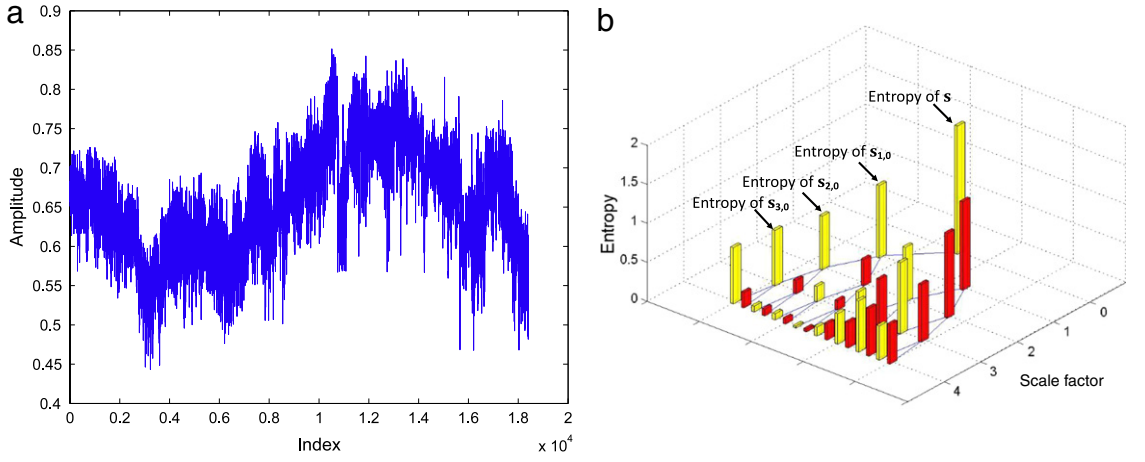

Fig. 2. HE analysis of simulative signal **s**.

To describe a multiscale analysis of the time series **s**, we need to repeatedly use these operators. Let N be a positive integer. For an $n \in \mathbb{Z}_N^+$ and $[\ell_1, \ell_2, \ldots, \ell_n] \in \mathbb{Z}_2^n$, the integer

$$e := \sum_{j=1}^{n} \ell_j 2^{n-j} \tag{2.3}$$

is nonnegative. On the other hand, when $n \in \mathbb{Z}_N^+$ is fixed, given a nonnegative integer e, there is a unique vector $[\ell_1, \ell_2, \ldots, \ell_n] \in \mathbb{Z}_2^n$ corresponding to it through Eq. (2.3). For $n \in \mathbb{Z}_N^+$ and a nonnegative integer e, we define the hierarchical components of a time series $\mathbf{s} \in \mathbb{R}^{2^N}$ by

$$\mathbf{s}_{n,e} := Q_{\ell_n} \circ Q_{\ell_{n-1}} \circ \cdots \circ Q_{\ell_1}(\mathbf{s}). \tag{2.4}$$

Note that for each $n \in \mathbb{Z}_N^+$, $\mathbf{s}_{n,0}$ is the low frequency component of the time series **s** in scale $n+1$. For $k \in \mathbb{Z}_{N+1}^+$, we define an index set $\mathbb{J}_k := \{(n, e) : n \in \mathbb{Z}_k, e \in \mathbb{Z}_{2^n}\}$. Let $\mathbf{s}_{0,0} := \mathbf{s}$. For a given $k \in \mathbb{Z}_{N+1}^+$, the signals $s_{n,e}$, $(n, e) \in \mathbb{J}_k$, consist of the hierarchical decomposition of signal **s** in k scales.

We find it convenient to arrange the hierarchical components of a time series **s** in a tree diagram. In this tree, the original time series $\mathbf{s}_{0,0}$ is the root node with the left child node $\mathbf{s}_{1,0}$ and the right child node $\mathbf{s}_{1,1}$, and for each $(n, e) \in \mathbb{J}_{k-1}$, node $\mathbf{s}_{n,e}$ has the left child node $\mathbf{s}_{n+1,2e}$ and the right child node $\mathbf{s}_{n+1,2e+1}$. We call this tree the *hierarchical tree* of **s**. In Fig. 1, we illustrate the hierarchical decomposition of **s** in three scales.

The hierarchical decomposition of a time series can be viewed as a generalization of both the multiscale decomposition developed in [2] and the Haar wavelet decomposition since these decompositions generate subtrees of the hierarchical tree. It preserves the strength of the multiscale decomposition with additional components of higher frequency in different scales. The hierarchical decomposition, unlike the wavelet decomposition which has no redundant component, contains redundant components. This allows us to catch the dynamical richness of the time series.

With a hierarchical tree $\mathbf{s}_{n,e}$, $(n, e) \in \mathbb{J}_k$, in place, we compute the sample entropy of each component $\mathbf{s}_{n,e}$ and use these entropies to measure the complexity of the biological system which has the time series **s** as its output variables. We call this process the *hierarchical entropy* (HE) analysis.

To close this section, we demonstrate the HE analysis by a simulative time series. The time series **s** that we consider in this example is shown in Fig. 2(a). The hierarchical entropy values of **s** are presented in Fig. 2(b). We observe from Fig. 2(b) that the entropy values of $\mathbf{s}$, $\mathbf{s}_{1,0}$, $\mathbf{s}_{2,0}$ and $\mathbf{s}_{3,0}$ decrease as the scale n increases. Although the entropy values of $\mathbf{s}_{n,e}$, for $n = 2, 3, 4$ and $e \in \mathbb{Z}_{2^{n-1}}^+$, are very small, in the subtree with root node $\mathbf{s}_{1,1}$, the entropy values of its nodes are significant. This shows

that the components $\mathbf{s}_{n,e}$, $n \in \mathbb{Z}_4^+$ and $e \in \{2^{n-1}, 2^{n-1}+1, \ldots, 2^n-1\}$ contain significant information of the original time series $\mathbf{s}$, which might have been ignored if only the multiscale entropy analysis were computed.

3. Analysis of the Gaussian white noise

In this section, we analyze the Gaussian white noise by employing the hierarchical entropy method. Both theoretical and numerical results will be presented.

We begin with recalling two results about the Gaussian white noise. They are needed for the analysis presented in this section. For a positive integer N, we suppose that $\mathbf{x} := [x_j : j \in \mathbb{Z}_{2^N}]$ is a random vector taking values in $\mathbb{R}^{2^N}$. When the components x_j, $j \in \mathbb{Z}_{2^N}$, of $\mathbf{x}$ are independent and have the same Gaussian distribution, that is, they have the same mean and standard deviation, we call $\mathbf{x}$ a real Gaussian random vector, a component of $\mathbf{x}$ a Gaussian random variable and an instance of $\mathbf{x}$ the Gaussian white noise. In this section, we let $\mathbf{g} := [g_j : j \in \mathbb{Z}_{2^N}]$ denote the real Gaussian random vector, with the mean of g_j being 0 and the standard derivation of g_j being δ. Recall that the sample entropy estimates the conditional probability that two vectors that are close to each other for m elements will remain close at the next element. In the case of Gaussian random vector $\mathbf{g}$, such a conditional probability equals to (see Appendix A of [2])

$$P(r) := \frac{1}{2\delta\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ \operatorname{erf}\left(\frac{x+r}{\sqrt{2}\delta}\right) - \operatorname{erf}\left(\frac{x-r}{\sqrt{2}\delta}\right) \right\} \mathrm{e}^{-x^2/2\delta^2} \mathrm{d}x, \quad r \in \mathbb{R}^+,$$

where erf denotes the error function defined by

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{e}^{-t^2} \mathrm{d}t, \quad x \in \mathbb{R}.$$

Thus, the theoretical value of the sample entropy of the Gaussian white noise, according to Appendix A of [2], is equal to

$$S_m(\mathbf{g}, r) = -\ln(P(r)). \tag{3.1}$$

Therefore, for $m \in \mathbb{Z}_{2^N-1}^+$ we obtain the sample entropy of the Gaussian white noise which is given by

$$S_m(\mathbf{g}, r) = -\ln\left(\frac{1}{2\delta\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left\{ \operatorname{erf}\left(\frac{x+r}{\sqrt{2}\delta}\right) - \operatorname{erf}\left(\frac{x-r}{\sqrt{2}\delta}\right) \right\} \mathrm{e}^{-x^2/2\delta^2} \mathrm{d}x \right). \tag{3.2}$$

Theorem 10.5 in [11] states that a linear combination of two independent real Gaussian random variables is also a real Gaussian random variable. To state this result precisely, we suppose that x and y are two real Gaussian random variables with mean 0 and standard deviation δ_x and δ_y, respectively. For $\alpha, \beta \in \mathbb{R}$, $\alpha x + \beta y$ is a real Gaussian random variable with mean 0 and standard deviation $\sqrt{\alpha^2\delta_x^2 + \beta^2\delta_y^2}$.

According to (2.4) the hierarchical decomposition of the real Gaussian random vector $\mathbf{g}$ is given for $n \in \mathbb{Z}_N^+$ and $[j_1, j_2, \ldots, j_n] \in \mathbb{Z}_2^n$, by

$$\mathbf{g}_{n,e} := Q_{j_n} \circ Q_{j_{n-1}} \circ \cdots \circ Q_{j_1}(\mathbf{g}),$$

where e is defined by (2.3) in terms of $j_1, j_2, \ldots, j_n$. We let $\mathbf{g}_{0,0} := \mathbf{g}$. By the definition of Q_j, $j \in \mathbb{Z}$ and the second fact that we just reviewed, we know that for each $j \in \mathbb{Z}$, the components of $\mathbf{g}_{1,j}$ are real Gaussian random variables with mean 0 and standard deviation $\delta/\sqrt{2}$. This implies that the random vectors $\mathbf{g}_{1,0}$ and $\mathbf{g}_{1,1}$ have the same statistical properties. Therefore, for a given $r \in \mathbb{R}^+$ and a given $m \in \mathbb{Z}_{2^N-1}^+$, we have that

$$S_m(\mathbf{g}_{1,0}, r) = S_m(\mathbf{g}_{1,1}, r).$$

On the other hand, since the standard deviation of $\mathbf{g}_{1,j}$ is smaller than that of $\mathbf{g}$, the instances of $\mathbf{g}_{1,j}$ are more regular than the instances of $\mathbf{g}$. Thus, we would expect that

$$S_m(\mathbf{g}_{1,j}, r) < S_m(\mathbf{g}, r).$$

The main purpose of this section is to present a general theorem that confirms the above conjecture. To this end, we first present a lemma that ensures that each component $\mathbf{g}_{n,e}$ of the hierarchical decomposition of $\mathbf{g}$ is also a real Gaussian random vector. As usual, the mathematical expectation of a random variable x taking values in $\mathbb{R}$ is defined by

$$E(x) := \int_{\mathbb{R}} tp(t)\mathrm{d}t,$$

where $p(t)$ is the probability density function for the random variable x. It is known from Theorem 10.5 in [11] that the components of a random vector $\mathbf{x}$ are independent if and only if $E(x_j x_{j'}) = 0$ for all $j \neq j'$.

Lemma 3.1. *If for a positive integer N, $\mathbf{g} := [g_j : j \in \mathbb{Z}_{2^N}]$ denotes a real Gaussian random vector with mean 0 and standard deviation δ, then for each $(n, e) \in \mathbb{J}_{N+1}$, $\mathbf{g}_{n,e}$ is a real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{n}{2}}$.*

Proof. We first show that for an arbitrary $\ell \in \mathbb{Z}_N^+$, if $\mathbf{x}_\ell := [x_j : j \in \mathbb{Z}_{2^\ell}]$ is a real Gaussian random vector with mean 0 and standard deviation δ_ℓ, then $\mathbf{y} := [y_j : j \in \mathbb{Z}_{2^{\ell-1}}]$ with $y_j := \alpha x_{2j} + \beta x_{2j+1}$, for some $\alpha, \beta \in \mathbb{R}$ (with $\alpha^2 + \beta^2 = 1/2$) is a real Gaussian random vectors with mean 0 and standard deviation $\delta_\ell/\sqrt{2}$. Since for each $i \in \mathbb{Z}_{2^\ell}$, x_j is a real Gaussian random variable with mean 0 and standard deviation δ_ℓ, from Theorem 10.5 in [11] we know that y_j is a real Gaussian random variable with mean 0 and standard deviation $\delta_\ell/\sqrt{2}$. It remains to show that the components of $\mathbf{y}$ are independent. Because the components of $\mathbf{x}_\ell$ are independent, we find that $E(x_j x_{j'}) = 0$, for $j \neq j'$. It follows that for $j, j' \in \mathbb{Z}_{2^{\ell-1}}$ with $j \neq j'$,

$$E(y_j y_{j'}) = \alpha^2 E(x_{2j}x_{2j'}) + \alpha\beta(E(x_{2j+1}x_{2j'}) + E(x_{2j}x_{2j'+1})) + \beta^2 E(x_{2j+1}x_{2j'+1}) = 0.$$

Thus, $y_j, j \in \mathbb{Z}_{2^{\ell-1}}$, are independent, and we conclude that $\mathbf{y} := [y_j : j \in \mathbb{Z}_{2^{\ell-1}}]$ is a real Gaussian random vector.

The proof of this lemma may be completed by induction on n. The statement of this lemma holds trivially for $n = 0$. We assume that the statement of this lemma holds for $n = \mu$, $\mu \in \mathbb{Z}_N$, that is, $\mathbf{g}_{\mu,\mathrm{e}'}$, $\mathrm{e}' \in \mathbb{Z}_{2^\mu}$, are real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{\mu}{2}}$. Note that for each $e \in \mathbb{Z}_{2^{\mu+1}}$, there exist integers $\mathrm{e}' \in \mathbb{Z}_{2^\mu}$ and $j \in \{0, 1\}$ such that

$$\mathbf{g}_{\mu+1,e} = Q_j \mathbf{g}_{\mu,\mathrm{e}'}.$$

In the definition of $\mathbf{y}$, by choosing $\alpha := 1/2$, $\beta := 1/2$, and $\alpha := 1/2$, $\beta := -1/2$, respectively, we obtain that both $Q_0(\mathbf{g}_{\mu,\mathrm{e}'})$ and $Q_1(\mathbf{g}_{\mu,\mathrm{e}'})$ are real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{\frac{\mu+1}{2}}$. That is, $\mathbf{g}_{\mu+1,e}$ is a real Gaussian random vector with mean 0 and standard deviation $\delta/2^{(\mu+1)/2}$. The induction principle ensures that for all $(n, e) \in \mathbb{J}_{N+1}$, $\mathbf{g}_{n,e}$ are real Gaussian random vectors with mean 0 and standard deviation $\delta/2^{n/2}$. □

We are now ready to present the main result of this section.

Theorem 3.2. *Suppose that $\mathbf{g} := [g_j : j \in \mathbb{Z}_{2^N}]$, with $N \in \mathbb{N}$, is a real Gaussian random vector with mean 0 and standard deviation δ and $m \in \mathbb{Z}^+_{2^N-1}$. Then the following statements hold:*

(i) *For all $r > 0$ and $(n, e) \in \mathbb{J}_{N+1}$,*

$$S_m(\mathbf{g}_{n,e}, r) = -\ln\left(\frac{1}{2\delta\sqrt{2\pi}}\int_{\mathbb{R}}\left\{\mathrm{erf}\left(\frac{y + 2^{n/2}r}{\delta}\right) - \mathrm{erf}\left(\frac{y - 2^{n/2}r}{\delta}\right)\right\}\mathrm{e}^{-y^2/\delta^2}\,\mathrm{d}y\right). \tag{3.3}$$

(ii) *For fixed r and n, $S_m(\mathbf{g}_{n,e}, r) = S_m(\mathbf{g}_{n,\mathrm{e}'}, r)$, for all $e, \mathrm{e}' \in \mathbb{Z}_{2^n}$.*

(iii) *For fixed r, if $n > n'$, then $S_m(\mathbf{g}_{n,e}, r) < S_m(\mathbf{g}_{n',\mathrm{e}'}, r)$, for all $e \in \mathbb{Z}_{2^n}$ and $\mathrm{e}' \in \mathbb{Z}_{2^{n'}}$.*

Proof. (i) From Lemma 3.1, we know that for each $(n, e) \in \mathbb{J}_{N+1}$, $\mathbf{g}_{n,e}$ is a real Gaussian random vector with mean 0 and standard deviation $\delta/2^{n/2}$. Thus, by Eq. (3.2) with a change of variables $y = 2^{n/2}x$, we obtain the formula (3.3).

(ii) Note that the right hand side of formula (3.3) depends only on n; it is independent of e. Hence, this statement follows immediately from (3.3).

(iii) By the definition of the error function erf, we know that erf is a strictly increasing function. Thus, we have that if $n < n'$,

$$\mathrm{erf}\left(\frac{y + 2^{n/2}r}{\delta}\right) < \mathrm{erf}\left(\frac{y + 2^{n'/2}r}{\delta}\right), \quad \text{and} \quad -\mathrm{erf}\left(\frac{y - 2^{n/2}r}{\delta}\right) < -\mathrm{erf}\left(\frac{y - 2^{n'/2}r}{\delta}\right).$$

It follows that for all $y \in \mathbb{R}$, if $n < n'$, then

$$\mathrm{erf}\left(\frac{y + 2^{n/2}r}{\delta}\right) - \mathrm{erf}\left(\frac{y - 2^{n/2}r}{\delta}\right) < \mathrm{erf}\left(\frac{y + 2^{n'/2}r}{\delta}\right) - \mathrm{erf}\left(\frac{y - 2^{n'/2}r}{\delta}\right).$$

Thus, the above inequality with formula (3.3) yields the inequality in (iii). □

To close this section, we present a numerical example which confirms the analytical results in Theorem 3.2. Specifically, we apply the HE method to a real Gaussian white noise $\mathbf{s}$, shown in Fig. 3(a), where entropies are calculated by Eq. (2.1), and compare the result of the HE analysis with the analytical result shown in Theorem 3.2. The numerical result is shown in Fig. 3(b). From Fig. 3(b), we can see that for fixed scale factor n, the sample entropy of $\mathbf{s}_{n,e}$, $e \in \mathbb{Z}_{2^n}$, is a constant with respect to e, and the values of $\mathbf{s}_{n,e}$ are decline as the scale factor n increases.

4. Analysis of the $1/f$ noise

The $1/f$ noise which can be observed in various physical systems is a signal whose power spectral density is proportional to the reciprocal of its frequency. We analyze in this section the $1/f$ noise by using the hierarchical entropy method.

We now describe the $1/f$ noise according to [12]. The $1/f$ noise is defined in terms of complex Gaussian random variables. As usual, we let $i = \sqrt{-1}$, the imaginary unit, and we use $\mathbb{C}$ to denote the complex plane. A complex variable $z := x + \mathrm{i}y$

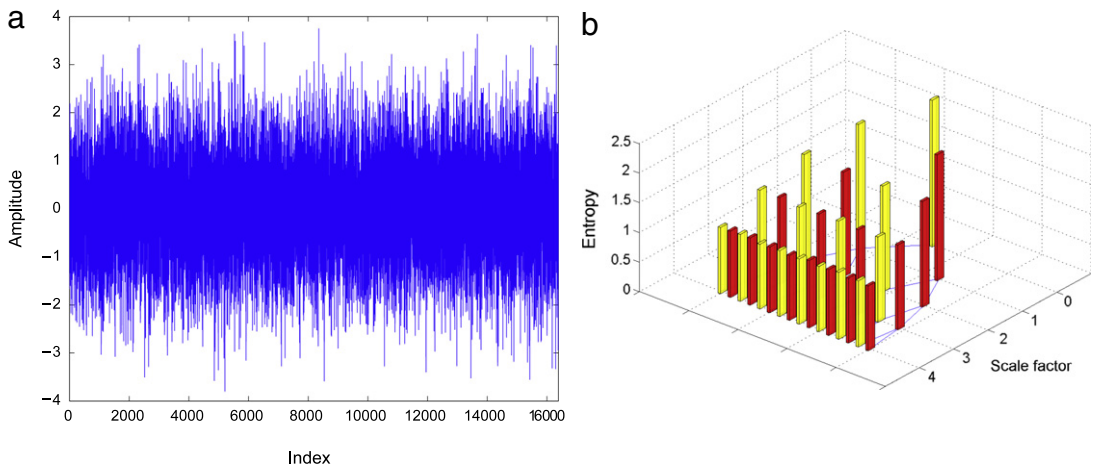

Fig. 3. HE analysis of Gaussian white noise.

is called a complex Gaussian random variable if both x and y are real independent Gaussian variables with the same mean 0 and standard deviation δ. The corresponding probability density function for the complex Gaussian random variable z is given by

$$\rho(z) := \frac{1}{\pi\,\delta_z^2} \mathrm{e}^{-|z|^2/\delta_z^2}, \quad z \in \mathbb{C},$$

where $\delta_z := \sqrt{2}\delta$. We need the notion of the discrete Fourier transform. For a given $n \in \mathbb{N}$, we let $\theta_n := \frac{2\pi}{2^n}$ and define the discrete Fourier transform F_n by

$$F_n := \frac{1}{2^n}\left[\mathrm{e}^{-\mathrm{i}\theta_n kl} : k \in \mathbb{Z}_{2^n}, l \in \mathbb{Z}_{2^n}\right]. \tag{4.1}$$

In particular, for a random vector $\mathbf{x}$ taking values in $\mathbb{R}^{2^N}$, we use $\hat{\mathbf{x}}$ to denote the discrete Fourier transform of $\mathbf{x}$, that is,

$$\hat{\mathbf{x}} := F_N \mathbf{x}.$$

We write $\hat{\mathbf{x}} := [z_k : k \in \mathbb{Z}_{2^N}]^T$. It can be verified by the definition of the discrete Fourier transform the following symmetric property

$$z_{2^{N-1}+k} = \overline{z}_{2^{N-1}-k}, \quad k \in \mathbb{Z}^+_{2^{N-1}}. \tag{4.2}$$

This symmetry property of the components of vector $\hat{\mathbf{x}}$ leads us to focusing only on its first $2^{N-1}+1$ components. If z_k, $k \in \mathbb{Z}^+_{2^{N-1}-1}$, are independent complex Gaussian random variables with mean 0, z_0 and $z_{2^{N-1}}$ are real Gaussian random variables with mean 0, and there is a positive constant c such that for all $k \in \mathbb{Z}_{2^{N-1}+1}$ the standard deviations δ_k of z_k satisfy $\delta_k \le \frac{c}{k+1}$, then we call $\mathbf{x}$ a $1/f$ random vector and call an instance of $\mathbf{x}$ a $1/f$ noise. We will use $\mathbf{f} := [f_k : k \in \mathbb{Z}_{2^N}]^T$ to denote a $1/f$ random vector and $\hat{\mathbf{f}} := [z_k : k \in \mathbb{Z}_{2^N}]^T$ to denote the discrete Fourier transform of $\mathbf{f}$.

The hierarchical entropy analysis of the $1/f$ noise requires understanding the statistical properties of $\mathbf{f}_{1,j} := Q_j(\mathbf{f})$, for $j \in \mathbb{Z}_2$. This leads us to investigate the composition

$$\widetilde{Q}_j := F_{N-1} Q_j F_N^{-1}, \quad j \in \mathbb{Z}_2, \tag{4.3}$$

where F_N^{-1} is the inverse discrete Fourier transform which takes the form

$$F_n^{-1} := \left[\mathrm{e}^{\mathrm{i}\theta_n kl} : k \in \mathbb{Z}_{2^n}, l \in \mathbb{Z}_{2^n}\right].$$

Note that for each $j \in \mathbb{Z}_2$ the matrix $\widetilde{Q}_j$ is a counter part of Q_j in the Fourier domain. In the next lemma, we identify them in terms of two diagonal matrices. For each $j \in \mathbb{Z}_2$, we introduce two sequences

$$d^+_{j,k} := 1 + (-1)^j \mathrm{e}^{\mathrm{i}\theta_N k}, \qquad d^-_{j,k} := 1 - (-1)^j \mathrm{e}^{\mathrm{i}\theta_N k}, \quad k \in \mathbb{Z}_{2^{N-1}}$$

and define two diagonal matrices accordingly

$$D^+_j := \mathrm{diag}[d^+_{j,k} : k \in \mathbb{Z}_{2^{N-1}}] \quad \text{and} \quad D^-_j := \mathrm{diag}[d^-_{j,k} : k \in \mathbb{Z}_{2^{N-1}}].$$

Lemma 4.1. *If $N \in \mathbb{N}$ is fixed, then for $j \in \mathbb{Z}_2$,*

$$\widetilde{Q}_j = \frac{1}{2}\left[D^+_j, D^-_j\right]. \tag{4.4}$$

Proof. For each $j \in \mathbb{Z}_2$, from the definition of F_N^{-1} and Q_j, a straightforward computation leads to

$$Q_j F_N^{-1} = \frac{1}{2}\left[e^{i\theta_{N-1}lk}(1 + (-1)^j e^{i\theta_N k}) : l \in \mathbb{Z}_{2^{N-1}}, k \in \mathbb{Z}_{2^N}\right]. \tag{4.5}$$

Let $p_j := \mathrm{diag}[D_j^+, D_j^-]$ and $\tilde{F}_N := \frac{1}{2}\left[F_{N-1}^{-1}, F_{N-1}^{-1}\right]$. Thus, from (4.5) we have that

$$Q_j^N F_N^{-1} = \tilde{F}_N p_j. \tag{4.6}$$

From (4.6) we have that

$$\widetilde{Q}_j = F_{N-1}\tilde{F}_N p_j = \frac{1}{2}[I_{N-1}, I_{N-1}]p_j, \tag{4.7}$$

where I_{N-1} denotes the identity matrix of order 2^{N-1}. Substituting the expression of p_j into (4.7) yields formula (4.4). □

For a real vector $\mathbf{x}$ of length 2^N, we write its discrete Fourier transform as $\hat{\mathbf{x}} := \tilde{\mathbf{x}} + i\tilde{\mathbf{y}}$, where $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{y}}$ are two real vectors of length 2^N. We define two operators T_l and T_r, respectively, which project a vector of length 2^N to a vector of length 2^{N-1} consisting of, respectively, the first and last 2^{N-1} components of the original vector, and let

$$\tilde{\mathbf{x}}_l := T_l\tilde{\mathbf{x}}, \qquad \tilde{\mathbf{x}}_r := T_r\tilde{\mathbf{x}}, \qquad \tilde{\mathbf{y}}_l := T_l\tilde{\mathbf{y}}, \qquad \tilde{\mathbf{y}}_r := T_r\tilde{\mathbf{y}}.$$

For $j \in \mathbb{Z}_2$, we let $\mathbf{x}_j := Q_j\mathbf{x}$ and we will express its discrete Fourier transform $\hat{\mathbf{x}}_j$ in terms of $\tilde{\mathbf{x}}_l$, $\tilde{\mathbf{x}}_r$, $\tilde{\mathbf{y}}_l$ and $\tilde{\mathbf{y}}_r$. To this end, we introduce three diagonal matrices

$$A_j^+ := \mathrm{diag}[1 + (-1)^j \cos(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}],$$
$$A_j^- := \mathrm{diag}[1 - (-1)^j \cos(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}],$$

and

$$B_j := \mathrm{diag}[(-1)^j \sin(\theta_N k) : k \in \mathbb{Z}_{2^{N-1}}].$$

Lemma 4.2. *If* $\mathbf{x}$ *is a real vector of length* 2^N*, then*

$$\hat{\mathbf{x}}_j = \frac{1}{2}\left(A_j^+\tilde{\mathbf{x}}_l + A_j^-\tilde{\mathbf{x}}_r - B_j\tilde{\mathbf{y}}_l + B_j\tilde{\mathbf{y}}_r\right) + \frac{i}{2}\left(A_j^+\tilde{\mathbf{y}}_l + A_j^-\tilde{\mathbf{y}}_r + B_j\tilde{\mathbf{x}}_l - B_j\tilde{\mathbf{x}}_r\right). \tag{4.8}$$

Proof. For each $j \in \mathbb{Z}_2$, from the definition of $\hat{\mathbf{x}}$, $\hat{\mathbf{x}}_j$ and $\mathbf{x}_j$, we have that

$$\hat{\mathbf{x}}_j = F_{N-1}Q_j\mathbf{x} = F_{N-1}Q_jF_N^{-1}\hat{\mathbf{x}} = \widetilde{Q}_j\hat{\mathbf{x}}. \tag{4.9}$$

In (4.9) we use Lemma 4.1 to conclude that

$$\hat{\mathbf{x}}_j = \frac{1}{2}\left(D_j^+, D_j^-\right)\hat{\mathbf{x}}. \tag{4.10}$$

According to the definition of D_j^+ and D_j^-, we know that

$$D_j^+ = A_j^+ + iB_j, \quad D_j^- = A_j^- - iB_j. \tag{4.11}$$

We partition the real and imaginary parts of the vector $\hat{\mathbf{x}}$ as

$$\hat{\mathbf{x}} = \begin{pmatrix}\tilde{\mathbf{x}}_l \\ \tilde{\mathbf{x}}_r\end{pmatrix} + i\begin{pmatrix}\tilde{\mathbf{y}}_l \\ \tilde{\mathbf{y}}_r\end{pmatrix}. \tag{4.12}$$

Substituting (4.11) and (4.12) into (4.10), we obtain (4.8). □

Lemma 4.3. *Suppose that* $\mathbf{x}$ *is a real random vector of length* 2^N*. If the first* $2^{N-1}+1$ *components of* $\hat{\mathbf{x}}$ *are independent, then the first* $2^{N-2}+1$ *components of* $\hat{\mathbf{x}}_j$ *are independent.*

Proof. We write $\hat{\mathbf{x}} := [z_k : k \in \mathbb{Z}_{2^N}]$ and $\hat{\mathbf{x}}_j := [z_{j,k} : k \in \mathbb{Z}_{2^{N-1}}]$. From (4.10) we have that for each $k \in \mathbb{Z}_{2^{N-2}+1}$, $z_{j,k}$ is a linear combination of z_k and $z_{k+2^{N-1}}$. Noting that $z_{k+2^{N-1}} = \bar{z}_{2^{N-1}-k}$, we know that $z_{j,k}$ is a linear combination of z_k and $\bar{z}_{2^{N-1}-k}$. Since for $k, k' \in \mathbb{Z}_{2^{N-2}+1}$, $\{k, 2^{N-1}-k\} \cap \{k', 2^{N-1}-k'\} = \emptyset$ when $k \neq k'$, and random variables z_k, $k \in \mathbb{Z}_{2^{N-1}+1}$, are independent, we conclude that random variables $z_{j,k}$, $k \in \mathbb{Z}_{2^{N-2}+1}$, are independent. □

A real random matrix is a matrix whose entries are random variables taking values in $\mathbb{R}$. The mathematical expectation of a random matrix is defined as a matrix of the expectation of its entries.

Lemma 4.4. *If z_1 and z_2 are two independent complex Gaussian random variables with mean 0, and standard derivations δ_1 and δ_2, respectively, then for each pair $\alpha := \alpha_r + \mathrm{i}\alpha_i$, $\beta := \beta_r + \mathrm{i}\beta_i$ with $\alpha_r, \alpha_i, \beta_r, \beta_i \in \mathbb{R}$, $\alpha z_1 + \beta z_2$ is a complex Gaussian random variable with mean 0 and standard derivation*

$$\delta := (\alpha_r^2\delta_1^2 + \alpha_i^2\delta_1^2 + \beta_r^2\delta_2^2 + \beta_i^2\delta_2^2)^{1/2}. \tag{4.13}$$

Proof. We let $z := \alpha z_1 + \beta z_2$ and prove that z is a complex Gaussian random variable. It suffices to show that the real part x and the imaginary part y of z are two independent real Gaussian random variables with the same real mean and standard derivation.

For $k \in \mathbb{Z}_2^+$, we let x_k and y_k be the real part and the imaginary part of z_k. From the definition of complex Gaussian random variables, we know that for all $k \in \mathbb{Z}_2^+$, x_k and y_k are two real Gaussian random variables with mean 0 and the same standard derivations $\delta_k/\sqrt{2}$. Since

$$x = \alpha_r x_1 - \alpha_i y_1 + \beta_r x_2 - \beta_i y_2, \qquad y = \alpha_i x_1 + \alpha_r y_1 + \beta_i x_2 + \beta_r y_2, \tag{4.14}$$

by Theorem 10.5 [11], we know that x and y are two real Gaussian random variables. Clearly, the mean of x and y is 0 and the standard derivation of x and y is

$$\frac{1}{\sqrt{2}}\left(\alpha_r^2\delta_1^2 + \alpha_i^2\delta_1^2 + \beta_r^2\delta_2^2 + \beta_i^2\delta_2^2\right)^{1/2}. \tag{4.15}$$

Thus, z has mean 0 and standard deviation δ given by (4.13)

It remains to show that x and y are independent. To this end, we will prove that the covariance between x and y is zero, since, by Theorem 4.1.2 in [13], this implies that x and y are independent. We calculate the covariance between x and y as follows. From (4.14), we have that $xy = \mathbf{v}^T D\mathbf{u}$, where $\mathbf{v} := [\alpha_r, -\alpha_i, \beta_r, -\beta_i]^T$, $\mathbf{u} := [\alpha_i, \alpha_r, \beta_i, \beta_r]^T$ and

$$D := \begin{bmatrix} x_1x_1 & x_1y_1 & x_1x_2 & x_1y_2 \\ y_1x_1 & y_1y_1 & y_1x_2 & y_1y_2 \\ x_2x_1 & x_2y_1 & x_2x_2 & x_2y_2 \\ y_2x_1 & y_2y_1 & y_2x_2 & y_2y_2 \end{bmatrix}.$$

Thus, we observe that

$$E(xy) = \mathbf{v}^T E(D)\mathbf{u}. \tag{4.16}$$

Since x_1, x_2, y_1, y_2 are independent and $E\left((x_k)^2\right) = E\left((y_k)^2\right) = \frac{\delta_k^2}{2}$, for $k \in \mathbb{Z}_2$, we have that $E(D) = \frac{1}{2}\mathrm{diag}[\delta_1^2, \delta_1^2, \delta_2^2, \delta_2^2]$. Substituting this equation into (4.16) yields $E(xy) = 0$. □

Lemma 4.5. *If the first $2^{N-1}+1$ components of $\hat{\mathbf{x}}$ are independent complex Gaussian random variables, then for each $j \in \{0, 1\}$, the first $2^{N-2}+1$ components of $\hat{\mathbf{x}}_j$ are independent complex Gaussian random variables, except for the first and the $(2^{N-2}+1)$th components which are real Gaussian random variables. Moreover, for each $k \in \mathbb{Z}_{2^{N-1}+1}$, if the mean and standard derivation of the $(k+1)$th component of $\hat{\mathbf{x}}$ are 0 and δ_k, respectively, then the mean of all components of $\hat{\mathbf{x}}_j$ is 0 and the standard deviation of the $(k+1)$th component of $\hat{\mathbf{x}}_j$ is*

$$\delta_{j,k} = \sqrt{2}\left[(1+\alpha_{j,k})\delta_k^2 + 2(1-\alpha_{j,k})\delta_{2^{N-1}-k}^2\right]^{1/2}, \tag{4.17}$$

where $\alpha_{j,k} := (-1)^j \cos(\theta_N k)$.

Proof. We write $\hat{\mathbf{x}}_j := \left[z_{j,k} : k \in \mathbb{Z}_{2^{N-2}+1}\right]$. We first prove that $z_{j,0}$ and $z_{j,2^{N-2}}$ are real Gaussian random variables. By Lemma 4.2, we have that

$$z_{j,0} = \frac{1}{2}\left(1 + (-1)^j\right)z_0 + \frac{1}{2}\left(1 - (-1)^j\right)z_{2^{N-1}}. \tag{4.18}$$

Since z_0 and $z_{2^{N-1}}$ are real Gaussian random variables, from (4.18) and Theorem 10.5 [11] we know that $z_{j,0}$ is a real Gaussian random variable. Furthermore, noting that $2^{N-1} - 2^{N-2} = 2^{N-2}$, again from Lemma 4.2 we have that

$$z_{j,2^{N-2}} = x_{2^{N-2}} - (-1)^j y_{2^{N-2}}, \tag{4.19}$$

where $x_{2^{N-2}}$ and $y_{2^{N-2}}$ are the real part and the image part of $z_{2^{N-2}}$, respectively. Hence, we conclude that $z_{j,2^{N-2}}$ is a real Gaussian random variable. Since from Lemma 4.2 we have that

$$z_{j,k} = \frac{1}{2}(1 + (-1)^j \mathrm{e}^{\mathrm{i}\theta_N k})z_k + \frac{1}{2}(1 - (-1)^j \mathrm{e}^{\mathrm{i}\theta_N k})z_{2^{N-1}+k}, \tag{4.20}$$

by Lemma 4.4 we obtain that $z_{j,k}$ is a complex Gaussian random variable.

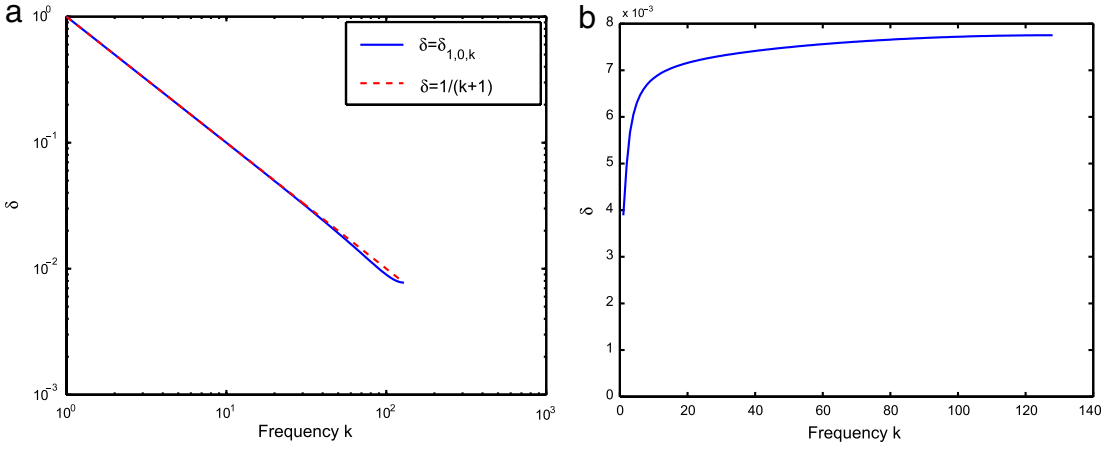

Fig. 4. Standard derivations of the components of random vectors $\hat{\mathbf{f}}_{1,0}$ and $\hat{\mathbf{f}}_{1,1}$.

Since the random variables $z_k, k \in \mathbb{Z}_{2^{N-1}+1}$, are independent, it follows from Lemma 4.3 that the random variables $z_{j,k}, k \in \mathbb{Z}_{2^{N-2}+1}$, are independent. By applying Lemma 4.4 to Eq. (4.20) with the symmetric property (4.2) and using Theorem 10.5 [11] to Eqs. (4.18) and (4.19), we obtain (4.17). □

We now return to the hierarchical entropy analysis of the $1/f$ noise $\mathbf{f}$. With the preparation presented above, we first prove a theoretical result which concerns $\mathbf{f}_{1,0}$.

Theorem 4.6. *If $\mathbf{f}$ is a $1/f$ random vector, and δ_k and $\delta_{0,k}$ are the standard derivations of the $(k+1)$-th random variable of $\hat{\mathbf{f}}$ and $\hat{\mathbf{f}}_{1,0}$, respectively, then $\mathbf{f}_{1,0}$ is also a $1/f$ random vector. Moreover, if there exists a positive constant c such that for all $k \in \mathbb{Z}_{2^{N-1}+1}$, $\delta_k^2 \le \frac{c}{1+k}$, then for all $k \in \mathbb{Z}_{2^{N-2}+1}$, $\delta_{0,k}^2 \le \frac{c}{1+k}$.*

Proof. Let $\hat{\mathbf{f}}_{1,0}$ be the Fourier transform of $\mathbf{f}_{1,0}$. To prove that $\mathbf{f}_{1,0}$ is a $1/f$ random vector, we will show that the first $2^{N-2}+1$ elements of $\hat{\mathbf{f}}_{1,0}$ are independent real or complex Gaussian random variables, and for all $k \in \mathbb{Z}_{2^{N-2}+1}$, $\delta_{0,k}^2 \le \frac{c}{1+k}$.

Since $\mathbf{f}$ be a $1/f$ random vector, from the definition of a $1/f$ random vector we know that the first $2^{N-1}+1$ elements of $\hat{\mathbf{f}}$ are independent real or complex Gaussian random variable with mean 0. Thus, from Lemma 4.5, we have that the first $2^{N-2}+1$ elements of $\hat{\mathbf{f}}_{1,0}$ are independent real or complex Gaussian random variables.

We next prove that for all $k \in \mathbb{Z}_{2^{N-2}+1}$, $\delta_{0,k}^2 \le \frac{c}{1+k}$. From (4.17), we have for all $k \in \mathbb{Z}_{2^{N-2}+1}$ that

$$\delta_{0,k}^2 = \frac{1}{2}\left((1+\cos(\theta_N k))\delta_k^2 + (1-\cos(\theta_N k))\delta_{2^{N-1}-k}^2\right). \tag{4.21}$$

Since for all $k \in \mathbb{Z}_{2^{N-1}+1}$, $\delta_k^2 \le \frac{c}{1+k}$, from (4.21) we have that

$$\delta_{0,k}^2 \le \frac{c}{1+k}\left(\frac{2^{N-2}+2^{N-2}\cos(\theta_N k) - k\cos(\theta_N k)+1}{1+2^{N-1}-k}\right). \tag{4.22}$$

Since $\cos(\theta_N k) \le 1$ and $k \in \mathbb{Z}_{2^{N-2}+1}$, it can be verified that the fractional expression in the parentheses of the right hand side of (4.22) is bounded below by $\frac{1}{2}$ and above by 1. This gives the desired result. □

In the remaining part of this section, we perform a numerical experiment which confirms the result regarding the lower frequency component $\mathbf{f}_{1,0}$, presented in Theorem 4.6 and provides us with additional insight into the higher frequency component $\mathbf{f}_{1,1}$.

Since $\mathbf{f}$ is a $1/f$ random vector, from Lemma 4.5 we know that for each $j \in \mathbb{Z}_2$, $\hat{\mathbf{f}}_{1,j}$ is a random vector of independent Gaussian random variables. We choose a $1/f$ random vector $\mathbf{f}$ so that the standard deviation δ_k of the components $z_k, k \in \mathbb{Z}_{2^{N-1}+1}$, of $\hat{\mathbf{f}}$ equals to $\frac{1}{k+1}$, and calculate the standard derivation $\delta_{1,j,k}$ of each component $z_{1,j,k}, k \in \mathbb{Z}_{2^{N-1}}$, of vector $\hat{\mathbf{f}}_{1,j}$ by using Eq. (4.17). Since $z_{1,j,k} = \bar{z}_{1,j,2^{N-1}-k}$ for $k \in \mathbb{Z}^+_{2^{N-2}}$, we only compute $\delta_{1,j,k}$ for $k \in \mathbb{Z}_{2^{N-2}+1}$. We show the values of $\delta_{1,j,k}, j \in \mathbb{Z}_2, k \in \mathbb{Z}_{2^{N-2}+1}$ in Fig. 4.

In Fig. 4(a), we plot the values of $\delta_{1,0,k}, k \in \mathbb{Z}_{2^{N-1}+1}$ by the solid line, and compare it with the values of $1/(k+1), k \in \mathbb{Z}_{2^{N-1}+1}$ which is shown by the dashed line. The numerical result shows that they are very close. In fact, Fig. 4(a) shows that the standard derivations of the elements of $\hat{\mathbf{f}}_{1,0}$ (the blue solid line) are proportional to the reciprocal of the frequency factor k (the red dashed line).

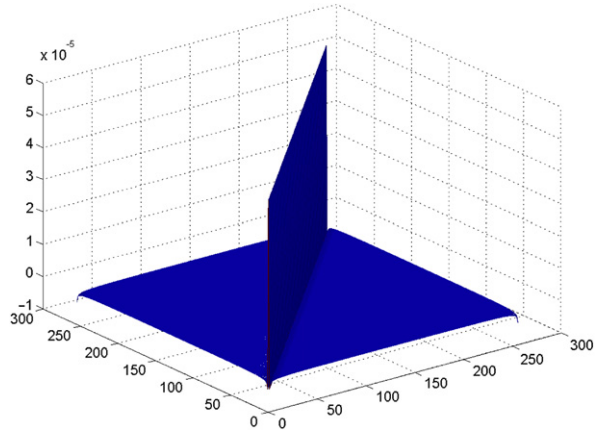

Fig. 5. Covariance matrix of $\mathbf{f}_{1,1}$.

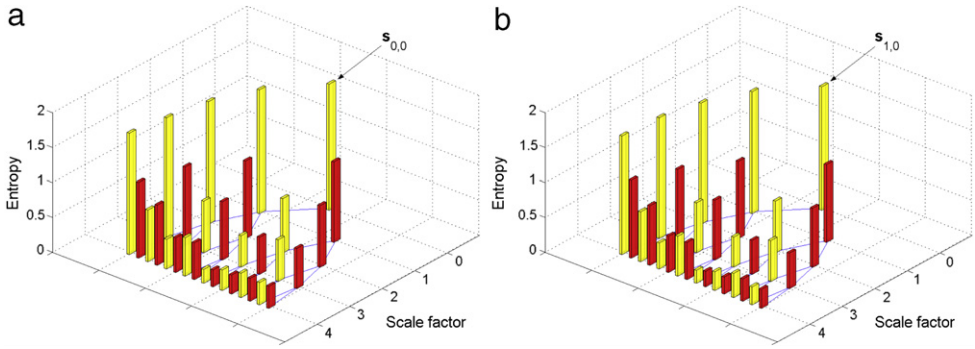

Fig. 6. HE analysis of $1/f$ noise $\mathbf{s}$ and $\mathbf{s}_{1,0}$.

In Fig. 4(b), we plot the values of $\delta_{1,1,k}$, $k \in \mathbb{Z}_{2^{N-1}+1}$ by the blue solid line, which shows that the standard derivations of the components of $\hat{\mathbf{f}}_{1,1}$ approximately equal to a constant as the frequency factor k increases. Noting that the components of $\hat{\mathbf{f}}_{1,1}$ are independent real or complex Gaussian random variables, the standard derivations of the components of $\hat{\mathbf{f}}_{1,1}$ with large frequencies are approximately equal to a constant, as shown in Fig. 4(b). This implies that $\mathbf{f}_{1,1}$ behaves pretty like a real Gaussian random vector. This can be further supported by the covariance matrix of $\mathbf{f}_{1,1}$ shown in Fig. 5. From Fig. 5 we see that the covariance matrix of $\mathbf{f}_{1,1}$ is very close to a diagonal matrix with constant diagonal entries. This means that the components of $\mathbf{f}_{1,1}$ are almost independent real Gaussian random variables with the same standard derivation.

We now consider the hierarchical entropy analysis of $1/f$ random vector $\mathbf{f}$. According to Theorem 4.6, we know that $\mathbf{f}_{1,0}$ is also a $1/f$ random vector. As shown in Fig. 4(a), the standard derivation of the components of $\hat{\mathbf{f}}_{1,0}$ is approximately equal to that of the components of $\hat{\mathbf{f}}$. Hence, we expect that *the result of the HE analysis of a* $1/f$ *random vector* $\mathbf{f}$ *equals to the result of HE analysis of* $\mathbf{f}_{1,0}$. We test this by applying the hierarchical entropy method to a $1/f$ noise $\mathbf{s}$ with 2^{15} data points and $\mathbf{s}_{1,0} := Q_0\mathbf{s}$. Here we use $\mathbf{s}$ to denote a $1/f$ noise, which is an instance of a $1/f$ random vector $\mathbf{f}$. The parameters used in calculating the hierarchical entropy are $m = 2$ and $r = 0.15$. The results are shown in Fig. 6. Fig. 6(a) shows the hierarchical entropy values of $\mathbf{s}_{0,0}$, while Fig. 6(b) shows those of $\mathbf{s}_{1,0}$. The difference of the entropy values of $\mathbf{s}_{0,0}$ and $\mathbf{s}_{1,0}$ can neglectable.

We next illustrate numerically that $\mathbf{f}_{1,1}$ is a random vector which is approximately equal to a real Gaussian random vector. We apply the hierarchical entropy method to $\mathbf{s}_{1,1} := Q_1\mathbf{s}$. The parameters used are $m = 2$ and $r = 0.15$. The results are shown in Fig. 7. From Fig. 7, we find that the entropies with the same scale factor is approximately equal to a constant, and the entropies decrease as the scale factor increases. This behavior of $\mathbf{s}_{1,1}$ is pretty much like that of the Gaussian white noise which is shown in Fig. 3(a).

5. Hierarchical entropy analysis of heartbeat data

In this section, we apply the hierarchical entropy method to analyze the cardiac interbeat interval time series derived from 24 h continuous electrocardiographic (ECG) Holter monitor recordings of 72 healthy subjects, 43 subjects with congestive heart failure (CHF) and 9 subjects with atrial fibrillation (AF). The parameters used in calculating the hierarchical entropy are $m = 2$ and $r = 0.15$. We demonstrate that the hierarchical entropy method extracts significant features of these time series which may help us distinguish the subjects.

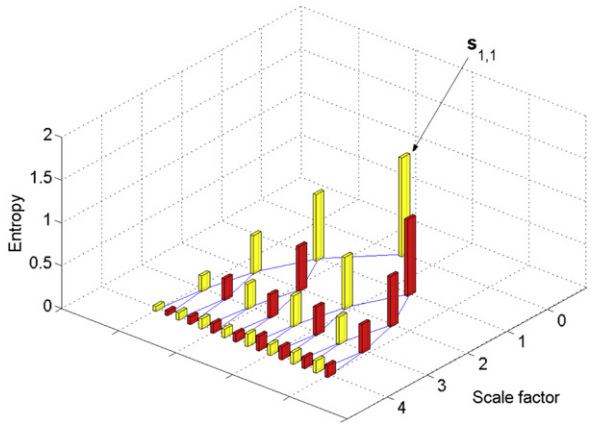

Fig. 7. HE analysis of $\mathbf{s}_{1,1}$.

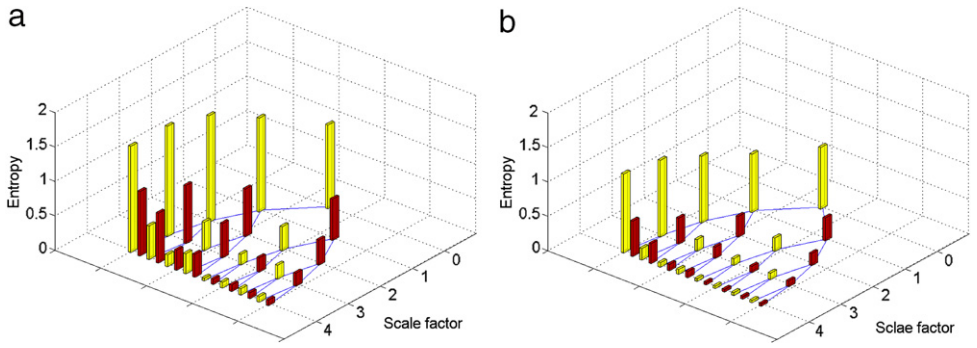

Fig. 8. HE analysis of heartbeat time series from 26 healthy young subjects and 46 healthy elderly subjects.

We calculate the hierarchical entropy of the subjects of each group. To present our results, we define necessary notations. We let $\{\mathbf{s}_l : l \in \mathbb{Z}_L^+\}$ denote a set of time series where L is a integer. For a given $K \in \mathbb{N}$, $\{\mathbf{s}_{l,n,e} : (n,e) \in \mathbb{J}_K\}$ denotes the hierarchical decomposition of signal $\mathbf{s}_l$ with K scales. Given the time series set $\{\mathbf{s}_l : l \in \mathbb{Z}_L^+\}$, for fixed $m \in \mathbb{N}, r \in \mathbb{R}$ we define the mean of the entropy of scale (n,e) by

$$M_{n,e} := \frac{1}{L}\sum_{l=1}^{L} S_m(\mathbf{s}_{l,n,e}, r).$$

We use a binary tree of column diagrams to represent the values of $M_{n,e}$. The high of a column at note (n,e) corresponds to the value of $M_{n,e}$. In this tree, $(0,0)$ denotes the root node, and for each node (n,e), we use $(n+1,2e)$ and $(n+1,2e+1)$ to denote its left and right child nodes, respectively.

Example 1 shown in Fig. 8 compares the hierarchical entropy analysis of the cardiac interbeat interval time series of the healthy young subjects with that of the healthy elderly subjects. We show the mean values $\{M_{n,e} : (n,e) \in \mathbb{J}_5\}$ of the 26 healthy young subjects in Fig. 8(a) and those of the 46 healthy elderly subjects in Fig. 8(b). The numerical results are reported in Table 1.

In both Fig. 8(a) and (b), the subtrees with the root node $(1,1)$ of the hierarchical entropy trees look pretty much like the hierarchical entropy tree of the Gaussian white noise. Hence, the first higher frequency component in both cases is most likely just noise. For this reason, we consider only the subtrees with root node $(1,0)$ in both Fig. 8(a) and (b). Comparing the values of the corresponding nodes in the two subtrees with root node $(1,0)$, we see that the values of nodes in the subtree in Fig. 8(a) are significantly larger than those of the corresponding nodes in the subtree in Fig. 8(b). Moreover, when we fix the scale factor $n = 3$ or 4, the values of the nodes in Fig. 8(b) decrease faster than the corresponding values in Fig. 8(a) as e increases. In other words, in the same scale, the higher frequency components of the time series derived from the young subjects are more irregular than the corresponding components of the time series derived from the elderly subjects.

In Example 2, we compare the hierarchical entropy analysis of the cardiac interbeat interval time series from the 72 healthy subjects and that from the 43 subjects with congestive heart failure (CHF). The numerical results are reported in Table 2 and shown in Fig. 9. The mean values $M_{n,e}$, $(n,e) \in \mathbb{J}_5$, of the hierarchical entropy values for the healthy subjects and for the CHF subjects are illustrated in Fig. 9(a) and (b), respectively.

Again, in both Fig. 9(a) and (b), the subtrees with root node $(1,1)$ of the hierarchical entropy trees looks pretty much like the hierarchical entropy tree of the Gaussian white noise and thus, as in the last example, the first higher frequency

Table 1
Hierarchical entropy values for the 26 healthy young subjects vs. those for the 46 healthy elderly subjects.

Scales	Healthy young subjects					Healthy elderly subjects				
	0	1	2	3	4	0	1	2	3	4
Entropy	1.21	1.36	1.55	1.60	1.54	0.89	0.85	0.97	1.11	1.15
					0.94					0.51
				0.84	0.49				0.38	0.17
					0.72					0.29
			0.69	0.42	0.17			0.30	0.17	0.07
					0.29					0.09
				0.50	0.29				0.17	0.07
					0.34					0.09
		0.60	0.35	0.16	0.04		0.33	0.20	0.10	0.04
					0.08					0.04
				0.21	0.08				0.09	0.04
					0.10					0.04
			0.37	0.20	0.09			0.18	0.10	0.04
					0.09					0.04
				0.19	0.08				0.09	0.03
					0.08					0.04

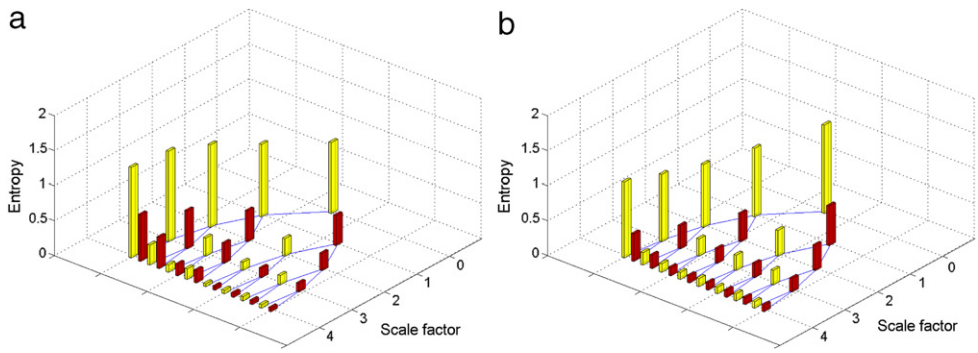

Fig. 9. Hierarchical entropy analysis of the heartbeat time series from the 72 healthy subjects and from the 43 subjects with congestive heart failure (CHF).

Table 2
Hierarchical entropy values of the heartbeat time series from the 72 healthy subjects vs. the 43 subjects with congestive heart failure (CHF).

Scales	Healthy subjects					CHF subjects				
	0	1	2	3	4	0	1	2	3	4
Entropy	1.01	1.03	1.18	1.29	1.29	1.26	0.98	0.90	0.96	1.08
					0.67					0.39
				0.55	0.29				0.33	0.18
					0.45					0.19
			0.44	0.26	0.10			0.40	0.25	0.12
					0.16					0.12
				0.29	0.15				0.21	0.10
					0.18					0.10
		0.43	0.25	0.12	0.04		0.57	0.37	0.22	0.11
					0.05					0.11
				0.14	0.05				0.20	0.10
					0.06					0.10
			0.25	0.13	0.06			0.36	0.21	0.10
					0.06					0.10
				0.12	0.05				0.19	0.09
					0.05					0.09

component in both cases is most likely just noise. In Fig. 9(a), the values $M_{n,1}$, $n \in \mathbb{Z}_4^+$ increase as the scale n increases. In Fig. 9(b), the values $M_{n,1}$, $n \in \mathbb{Z}_4^+$, have the minimum value at $n = 3$. Note that $M_{n,1}$ describes the heartbeat acceleration with the scale $n-1$. Thus, from Fig. 9(a), we know that the heartbeat accelerations of the healthy subjects becomes more irregular as the scale factor increasing. This indicates that the time series derived from the healthy subjects are more "complex" than that derived from the CHF subjects.

Example 3 is about the hierarchical entropy analysis of the cardiac interbeat interval time series from the 9 subjects with atrial fibrillation (AF). We report the numerical results in Table 3 and illustrate them in Fig. 10. For the comparison purpose,

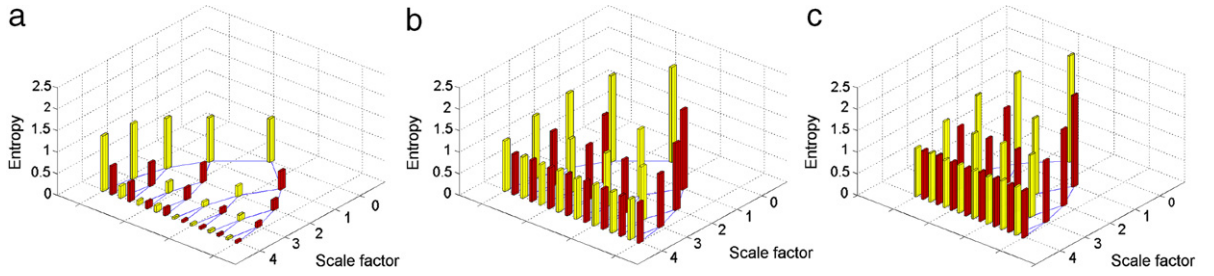

Fig. 10. Hierarchical entropy analysis of the heartbeat time series from the 72 healthy subjects and from the 9 subjects with atrial fibrillation (AF) and the Gaussian noise.

Table 3
Hierarchical entropy values of the 9 subjects with atrial fibrillation (AF) vs. the Gaussian white noise.

Scales	AF subjects					Gaussian white noise				
	0	1	2	3	4					
Entropy	2.21	2.01	1.76	1.47	1.18	2.47	2.12	1.81	1.46	1.12
					0.94					1.12
				1.27	0.96				1.46	1.13
					0.96					1.14
			1.58	1.26	0.94			1.76	1.43	1.10
					0.94					1.11
				1.27	0.96				1.44	1.11
					0.96					1.10
		1.86	1.57	1.25	0.93		2.12	1.79	1.46	1.10
					0.95					1.14
				1.26	0.95				1.44	1.11
					0.94					1.12
			1.56	1.26	0.95			1.78	1.45	1.13
					0.94					1.10
				1.25	0.95				1.44	1.12
					0.94					1.10

in Fig. 10(a), we review the results of the hierarchical entropy analysis of the time series from the 72 healthy subjects (shown in Fig. 9(a)). In Fig. 10(b), we present the mean values $M_{n,e}$, $(n, e) \in \mathbb{J}_5$, of the hierarchical entropy analysis of the AF subjects while in Fig. 10(c), we review the hierarchical entropy analysis of the Gaussian noise which is shown in Fig. 3(a).

Comparing Fig. 10(a) and (b), we see that the hierarchical entropy results of the time series from the subjects with AF are significantly different from those from the healthy subjects. In Fig. 10(b), we see that the values $M_{n,0}$, $n \in \mathbb{Z}_5$, decrease as the scale n increases, in the same manner as the hierarchical entropy values of the Gaussian white noise do (see, Fig. 10(c)).

It is difficult to distinguish the time series for AF subjects from the Gaussian white noise by only using the multiscale entropy, the values at nodes $(n, 0)$, $n \in \mathbb{Z}_5$, since the multiscale entropy for AF data behaves very similar to that for the Gaussian noise [2]. The hierarchical entropy analysis seems to have significant improvement over the multiscale entropy analysis. Observing the hierarchical entropy values shown in Fig. 10(b), we see that there is a notable feature that for each $n \in \mathbb{Z}_4^+$, the value $M_{n,0}$ is different from the values $M_{n,e}$, $e \in \mathbb{Z}_{2^n-1}^+$. On the other hand, from Fig. 10(c), we observe that the hierarchical entropy values of the Gaussian noise at the same scale are a constant. By using this feature, we can clearly distinguish the time series for the atrial fibrillation (AF) subjects from the Gaussian white noise.

6. Conclusion

The hierarchical entropy method introduces entropy of higher frequency components of a time series in addition to entropy of its lower frequency components provided by the multiscale entropy analysis for the series. Theoretical study and numerical experiments show that the HE methods provides useful information in quantifying the complexity of a time series.

Acknowledgments

The authors of this paper are grateful to Drs. Madalena Costa and Ary L. Goldberger of Harvard Medical School for useful discussion on the subject of multiscale entropy.

The third author was supported in part by Guangdong Provincial Government of China through the "Computational Science Innovative Research Team" program, by US Air Force Office of Scientific Research under grant FA9550-09-1-0511, by the US National Science Foundation under grant DMS-0712827, and by the Natural Science Foundation of China under grant 11071286.

References

[1] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (2002) 068102.
[2] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Phys. Rev. E 71 (2005) 021906.
[3] M. Costa, A.L. Goldberger, C.-K. Peng, Multiscale entropy to distinguish physiologic and synthetic RR time series, Comput. Cardiol. 29 (2002) 137–140.
[4] M. Costa, A.L. Goldberger, C.-K. Peng, Costa, Goldberger, and Peng reply, Phys. Rev. Lett. 92 (2004) 089804.
[5] M. Costa, C.-K. Peng, A.L. Goldberger, J.M. Hausdorff, Multiscale entropy analysis of human gait dynamics, Physica A 330 (2003) 53–60.
[6] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. 278 (2000) H2039.
[7] Y. Jiang, D. Mao, Y. Xu, A fast algorithm for computing sample entropy, Adv. Adapt. Data Anal. (2011) (in press).
[8] D.P. Feldman, J.P. Crutchfield, Statistical measures of complexity: why? Phys. Lett. A 238 (1998) 244–252.
[9] I. Daubechies, Ten Lectures on Wavelets, in: CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
[10] C.A. Micchelli, Y. Xu, Using the matrix refinement equation for the construction of wavelets on invariant sets, Appl. Comput. Harmon. Anal. 1 (1994) 391–401.
[11] P.L. Meyer, Introductory Probability and Statistical Applications, second ed., Addison-Wesley Publishing Co., Massachusetts, 1970.
[12] T. Antal, M. Droz, G. Gyorgyi, Z. Racz, Roughness distributions for $1/f^{\alpha}$ signals, Phys. Rev. E 65 (2002) 046140.
[13] P. Grassberger, Information and complexity measures in dynamical systems, in: H. Atmanspacher, H. Scheingraber (Eds.), Information Dynamics, Plenum Press, New York, 1991, pp. 15–33.