# Distributed Simulation of Continuous Random Variables

Cheuk Ting Li and Abbas El Gamal
Department of Electrical Engineering, Stanford University
Email: ctli@stanford.edu, abbas@ee.stanford.edu



**Abstract**

We establish the first known upper bound on the exact and Wyner's common information of $n$ continuous random variables in terms of the dual total correlation between them (which is a generalization of mutual information). In particular, we show that when the pdf of the random variables is log-concave, there is a constant gap of $n^2 \log e + 9n \log n$ between this upper bound and the dual total correlation lower bound that does not depend on the distribution. The upper bound is obtained using a computationally efficient dyadic decomposition scheme for constructing a discrete common randomness variable $W$ from which the $n$ random variables can be simulated in a distributed manner. We then bound the entropy of $W$ using a new measure, which we refer to as the erosion entropy.

**Index Terms**

Exact common information, Wyner's common information, log-concave distribution, dual total correlation, channel synthesis.

## I. Introduction

This paper is motivated by the following question. Alice would like to simulate a random variable $X_1$ and Bob would like to simulate another random variable $X_2$ such that $(X_1, X_2)$ are jointly Gaussian with a prescribed mean and covariance matrix. Can these two random variables be simulated in a distributed manner with only a finite amount of common randomness between Alice and Bob?

We answer this question in the affirmative for $n$ continuous random variables under certain conditions on their joint pdf, including when it is log-concave such as Gaussian.

The general distributed randomness generation setup we consider is depicted in Figure 1. There are $n$ agents (e.g., processors in a computer cluster or nodes in a communication network) that have access to *common randomness* $W \in \{0,1\}^*$. Agent $i \in [1:n]$ wishes to simulate the random variable $X_i$ using $W$ and its local randomness, which is independent of $W$ and the local randomness at other agents, such that $X^n = (X_1, \ldots, X_n)$ follows a prescribed distribution *exactly*. The distributed randomness simulation problem is to find the common randomness $W^*$ with the minimum average description length $R^*$, referred to in [1] as the *exact common information* between $X^n$, and the scheme that achieves this exact common information.

Since $W$ can be represented by an optimal prefix-free code, e.g., a Huffman code or the code in [2] if the alphabet is infinite, the average description length $R^*$ can be upper bounded as $H(\tilde{W}) \le R^* < H(\tilde{W}) + 1$, where $\tilde{W}$ minimizes $H(W)$. Hence in this paper we will focus on investigating $W$ that minimizes $H(W)$ instead of $R^*$.

The above setting was introduced in [1] for two discrete random variables and the minimum entropy of $W$, referred to as the *common entropy*, is given by

$$G(X_1; X_2) = \min_{W:\, X_1 \perp\!\!\!\perp X_2 | W} H(W). \tag{1}$$

Computing $G(X_1; X_2)$, even for moderate size random variable alphabets, can be computationally difficult since it involves minimizing a concave function over a non-convex set; see [1] for some cases where $G$ can be computed and for some properties that can be exploited to compute it. Hence the main difficulty in constructing a scheme that achieves $G$ (within 1-bit) for a given $(X_1, X_2)$ distribution is finding the optimal common randomness $W$ that achieves it.

It can be readily shown that

$$I(X_1; X_2) \le J(X_1; X_2) \le G(X_1; X_2) \le \min\{H(X_1), H(X_2)\}, \tag{2}$$

where

$$J(X_1; X_2) = \min_{W:\, X_1 \perp\!\!\!\perp X_2 | W} I(W; X_1, X_2) \tag{3}$$

is Wyner's common information [3], which is the minimum amount of common randomness rate needed to generate the discrete memoryless source (DMS) $(X_1, X_2)$ with asymptotically vanishing total variation. The notion of exact common information

The work of C. T. Li was partially supported by a Hong Kong Alumni Stanford Graduate Fellowship. This paper was presented in part at the IEEE International Symposium on Information Theory, Barcelona, July 2016.

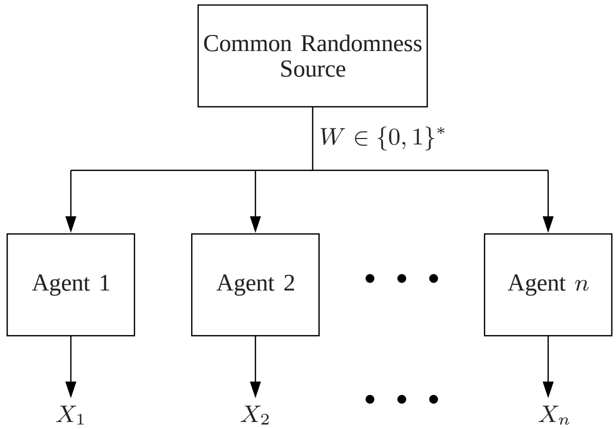


Figure 1. Distributed randomness generation setting. Common randomness $W$ is broadcast to $n$ agents and agent $i \in [1:n]$ generates $X_i$ using $W$ and it local randomness.

rate $\bar{G}(X_1;X_2)$, which is the minimum amount of common randomness rate needed to generate the DMS $(X_1, X_2)$ exactly, was also introduced in [1]. It was shown that: (i) in general $J \le \bar{G} \le G$, (ii) $G$ can be strictly larger than $\bar{G}$, and (iii) in some cases $\bar{G}(X_1;X_2) = J(X_1;X_2)$. It is not known, however, if $\bar{G}(X_1;X_2) = J(X_1;X_2)$ in general. As such, we do not consider $\bar{G}$ further in this paper.

The above results can be extended to $n$ random variables. First, it is straightforward to extend the common entropy in (1) to $n$ general random variables to obtain

$$G(X_1;\ldots;X_n) = \min_{W:\, X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_n | W} H(W). \tag{4}$$

Second, in [4], [5] Wyner's common information was extended to $n$ discrete random variables to obtain

$$J(X_1;\ldots;X_n) = \min_{W:\, X_1 \perp\!\!\!\perp X_2 \perp\!\!\!\perp \ldots \perp\!\!\!\perp X_n | W} I(W; X_1,\ldots,X_n).$$

The operational implications of Wyner's common information for two continuous random variables was studied in [6]. Wyner's common information between scalar jointly Gaussian random variables is computed in [6], and the result is extended to Gaussian vectors in [7], and to outputs of additive Gaussian channels in [8].

We can also generalize the bounds in (2) to $n$ random variables to obtain

$$I_{\mathrm{D}}(X_1;X_2;\ldots;X_n) \le J(X_1;X_2;\ldots;X_n) \le G(X_1;X_2;\ldots;X_n) \le \min_i H(X_1,\ldots,X_{i-1},X_{i+1},\ldots X_n), \tag{5}$$

where $I_{\mathrm{D}}$ is the *dual total correlation* [9] — a generalization of mutual information defined as

$$I_{\mathrm{D}}(X_1;X_2;\ldots;X_n) = H(X_1,\ldots,X_n) - \sum_{i=1}^{n} H(X_i | X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n).$$

Details of the derivation of the lower bound in (5) can be found in Appendix A. Note that the lower bound on $J$ continues to hold for continuous random variables after replacing the entropy $H$ in the definition of $I_{\mathrm{D}}$ with the differential entropy $h$. There is no corresponding upper bound to (5) for continuous random variables, however, and it is unclear under what conditions $G$ is finite.

In this paper we devise a computationally efficient scheme for constructing a common randomness variable $W$ for distributed simulation of $n$ continuous random variables and establish upper bounds on its entropy, which in turn provide upper bounds on $G$. In particular we establish the following upper bound on $G$ when the pdf of $X^n$ is *log-concave*

$$I_{\mathrm{D}} \le J \le G \le I_{\mathrm{D}} + n^2 \log e + 9n \log n.$$

For $n = 2$, this bound reduces to

$$I(X_1;X_2) \le J(X_1;X_2) \le G(X_1;X_2) \le I(X_1;X_2) + 24.$$

Applying this result to two jointly Gaussian random variables shows that only a finite amount of common randomness is needed for their distributed simulation. The above upper bound also provides an upper bound on Wyner's common information between $n$ continuous random variables with log-concave pdf. This is an interesting result since computing Wyner's common information for $n$ continuous random variables is very difficult in general and there is no previously known upper bound on it.

Our distributed randomness simulation scheme uses a dyadic decomposition procedure to construct the common randomness variable $W$. For $X^n$ uniformly distributed over a set $A$, our decomposition method partitions $A$ into hypercubes. The common randomness $W$ is defined as the position and size of the hypercube that contains $X_1, \ldots, X_n$ represented via an optimal prefix-free code. Conditioned on $W$, the random variables $X^n$ are independent and uniformly distributed over line segments, which when combined with local randomness facilitates distributed exact simulation. This scheme is extended to non-uniformly distributed $X^n$ by performing the same dyadic decomposition on the positive part of the hypograph of the pdf of $X^n$. Since bounding $H(W)$ directly is quite difficult, we bound it using the *erosion entropy* of the set, which is a new measure that is shift invariant.

The cardinality of the random variable $W$ needed for exact distributed simulation of continuous random variables is in general infinite. By terminating the dyadic decomposition at a finite iteration, however, we show that the random variables can be approximately simulated using a fixed length code such that for log-concave pdfs, the total variation distance between the simulated and prescribed pdfs can be bounded as a function of the dual total correlation and the cardinality of $W$. This result provides an upper bound on the one-shot version of Wyner's common information with total variation constraint.

The rest of the paper is organized as follows. In Section II, we introduce the aforementioned dyadic decomposition scheme and establish an upper bound on $G$ when the random variables are uniformly distributed over an orthogonally convex set. In Section III, we extend this bound to non-uniform distributions with orthogonally concave pdf and establish our main result on log-concave pdfs. In Section IV, we establish an upper bound on the one-shot version of Wyner's common information with total variation constraint. In Appendix B, we provide details on the implementation of the coding scheme for constructing the common randomness variable.

### A. Notation

Throughout this paper, we assume that $\log$ is base 2 and the entropy $H$ is in bits. We use the notation: $[a:b] = [a,b] \cap \mathbb{Z}$ and $X_{[1:n]\setminus i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

A set $A \subseteq \mathbb{R}^n$ is said to be *orthogonally convex* if for any line $L$ parallel to one of the $n$ axes, $L \cap A$ is a connected set (empty, a point, or an interval). A function $f$ is said to be orthogonally concave if its *hypograph* $\{(x, \alpha) : x \in \mathbb{R}^n, \, \alpha \le f(x)\}$ is orthogonally convex.

We denote the $i$-th standard basis vector of $\mathbb{R}^n$ by $\mathrm{e}_i$. We denote the volume of a Lebesgue measurable set $A \subseteq \mathbb{R}^n$ by $\mathrm{V}_n(A) = \int_{\mathbb{R}^n} \mathbf{1}_A(x) dx$. If $A \subseteq B \subseteq \mathbb{R}^n$, where $B$ is an $m$-dimensional affine subspace, we denote the $m$-dimensional volume of $A$ by $\mathrm{V}_m(A) = \int_B \mathbf{1}_A(x) dx$.

We define the projection of a point $x \in \mathbb{R}^n$ as

$$\mathrm{P}_{i_1,\ldots,i_k}(x) = (x_{i_1}, \ldots, x_{i_k}) \in \mathbb{R}^k,$$

and the projection of a set $A \subseteq \mathbb{R}^n$ onto the dimensions $i_1, \ldots, i_k$ as

$$\mathrm{P}_{i_1,\ldots,i_k}(A) = \{(x_{i_1}, \ldots, x_{i_k}) : x \in A\} \subseteq \mathbb{R}^k.$$

We use the shorthand notation

$$\begin{aligned}
\mathrm{P}_{\setminus i}(A) &= \mathrm{P}_{1,2,\ldots,i-1,i+1,\ldots,n}(A), \\
\mathrm{VP}_{i_1,\ldots,i_k}(A) &= \mathrm{V}_k(\mathrm{P}_{i_1,\ldots,i_k}(A)), \\
\mathrm{VP}_{\setminus i}(A) &= \mathrm{V}_{n-1}(\mathrm{P}_{\setminus i}(A)).
\end{aligned}$$

For $A, B \subseteq \mathbb{R}^n$, $A + B$ denotes the Minkowski sum $\{a + b : \, a \in A, \, b \in B\}$, and for $x \in \mathbb{R}^n$, $A + x = \{a + x : \, a \in A\}$. For $\gamma \in \mathbb{R}$, $\gamma A = \{\gamma a : \, a \in A\}$. For $M \in \mathbb{R}^{n \times n}$, $MA = \{Ma : \, a \in A\}$. The *erosion* of the set $A$ by the set $B$ is defined as $A \ominus B = \{x \in \mathbb{R}^n : \, B + x \subseteq A\}$.

For a set $A \subseteq \mathbb{R}^n$ where $0 \in A$, the radial function $\rho_A : \mathbb{R}^n \to \mathbb{R}$ is defined as $\rho_A(x) = \sup\{\lambda \ge 0 : \, \lambda x \in A\}$.

## II. Uniform Distribution over a Set

We first define the dyadic decomposition of a set, which is the building block of our distributed randomness simulation scheme.

**Definition 1** (Dyadic decomposition)**.** For $v \in \mathbb{Z}^n$ and $k \in \mathbb{Z}$, we define the hypercube $C_{k,v} = 2^{-k}([0,1]^n + v) \subset \mathbb{R}^n$. For a set $A \subseteq \mathbb{R}^n$ with a boundary of measure zero and $k \in \mathbb{Z}$, define the set

$$D_k(A) = \left\{ v \in \mathbb{Z}^n \, : \, C_{k,v} \subseteq A \text{ and } C_{k-1,\lfloor v/2 \rfloor} \nsubseteq A \right\},$$

where $\lfloor v/2 \rfloor$ is the vector formed by the entries $\lfloor v_i/2 \rfloor$.

The *dyadic decomposition* of $A$ is the partitioning of $A$ into hypercubes $\{C_{k,v}\}$ such that $v \in D_k(A)$ and $k \in \mathbb{Z}$. Since every point $x$ in the interior of $A$ is contained in some hypercube in $A$, the interior is contained in $\cup_{k \in \mathbb{Z}, \, v \in D_k(A)} C_{k,v}$, and the set of points in $A$ not covered by the hypercubes has measure zero.

For $X^n \sim \text{Unif}(A)$, denote by $C_{K,V}$, $V \in D_K(A)$, the hypercube that contains $X^n$ and let the *dyadic decomposition random variable* $W_A = (K, V)$. Then conditioned on $W_A = (k, v)$, $X^n \sim \text{Unif}(C_{k,v})$, that is, $X_1, \ldots, X_n$ are conditionally independent given $W_A$. Hence, we can use the dyadic decomposition to perform distributed randomness simulation as follows.

1) The common randomness source generates $\tilde{x}^n$ according to a uniform pdf over $A$ and finds $w_A = (k, v)$ such that $v \in D_k(A)$ and $\tilde{x}^n \in C_{k,v}$.
2) The common randomness source represents $w_A$ by a codeword from an agreed upon optimal prefix-free code and sends it to the processors (from this point on, we will assume that $W_A$ is always represented by an optimal prefix-free code).
3) Upon receiving and recovering $w_A = (k, v)$, agent $i$ generates $X_i \sim \text{Unif}[2^{-k} v_i,\, 2^{-k}(v_i + 1)]$.

The implementation details of this scheme are provided in Appendix B.

To illustrate the above dyadic decomposition scheme, consider the following.

**Example 1.** Let $X^n \sim \text{Unif}(A)$, where $A$ is an ellipse, i.e., $A = \left\{x \in \mathbb{R}^2 : x^T K x < 1\right\}$ and $K$ is a positive definite matrix. Figure 2 illustrates the dyadic decomposition for $K = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}$, and the codewords assigned to the larger squares. Figure 3 plots the pmf of the constructed $W_A$ for the same $K$ in a log-log scale ($w_i$ is the $i$-th most probable $w$). As can be seen, the tail of the pmf of $W_A$ roughly follows a straight line, that is, the pmf of $W_A$ follows a power law tail $p(w_i) \propto i^{-\alpha}$ with $\alpha \approx 2$). Hence $H(W_A)$ is finite.

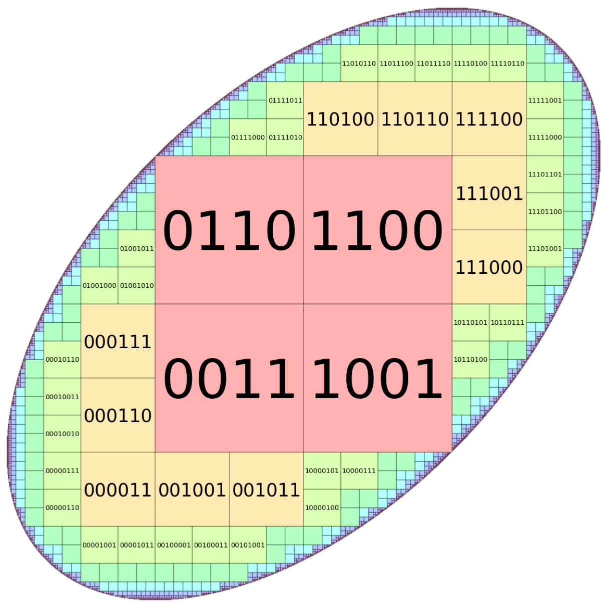


Figure 2. Dyadic decomposition of the uniform pdf over the ellipse in Example 1.

The entropy of $W_A$ can be expressed as

$$\begin{aligned} H(W_A) &= -\sum_{k \in \mathbb{Z}} \sum_{v \in D_k(A)} \mathsf{P}\left\{W_A = (k, v)\right\} \log \mathsf{P}\left\{W_A = (k, v)\right\} \\ &= -\sum_{k \in \mathbb{Z}} \sum_{v \in D_k(A)} \frac{2^{-nk}}{\mathrm{V}_n(A)} \log \frac{2^{-nk}}{\mathrm{V}_n(A)} \\ &= \sum_{k \in \mathbb{Z}} \frac{2^{-nk} \left|D_k(A)\right|}{\mathrm{V}_n(A)} \left(nk + \log \mathrm{V}_n(A)\right) \\ &= \log \mathrm{V}_n(A) + \frac{1}{\mathrm{V}_n(A)} \sum_{k \in \mathbb{Z}} nk 2^{-nk} \left|D_k(A)\right|. \end{aligned}$$

Since $X_1, \ldots, X_n$ are conditionally independent given $W_A$, we have

$$G(X_1; \ldots; X_n) \le H(W_A),$$

and since an optimal prefix-free code is used to represent the hypercubes resulting from the dyadic decomposition, the average code length is upper-bounded by $H(W_A) + 1$.

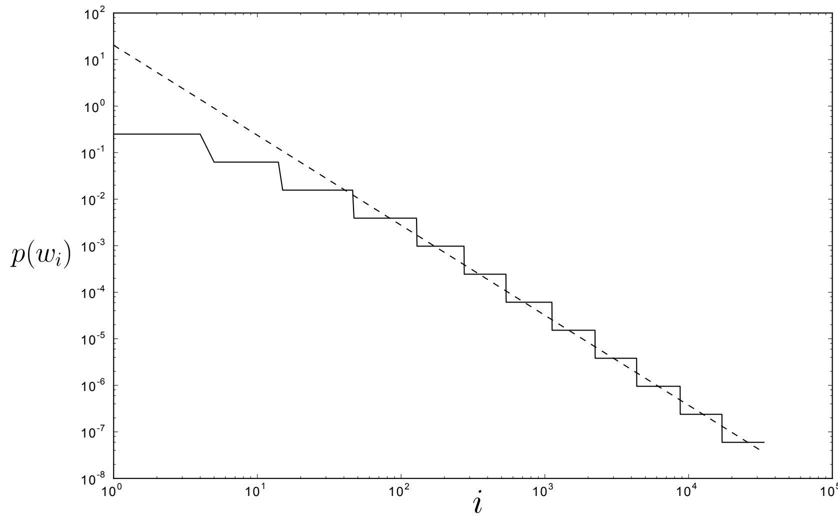


Figure 3. The pmf of $W$ for the dyadic decomposition of the uniform pdf over the ellipse in Example 1.

The exact value of $H(W_A)$ is very difficult to compute in general. It also varies as we shift and scale $A$, which should not matter in the context of distributed randomness simulation, since we can simply shift and scale the random variables before applying the scheme. Hence, we bound $H(W_A)$ using the following quantity, which is shift invariant and easier to analyze.

**Definition 2** (Erosion entropy)**.** The *erosion entropy* of a set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$ by a convex set $B \subseteq \mathbb{R}^n$ is defined as

$$h_{\ominus B}(A) = \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \geq 0\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-t} B\right)}{\mathrm{V}_n(A)} \right) dt,$$

where $A \ominus B = \{x \in \mathbb{R}^n : B + x \subseteq A\}$ is the erosion of $A$ by $B$.

The erosion entropy roughly measures the ratio of the surface area to the volume of the set $A$. To see this, assume that $0 \in B$ and let $X^n \sim \mathrm{Unif}(A)$ and $\Phi = \sup\{\phi : \phi B + X \subseteq A\}$, then we can rewrite the erosion entropy as $h_{\ominus B}(A) = \mathsf{E}(-\log(\Phi))$. If we further assume that $B = [0,1]^n$, then the erosion entropy is the expectation of the negative logarithm of the side length of the largest hypercube centered at a randomly distributed point in $A$. Hence, a large erosion entropy means that the largest hypercube centered at a random point in $A$ is small, which suggests that $A$ has a large surface area to volume ratio. We now state some basic properties of the erosion entropy.

**Proposition 1.** *For a set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$, convex sets $B, B_1, B_2 \subseteq \mathbb{R}^n$, and nonsingular $M \in \mathbb{R}^{n \times n}$, the erosion entropy satisfies the following.*

1) *Monotonicity. If $B_1 \subseteq B_2$, then $h_{\ominus B_1}(A) \leq h_{\ominus B_2}(A)$.*
2) *Scaling. $h_{\ominus \beta B}(\alpha A) = h_{\ominus B}(A) + \log(\beta/\alpha)$.*
3) *Linear transformation. $h_{\ominus MB}(MA) = h_{\ominus B}(A)$.*
4) *Union. If $A_1, \ldots, A_k \subseteq \mathbb{R}^n$ are disjoint, then*

$$h_{\ominus B}\left( \bigcup_{i=1}^{k} A_i \right) \leq \sum_{i=1}^{k} \frac{\mathrm{V}_n(A_i)}{\mathrm{V}_n(\bigcup_j A_j)} \cdot h_{\ominus B}(A_i).$$

   *Equality holds when the closures of $A_1, \ldots, A_k \subseteq \mathbb{R}^n$ are disjoint.*

5) *Reduction to differential entropy. If $X^n \sim \mathrm{Unif}(A)$, and $A \cap L$ is connected for any line $L$ parallel to the $n$-th axis, then*

$$h_{\ominus \{0\}^{n-1} \times [0,1]}(A) = h(X^{n-1}) + \log e.$$

   *As a result, for general continuous random variables $X^n$ with pdf $f$,*

$$h_{\ominus \{0\}^{n} \times [0,1]}(\mathrm{hyp}_+ f) = h(X^n) + \log e,$$

   *where $\mathrm{hyp}_+ f = \{(x, \alpha) : x \in \mathbb{R}^n,\, 0 < \alpha < f(x)\} \subseteq \mathbb{R}^{n+1}$.*

The proofs of these properties are given in Appendix C.

In the following proposition we show that $H(W_A)$ can be upper bounded using the erosion entropy. Moreover we show that the erosion entropy is the average of the dyadic decomposition entropy under random shifting and scaling.

**Proposition 2.** *For a set $A \subseteq \mathbb{R}^n$ with a boundary of measure zero, we have*

$$H(W_A) \le \log \mathrm{V}_n(A) + nh_{\ominus[0,1]^n}(A) + 2n.$$

*Moreover, for any $T \in \mathbb{Z}$, $T > (1/n)\log \mathrm{V}_n(A) + 1$, when $U^n \sim \mathrm{Unif}[0, 2^T]$ i.i.d., $\Theta \sim \mathrm{Unif}[0,1]$ independent of $U^n$ and $\Lambda = 2^\Theta$, we have*

$$\mathsf{E}\left[H(W_{\Lambda A+U})\right] = \log \mathrm{V}_n(A) + nh_{\ominus[0,1]^n}(A).$$

The proof of this proposition is given in Appendix D.

For an orthogonally convex set $A$, the entropy of the dyadic decomposition can be bounded by the volume of $A$ and the volume of its projection as follows.

**Theorem 1.** *Let $A \subseteq \mathbb{R}^n$ be an orthogonally convex set with $0 < \mathrm{V}_n(A) < \infty$ and $X^n \sim \mathrm{Unif}(A)$, then*

$$H(W_A) \le n\log\left(\sum_{i=1}^{n} \mathrm{VP}_{\backslash i}(A)\right) - (n-1)\log \mathrm{V}_n(A) + (2+\log e)n.$$

*Moreover, by applying the randomization in Proposition 2, we obtain*

$$G(X_1;\cdots;X_n) \le n\log\left(\sum_{i=1}^{n} \mathrm{VP}_{\backslash i}(A)\right) - (n-1)\log \mathrm{V}_n(A) + n\log e. \tag{6}$$

If the set $A$ is not orthogonally convex but can be partitioned into orthogonally convex sets, then the property of erosion entropy of union of sets in Proposition 1 can be used to bound $H(W_A)$. We now use the above theorem to upper bound $G$ for the uniform pdf over an ellipse example.

**Example 1** (continued). Applying (6) to the uniform pdf over the ellipse $A = \left\{x \in \mathbb{R}^2 : x^T K x \le 1\right\}$, we obtain

$$\begin{aligned} H(W_A) &\le 2\log\left(\sum_{i=1}^{2} \mathrm{VP}_{\backslash i}(A)\right) - \log \mathrm{V}_2(A) + 4 + 2\log e \\ &= 2\log\left(2\sqrt{\frac{K_{11}}{\det K}} + 2\sqrt{\frac{K_{22}}{\det K}}\right) - \log\left(\pi\sqrt{\frac{1}{\det K}}\right) + 4 + 2\log e \\ &= \log\left(\pi^{-1}\frac{\left(\sqrt{K_{11}}+\sqrt{K_{22}}\right)^2}{\sqrt{\det K}}\right) + 6 + 2\log e. \end{aligned}$$

Comparing this to the mutual information for the uniform pdf over the ellipse, we have

$$I(X_1;X_2) = \log\left(\pi e^{-1}\sqrt{\frac{K_{11}K_{22}}{\det K}}\right).$$

Note that the gap between $H(W_A)$ and $I(X_1;X_2)$ depends on the ratio between $\left(\sqrt{K_{11}}+\sqrt{K_{22}}\right)^2$ and $\sqrt{K_{11}K_{22}}$, which becomes very large when $K_{11} \gg K_{22}$. For example, if $K = \mathrm{diag}(10000, 1)$, then $\sqrt{K_{11}K_{22}} = 100$ and $I(X_1;X_2) \approx 0.21$. On the other hand, $\left(\sqrt{K_{11}}+\sqrt{K_{22}}\right)^2 = 10201$ and the bound on $H(W_A)$ is 13.02. In the next section we show that this gap can be reduced and bounded by a constant by appropriately scaling $A$.

To prove Theorem 1, we need the following lemma, which bounds the volume of the erosion of $A$ by a hypercube.

**Lemma 1.** *For any orthogonally convex set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$ and $\gamma \ge 0$, the set $A \ominus \left([0,\gamma]\times\{0\}^{n-1}\right) \subseteq A$ is orthogonally convex, and*

$$\mathrm{V}_n\left(A \ominus \left([0,\gamma]\times\{0\}^{n-1}\right)\right) \ge \mathrm{V}_n(A) - \int_{\mathrm{P}_{[2:n]}(A)} \min\left\{\gamma\,,\ \mathrm{V}_1\left(A \cap (\mathrm{span}(\mathrm{e}_1)+x)\right)\right\} dx_2^n,$$

*where $\mathrm{span}(\mathrm{e}_1) + x = \{(\alpha, x_2, x_3, \ldots, x_n) : \alpha \in \mathbb{R}\} \subseteq \mathbb{R}^n$. As a result,*

$$\mathrm{V}_n\left(A \ominus [0,\gamma]^n\right) \ge \mathrm{V}_n(A) - \sum_{i=1}^{n}\int_{\mathrm{P}_{\backslash i}(A)} \min\left\{\gamma\,,\ \mathrm{V}_1\left(A \cap (\mathrm{span}(\mathrm{e}_i)+x)\right)\right\} dx_{[1:n]\backslash i}.$$

*Proof:* We first prove the following result on the erosion of $A$ by a line segment: for any orthogonally convex set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$ and $\gamma \ge 0$, the set $A \ominus \left([0,\gamma]\times\{0\}^{n-1}\right) \subseteq A$ is orthogonally convex, and

$$\mathrm{V}_n\left(A \ominus \left([0,\gamma]\times\{0\}^{n-1}\right)\right) \ge \mathrm{V}_n(A) - \int_{\mathrm{P}_{[2:n]}(A)} \min\left\{\gamma\,,\ \mathrm{V}_1\left(A \cap (\mathrm{span}(\mathrm{e}_1)+x)\right)\right\} dx.$$

Note that

$$\begin{aligned} A \ominus \left([0,\gamma]\times\{0\}^{n-1}\right) &= \{x:\, x+\alpha \mathrm{e}_1 \in A \text{ for all } \alpha\in[0,\gamma]\} \\ &= \{x:\, x\in A-\alpha \mathrm{e}_1 \text{ for all } \alpha\in[0,\gamma]\} \\ &= \bigcap_{\alpha\in[0,\gamma]} (A-\alpha \mathrm{e}_1) \end{aligned}$$

is the intersection of orthogonally convex sets, and therefore is orthogonally convex. Also

$$\begin{aligned} &\mathrm{V}_n(A) - \mathrm{V}_n\left(A\ominus\left([0,\gamma]\times\{0\}^{n-1}\right)\right) \\ &= \mathrm{V}_n\left\{x\in A:\, x+\alpha\mathrm{e}_1\notin A \text{ for some } \alpha\in[0,\gamma]\right\} \\ &= \int_{\mathrm{P}_{[2:n]}(A)} \mathrm{V}_1\left\{x_1\in\mathbb{R}:\, (x_1,\tilde{x}_2,\ldots,\tilde{x}_n)\in A,\, (x_1+\alpha,\, \tilde{x}_2,\ldots,\tilde{x}_n)\notin A \text{ for some } \alpha\in[0,\gamma]\right\} d\tilde{x}_2^n \\ &= \int_{\mathrm{P}_{[2:n]}(A)} \mathrm{V}_1\left\{x\in A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x}):\, x+\alpha\mathrm{e}_1\notin A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x}) \text{ for some } \alpha\in[0,\gamma]\right\} d\tilde{x}_2^n \\ &\le \int_{\mathrm{P}_{[2:n]}(A)} \min\left\{\gamma\,,\, \mathrm{V}_1\left(A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x})\right)\right\} d\tilde{x}_2^n, \end{aligned}$$

where the last inequality follows since $A\cap(\mathrm{span}(\mathrm{e}_1)+\tilde{x})$ is connected.

By repeating this result for each axis, and observing that $\int_{\mathrm{P}_{\setminus i}(A)} \min\left\{\gamma\,,\, \mathrm{V}_1\left(A\cap(\mathrm{span}(e_i)+x)\right)\right\} dx_{[1:n]\setminus i}$ cannot increase when $A$ is replaced with an orthogonally convex subset of $A$, we obtain the second bound. ∎

We are now ready to prove Theorem 1.

*Proof of Theorem 1:* By Proposition 2, the theorem can be proved by bounding $h_{\ominus[0,1]^n}(A)$. Note that by Lemma 1,

$$\begin{aligned} h_{\ominus[0,1]^n}(A) &= \int_{-\infty}^{\infty}\left(\mathbf{1}\{t\ge 0\} - \frac{\mathrm{V}_n\left(A\ominus[0,\,2^{-t}]^n\right)}{\mathrm{V}_n(A)}\right)dt \\ &\le \int_{-\infty}^{\infty}\left(\mathbf{1}\{t\ge 0\} - \frac{1}{\mathrm{V}_n(A)}\max\left(0,\, \mathrm{V}_n(A) - \sum_{i=1}^{n}\int_{\mathrm{P}_{\setminus i}(A)} \min\left\{2^{-t}\,,\, \mathrm{V}_1\left(A\cap(\mathrm{span}(\mathrm{e}_i)+x)\right)\right\}dx_{[1:n]\setminus i}\right)\right)dt \\ &\le \int_{-\infty}^{\infty}\left(\mathbf{1}\{t\ge 0\} - \max\left(0,\, 1-\frac{1}{\mathrm{V}_n(A)}\sum_{i=1}^{n}\int_{\mathrm{P}_{\setminus i}(A)} 2^{-t}dx_{[1:n]\setminus i}\right)\right)dt \\ &= \int_{-\infty}^{\infty}\left(\mathbf{1}\{t\ge 0\} - \max\left(0,\, 1-2^{-t}\frac{\sum_{i=1}^{n}\mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right)\right)dt \\ &= \log\left(\frac{\sum_{i=1}^{n}\mathrm{VP}_{\setminus i}(A)}{\mathrm{V}_n(A)}\right) + \log e. \end{aligned}$$

For the second result, note that the randomization in Proposition 2 does not affect the right hand side of Theorem 1, which completes the proof of the theorem. ∎

### *A. Scaling*

In this section, we present a tighter bound on the common entropy between continuous random variables by first scaling $A$ along each dimension, that is, by performing a linear transformation $DA$ where $D$ is a diagonal matrix. This corresponds to scaling the random variable $X_i$ by $D_{ii}$, $i\in[1:n]$, before applying the scheme. This new bound will be in terms of the following.

**Definition 3** (Truncated differential entropy)**.** Let $X^n \sim f(x^n)$ and define its truncated differential entropy $\tilde{h}_\zeta(X^n)$ for $\zeta\in(0,1]$, as

$$\tilde{h}_\zeta(X^n) = \int_{\mathbb{R}^n} -\zeta^{-1}\min\{\xi,\, f(x)\}\log\left(\zeta^{-1}\min\{\xi,\, f(x)\}\right)dx,$$

where $\xi>0$ such that

$$\int_{\mathbb{R}^n}\min\{\xi,\, f(x)\}\,dx = \zeta.$$

And define

$$\tilde{h}_0(X^n) = \lim_{\zeta\to 0}\tilde{h}_\zeta(X^n) = \log \mathrm{V}_n\{x:\, f(x)>0\}\,.$$

Note that $\tilde{h}_\zeta(X^n)$ is decreasing in $\zeta$ from $\tilde{h}_0(X^n)$ (the entropy of the uniform pdf on the support of $X^n$) to $\tilde{h}_1(X^n) = h(X^n)$.

We now state the main result of this section, which shows that the gap between $H(W_{DA})$ and $I_{\mathrm{D}}$ depends on how close $\tilde{h}_{1/n}(X_{[1:n]\setminus i})$ and $h(X_{[1:n]\setminus i})$, $i \in [1:n]$, are to each other.

**Theorem 2.** *For any orthogonally convex set $A \subseteq \mathbb{R}^n$ with $0 < \mathrm{V}_n(A) < \infty$, there exists a diagonal matrix $D \in \mathbb{R}^{n\times n}$ with positive diagonal entries such that the entropy of the dyadic decomposition of $DA = \{Dx : x \in A\}$ is bounded by*

$$H(W_{DA}) \le \sum_{i=1}^{n} \tilde{h}_{1/n}(X_{[1:n]\setminus i}) - (n-1)\log \mathrm{V}_n(A) + n\log n + (2+\log e)n.$$

*Equivalently, when $X^n \sim \mathrm{Unif}(A)$,*

$$H(W_{DA}) \le I_{\mathrm{D}}(X_1;\cdots;X_n) + \sum_{i=1}^{n}\left(\tilde{h}_{1/n}(X_{[1:n]\setminus i}) - h(X_{[1:n]\setminus i})\right) + n\log n + (2+\log e)n.$$

*Moreover, by applying the randomization in Proposition 2, we obtain*

$$I_{\mathrm{D}} \le J \le G \le I_{\mathrm{D}} + \sum_{i=1}^{n}\left(\tilde{h}_{1/n}(X_{[1:n]\setminus i}) - h(X_{[1:n]\setminus i})\right) + n\log n + n\log e.$$

The proof of this theorem and a method for finding find $D$ are given in Appendix E.

We illustrate the above bound in the following.

**Example 1** (continued). Applying Theorem 2 to the uniform pdf over the ellipse $A = \left\{x \in \mathbb{R}^2 : x^T K x \le 1\right\}$, we have

$$\begin{aligned} H(W_{DA}) &\le \sum_{i=1}^{2} \tilde{h}_{1/2}(X_{[1:2]\setminus i}) - \log \mathrm{V}_2(A) + 6 + 2\log e \\ &\le \sum_{i=1}^{2} \log\left(\mathrm{VP}_{\setminus i}(A)\right) - \log \mathrm{V}_2(A) + 6 + 2\log e \\ &= \log\left(2\sqrt{\frac{K_{11}}{\det K}}\right) + \log\left(2\sqrt{\frac{K_{22}}{\det K}}\right) - \log\left(\pi\sqrt{\frac{1}{\det K}}\right) + 6 + 2\log e \\ &= \log\left(\pi^{-1}\sqrt{\frac{K_{11}K_{22}}{\det K}}\right) + 8 + 2\log e. \end{aligned}$$

In comparison, the mutual information is

$$I(X_1;X_2) = \log\left(\pi e^{-1}\sqrt{\frac{K_{11}K_{22}}{\det K}}\right),$$

and the gap between $H(W_{DA})$ and $I(X_1;X_2)$ is bounded by a constant. Figure 4 plots the values of $H(W_{DA})$ (calculated by finding all squares in the dyadic decomposition with side length at least $2^{-11}$, which yields a precise estimate), the upper bound in Theorem 2, and $I(X_1;X_2)$ for $K = \frac{1}{1-t^2}\begin{bmatrix} 1 & -t \\ -t & 1 \end{bmatrix}$, $t \in [0,1]$.

## III. Nonuniform Distributions

In this section, we extend our results to the case in which the pdf of $X^n$ is not necessarily uniform. Let $X^n \sim f(x^n)$ and let the support of $f$ be $A$. We add a random variable $Z$ such that $(X_1,\ldots,X_n,Z) \sim \mathrm{Unif}(\mathrm{hyp}_+ f)$, where $\mathrm{hyp}_+ f$ is the *positive strict hypograph* defined as

$$\mathrm{hyp}_+ f = \{(x,\alpha) : x \in \mathbb{R}^n,\, 0 < \alpha < f(x)\} \subseteq \mathbb{R}^{n+1}.$$

Note that the marginal pdf of $X^n$ is $f$. Assuming that $\mathrm{hyp}_+ f$ is orthogonally convex, i.e., $f$ is orthogonally concave, we can apply the results for the uniform pdf case in Section II. To illustrate this extension, consider the following.

**Example 2.** Let $(X_1,X_2)$ be zero mean Gaussian with covariance matrix $K = \begin{bmatrix} 1/8 & 1/16 \\ 1/16 & 1/8 \end{bmatrix}$. Figure 5 plots the cubes with side length $\ge 2^{-3}$ of the dyadic decomposition of the positive strict hypograph of this pdf. Note that the cubes are scaled down so as to show the ones behind them. Figure 6 plots the pmf of $W_{\mathrm{hyp}_+ f}$ in log-log scale. As in Example 1, the tail of the pmf follows a power law with power approximately $1.12$ and $H(W_{\mathrm{hyp}_+ f})$ is finite.

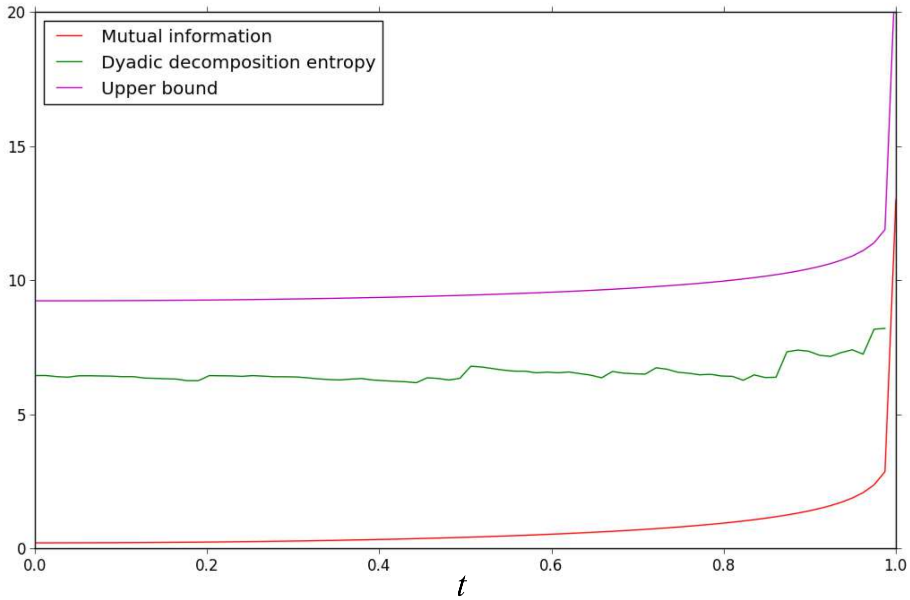


Figure 4. Plot of the entropy of the dyadic decomposition, mutual information, and the upper bound in Theorem 2 for Example 1.

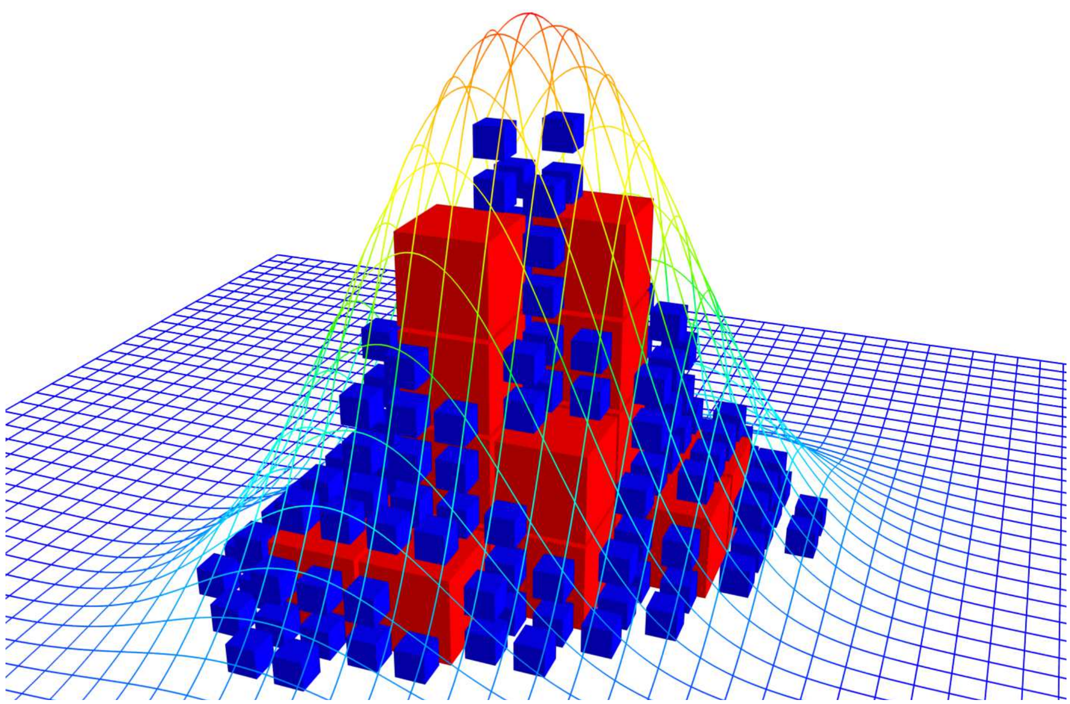

Figure 5. Dyadic decomposition of $\mathrm{hyp}_+ f$ for the Gaussian pdf $f$ in Example 2.

We now show that if $f$ is log-concave (i.e., $x \mapsto \log f(x)$ is concave), then the difference between the entropy of the dyadic decomposition and the dual total correlation is bounded by a constant that depends only on $n$.

**Theorem 3.** *If the pdf of $X^n$ is log-concave, then there exists a diagonal matrix $D \in \mathbb{R}^{(n+1)\times(n+1)}$ with positive diagonal entries such that the entropy of the dyadic decomposition of $D\mathrm{hyp}_+ f$ satisfies*

$$\begin{aligned} H(W_{D\mathrm{hyp}_+ f}) &\le I_{\mathrm{D}}(X_1; \cdots ; X_n) + n^2 \log e + n\left(\log n + \log(n+1) + e + 2\log e + 2\right) + 2 + \log e \\ &\le I_{\mathrm{D}}(X_1; \cdots ; X_n) + n^2 \log e + 12 n \log n. \end{aligned}$$

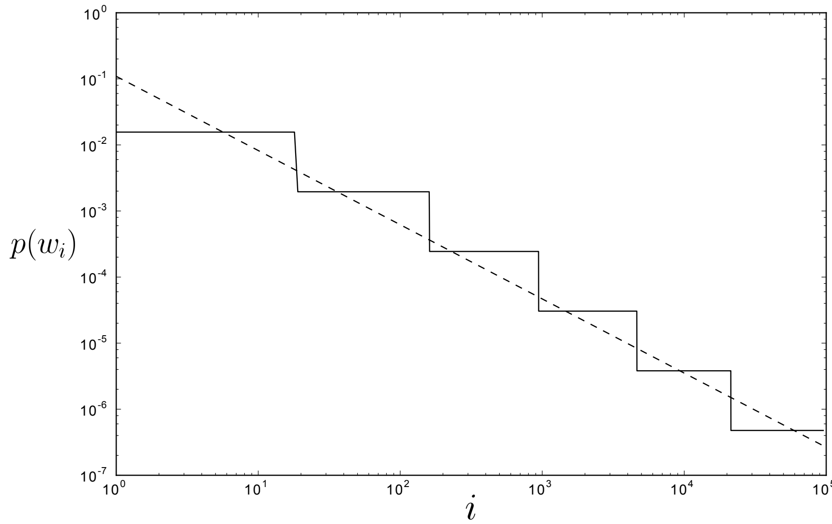


Figure 6. The pmf of $W$ for the dyadic decomposition of the hypograph of the Gaussian pdf in Example 2.

*Moreover, by applying the randomization in Proposition 2, we obtain*

$$I_{\mathrm{D}} \leq J \leq G \leq I_{\mathrm{D}} + n^2 \log e + 9n \log n.$$

To prove Theorem 3, we first state a result in [10], which bounds the differential entropy of a log-concave pdf by its maximum density.

**Lemma 2.** *For any log-concave pdf $f$ of $X^n$,*

$$-\log\Big(\sup_{x\in\mathbb{R}^n} f(x)\Big) \leq h(X^n) \leq -\log\Big(\sup_{x\in\mathbb{R}^n} f(x)\Big) + n \log e.$$

We can generalize Lemma 2 to bound the conditional differential entropy as follows.

**Lemma 3.** *For any log-concave pdf $f$ of $X^n$, and $1 \leq m \leq n$,*

$$h(X_{m+1}^n \mid X^m) - \left(-\log\Big(\int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m,\, \tilde{x}_{m+1}^n) dx^m\Big)\right) \in \left[0,\, n\log e - \log\left(m \cdot \mathrm{B}(m,\, n-m+1)\right)\right],$$

*where*

$$\mathrm{B}(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$$

*is the beta function.*

The proof of this lemma is given in Appendix F.

Next, we establish a bound on the difference between the differential entropy and the truncated differential entropy.

**Lemma 4.** *For any log-concave pdf $f$,*

$$\tilde{h}_\zeta(X^n) - h(X^n) \leq \log\zeta + \nu + n\log e,$$

*where $\nu \geq 0$ satisfies*

$$\Gamma(n+1,\, \nu) = \zeta \cdot \Gamma(n+1),$$

*and $\Gamma(n,z) = \int_z^\infty t^{n-1}e^{-t}dt$ is the incomplete gamma function, and $\Gamma(n) = \Gamma(n,0)$ is the gamma function. Moreover, if $\zeta \geq e^{-(e-2)n}$, then*

$$\tilde{h}_\zeta(X^n) - h(X^n) \leq \log\zeta + (e + \log e)n.$$

The proof of this lemma is given in Appendix G.

We now proceed to the proof of Theorem 3.

*Proof of Theorem 3:* Let $(X^n, Z) \sim \mathrm{Unif}(\mathrm{hyp}_+ f)$. Applying Theorem 2 on $\mathrm{hyp}_+ f$, we have

$$\begin{aligned} H(W_{D\mathrm{hyp}_+ f}) &\le \tilde{h}_{1/(n+1)}(X^n) + \sum_{i=1}^{n} \tilde{h}_{1/(n+1)}(X_{[1:n]\setminus i},\, Z) + (n+1)\log(n+1) + (2+\log e)(n+1) \\ &\le \tilde{h}_{1/(n+1)}(X^n) + \sum_{i=1}^{n} \log\left(\mathrm{VP}_{\setminus i}(\mathrm{hyp}_+ f)\right) + (n+1)\log(n+1) + (2+\log e)(n+1). \end{aligned}$$

Consider the term $\tilde{h}_{1/(n+1)}(X^n)$. Since $1/(n+1) \ge e^{-(e-2)n}$, by Lemma (4) we have

$$\tilde{h}_{1/(n+1)}(X^n) - h(X^n) \le -\log(n+1) + (e+\log e)n.$$

Consider the term $\log\left(\mathrm{VP}_{\setminus n}(\mathrm{hyp}_+ f)\right)$. By Lemma (3), we have

$$\begin{aligned} &\log\left(\mathrm{VP}_{\setminus n}(\mathrm{hyp}_+ f)\right) \\ &= \log \int_{\mathbb{R}^{n-1}} \sup_{\tilde{x}_n} f(x_1^{n-1}, \tilde{x}_n)\, dx_1^{n-1} \\ &\le -h(X_n | X_1^{n-1}) + n\log e - \log\left((n-1)\cdot \mathrm{B}(n-1,\, 2)\right) \\ &= -h(X_n | X_1^{n-1}) + n\log e - \log\left((n-1)\cdot \frac{1}{n(n-1)}\right) \\ &= -h(X_n | X_1^{n-1}) + n\log e + \log n. \end{aligned}$$

Hence

$$\begin{aligned} &H(W_{D\mathrm{hyp}_+ f}) \\ &\le h(X^n) - \log(n+1) + (e+\log e)n \\ &\quad + \sum_{i=1}^{n} \left(-h(X_i | X_{[1:n]\setminus i}) + n\log e + \log n\right) + (n+1)\log(n+1) + (2+\log e)(n+1) \\ &= I_{\mathrm{D}}(X_1; \cdots; X_n) + n^2 \log e + n\left(\log n + \log(n+1) + e + 2\log e + 2\right) + 2 + \log e. \end{aligned}$$

■

Note that the result in Theorem 3 can be readily extended to mixtures of log-concave pdfs using the union property in Proposition 1. It is not possible to obtain a constant bound on the gap between $I_{\mathrm{D}}$ and $G$ for arbitrary pdfs, however. To see this, let $(\tilde{X}_1, \tilde{X}_2) \in \{1, \dots, 2^m\}^2$ be two discrete random variables with pmf $\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 0 \end{bmatrix}^{\otimes m}$, i.e., $(\tilde{X}_1, \tilde{X}_2)$ consists of $m$ i.i.d. copies of two random variables with pmf $\begin{bmatrix} 1/3 & 1/3 \\ 1/3 & 0 \end{bmatrix}$. Now let $X_1 = \tilde{X}_1 + Z_1$ and $X_2 = \tilde{X}_2 + Z_2$, where $Z_1, Z_2 \sim \mathrm{Unif}[0,1]$. Then we have $I(X_1; X_2) = I(\tilde{X}_1; \tilde{X}_2) = (\log 3 - 4/3)m$, $J(X_1; X_2) = J(\tilde{X}_1; \tilde{X}_2) = (2/3)m$, and the gap between $I$ and $J$ grows linearly in $m$. Since $G \ge J$, the gap between $I$ and $G$ grows at least linearly in $m$.

## IV. Approximate Distributed Simulation

The cardinality of $W$ for exact simulation of $n$ continuous random variables is in general infinite, hence the length of the codeword for $W$ is unbounded. We show that if the exact simulation requirement is relaxed by only requiring that the total variation distance between the distributions of the simulated and the prescribed random variables to be small, then distributed simulation is possible with a fixed length code.

We define the approximate distributed simulation problem as follows. There are $n$ agents that have access to common randomness $W \in \{0,1\}^N$. Agent $i \in [1:n]$ wishes to simulate the random variable $\tilde{X}_i$ using $W$ and its local randomness, which is independent of $W$ and local randomness at other agents, such that the total variation between the distributions of $\tilde{X}^n$ and $X^n$ is bounded as

$$d_{\mathrm{TV}}\left((\tilde{X}_1, \dots, \tilde{X}_n),\, (X_1, \dots, X_n)\right) \le \epsilon,$$

for some $\epsilon > 0$. The problem is to find the conditions under which the length-distance pair $(N, \epsilon)$ is achievable.

We can find sufficient conditions under which $(N, \epsilon)$ is achievable by terminating the dyadic decomposition scheme described in the previous sections after a finite number of iterations, that is, by discarding all hypercubes smaller than a prescribed size. The following proposition gives the length-distance pairs achievable by this truncated dyadic decomposition scheme in terms of $H(W_A)$ for uniform pdf over $A$ (or $H(W_{D\mathrm{hyp}_+ f})$ for non-uniform pdfs), which in turn can be bounded by Theorem 1, 2 or 3.

**Theorem 4.** *Let $X^n \sim \mathrm{Unif}(A)$, where $A \subseteq \mathbb{R}^n$ with a boundary of measure zero. The truncated dyadic decomposition scheme can achieve the length-distance pair $(N, \epsilon)$ if*

$$N \geq \log\left(\epsilon 2^{\epsilon^{-1} H(W_A)} + 1\right).$$

*Proof:* Let $W_A$ be the dyadic decomposition common randomness variable for $X^n \mathrm{Unif}(A)$ and define

$$\tilde{W}_A = \begin{cases} (k, v) & \text{if } k < l \\ (k_0, v_0) & \text{if } k \geq l, \end{cases}$$

where $W_A = (k, v)$, and $(k_0, v_0)$ is any hypercube with side length $> 2^{-l}$, where $l = n^{-1}(\epsilon^{-1} H(W_A) - \log \mathrm{V}_n(A))$. The truncated scheme uses $\tilde{W}_A$ as common randomness variable, and the operations performed by the agents are unchanged. Since the truncated scheme differs from the original scheme only when $\tilde{W}_A \neq W_A$, we have

$$\begin{aligned}
&d_{\mathrm{TV}}\left((\tilde{X}_1, \ldots, \tilde{X}_n), (X_1, \ldots, X_n)\right) \\
&\leq \mathsf{P}\{K \geq l\} \\
&= \mathsf{P}\{\log \mathrm{V}_n(A) + nK \geq \log \mathrm{V}_n(A) + nl\} \\
&\leq \frac{\mathsf{E}[\log \mathrm{V}_n(A) + nK]}{\log \mathrm{V}_n(A) + nl} \\
&= \frac{H(W_A)}{\log \mathrm{V}_n(A) + nl} \\
&= \epsilon.
\end{aligned}$$

It is left to bound the cardinality of $\tilde{W}_A$. Consider the probability vector $p_{\tilde{W}_A} \in \mathbb{R}^{|\tilde{W}_A|}$. Since $p_{\tilde{W}_A}(w) \geq 2^{-nl} \mathrm{V}_n^{-1}(A)$, the vector $p_{\tilde{W}_A}$ can be expressed as a convex combination of the vectors $2^{-nl} \mathrm{V}_n^{-1}(A) \mathbf{1} + \left(1 - 2^{-nl} \mathrm{V}_n^{-1}(A) |\tilde{W}_A|\right) \mathrm{e}_i$ for $i = 1, \ldots, |\tilde{W}_A|$. We have

$$\begin{aligned}
&H\left(2^{-nl} \mathrm{V}_n^{-1}(A) \mathbf{1} + \left(1 - 2^{-nl} \mathrm{V}_n^{-1}(A) |\tilde{W}_A|\right) \mathrm{e}_i\right) \\
&= 2^{-nl} \mathrm{V}_n^{-1}(A)(|\tilde{W}_A| - 1) \cdot (\log \mathrm{V}_n(A) + nl) - \left(1 - 2^{-nl} \mathrm{V}_n^{-1}(A)(|\tilde{W}_A| - 1)\right) \log\left(1 - 2^{-nl} \mathrm{V}_n^{-1}(A)(|\tilde{W}_A| - 1)\right) \\
&\geq 2^{-nl} \mathrm{V}_n^{-1}(A)(|\tilde{W}_A| - 1) \cdot (\log \mathrm{V}_n(A) + nl).
\end{aligned}$$

Since entropy is concave, $H(W_A) \geq H(\tilde{W}_A) \geq 2^{-nl} \mathrm{V}_n^{-1}(A)(|\tilde{W}_A| - 1) \cdot (\log \mathrm{V}_n(A) + nl)$. By Theorem 3,

$$\begin{aligned}
|\tilde{W}_A| &\leq \frac{1}{\log \mathrm{V}_n(A) + nl} 2^{nl} \mathrm{V}_n(A) H(W_A) + 1 \\
&= \frac{1}{\epsilon^{-1} H(W_A)} 2^{\epsilon^{-1} H(W_A)} H(W_A) + 1 \\
&= \epsilon 2^{\epsilon^{-1} H(W_A)} + 1 \\
&\leq 2^N.
\end{aligned}$$

The result follows. ∎

## V. Conclusion

We proposed a scheme for distributed simulation of continuous random variables based on dyadic decomposition. We established a bound on the entropy of the constructed common randomness in terms of the dual total correlation for the class of log concave pdfs. As a result, the gap between exact and Wyner's common information and dual total correlation can be bounded for this set of distributions.

Our results readily translate to the exact, one-shot version of the channel synthesis problem in [4], [11] without common randomness in which we wish to simulate a channel $f_{Y|X}(y|x)$ with input distribution $f_X(x)$. Given the input $X \sim f_X$, the encoder produces the codeword $W$ using a prefix-free code. Upon receiving $W$, the decoder produces the output $\hat{Y}$ such that $(X, \hat{Y}) \sim f_X f_{Y|X}$. The problem again is to find the minimum entropy of $W$. A consequence of our results is that an additive Gaussian noise channel with Gaussian input, can be exactly simulated using only a finite amount of common randomness.

We have seen in Section II-A that performing different scalings on each $X_i$ can reduce $H(W)$. More generally, applying a bijective transformation $g_i(x_i)$ to each random variable before using the dyadic decomposition scheme may help reduce $H(W)$ further. For example, applying the copula transform [12] $g_i(x) = F_{X_i}(x)$ such that $g_i(X_i) \sim \mathrm{Unif}[0, 1]$ has the benefit that when the $X_i$'s are close to independent, the pdf is close to a constant function over the unit hypercube, which is likely to result in a smaller $H(W)$.

Finally, our results readily apply to the distributed randomness generation setting in which the agents share a stream of uniformly random bits instead of a codeword generated by an active encoder. In this setting, the agents would need to agree on the number of random bits used (i.e., they recover $W$ from the stream using the an optimal prefix-free code). In this case, the optimal expected number of random bits used is between $H(\tilde{W})$ and $H(\tilde{W})+2$ (see [**?**]). Hence it is also sufficient to consider $H(W)$.

## Appendix

### A. Bounding Common Information by Dual Total Correlation

We show that $I_{\mathrm{D}}(X_1;\cdots;X_n) \le J(X_1;\cdots;X_n)$. For general random variables, the dual total correlation is defined as

$$I_{\mathrm{D}}(X_1;\cdots;X_n) = \sum_{i=1}^{n-1} I(X_i; X_{i+1}^n \mid X_1^{i-1}).$$

To prove the inequality, let $W$ be a random variable such that $X^n$ are conditionally independent given $W$, then

$$\begin{aligned} I(W;X^n) &= I(X_1;X_2^n) + I(W;X_2^n|X_1) + I(W;X_1|X_2^n) \\ &\ge I(X_1;X_2^n) + I(W;X_2^n|X_1) \\ &= I(X_1;X_2^n) + I(X_2;X_3^n|X_1) + I(W;X_3^n|X_1^2) + I(W;X_2|X_1,X_3^n) \\ &\ge I(X_1;X_2^n) + I(X_2;X_3^n|X_1) + I(W;X_3^n|X_1^2) \\ &\ \vdots \\ &\ge I(X_1;X_2^n) + I(X_2;X_3^n|X_1) + \cdots + I(X_{n-1};X_n|X_1^{n-2}). \end{aligned}$$

### B. Dyadic Decomposition Algorithm Details

We present the common randomness generation and simulation algorithms for $X^n \sim \mathrm{Unif}(A)$ using arithmetic coding. The common random source runs the generation algorithm to produce $W$ and agent $i \in [1:n]$ runs the simulation algorithm with input $W$ to produce $x_i$.

**Common randomness generation algorithm.**
Input: $A \subseteq [0,1]^n$
Output: codeword $w$

1) $v \leftarrow (0,\ldots,0)$, $k \leftarrow 0$, $\alpha \leftarrow 0$, $(\mu,\nu) \leftarrow (0,1)$, $w \leftarrow \emptyset \in \{0,1\}^*$
2) While $C_{k,v} = 2^{-k}([0,1]^n + v) \nsubseteq A$:
3) For each $\tilde{v} \in \{0,1\}^n + 2v = \{2v_1, 2v_1+1\} \times \cdots \times \{2v_n, 2v_n+1\}$:
4) $p_{\tilde{v}} \leftarrow \mathrm{V}_n(A \cap C_{k+1,\tilde{v}})/\mathrm{V}_n(A)$
5) While $[\mu,\nu] \nsubseteq [\alpha + \sum_{\hat{v}\prec\tilde{v}} p_{\hat{v}},\, \alpha + \sum_{\hat{v}\preceq\tilde{v}} p_{\hat{v}}]$ for all $\tilde{v} \in \{0,1\}^n + 2v$ ($\prec$ is lexicographical order)
6) Generate $\tilde{w} \in \{0,1\}$ uniformly at random
7) $w \leftarrow w \,\|\, \tilde{w}$ (append $\tilde{w}$ to $w$)
8) $(\mu,\nu) \leftarrow (\mu + (\nu-\mu)\tilde{w}/2,\, \mu + (\nu-\mu)(\tilde{w}+1)/2)$
9) $v \leftarrow \tilde{v}$ where $[\mu,\nu] \subseteq [\alpha + \sum_{\hat{v}\prec\tilde{v}} p_{\hat{v}},\, \alpha + \sum_{\hat{v}\preceq\tilde{v}} p_{\hat{v}}]$
10) $\alpha \leftarrow \alpha + \sum_{\hat{v}\prec\tilde{v}} p_{\hat{v}}$
11) Output $w$

**Simulation algorithm.**
Input: Agent $i$, common randomness $w$, $A \subseteq [0,1]^n$
Output: random variate $x_i$

1) $v \leftarrow (0,\ldots,0)$, $k \leftarrow 0$, $\alpha \leftarrow 0$, $(\mu,\nu) \leftarrow (0,1)$
2) While $C_{k,v} = 2^{-k}([0,1]^n + v) \nsubseteq A$:
3) For each $\tilde{v} \in \{0,1\}^n + 2v = \{2v_1, 2v_1+1\} \times \cdots \times \{2v_n, 2v_n+1\}$:
4) $p_{\tilde{v}} \leftarrow \mathrm{V}_n(A \cap C_{k+1,\tilde{v}})/\mathrm{V}_n(A)$
5) While $[\mu,\nu] \nsubseteq [\alpha + \sum_{\hat{v}\prec\tilde{v}} p_{\hat{v}},\, \alpha + \sum_{\hat{v}\preceq\tilde{v}} p_{\hat{v}}]$ for all $\tilde{v} \in \{0,1\}^n + 2v$ ($\prec$ is lexicographical order)
6) $\tilde{w} \leftarrow$first bit of $w$, discard first bit of $w$
7) $(\mu,\nu) \leftarrow (\mu + (\nu-\mu)\tilde{w}/2,\, \mu + (\nu-\mu)(\tilde{w}+1)/2)$
8) $v \leftarrow \tilde{v}$ where $[\mu,\nu] \subseteq [\alpha + \sum_{\hat{v}\prec\tilde{v}} p_{\hat{v}},\, \alpha + \sum_{\hat{v}\preceq\tilde{v}} p_{\hat{v}}]$
9) $\alpha \leftarrow \alpha + \sum_{\hat{v}\prec\tilde{v}} p_{\hat{v}}$
10) Output randomly generated $x_i$ according to $\mathrm{Unif}[2^{-k}v_i,\, 2^{-k}(v_i+1)]$

The above algorithms assume $A \subseteq [0,1]^n$. The case where $A$ is unbounded can be handled by first encoding the integer parts $\lfloor X^n \rfloor = (\lfloor X_1 \rfloor, \ldots, \lfloor X_n \rfloor)$ of $X^n \sim \mathrm{Unif}(A)$, then run the algorithms on $A \cap ([0,1]^n + \lfloor X^n \rfloor)$. The algorithms can be applied to non-uniform pdfs by letting $A$ to be $\mathrm{hyp}_+ f$ scaled according to Theorem 2.

An advantage of the generation and simulation algorithms based on arithmetic coding is that each bit of $w$ is generated uniformly, and hence can be applied to the situation where the agents share a stream of uniformly random bits [**?**].

### C. Proof of Proposition 1

The monotonicity property and the linear transformation property follow directly from the definition of erosion entropy. For the scaling property, consider

$$\begin{aligned} h_{\ominus\beta B}(\alpha A) &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(\alpha A \ominus 2^{-t}\beta B\right)}{\mathrm{V}_n(\alpha A)} \right) dt \\ &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-t+\log(\beta/\alpha)} B\right)}{\mathrm{V}_n(A)} \right) dt \\ &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge -\log(\beta/\alpha)\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-t} B\right)}{\mathrm{V}_n(A)} \right) dt \\ &\ge h_{\ominus B}(A) + \log(\beta/\alpha). \end{aligned}$$

For the union property, consider

$$\begin{aligned} h_{\ominus B}\left( \bigcup_{i=1}^{k} A_i \right) &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left((\bigcup_i A_i) \ominus 2^{-t} B\right)}{\sum_i \mathrm{V}_n(A_i)} \right) dt \\ &\le \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(\bigcup_i (A_i \ominus 2^{-t} B)\right)}{\sum_i \mathrm{V}_n(A_i)} \right) dt \\ &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \frac{\sum_i \mathrm{V}_n\left(A_i \ominus 2^{-t} B\right)}{\sum_i \mathrm{V}_n(A_i)} \right) dt \\ &= \int_{-\infty}^{\infty} \left( \sum_i \frac{\mathrm{V}_n(A_i)}{\sum_j \mathrm{V}_n(A_j)} \left( \mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(A_i \ominus 2^{-t} B\right)}{\mathrm{V}_n(A_i)} \right) \right) dt \\ &= \sum_i \frac{\mathrm{V}_n(A_i)}{\sum_j \mathrm{V}_n(A_j)} \cdot h_{\ominus B}(A_i). \end{aligned}$$

Equality holds if $(\bigcup_i A_i) \ominus 2^{-t} B = \bigcup_i (A_i \ominus 2^{-t} B)$, which is true when the closures of $A_1, \ldots, A_k$ are disjoint.

For the reduction to differential entropy property, let $X^n \sim \mathrm{Unif}(A)$, and $A \cap L$ is connected for any line $L$ parallel to the $n$-th axis, then

$$\begin{aligned} h_{\ominus \{0\}^{n-1} \times [0,1]}(A) &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \frac{\mathrm{V}_n\left(A \ominus \left(\{0\}^{n-1} \times [0, 2^{-t}]\right)\right)}{\mathrm{V}_n(A)} \right) dt \\ &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \int_{\mathbb{R}^{n-1}} \frac{\mathrm{V}_1\left(\left(A \ominus \{0\}^{n-1} \times [0, 2^{-t}]\right) \cap \left(\{x_1^{n-1}\} \times \mathbb{R}\right)\right)}{\mathrm{V}_n(A)} dx_1^{n-1} \right) dt \\ &= \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \int_{\mathbb{R}^{n-1}} \max\left\{ f_{X_1^{n-1}}(x_1^{n-1}) - 2^{-t},\, 0 \right\} dx_1^{n-1} \right) dt \\ &= \int_{\mathbb{R}^{n-1}} f_{X_1^{n-1}}(x_1^{n-1}) \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \max\left\{ 1 - 2^{-t}/f_{X_1^{n-1}}(x_1^{n-1}),\, 0 \right\} \right) dt dx_1^{n-1} \\ &= \int_{\mathbb{R}^{n-1}} f_{X_1^{n-1}}(x_1^{n-1}) \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \max\left\{ 1 - 2^{-t}/f_{X_1^{n-1}}(x_1^{n-1}),\, 0 \right\} \right) dt dx_1^{n-1} \\ &= \int_{\mathbb{R}^{n-1}} f_{X_1^{n-1}}(x_1^{n-1}) \left( -\log f_{X_1^{n-1}}(x_1^{n-1}) + \int_{-\infty}^{\infty} \left( \mathbf{1}\{t \ge 0\} - \max\{1 - 2^{-t},\, 0\} \right) dt dx_1^{n-1} \right) \\ &= h(X_1^{n-1}) + \log e. \end{aligned}$$

### D. Proof of Proposition 2

Note that

$$\sum_{l=-\infty}^{k} 2^{-nl} \left| D_l(A) \right| = 2^{-nk} \left| \{ v \in \mathbb{Z}^n : C_{k,v} \subseteq A \} \right| \ge 2^{-nk} \left| \left\{ v \in \mathbb{Z}^n : C_{k-1,(v-w)/2} \subseteq A \right\} \right|.$$

for any $w \in [0,1]^n$, since $C_{k,v} \subseteq C_{k-1,(v-w)/2}$. Note that the $(v-w)/2$ in the subscript may not have integer entries, but still the same definition $C_{k,v} = 2^{-k}([0,1]^n + v)$ can be applied. Also

$$\begin{aligned}\int_{[0,1]^n} \left|\left\{v \in \mathbb{Z}^n : C_{k-1,(v-w)/2} \subseteq A\right\}\right| dw &= \sum_{v\in\mathbb{Z}^n} \int_{[0,1]^n} \mathbf{1}\left\{C_{k-1,(v-w)/2} \subseteq A\right\} dw \\ &= 2^n \int_{\mathbb{R}^n} \mathbf{1}\left\{C_{k-1,w} \subseteq A\right\} dw \\ &= 2^n 2^{n(k-1)} \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right) \\ &= 2^{nk} \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right).\end{aligned}$$

Hence

$$\sum_{l=-\infty}^{k} 2^{-nl}\left|D_l(A)\right| \geq \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right),$$

$$\sum_{l=k+1}^{\infty} 2^{-nl}\left|D_l(A)\right| \leq \mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right).$$

Note that $H(W_A) = H(W_{(1/2)A})$, and also the right-hand-side of the proposition remains the same when $A$ is replaced by $(1/2)A$. Without loss of generality assume $A$ is small enough such that $\mathrm{V}_n(A) \leq 1$, so $D_k(A) = \emptyset$ for $k < 0$.

$$\begin{aligned}H(W_A) &= \log \mathrm{V}_n(A) + \frac{1}{\mathrm{V}_n(A)} \sum_{k=0}^{\infty} nk2^{-nk}\left|D_k(A)\right| \\ &= \log \mathrm{V}_n(A) + \frac{n}{\mathrm{V}_n(A)} \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty} 2^{-nl}\left|D_l(A)\right| \\ &\leq \log \mathrm{V}_n(A) + \frac{n}{\mathrm{V}_n(A)} \sum_{k=0}^{\infty}\left(\mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-(k-1)}]^n\right)\right) \\ &\leq \log \mathrm{V}_n(A) + \frac{n}{\mathrm{V}_n(A)} \int_{-2}^{\infty}\left(\mathrm{V}_n(A) - \mathrm{V}_n\left(A \ominus [0,\, 2^{-t}]^n\right)\right) dt \\ &= \log \mathrm{V}_n(A) + n \cdot h_{\ominus[0,1]^n}(A) + 2n.\end{aligned}$$

To prove the second result, consider

$$\begin{aligned}\sum_{l=-\infty}^{k} 2^{-nl}\left|D_l\left(\Lambda(A+U)\right)\right| &= 2^{-nk}\left|\{v \in \mathbb{Z}^n : C_{k,v} \subseteq \Lambda(A+U)\}\right| \\ &= 2^{-nk}\left|\left\{v \in \mathbb{Z}^n : \Lambda^{-1}C_{k,v} - U \subseteq A\right\}\right|.\end{aligned}$$

Assuming $k \geq -T$ and taking expectation over $U$, we obtain

$$\begin{aligned}\mathsf{E}\left[\sum_{l=-\infty}^{k} 2^{-nl}\left|D_l\left(\Lambda A+U\right)\right| \,\middle|\, \Lambda\right] &= 2^{-nT}\int_{[0,2^T]^n} 2^{-nk}\left|\{v \in \mathbb{Z}^n : C_{k,v} - u \subseteq \Lambda A\}\right| du \\ &= 2^{-nT}\int_{[0,2^T]^n} 2^{-nk}\left|\left\{v \in \mathbb{Z}^n : 2^{-k}\left([0,1]^n + v - 2^k u\right) \subseteq \Lambda A\right\}\right| du \\ &= \int_{[0,1]^n} 2^{-nk}\left|\left\{v \in \mathbb{Z}^n : 2^{-k}\left([0,1]^n + v - 2^{T+k} u\right) \subseteq \Lambda A\right\}\right| du \\ &\overset{(a)}{=} \int_{\mathbb{R}^n} 2^{-nk}\mathbf{1}\left\{2^{-k}\left([0,1]^n + u\right) \subseteq \Lambda A\right\} du \\ &= \int_{\mathbb{R}^n} \mathbf{1}\left\{2^{-k}[0,1]^n + u \subseteq \Lambda A\right\} du \\ &= \mathrm{V}_n\left(\Lambda A \ominus 2^{-k}[0,1]^n\right) \\ &= \Lambda^n \mathrm{V}_n\left(A \ominus \Lambda^{-1}2^{-k}[0,1]^n\right),\end{aligned}$$

where $(a)$ follows since $2^{T+k}$ is a non-negative integer. Since $2^{nT} > \mathrm{V}_n(2A)$, $D_k(\Lambda A+U) = \emptyset$ for $k < -T$. Hence, we have

$$H(W_{\Lambda A+U}) = \log \mathrm{V}_n(\Lambda A) + \frac{1}{\mathrm{V}_n(\Lambda A)} \sum_{k=-T}^{\infty} nk2^{-nk}\left|D_k(\Lambda A+U)\right|$$

$$= \log \mathrm{V}_n(A) + n \log \Lambda + \frac{n}{\Lambda^n \mathrm{V}_n(A)} \sum_{k=-T}^{\infty} \left( \mathbf{1}\left\{k \geq 0\right\} \Lambda^n \mathrm{V}_n(A) - \sum_{l=-\infty}^{k} 2^{-nl} \left| D_l(\Lambda A + U) \right| \right).$$

Taking expectation over $U$, we obtain

$$\mathsf{E}\left[ H(W_{\Lambda A+U}) \,\middle|\, \Lambda \right] = \log \mathrm{V}_n(A) + n \log \Lambda + \frac{n}{\Lambda^n \mathrm{V}_n(A)} \sum_{k=-T}^{\infty} \left( \mathbf{1}\left\{k \geq 0\right\} \Lambda^n \mathrm{V}_n(A) - \Lambda^n \mathrm{V}_n\left(A \ominus \Lambda^{-1} 2^{-k} [0,1]^n\right) \right)$$

$$= \log \mathrm{V}_n(A) + n \log \Lambda + n \sum_{k=-T}^{\infty} \left( \mathbf{1}\left\{k \geq 0\right\} - \frac{\mathrm{V}_n\left(A \ominus \Lambda^{-1} 2^{-k} [0,1]^n\right)}{\mathrm{V}_n(A)} \right).$$

Taking expectation over $\Lambda$, we have

$$\begin{aligned}
\mathsf{E}\left[H(W_{\Lambda A+U})\right] &= \log \mathrm{V}_n(A) + \mathsf{E}\left[n \log \Lambda\right] + n\,\mathsf{E}\left[ \sum_{k=-T}^{\infty} \left( \mathbf{1}\left\{k \geq 0\right\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-(k+\Theta)} [0,1]^n\right)}{\mathrm{V}_n(A)} \right) \right] \\
&= \log \mathrm{V}_n(A) + \frac{n}{2} + n \left( \int_{-T}^{\infty} \left( \mathbf{1}\left\{\theta \geq 0\right\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-\theta} [0,1]^n\right)}{\mathrm{V}_n(A)} \right) d\theta - \frac{1}{2} \right) \\
&= \log \mathrm{V}_n(A) + n \int_{-T}^{\infty} \left( \mathbf{1}\left\{\theta \geq 0\right\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-\theta} [0,1]^n\right)}{\mathrm{V}_n(A)} \right) d\theta \\
&\overset{\text{(a)}}{=} \log \mathrm{V}_n(A) + n \int_{-\infty}^{\infty} \left( \mathbf{1}\left\{\theta \geq 0\right\} - \frac{\mathrm{V}_n\left(A \ominus 2^{-\theta} [0,1]^n\right)}{\mathrm{V}_n(A)} \right) d\theta \\
&= \log \mathrm{V}_n(A) + n \cdot h_{\ominus [0,1]^n}(A).
\end{aligned}$$

where $(a)$ follows since $2^{nT} > \mathrm{V}_n(2A)$, hence $A \ominus 2^{-\theta}[0,1]^n = \emptyset$ for $\theta \leq -T$.

### *E. Proof of Theorem 2*

We first prove the following claim on $H(W_A)$ involving truncated differential entropy

$$H(W_A) \leq n \left( H(\zeta_1, \ldots, \zeta_n) + \sum_{i=1}^{n} \zeta_i \tilde{h}_{\zeta_i}(X_{[1:n]\backslash i}) \right) - (n-1) \log \mathrm{V}_n(A) + (2 + \log e) n,$$

where

$$\zeta_i = \int_{\mathbb{R}^{n-1}} \min\left\{ f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \,,\, \xi \right\} dx_{[1:n]\backslash i}$$

for a suitable $\xi > 0$ such that $\sum \zeta_i = 1$. By Proposition 2, the claim can be proved by bounding $h_{\ominus[0,1]^n}(A)$. Note that by Lemma 1,

$$\begin{aligned}
h_{\ominus [0,1]^n}(A) &= \int_{-\infty}^{\infty} \left( \mathbf{1}\left\{t \geq 0\right\} - \frac{\mathrm{V}_n\left(A \ominus [0,\, 2^{-t}]^n\right)}{\mathrm{V}_n(A)} \right) dt \\
&\leq \int_{-\infty}^{\infty} \left( \mathbf{1}\left\{t \geq 0\right\} - \frac{1}{\mathrm{V}_n(A)} \max\left( 0,\, \mathrm{V}_n(A) - \sum_{i=1}^{n} \int_{\mathrm{P}_{\backslash i}(A)} \min\left\{ 2^{-t} \,,\, \mathrm{V}_1\left(A \cap (\mathrm{span}(\mathrm{e}_i) + x)\right) \right\} dx_{[1:n]\backslash i} \right) \right) dt \\
&= \int_{-\infty}^{\infty} \left( \mathbf{1}\left\{t \geq 0\right\} - \max\left( 0,\, 1 - \sum_{i=1}^{n} \int_{\mathrm{P}_{\backslash i}(A)} \min\left\{ \frac{2^{-t}}{\mathrm{V}_n(A)} \,,\, f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \right\} dx_{[1:n]\backslash i} \right) \right) dt \\
&\overset{(a)}{=} \int_{-\infty}^{\infty} \left( -\mathbf{1}\left\{t < 0\right\} + \sum_{i=1}^{n} \int_{\mathrm{P}_{\backslash i}(A)} \min\left\{ \frac{2^{-t}}{\mathrm{V}_n(A)},\, \xi,\, f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \right\} dx_{[1:n]\backslash i} \right) dt \\
&= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \left( -\mathbf{1}\left\{t < 0\right\} \cdot \zeta_i + \int_{\mathrm{P}_{\backslash i}(A)} \min\left\{ \frac{2^{-t}}{\mathrm{V}_n(A)},\, \xi,\, f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \right\} dx_{[1:n]\backslash i} \right) dt \\
&= \sum_{i=1}^{n} \int_{\mathrm{P}_{\backslash i}(A)} \int_{-\infty}^{\infty} \left( -\mathbf{1}\left\{t < 0\right\} \cdot \min\left\{ \xi,\, f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \right\} + \min\left\{ \frac{2^{-t}}{\mathrm{V}_n(A)},\, \xi,\, f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \right\} \right) dt dx_{[1:n]\backslash i} \\
&= \sum_{i=1}^{n} \int_{\mathrm{P}_{\backslash i}(A)} - \min\left\{ \xi,\, f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \right\} \cdot \left( \log\left( \mathrm{V}_n(A) \min\left\{ \xi,\, f_{X_{[1:n]\backslash i}}(x_{[1:n]\backslash i}) \right\} \right) - \log e \right) dx_{[1:n]\backslash i} \\
&= -\log \mathrm{V}_n(A) + H(\zeta_1, \ldots, \zeta_n)
\end{aligned}$$

$$+\sum_{i=1}^{n}\zeta_i\int_{\mathrm{P}_{\setminus i}(A)}-\zeta_i^{-1}\min\left\{\xi,\,f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}\cdot\log\left(\zeta_i^{-1}\min\left\{\xi,\,f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\right\}\right)dx_{[1:n]\setminus i}+\log e$$
$$=-\log\mathrm{V}_n(A)+H(\zeta_1,\ldots,\zeta_n)+\sum_{i=1}^{n}\zeta_i\tilde{h}_{\zeta_i}(X_{[1:n]\setminus i})+\log e,$$

where $(a)$ follows by the definition of $\xi$. The claim follows.

We proceed to prove Theorem 2. Let $D=\mathrm{diag}(d_1,\ldots,d_n)$. Assuming $\prod_i d_i=1$, then by the claim,

$$H(W_{DA})\le n\left(H(\zeta_1,\ldots,\zeta_n)+\sum_{i=1}^{n}\zeta_i\left(\tilde{h}_{\zeta_i}(X_{[1:n]\setminus i})+\sum_{j\neq i}\log d_j\right)\right)-(n-1)\log\mathrm{V}_n(A)+(2+\log e)n$$
$$\le n\sum_{i=1}^{n}\zeta_i\left(\tilde{h}_{\zeta_i}(X_{[1:n]\setminus i})-\log d_i\right)-(n-1)\log\mathrm{V}_n(A)+n\log n+(2+\log e)n,$$

where

$$\zeta_i=\left(\prod_{j\neq i}d_j\right)\int_{\mathbb{R}^{n-1}}\min\left\{\left(\prod_{j\neq i}d_j\right)^{-1}f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\,,\,\xi\right\}dx_{[1:n]\setminus i}$$
$$=\int_{\mathbb{R}^{n-1}}\min\left\{f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\,,\,\xi d_i^{-1}\right\}dx_{[1:n]\setminus i}$$

for a suitable $\xi>0$ such that $\sum\zeta_i=1$. Let $\alpha_1,\ldots,\alpha_n>0$ such that

$$\int_{\mathbb{R}^{n-1}}\min\left\{f_{X_{[1:n]\setminus i}}(x_{[1:n]\setminus i})\,,\,\alpha_i\right\}dx_{[1:n]\setminus i}=\frac{1}{n}.$$

Set $\xi=\left(\prod_j a_j\right)^{1/n}$, $d_i=\alpha_i^{-1}\xi$, then we have $\zeta_i=1/n$,

$$H(W_{DA})\le\sum_{i=1}^{n}\tilde{h}_{1/n}(X_{[1:n]\setminus i})-(n-1)\log\mathrm{V}_n(A)+n\log n+(2+\log e)n.$$

### *F. Proof of Lemma 3*

Before proving this lemma, we first prove the following claim on the volume of a convex set. For any convex set $A\subseteq\mathbb{R}^n$ where $0\in A$ and $1\le m\le n$, let $\tilde{A}=\left\{x_{m+1}^n:\,(0^m,\,x_{m+1}^n)\in A\right\}$, then

$$\mathrm{V}_n(A)\ge m\cdot\mathrm{B}(m,\,n-m+1)\cdot\mathrm{V}_{n-m}(\tilde{A})\cdot\mathrm{VP}_{[1:m]}(A),$$

where

$$\mathrm{B}(\alpha,\beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$

is the beta function. Now we prove the claim. Denote the section of $A$ as

$$S_A(\tilde{x}_{m+1}^n)=\left\{x^m:\,(x^m,\,\tilde{x}_{m+1}^n)\in A\right\}\subseteq\mathbb{R}^m.$$

Note that

$$\mathrm{V}_n(A)=\int_{S^{m-1}}\int_0^\infty\left(\int_{\left\{\tilde{x}_{m+1}^n:\,(rx^m,\,\tilde{x}_{m+1}^n)\in A\right\}}d\tilde{x}_{m+1}^n\right)r^{m-1}drdx^m$$
$$=\int_{S^{m-1}}\int_{\left\{(r,\,\tilde{x}_{m+1}^n):\,r\ge0,\,(rx^m,\,\tilde{x}_{m+1}^n)\in A\right\}}r^{m-1}d(r,\,\tilde{x}_{m+1}^n)dx^m.$$

Consider the set

$$S_{\mathrm{rad},A}(x^m)=\left\{(r,\,\tilde{x}_{m+1}^n):\,r\ge0,\,(rx^m,\,\tilde{x}_{m+1}^n)\in A\right\}.$$

It is the intersection of $A$ and a half-space, and hence it is convex. By definition of radial function and projection, there exists $\hat{x}_{m+1}^n(x^n)$ such that

$$(\rho_{\mathrm{P}_{[1:m]}(A)}(x^m),\,\hat{x}_{m+1}^n(x^n))\in S_{\mathrm{rad},A}(x^m).$$

Also by definition of $\tilde{A}$,

$$\{0\}\times\tilde{A}\subseteq S_{\mathrm{rad},A}(x^m).$$

Hence the convex hull of $\left\{(\rho_{\mathrm{P}_{[1:m]}(A)}(x^m),\, \hat{x}_{m+1}^n(x^n))\right\} \cup \left(\{0\} \times \tilde{A}\right)$ is a subset of $S_{\mathrm{rad},A}(x^m)$. The convex hull can be expressed as

$$\left\{(r,\, \tilde{x}_{m+1}^n) :\, 0 \le r \le \rho_{\mathrm{P}_{[1:m]}(A)}(x^m),\, \tilde{x}_{m+1}^n \in \left(1 - r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\right) \tilde{A} + r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m) \cdot \hat{x}_{m+1}^n(x^n)\right\}.$$

Therefore,

$$\begin{aligned}
\mathrm{V}_n(A) &\ge \int_{S^{m-1}} \int_{S_{\mathrm{rad},A}(x^m)} r^{m-1} d(r,\, \tilde{x}_{m+1}^n) dx^m \\
&\ge \int_{S^{m-1}} \int_0^{\rho_{\mathrm{P}_{[1:m]}(A)}(x^m)} \int_{\left(1-r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\right)\tilde{A} + r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\cdot \hat{x}_{m+1}^n(x^n)} d\tilde{x}_{m+1}^n \cdot r^{m-1} dr dx^m \\
&= \int_{S^{m-1}} \int_0^{\rho_{\mathrm{P}_{[1:m]}(A)}(x^m)} \mathrm{V}_{n-m}(\tilde{A}) \left(1 - r\rho_{\mathrm{P}_{[1:m]}(A)}^{-1}(x^m)\right)^{n-m} r^{m-1} dr dx^m \\
&= \int_{S^{m-1}} \rho_{\mathrm{P}_{[1:m]}(A)}^m (x^m) \mathrm{V}_{n-m}(\tilde{A}) \cdot \mathrm{B}(m,\, n-m+1) dx^m \\
&= m \cdot \mathrm{B}(m,\, n-m+1) \cdot \mathrm{V}_{n-m}(\tilde{A}) \cdot \mathrm{VP}_{[1:m]}(A).
\end{aligned}$$

The claim follows.

We proceed to prove Lemma 3. To prove the lower bound, consider

$$\begin{aligned}
h(X_{m+1}^n \mid X^m) &= \int_{\mathbb{R}^m} f_{X^m}(x^m) h(X_{m+1}^n \mid X^m = x^m) dx^m \\
&\ge \int_{\mathbb{R}^m} f_{X^m}(x^m) \cdot -\log \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(\tilde{x}_{m+1}^n \mid x^m) dx^m \\
&\ge -\log \int_{\mathbb{R}^m} f_{X^m}(x^m) \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(\tilde{x}_{m+1}^n \mid x^m) dx^m \\
&= -\log \int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m,\, \tilde{x}_{m+1}^n) dx^m.
\end{aligned}$$

Now we prove the upper bound. By Lemma 2,

$$h(X^n) \le -\log\Big(\sup_{x^n \in \mathbb{R}^n} f(x^n)\Big) + n \log e, \tag{7}$$

$$h(X^m) \ge -\log\Big(\sup_{x^m \in \mathbb{R}^m} f_{X^m}(x^m)\Big). \tag{8}$$

Without loss of generality, assume that $\sup_{x^m \in \mathbb{R}^m} f_{X^m}(x^m)$ is attained at $x^m = 0$ and $\sup_{x_{n+1}^m} f(0^m, x_{n+1}^m)$ is attained at $x_{n+1}^m = 0$. Denote the super level set

$$L_z^+(f) = \{x^n :\, f(x^n) \ge z\}.$$

Since $f$ is log-concave, $L_z^+(f)$ is convex. Define

$$\tilde{L}_z^+(f) = \left\{x_{m+1}^n :\, (0^m,\, x_{m+1}^n) \in L_z^+(f)\right\}.$$

By the claim we proved earlier,

$$\begin{aligned}
\int_{\mathbb{R}^n} f(x^n) &= \int_0^\infty \mathrm{V}_n\left(L_z^+(f)\right) dz \\
&\ge \int_{\left\{z:\, 0 \in L_z^+(f)\right\}} \mathrm{V}_n\left(L_z^+(f)\right) dz \\
&\ge \int_{\left\{z:\, 0 \in L_z^+(f)\right\}} m \cdot \mathrm{B}(m,\, n-m+1) \cdot \mathrm{V}_{n-m}(\tilde{L}_z^+(f)) \cdot \mathrm{VP}_{[1:m]}(L_z^+(f)) dz \\
&= m \cdot \mathrm{B}(m,\, n-m+1) \cdot \int_0^{f(0)} \mathrm{V}_{n-m}(\tilde{L}_z^+(f)) \cdot \mathrm{VP}_{[1:m]}(L_z^+(f)) dz \\
&\overset{(a)}{\ge} m \cdot \mathrm{B}(m,\, n-m+1) \cdot \left(\int_0^{f(0)} \mathrm{V}_{n-m}(\tilde{L}_z^+(f)) dz\right) \cdot \left(\frac{1}{f(0)} \int_0^{f(0)} \mathrm{VP}_{[1:m]}(L_z^+(f)) dz\right) \\
&\overset{(b)}{\ge} m \cdot \mathrm{B}(m,\, n-m+1) \cdot f_{X^m}(0) \cdot \left(\frac{1}{\sup_x f(x)} \int_0^{\sup_x f(x)} \mathrm{VP}_{[1:m]}(L_z^+(f)) dz\right)
\end{aligned}$$

$$= m \cdot \mathrm{B}(m,\, n-m+1) \cdot f_{X^m}(0) \cdot \frac{1}{\sup_x f(x)} \int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m,\, \tilde{x}_{m+1}^n) dx^m$$

$$\overset{(c)}{\geq} m \cdot \mathrm{B}(m,\, n-m+1) \cdot 2^{-h(X^m)} \cdot 2^{h(X^n)} e^{-n} \int_{\mathbb{R}^m} \sup_{\tilde{x}_{m+1}^n \in \mathbb{R}^{n-m}} f(x^m,\, \tilde{x}_{m+1}^n) dx^m.$$

where $(a)$ is due to Chebyshev's sum inequality, since both $\mathrm{V}_{n-m}(\tilde{L}_z^+(f))$ and $\mathrm{VP}_{[1:m]}(L_z^+(f))$ are non-increasing in $z$, $(b)$ is due to $\int_0^{f(0)} \mathrm{V}_{n-m}(\tilde{L}_z^+(f))dz = \int_{\mathbb{R}^{n-m}} f(0^m,\, \tilde{x}_{m+1}^n) d\tilde{x}_{m+1}^n = f_{X^m}(0)$ since $\sup_{\tilde{x}_{n+1}^m} f(0^m, \tilde{x}_{n+1}^m) = f(0)$, and $\mathrm{VP}_{[1:m]}(L_z^+(f))\, dz$ is non-increasing in $z$, and $(c)$ is due to (7) and (8). The result follows from $\int_{\mathbb{R}^n} f(x^n) = 1$.

### G. *Proof of Lemma 4*

As in the definition of $\tilde{h}_\zeta$, let $\xi > 0$ such that

$$\int_{\mathbb{R}^n} \min\{\xi,\, f(x)\}\, dx = \zeta.$$

Without loss of generality, assume $\sup_{x\in\mathbb{R}^n} f(x) = f(0)$ and let $\alpha = f(0)$. Let

$$A = \{x \in \mathbb{R}^n :\, f(x) \geq \xi\}\,.$$

By log-concavity of $f$, we know $A$ is convex, and we have

$$\mathrm{V}_n(A) = \frac{1}{n} \cdot \int_{S^{n-1}} \rho_A^n(x) dx,$$

where $S^{n-1} = \{x \in \mathbb{R}^n \,:\, \|x\|_2 = 1\}$ is the unit sphere. For $x \in A$, by definition of $\rho_A$, we know $x/(\rho_A^{-1}(x) + \epsilon) \in A$ for any $\epsilon > 0$,

$$\begin{aligned} f(x) &= f\left(\left(1 - \rho_A^{-1}(x) - \epsilon\right) \cdot 0 + \left(\rho_A^{-1}(x) + \epsilon\right) \cdot \frac{x}{\rho_A^{-1}(x) + \epsilon}\right) \\ &\geq (f(0))^{1-\rho_A^{-1}(x)-\epsilon} \cdot \left(f\left(\frac{x}{\rho_A^{-1}(x)+\epsilon}\right)\right)^{\rho_A^{-1}(x)+\epsilon} \\ &\geq \alpha^{1-\rho_A^{-1}(x)-\epsilon} \xi^{\rho_A^{-1}(x)+\epsilon}. \end{aligned}$$

Therefore

$$f(x) \geq \alpha^{1-\rho_A^{-1}(x)} \xi^{\rho_A^{-1}(x)}.$$

Hence

$$\begin{aligned} \int_A f(x)dx &= \int_{S^{n-1}} \int_0^{\rho_A(x)} f(rx) \cdot r^{n-1} dr dx \\ &\geq \int_{S^{n-1}} \int_0^{\rho_A(x)} \alpha^{1-\rho_A^{-1}(rx)} \xi^{\rho_A^{-1}(rx)} r^{n-1} dr dx \\ &= \int_{S^{n-1}} \int_0^{\rho_A(x)} \alpha^{1-r\rho_A^{-1}(x)} \xi^{r\rho_A^{-1}(x)} r^{n-1} dr dx \\ &= \int_{S^{n-1}} \rho_A^n(x) \int_0^1 \alpha^{1-r} \xi^r r^{n-1} dr dx \\ &= \int_{S^{n-1}} \rho_A^n(x) \left(\alpha\left(-\log(\xi/\alpha)\right)^{-n} \left(\Gamma(n) - \Gamma(n,\, -\log(\xi/\alpha))\right)\right) dx \\ &= n\alpha\left(-\log(\xi/\alpha)\right)^{-n} \left(\Gamma(n) - \Gamma(n,\, -\log(\xi/\alpha))\right) \mathrm{V}_n(A), \end{aligned}$$

where $\Gamma(n,z) = \int_z^\infty t^{n-1} e^{-t} dt$ is the incomplete gamma function, and $\Gamma(n) = \Gamma(n,0)$ is the gamma function.

On the other hand, for $x \notin A$, then for any $\epsilon > 0$, we have $x/(\rho_A^{-1}(x) - \epsilon) \notin A$,

$$\begin{aligned} \xi \geq f\left(\frac{x}{\rho_A^{-1}(x)-\epsilon}\right) &= f\left(\left(1 - \frac{1}{\rho_A^{-1}(x)-\epsilon}\right) \cdot 0 + \frac{1}{\rho_A^{-1}(x)-\epsilon} \cdot x\right) \\ &\geq \alpha^{1-1/(\rho_A^{-1}(x)-\epsilon)} \cdot (f(x))^{1/(\rho_A^{-1}(x)-\epsilon)}, \end{aligned}$$

Therefore

$$f(x) \leq \alpha^{1-\rho_A^{-1}(x)} \xi^{\rho_A^{-1}(x)}.$$

Hence

$$\begin{aligned}
\int_{\mathbb{R}^n\setminus A} f(x)dx &= \int_{S^{n-1}}\int_{\rho_A(x)}^{\infty} f(rx)\cdot r^{n-1}drdx \\
&\leq \int_{S^{n-1}}\int_{\rho_A(x)}^{\infty} \alpha^{1-\rho_A^{-1}(rx)}\xi^{\rho_A^{-1}(rx)}r^{n-1}drdx \\
&= \int_{S^{n-1}}\int_{\rho_A(x)}^{\infty} \alpha^{1-r\rho_A^{-1}(x)}\xi^{r\rho_A^{-1}(x)}r^{n-1}drdx \\
&= \int_{S^{n-1}}\rho_A^n(x)\int_{1}^{\infty} \alpha^{1-r}\xi^{r}r^{n-1}drdx \\
&= \int_{S^{n-1}}\rho_A^n(x)\left(\alpha\left(-\log(\xi/\alpha)\right)^{-n}\Gamma(n,\,-\log(\xi/\alpha))\right)dx \\
&= n\alpha\left(-\log(\xi/\alpha)\right)^{-n}\Gamma(n,\,-\log(\xi/\alpha))\mathrm{V}_n(A).
\end{aligned}$$

Recall that $\int_{\mathbb{R}^n}\min\{\xi,\,f(x)\}\,dx=\zeta$,

$$\begin{aligned}
\zeta &= \int_{\mathbb{R}^n}\min\{\xi,\,f(x)\}\,dx \\
&= \int_{\mathbb{R}^n\setminus A} f(x)dx+\xi\mathrm{V}_n(A) \\
&\leq \left(n\alpha\left(-\log(\xi/\alpha)\right)^{-n}\Gamma(n,\,-\log(\xi/\alpha))+\xi\right)\mathrm{V}_n(A).
\end{aligned}$$

Also

$$\begin{aligned}
\zeta &= \int_{\mathbb{R}^n}\min\{\xi,\,f(x)\}\,dx \\
&= 1-\int_{A} f(x)dx+\xi\mathrm{V}_n(A) \\
&\leq 1-\left(n\alpha\left(-\log(\xi/\alpha)\right)^{-n}\left(\Gamma(n)-\Gamma(n,\,-\log(\xi/\alpha))\right)-\xi\right)\mathrm{V}_n(A).
\end{aligned}$$

Let $\nu=-\log(\xi/\alpha)$. Since $\zeta\leq ac$ and $\zeta\leq 1-bc$ implies $\zeta\leq a/(a+b)$, we have

$$\begin{aligned}
\zeta &\leq \frac{n\alpha\nu^{-n}\Gamma(n,\nu)+\xi}{(n\alpha\nu^{-n}\Gamma(n,\nu)+\xi)+(n\alpha\nu^{-n}\left(\Gamma(n)-\Gamma(n,\nu)\right)-\xi)} \\
&= \frac{n\Gamma(n,\nu)+e^{-\nu}\nu^n}{n\Gamma(n)} \\
&= \frac{\Gamma(n+1,\,\nu)}{\Gamma(n+1)}.
\end{aligned}$$

By Lemma 2, $h(X^n)\geq-\log\alpha$. Recall that $\tilde{h}_\zeta(X^n)$ is the entropy of the pdf $\tilde{f}(x)=\zeta^{-1}\min\{\xi,\,f(x)\}$, which is also log-concave. Hence by Lemma 2, $\tilde{h}_\zeta(X^n)\leq-\log(\zeta^{-1}\xi)+n\log e$. As a result,

$$\begin{aligned}
\tilde{h}_\zeta(X^n)-h(X^n) &\leq -\log\left(\zeta^{-1}\xi/\alpha\right)+n\log e \\
&\leq \log\zeta+\nu+n\log e.
\end{aligned}$$

To prove the second bound, assume that $\zeta\geq e^{-(e-2)n}$ and $\nu>en$. We use the bound

$$\Gamma(a,z)<Bz^{a-1}e^{-z}$$

for $a>1$, $B>1$, $z>B(a-1)/(B-1)$ due to [13].

Substituting $a=n+1$, $z=en$, and $B=e$, we have $\Gamma(n+1,\nu)<\Gamma(n+1,en)<e(en)^ne^{-en}$. We also know that $\Gamma(n+1)\geq n^ne^{-(n-1)}$, hence

$$\begin{aligned}
\frac{\Gamma(n+1,\,\nu)}{\Gamma(n+1)} &< \frac{e(en)^ne^{-en}}{n^ne^{-(n-1)}} \\
&= e^{-(e-2)n},
\end{aligned}$$

which leads to a contradiction and $\nu\leq en$. The result follows.

## REFERENCES


[1] G. R. Kumar, C. T. Li, and A. El Gamal, “Exact common information,” in *Proc. IEEE Symp. Info. Theory*. IEEE, 2014, pp. 161–165.

[2] T. Linder, V. Tarokh, and K. Zeger, “Existence of optimal prefix codes for infinite source alphabets,” *IEEE Trans. Info. Theory*, vol. 43, no. 6, pp. 2026–2028, 1997.

[3] A. D. Wyner, “The common information of two dependent random variables,” *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[4] P. Cuff, H. Permuter, and T. M. Cover, “Coordination capacity,” *IEEE Trans. Info. Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.

[5] W. Liu, G. Xu, and B. Chen, “The common information of $N$ dependent random variables,” in *Proc. 48th Ann. Allerton Conf. Commun., Contr., and Comput.* IEEE, 2010, pp. 836–843.

[6] G. Xu, W. Liu, and B. Chen, “Wyner’s common information for continuous random variables–a lossy source coding interpretation,” in *Proc. 45th Ann. Conf. Information Sciences and Systems*. IEEE, 2011, pp. 1–6.

[7] S. Satpathy and P. Cuff, “Gaussian secure source coding and Wyner’s common information,” *arXiv preprint arXiv:1506.00193*, 2015.

[8] P. Yang and B. Chen, “Wyner’s common information in Gaussian channels,” in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 3112–3116.

[9] T. S. Han, “Nonnegative entropy measures of multivariate symmetric correlations,” *Information and Control*, vol. 36, pp. 133–156, 1978.

[10] S. Bobkov and M. Madiman, “The entropy per coordinate of a random vector is highly constrained under convexity conditions,” *IEEE Trans. Info. Theory*, vol. 57, no. 8, pp. 4940–4954, 2011.

[11] P. Cuff, “Distributed channel synthesis,” *IEEE Trans. Info. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[12] M. Sklar, *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8, 1959.

[13] P. Natalini and B. Palumbo, “Inequalities for the incomplete gamma function,” *Math. Inequal. Appl*, vol. 3, no. 1, pp. 69–77, 2000.