On the Information Dimension Rate of Multivariate Gaussian Processes

Bernhard C. Geiger*, Tobias Koch†‡
*Signal Processing and Speech Communication Laboratory, Graz University of Technology, Graz, Austria
†Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911, Leganés, Spain
‡Gregorio Marañón Health Research Institute, 28007, Madrid, Spain.
Emails: geiger@ieee.org, koch@tsc.uc3m.es

Abstract—**The authors have recently defined the Rényi information dimension rate $d(\{X_t\})$ of a stationary stochastic process $\{X_t,\, t \in \mathbb{Z}\}$ as the entropy rate of the uniformly-quantized process divided by minus the logarithm of the quantizer step size $1/m$ in the limit as $m \to \infty$ (B. Geiger and T. Koch, "On the information dimension rate of stochastic processes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, June 2017). For Gaussian processes with a given spectral distribution function F_X, they showed that the information dimension rate is given by the Lebesgue measure of the set of harmonics where the derivative of F_X is positive. This paper extends this result to multivariate Gaussian processes with a given matrix-valued spectral distribution function $F_{\mathbf{X}}$. It is demonstrated that the information dimension rate equals the average rank of the derivative of $F_{\mathbf{X}}$. As side results, it is shown that the scale and translation invariance of information dimension carries over from random variables to stochastic processes.**

I. Introduction

In 1959, Rényi [1] proposed the information dimension and the d-dimensional entropy to measure the information content of general random variables (RVs). In recent years, it was shown that the information dimension is of relevance in various areas of information theory, including rate-distortion theory, almost lossless analog compression, or the analysis of interference channels. For example, Kawabata and Dembo [2] showed that the information dimension of a RV is equal to its rate-distortion dimension, defined as twice the rate-distortion function $R(D)$ divided by $-\log(D)$ in the limit as $D \downarrow 0$. Koch [3] demonstrated that the rate-distortion function of a source with infinite information dimension is infinite, and that for any source with finite information dimension and finite differential entropy the Shannon lower bound on the rate-distortion function is asymptotically tight. Wu and Verdú [4] analyzed both linear encoding and Lipschitz decoding of discrete-time, independent and identically distributed (i.i.d.), stochastic processes and showed that the information dimension plays a fundamental role in achievability and converse results. Wu *et al.* [5] showed that the degrees of freedom of the K-user Gaussian interference channel can be characterized through the sum of information dimensions. Stotz and Bölcskei [6] later generalized this result to vector interference channels.

In [7], [8], we proposed the information dimension rate as a generalization of information dimension from RVs to univariate (real-valued) stochastic processes. Specifically, consider the stationary process $\{X_t, t \in \mathbb{Z}\}$, and let $\{[X_t]_m,\, t \in \mathbb{Z}\}$ be the process obtained by uniformly quantizing $\{X_t\}$ with step size $1/m$. We defined the information dimension rate $d(\{X_t\})$ of $\{X_t\}$ as the entropy rate of $\{[X_t]_m\}$ divided by $\log m$ in the limit as $m \to \infty$ [8, Def. 2]. We then showed that, for any stochastic process, $d(\{X_t\})$ coincides with the rate-distortion dimension of $\{X_t\}$ [8, Th. 5]. We further showed that for stationary Gaussian processes with spectral distribution function F_X, the information dimension rate $d(\{X_t\})$ equals the Lebesgue measure of the set of harmonics on $[-1/2, 1/2]$ where the derivative of F_X is positive [8, Th. 7]. This implies an intuitively appealing connection between the information dimension rate of a stochastic process and its bandwidth.

In this work, we generalize our definition of $d(\{X_t\})$ to multivariate processes. Consider the L-variate (real-valued) stationary process $\{\mathbf{X}_t\}$, and let $\{[\mathbf{X}_t]_m\}$ be the process obtained by quantizing every component process of $\{\mathbf{X}_t\}$ uniformly with step size $1/m$. As in the univariate case, the information dimension rate $d(\{\mathbf{X}_t\})$ of $\{\mathbf{X}_t\}$ is defined as the entropy rate of $\{[\mathbf{X}_t]_m\}$ divided by $\log m$ in the limit as $m \to \infty$. Our main result is an evaluation of $d(\{\mathbf{X}_t\})$ for L-variate Gaussian processes with spectral distribution matrix $F_{\mathbf{X}}$. We demonstrate that for such processes $d(\{\mathbf{X}_t\})$ equals the Lebesgue integral of the rank of the derivative of $F_{\mathbf{X}}$. As a corollary, we show that the information dimension rate of univariate complex-valued Gaussian processes is maximized if the process is proper, in which case it is equal to twice the Lebesue measure of the set of harmonics where the derivative of its spectral distribution function F_X is positive.

As side results, we show that $d(\{\mathbf{X}_t\})$ is scale and translation invariant. These properties are known for the information dimension of RVs (cf. [9, Lemma 3]), but they do not directly carry over to our definition of $d(\{\mathbf{X}_t\})$, which is why we state them explicitly in this paper.

Due to space limitations, some of the proofs are only sketched or omitted altogether. The full proofs appear in [10].

The work of Bernhard C. Geiger has been funded by the Erwin Schrödinger Fellowship J 3765 of the Austrian Science Fund. The work of Tobias Koch has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 714161), from the Spanish Ministerio de Economía y Competitividad under Grants TEC2013-41718-R, RYC-2014-16332 and TEC2016-78434-C3-3-R (AEI/FEDER, EU), and from the Comunidad de Madrid under Grant S2103/ICE-2845.

II. Notation and Preliminaries

We denote by $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{Z}$ the set of real numbers, the set of complex numbers, and the set of integers, respectively. We use a calligraphic font, such as $\mathcal{F}$, to denote other sets, and we denote complements as $\mathcal{F}^{\mathsf{c}}$.

We denote RVs by upper case letters, e.g., X. For a finite or countably infinite collection of RVs we abbreviate $X_\ell^k \triangleq (X_\ell, \ldots, X_{k-1}, X_k)$, $X_\ell^\infty \triangleq (X_\ell, X_{\ell+1}, \ldots)$, and $X_{-\infty}^k \triangleq (\ldots, X_{k-1}, X_k)$. Univariate discete-time stochastic processes are denoted as $\{X_t, t \in \mathbb{Z}\}$ or, in short, as $\{X_t\}$. For L-variate stochastic processes we use the same notation but with X_t replaced by $\mathbf{X}_t \triangleq (X_{1,t}, \ldots, X_{L,t})$. We call $\{X_{i,t}, t \in \mathbb{Z}\}$ a *component process*.

We define the quantization of X with precision m as

$$[X]_m \triangleq \frac{\lfloor mX \rfloor}{m} \tag{1}$$

where $\lfloor a \rfloor$ is the largest integer less than or equal to a. Likewise, $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. We denote by $[X_\ell^k]_m = ([X_\ell]_m, \ldots, [X_k]_m)$ the component-wise quantization of X_ℓ^k (and similarly for other collections of RVs or random vectors). Likewise, for complex RVs Z with real part R and imaginary part I, the quantization $[Z]_m$ is equal to $[R]_m + \imath [I]_m$ where $\imath \triangleq \sqrt{-1}$.

Let $H(\cdot)$, $h(\cdot)$, and $D(\cdot\|\cdot)$ denote entropy, differential entropy, and relative entropy, respectively, and let $I(\cdot;\cdot)$ denote the mutual information [11]. We take logarithms to base $e \approx 2.718$, so mutual informations and entropies have dimension *nats*. The entropy rate of a discrete-valued, stationary, L-variate stochastic process $\{\mathbf{X}_t\}$ is [11, Th. 4.2.1]

$$H'(\{\mathbf{X}_t\}) \triangleq \lim_{k\to\infty} \frac{H(\mathbf{X}_1^k)}{k}. \tag{2}$$

Rényi defined the information dimension of a collection of RVs X_ℓ^k as [1]

$$d(X_\ell^k) \triangleq \lim_{m\to\infty} \frac{H([X_\ell^k]_m)}{\log m} \tag{3}$$

provided the limit exists. If the limit does not exist, one can define the upper and lower information dimension $\overline{d}(X_\ell^k)$ and $\underline{d}(X_\ell^k)$ by replacing the limit with the limit superior and limit inferior, respectively. If a result holds for both the limit superior and the limit inferior but it is unclear whether the limit exists, then we shall write $\overline{\underline{d}}(X_\ell^k)$. We shall follow this notation throughout this document: an overline $\overline{(\cdot)}$ indicates that the quantity in the brackets has been computed using the limit superior over m, an underline $\underline{(\cdot)}$ indicates that it has been computed using the limit inferior, both an overline and an underline $\overline{\underline{(\cdot)}}$ indicates that a result holds irrespective of whether the limit superior or limit inferior over m is taken.

If $H([X_\ell^k]_1) < \infty$, then [1, Eq. 7], [4, Prop. 1]

$$0 \le \underline{d}(X_\ell^k) \le \overline{d}(X_\ell^k) \le k - \ell + 1. \tag{4}$$

If $H([X_\ell^k]_1) = \infty$, then $\overline{\underline{d}}(X_\ell^k) = \infty$. As shown in [9, Lemma 3], information dimension is invariant under scaling and translation, i.e., $\overline{\underline{d}}(a \cdot X_\ell^k) = \overline{\underline{d}}(X_\ell^k)$ and $\overline{\underline{d}}(X_\ell^k + c) = \overline{\underline{d}}(X_\ell^k)$ for every $a \neq 0$ and $c \in \mathbb{R}^{k-\ell+1}$.

III. Information Dimension of Univariate Processes

In [7], [8], we generalized (3) by defining the information dimension rate of a univariate stationary process $\{X_t\}$ as

$$d(\{X_t\}) \triangleq \lim_{m\to\infty} \frac{H'(\{[X_t]_m\})}{\log m} = \lim_{m\to\infty} \lim_{k\to\infty} \frac{H([X_1^k]_m)}{k \log m} \tag{5}$$

provided the limit exists. (The limit over k exists by stationarity.)

If $H([X_1]_1) < \infty$, then [8, Lemma 4]

$$0 \le \underline{d}(\{X_t\}) \le \overline{d}(\{X_t\}) \le 1. \tag{6}$$

If $H([X_1]_1) = \infty$, then $\overline{\underline{d}}(\{X_t\}) = \infty$. Moreover, the information dimension rate of the process cannot exceed the information dimension of the marginal RV, i.e.,

$$\overline{\underline{d}}(\{X_t\}) \le \overline{\underline{d}}(X_1). \tag{7}$$

Kawabata and Dembo [2, Lemma 3.2] showed that the information dimension of a RV equals its rate-distortion dimension. By emulating the proof of [2, Lemma 3.2], we generalized this result to stationary processes by demonstrating that the information dimension rate is equal to the rate-distortion dimension. Specifically, let $R(X_1^k, D)$ denote the rate-distortion function of the k-dimensional source X_1^k, i.e.,

$$R(X_1^k, D) \triangleq \inf_{\mathsf{E}[\|\hat{X}_1^k - X_1^k\|^2] \le D} I(X_1^k; \hat{X}_1^k) \tag{8}$$

where the infimum is over all conditional distributions of $\hat{X}_1^k$ given X_1^k such that $\mathsf{E}[\|\hat{X}_1^k - X_1^k\|^2] \le D$ (where $\|\cdot\|$ denotes the Euclidean norm). The rate-distortion dimension of the stationary process $\{X_t\}$ is defined as

$$\mathsf{dim}_R(\{X_t\}) \triangleq 2 \lim_{D\downarrow 0} \lim_{k\to\infty} \frac{R(X_1^k, kD)}{-k \log D} \tag{9}$$

provided the limit as $D \downarrow 0$ exists. By stationarity, the limit over k always exists [12, Th. 9.8.1]. We showed that [8, Th. 5]

$$\overline{\underline{\mathsf{dim}}}_R(\{X_t\}) = \overline{\underline{d}}(\{X_t\}). \tag{10}$$

This result directly generalizes to non-stationary process (possibly with the limit over k replaced by the limit superior or limit inferior).

IV. Information Dimension of Multivariate Processes

In this section, we generalize the definition of the information dimension rate (5) to multivariate (real-valued) processes and study its properties.

Definition 1 (Information Dimension Rate): The information dimension rate of the stationary, L-variate process $\{\mathbf{X}_t\}$ is

$$\begin{aligned} d(\{\mathbf{X}_t\}) &\triangleq \lim_{m\to\infty} \frac{H'(\{[\mathbf{X}_t]_m\})}{\log m} \\ &= \lim_{m\to\infty} \lim_{k\to\infty} \frac{H([X_{1,1}^k]_m, \ldots, [X_{L,1}^k]_m)}{k \log m} \end{aligned} \tag{11}$$

provided the limit over m exists.

We next summarize some basic properties of the information dimension rate.

Lemma 1 (Finiteness and Bounds): Let $\{\mathbf{X}_t\}$ be a stationary, L-variate process. If $H([\mathbf{X}_1]_1) < \infty$, then

$$0 \le \underline{d}(\{\mathbf{X}_t\}) \le \underline{d}(\mathbf{X}_1) \le L. \tag{12}$$

If $H([\mathbf{X}_1]_1) = \infty$, then $\underline{d}(\{\mathbf{X}_t\}) = \infty$.

Proof: Suppose first that $H([\mathbf{X}_1]_1) < \infty$. Then, the rightmost inequality in (12) follows from (4). The leftmost inequality follows from the nonnegativity of entropy. Finally, the center inequality follows since conditioning reduces entropy, hence $H'(\{[\mathbf{X}_t]_m\}) \le H([\mathbf{X}_1]_m)$.

Now suppose that $H([\mathbf{X}_1]_1) = \infty$. By stationarity and since $[\mathbf{X}_1]_1$ is a function of $[\mathbf{X}_1^k]_m$ for every m and every k, we have

$$H([\mathbf{X}_1]_1) \le H([\mathbf{X}_1^k]_m). \tag{13}$$

This implies that $H'(\{[\mathbf{X}_t]_m\}) = \infty$ and the claim $\underline{d}(\{\mathbf{X}_t\}) = \infty$ follows from Definition 1. ∎

It was shown in [9, Lemma 3] that information dimension is invariant under scaling and translation. The same properties hold for the information dimension rate.

Lemma 2 (Scale Invariance): Let $\{\mathbf{X}_t\}$ be a stationary, L-variate process and let $a_i > 0$, $i = 1, \ldots, L$. Further let $Y_{i,t} \triangleq a_i X_{i,t}$, $i = 1, \ldots, L$, $t \in \mathbb{Z}$. Then, $\underline{d}(\{\mathbf{Y}_t\}) = \underline{d}(\{\mathbf{X}_t\})$.

Proof: The proof is based on [4, Lemma 16] and appears in [10]. For brevity, let us focus on the case $L = 2$. The case $L > 2$ follows analogously. For $L = 2$, we have

$$\begin{aligned} &H([a_1 X_{1,1}^k]_m, [a_2 X_{2,1}^k]_m) \\ &\quad\le H([X_{1,1}^k]_m, [X_{2,1}^k]_m) + H([a_1 X_{1,1}^k]_m | [X_{1,1}^k]_m) \\ &\qquad + H([a_2 X_{2,1}^k]_m | [X_{2,1}^k]_m) \\ &\quad\le H([X_{1,1}^k]_m, [X_{2,1}^k]_m) \\ &\qquad + k\log(\lceil a_1 \rceil + 1) + k\log(\lceil a_2 \rceil + 1) \end{aligned} \tag{14}$$

where the second step follows because, given $[X_{i,1}^k]_m$, $[a_i X_{i,1}^k]_m$ can have at most $\lceil a_i \rceil + 1$ possible values. By following the same steps with a_i replaced by $1/a_i$, we obtain the reverse inequality

$$\begin{aligned} H([a_1 X_{1,1}^k]_m, [a_2 X_{2,1}^k]_m) &\ge H([X_{1,1}^k]_m, [X_{2,1}^k]_m) \\ &\quad - k\log(\lceil 1/a_1 \rceil + 1) - k\log(\lceil 1/a_2 \rceil + 1). \end{aligned} \tag{15}$$

The lemma then follows by dividing (14) and (15) by $k \log m$ and by letting k and m tend to infinity. ∎

Lemma 3 (Translation Invariance): Let $\{\mathbf{X}_t\}$ be a stationary, L-variate process and let $\{\mathbf{c}_t\}$, $t \in \mathbb{Z}$ be a sequence of L-dimensional vectors. Then, $\underline{d}(\{\mathbf{X}_t + \mathbf{c}_t\}) = \underline{d}(\{\mathbf{X}_t\})$.

Proof: The lemma follows from [9, Lemma 30], which states that

$$|H(U_1^{kL}) - H(V_1^{kL})| \le \sum_{i=1}^{kL} \log(1 + A_i + B_i) \tag{16}$$

for any collection of integer-valued RVs U_1^{kL} and V_1^{kL} satisfying almost surely $-B_i \le U_i - V_i \le A_i$, $i = 1, \ldots, kL$. Applying this result with $U_{\ell L+j} = \lfloor mX_{\ell,j} + mc_{\ell,j} \rfloor$ and $V_{\ell L+j} = \lfloor mX_{\ell,j} \rfloor + \lfloor mc_{\ell,j} \rfloor$ gives the desired result. Indeed, we have that $-1 \le U_{\ell L+j} - V_{\ell L+j} \le 2$, so (16) yields

$$\left| H([\mathbf{X}_1^k]_m) - H([\mathbf{X}_1^k + \mathbf{c}_1^k]_m) \right| \le kL\log(4). \tag{17}$$

We thus obtain $|d(\{\mathbf{X}_t\}) - d(\{\mathbf{X}_t + \mathbf{c}_t\})| = 0$ by dividing (17) by $k \log m$ and by letting k and m tend to infinity. ∎

We finally observe that the information dimension rate of a stationary stochastic process equals its rate-distortion dimension. This generalizes [8, Th. 5] to multivariate processes.

Theorem 1: Let $\{\mathbf{X}_t\}$ be a stationary, L-variate process. Then,

$$\underline{d}(\{\mathbf{X}_t\}) = \overline{\underline{\mathsf{dim}}}_R\{\mathbf{X}_t\} \tag{18}$$

where $\mathsf{dim}_R\{\mathbf{X}_t\}$ is defined as in (9) but with $\{X_t\}$ replaced by $\{\mathbf{X}_t\}$.

Proof: The proof is analog to that of [2, Lemma 3.2] and [8, Th. 5] and is therefore omitted. ∎

V. Information Dimension of Gaussian Processes

Let $\{\mathbf{X}_t\}$ be a stationary, L-variate, real-valued Gaussian process with mean vector $\boldsymbol{\mu}$ and (matrix-valued) spectral distribution function (SDF) $\theta \mapsto F_{\mathbf{X}}(\theta)$. Thus, $F_{\mathbf{X}}$ is bounded, non-decreasing, and right-continuous on $[-1/2, 1/2]$, and it satisfies [13, (7.3), p. 141]

$$K_{\mathbf{X}}(\tau) = \int_{-1/2}^{1/2} e^{-\imath 2\pi\tau\theta} \mathrm{d}F_{\mathbf{X}}(\theta), \quad \tau \in \mathbb{Z} \tag{19}$$

where $K_{\mathbf{X}}(\tau) \triangleq \mathsf{E}\left[(\mathbf{X}_{t+\tau} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})^{\mathsf{T}}\right]$ denotes the autocovariance function and $(\cdot)^{\mathsf{T}}$ denotes the transpose. It can be shown that $\theta \mapsto F_{\mathbf{X}}(\theta)$ has a derivative almost everywhere, which has positive semi-definite, Hermitian values [13, (7.4), p. 141]. We shall denote the derivative of $F_{\mathbf{X}}$ by $F'_{\mathbf{X}}$.

For univariate stationary Gaussian processes with SDF F_X, we have shown that the information dimension rate is equal to the Lebesgue measure of the set of harmonics on $[-1/2, 1/2]$ where the derivative of F_X is is positive [8, Th. 7], i.e.,

$$d(\{X_t\}) = \lambda(\{\theta\colon F'_X(\theta) > 0\}) \tag{20}$$

where $\lambda(\cdot)$ denotes the Lebesgue measure on $[-1/2, 1/2]$. This result can be directly generalized to the multivariate case where the component processes are independent. Indeed, suppose that $\{\mathbf{X}_t\}$ is a collection of L independent Gaussian processes $\{X_{i,t},\ t \in \mathbb{Z}\}$ with SDFs F_{X_i}. This corresponds to the case where the (matrix-valued) SDF is a diagonal matrix with the SDFs of the individual processes on the main diagonal. For independent processes, the joint entropy rate can be written as the sum of the entropy rates of the component processes. It follows that

$$d(\{\mathbf{X}_t\}) = \sum_{i=1}^{L} d(\{X_{i,t}\}) = \sum_{i=1}^{L} \lambda(\{\theta\colon F'_{X_i}(\theta) > 0\}). \tag{21}$$

The expression on the right-hand side (RHS) of (21) can alternatively be written as

$$\int_{-1/2}^{1/2} \sum_{i=1}^{L} \mathbf{1}\{F'_{X_i}(\theta) > 0\} \mathrm{d}\theta = \int_{-1/2}^{1/2} \mathrm{rank}(F'_{\mathbf{X}}(\theta)) \mathrm{d}\theta \tag{22}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function. Observe that it is immaterial at which frequencies the component processes contain signal power. For example, the information dimension rate of two independent Gaussian processes with bandwidth $1/4$ equals 1 regardless of where the derivatives of their SDFs have their support. The following theorem shows that this result continuous to hold for general L-variate Gaussian processes.

Theorem 2: Let $\{\mathbf{X}_t\}$ be a stationary, L-variate Gaussian process with mean vector $\boldsymbol{\mu}$ and SDF $F_{\mathbf{X}}$. Then,

$$d(\{\mathbf{X}_t\}) = \int_{-1/2}^{1/2} \operatorname{rank}(F'_{\mathbf{X}}(\theta))\mathrm{d}\theta. \tag{23}$$

Proof: Due to space limitations, we only provide a proof outline. The full proof can be found in [10].

We first note that we can assume, without loss of optimality, that $\{\mathbf{X}_t\}$ has zero mean and that every component process of $\{\mathbf{X}_t\}$ has unit variance. Indeed, by Lemma 3, the information dimension rate of $\{\mathbf{X}_t\}$ is translation invariant, so we can subtract the mean without affecting the information dimension rate. Likewise, by Lemma 2, the information dimension rate of $\{\mathbf{X}_t\}$ is scale invariant, so any component process with positive variance can be normalized to a unit-variance process without affecting the information dimension rate. Furthermore, zero-variance component processes can be omitted without affecting neither the left-hand side (LHS) nor the RHS of (23).

We next write the entropy of $[\mathbf{X}_1^k]_m$ as

$$H([\mathbf{X}_1^k]_m) = h(\mathbf{W}_1^k) + kL\log m \tag{24}$$

where $\mathbf{W}_t \triangleq [\mathbf{X}_t]_m + \mathbf{U}_t$, $t \in \mathbb{Z}$ and $\{\mathbf{U}_t\}$ is a sequence of i.i.d. random vectors that are uniformly distributed on the L-dimensional hypercube $[0, 1/m)^L$. Denoting by $(\mathbf{W}_1^k)_G$ a Gaussian vector with the same mean and covariance matrix as $\mathbf{W}_1^k$, and denoting by $f_{\mathbf{W}_1^k}$ and $g_{\mathbf{W}_1^k}$ the probability density functions of $\mathbf{W}_1^k$ and $(\mathbf{W}_1^k)_G$, respectively, this can be expressed as

$$H([\mathbf{X}_1^k]_m) = h\big((\mathbf{W}_1^k)_G\big) + D(f_{\mathbf{W}_1^k}\|g_{\mathbf{W}_1^k}) + kL\log m. \tag{25}$$

The entropy rate of a stationary, multivariate, Gaussian process is given by [13, Th. 7.10]

$$\lim_{k\to\infty} \frac{h((\mathbf{W}_1^k)_G)}{k} = \frac{1}{2}\int_{-1/2}^{1/2} \log\big(2\pi e \det F'_{\mathbf{W}}(\theta)\big)\mathrm{d}\theta. \tag{26}$$

Furthermore, the relative entropy $D(f_{\mathbf{W}_1^k}\|g_{\mathbf{W}_1^k})$ is bounded by [10, Lemma 6]

$$\frac{D(f_{\mathbf{W}_1^k}\|g_{\mathbf{W}_1^k})}{k} \le L\left(\frac{\log\big(2\pi(1+\frac{1}{12})\big)}{2} + \frac{75}{2} + \frac{24}{\pi}\right). \tag{27}$$

Thus, dividing (25) by $k\log m$, and letting first k and then m tend to infinity yields

$$d(\{\mathbf{X}_t\}) = L + \lim_{m\to\infty}\int_{-1/2}^{1/2} \frac{\log\det F'_{\mathbf{W}}(\theta)}{2\log m}\mathrm{d}\theta. \tag{28}$$

It remains to show that the RHS of (28) is equal to the RHS of (23). To this end, we use that for zero-mean processes $\{\mathbf{X}_t\}$ with unit-variance component processes the SDF of $\{[\mathbf{X}_t]_m\}$ can be expressed as [10, Lemma 4]

$$F_{[\mathbf{X}]_m}(\theta) = (2a-1)F_{\mathbf{X}}(\theta) + F_{\mathbf{N}}(\theta) \tag{29}$$

where $a \triangleq \mathsf{E}\left[X_{1,1}[X_{1,1}]_m\right]$ and the diagonal elements of $F_{\mathbf{N}}(\theta)$ satisfy

$$\int_{-1/2}^{1/2} \mathrm{d}F_{N_i}(\theta) \le \frac{1}{m^2}. \tag{30}$$

We can thus express the derivative of the SDF of $\{\mathbf{W}_t\}$ as

$$F'_{\mathbf{W}}(\theta) = (2a-1)F'_{\mathbf{X}}(\theta) + F'_{\mathbf{N}}(\theta) + \frac{1}{12m^2}I_L \tag{31}$$

where I_L denotes the $L\times L$ identity matrix. By performing an analysis similar to that in [8, App. C-A], one can show that

$$\lim_{m\to\infty}\int_{-1/2}^{1/2}\frac{\log\det F'_{\mathbf{W}}(\theta)}{2\log m}\mathrm{d}\theta = -\sum_{i=1}^{L}\lambda(\{\theta\colon \mu_i(\theta)=0\}) \tag{32}$$

where $\mu_i(\theta)$ denotes the i-th eigenvalue of $F'_{\mathbf{X}}(\theta)$. (For the details, see [10, App. A].). Combining (32) with (28) gives

$$\begin{aligned} d(\{\mathbf{X}_t\}) &= \sum_{i=1}^{L}\big[1-\lambda(\{\theta\colon \mu_i(\theta)=0\})\big] \\ &= \sum_{i=1}^{L}\lambda(\{\theta\colon \mu_i(\theta)>0\}) \end{aligned} \tag{33}$$

which as in (21) and (22) can be shown to be equal to the RHS of (23). ∎

VI. Information Dimension of Complex Gaussian Processes

Theorem 2 allows us to study the information dimension of stationary, univariate, complex-valued Gaussian processes by treating them as bivariate, real-valued processes. Let $\{Z_t\}$ be a stationary, univariate, complex-valued, Gaussian process with mean μ and SDF F_Z, i.e.,

$$K_Z(\tau) = \int_{-1/2}^{1/2} e^{-\imath 2\pi\tau\theta}\mathrm{d}F_Z(\theta), \quad \tau\in\mathbb{Z} \tag{34}$$

where $K_Z(\tau) \triangleq \mathsf{E}\left[(Z_{t+\tau}-\mu)(Z_t-\mu)^*\right]$ is the autocovariance function, and $(\cdot)^*$ denotes complex conjugation.

Alternatively, $\{Z_t\}$ can be expressed in terms of its real and imaginary part. Indeed, let $Z_t = R_t + \imath I_t$, $t\in\mathbb{Z}$. The stationary, bivariate, real-valued process $\{(R_t, I_t),\, t\in\mathbb{Z}\}$ is jointly Gaussian and has SDF

$$F_{(R,I)}(\theta) = \begin{pmatrix} F_R(\theta) & F_{RI}(\theta) \\ F_{IR}(\theta) & F_I(\theta)\end{pmatrix}, \quad -\frac{1}{2}\le\theta\le\frac{1}{2} \tag{35}$$

where F_R and F_I are the SDFs of $\{R_t\}$ and $\{I_t\}$, respectively, and F_{RI} and F_{IR} are the cross SDFs between $\{R_t\}$ and $\{I_t\}$. The derivatives of F_Z and $F_{(R,I)}$ are connected as follows:

$$\begin{aligned} F'_Z(\theta) &= F'_R(\theta) + F'_I(\theta) + \imath\big(F'_{IR}(\theta) - F'_{RI}(\theta)\big) \\ &= F'_R(\theta) + F'_I(\theta) + 2\mathfrak{Im}\big(F'_{RI}(\theta)\big) \end{aligned} \tag{36}$$

where the last equality follows because $F'_{(R,I)}$ is Hermitian. Here we use $\mathfrak{Im}(\cdot)$ to denote the imaginary part. It can be further shown that $\theta \mapsto F'_R(\theta)$ and $\theta \mapsto F'_I(\theta)$ are real-valued and symmetric, and that $\theta \mapsto \mathfrak{Im}\big(F'_{RI}(\theta)\big)$ is anti-symmetric.

A stationary, complex-valued process $\{Z_t\}$ is said to be *proper* if its mean μ and its pseudo-autocovariance function

$$\overline{K_Z}(\tau) \triangleq \mathsf{E}\left[(Z_{t+\tau}-\mu)(Z_t-\mu)\right], \quad \tau \in \mathbb{Z}$$

are both zero [14, Def. 17.5.4]. Since, by Lemma 3, the information dimension rate is independent of μ, we shall slightly abuse notation and say that a stationary, complex-valued process is proper if its pseudo-autocovariance function is identically zero, irrespective of its mean. Properness implies that, for all θ, $F_R(\theta) = F_I(\theta)$ and $F_{RI}(\theta) = -F_{IR}(\theta)$. Since $\theta \mapsto F'_{(R,I)}(\theta)$ is Hermitian, this implies that for a proper process the function $\theta \mapsto F'_{RI}(\theta)$ is purely imaginary.

The following corollary to Theorem 2 shows that proper Gaussian processes maximize information dimension. This parallels the result that proper Gaussian vectors maximize differential entropy [15, Th. 2].

Corollary 1: Let $\{Z_t\}$ be a stationary, complex-valued Gaussian process with mean μ and SDF F_Z. Then

$$d(\{Z_t\}) \le 2 \cdot \lambda(\{\theta\colon F'_Z(\theta) > 0\}) \tag{37}$$

with equality if $\{Z_t\}$ is proper.

Proof: We know from Theorem 2 that

$$d(\{Z_t\}) = \int_{-1/2}^{1/2} \operatorname{rank}(F'_{(R,I)}(\theta))\mathrm{d}\theta. \tag{38}$$

For a given θ, the eigenvalues of $F'_{(R,I)}(\theta)$ are given by

$$\frac{F'_R(\theta)+F'_I(\theta)}{2} \pm \sqrt{\frac{(F'_R(\theta)-F'_I(\theta))^2}{4} + |F'_{RI}(\theta)|^2}. \tag{39}$$

Since $F'_{(R,I)}(\theta)$ is positive semi-definite, these eigenvalues are nonnegative and

$$F'_R(\theta)F'_I(\theta) \ge |F'_{RI}(\theta)|^2. \tag{40}$$

The larger of these eigenvalues, say $\mu_1(\theta)$, is zero on

$$\mathcal{F}_1 \triangleq \{\theta\colon F'_R(\theta) = F'_I(\theta) = 0\}. \tag{41}$$

The smaller eigenvalue, $\mu_2(\theta)$, is zero on

$$\mathcal{F}_2 \triangleq \big\{\theta\colon F'_R(\theta)F'_I(\theta) = |F'_{RI}(\theta)|^2\big\}. \tag{42}$$

Clearly, $\mathcal{F}_1 \subseteq \mathcal{F}_2$. By (38), we have that

$$\begin{aligned} d(\{Z_t\}) &= \lambda(\{\theta\colon \mu_1(\theta) > 0\}) + \lambda(\{\theta\colon \mu_2(\theta) > 0\}) \\ &= 1 - \lambda(\mathcal{F}_1) + 1 - \lambda(\mathcal{F}_1) - \lambda(\mathcal{F}_1^{\mathsf{c}} \cap \mathcal{F}_2). \end{aligned} \tag{43}$$

We next note that, by (36) and (40), the derivative $F'_Z(\theta)$ is zero if either $F'_R(\theta) = F'_I(\theta) = 0$ or if $F'_R(\theta) + F'_I(\theta) > 0$ and $F'_R(\theta) + F'_I(\theta) = -2\mathfrak{Im}(F'_{RI}(\theta))$. Since $\theta \mapsto F'_R(\theta)$ and $\theta \mapsto F'_I(\theta)$ are symmetric and $\theta \mapsto \mathfrak{Im}(F'_{RI}(\theta))$ is anti-symmetric, it follows that for any $\theta \in \mathcal{F}_1^c$ satisfying $F'_R(\theta) + F'_I(\theta) = -2\mathfrak{Im}(F'_{RI}(\theta))$ we have that $F'_R(-\theta) + F'_I(-\theta) = 2\mathfrak{Im}(F'_{RI}(-\theta))$. Thus, defining

$$\mathcal{F}_3 \triangleq \big\{\theta\colon F'_R(\theta) + F'_I(\theta) = 2|\mathfrak{Im}(F'_{RI}(\theta))|\big\} \tag{44}$$

we can express the Lebesgue measure of the set of harmonics where $F'_Z(\theta) = 0$ as

$$\lambda(\{\theta\colon F'_Z(\theta) = 0\}) = \lambda(\mathcal{F}_1) + \frac{1}{2}\lambda(\mathcal{F}_1^{\mathsf{c}} \cap \mathcal{F}_3). \tag{45}$$

Combining (43) and (45), we obtain

$$\begin{aligned} d(\{Z_t\}) &= 2\lambda(\{\theta\colon F'_Z(\theta) > 0\}) \\ &\quad + \lambda(\mathcal{F}_1^{\mathsf{c}} \cap \mathcal{F}_3) - \lambda(\mathcal{F}_1^{\mathsf{c}} \cap \mathcal{F}_2). \end{aligned} \tag{46}$$

Since the arithmetic mean is greater than or equal to the geometric mean, and with (40), we have that

$$\begin{aligned} (F'_R(\theta) + F'_I(\theta))^2 &\ge 4F'_R(\theta)F'_I(\theta) \\ &\ge 4|F'_{RI}(\theta)|^2 \ge 4\mathfrak{Im}(F'_{RI}(\theta))^2. \end{aligned} \tag{47}$$

Hence, $\mathcal{F}_3 \subseteq \mathcal{F}_2$ and the second line in (46) is less than or equal to zero. This proves (37).

If $\{Z_t\}$ is proper, then we have $F'_R(\theta) = F'_I(\theta)$ and $|F'_{RI}(\theta)| = |\mathfrak{Im}(F'_{RI}(\theta))|$. In this case, $F'_R(\theta)F'_I(\theta) = |F'_{RI}(\theta)|^2$ implies $F'_R(\theta) + F'_I(\theta) = 2|\mathfrak{Im}(F'_{RI}(\theta))|$, so $\mathcal{F}_2 \subseteq \mathcal{F}_3$. It follows that $\mathcal{F}_2 = \mathcal{F}_3$ and the second line in (46) is zero. Hence, (37) holds with equality. ■

Remark 1: There are also non-proper processes for which (37) holds with equality. For example, this is the case for any stationary Gaussian process for which real and imaginary parts are independent and F'_R and F'_I have matching support but are different otherwise.

References

[1] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, no. 1-2, pp. 193–215, Mar. 1959.

[2] T. Kawabata and A. Dembo, "The rate-distortion dimension of sets and measures," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1564–1572, Sep. 1994.

[3] T. Koch, "The Shannon lower bound is asymptotically tight," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 6155–6161, Nov. 2016.

[4] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.

[5] Y. Wu, S. Shamai (Shitz), and S. Verdú, "Information dimension and the degrees of freedom of the interference channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 256–279, Jan. 2015.

[6] D. Stotz and H. Bölcskei, "Degrees of freedom in vector interference channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 4172–4197, Jul. 2016.

[7] B. C. Geiger and T. Koch, "On the information dimension rate of stochastic processes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 888–892.

[8] ——, "On the information dimension of stochastic processes," `arXiv:1702.00645v2 [cs.IT]`, Feb. 2017.

[9] Y. Wu, "Shannon theory for compressed sensing," Ph.D. dissertation, Princeton University, 2011.

[10] B. C. Geiger and T. Koch, "On the information dimension of multivariate Gaussian processes," `arXiv:1712.07863v1 [cs.IT]`, Dec. 2017.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. Wiley Interscience, 1991.

[12] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[13] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes," *Acta Mathematica*, vol. 98, no. 1, pp. 111–150, 1957.

[14] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge: Cambridge University Press, 2009.

[15] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.