

# A Lower Bound on the Entropy Rate for a Large Class of Stationary Processes and its Relation to the Hyperplane Conjecture

Meik Dörpinghaus  
 Vodafone Chair Mobile Communications Systems  
 Technische Universität Dresden  
 01062 Dresden, Germany  
 meik.doerpinghaus@tu-dresden.de

**Abstract**—We present a new lower bound on the differential entropy rate of stationary processes whose sequences of probability density functions fulfill certain regularity conditions. This bound is obtained by showing that the gap between the differential entropy rate of such a process and the differential entropy rate of a Gaussian process with the same autocovariance function is bounded. This result is based on a recent result on bounding the Kullback-Leibler divergence by the Wasserstein distance given by Polyanskiy and Wu. Moreover, it is related to the famous *hyperplane conjecture*, also known as *slicing problem*, in convex geometry originally stated by J. Bourgain. Based on an entropic formulation of the hyperplane conjecture given by Bobkov and Madiman we discuss the relation of our result to the hyperplane conjecture.

## I. INTRODUCTION

We consider a discrete-time stationary stochastic process  $\{X\}$ . A finite sequence of this process is given by the vector  $X^n = [X_1, \dots, X_n]$  in  $\mathbb{R}^n$  with density  $f^{(n)}$ . The differential entropy rate of the process  $\{X\}$  is defined as

$$h'(\{X\}) = \lim_{n \rightarrow \infty} \frac{1}{n} h(X^n) \quad (1)$$

when the limit exists and where  $h(\cdot)$  denotes the differential entropy.

In general, no closed form solutions for the entropy rate are available. However, for a stationary Gaussian random process  $\{X_{\text{Gauss}}\}$  having the same autocovariance function and the same mean as the process  $\{X\}$  the entropy rate can be expressed in closed form. Let

$$r(l) = E[(X_{k+l} - E[X_{k+l}])(X_k - E[X_k])] \quad (2)$$

be the autocovariance function of the process  $\{X\}$ . For a blocklength of  $n$  the elements of its covariance matrix of size

This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the cluster of excellence EXC 1056 ‘Center for Advancing Electronics Dresden (cfaed)’ and by the DFG within the Collaborative Research Center SFB 912 ‘Highly Adaptive Energy-Efficient Computing (HAEC)’.

During the course of this work the author was a visiting assistant professor at the Department of Electrical Engineering, Stanford University, CA, U.S.A.

$n \times n$  are given by

$$[\mathbf{R}_X^{(n)}]_{kl} = r(k-l). \quad (3)$$

Then the entropy rate of  $\{X_{\text{Gauss}}\}$  is given by, see, e.g., [1, Ch. 12.5]<sup>1</sup>

$$\begin{aligned} h'(\{X_{\text{Gauss}}\}) &= \lim_{n \rightarrow \infty} \frac{1}{2n} \log((2\pi e)^n \det(\mathbf{R}_X^{(n)})) \\ &= \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(2\pi e S_X(f)) df \end{aligned} \quad (4)$$

where (4) follows with Szegő’s theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [2, pp. 64-65], [3] with the power spectral density  $S_X(f)$  given by

$$S_X(f) = \sum_{m=-\infty}^{\infty} r(m) e^{-j2\pi m f}, \quad -\frac{1}{2} < f < \frac{1}{2} \quad (5)$$

with  $j = \sqrt{-1}$ .

As for a given covariance matrix Gaussian random variables maximize the differential entropy [1, Th. 8.6.5], it follows that the entropy rate  $h'(\{X\})$  is upper-bounded by

$$h'(\{X\}) \leq h'(\{X_{\text{Gauss}}\}). \quad (6)$$

While for given second moments an upper bound on the differential entropy and the corresponding rate can be easily given by considering the maximum entropy property of the Gaussian distribution as in (6), there exist not many general lower bounds on differential entropy and the corresponding rate. One approach of lower-bounding the differential entropy has been presented in [4, Theorem I.1] where a lower bound on the entropy per coordinate is constructed based on a Gaussian density having the same maximum density as the actual density of the considered random vector. I.e., differently to matching the first two moments when using the maximum entropy property of the Gaussian distribution for upper-bounding, here the maximum of the density between the actual distribution and a Gaussian distribution is matched.

<sup>1</sup>Unless otherwise stated, in this paper  $\log$  is to an arbitrary but fixed base.

Related to this, in [4, Corollary IX.1] under specific conditions the entropy rate of a stationary process can be bounded away from  $-\infty$ .

Differently, the lower bound on the entropy rate presented in this paper is based on an upper bound on the Kullback-Leibler divergence between the actual density and a Gaussian density with the same first two moments. In this regard, note that

$$\begin{aligned} h'(\{X_{\text{Gauss}}\}) - h'(\{X\}) &= \lim_{n \rightarrow \infty} \frac{1}{n} [h(X_{\text{Gauss}}^n) - h(X^n)] \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} D(f^{(n)}) \end{aligned} \quad (7)$$

where  $X_{\text{Gauss}}^n = [X_{\text{Gauss},1}, \dots, X_{\text{Gauss},n}]$  is the vector containing the elements  $1, \dots, n$  of the process  $\{X_{\text{Gauss}}\}$ . Moreover,  $D(f^{(n)})$  is the Kullback-Leibler divergence or relative entropy between the density  $f^{(n)}$  and a corresponding Gaussian distribution with the same mean and the same covariance matrix as  $X^n \sim f^{(n)}$  and with density  $g^{(n)}$ , i.e.

$$D(f^{(n)}) = D(f^{(n)} \| g^{(n)}) = \int f^{(n)}(x^n) \log \frac{f^{(n)}(x^n)}{g^{(n)}(x^n)} dx^n \quad (8)$$

where the superscript at  $x^n$  denotes a vector of dimension  $n$ . For (7) we have used that, see, e.g., [1, pp. 254-255]

$$D(f^{(n)}) = h(X_{\text{Gauss}}^n) - h(X^n) \quad (9)$$

as  $X_{\text{Gauss}}^n$  is Gaussian with the same mean and the same covariance matrix as  $X^n$ .

In the present paper, we show that

$$\frac{1}{n} D(f^{(n)}) \leq c < \infty \quad (10)$$

for all  $n$  if the sequence of densities  $f^{(n)}$  fulfills certain regularity conditions and, hence, that

$$\lim_{n \rightarrow \infty} \frac{1}{n} D(f^{(n)}) \leq c < \infty. \quad (11)$$

Here,  $c$  is a uniform constant independent of  $n$ .

The proof of (10) is based on a recent result of bounding the Kullback-Leibler divergence by the Wasserstein distance given by Polyanskiy and Wu [5]. More specifically, we show that the LHS of (10) is bounded away from infinity for all  $n$  if the sequence of probability density functions  $f^{(n)}$  is  $(c_1, c_2)$ -regular with respect to the following definition given in [5].

A probability density function  $f^{(n)}$  on  $\mathbb{R}^n$  is  $(c_1, c_2)$ -regular if  $c_1 > 0$ ,  $c_2 \geq 0$ , and

$$\|\nabla \log f^{(n)}(x^n)\| \leq c_1 \|x^n\| + c_2, \quad \forall x^n \in \mathbb{R}^n \quad (12)$$

where  $\|\cdot\|$  denotes the Euclidean distance and with the differential operator  $\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right)$  for  $x^n = [x_1, \dots, x_n]$ .

We state that a sequence of densities  $f^{(n)}$  or a process  $\{X\}$  are  $(c_1, c_2)$ -regular if there exist constants  $c_1$  and  $c_2$  such that (12) holds for all  $n \in \mathbb{N}$ .

In conclusion, we state the following theorem.

**Theorem 1.** *Let  $\{X\}$  be a stationary stochastic process with autocovariance function  $r(l)$  and probability density function*

*$f^{(n)}$  for a sequence of length  $n$ . Moreover, let  $\{X_{\text{Gauss}}\}$  be a Gaussian process with the same mean and the same autocovariance function as the process  $\{X\}$ . If the sequence of densities  $f^{(n)}$  is  $(c_1, c_2)$ -regular, it holds that*

$$\frac{1}{n} D(f^{(n)}) = \frac{1}{n} [h(X_{\text{Gauss}}^n) - h(X^n)] \leq c < \infty, \quad \forall n \quad (13)$$

where  $c$  is a uniform constant independent of  $n$ .

With Theorem 1 it follows that

$$h'(\{X_{\text{Gauss}}\}) - h'(\{X\}) = \lim_{n \rightarrow \infty} \frac{1}{n} [h(X_{\text{Gauss}}^n) - h(X^n)] \leq c \quad (14)$$

and we get the following lower bound on the differential entropy rate of  $\{X\}$

$$h'(\{X\}) \geq h'(\{X_{\text{Gauss}}\}) - c. \quad (15)$$

In Section II we give the proof of Theorem 1 and provide an example on the application of this lower bound.

The result stated in Theorem 1 is related to the so called *hyperplane conjecture*, also referred to as *slicing problem*, which is a longstanding open problem in convex geometry. The hyperplane conjecture has been initially formulated by J. Bourgain [6] and has raised a lot of attention in almost the last 30 years. Its initial version by Bourgain was a slight variant of the following form [4, Conj. V.2].

**Conjecture 1** (*Hyperplane Conjecture or Slicing Problem*). *There exists a universal, positive constant  $c$  (not depending on  $n$ ) such that for any convex set  $K$  of unit volume in  $\mathbb{R}^n$ , there exists a hyperplane  $H$  such that the  $(n-1)$ -dimensional volume of the section  $K \cap H$  is bounded below by  $c$ .*

There exist several equivalent formulations of the hyperplane conjecture having a geometric or a functional analytic flavor [4]. As summarized in [4], Milman and Pajor [7] looked like Bourgain [6] on a setting of centrally symmetric, convex bodies, while Ball's formulation of the conjecture [8] states that the isotropic constant of a log-concave measure in any Euclidean space is bounded above by a universal constant independent of the dimension. In addition to these formulations, Bobkov and Madiman gave the following entropic form of the hyperplane conjecture.

**Conjecture 2** (*Entropic Form of the Hyperplane Conjecture*). *[4, Conjecture V.4] For any log-concave density  $f^{(n)}$  on  $\mathbb{R}^n$  and some universal constant  $c$*

$$\frac{D(f^{(n)})}{n} \leq c. \quad (16)$$

As stated in [4] Conjecture 2 gives a formulation of the hyperplane conjecture as a statement about the (dimension-free) closeness of a log-concave measure to a Gaussian measure.

While the proof of the hyperplane conjecture is in general open, partial results are known, see, e.g., [6], [7], [9] and references therein.

Obviously, Conjecture 2 and Theorem 1 have some similarities as they both upper-bound  $D(f^{(n)})/n$ . However, there

are also significant differences. We will discuss the relation between both in Section III.

## II. LOWER-BOUNDING OF ENTROPY RATES

In the following, we present a proof of Theorem 1 based on a result by Polyanskiy and Wu given in [5, Proposition 1]. Before stating this proposition we recall the definition of the Wasserstein distance as given in [5].

The Wasserstein distance on the Euclidean space is defined as follows. Let  $\mathbf{X}$  and  $\mathbf{Y}$  be random vectors in  $\mathbb{R}^n$ . Given probability measures  $\mu, \nu$  on  $\mathbb{R}^n$ , their  $p$ -Wasserstein distance ( $p \geq 1$ ) is given by

$$W_p(\mu, \nu) = \inf (\mathbb{E}[\|\mathbf{X} - \mathbf{Y}\|^p])^{1/p} \quad (17)$$

where the infimum is taken over all couplings of  $\mu$  and  $\nu$ , i.e., joint distributions  $P_{\mathbf{X}\mathbf{Y}}$  whose marginals satisfy  $P_{\mathbf{X}} = \mu$  and  $P_{\mathbf{Y}} = \nu$ .

Now we are ready to state the proposition given by Polyanskiy and Wu.

**Proposition 1.** [5, Prop. 1] *Let  $\mathbf{U}$  and  $\mathbf{V}$  be random vectors with finite second moments. If  $\mathbf{V}$  has a  $(c_1, c_2)$ -regular density  $p_{\mathbf{V}}$ , then there exists a coupling  $P_{\mathbf{U}\mathbf{V}}$ , such that<sup>2</sup>*

$$\mathbb{E} \left[ \left| \log \frac{p_{\mathbf{V}}(\mathbf{V})}{p_{\mathbf{V}}(\mathbf{U})} \right| \right] \leq \Delta \quad (18)$$

where

$$\Delta = \left( \frac{c_1}{2} \sqrt{\mathbb{E}[\|\mathbf{U}\|^2]} + \frac{c_1}{2} \sqrt{\mathbb{E}[\|\mathbf{V}\|^2]} + c_2 \right) W_2(P_{\mathbf{U}}, P_{\mathbf{V}}). \quad (19)$$

Consequently,

$$h(\mathbf{U}) - h(\mathbf{V}) \leq \Delta. \quad (20)$$

If both  $\mathbf{U}$  and  $\mathbf{V}$  are  $(c_1, c_2)$ -regular, then

$$|h(\mathbf{U}) - h(\mathbf{V})| \leq \Delta \quad (21)$$

$$D(P_{\mathbf{U}}\|P_{\mathbf{V}}) + D(P_{\mathbf{V}}\|P_{\mathbf{U}}) \leq 2\Delta. \quad (22)$$

Based on Proposition 1 we now prove Theorem 1.

*Proof of Theorem 1:* We assume that the sequence of densities  $f^{(n)}$  is  $(c_1, c_2)$ -regular. Hence, there exist a finite  $c_1 > 0$  and a finite  $c_2 \geq 0$  such that (12) holds for all  $n \in \mathbb{N}$ . We further assume that  $g^{(n)}$  is a Gaussian density having the same mean and the same covariance matrix as  $f^{(n)}$ . As  $g^{(n)}$  is Gaussian the Kullback-Leibler divergence between  $f^{(n)}$  and  $g^{(n)}$  can be expressed as a difference of differential entropies, see (9)

$$\begin{aligned} D(f^{(n)}) &= D(f^{(n)}\|g^{(n)}) = h(g^{(n)}) - h(f^{(n)}) \\ &= h(\mathbf{X}_{\text{Gauss}}^n) - h(\mathbf{X}^n) \end{aligned} \quad (23)$$

<sup>2</sup>Note, on the LHS of (18) the expectation is taken with respect to the random variable  $(\mathbf{U}, \mathbf{V})$ , i.e.,

$$\mathbb{E} \left[ \left| \log \frac{p_{\mathbf{V}}(\mathbf{V})}{p_{\mathbf{V}}(\mathbf{U})} \right| \right] = \int \int p_{\mathbf{U}\mathbf{V}}(u, v) \left| \log \frac{p_{\mathbf{V}}(v)}{p_{\mathbf{V}}(u)} \right| dudv$$

where  $p_{\mathbf{U}\mathbf{V}}$  is the joint density of  $(\mathbf{U}, \mathbf{V})$ .

where  $\mathbf{X}^n \sim f^{(n)}$  and  $\mathbf{X}_{\text{Gauss}}^n \sim g^{(n)}$ . The Kullback-Leibler divergence is always nonnegative [1, Theorem 8.6.1].

Identifying  $\mathbf{U}$  and  $\mathbf{V}$  in Prop. 1 with  $\mathbf{X}_{\text{Gauss}}^n$  and  $\mathbf{X}^n$ , respectively, with (20) and (19)  $D(f^{(n)})$  is upper-bounded by

$$\begin{aligned} D(f^{(n)}) &\leq \left( \frac{c_1}{2} \sqrt{\mathbb{E}[\|\mathbf{X}_{\text{Gauss}}^n\|^2]} + \frac{c_1}{2} \sqrt{\mathbb{E}[\|\mathbf{X}^n\|^2]} + c_2 \right) \\ &\quad \times W_2(g^{(n)}, f^{(n)}) \\ &= \left( c_1 \sqrt{\mathbb{E}[\|\mathbf{X}^n\|^2]} + c_2 \right) \inf \sqrt{\mathbb{E}[\|\mathbf{X}_{\text{Gauss}}^n - \mathbf{X}^n\|^2]} \end{aligned} \quad (24)$$

where for (24) we have substituted the Wasserstein distance  $W_2(g^{(n)}, f^{(n)})$  using its definition in (17). The infimum is taken over all couplings of  $g^{(n)}$  and  $f^{(n)}$ . Note, as we assume that the densities  $f^{(n)}$  and  $g^{(n)}$  exist, we have substituted the distributions in the argument of Wasserstein distance in (19) by the corresponding densities. Moreover, we have used that  $\mathbf{X}^n$  and  $\mathbf{X}_{\text{Gauss}}^n$  have the same mean and the same covariance. Using the triangle inequality we can further upper-bound the RHS of (24) yielding

$$\begin{aligned} D(f^{(n)}) &\leq \left( c_1 \sqrt{\mathbb{E}[\|\mathbf{X}^n\|^2]} + c_2 \right) \inf \sqrt{\mathbb{E}[(\|\mathbf{X}_{\text{Gauss}}^n\| + \|\mathbf{X}^n\|)^2]} \\ &= \left( c_1 \sqrt{\mathbb{E}[\|\mathbf{X}^n\|^2]} + c_2 \right) \\ &\quad \times \inf \sqrt{\mathbb{E}[\|\mathbf{X}_{\text{Gauss}}^n\|^2 + 2\|\mathbf{X}_{\text{Gauss}}^n\|\|\mathbf{X}^n\| + \|\mathbf{X}^n\|^2]} \\ &\leq \left( c_1 \sqrt{\mathbb{E}[\|\mathbf{X}^n\|^2]} + c_2 \right) \sqrt{4\mathbb{E}[\|\mathbf{X}^n\|^2]} \end{aligned} \quad (25)$$

where for (25) we have applied the Cauchy-Schwarz inequality and the equality of the second moments of  $\mathbf{X}^n$  and  $\mathbf{X}_{\text{Gauss}}^n$ . With (25) we get

$$\begin{aligned} \frac{D(f^{(n)})}{n} &\leq \frac{2c_1 \mathbb{E}[\|\mathbf{X}^n\|^2]}{n} + \frac{2c_2 \sqrt{\mathbb{E}[\|\mathbf{X}^n\|^2]}}{n} \\ &= 2c_1 \mathbb{E}[\|\mathbf{X}_k^n\|^2] + \frac{2c_2 \sqrt{\mathbb{E}[\|\mathbf{X}_k^n\|^2]}}{\sqrt{n}} \\ &\leq c < \infty \quad \forall n \end{aligned} \quad (26)$$

where in (26)  $\mathbf{X}_k^n$  is the  $k$ -th element of the vector  $\mathbf{X}^n$  and where we have used that  $\{\mathbf{X}\}$  is stationary. Moreover, (27) holds as the second moments of  $\mathbf{X}^n \sim f^{(n)}$  as well as  $c_1$  and  $c_2$  are finite and with  $c$  being a uniform constant independent of  $n$ , which completes the proof.  $\blacksquare$

Theorem 1 states that for a stationary  $(c_1, c_2)$ -regular process  $\{\mathbf{X}\}$  the differential entropy rate is upper-bounded by the differential entropy rate of a Gaussian process with the same autocovariance function, such that  $\{h(\mathbf{X}_{\text{Gauss}}^n) - h(\mathbf{X}^n)\}/n$  is not diverging for  $n \rightarrow \infty$ . For  $n \rightarrow \infty$  we get with (26), cf. (7)

$$h'(\{\mathbf{X}_{\text{Gauss}}\}) - h'(\{\mathbf{X}\}) \leq 2c_1 \mathbb{E}[\|\mathbf{X}_k^n\|^2] = 2c_1(\mu^2 + \sigma^2) \quad (28)$$

with  $\mu$  and  $\sigma^2$  being the mean and the variance of the elements of the process  $\{X\}$ .<sup>3</sup> Differently stated, with (4) it follows that the differential entropy rate  $h'(\{X\})$  is lower-bounded by

$$h'(\{X\}) \geq \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(2\pi e S_X(f)) df - 2c_1(\mu^2 + \sigma^2). \quad (29)$$

As  $h'(\{X\})$  does not depend on the mean value of  $\{X\}$ , we can set  $\mu = 0$ .

#### A. Example

We consider the stationary random process  $\{Y\}$ . Let  $Y^n$  be a vector containing a sequence of length  $n$  of  $\{Y\}$ . It is given by

$$Y^n = H^n \odot X^n + Z^n \quad (30)$$

where  $\odot$  denotes the Hadamard product, i.e., the element wise product. Moreover,  $H^n$  and  $X^n$  are vectors containing sequences of length  $n$  of the two zero-mean stationary Gaussian processes  $\{H\}$  and  $\{X\}$ , respectively. These processes have the autocorrelation functions  $r_H(l)$  and  $r_X(l)$ , cf. (2). The corresponding power spectral densities are denoted by  $S_H(f)$  and  $S_X(f)$ , cf. (5). Finally, the entries of the vector  $Z^n$  correspond to a sequence of length  $n$  of the i.i.d. zero-mean Gaussian process  $\{Z\}$  and have variance  $\sigma_Z^2$ . The processes  $\{H\}$ ,  $\{X\}$ , and  $\{Z\}$  are mutually independent. The mean and the variance of the elements of the process  $\{Y\}$  are given by  $\mu = 0$  and  $\sigma^2 = r_H(0)r_X(0) + \sigma_Z^2$ . Moreover, the power spectral density  $S_Y(f)$  is given by

$$S_Y(f) = (S_H \star S_X)(f) + \sigma_Z^2, \quad -\frac{1}{2} < f < \frac{1}{2} \quad (31)$$

where  $\star$  denotes convolution.

The vector  $Y^n$  is not Gaussian distributed and no closed form for the probability density function is available. Moreover, no closed form solution for the differential entropy rate of the process  $\{Y\}$  is available. Thus, our aim is to lower-bound the differential entropy rate of  $\{Y\}$  using the approach described above. For this purpose we use the following proposition given by Polyanskiy and Wu.

**Proposition 2.** [5, Prop. 2] *Let  $Y^n = B^n + Z^n$  where  $B^n$  is independent of  $Z^n \sim \mathcal{N}(0, \sigma_Z^2 \mathbf{I}_n)$ , with  $\mathbf{I}_n$  being the identity matrix of size  $n \times n$ . Then the density of  $Y^n$  is  $(c_1, c_2)$ -regular with  $c_1 = \frac{3 \log e}{\sigma_Z^2}$  and  $c_2 = \frac{4 \log e}{\sigma_Z^2} \mathbb{E}[\|B^n\|]$ .*

Using Proposition 2 it follows that  $Y^n$  in (30) is  $(c_1, c_2)$ -regular with  $c_1 = \frac{3 \log e}{\sigma_Z^2}$  and  $c_2 = \frac{4 \log e}{\sigma_Z^2} \sqrt{nr_H(0)r_X(0)} \leq \frac{4 \log e}{\sigma_Z^2} \mathbb{E}[\|H^n \odot X^n\|]$ . However, as  $c_2$  depends on  $n$  we cannot directly apply (29) but still have to consider the second term on the RHS of (26) for bounding, yielding

$$h'(\{Y\}) \geq \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(2\pi e S_Y(f)) df - 2c_1 \sigma^2 - \lim_{n \rightarrow \infty} \frac{2c_2 \sqrt{\sigma^2}}{\sqrt{n}}. \quad (32)$$

<sup>3</sup>This holds in case  $c_1$  and  $c_2$  are independent of  $n$ , which is fulfilled by our definition of a  $(c_1, c_2)$ -regular process.

Introducing  $c_1, c_2, \sigma^2$ , and  $S_Y(f)$  yields

$$\begin{aligned} h'(\{Y\}) &\geq \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(2\pi e ((S_H \star S_X)(f) + \sigma_Z^2)) df \\ &\quad - 6 \log e \left( \frac{r_H(0)r_X(0)}{\sigma_Z^2} + 1 \right) \\ &\quad - \frac{8 \log e}{\sigma_Z^2} \sqrt{r_H(0)r_X(0)} \sqrt{r_H(0)r_X(0) + \sigma_Z^2}. \end{aligned} \quad (33)$$

The larger  $\sigma_Z^2$  becomes the more Gaussian the process  $\{Y\}$  gets and the closer the lower bound on the RHS of (33) comes to the differential entropy rate of a Gaussian process with the same power spectral density which is given by  $\frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \log(2\pi e ((S_H \star S_X)(f) + \sigma_Z^2)) df$ . For large  $\sigma_Z^2$  the difference between the lower bound and the actual value of  $h'(Y)$  converges to 6 nats.

### III. RELATION TO THE HYPERPLANE CONJECTURE

The methods applied to prove Theorem 1 allow also to state the following theorem.

**Theorem 2.** *For any sequence of log-concave probability density functions  $f^{(n)}$  on  $\mathbb{R}^n$  for which the second moments are uniformly bounded and which is  $(c_1, c_2)$ -regular it holds that*

$$\frac{D(f^{(n)})}{n} \leq c \quad (34)$$

with  $c$  being a uniform constant independent of  $n$ .

The difference between Theorem 1 and Theorem 2 is that in the latter one we do not assume that the sequence of probability density functions  $f^{(n)}$  arises from a stationary process. Instead, we make the assumption that the sequence of  $f^{(n)}$  has a uniform bound for all second moments. In this regard, note that in the proof of Theorem 1 the assumption on stationarity is only used in (26). However, with the constraint on uniformly bounded second moments it similarly to (26) holds that

$$\begin{aligned} \frac{1}{n} D(f^{(n)}) &\leq 2c_1 \max_{k \in \{1, \dots, n\}} \mathbb{E}[\|X_k^n\|^2] \\ &\quad + \frac{2c_2 \sqrt{\max_{k \in \{1, \dots, n\}} \mathbb{E}[\|X_k^n\|^2]}}{\sqrt{n}} \\ &\leq c < \infty \quad \forall n. \end{aligned} \quad (35)$$

Hence, with this slight change of the proof of Theorem 1 Theorem 2 follows.

Although Theorem 2 looks to some extent similar to the entropic form of the hyperplane conjecture (Conjecture 2) there are some significant differences. On the one hand, Theorem 2 is limited to sequences of log-concave probability density functions which have a uniform bound for all second moments and which are  $(c_1, c_2)$ -regular. On the other hand, the constant  $c$  in Theorem 2 is a uniform constant independent of  $n$ , however, it depends on the individual sequence of  $f^{(n)}$ . Differently, the constant  $c$  in Conjecture 2 is universal and holds for any log-concave density  $f^{(n)}$  on  $\mathbb{R}^n$ . Thus, by the methods applied in the present paper we do not obtain a proof of the hyperplane conjecture.

The hyperplane conjecture and, thus, Theorem 2 are formulated for probability density functions  $f^{(n)}$  which are log-concave. The question is, which log-concave density functions are  $(c_1, c_2)$ -regular such that Theorem 2 can be applied? First of all note that a regular density fulfilling (12) has infinite support on  $\mathbb{R}^n$ , i.e.,  $f^{(n)}$  is never zero. However, even for  $f^{(n)}$  having infinite support neither does log-concavity imply  $(c_1, c_2)$ -regularity, nor does  $(c_1, c_2)$ -regularity imply log-concavity. Hence, the question remains, which log-concave densities  $f^{(n)}$  with infinite support are  $(c_1, c_2)$ -regular?

If we assume that  $f^{(n)}$  is differentiable on  $\mathbb{R}^n$ , for any finite  $x^n$  we can find a  $c_1$  and  $c_2$  such that (12) is fulfilled. In order to prove this consider that the derivative of  $f^{(n)}$  is finite on any closed set. What remains is to check the case  $\|x^n\| \rightarrow \infty$ . In this regard, note that (12) implies that

$$|\log f^{(n)}(x^n) - \log f^{(n)}(0)| \leq \frac{c_1}{2} \|x^n\|^2 + c_2 \|x^n\|. \quad (37)$$

Without loss of generality we can assume that the maximum of  $f^{(n)}$  is attained at  $x^n = 0$ , as a shift of the density  $f^{(n)}$  has no impact on its log-concavity. As log-concavity of  $f^{(n)}$  implies that  $f^{(n)}$  is unimodal, it holds that  $f^{(n)}(x^n) \leq f^{(n)}(0)$ . Hence, (37) yields<sup>4</sup>

$$f^{(n)}(x^n) \geq f^{(n)}(0) e^{-\left(\frac{c_1}{2} \|x^n\|^2 + c_2 \|x^n\|\right)}. \quad (38)$$

On the other hand, log-concave densities can be bounded from above using the following lemma given in [10, Lemma 1].

**Lemma 1.** *Let  $f^{(n)}$  be a log-concave density on  $\mathbb{R}^n$ . There exist  $a > 0$  and  $b \in \mathbb{R}$  such that  $f^{(n)}(x^n) \leq e^{-a\|x^n\|+b}$  for all  $x^n \in \mathbb{R}^n$ .*

As the upper bound on log-concave densities in Lemma 1 decays only with  $e^{-a\|x^n\|}$  whereas  $(c_1, c_2)$ -regularity requires that the density decays not faster than with  $e^{-\frac{c_1}{2}\|x^n\|^2}$ , for every  $(c_1, c_2)$ -regular and log-concave density there exist constants  $a$  and  $b$  such that

$$f^{(n)}(0) e^{-\left(\frac{c_1}{2} \|x^n\|^2 + c_2 \|x^n\|\right)} \leq f^{(n)}(x^n) \leq e^{-a\|x^n\|+b} \quad (39)$$

holds. I.e., densities being log-concave as well as  $(c_1, c_2)$ -regular have to be characterized by an exponential decay with a rate between the bounds given in (39). That means a log-concave density  $f^{(n)}$  with infinite support on  $\mathbb{R}^n$  which is differentiable and for which we can find constants  $c_1$  and  $c_2$  is  $(c_1, c_2)$ -regular in addition to being log-concave.

One example for a density falling into this class is the *Logistic*-distribution whose density is given by

$$f(x) = \frac{e^x}{(1 + e^x)^2}, \quad x \in \mathbb{R}. \quad (40)$$

Also the multivariate Gaussian density falls into this class.

In conclusion, although Theorem 2 is related to the entropic form of the hyperplane conjecture (both upper-bound  $D(f^{(n)})/n$ ) and can be applied to any sequence of log-concave density functions  $f^{(n)}$  which is  $(c_1, c_2)$ -regular and for which the second moments are uniformly bounded, there

is the significant difference that the constant  $c$  in Theorem 2 depends on the sequence of  $f^{(n)}$  whereas it is a universal constant in the hyperplane conjecture.

*Remark:* It has to be stated that we have only one concrete example of a sequence of  $(c_1, c_2)$ -regular log-concave probability density functions. Namely if the elements of  $X^n \sim f^{(n)}$  correspond to a sequence of a discrete-time zero-mean stationary Gaussian process  $\{X\}$  with autocovariance function  $r(l)$  and corresponding power spectral density  $S_X(f)$ , see (5), being supported in the entire interval  $[-\frac{1}{2}, \frac{1}{2}]$ , then the sequence of densities  $f^{(n)}$  is  $(c_1, c_2)$ -regular and log-concave. The condition on  $S_X(f)$  assures that the eigenvalues of the covariance matrices  $\mathbf{R}_X^{(n)}$  of  $X^n$  are bounded away from zero for any  $n$ . In this case we can find a  $c_1$  such that (12) is fulfilled for all  $n \in \mathbb{N}$ . In this regard, consider that

$$\begin{aligned} \|\nabla \log f^{(n)}(x^n)\| &= \|(\mathbf{R}_X^{(n)})^{-1} x^n\| \\ &\leq \|(\mathbf{R}_X^{(n)})^{-1}\| \|x^n\| \\ &= \lambda_{\min}^{-1} \|x^n\|. \end{aligned} \quad (41)$$

$$(42)$$

where (41) follows with [11, Theorem 5.6.2] with  $\|(\mathbf{R}_X^{(n)})^{-1}\|$  being the spectral norm of  $(\mathbf{R}_X^{(n)})^{-1}$  [11, p. 295]. Thus,  $\lambda_{\min}$  is the minimal eigenvalue of  $\mathbf{R}_X^{(n)}$ , which as stated above is bounded away from zero for all  $n$ . However, for a multivariate Gaussian density it is obvious that  $D(f^{(n)}) = 0$  and, thus, the application of Theorem 2 is not necessary.

#### ACKNOWLEDGEMENT

The author thanks Heinrich Meyr for proofreading and discussing the paper and Martin Mittelbach and Tsachy Weissman for helpful discussions. Moreover, he thanks the reviewers for useful comments.

#### REFERENCES

- [1] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd edition. New York: Wiley & Sons, 2006.
- [2] U. Grenander and G. Szegő, *Toeplitz Forms and Their Applications*. Berkeley, CA, U.S.A.: Univ. Calif. Press, 1958.
- [3] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.
- [4] S. Bobkov and M. Madiman, "The entropy per coordinate of a random vector is highly constrained under convexity conditions," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4940–4954, 2011.
- [5] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *arXiv preprint arXiv:1504.04419*, 2015.
- [6] J. Bourgain, "On high dimensional maximal functions associated to convex bodies," *American Journal of Mathematics*, vol. 108, no. 6, pp. 1467–1476, 1986.
- [7] V. D. Milman and A. Pajor, "Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed  $n$ -dimensional space," in *Geometric Aspects of Functional Analysis*, ser. Lecture Notes in Mathematics. Springer, 1989, vol. 1376, pp. 64–104.
- [8] K. Ball, "Logarithmic concave functions and sections of convex sets in  $\mathbb{R}^n$ ," *Studia Math.*, vol. 88, no. 1, pp. 69–84, 1988.
- [9] B. Klartag and G. Kozma, "On the hyperplane conjecture for random convex sets," *Israel Journal of Mathematics*, vol. 170, no. 1, pp. 253–268, 2009.
- [10] M. Cule, R. Samworth *et al.*, "Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density," *Electronic Journal of Statistics*, vol. 4, pp. 254–270, 2010.
- [11] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

<sup>4</sup>Interpreting the log as the natural logarithm.