Published in final edited form as:

IEEE Trans Signal Process. 2010 April 1; 58(4): 2409–2414. doi:10.1109/TSP.2010.2040411.

On entropy rate for the complex domain

Wei Xiong [Student Member, IEEE][†], Tülay Adalı [Fellow, IEEE][†], Yi-Ou Li [Student Member, IEEE][†], Hualiang Li[†], and Vince D. Calhoun [Senior Member, IEEE][‡]

[†] University of Maryland Baltimore County, Baltimore, MD 21250, USA

[‡] The MIND Institute and University of New Mexico, Albuquerque, NM 87131, USA

Abstract

We derive the entropy rate formula for a complex Gaussian random process by using a widely linear model. The resulting expression is general and applicable to both circular and noncircular Gaussian processes, since any second-order stationary process can be modeled as the output of a widely linear system driven by a circular white noise. Furthermore, we demonstrate application of the derived formula to an order selection problem. We extend a scheme for independent and identically distributed (i.i.d.) sampling to the complex domain to improve the estimation performance of information-theoretic criteria when samples are correlated. We show the effectiveness of the approach for order selection for simulated and actual functional magnetic resonance imaging (fMRI) data that are inherently complex valued.

Index Terms

Entropy rate; order selection; complex-valued signal processing

I. Introduction

Entropy rate is a measure of the average information carried by each sample in a random sequence, and is a useful measure with applications in spectral estimation, coding, and biomedical signal processing among others [1], [2]. For example, it has been used to quantify heart variability for anomaly detection [3] and to determine independent samples in functional magnetic resonance data [4].

The entropy rate formula for a real-valued stationary Gaussian random process is developed by Kolmogorov [2], and a practical derivation using a linear model is given by Papoulis in [1]. As complex-valued signals arise typically in many signal processing problems such as communications and medical imaging, it is desirable to extend the formula to the complex domain. One such extension is possible assuming a Gaussian process with uncorrelated real and imaginary parts [5]. However, this is a limiting assumption as it ignores the possible correlation in the real and imaginary parts. As it has been noted recently, many applications such as communications and medical imaging involve noncircular signals, *i.e.* those with correlation in the real and imaginary parts of the signal [6]. Thus, the usefulness of a formula that ignores such correlation is, in general, limited.

In this paper, we first derive the general formula for the entropy rate of a general complex-valued Gaussian process by using a widely linear model [7]. Since any second-order stationary process can be modeled as the output of a widely linear system driven by a circular white noise, the resulting expression is general and takes the possible correlation in the real and imaginary parts into account. The expression reduces to the entropy rate of a real-valued Gaussian random process for signals with independent real and imaginary parts.

Next, we demonstrate the application of the complex entropy rate to an order selection problem and extend an independent sampling scheme developed for the real domain [4] to the complex domain. The scheme improves the performance of order selection with information-theoretic criteria when the samples are correlated. Entropy rate is used as the measure of independence such that the samples used in the order selection formulas are approximately independent.

Such an order selection scheme is useful in a number of scenarios one of which is the analysis of functional magnetic resonance imaging (fMRI) data. The fMRI data is acquired as complex-valued spatial-temporal data and order selection is a typical step used prior to the use of analysis techniques such as independent component analysis (ICA) and principal component analysis (PCA). FMRI voxels typically have spatial correlation, hence such a sampling scheme significantly improves the accuracy of order selection results for the analysis of fMRI data when used in its native complex-valued form.

In the next section, we introduce the basic statistical relationships for complex random process and derive the entropy rate formula for a complex Gaussian random process. In Section III, we present the application of entropy rate to i.i.d. sampling for order selection in the complex domain. We present experimental results of the order selection scheme using i.i.d. sampling, on simulated and actual fMRI data in Section IV, and provide a discussion of the results in the final section.

II. Entropy Rate of Complex Gaussian Random Process

A. Preliminaries

A complex random variable Z is written as $Z = Z_r + jZ_i$ where $j = \sqrt{-1}$ and Z_r, Z_i refer to the real and imaginary parts. The probability density function (pdf) of a complex random variable is defined as the joint pdf $p(\mathbf{Z}) \triangleq p(Z_r, Z_i)$. Given a complex random vector $\mathbf{Z} \in \mathbb{C}^N$ with the pdf $p(\mathbf{Z}) \triangleq p(\mathbf{Z}_r, \mathbf{Z}_i)$ where $\mathbf{Z}_r, \mathbf{Z}_i$ are the real and imaginary parts, the entropy is written as

$$H(\mathbf{Z}) \triangleq H(\mathbf{Z}_r, \mathbf{Z}_i) = -E\{\log p(\mathbf{Z}_r, \mathbf{Z}_i)\}.$$

A complex random process Z_k is written as a sequence of complex random variables indexed by k , with real and imaginary parts $Z_{r,k}$ and $Z_{i,k}$. Given a *second-order stationary* (SOS) and zero mean random process Z_k , the *covariance function* is defined by $R(m) = E\{Z_{k+m}Z_k^*\}$ and the *pseudo covariance function* [8], also called the *relation function* [9], as $\tilde{R}(m) = E\{Z_{k+m}Z_k\}$. Without loss of generality, the random processes and vectors discussed in this paper are assumed to be zero mean. A random process is called SOS if it is *wide sense stationary* and its pseudo covariance function only depends on the index difference. In the complex case, the covariance function is not sufficient to entirely describe the second-order statistics of a random process, hence both the covariance function and pseudo covariance function need to be taken into account in the analysis. The pseudo covariance function vanishes when the random process is *second-order circular*.

The Fourier transform of the covariance function yields the *power spectrum* (or *spectral density*) function $S(\omega)$. Similarly, we define the Fourier transform of the pseudo covariance function as the *pseudo power spectrum function* $\tilde{S}(\omega)$. The necessary conditions for a pair of functions $S(\omega)$ and $\tilde{S}(\omega)$ to be the power spectrum function and pseudo power spectrum function of a complex SOS random process are given by [9], [10]:

$$S(\omega) \geq 0, \quad \tilde{S}(\omega) = \tilde{S}(-\omega), \quad \text{and} \quad |\tilde{S}(\omega)|^2 \leq S(\omega)S(-\omega). \quad (1)$$

A stationary random process is said to be *white* if its covariance function satisfies $R(m) = R(0)\delta(m)$ where $\delta(\cdot)$ is the Kronecker-delta function, while there is no restriction on the pseudo covariance function. A *doubly white* [9] process has restrictions on both the covariance function and the pseudo covariance function, *i.e.*, one should have both $R(m) = R(0)\delta(m)$ and $\tilde{R}(m) = \tilde{R}(0)\delta(m)$ satisfied. The value of $\tilde{R}(0)$ is in general complex valued. When the real and imaginary parts are uncorrelated, however it assumes a real value. Furthermore, when the pseudo covariance function vanishes, *i.e.*, $\tilde{R}(m) = 0$, the process is a *circular white* process.

Entropy rate is a measure of average information in a random sequence, which can be written for a complex random process Z_k as

$$h_c(Z) = \lim_{n \rightarrow \infty} \frac{H(Z_1, Z_2, \dots, Z_n)}{n} \quad (2)$$

when the limit exists. As in the real case, $H(Z_1, Z_2, \dots, Z_n) \leq \sum_{k=1}^n H(Z_k)$, with equality if and only if the random variables Z_k are independent. Therefore the entropy rate can be used to measure the sample dependence and it reaches the upper bound when all samples of the process are independent.

B. Widely Linear Model

The widely linear filter is introduced in [7], and it improves minimum-mean square estimation performance with respect to the strictly linear filter that is generally used, when the commonly used circularity assumptions do not hold. Also, any SOS complex signal can be modeled as the output of a widely linear system driven by a circular white noise, which cannot be achieved by a strictly linear system [9]. Given the input and output vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^N$, a widely linear system is given by

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{G}\mathbf{x}^* \quad (3)$$

where $\mathbf{F}$ and $\mathbf{G}$ are complex-valued impulse responses in matrix form. The system function of a widely linear system is the pair of functions $[F(\omega), G(\omega)]$. The input-output relationship for the two power spectrum functions of a widely linear system can be written as

$$S_y(\omega) = |F(\omega)|^2 S_x(\omega) + |G(\omega)|^2 S_x^*(-\omega) + F(\omega)G^*(\omega)\tilde{S}_x(\omega) + F^*(\omega)G(\omega)\tilde{S}_x^*(-\omega) \quad (4)$$

$$\tilde{S}_y(\omega) = F(\omega)G(-\omega)S_x(\omega) + F(-\omega)G(\omega)S_x^*(-\omega) + F(\omega)F(-\omega)\tilde{S}_x(\omega) + G(\omega)G(-\omega)\tilde{S}_x^*(-\omega). \quad (5)$$

To write the joint entropy between the input and output of a widely linear system, we first write the relationship of input and output in $\mathbb{R}^{2N}$ using (3) as $\tilde{\mathbf{y}} = \mathbf{A}\tilde{\mathbf{x}}$ where $\tilde{\mathbf{y}} = [\mathbf{y}_r, \mathbf{y}_i]^T$, $\tilde{\mathbf{x}} = [\mathbf{x}_r, \mathbf{x}_i]^T$,

$$\mathbf{A} = \begin{bmatrix} \mathbf{F}_r + \mathbf{G}_r & \mathbf{G}_i - \mathbf{F}_i \\ \mathbf{F}_i + \mathbf{G}_i & \mathbf{F}_r - \mathbf{G}_r \end{bmatrix} \quad (6)$$

where the subscripts r and i refer to the real and imaginary parts respectively. The output entropy is written as

$$H(\mathbf{y}) = H(\mathbf{x}) + \log|\det(\mathbf{A})|. \quad (7)$$

since the output pdf can be expressed using Jacobian as $p(\mathbf{y}) = p(\mathbf{x})/|\det(\mathbf{A})|$.

C. Derivation of Entropy Rate for a Complex Gaussian Random Process

In this section, we study the entropy rate of a complex SOS Gaussian random process through a widely linear model using an approach similar to the one given in [1].

Proposition 1—The entropy rate of output $y(n)$ of a widely linear system $[F(\omega), G(\omega)]$, where $F(\omega)$ and $G(\omega)$ are minimum phase, is given by

$$h_c(Y) = h_c(X) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log\{[|F(e^{j\omega})|^2 - |G(e^{j\omega})|^2][|F(e^{-j\omega})|^2 - |G(e^{-j\omega})|^2]\} d\omega$$

where $h_c(X)$ is the entropy rate of input $x(n)$.

Proof—We assume that $x(n)$ is applied at $n = 0$, then the resulting system output is

$$y(n) = \sum_{k=0}^n x(n-k)f(k) + \sum_{k=0}^n x^*(n-k)g(k).$$

The widely linear transform for input $\mathbf{x} = [x(0), x(1), \dots, x(n)]^T$ and output $\mathbf{y} = [y(0), y(1), \dots, y(n)]^T$ is expressed as $\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{G}\mathbf{x}^*$, where

$$\mathbf{F} = \begin{bmatrix} f(0) & 0 & \cdots & 0 \\ f(1) & f(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ f(n) & f(n-1) & \cdots & f(0) \end{bmatrix}$$

and $\mathbf{G}$ is written similarly.

The entropy relation for input $\mathbf{x}$ and output $\mathbf{y}$ is given by (7) and the determinant of $\mathbf{A}$ given in (6) is

$$\det(\mathbf{A}) = [|f(0)|^2 - |g(0)|^2]^{n-1}. \quad (8)$$

From (2), (7) and (8), the relationship of entropy rate between the input and the output is given by

$$h_c(Y) = h_c(X) + \log[|f(0)|^2 - |g(0)|^2]. \quad (9)$$

Therefore, to complete the proof, we need to express (9) in terms of the integral of the system functions $F(\omega)$ and $G(\omega)$. Using $z = e^{j\omega}$, we obtain

$$j \int_{-\pi}^{\pi} \log[|F(e^{j\omega})|^2 - |G(e^{j\omega})|^2][|F(e^{-j\omega})|^2 - |G(e^{-j\omega})|^2] d\omega = 2 \oint \frac{1}{z} \log[|F(z)|^2 - |G(z)|^2] dz \quad (10)$$

where the path of the contour integral is the unit circle. Assuming $F(z)$ and $G(z)$ are minimum-phase systems, the circle of integration can be made arbitrarily large, thus yielding $F(z) = f(0)$, $G(z) = g(0)$ as $z \rightarrow \infty$. It follows that

$$\oint \frac{1}{z} \log[|F(z)|^2 - |G(z)|^2] dz = j2\pi \log[|f(0)|^2 - |g(0)|^2]. \quad (11)$$

Hence, (9), (10) and (11) lead to the conclusion.

The same result can be obtained using the entropy rate formula for a vector of real-valued random processes given in [11], which is subject to the lower semicontinuity condition defined in [11].

Theorem 1—If $z(n)$ is a complex SOS Gaussian random process with power spectrum function $S(\omega)$ and pseudo power spectrum function $\tilde{S}(\omega)$, its entropy rate h_c is given by

$$h_c = \log(\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \left[S(\omega)S(-\omega) - |\tilde{S}(\omega)|^2 \right] d\omega. \quad (12)$$

Proof—Since the real and imaginary parts of $z(n)$ are real-valued stationary Gaussian random processes, the joint entropy of $z(n)$ is given by

$$H(\mathbf{z}) = \frac{1}{2} \log[(2\pi e)^{2n} \det(\mathbf{K})] \quad (13)$$

where $\mathbf{K}$ is the covariance matrix of the real-valued random vector $\mathbf{z} = [z_{r,1}, z_{r,2}, \dots, z_{r,n}, z_{i,1}, z_{i,2}, \dots, z_{i,n}]^T$. The entropy rate of a complex Gaussian random process $z(n)$ is completely determined by its covariance function $R(m)$ and pseudo covariance function $\tilde{R}(m)$, or their Fourier transforms, *i.e.*, power spectrum function $S(\omega)$ and pseudo power spectrum function $\tilde{S}(\omega)$.

We can construct a complex Gaussian random sequence $y(n)$ as the output of a widely linear system $[F(\omega), G(\omega)]$. Let the input $x(n)$ be a circular white Gaussian random sequence with

$\mathbf{K} = (1/2)\mathbf{I}$, where $\mathbf{I}$ is the $2n \times 2n$ identity matrix. From (2) and (13), the entropy rate of input $x(n)$ can be written as

$$h(X) = \log(\pi e). \quad (14)$$

The power spectrum function and pseudo power spectrum function of $x(n)$ are given as $S_x(\omega) = 1$ and $\tilde{S}_x(\omega) = 0$. From (4) and (5), we obtain the power spectrum function and pseudo power spectrum function of output $y(n)$, $S_y(\omega) = |F(\omega)|^2 + |G(\omega)|^2$ and $\tilde{S}_y(\omega) = F(\omega)G(-\omega) + F(-\omega)G(\omega)$. Hence, if we choose the widely linear system as

$$|F(\omega)|^2 + |G(\omega)|^2 = S(\omega) \quad (15)$$

$$F(\omega)G(-\omega) = F(-\omega)G(\omega) = \frac{1}{2} \tilde{S}(\omega), \quad (16)$$

the output sequence $y(n)$ has the same power spectrum function and pseudo power spectrum function as $z(n)$, which indicates that the entropy rate of $z(n)$ is equal to that of $y(n)$. In [9], it is shown that widely linear system defined by $F(\omega)$ and $G(\omega)$, which satisfies the equalities (15) and (16), exists and fulfills the necessary conditions (1) for $S(\omega)$, $\tilde{S}(\omega)$. Furthermore, using (15) and (16), we obtain

$$[|F(\omega)|^2 - |G(\omega)|^2][|F(-\omega)|^2 - |G(-\omega)|^2] = S(\omega)S(-\omega) - |\tilde{S}(\omega)|^2. \quad (17)$$

Therefore, the conclusion is followed by (14), (17) and *Proposition 1*.

For a second-order circular process, we have $\tilde{S}(\omega) = 0$, thus yielding the entropy rate of a second-order circular Gaussian random process as

$$h_{\text{circ}} = \log(\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log[S(\omega)S(-\omega)] d\omega. \quad (18)$$

For the general case given in (12), $|\tilde{S}(\omega)|^2 \geq 0$. Hence, for the second-order circular and noncircular Gaussian random sequences with the same covariance function $R(m)$, we have

$$h_{\text{noncirc}} \leq h_{\text{circ}},$$

which can be also verified using the result for complex entropy [8], [12] and the definition of entropy rate given in (2).

Next, we show that for a doubly white Gaussian random process with $\tilde{R}(0) \in \mathbb{R}$, the complex entropy rate given in (12) is equal to twice the entropy rate of a real-valued Gaussian random process. For a doubly white Gaussian random process with $\tilde{R}(0) \in \mathbb{R}$, the real and imaginary parts are uncorrelated and white. Since the real and imaginary parts are decoupled in this case,

we can directly use the entropy rate formula for real-valued Gaussian process respectively, and the entropy rate of the complex Gaussian process is

$$h_c = h_r + h_i \quad (19)$$

where the subscripts r and i refer to the functions of the real and imaginary parts.

Since the real and imaginary parts are white, their power spectrum functions are given as $S_r(\omega) = R_r(0)$ and $S_i(\omega) = R_i(0)$. Furthermore, we have the conditions $R(0) = R_r(0) + R_i(0)$ and $R(0) = R_r(0) - R_i(0)$, since the real and imaginary parts are uncorrelated. A doubly white Gaussian process with $\tilde{R}(0) \in \mathbb{R}$ satisfies that $S(\omega) = S(-\omega)$, and $\tilde{S}(\omega)$ is real valued. Thus, we obtain

$$S(\omega)S(-\omega) - |\tilde{S}(\omega)|^2 = 4S_r(\omega)S_i(\omega). \quad (20)$$

By incorporating (20) into (12), we obtain (19), where

$$h_r = \frac{1}{2} \log(2\pi e) + \frac{1}{4\pi} \int_{-\pi}^{\pi} \log S_r(\omega) d\omega$$

and h_i is written similarly [2]. Hence, the entropy rate calculated from formulas in the real and the complex domain are same for a doubly white Gaussian process with $R(0) \in \mathbb{R}$.

For a circular white Gaussian random process, we have $\tilde{R}(m) = 0$, and thus the entropy rate can be simply calculated using the entropy rate formula in the real domain, leading to the result given in (18).

III. Application to IID Sampling for Order Selection

A. Information-theoretic Criteria

Information-theoretic criteria are commonly used for order selection in many signal processing problems. There are a number of information-theoretic criteria commonly used for order selection, such as, Akaike's information criterion (AIC) [13], Kullback-Leibler information criterion (KIC) [14], and the minimum description length (MDL) criterion [15] or the Bayesian information criterion [16]. The formulas for AIC, KIC and MDL criteria assume similar forms such that the last term in

$$\mathcal{E}(k) = -2\mathcal{L}(\mathbf{x}|\Theta_k) + \mathcal{P}(\Theta_k) \quad (21)$$

is given by $2\mathcal{G}(\Theta_k)$ for AIC, $3\mathcal{G}(\Theta_k)$ for KIC, and by $\mathcal{G}(\Theta_k)(\log N)/2$ for the MDL criterion. In (21), where $\mathcal{L}(\mathbf{x}|\Theta_k)$ is the maximum log-likelihood of i.i.d. observations $\mathbf{x}$ based on the model parameter set Θ_k , and $\mathcal{G}(\Theta_k)$ is the penalty for model complexity given by the total number of free parameters in Θ_k . For MDL, the penalty term is scaled by $\log N$ where N is the sample size.

In [17], Wax and Kailath provide a practical form of the maximum log-likelihood for complex-valued data, using multivariate Gaussian model, as

$$\mathcal{L}(\mathbf{x}|\mathbf{\Theta}_k) = N \log \left(\frac{\prod_{i=k+1}^T \lambda_i^{\frac{1}{T-k}}}{\frac{1}{T-k} \sum_{i=k+1}^T \lambda_i} \right)^{T-k} \quad (22)$$

where T is the original dimension of the multivariate data, k is the candidate order, N is the sample size, and λ_i 's are the eigenvalues of the sample covariance matrix of the multivariate observations. The number of free parameters for complex-valued data is given by $\mathcal{G}(\mathbf{\Theta}_k) = 1 + 2Tk - k^2$.

Order selection formulations are based on the assumption of i.i.d. samples. When dependent samples are used, the actual number of i.i.d. samples is less than N , and the likelihood term given by (22) improperly dominates the information-theoretic criteria, resulting in an over-estimation of the order. In the experiments section, we show the over-estimation of order number on simulated fMRI data.

B. IID sampling in the complex domain

Commonly there is sample dependence among the samples, which violates the i.i.d. assumption of information-theoretic criteria. An example is in fMRI data where samples are not i.i.d. due to the point spread function of the scanner as well as the use of spatial smoothing as a preprocessing step. However, the dependence among the data is typically localized, *i.e.*, lies within few adjacent samples. Hence, an i.i.d. sampling scheme can be used to identify an effectively i.i.d. sample set in the complex domain, as shown in [4] for real-valued data.

We model the data as a complex-valued finite-order moving average (MA) process, *i.e.*, an SOS Gaussian random process, which is the output of a linear system with an i.i.d. complex Gaussian input. The second-order statistics of the finite-order MA sequence $z(n)$ has finite length, *i.e.*, $R(m) = E\{z(n+m)z^*(n)\} = 0$ and $\hat{R}(m) = E\{z(n+m)z(n)\} = 0$ for $|m| \geq L$, where L is a small positive integer. The subsampled sequence $z_s(n) = z(Ln)$ is a doubly white Gaussian random sequence, *i.e.*, after normalization, $R_s(m) = \delta(m)$ and $\hat{R}_s(m) = c\delta(m)$, where $c \in \mathbb{C}$, $|c| \leq 1$. Here c is a measure for the degree of noncircularity and $c = 0$ in circular case.

Entropy rate can be used to measure the sample dependence, and it reaches the upper bound when the samples are i.i.d. From (12), the theoretical upperbound value for the entropy rate to the normalized complex Gaussian random sequence $z(n)$, *i.e.*, $R(0) = 1$ and $\hat{R}(0) = c$, where $c \in \mathbb{C}$, $|c| \leq 1$, is given by

$$h_c(Z) \leq \log(\pi e) + \frac{1}{2} \log(1 - |c|^2).$$

The grid of locations on which the data samples are considered to be effectively independent is determined when the entropy rate reaches its upperbound. Therefore, an effective i.i.d. sample set is obtained on this grid of spatial locations at which the dependence among the samples is small enough to be ignored. Since the subsampling procedure decreases the number of samples for estimation, an eigenspectrum adjustment scheme [18] is used to mitigate the finite sample effect.

IV. Experiments

A. Simulated data

We generate eight complex-valued spatial maps to simulate the fMRI sources and corresponding time courses, the magnitudes of which are similar to the ones used in [19]. In an fMRI experiment, the phase difference induced by the task activation is typically less than $\pi/9$ [20], [21]. Therefore, we keep the phase of each pixel uniformly distributed in the range $[-\pi/18, \pi/18]$. The phase of each complex-valued time point is generated proportional to its magnitude, but is again restricted to a small range [22], which in our case is $[-\pi/18, \pi/18]$. The spatial sources are rearranged into one-dimensional vectors and mixed by the corresponding time courses as the spatial ICA model for fMRI data [23]:

$$\mathbf{X} = \sum_{k=1}^M \mathbf{a}_k \mathbf{s}_k^T + \mathbf{V}.$$

Here, $\mathbf{s}_k, \mathbf{a}_k \in \mathbb{C}^N$, represent the activation intensity of each pixel, for the k th spatial map and the k th corresponding time course, $M = 8$ is the number of informative spatial map sources, $N = 3600$ is the number of pixels in each spatial map source, $T = 100$ is the number of time points in the time course, and $\mathbf{V}$ is the $T \times N$ matrix of Gaussian noise.

Complex-valued Gaussian noise is added to the data set with a specified contrast to noise ratio (CNR), calculated as the ratio of the standard deviation of the mixed data set without noise to the standard deviation of the Gaussian noise. The mixture data are spatially smoothed, separately for the real and imaginary parts, by a Gaussian kernel with a full-width at half maximum (FWHM) of 2 pixels.

The mixtures of eight sources, with noise levels of $\text{CNR} = -3, 0, 3$ and 6 dB, are created and the complex-valued order selection with i.i.d. sampling scheme is applied to these mixtures. To study the effect of sample dependence on the estimated order, we also apply order selection without subsampling to simulated data with $\text{CNR} = 3$ dB. The criteria used in the experiment are AIC, KIC and MDL. Fig. 1 shows the results of 10 Monte Carlo simulations where a different noise realization is used for each run. The standard deviation is stacked on the mean value in each bar plot.

Without subsampling, the order number is significantly over-estimated as observed in Fig. 1 (a), since the samples become correlated after spatial smoothing. As shown in Fig. 1(b), the criteria based on the effective i.i.d. samples yield accurate estimates (8 sources) when the CNR is higher than 0 dB. CNR of actual fMRI data is typically in the range $[0, 3]$ dB, and the complex-valued order selection scheme is thus effective in this CNR range.

B. FMRI data

FMRI data from ten subjects performing a finger-tapping task are used in this study. Details of the data are given in [24]. The real and imaginary images are spatially smoothed using a $10 \times 10 \times 10$ mm full-width half-maximum Gaussian kernel. To study the effect of sample dependence on order selection, the data without spatial smoothing is used in the experiments as the “unsmoothed” fMRI data set in contrast to the “smoothed” fMRI data which are preprocessed using the Gaussian kernel.

To show the effect of i.i.d. sampling, we plot the entropy rate of the subsampled data with different subsampling depths on 10 sets of unsmoothed and smoothed fMRI data sets in Fig. 2. It can be observed that as the subsampling depth increases, the entropy rate increases and

converges to its upperbound. Thus the subsampling depth is determined when all the samples are independent. Since the smoothed data have greater sample dependence compared to the unsmoothed data, greater subsampling depth is required to achieve convergence for the smoother data.

The results of order selection using i.i.d. sampling is shown in [5] using actual fMRI data. It is shown that the subsampling scheme effectively addresses the over-estimation problem and the range of values indicated by KIC and MDL provide a good regime for the order numbers to be used in complex ICA of fMRI data.

V. Discussion

In this paper, we derive the entropy rate for complex SOS random processes using a widely linear model. The resulting expression is general such that it is applicable to both circular and noncircular Gaussian processes, and can be employed in various applications. For example, the traditional spectral estimation problem in the real domain focuses on estimating the spectral density function, and one approach is through the maximization of the entropy rate [1]. When extending this approach to the complex domain, as we have demonstrated, both the spectral density function and the *pseudo* spectral density functions need to be estimated for the entropy rate when the process is noncircular. In other application domains such as biomedical image analysis [23] and communications that generate complex-valued data the given entropy rate can be employed as an effective measure.

In summary, we introduce an application of complex entropy rate in biomedical signal processing. By modeling the complex data as a finite-order MA process, we develop an i.i.d. sampling scheme in the complex domain to improve the performance of order selection with information-theoretic criteria when the samples are correlated. Entropy rate is used to measure the sample dependence in this order selection scheme. The order selection scheme with i.i.d. sampling is implemented for analysis of complex fMRI data, which is useful for data-driven analysis approaches such as PCA and ICA. The method can be further utilized in other signal processing fields, where a lower dimensional informative subspace needs to be identified.

Acknowledgments

This work is supported by the NSF grants NSF-IIS 0612076 and NSF-CCF 0635129.

References

1. Papoulis A. Maximum entropy and spectral estimation: a review. *IEEE Trans Acoust Speech, Signal Process* 1982;29:1176–1186.
2. Cover, TM.; Thomas, JA. *Elements of Information Theory*. Wiley; 1991.
3. Porta A, Guzzetti S, Montano N, Furlan R, Pagani M, Malliani A, Cerutti S. Entropy, entropy rate and pattern classification as tools to typify complexity in short heart period variability series. *IEEE Trans Biomed Eng* 2001;48:1282–1291. [PubMed: 11686627]
4. Li YO, Adalı T, Calhoun VD. Estimating the number of independent components for fMRI data. *Hum Brain Mapp* 2007;28:1251–1266. [PubMed: 17274023]
5. Xiong, W.; Li, Y.-O.; Li, H.; Adalı, T.; Calhoun, VD. On ICA of complex-valued fMRI: advantages and order selection. *Proc. ICASSP; Las Vegas, NV. 2008. p. 529-532.*
6. Adalı T, Li H, Novey M, Cardoso JF. Complex ICA using nonlinear functions. *IEEE Trans Signal Process* 2008;56:4536–4544.
7. Picinbono B, Chevalier P. Widely linear systems for estimation. *IEEE Trans Signal Process* 1995;43:2030–2033.
8. Neeser FD, Massey JL. Proper complex random processes with applications to information theory. *IEEE Trans Inf Theory* 1993;39:1293–1302.

9. Picinbono B, Bondon P. Second-order statistics of complex signals. *IEEE Trans Signal Process* 1997;45:411–420.
10. Schreier PJ, Scharf LL. Second-order analysis of improper complex random vectors and processes. *IEEE Trans Signal Process* 2003;51:714–725.
11. Pham DT. Mutual information approach to blind separation of stationary sources. *IEEE Trans Inf Theory* 2002;48:1935–1946.
12. Schreier PJ. Bounds on the degree of impropriety of complex random vectors. *IEEE Signal Process Lett* 2008;15:190–193.
13. Akaike H. A new look at statistical model identification. *IEEE Trans Autom Control* 1974;19:716–723.
14. Cavanaugh JE. A large-sample model selection criterion based on Kullback's symmetric divergence. *Stat Probab Lett* 1999;44:333–344.
15. Rissanen J. Modeling by the shortest data description. *Automatica* 1978;14:465–471.
16. Schwartz G. Estimating the dimension of a model. *Ann Stat* 1978;6:461–464.
17. Wax M, Kailath T. Detection of signals by information theoretic criteria. *IEEE Trans Acoust, Speech, Signal Process* 1985;33:387–392.
18. Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans Med Imag* 2004;23:137–152.
19. Correa, N.; Li, Y-O.; Adalı, T.; Calhoun, VD. Comparison of blind source separation algorithms for fMRI using a new Matlab toolbox: GIFT. *Proc. ICASSP*; Philadelphia, PA. 2005. p. 401-404.
20. Hoogenrad FG, Reichenbach JR, Haacke EM, Lai S, Kuppusamy K, Sprenger M. In vivo measurement of changes in venous blood-oxygenation with high resolution functional MRI at .95 Tesla by measuring changes in susceptibility and velocity. *Magn Reson Med* 1998;39:97–107. [PubMed: 9438443]
21. Calhoun VD, Adalı T, van Zijl PCM, Pekar JJ. Independent component analysis of fMRI data in the complex domain. *Magn Reson Med* 2002;48:180–192. [PubMed: 12111945]
22. Rowe DB. Modeling both the magnitude and phase of complex-valued fMRI data. *Neuroimage* 2005;25:1310–1324. [PubMed: 15850748]
23. Adalı T, Calhoun VD. Complex ICA of brain imaging data. *IEEE Signal Processing Mag* 2007;24:136–139.
24. Feng, Z.; Caprihan, A.; Blagoev, K.; Zhao, F.; Calhoun, VD. Modeling of phase changes in BOLD fMRI. *Proc. ISMRM*; Toronto, Canada. 2008.

Fig. 1. Order selection, (a) with and without i.i.d. sampling on simulated data (CNR = 3 dB), (b) on simulated data with different CNR values: -3, 0, 3 and 6 dB, using effectively i.i.d. samples

Fig. 2.

The entropy rate of the subsampled data with different subsampling depths, (a) on unsmoothed fMRI data sets, (b) on smoothed fMRI data sets