Universidad
Carlos III de Madrid

e-Archivo
Institutional Repository

This is a postprint version of the following published document:

On the Information Dimension Rate
of Stochastic Processes

Bernhard C. Geiger[*], Tobias Koch[†‡]

[*]Institute for Communications Engineering, Technical University of Munich, 80333, Munich, Germany
[†]Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911, Leganés, Spain
[‡]Gregorio Marañón Health Research Institute, 28007, Madrid, Spain.
Emails: bernhard.geiger@tum.de, koch@tsc.uc3m.es

Abstract—**Jalali and Poor ("Universal compressed sensing," arXiv:1406.7807v3, Jan. 2016) have recently proposed a generalization of Rényi's information dimension to stationary stochastic processes by defining the information dimension of the stochastic process as the information dimension of k samples divided by k in the limit as $k \to \infty$. This paper proposes an alternative definition of information dimension as the entropy rate of the uniformly-quantized stochastic process divided by minus the logarithm of the quantizer step size $1/m$ in the limit as $m \to \infty$. It is demonstrated that both definitions are equivalent for stochastic processes that are ψ^*-mixing, but that they may differ in general. In particular, it is shown that for Gaussian processes with essentially-bounded power spectral density (PSD), the proposed information dimension equals the Lebesgue measure of the PSD's support. This is in stark contrast to the information dimension proposed by Jalali and Poor, which is 1 if the process's PSD is positive on a set of positive Lebesgue measure, irrespective of its support size.**

I. INTRODUCTION

Information dimension, proposed by Rényi in [1], was recently given an operational characterization in (almost) lossless compressed sensing. In [2], Wu and Verdú analyzed both linear encoding and Lipschitz decoding of discrete-time, independent and identically distributed (i.i.d.), stochastic processes and showed that information dimension plays a fundamental role in achievability and converse results.

Jalali and Poor proposed a generalization of information dimension to stationary, discrete-time, stochastic processes [3], [4] by defining the *information dimension* $d'(\mathbf{X})$ of the stochastic process $\{X_t,\, t \in \mathbb{Z}\}$ as the information dimension of $(X_1, \ldots, X_k)$ divided by k in the limit as $k \to \infty$.[1] They showed that, for ψ^*-mixing processes, the information dimension $d'(\mathbf{X})$ is a fundamental limit for universal compressed

sensing with linear encoding and decoding via Lagrangian minimum entropy pursuit [3, Th. 8]. Rezagah *et al.* showed in [5], [6] that $d'(\mathbf{X})$ coincides, under certain conditions, with the *rate-distortion dimension* $\dim_R(\{X_t\})$, defined as twice the rate-distortion function of the stochastic process $\{X_t,\, t \in \mathbb{Z}\}$ divided by $-\log D$ in the limit as $D \downarrow 0$. This generalizes to stochastic processes the result by Kawabata and Dembo that the rate-distortion dimension of a random variable (RV) equals its information dimension [7].

In this work, we propose a different definition for the information dimension of stationary, discrete-time, stochastic processes. Specifically, let $\{[X_t]_m,\, t \in \mathbb{Z}\}$ denote the stochastic process $\{X_t,\, t \in \mathbb{Z}\}$ uniformly quantized with step size $1/m$. We define the *information dimension rate* $d(\{X_t\})$ of $\{X_t,\, t \in \mathbb{Z}\}$ as the entropy rate of $\{[X_t]_m,\, t \in \mathbb{Z}\}$ divided by $\log m$ in the limit as $m \to \infty$. For i.i.d. processes, our definition of information dimension coincides with that of Jalali and Poor (and, in fact, evaluates to Rényi's information dimension of the marginal RV X_t). More generally, we show that these definitions are equivalent for ψ^*-mixing processes.

Nevertheless, there are stochastic processes for which the two definitions disagree. In particular, for Gaussian processes with essentially-bounded power spectral density (PSD) S_X, we demonstrate that $d(\{X_t\})$ is equal to the Lebesgue measure of the set of harmonics on $[-1/2, 1/2]$ for which S_X is positive. Thus, for processes with a bandlimited PSD, the information dimension rate $d(\{X_t\})$ is equal to twice the PSD's bandwidth. This is consistent with the intuition that for such processes not all samples contain information. For example, if the bandwidth of the PSD is $1/4$, then we expect that half of the samples in $\{X_t,\, t \in \mathbb{Z}\}$ can be expressed as linear combinations of the other samples and, hence, do not contain information. In contrast, we show that the information dimension $d'(\mathbf{X})$ is 1 if S_X is positive on a set of positive Lebesgue measure. In other words, $d'(\mathbf{X})$ does not capture the dependence of the information dimension rate on the support size of S_X.

By emulating the proof of [7, Lemma 3.2], we show that for any stochastic process $\{X_t,\, t \in \mathbb{Z}\}$, the information dimension rate $d(\{X_t\})$ coincides with the rate-distortion dimension $\dim_R(\{X_t\})$. This implies that $d'(\mathbf{X}) = \dim_R(\{X_t\})$ only for those stochastic processes for which $d'(\mathbf{X}) = d(\{X_t\})$.

Due to space limitations, some proofs are deferred to a longer version of our paper [8].

[1]More precisely, Jalali and Poor define the information dimension via a conditional entropy of the uniformly-quantized process. For stationary processes, their definition coincides with the above-mentioned definition [3, Lemma 3].

II. Notation and Preliminaries

We denote by $\mathbb{R}$ and $\mathbb{Z}$ the set of real numbers and the set of integers, respectively. For a finite or countably infinite collection of RVs we abbreviate $\mathbf{X}_\ell^k \triangleq (X_\ell, \ldots, X_{k-1}, X_k)$, $\mathbf{X}_\ell \triangleq (X_\ell, X_{\ell+1}, \ldots)$, and $\mathbf{X}^k \triangleq (\ldots, X_{k-1}, X_k)$. If a collection $\mathbf{X}_\ell^k$ of RVs has a probability density function (PDF), then we denote it as $f_{\mathbf{X}_\ell^k}$. We denote stochastic processes as $\{X_t, t \in \mathbb{Z}\}$ or, in short, as $\{X_t\}$.

Let $\lfloor \cdot \rfloor$ denote the floor function. We define the quantization of X with precision m as

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

and denote by $[\mathbf{X}_\ell^k]_m = ([X_\ell]_m, \ldots, [X_k]_m)$ the component-wise quantization of $\mathbf{X}_\ell^k$. Quantizations such as $[\mathbf{X}_\ell]_m$ and $[\mathbf{X}^k]_m$ are defined accordingly.

Let $H(\cdot)$, $h(\cdot)$, and $D(\cdot\|\cdot)$ denote entropy, differential entropy, and relative entropy, respectively, and let $I(\cdot; \cdot)$ denote mutual information [9]. We take logarithms to base $e \approx 2.718$, i.e., mutual informations and entropies have dimension *nats*. The entropy rate of a discrete-valued, stationary, stochastic process $\{X_t\}$ is defined as

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

If $H(X_t) < \infty$, then the stationarity of $\{X_t\}$ guarantees that the limit in (2) exists and is equal to [9, Th. 4.2.1]

$$\lim_{k \to \infty} \frac{H(\mathbf{X}_1^k)}{k} = \lim_{k \to \infty} H(X_k | \mathbf{X}_1^{k-1}). \tag{3}$$

We say that $\{X_t\}$ is ψ^*-*mixing* if

$$\lim_{k \to \infty} \sup \frac{P_{\mathbf{X}^0, \mathbf{X}_k}(A \cap B)}{P_{\mathbf{X}^0}(A) P_{\mathbf{X}_k}(B)} = 1 \tag{4}$$

where the supremum is over all $A \in \mathcal{F}^0$ and $B \in \mathcal{F}_k$ satisfying $P_{\mathbf{X}^0}(A) P_{\mathbf{X}_k}(B) > 0$. (Here, $\mathcal{F}^0$ and $\mathcal{F}_k$ denote the σ-fields generated by $\mathbf{X}^0$ and $\mathbf{X}_k$, respectively.)

The *Rényi information dimension* of a collection of RVs $\mathbf{X}_\ell^k$ is defined as [1]

$$d(\mathbf{X}_\ell^k) \triangleq \lim_{m \to \infty} \frac{H([\mathbf{X}_\ell^k]_m)}{\log m}, \quad \text{if the limit exists.} \tag{5}$$

The *upper Rényi information dimension* $\bar{d}(\mathbf{X}_\ell^k)$ is defined as in (5), but with the limit replaced by the limit superior.

The Rényi information dimension is finite if, and only if, $H([\mathbf{X}_\ell^k]_1) < \infty$ [1], [2]. If $H([\mathbf{X}_\ell^k]_1) < \infty$, then the Rényi information dimension of a collection of $(k-\ell+1)$ real-valued (scalar) RVs satisfies [1, Eq. (7)], [2, Prop. 1]

$$0 \leq \bar{d}(\mathbf{X}_\ell^k) \leq k - \ell + 1. \tag{6}$$

Furthermore, if X is a real-valued (scalar) RV satisfying $H([X]_1) < \infty$ and with probability measure

$$P_X = (1 - \rho) P_d + \rho P_c \tag{7}$$

where P_d is a discrete measure, P_c is an absolutely continuous measure, and $0 \leq \rho \leq 1$, then $d(X) = \rho$ [1, Th. 3], [2,

Th. 1]. This implies that the Rényi information dimension of a discrete RV is zero, and the Rényi information dimension of a continuous RV with a PDF is one.

III. Information Dimension Rate of Stochastic Processes

We next generalize the definition of information dimension to stationary stochastic processes.

Definition 1: Let $\{X_t\}$ be a stationary stochastic process. Then, the *information dimension rate* is defined as

$$d(\{X_t\}) \triangleq \lim_{m \to \infty} \frac{H'([\mathbf{X}]_m)}{\log m}, \quad \text{if the limit exists.} \tag{8}$$

We define the *upper information dimension rate* $\bar{d}(\{X_t\})$ as in (8), but with the limit replaced by the limit superior.

It can be shown that $\bar{d}(\{X_t\})$ is finite if, and only if, $H([X_t]_1) < \infty$. If $H([X_t]_1) < \infty$, then

$$0 \leq \bar{d}(\{X_t\}) \leq 1. \tag{9}$$

The information dimension rate can be related to the rate-distortion dimension of $\{X_t\}$. Let $R(\mathbf{X}_1^k, D)$ denote the rate-distortion function of the k-dimensional source $\mathbf{X}_1^k$, i.e.,

$$R(\mathbf{X}_1^k, D) \triangleq \inf_{\substack{P_{\hat{\mathbf{X}}_1^k | \mathbf{x}_1^k}: \\ \mathsf{E}[\|\hat{\mathbf{X}}_1^k - \mathbf{X}_1^k\|^2] \leq D}} I(\mathbf{X}_1^k; \hat{\mathbf{X}}_1^k) \tag{10}$$

where the infimum is over all conditional distributions of $\hat{\mathbf{X}}_1^k$ given $\mathbf{X}_1^k$ such that $\mathsf{E}[\|\hat{\mathbf{X}}_1^k - \mathbf{X}_1^k\|^2] \leq D$ (where $\|\cdot\|$ denotes the Euclidean norm).

Definition 2: The *rate-distortion dimension* of the stochastic process $\{X_t\}$ is defined as

$$\dim_R(\{X_t\}) \triangleq 2 \lim_{D \downarrow 0} \lim_{k \to \infty} \frac{R(\mathbf{X}_1^k, kD)}{-k \log D} \tag{11}$$

provided the limits exist. If the limits do not exist, we define the *upper rate-distortion dimension* $\overline{\dim}_R(\{X_t\})$ by replacing the limits with limits superior.

Kawabata and Dembo showed that the rate-distortion dimension of a RV equals its information dimension [7]. This result can be generalized to stochastic processes.

Theorem 1: For any stochastic process $\{X_t\}$,

$$2 \varlimsup_{D \downarrow 0} \varlimsup_{k \to \infty} \frac{R(\mathbf{X}_1^k, kD)}{-k \log D} = \varlimsup_{m \to \infty} \varlimsup_{k \to \infty} \frac{H([\mathbf{X}_1^k]_m)}{k \log m}. \tag{12}$$

Here, $\varlimsup$ denotes the limit superior.

Proof: See Section V. $\blacksquare$

Note that Theorem 1 does not require $\{X_t\}$ to be stationary. If $\{X_t\}$ is stationary, then the limits over k exist (see [10, Th. 9.8.1] and [9, Th. 4.2.1]), and we have $\overline{\dim}_R(\{X_t\}) = \bar{d}(\{X_t\})$. Thus, the (upper) information dimension rate is equal to the (upper) rate-distortion dimension.

We next derive an expression for the information dimension rate of stationary Gaussian processes $\{X_t\}$ with mean μ, variance σ^2, and PSD $\theta \mapsto \mathsf{S}_X(\theta)$. Thus, S_X is a nonnegative, symmetric function satisfying

$$K_X(\tau) = \int_{-1/2}^{1/2} e^{-i2\pi\tau\theta} \mathsf{S}_X(\theta) d\theta, \quad \tau \in \mathbb{Z} \tag{13}$$

where $K_X(\tau) \triangleq \mathsf{E}\left[(X_{k+\tau} - \mu)(X_k - \mu)\right]$ denotes the auto-covariance function of $\{X_t\}$ and $\mathsf{i} \triangleq \sqrt{-1}$.

Theorem 2: Let $\{X_t\}$ be a stationary Gaussian process with mean μ, variance σ^2, and essentially-bounded PSD S_X. Then,

$$d(\{X_t\}) = \lambda\left(\{\theta : \mathsf{S}_X(\theta) > 0\}\right) \tag{14}$$

where $\lambda(\cdot)$ denotes the Lebesgue measure on $[-1/2, 1/2]$.

Proof: See Section VI. ∎

Thus, for stationary Gaussian processes, the information dimension rate is equal to the support size of its PSD.

IV. ANOTHER DEFINITION OF INFORMATION DIMENSION

Jalali and Poor [3], [4] proposed a different definition for the information dimension of stochastic processes. We shall refer to this information dimension as the *block-average information dimension* and denote it by $d'(\mathbf{X})$.

Definition 3: Let $\{X_t\}$ be a stationary stochastic process. Then, the *block-average information dimension* is defined as

$$d'(\mathbf{X}) \triangleq \lim_{k \to \infty} \frac{d(\mathbf{X}_1^k)}{k} \tag{15}$$

if the information dimension $d(\mathbf{X}_1^k)$ exists. We define the *upper block-average information dimension* $\bar{d}'(\mathbf{X})$ as in (15), but with $d(\mathbf{X}_1^k)$ replaced by $\bar{d}(\mathbf{X}_1^k)$.

Remark 1: More precisely, Jalali and Poor define the information dimension of the stationary process $\{X_t\}$ as [3, Def. 2]

$$d'(\mathbf{X}) \triangleq \lim_{k \to \infty} \lim_{m \to \infty} \frac{H([X_k]_m | [\mathbf{X}_1^{k-1}]_m)}{\log m} \tag{16}$$

if the limit over m exists (the limit over k always exists). They further define the upper information dimension $\bar{d}'(\mathbf{X})$ by replacing the limit over m by the limit superior. As shown in [3, Lemma 3], this definition coincides with Definition 3.

For the class of ψ^*-mixing processes, Jalali and Poor [3], [4] demonstrated that $\bar{d}'(\mathbf{X})$ is a fundamental limit for universal compressed sensing with linear encoding and decoding via Lagrangian minimum entropy pursuit. Thus, for this class of stochastic processes, $\bar{d}'(\mathbf{X})$ has an operational meaning.

We next demonstrate that, for ψ^*-mixing processes, the information dimension rate $\bar{d}(\{X_t\})$ coincides with the block-average information dimension $\bar{d}'(\mathbf{X})$. Thus, for such processes the information dimension rate inherits the operational meaning of the block-average information dimension. However, in general the two definitions do not coincide, but there exists an ordering between them.

Theorem 3: Let $\{X_t\}$ be a stationary stochastic process. Then,

$$\bar{d}(\{X_t\}) \leq \bar{d}'(\mathbf{X}). \tag{17}$$

Moreover, if there exists a nonnegative integer n such that

$$I(\mathbf{X}_1^k; \mathbf{X}^{-n}) < \infty, \quad k = 1, 2, \ldots \tag{18}$$

then $\bar{d}(\{X_t\}) = \bar{d}'(\mathbf{X})$.

Proof: See [8]. ∎

Condition (18) is satisfied if $\{X_t\}$ is ψ^*-mixing. It further holds if $\{X_t\}$ is a sequence of i.i.d. RVs, if it is a discrete-valued process with finite marginal entropy, or if it is a continuous-valued process with finite marginal differential entropy and finite differential entropy rate. Example 1 in [8] further demonstrates that $\bar{d}(\{X_t\}) = \bar{d}'(\mathbf{X})$ for the piecewise constant processes studied in [3]–[6], even though they do not satisfy (18). However, in many cases, the inequality (17) can be strict, as demonstrated by the following example.

Example 1: Let $\{X_t\}$ be a zero-mean, variance-σ^2, stationary Gaussian process with bounded PSD S_X having support $\mathcal{I}_X$. It follows from Theorem 2 that $d(\{X_t\}) = \lambda(\mathcal{I}_X)$. We next argue that the assumption $\lambda(\mathcal{I}_X) > 0$ implies that $d'(\mathbf{X}) = 1$. Consequently, $d(\{X_t\}) < d'(\mathbf{X})$ unless S_X has full support. Indeed, denoting by $\underline{\lim}$ the limit inferior, we obtain

$$\varliminf_{m \to \infty} \frac{H([\mathbf{X}_1^k]_m)}{k \log m} \geq \varliminf_{m \to \infty} \frac{H([X_0]_m \mid \mathbf{X}_{-k}^{-1})}{\log m} \tag{19}$$

by the chain rule, the stationarity of $\{X_t\}$, and because conditioning reduces entropy. Since $\{X_t\}$ is Gaussian, it follows that, conditioned on $\mathbf{X}_{-k}^{-1}$, the RV X_0 is Gaussian with mean $\mathsf{E}[X_0|\mathbf{X}_{-k}^{-1}]$ and variance σ_k^2, which is independent of $\mathbf{X}_{-k}^{-1}$. It can be further shown that if $\lambda(\mathcal{I}_X) > 0$ then $\sigma_k^2 > 0$ for every finite k; see, e.g., [8, Lemma 4]. It follows that, conditioned on any $\mathbf{X}_{-k}^{-1} = \mathbf{x}_{-k}^{-1}$, the RV X_0 has a PDF, so by [1, Th. 3], [2, Th. 1]

$$\lim_{m \to \infty} \frac{H([X_0]_m \mid \mathbf{X}_{-k}^{-1} = \mathbf{x}_{-k}^{-1})}{\log m} = 1. \tag{20}$$

With Fatou's lemma, this shows that the right-hand side (RHS) of (19) is 1. Together with Definition 3 and (6), this gives $d'(\mathbf{X}) = 1$.

As a final remark, Rezagah *et al.* showed that if $\lim_{D \to 0} \frac{R(\mathbf{X}_1^k, kD)}{-k \log D}$ exists for all k, then [5], [6, Th. 2]

$$\mathsf{dim}_R(\{X_t\}) = \bar{d}'(\mathbf{X}). \tag{21}$$

This may appear as a contradiction to our results, since we have demonstrated in Theorem 1 that $\mathsf{dim}_R(\{X_t\}) = d(\{X_t\})$ (if the limits exist), and Example 1 shows that there are stochastic processes for which $d(\{X_t\}) < d'(\mathbf{X})$. However, the proof of (21) relies on the fact that [5, Sec. VI-E]

$$\left| \frac{1}{k} R(\mathbf{X}_1^k, kD) - \lim_{\kappa \to \infty} \frac{1}{\kappa} R(\mathbf{X}_1^\kappa, \kappa D) \right| \leq \frac{1}{k} I(\mathbf{X}_1^k; \mathbf{X}^0) \tag{22}$$

and that the RHS of (22) vanishes as $k \to \infty$. If (18) holds for $n = 0$, then this is indeed the case. As shown in Theorem 3, in this case we also have $\bar{d}(\{X_t\}) = \bar{d}'(\mathbf{X})$. In contrast, the RHS of (22) is infinite, e.g., if $\{X_t\}$ is a stationary stochastic process with a PSD that is zero on a set of positive Lebesgue measure, since for such processes $h(X_1|\mathbf{X}^0) = -\infty$. Our proof of Theorem 1 does not rely on (22). We thus conclude that $\mathsf{dim}_R(\{X_t\}) = d(\{X_t\})$ for any stochastic process $\{X_t\}$ for which the limits in (12) exist, but that $\mathsf{dim}_R(\{X_t\}) = d'(\mathbf{X})$ only for those processes for which $d'(\mathbf{X}) = d(\{X_t\})$.

V. Proof of Theorem 1

The proof of Theorem 1 is essentially identical to the proof of [7, Lemma 3.2] by Kawabata and Dembo, so we only provide a brief outline. More details can be found in [8]. Indeed, choosing in (10) $\hat{\mathbf{X}}_1^k = [\mathbf{X}_1^k]_m$, $m = \frac{1}{\sqrt{D}}$ yields

$$R(\mathbf{X}_1^k, kD) \leq H([\mathbf{X}_1^k]_m) \tag{23}$$

since for this choice, $\|\mathbf{X}_1^k - \hat{\mathbf{X}}_1^k\|^2 \leq kD$ almost surely, hence it satisfies the distortion constraint. That the left-hand side (LHS) of (12) is upper-bounded by the RHS of (12) then follows by dividing by $-k \log D$ and by taking limits over k and D.

To show that the LHS of (12) is lower-bounded by the RHS of (12), we use the following bound given in [7, Eq. (A.1)]:

$$R(\mathbf{X}_1^k, D) \geq \sup_{s \leq 0, \lambda_s} \left\{ sD + \mathsf{E}\left[\log \lambda_s(\mathbf{X}_1^k)\right] \right\} \tag{24}$$

where $\lambda_s \colon \mathbb{R}^k \to [0, \infty)$ is any nonnegative measurable function satisfying

$$\sup_{\mathbf{y}_1^k \in \mathbb{R}^k} \mathsf{E}\left[\lambda_s(\mathbf{X}_1^k) e^{s \sum_{\ell=1}^k (y_\ell - X_\ell)^2}\right] \leq 1. \tag{25}$$

As in the proof of [7, Lemma 3.2], we apply (24) with

$$s = -m^2 \tag{26a}$$

$$\lambda_s(\mathbf{x}_1^k) = \frac{1}{N^k} \sum_{\mathbf{i}_1^k \in \mathbb{Z}^k} \frac{\mathbf{1}\{[\mathbf{x}_1^k]_m = \mathbf{i}_1^k/m\}}{\Pr([\mathbf{x}_1^k]_m = \mathbf{i}_1^k/m)} \tag{26b}$$

$$N = 1 + 2 \sum_{i=0}^{\infty} e^{-i^2} \tag{26c}$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Evaluating (24) for this choice gives

$$R(\mathbf{X}_1^k, kD) \geq H([\mathbf{X}_1^k]_m) - k(m^2 D + \log N). \tag{27}$$

The claim thus follows by choosing $m = \frac{1}{\sqrt{D}}$, by dividing both sides of (27) by $-k \log D$ and by taking limits over k and D. This proves Theorem 1.

VI. Proof of Theorem 2

Let $\{X_t\}$ be a stationary Gaussian process with mean μ, variance σ^2, and essentially-bounded PSD S_X, and let $Z_t = [X_t]_m$, $t \in \mathbb{Z}$. To compute $d(\{X_t\})$, we express $H(\mathbf{Z}_1^k)$ as

$$H(\mathbf{Z}_1^k) = h(Z_1 + U_{1,m}, \ldots, Z_k + U_{k,m}) + k \log m \tag{28}$$

where $\{U_{t,m}, t \in \mathbb{Z}\}$ is a sequence of i.i.d. RVs with $U_{t,m}$ uniformly distributed on the interval $[0, 1/m]$. Defining $W_t \triangleq Z_t + U_{t,m}$, and using that $h(\mathbf{W}_1^k)$ can be written as $h((\mathbf{W}_1^k)_G) - D(f_{\mathbf{W}_1^k} \| g_{\mathbf{W}_1^k})$ (where $(\mathbf{W}_1^k)_G$ denotes a Gaussian vector with the same mean and covariance matrix as $\mathbf{W}_1^k$, and $g_{\mathbf{W}_1^k}$ denotes its PDF) we obtain

$$H'(\mathbf{Z}) = \lim_{k \to \infty} \frac{h((\mathbf{W}_1^k)_G) - D(f_{\mathbf{W}_1^k} \| g_{\mathbf{W}_1^k})}{k} + \log m. \tag{29}$$

The first term on the RHS of (29) is the differential entropy rate of a stationary Gaussian process, which can be expressed

in terms of its PSD S_W [9, Sec. 11.5]. By the independence of $\{Z_t\}$ and $\{U_{t,m}\}$, S_W is given by $\mathsf{S}_Z + \mathsf{S}_U$, where S_Z and S_U denote the PSDs of $\{Z_t\}$ and $\{U_{t,m}\}$, respectively. Since $\{U_{t,m}\}$ is i.i.d., we have that

$$\mathsf{S}_U(\theta) = \frac{1}{12m^2}, \quad -\frac{1}{2} \leq \theta \leq \frac{1}{2}. \tag{30}$$

The following lemma provides a characterization of S_Z.

Lemma 1: Let $\{X_t\}$ be a stationary Gaussian process with mean μ, variance σ^2, and essentially-bounded PSD S_X, and let $Z_t = [X_t]_m$, $t \in \mathbb{Z}$. Then, the PSD S_Z of $\{Z_t\}$ satisfies

$$a_1^2 \mathsf{S}_X(\theta) \leq \mathsf{S}_Z(\theta) \leq a_1^2 \mathsf{S}_X(\theta) + \frac{c}{m^2}, \quad -\frac{1}{2} \leq \theta \leq \frac{1}{2} \tag{31}$$

where $a_1 \triangleq \mathsf{E}\left[(Z_t - \mathsf{E}[Z_t])(X_t - \mu)\right]/\sigma^2$ and c is a positive constant. Moreover, $|1 - a_1| \leq \frac{1}{m}\sqrt{2/(\pi\sigma^2)}$.

Proof: The proof is based on [11] and appears in [8]. ∎

Combining Lemma 1 with (30), it follows that

$$\mathsf{S}_W(\theta) = a_1^2 \mathsf{S}_X(\theta) + g_m(\theta), \quad -\frac{1}{2} \leq \theta \leq \frac{1}{2} \tag{32}$$

where the function $\theta \mapsto g_m(\theta)$ satisfies

$$\frac{\mathsf{k}_1}{m^2} \leq g_m(\theta) \leq \frac{\mathsf{K}_1}{m^2}, \quad -\frac{1}{2} \leq \theta \leq \frac{1}{2} \tag{33}$$

for some nonnegative constants k_1 and K_1. Together with [9, Sec. 11.5], this yields

$$\lim_{k \to \infty} \frac{h((\mathbf{W}_1^k)_G)}{k} = \frac{1}{2} \log(2\pi e) + \frac{1}{2} \int_{-1/2}^{1/2} \log\left(a_1^2 \mathsf{S}_X(\theta) + g_m(\theta)\right) \mathrm{d}\theta. \tag{34}$$

We next consider the second term on the RHS of (29). We show that, for some positive constant K_2,

$$\lim_{k \to \infty} \frac{D(f_{\mathbf{W}_1^k} \| g_{\mathbf{W}_1^k})}{k} \leq \mathsf{K}_2. \tag{35}$$

To this end, we first note that $\mathbf{X} = (X_1, \ldots, X_k)^\mathsf{T}$ (where $(\cdot)^\mathsf{T}$ denotes transpose) can be written as [12, Th. 23.6.14]

$$\mathbf{X} = A\mathbf{N} + \boldsymbol{\mu} \tag{36}$$

where $\mathbf{N}$ is a k'-dimensional, zero-mean, Gaussian vector $(k' \leq k)$ with independent components whose variances are the nonzero eigenvalues of the covariance matrix $C_{\mathbf{X}}$ of $\mathbf{X}$, A is a $k \times k'$ matrix satisfying $A^\mathsf{T}A = I$ (where I denotes the identity matrix), and $\boldsymbol{\mu} = (\mu, \ldots, \mu)^\mathsf{T}$. Denoting $\mathbf{W} = (W_1, \ldots, W_k)^\mathsf{T}$, the chain rule for relative entropy yields

$$D(f_{\mathbf{W}_1^k} \| g_{\mathbf{W}_1^k}) \leq D(f_{\mathbf{W},\mathbf{N}} \| g_{\mathbf{W},\mathbf{N}})$$
$$= \int D(f_{\mathbf{W}|\mathbf{N}=\mathbf{n}} \| g_{\mathbf{W}|\mathbf{N}=\mathbf{n}}) f_{\mathbf{N}}(\mathbf{n}) \mathrm{d}\mathbf{n} \tag{37}$$

where $g_{\mathbf{W},\mathbf{N}}$ denotes the PDF of a Gaussian vector with the same mean and covariance matrix as $(\mathbf{W}, \mathbf{N})$, and $g_{\mathbf{W}|\mathbf{N}=\mathbf{n}}(\mathbf{w}) \triangleq \frac{g_{\mathbf{W},\mathbf{N}}(\mathbf{w},\mathbf{n})}{f_{\mathbf{N}}(\mathbf{n})}$. The inequality in (37) follows because the conditional relative entropy between $f_{\mathbf{N}|\mathbf{W}}$ and $g_{\mathbf{N}|\mathbf{W}}$ is nonnegative; the equality follows because $\mathbf{N}$ is Gaussian, so $D(f_{\mathbf{N}} \| g_{\mathbf{N}}) = 0$.

To evaluate the relative entropy on the RHS of (37), we note that $\mathbf{x}$ can be obtained from $\mathbf{n}$ via (36), so for $\mathbf{w} \in \mathbb{R}^k$,

$$f_{\mathbf{W}|\mathbf{N}=\mathbf{n}}(\mathbf{w}) = m^k \mathbf{1}\{[\mathbf{w}]_m = [A\mathbf{n} + \boldsymbol{\mu}]_m\}. \quad (38)$$

Thus, denoting $\mathcal{W}(\mathbf{n}) = \{\mathbf{w} \in \mathbb{R}^k \colon [\mathbf{w}]_m = [A\mathbf{n} + \boldsymbol{\mu}]_m\}$,

$$D(f_{\mathbf{W}|\mathbf{N}=\mathbf{n}}\|g_{\mathbf{W}|\mathbf{N}=\mathbf{n}}) = \log\left(m^k\sqrt{(2\pi)^k \det C_{\mathbf{W}|\mathbf{N}}}\right)$$
$$+ \frac{m^k}{2}\int_{\mathcal{W}(\mathbf{n})}(\mathbf{w} - \mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}})^\mathsf{T}C_{\mathbf{W}|\mathbf{N}}^{-1}(\mathbf{w} - \mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}})d\mathbf{w} \quad (39)$$

where $\mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}}$ and $C_{\mathbf{W}|\mathbf{N}}$ denote the conditional expectation and the conditional covariance matrix of $\mathbf{W}$ given $\mathbf{N} = \mathbf{n}$. These can be computed as [12, Th. 23.7.4]

$$\mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}} = \mathsf{E}[\mathbf{W}] + C_{\mathbf{WN}}C_{\mathbf{N}}^{-1}\mathbf{n} \quad (40a)$$
$$C_{\mathbf{W}|\mathbf{N}} = C_{\mathbf{W}} - C_{\mathbf{WN}}C_{\mathbf{N}}^{-1}C_{\mathbf{WN}}^\mathsf{T} \quad (40b)$$

where $C_{\mathbf{WN}}$ denotes the cross-covariance matrix of $\mathbf{W}$ and $\mathbf{N}$, and $C_{\mathbf{W}}$ and $C_{\mathbf{N}}$ denote the covariance matrices of $\mathbf{W}$ and $\mathbf{N}$, respectively.

By Bussgang's theorem, the cross-covariance matrix of $\mathbf{W}$ and $\mathbf{X}$ is $C_{\mathbf{WX}} = a_1 C_{\mathbf{X}}$, where a_1 is as in Lemma 1. From (36) we get $C_{\mathbf{X}} = AC_{\mathbf{N}}A^\mathsf{T}$ and $C_{\mathbf{WN}} = C_{\mathbf{WX}}A$, hence $C_{\mathbf{WN}} = a_1 AC_{\mathbf{N}}$. Together with (40), this yields

$$\mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}} = \mathsf{E}[\mathbf{W}] + a_1 A\mathbf{n} \quad (41a)$$
$$C_{\mathbf{W}|\mathbf{N}} = C_{\mathbf{W}} - a_1^2 C_{\mathbf{X}}. \quad (41b)$$

Combining (41a) with (36), we can upper-bound each component of $\mathbf{w} - \mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}}$ as

$$|w_\ell - \mathsf{E}[W_\ell] - a_1(x_\ell - \mu)|$$
$$\leq |z_\ell - x_\ell| + |u_\ell - \mathsf{E}[U_\ell]|$$
$$+ |\mathsf{E}[Z_\ell] - \mu| + |1 - a_1||x_\ell - \mu|. \quad (42)$$

The first and the third term on the RHS of (42) are both upper-bounded by $\frac{1}{m}$, and the second term is upper-bounded by $\frac{1}{2m}$. With Lemma 1 we thus get

$$\|\mathbf{w} - \mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}}\|^2 \leq \frac{1}{m^2}\left(\frac{25k}{2} + \frac{4}{\pi\sigma^2}\|\mathbf{x} - \boldsymbol{\mu}\|^2\right). \quad (43)$$

We next note that, since $\mathsf{S}_W = \mathsf{S}_Z + \mathsf{S}_U$ and $\mathsf{S}_U(\theta) = \frac{1}{12m^2}$,

$$C_{\mathbf{W}|\mathbf{N}} = C_{\mathbf{Z}} - a_1^2 C_{\mathbf{X}} + \frac{1}{12m^2}I \quad (44)$$

where $\mathbf{Z} = (Z_1, \ldots, Z_k)^\mathsf{T}$. Noting that $C_{\mathbf{Z}} - a_1^2 C_{\mathbf{X}}$ is the conditional covariance matrix of $\mathbf{Z}$ given $\mathbf{N}$ and, therefore, is positive semidefinite, it follows that the smallest eigenvalue of $C_{\mathbf{W}|\mathbf{N}}$ is lower-bounded by $\frac{1}{12m^2}$. Together with (43), this yields for the second term on the RHS of (39)

$$\frac{m^k}{2}\int_{\mathcal{W}(\mathbf{n})}(\mathbf{w} - \mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}})^\mathsf{T}C_{\mathbf{W}|\mathbf{N}}^{-1}(\mathbf{w} - \mu_{\mathbf{W}|\mathbf{N}=\mathbf{n}})d\mathbf{w}$$
$$\leq 6m^{k+2}\frac{1}{m^k}\frac{1}{m^2}\left(\frac{25k}{2} + \frac{4}{\pi\sigma^2}\|\mathbf{x} - \boldsymbol{\mu}\|^2\right)$$
$$= k\eta_1 + \eta_2\|\mathbf{x} - \boldsymbol{\mu}\|^2 \quad (45)$$

for some constants η_1 and η_2 that do not depend on (k, m).

To upper-bound the first term on the RHS of (39), we use that (44) and Lemma 1 imply that every diagonal element of $C_{\mathbf{W}|\mathbf{N}}$ is upper-bounded by Υ/m^2, where $\Upsilon \triangleq c + \frac{1}{12}$. It follows from Hadamard's inequality that

$$\log\left(m^k\sqrt{(2\pi)^k \det C_{\mathbf{W}|\mathbf{N}}}\right) \leq \frac{k}{2}\log(2\pi\Upsilon). \quad (46)$$

Combining (45) and (46) with (39) and (37) yields

$$D(f_{\mathbf{W}_1^k}\|g_{\mathbf{W}_1^k}) \leq \frac{k}{2}\log(2\pi\Upsilon) + k\eta_1 + \eta_2\mathsf{E}\left[\|\mathbf{X} - \boldsymbol{\mu}\|^2\right] \quad (47)$$

which together with $\mathsf{E}\left[\|\mathbf{X} - \boldsymbol{\mu}\|^2\right] = k\sigma^2$ proves (35).

To conclude the proof of Theorem 2, we divide $H'(\mathbf{Z})$ in (29) by $\log m$ and let m tend to infinity. The nonnegativity of relative entropy and the upper bound (35) imply that the relative entropy on the RHS of (29) does not contribute to the information dimension rate. Thus, by (34),

$$d(\{X_t\}) = 1 + \lim_{m\to\infty}\frac{1}{2}\int_{-1/2}^{1/2}\frac{\log\left(a_1^2\mathsf{S}_X(\theta) + g_m(\theta)\right)}{\log m}d\theta. \quad (48)$$

To evaluate the RHS of (48), we follow along the lines of [13, Sec. VI] and swap the order of limit and integral (which can be justified by the dominated and monotone convergence theorem). The integrand vanishes as $m \to \infty$ for those frequencies θ for which $\mathsf{S}_X(\theta) > 0$. Moreover, by (33), $\lim_{m\to\infty}\frac{\log g_m(\theta)}{\log m} = -2$. Consequently,

$$\lim_{m\to\infty}\frac{1}{2}\int_{-1/2}^{1/2}\frac{\log\left(a_1^2\mathsf{S}_X(\theta) + g_m(\theta)\right)}{\log m}d\theta$$
$$= -\lambda(\{\theta\colon \mathsf{S}_X(\theta) = 0\}). \quad (49)$$

Together with (48), this demonstrates that $d(\{X_t\}) = \lambda(\{\theta\colon \mathsf{S}_X(\theta) > 0\})$ and concludes the proof.

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] 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.

[3] S. Jalali and H. V. Poor, "Universal compressed sensing," `arXiv:1406.7807v3 [cs.IT]`, Jan. 2016.

[4] ——, "Universal compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Jul. 2016, pp. 2369–2373.

[5] F. E. Rezagah, S. Jalali, E. Erkip, and H. V. Poor, "Compression-based compressed sensing," `arXiv:1601.01654 [cs.IT]`, Jan. 2016.

[6] ——, "Rate-distortion dimension of stochastic processes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Jul. 2016, pp. 2079–2083.

[7] 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.

[8] B. C. Geiger and T. Koch, "On the information dimension of stochastic processes," `arXiv:1702.00645v1 [cs.IT]`, Feb. 2017.

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

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

[11] D. Chan and R. W. Donaldson, "Correlation functions and reconstruction error for quantized Gaussian signals transmitted over discrete memoryless channels," *IEEE Trans. Inf. Theory*, pp. 519–523, Jul. 1972.

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

[13] ——, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.