Rate-Distortion Bounds for High-Resolution Vector Quantization via Gibbs’s Inequality

Tobias Koch and Gonzalo Vazquez-Vilar

Abstract

Gibbs’s inequality states that the differential entropy of a random variable with probability density function \( f \) is less than or equal to its cross entropy with any other pdf \( g \) defined on the same alphabet, i.e., \( h(X) \leq -E[\log g(X)] \). Using this inequality with a cleverly chosen \( g \), we derive a lower bound on the smallest output entropy that can be achieved by quantizing a \( d \)-dimensional source with given expected \( r \)th-power distortion. Specialized to the one-dimensional case, and in the limit of vanishing distortion, this lower bound converges to the output entropy achieved by a uniform quantizer, thereby recovering the result by Gish and Pierce that uniform quantizers are asymptotically optimal as the allowed distortion tends to zero. Our lower bound holds for any \( d \)-dimensional memoryless source that has a pdf and whose differential entropy and Rényi information dimension are finite. In contrast to Gish and Pierce, we do not require any additional constraints on the continuity or decay of the source pdf.

1 Introduction

Suppose we wish to quantize a memoryless source with an \( r \)th-power distortion not larger than \( D \). More specifically, suppose a source produces the sequence of independent and identically distribution (IID), \( d \)-dimensional, real-valued vectors \( \{X_k, k \in \mathbb{Z}\} \) according to the distribution \( P_X \) and we employ a vector quantizer that produces a sequence of quantized symbols \( \{\hat{X}_k, k \in \mathbb{Z}\} \) satisfying

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} E[\|X_k - \hat{X}_k\|^r] \leq D
\]

for some norm \( \| \cdot \| \) and some exponent \( r > 0 \). (We use \( \lim \) to denote the limit superior and \( \lim \) to denote the limit inferior.) Rate-distortion theory states that if for every blocklength \( n \) and distortion constraint \( D \) we quantize the sequence of source vectors \( X_1, \ldots, X_n \) to one of \( e^{nR} \) possible sequences of quantized symbols \( \hat{X}_1, \ldots, \hat{X}_n \), then the smallest rate \( R \) (in nats per source symbol) for which there exists a vector quantizer satisfying (1) is given by [1,2]

\[
R(D) = \inf_{P_{\hat{X}|X}} I(X; \hat{X})
\]

where the infimum is over all conditional distributions of \( \hat{X} \) given \( X \) for which

\[
E[\|X - \hat{X}\|^r] \leq D
\]

and where the expectation in (3) is computed with respect to the joint distribution \( P_X P_{\hat{X}|X} \). Here and throughout this paper we omit the time indices where they are immaterial. The rate \( R(D) \) as a function of \( D \) is referred to as the rate-distortion function.

While \( R(D) \) characterizes the rate of the optimal vector quantizer that quantizes the source with \( r \)th-power distortion not exceeding \( D \), sometimes quantizing blocks of \( n \) source symbols may not be feasible, especially if \( n \) is large (which is typically required to achieve (2)). In this case, it might be more practical to quantize

---

The authors are with the Signal Theory and Communications Department, Universidad Carlos III de Madrid, 28911, Leganés, Spain and also with the Gregorio Maraño Health Research Institute (e-mails: koch@tsc.uc3m.es and gonzalo.vazquez@uc3m.es).

This work has been supported by the European Community’s Seventh Framework Programme FP7/2007-2013 under Grant 333680, by the Spanish Ministry of Economy and Competitiveness under Grants TEC2015-69648-REDC, RYC-2014-16322, TEC2013-41718-R, and FPDI-2013-18602, and by the Comunidad de Madrid under Grant S2013/ICE-2845.
each source symbol separately using a vector quantizer, defined as a (deterministic) mapping \( q(\cdot) \) from the source alphabet \( \mathcal{X} \) to the countable reconstruction alphabet \( \hat{\mathcal{X}} \). In this paper, we consider the symbol-wise quantization of \( d \)-dimensional source vectors, i.e., \( \mathcal{X}, \hat{\mathcal{X}} \subseteq \mathbb{R}^d \). This setup is sufficiently general to comprise various problems of interest in high-resolution vector quantization. For example, it directly describes the quantization of a sequence of \( d \) consecutive symbols of a one-dimensional memoryless source using a vector quantizer. Furthermore, the quantization of stationary sources with memory can be studied by combining the analysis of symbol-wise, \( d \)-dimensional quantization with a limiting argument where \( d \to \infty \).

We define the rate of the vector quantizer as the entropy of the quantized source symbol \( \hat{X} = q(X) \). Thus, the smallest rate of a symbol-wise quantizer satisfying the distortion constraint \( D \) is given by

\[
R_{r,d}(D) \triangleq \inf_{q(\cdot)} H(q(X))
\]  

where the infimum is over the set of quantizers \( q(\cdot) \) satisfying (3). Since \( X \) determines the quantizer output \( q(X) \), we have \( H(q(X)|X) = 0 \) and the rate \( R_{r,d}(D) \) can be written in the same form as (2) but with \( P_{X|X} \) replaced by \( q(\cdot) \):

\[
R_{r,d}(D) = \inf_{q(\cdot)} I(X; q(X)).
\]  

Since \( \hat{X} = q(X) \) corresponds to a deterministic \( P_{\hat{X}|X} \), it follows that \( R_{r,s}(D) \geq R(D) \).

By Shannon’s source coding theorem, any discrete memoryless source can be losslessly described by a variable-length code whose expected length is roughly the entropy of the source [1]. Consequently, \( R_{r,d}(D) \) is the smallest expected length of a vector quantization scheme that first quantizes each source symbol using a vector quantizer and then compresses the resulting sequence of quantized symbols using a lossless variable-length code.

In this paper, we focus on the asymptotic rate-distortion tradeoff in the limit as the permitted distortion tends to zero. Specifically, we study the asymptotic excess rate with respect to the rate-distortion function defined, as

\[
R_{r,d} \triangleq \lim_{D \to 0} \{ R_{r,d}(D) - R(D) \}
\]

For one-dimensional sources \((d = 1)\) and quadratic distortion \((r = 2)\), Gish and Pierce demonstrated that the excess rate is equal to [3]

\[
R_{2,1} = \frac{1}{2} \log \frac{\pi e}{6}
\]  

where \( \log(\cdot) \) denotes the natural logarithm. For multi-dimensional sources, only bounds on \( R_{r,d} \) are available. To obtain (7), Gish and Pierce [3] imposed constraints on the continuity and decay of the probability density function (pdf) of \( X \). Furthermore, they merely provide an intuitive explanation of their result together with an outline of the proof—at the end of Appendix II of [3] they write “The complete proof is surprisingly long and will not be given here.”

The result (7) is equivalent to a result by Zador, which concerns the asymptotic excess distortion with respect to the distortion-rate function as the rate tends to infinity [4]. Indeed, let \( D_{r,d}(R) \) denote the minimum distortion achievable with a symbol-wise quantizer whose output has an entropy not exceeding \( R \), i.e.,

\[
D_{r,d}(R) \triangleq \inf_{q(\cdot)} \mathbb{E} [\| X - q(X) \|^r]
\]  

where the infimum is over the set of quantizers \( q(\cdot) \) satisfying \( H(q(X)) \leq R \). Zador’s theorem states that

\[
\lim_{R \to \infty} e^{R D_{r,d}(R)} = b_{r,d} e^{\frac{1}{2} h(X)}
\]  

where \( b_{r,d} \) is a constant that only depends on \( r \) and \( d \) but not on the distribution of \( X \). Zador did not evaluate the constant \( b_{r,d} \), but he did provide upper and lower bounds on \( b_{r,d} \) that become tight for large \( d \). Furthermore, for one-dimensional sources and quadratic distortion, it can be shown that \( b_{2,1} = 1/12 \). Taking logarithms on both sides of (9), and replacing \( R \leftrightarrow R_{r,d}(D) \) and \( D_{r,d}(R) \leftrightarrow D \), we thus obtain that

\[
R_{2,1}(D) = h(X) + \frac{1}{2} \log \frac{1}{D} - \frac{1}{2} \log 12 + o_R(1)
\]  

2
where \( o_R(1) \) denotes error terms that vanish as \( R \) tends to infinity. By the asymptotic tightness of the Shannon lower bound, we further have [5–8]

\[
R(D) = h(X) + \frac{1}{2} \log \frac{1}{D} - \frac{1}{2} \log(2\pi e) + o_D(1)
\]

where \( o_D(1) \) denotes error terms that vanish as \( D \) tends to zero. Hence, the equivalence of Zador’s theorem (9) and Gish and Pierce’s result (7) follows.

While Zador’s original proof of (9) was flawed, a rigorous proof for quadratic distortion was given by Gray, Linder, and Li by using a Langrangian formulation of variable-rate vector quantization [9]. Their proof follows Zador’s approach of 1) proving the result for sources with a uniform pdf on the unit cube; 2) extending it to piecewise constant pdfs on disjoint cubes of equal sides; 3) prove the result for a general pdf on a cube; and 4) proving the result for general pdfs. Gray et al. do not impose any constraints on the continuity or decay of the pdf of \( X \), so their proof is more general than both the proof by Zador [4] and the proof by Gish and Pierce [3].

In this paper, we derive a lower bound on \( R_{r,d} \) that recovers (7) for one-dimensional sources and quadratic distortion. In contrast to [9], our proof follows essentially along the lines outlined by Gish and Pierce [3]. We do not impose any constraints on the continuity or decay of the pdf of \( X \), so our proof is as general as the proof by Gray et al., and it is more general than the proof by Gish and Pierce.

The rest of this paper is organized as follows. Section 2 introduces the problem setup and presents the main result of this paper, Theorem 1. Section 3 provides a back-of-the-envelope derivation of Theorem 1 that serves as an outline for the proof. Section 4 contains the complete proof of this theorem. Section 5 assesses the tightness of the lower bound presented in Theorem 1 by numerically comparing it to several upper bounds achievable by lattice quantizers. Section 6 concludes the paper with a summary and discussion of the results.

## 2 Problem Setup and Main Result

We consider a \( d \)-dimensional, real-valued source \( X \) with support \( \mathcal{X} \subseteq \mathbb{R}^d \) whose distribution is absolutely continuous with respect to the Lebesgue measure and we denote its pdf by \( f_X \). We require the source to satisfy the following two conditions:

**C1** \( x \mapsto f_X(x) \log f_X(x) \) is integrable, ensuring that the differential entropy

\[
h(X) \triangleq - \int_{\mathcal{X}} f_X(x) \log f_X(x) \, dx
\]

is well-defined and finite;

**C2** the integer part of the source \( X \) has a finite entropy, i.e.,

\[
H([X]) < \infty.
\]

Here \( [\mathbf{a}] \), \( \mathbf{a} = (a_1, \ldots, a_d) \in \mathbb{R}^d \) denotes the element-wise floor function, i.e., \( [\mathbf{a}] = ([a_1], \ldots, [a_d]) \in \mathbb{Z}^d \) where \( [a_\ell] \), \( \ell = 1, \ldots, d \) denotes the largest integer not larger than \( a_\ell \).

Condition C2 requires that quantizing the source with a cubic lattice quantizer of unit-volume cells gives rise to a discrete random variable of finite entropy. It is necessary for the asymptotic excess rate \( R_{r,d} \) to be well-defined. Indeed, as demonstrated in [7], if \( H([X]) = \infty \) then the rate-distortion function \( R(D) \) is infinite for any finite \( D \). Since \( R_{r,d}(D) \geq R(D) \), this implies that in this case \( R_{r,d}(D) = R(D) \) is of the form \( \infty - \infty \). Fortunately, Condition C2 is very mild. For example, by generalizing Proposition 1 in [10] to the vector case, it can be shown that it is satisfied if \( E[\log(1+||X||)] < \infty \). This in turn is true, for example, for sources for which \( E[||X||^\alpha] < \infty \) for some \( \alpha > 0 \).

The quantity \( H([X]) \) is intimately related with the Rényi information dimension defined in [11]; see also [10,12]. Indeed, generalizing Proposition 1 in [10] to the vector case, it can be shown that a source vector has a finite Rényi information dimension if, and only if, (13) is satisfied.

The quantizer is characterized by the (Borel measurable) function \( q : \mathcal{X} \rightarrow \hat{\mathcal{X}} \) for some (countable) reconstruction alphabet \( \hat{\mathcal{X}} \subseteq \mathbb{R}^d \). Equivalently, we characterize \( q \) by the quantization regions \( S_i, i \in \mathbb{Z} \) and
corresponding reconstruction values $\hat{x}_i$, $i \in \mathbb{Z}$. Specifically, $S_i$, $i \in \mathbb{Z}$ are disjoint (Borel measurable) subsets of $\mathbb{R}^d$ that together with the reconstruction values $\hat{x}_i$, $i \in \mathbb{Z}$ satisfy

$$\bigcup_i S_i = \mathcal{X} \quad (14a)$$

$$q(x) = \sum_i \hat{x}_i \mathbb{1}\{x \in S_i\}, \quad \text{for } x \in \mathcal{X} \quad (14b)$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. To simplify notation, we denote the Lebesgue measure of the quantization region $S_i$ by $\Delta_i$ and the probability of $X$ being in $S_i$ by $p_i$.

The main result of this paper is a lower bound on the excess rate $R_{r,d}$ for general $r$ and $d$. For one-dimensional sources and quadratic distortion, it recovers the excess rate (7) by Gish and Pierce. However, in contrast to Gish and Pierce’s result, our bound does not require any continuity or decay conditions on the behavior of the source pdf—it holds for all source vectors having a pdf, having a finite differential entropy, and having a finite Rényi information dimension.

**Theorem 1** (Main Result). *Let the source vector $X$ have a pdf and assume that $h(X)$ and $H(|X|)$ are finite. Then, the excess rate $R_{r,d}$, as defined in (6), is lower-bounded by

$$R_{r,d} \geq \frac{d}{r} \log \left( \frac{\Gamma(1 + d/r)^{d/r}}{1 + d/r} \right)$$

(15)

where $\Gamma(\cdot)$ denotes the Gamma function.*

**Proof.** See Section 4. \hfill \Box

In the one-dimensional case, (15) becomes

$$R_{r,1} \geq \frac{1}{r} \log \left( \frac{\Gamma(1 + 1/r)^{1/r}}{1 + 1/r} \right).$$

(16)

As we shall see next, this lower bound can be achieved by a uniform quantizer, so in the one-dimensional case it is tight. Furthermore, for quadratic distortion, (16) is equal to $1/2 \log(\pi e/6)$, hence it recovers the excess rate obtained by Gish and Pierce.

To demonstrate the tightness of (15) in the one-dimensional case, and to assess the accuracy of (15) in higher-dimensional cases, we consider an upper bound on the excess rate that follows by restricting ourselves to the class of **tessellating quantizers**. A polytope $P$ is **tessellating** if there exists a partition of $\mathbb{R}^d$ consisting of translated and/or rotated copies of $P$; a **tessellating quantizer**, denoted by $q_P: \mathcal{X} \to \mathcal{X}$, is a quantizer whose quantization regions $S_i$ are translated and/or rotated copies of a tessellating convex polytope $P$ and the corresponding reconstruction values $\hat{x}_i$ are the centroids of $S_i$. A special case of a tessellating quantizer is a **lattice quantizer**, i.e., a quantizer whose quantization regions are the Voronoi cells of a $d$-dimensional lattice. Note that in the one-dimensional case the only convex polytope is the interval, so in this case the tessellating quantizer is the uniform quantizer. For the class of tessellating quantizers, Linder and Zeger [13] derived an asymptotic expression equivalent to (9).

**Theorem 2** (Linder and Zeger [13, Th. 1]). *Let the source vector $X$ have a pdf and assume that $h(X)$ and $H(|X|)$ are finite. Then, a tessellating quantizer $q_P(\cdot)$ with $r$th-power distortion $\mathbb{E}[\|X - q_P(X)\|^r] = D$ and rate $R_P(D) \triangleq H(q_P(X))$ satisfies

$$\lim_{D \downarrow 0} D e^{\frac{1}{D} R_P(D)} = \ell(P) e^{\frac{1}{D} h(X)}$$

(17)

where $\ell(P)$ denotes the normalized $r$-th moment of $P$, defined as

$$\ell(P) \triangleq \frac{\int_P \|x - \hat{x}\|^r \, dx}{V(P)^{1+r/d}}$$

(18)

and $V(P)$ denotes the volume of $P$.\]
Remark 1. To be precise, Theorem 1 in [13] requires that $H(q_{P_0}(X)) < \infty$ for some $\alpha > 0$ rather than (13), i.e., $H(|X|) < \infty$. (Here $P_0 = \{x \in \mathbb{R}^d : x/\alpha \in P\}$ denotes the polytope $P$ rescaled by $\alpha$.) Nevertheless, its proof hinges on a lemma by Csiszár (cf. [13, Lemma 2]), which also applies if the condition $H(q_{P_0}(X)) < \infty$ is replaced by (13). Specifically, by setting in [13, Lemma 2] the partition $B_0 = \{B_1, B_2, \ldots\}$ of $\mathbb{R}^d$ to be the set of $d$-dimensional cubes of unit-volume with the lower-most cornerpoint located at coordinates $i \in \mathbb{Z}^d$, this partition satisfies the lemma’s conditions provided that (13) holds.

Taking logarithms on both sides of (17), we obtain

$$R_P(D) = h(X) + \frac{d}{r} \log \frac{1}{D} + \frac{d}{r} \log \ell(P) + o_D(1).$$

(19)

Since a tessellating quantizer with $r$-th-power distortion $D$ satisfies (3), the rate-distortion function $R_P(D)$ upper-bounds $R_{r,d}(D)$. An upper bound on $R_{r,d}$ thus follows from (19) and by expressing $R(D)$ as [14]

$$R(D) = h(X) + \frac{d}{r} \log \frac{1}{D} - \frac{d}{r} \log \left( \frac{r}{d} V_d (1 + d/r)^{r/d} \right) + o_D(1)$$

(20)

where $V_d$ denotes the volume of the unit ball $\{x \in \mathbb{R}^d : \|x\| \leq 1\}$. The right-hand side (RHS) of (20) (without the error term $o_D(1)$) is referred to as Shannon lower bound. It lower-bounds the rate-distortion function $R(D)$ for every $D > 0$, and its difference to $R(D)$ vanishes as $D$ tends to zero, provided that the source distribution satisfies some conditions; see, e.g., [5–8]. A finite-blocklength refinement of this bound can be found in [15, 16]. Recently, it has been demonstrated that the Shannon lower bound is asymptotically tight if, and only if, $h(X)$ and $H(|X|)$ are finite [7]. Thus, (20) holds for the class of sources considered in this paper.

Combining (19) with (20), we obtain

$$\lim_{D \downarrow 0} \{R_P(D) - R(D)\} = \frac{d}{r} \log \left( \frac{r}{d} V_d (1 + d/r)^{r/d} \right) + \frac{d}{r} \log \ell(P).$$

(21)

Recalling that $R_{r,d}(D) \leq R_P(D)$ for every $D$, this yields

$$R_{r,d} \leq \frac{d}{r} \log \left( \frac{r}{d} V_d (1 + d/r)^{r/d} \right) + \inf_{P} \frac{d}{r} \log \ell(P)$$

(22)

where the infimum is over all $d$-dimensional, tessellating, convex polytopes $P$.

Using that in the one-dimensional case the only convex polytope is the interval, and noting that the interval has the normalized $r$-th moment

$$\ell(P) = \frac{1}{2^r(1+r)}$$

(23)

the upper bound (22) becomes in this case

$$R_{r,1} \leq \frac{d}{r} \log \left( \frac{\Gamma(1 + 1/r)^r}{1 + 1/r} \right)$$

(24)

which coincides with (16). Thus, in the one-dimensional case, a tessellating quantizer (which in this case is the uniform quantizer) is asymptotically optimal.

3 Derivation for One-Dimensional Sources and Certain Quantizers

Before proving Theorem 1, we provide a simplified derivation of the lower bound (15) for one-dimensional sources ($d = 1$) and quadratic distortion ($r = 2$) that will serve as an outline for the complete proof of Theorem 1 given in Section 4. In this derivation, we shall only consider quantizers satisfying

$$\sup_{i} \sup_{x \in S_i} (x - \hat{x}_i)^2 \leq \alpha D, \quad \text{for some constant } \alpha.$$

(25)

These simplifying assumptions are, for example, satisfied by the uniform quantizer when $\hat{x}_i$ is the midpoint of $S_i$ and the cell length $\Delta$ vanishes proportionally to $D$. Nevertheless, it is *prima facie* unclear whether (25) holds without loss of optimality.
By (5), we have
\[ R_{2,1}(D) = \inf_{q(\cdot)} I(X; \hat{X}) = h(X) - \sup_{q(\cdot)} h(X|\hat{X}). \] (26)

In order to lower-bound \( R_{2,1}(D) \), we upper-bound \( h(X|\hat{X}) \) by using that, conditioned on \( \hat{X} = \hat{x}_i \), the support of \( X \) is \( S_i \), so a uniform distribution over \( S_i \) maximizes the differential entropy [17, Th. 11.1.1]:
\[ h(X|\hat{X} = \hat{x}_i) \leq \log \Delta_i. \] (27)

Averaging over \( \hat{X} \) then yields
\[ R_{2,1}(D) \geq h(X) - \sup_{q(\cdot)} \sum_i p_i \log \Delta_i, \] (28)
which, by Jensen’s inequality, can be further lower-bounded by
\[ R_{2,1}(D) \geq h(X) - \frac{1}{2} \log \left( \sup_{q(\cdot)} \sum_i p_i \Delta_i^2 \right). \] (29)

Together with (11), this yields
\[ \lim_{D \uparrow 0} \left\{ R_{2,1}(D) - R(D) \right\} \geq \lim_{D \uparrow 0} \left\{ \frac{1}{2} \log D + \frac{1}{2} \log(2\pi e) - \frac{1}{2} \log \left( \sup_{q(\cdot)} \sum_i p_i \Delta_i^2 \right) \right\}. \] (30)

It remains to show that, for any sequence of quantizers (parametrized by \( D \)),
\[ \lim_{D \uparrow 0} \frac{1}{D} \sum \frac{1}{\Delta_i^2} \leq 12. \] (31)

Then the RHS of (30) is lower-bounded by \( 1/2 \log(\pi e/6) \) and, upon noting that the left-hand side (LHS) of (30) is equal to \( R_{2,1} \), we obtain \( R_{2,1} \geq 1/2 \log(\pi e/6) \). Hence we recover Theorem 1 for one-dimensional sources and quadratic distortion.

The upper bound (31) follows along the lines of the proof of Lemma 1 in [13]. We first express \( \mathbb{E}[(X - \hat{X})^2] \) as
\[ \mathbb{E}[(X - \hat{X})^2] = \sum_i \int_{S_i} f_X(x)(x - \hat{x}_i)^2 \, dx 
= \sum_i p_i \frac{1}{\Delta_i} \int_{S_i} (x - \hat{x}_i)^2 \, dx - \sum_i \int_{S_i} \left[ \frac{p_i}{\Delta_i} - f_X(x) \right] (x - \hat{x}_i)^2 \, dx. \] (32)

We next note that the region \( S_i \) of measure \( \Delta_i \) that minimizes \( \int_{S_i} (x - \hat{x}_i)^2 \, dx \) is the interval \( [\hat{x}_i - \frac{\Delta_i}{2}, \hat{x}_i + \frac{\Delta_i}{2}] \), so
\[ \frac{1}{\Delta_i} \int_{S_i} (x - \hat{x}_i)^2 \, dx \geq \frac{\Delta_i^2}{12}. \] (33)

The first term on the RHS of (32) can therefore be lower-bounded by
\[ \sum_i p_i \frac{1}{\Delta_i} \int_{S_i} (x - \hat{x}_i)^2 \, dx \geq \sum_i p_i \frac{\Delta_i^2}{12}. \] (34)

To evaluate the second term on the RHS of (32), we introduce the piecewise-constant pdf
\[ f_X(x) = \sum_i \frac{p_i}{\Delta_i} \mathbbm{1}\{ x \in S_i \}, \quad x \in \mathbb{R}. \] (35)

With this, we can express the second term on the RHS of (32) as
\[ \sum_i \int_{S_i} \left[ \frac{p_i}{\Delta_i} - f_X(x) \right] (x - \hat{x}_i)^2 \, dx = \sum_i \int_{S_i} \left[ f_X^{(\Delta)}(x) - f_X(x) \right] (x - \hat{x}_i)^2 \, dx \] (36)
\[ \leq \alpha D \int \left| f_X^{(\Delta)}(x) - f_X(x) \right| \, dx \] (37)
since by (25) and the definition of the quantization regions, we have \( \sup_{x} \sup_{x \in S_{i}} (x - \hat{x}_{i})^{2} \leq \alpha D \) and \( \bigcup_{i} S_{i} = \mathbb{R} \).

By Lebesgue’s Differentiation Theorem, \( f_{X}^{(\Delta)} \) converges to \( f_{X} \) almost everywhere as \( \sup_{i} \Delta_{i} \to 0 \). Since \( f_{X}^{(\Delta)} \) and \( f_{X} \) both integrate to one, it follows from Scheffe’s Lemma [18, Th. 16.12] that

\[
\lim_{D \downarrow 0} \int \left| f_{X}^{(\Delta)}(x) - f_{X}(x) \right| \, dx = 0. \tag{38}
\]

Combining (34) and (37) with (32), and using that \( \mathbb{E} \left[ (X - \hat{X})^{2} \right] \leq D \), we obtain

\[
\sum_{i} p_{i} \Delta_{i}^{2} \leq 12D \left( 1 + \alpha \int \left| f_{X}^{(\Delta)}(x) - f_{X}(x) \right| \, dx \right) \tag{39}
\]

which together with (38) proves (31).

4 Proof of Theorem 1

The above back-of-the-envelope derivation directly generalizes to multi-dimensional sources and rth-power distortion. In order to prove Theorem 1, it would remain to show that any optimal quantizer satisfies (25). Unfortunately, for general sources this appears to be a difficult task. Indeed, the quantization regions of the optimal quantizer are difficult to characterize since the optimal quantizer (and hence the number of quantization regions together with their locations and volumes) changes with \( D \). To sidestep this problem, we replace (27) by an upper bound on \( h(X|\bar{X} = \bar{x}_{i}) \) that is based on the following lemma.

**Lemma 1.** Let \( f \) and \( g \) be arbitrary pdfs. If \( -\int f(x) \log f(x) \, dx \) is finite, then \( -\int f(x) \log g(x) \, dx \) exists and

\[
-\int f(x) \log f(x) \, dx \leq -\int f(x) \log g(x) \, dx \tag{40}
\]

with equality if, and only if, \( f(x) = g(x) \) almost everywhere.

**Proof.** See [19, Lemma 8.3.1]. \( \square \)

The inequality (40), which is a direct consequence of the information inequality, is sometimes referred to as *Gibbs’s inequality*. Lemma 1 is also reminiscent of Theorem 5.1 in [20], which provides an upper bound on the mutual information between a channel input \( X \) and a channel output \( Y \) and holds for general random variables. In fact, when \( Y \) is a real-valued random variable and the conditional distribution of \( Y \) given \( X \) is absolutely continuous with respect to the Lebesgue measure, then [20, Th. 5.1] essentially provides an upper bound on \( h(Y) \) that is of the form (40).

Lemma 1 allows us to upper-bound a differential entropy by replacing the true pdf \( f \) inside the logarithm by an auxiliary pdf \( g \). In order to upper-bound the conditional differential entropy \( h(X|\bar{X} = \bar{x}_{i}) \), we apply Lemma 1 with the conditional pdf

\[
g_{X|\bar{X}}(x|\bar{x}_{i}) = \begin{cases} 
\frac{1}{K_{i,\epsilon}}, & x \in B_{i,\epsilon} \\
\frac{1}{K_{i,\epsilon}} \frac{r}{\delta^{d} r}, e^{-\frac{|x - \bar{x}_{i}|^{r}}{\delta^{d} r}}, & x \in \bar{B}_{i,\epsilon} 
\end{cases} \tag{41}
\]

where

\[
B_{i,\epsilon} \triangleq \{ x \in S_{i} : \| x - \bar{x}_{i} \| \leq \epsilon \} \tag{42a}
\]

\[
\bar{B}_{i,\epsilon} \triangleq \{ x \in S_{i} : \| x - \bar{x}_{i} \| > \epsilon \} \tag{42b}
\]

\[
K_{i,\epsilon} \triangleq \Lambda_{i,\epsilon} + \frac{r}{\delta^{d} r} \int_{\bar{B}_{i,\epsilon}} e^{-\frac{|x - \bar{x}_{i}|^{r}}{\delta^{d} r}} \, dx \tag{42c}
\]

\( \Lambda_{i,\epsilon} \) denotes the Lebesgue measure of \( B_{i,\epsilon} \), and \( \delta \) and \( \epsilon \) are parameters to be specified later.

This conditional pdf of \( X \) given \( \bar{X} \) is uniform on a set of Volume \( \Lambda_{i,\epsilon} \) around \( \bar{x}_{i} \) and then decays exponentially. Intuitively, if \( \epsilon \) decays more slowly than \( \Delta_{i} \) as \( D \) tends to zero, then with high probability \( X \) lies in \( B_{i,\epsilon} \) and the upper bound obtained from Lemma 1 is essentially equivalent to (27) but with \( \Delta_{i} \) replaced
by $\Lambda_{i, \epsilon}$. Our choice of $g_{\mathbf{X} | \mathbf{X}}$ for $\mathbf{x} \in \mathcal{B}_{i, \epsilon}$ allows us to control the contribution of $\mathbf{x}$'s lying outside of $\mathcal{B}_{i, \epsilon}$. We then need to show that
\[
\lim_{\delta \downarrow 0} \frac{1}{D} \sum_{i} p_i \Lambda_{i, \epsilon} \leq V_d^{r/d} \left(1 + \frac{r}{d}\right)
\]
which for one-dimensional sources and quadratic distortion is equivalent to (31). By construction of $\mathcal{B}_{i, \epsilon}$, we have that $\sup_i \sup_{\mathbf{x} \in \mathcal{B}_{i, \epsilon}} \|\mathbf{x} - \mathbf{\tilde{x}}_i\| \leq \epsilon'$, so $\mathcal{B}_{i, \epsilon}$ satisfies (25) upon choosing $\epsilon' = D/\kappa$ (for some constant $\kappa$). The claim (43) follows therefore immediately from the steps (32)–(39). Thus, by using Lemma 1 together with (41), we can replace $\Delta_i$ (whose behavior as a function of $D$ is unknown) by $\Lambda_{i, \epsilon}$ (whose behavior can be controled by cleverly choosing $\epsilon$).

Before we set out to prove Theorem 1, we first provide a number of auxiliary results that we shall need throughout the proof. The proof of Theorem 1 is then given in Section 4.2.

### 4.1 Auxiliary Lemmas

**Lemma 2.** The normalizing constant $K_{i, \epsilon}$ is upper-bounded by
\[
K_{i, \epsilon} \leq \Lambda_{i, \epsilon} + dV_d D^{d/r} \Gamma \left(\frac{d}{r}, \epsilon \right) \leq \epsilon^d V_d + dV_d D^{d/r} \Gamma \left(\frac{d}{r}\right)
\]
where $\Gamma(\cdot, \cdot)$ denotes the upper incomplete Gamma function.

**Proof.** The first inequality in (44) follows from the definition of $K_{i, \epsilon}$ (42c) and by upper-bounding $\frac{r}{d/r} \int_{\mathcal{B}_{i, \epsilon}} e^{-\frac{||\mathbf{x} - \mathbf{\tilde{x}}_i||^r}{d}} d\mathbf{x}$. Indeed, since $\mathcal{B}_{i, \epsilon} \subseteq \{ \mathbf{x} \in \mathbb{R}^d : ||\mathbf{x} - \mathbf{\tilde{x}}_i|| > \epsilon \}$, it follows that
\[
\frac{r}{d/r} \int_{\mathcal{B}_{i, \epsilon}} e^{-\frac{||\mathbf{x} - \mathbf{\tilde{x}}_i||^r}{d}} d\mathbf{x} \leq \frac{r}{d/r} \int_{||\mathbf{x} - \mathbf{\tilde{x}}_i|| > \epsilon} e^{-\frac{||\mathbf{x} - \mathbf{\tilde{x}}_i||^r}{d}} d\mathbf{x} = dV_d \frac{r}{d/r} \int_{\rho > \epsilon} \rho^{d-1} e^{-\frac{\rho^r}{d}} d\rho = dV_d \frac{r}{d/r} \int_{\xi > \frac{\epsilon}{\rho^{r/d}}} \xi^{d/r} e^{-\xi} d\xi = dV_d \frac{r}{d/r} \Gamma \left(\frac{d}{r}, \epsilon \right)
\]
where the second step follows by writing $\mathbf{x} - \mathbf{\tilde{x}}_i$ in polar coordinates and by using that the surface area of the $d$-dimensional ball of radius $\rho = ||\mathbf{x} - \mathbf{\tilde{x}}_i||$ is $dV_d \rho^{d-1}$ (cf. [14, Eq. (10)]), and the third step follows by the change of variable $\xi = \rho^r/\rho^{r/d}$.

The second inequality in (44) follows by upper-bounding $\Lambda_{i, \epsilon} \leq \int_{||\mathbf{x} - \mathbf{\tilde{x}}_i|| \leq \epsilon} d\mathbf{x} = \epsilon^d V_d$ (cf. [14, Eq. (7)]) and $\Gamma(d/r, x) \leq \Gamma(d/r, x)$, $x \geq 0$.

**Lemma 3.** The set $\mathcal{B}_{i, \epsilon}$ satisfies
\[
\sum_i \Pr(\mathbf{X} \in \mathcal{B}_{i, \epsilon}) \leq \frac{D}{e^{\epsilon'}}
\]
\[
\sum_i \mathbb{E}[\|\mathbf{X} - \mathbf{\tilde{x}}_i\|^r \mathbb{1}\{ \mathbf{X} \in \mathcal{B}_{i, \epsilon} \}] \leq D.
\]

**Proof.** We first prove (46a). By (3), and since $\mathcal{S}_i \supseteq \mathcal{B}_{i, \epsilon}$ and $\|\mathbf{x} - \mathbf{\tilde{x}}_i\| > \epsilon$ for $\mathbf{x} \in \mathcal{B}_{i, \epsilon}$, we have
\[
D \geq \sum_i \int_{\mathcal{S}_i} f_X(\mathbf{x}) \|\mathbf{x} - \mathbf{\tilde{x}}_i\|^r d\mathbf{x} \geq \int_{\mathcal{B}_{i, \epsilon}} f_X(\mathbf{x}) \|\mathbf{x} - \mathbf{\tilde{x}}_i\|^r d\mathbf{x} \geq \int_{\mathcal{B}_{i, \epsilon}} f_X(\mathbf{x}) e^r d\mathbf{x}
\]
Using that $\epsilon$ neither depends on $i$ nor on $\mathbf{x}$, (46a) follows by diving both sides of (47) by $e^r$. 


To prove (46b) we use again $\bar{B}_{i,e} \subseteq S_i$ to obtain
\[
\sum_i E[\|X - \hat{x}_i\|^r I \{X \in \bar{B}_{i,e}\}] = \sum_i \int_{\bar{B}_{i,e}} f_X(x)\|x - \hat{x}_i\|^r \, dx \\
\leq E[\|X - \bar{X}\|^r] \\
\leq D \tag{48}
\]
where the first inequality follows because $\bigcup_i \bar{B}_{i,e} \subseteq \mathbb{R}^d$ and the second inequality follows from (3).

\[\square\]

4.2 Proof of Theorem 1

Expanding $I(X; \hat{X})$ as $h(X) - h(X|\hat{X})$, we obtain from (5) and (20) that the excess rate can be expressed as
\[
R_{r,d} = \lim_{D \downarrow 0} \left\{ \frac{d}{r} \log D + \frac{d}{r} \log \left( \frac{r}{d} \Gamma(1 + d/r) \right) - \sup_{q(\cdot)} h(X|\hat{X}) \right\}. \tag{49}
\]
To derive the lower bound (15) given in Theorem 1, it remains to show that
\[
\lim_{D \downarrow 0} \left\{ \sup_{q(\cdot)} h(X|\hat{X}) - \frac{d}{r} \log D \right\} \leq \frac{d}{r} \log \left( \frac{r}{d} \Gamma(1 + r/d) \right). \tag{50}
\]
To this end, we upper-bound the conditional differential entropy $h(X|\hat{X})$ using Lemma 1 together with (41). This yields for every $\hat{X} = \hat{x}_i$
\[
h(X|\hat{X} = \hat{x}_i) \leq \log K_{i,e} - E \left[ \log \left( \frac{r}{D \delta/r} e^{-\frac{r}{D \delta/r} \|x - \hat{x}_i\|^r} \right) I \{X \in \bar{B}_{i,e}\} \mid X \in S_i \right] \\
\leq \log \left( A_{i,e} + D \Gamma \left( \frac{d}{r} \frac{e^r}{D \delta} \right) \right) + \log \left( \frac{r}{D \delta/r} \right) \left| \frac{1}{D \delta} E[\|X - \hat{x}_i\|^r I \{X \in \bar{B}_{i,e}\} \mid X \in S_i] \right| \\
+ \frac{1}{D \delta} E[\|X - \hat{x}_i\|^r I \{X \in \bar{B}_{i,e}\} \mid X \in S_i] \tag{51}
\]
where the second inequality follows from the bound on $K_{i,e}$ presented in Lemma 2 and by upper-bounding $-\log(r/D \delta/r) \leq \log(r/D \delta/r)$. Averaging over $\hat{X}$ then yields
\[
h(X|\hat{X}) \leq \sum_i p_i \log \left( A_{i,e} + D \Gamma \left( \frac{d}{r} \frac{e^r}{D \delta} \right) \right) \\
+ \log \left( \frac{r}{D \delta/r} \right) \left| \sum_i \text{Pr}(X \in \bar{B}_{i,e}) \frac{1}{D \delta} \sum_i E[\|X - \hat{x}_i\|^r I \{X \in \bar{B}_{i,e}\}] \right|. \tag{52}
\]
By Lemma 3, this can be further upper-bounded by
\[
h(X|\hat{X}) \leq \sum_i p_i \log \left( A_{i,e} + D \Gamma \left( \frac{d}{r} \frac{e^r}{D \delta} \right) \right) + \log \frac{r}{D \delta/r} \left| D \frac{e^r}{D \delta} + \frac{1}{D \delta} \right|. \tag{53}
\]
We next choose
\[
e^r = \frac{D}{\kappa} \tag{54}
\]
for some $\kappa > 0$ that we will let tend to zero at the end of the proof. For ease of exposition, we do not always make this choice explicit in the notation but write $e^r$ or $D/\kappa$ depending on which is more convenient.

With this choice, the second term on the RHS of (53) becomes $\kappa \log(r/D \delta/r)$. To evaluate the first term on the RHS of (53), we express $p_i$ as
\[
p_i = \text{Pr}(X \in \bar{B}_{i,e}) + \text{Pr}(X \in \bar{B}_{i,e}) \tag{55}
\]
and define
\[
\varphi_e \triangleq \sum_i \text{Pr}(X \in \bar{B}_{i,e}). \tag{56}
\]
By Lemma 3, we have
\[ \varphi_\epsilon \leq \kappa \]  
(57)
which vanishes as we let \( \kappa \) tend to zero. With the above definition, and applying the second inequality in (44) (Lemma 2), we obtain for the first term on the RHS of (53) that
\[
\sum_i p_i \log \left( \Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) = \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left( \Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) \\
+ \sum_i \Pr(X \in \bar{\mathcal{B}}_{i,\epsilon}) \log \left( \Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) \\
\leq \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left( \Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) \\
+ \varphi_\epsilon \log \left( V_d D^{d/r} + dV_d D^{d/r} \Gamma \left( d/r \right) \right).
\]  
(58)
Using (57) and that \( \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) + \varphi_\epsilon = 1 \), (58) becomes
\[
\sum_i p_i \log \left( \Lambda_{i,\epsilon} + dV_d D^{d/r} \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) \\
\leq \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left( \frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) + \varphi_\epsilon \log \left( V_d \frac{1}{\kappa \delta} + dV_d \Gamma \left( d/r \right) \right) \\
\leq \frac{d}{r} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left( \frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right)^{r/d} + \epsilon \log \left( V_d \frac{1}{\kappa \delta} + dV_d \Gamma \left( d/r \right) \right) + \frac{d}{r} \log D. 
\]  
(59)
By Jensen’s inequality, the first term on the RHS of (59) is upper-bounded by
\[
\frac{d}{r} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \log \left( \frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right)^{r/d} \\
\leq (1 - \varphi_\epsilon) \frac{d}{r} \log \left( \frac{1}{1 - \varphi_\epsilon} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \left[ \frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right]^{r/d} \right). 
\]  
(60)
For \( r/d < 1 \), we have \( (x + \alpha)^{r/d} \leq x^{r/d} + \alpha^{r/d} \) for every \( x, \alpha \geq 0 \); for \( r/d \geq 1 \), the function \( x \mapsto (x^{d/r} + \alpha)^{r/d} \) is concave for every \( \alpha \geq 0 \). Consequently,
\[
\frac{1}{1 - \varphi_\epsilon} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \left[ \frac{\Lambda_{i,\epsilon}}{D^{d/r}} + dV_d \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right]^{r/d} \\
\leq \left\{ \left( \frac{1}{1 - \varphi_\epsilon} \sum_i \Pr(X \in \mathcal{B}_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}}{D^{d/r}} \right)^{d/r} + dV_d \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right\}^{r/d}, \\
r/d < 1
\]  
(61)
where the upper bound for \( r/d \geq 1 \) follows from Jensen’s inequality.

We next generalize (31), namely,
\[
\lim_{D \downarrow 0} \frac{1}{D} \sum_i p_i \Delta_i^2 \leq 12 
\]  
(62)
to the \( d \)-dimensional sets \( \mathcal{B}_{i,\epsilon} \) of Lebesgue measure \( \Lambda_{i,\epsilon} \). To this end, we follow essentially the steps (32)–(39) in Section 3 with \( \mathcal{S}_i \) replaced by \( \mathcal{B}_{i,\epsilon} \) and with \( \Delta_i \) replace by \( \Lambda_{i,\epsilon} \). However, (38) in Section 3 is based on Lebesgue’s differentiation theorem, which requires that the families of sets \( \mathcal{B}_{i,\epsilon} \) (parametrized by \( D \)) have \textit{bounded eccentricity}.\(^1\) Since \( \mathcal{B}_{i,\epsilon} \) is the intersection of \( \mathcal{S}_i \) with the \( d \)-dimensional ball of radius \( \epsilon \) centered at \( \hat{x}_i \),

\(^1\)A family \( \mathcal{F} \) of sets is said to have bounded eccentricity if there exists a constant \( c > 0 \) such that for every \( \mathcal{S} \in \mathcal{F} \) the Lebesgue measure of \( \mathcal{S} \) is not smaller than \( c \) times the volume of the smallest ball containing \( \mathcal{S} \).
cf. (42a), and since $S_i$ is arbitrary, the sets $B_{i,\epsilon}$ may not fulfill this condition. In the one-dimensional case, a sufficient condition would be that, for every distortion $D$, the quantization regions $S_i$ are convex. This can be assumed without loss of optimality, e.g., if the source has a finite second moment, but the assumption may be too restrictive for general sources [21]. Fortunately, the families of sets $B_{i,\epsilon}$ that have not bounded eccentricity can be disregarded without affecting the final result. The inequality (31) can therefore be generalized to the case at hand without imposing any additional constraints on the quantization regions $S_i$, $i \in \mathbb{Z}$ or the source pdf $f_X$. The result is stated in the following lemma.

**Lemma 4.** Let the sets $B_{i,\epsilon}$, $i \in \mathbb{Z}$ be defined in (42a), and let $\Lambda_{i,\epsilon}$, $i \in \mathbb{Z}$ denote the Lebesgue measures of these sets. Assume that $\epsilon^r = D/\kappa$. Then, for every $\kappa > 0$,

$$
\lim_{D \downarrow 0} \sup_{q(i)} \sum_i \Pr(X \in B_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}}{D} \leq V_d^{r/d}(1 + \frac{r}{d}).
$$

*(Proof.)* See appendix.

Combining Lemma 4 with (53)–(61), and bounding $0 \leq \varphi_\epsilon \leq \kappa$, we obtain that

$$
\lim_{D \downarrow 0} \left\{ \sup_{q(i)} h(X|\hat{X}) - \frac{d}{r} \log D \right\} \leq \frac{d}{r} \log \left( \frac{V_d^{r/d}(1 + r/d)}{1 - \kappa} + d^{r/d} V_d^{r/d} \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) + \kappa \log \left( \frac{V_d}{\kappa d/r} + d V_d \Gamma(d/r) \right) + \kappa \left| \log \frac{r}{\delta d/r} \right| + \frac{1}{\delta}, \quad \text{for } r/d < 1
$$

and

$$
\lim_{D \downarrow 0} \left\{ \sup_{q(i)} h(X|\hat{X}) - \frac{d}{r} \log D \right\} \leq \log \left( \frac{V_d(1 + r/d)^{d/r}}{(1 - \kappa)^{d/r}} + d V_d \Gamma \left( \frac{d}{r}, \frac{1}{\kappa \delta} \right) \right) + \kappa \log \left( \frac{V_d}{\kappa d/r} + d V_d \Gamma(d/r) \right) + \kappa \left| \log \frac{r}{\delta d/r} \right| + \frac{1}{\delta}, \quad \text{for } r/d \geq 1.
$$

Using that $\lim_{\xi \to \infty} \Gamma(d/r, \xi) = 0$ and $\lim_{\xi \to 0} \xi \log(\alpha/\xi^{d/r} + \beta) = 0$ (for any $\alpha, \beta > 0$), it follows by letting $\kappa \to 0$ that

$$
\lim_{D \downarrow 0} \left\{ \sup_{q(i)} h(X|\hat{X}) - \frac{d}{r} \log D \right\} \leq \frac{d}{r} \log \left( V_d^{r/d}(1 + r/d) \right) + \frac{1}{\delta}
$$

which in turn proves (50) upon letting $\delta \to \infty$. This concludes the proof of Theorem 1.

## 5 Balls versus Tessellating Polytopes

The lower bound (15) on the excess rate presented in Theorem 1 hinges on the fact that the distortion over the quantization region $S_i$, i.e., $\int_{S_i} \|x - \hat{x}\|^r \, dx$, is lower-bounded by the distortion over a ball around $\hat{x}_i$ with the same volume; cf. (74) with $B_{i,\epsilon}$ replaced by $S_i$ and with $\Lambda_{i,\epsilon}$ replaced by $\Delta_i$. Since the one-dimensional ball is an interval and, hence, tessellates $\mathbb{R}$, it follows that for one-dimensional sources the lower bound (15) is achieved by a tessellating quantizer, so in this case it is tight. However, it is expected that this is no longer true for multi-dimensional sources, since in general balls do not tessellate the space. In fact, it is unclear whether there exists any (possibly non-tessellating) vector quantizer that achieves (15) for multi-dimensional sources.

To assess the tightness of the obtained lower bound, we compare it numerically with the excess rates achievable by several lattice quantizers. To this end, we use Linder and Zeger’s upper bound for tessellating quantizers (22) together with the normalized second moments $t(\mathcal{P})$ of various lattice quantizers tabulated in [22, Table I]. In order to better compare our results with previous works, in this section we consider the *excess rate per dimension*, defined as $R_{r,d} \equiv R_{r,d}/d$. The excess rate per dimension is often considered, for example, in the analysis of quantization schemes that buffer $d$ consecutive symbols of a one-dimensional memoryless source and then quantize them using a $d$-dimensional vector quantizer.
Figure 1: Bounds on the excess rate per dimension $\bar{R}_{2,d}$ (in bits per source dimension) of a $d$-dimensional vector quantizer. Lower bounds correspond to (66) (Theorem 1) and Conway and Sloane’s conjecture in [22, Eq. (4)]. Circle markers indicate the upper bounds corresponding to scalar ($d = 1$), hexagonal ($d = 2$), and cuboctahedron ($d = 3$) quantization regions. Lattice upper bounds were obtained by applying to (67) the normalized second moments tabulated in [22, Table I].

For the sake of simplicity, we only consider quadratic distortion and the Euclidean norm. In this case, the lower bound (15) gives

$$\bar{R}_{2,d} \geq \frac{1}{2} \log \left( 2\pi e \frac{\Gamma\left(1 + d/2\right)}{\pi^{d/2}} \right).$$

(66)

Furthermore, the upper bound corresponding to tessellating quantizers (22) yields

$$\bar{R}_{2,d} \leq \frac{1}{2} \log \left( 2\pi e \frac{1}{d} \inf_{P} \ell(P) \right).$$

(67)

Another upper bound on $\bar{R}_{2,d}$ follows from an upper bound on $b_{r,d}$ in Zador’s theorem (9) that was presented in [4] (see also [23]). This upper bound is based on random coding arguments and yields for quadratic distortion and the Euclidean norm

$$\bar{R}_{2,d} \leq \frac{1}{2} \log \left( 2\pi e \frac{\Gamma\left(1 + 2/d\right)\Gamma\left(1 + d/2\right)}{\pi d} \right).$$

(68)

The bounds (66) and (68) demonstrate that $\bar{R}_{2,d}$ vanishes as $d$ tends to infinity. This is perhaps not very surprising, since the rate-distortion function $R(D)$ is essentially achieved by a vector quantizer whose dimension tends to infinity.

In Figure 1, we depict the bounds (66) and (68) as a function of the dimension $d$. We further show several achievability results based on lattice quantizers (67). The normalized second moments $\ell(P)$ corresponding to
these lattice quantizers were tabulated by Conway and Sloane in [22, Table I]. In fact, Figure 1 is essentially equal to Figure 1 in [22] with the difference that here we plot the excess rate per dimension whereas Conway and Sloane plot the normalized second moment. Specifically, we include the excess rates per dimension incurred by a (one-dimensional) uniform quantizer, by a (two-dimensional) hexagonal quantizer, and by the (three-dimensional) tessellating quantizer whose regions are cuboctahedrons. These quantizers correspond to cases that can be imposed in the analysis of high-resolution quantizers.

Our result holds for any $d$-dimensional memoryless source $X$ that satisfies $|h(X)| < \infty$ and $H(\lfloor X \rfloor) < \infty$. The presented proof is thus as general as the proof by Gray et al., and it is more general than the proof by Gish and Pierce. In fact, it has recently been shown in [7] that these conditions are necessary and sufficient for the Shannon lower bound to be asymptotically tight for vanishing distortion, and that $H(\lfloor X \rfloor) < \infty$ is a necessary condition for the rate-distortion function to be finite. Our result thus holds for the most general conditions that can be imposed in the analysis of high-resolution quantizers.

6 Conclusions

The nonnegativity of relative entropy implies that the differential entropy of a random variable $X$ with pdf $f$ is upper-bounded by $-\mathbb{E}[\log g(X)]$ for any arbitrary pdf $g$. Using this inequality with a cleverly chosen $g$, we derived a lower bound on the asymptotic excess rate of entropy-constrained scalar quantization. Specialized to the one-dimensional case and quadratic distortion, this bound coincides with the excess rate obtained by Gish and Pierce in [3], and by Gray et al. in [9] particularized for scalar quantizers. The proposed derivation thus recovers the well-known result that uniform quantizers are asymptotically optimal as the allowed distortion vanishes.

While the presented bound is tight for scalar sources, it is unclear whether the same is true for multi-dimensional sources. Indeed, its derivation hinges on the fact that the distortion over the quantization region $S_i$ is lower-bounded by the distortion over a ball around $\hat{x}_i$ with the same volume, cf. (74). Since the one-dimensional ball is an interval and, hence, tessellates $\mathbb{R}$, it follows that for one-dimensional sources the converse bound (15) is achieved by a tessellating quantizer (which in this case is the uniform quantizer). However, it is expected that this is no longer true for multi-dimensional sources, since in general balls do not tessellate the space. It is yet unclear whether there exists any (possibly non-tessellating) vector quantizer that achieves our converse bound for multi-dimensional sources.
A  Proof of Lemma 4

To prove Lemma 4, we first fix an arbitrary constant $\eta > 0$ and divide the indices $i$ according to whether $\Lambda_{i,\epsilon} \geq \eta V_d \epsilon^d$ or not. Specifically, let

$$\mathcal{I} \triangleq \{ i \in \mathbb{Z} : \Lambda_{i,\epsilon} \geq \eta V_d \epsilon^d \}$$

(70)

and divide the sum on the LHS of (63) into

$$\sum_i \Pr(X \in B_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} = \sum_{i \in \mathcal{I}} \Pr(X \in B_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} + \sum_{i \in \mathcal{I}^c} \Pr(X \in B_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D}$$

(71)

where $\mathcal{I}^c$ denotes the complement of $\mathcal{I}$. For every $i \in \mathcal{I}^c$ we have $\Lambda_{i,\epsilon} < \eta V_d \epsilon^d$, so the second sum on the RHS of (71) can be upper-bounded as

$$\sum_{i \in \mathcal{I}^c} \Pr(X \in B_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \leq \eta^{r/d} V_d^{r/d} \epsilon^r \sum_{i \in \mathcal{I}^c} \Pr(X \in B_{i,\epsilon})$$

$$\leq \eta^{r/d} V_d^{r/d} \frac{\epsilon}{\kappa}$$

(72)

where the second step follows because $\epsilon^r = D/\kappa$ and because, by definition, the sets $B_{i,\epsilon}$ are disjoint, so the sum of the probabilities $\Pr(X \in B_{i,\epsilon})$ is equal to the probability of $\bigcup_{i \in \mathcal{I}} B_{i,\epsilon}$, which is upper-bounded by 1.

To upper-bound the first sum on the RHS of (71), we begin by lower-bounding $E \left[ \|X - \hat{X}\|^r \right]$ as

$$E \left[ \|X - \hat{X}\|^r \right] = \sum_i \int_{S_i} f_X(x) \|x - \hat{x}_i\|^r \, dx$$

$$\geq \sum_{i \in \mathcal{I}} \int_{B_{i,\epsilon}} f_X(x) \|x - \hat{x}_i\|^r \, dx$$

$$= \sum_{i \in \mathcal{I}} \Pr(X \in B_{i,\epsilon}) \frac{1}{\Lambda_{i,\epsilon}} \int_{B_{i,\epsilon}} \|x - \hat{x}_i\|^r \, dx$$

$$- \sum_{i \in \mathcal{I}} \int_{B_{i,\epsilon}} \left[ \frac{1}{\Lambda_{i,\epsilon}} \Pr(X \in B_{i,\epsilon}) - f_X(x) \right] \|x - \hat{x}_i\|^r \, dx. \quad (73)$$

The region $B_{i,\epsilon}$ of volume $\Lambda_{i,\epsilon}$ that minimizes $\int_{B_{i,\epsilon}} \|x - \hat{x}\|^r \, dx$ is a ball around $\hat{x}$. We thus have the lower bound [14, Sec. III]

$$\frac{1}{\Lambda_{i,\epsilon}} \int_{B_{i,\epsilon}} \|x - \hat{x}_i\|^r \, dx \geq \frac{d}{d + r} \frac{\Lambda_{i,\epsilon}^{r/d}}{V_d^{r/d}}$$

(74)

which yields for the first term on the RHS of (73)

$$\sum_{i \in \mathcal{I}} \Pr(X \in B_{i,\epsilon}) \frac{1}{\Lambda_{i,\epsilon}} \int_{B_{i,\epsilon}} \|x - \hat{x}_i\|^r \, dx \geq \sum_{i \in \mathcal{I}} \Pr(X \in B_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{V_d^{r/d} (1 + r/d)}.$$  \quad (75)

Multiplying both sides of (73) by $V_d^{r/d} (1 + r/d)/D$, applying (75) to (73), and using that $E \left[ \|X - \hat{X}\|^r \right] \leq D$, we obtain

$$\sum_{i \in \mathcal{I}} \Pr(X \in B_{i,\epsilon}) \frac{\Lambda_{i,\epsilon}^{r/d}}{D} \leq V_d^{r/d} \left( 1 + \frac{r}{d} \right) \left( 1 + \frac{1}{D} \sum_{i \in \mathcal{I}} \int_{B_{i,\epsilon}} \left[ \frac{1}{\Lambda_{i,\epsilon}} \Pr(X \in B_{i,\epsilon}) - f_X(x) \right] \|x - \hat{x}_i\|^r \, dx \right). \quad (76)$$

We next introduce the pdf

$$f_X^{(\Lambda)}(x) \triangleq \sum_{i \in \mathcal{I}} \frac{1}{\Lambda_{i,\epsilon}} \Pr(X \in B_{i,\epsilon}) \mathbb{I} \{ x \in B_{i,\epsilon} \} + f_X(x) \left[ \sum_{i \in \mathcal{I}} \mathbb{I} \{ x \in B_{i,\epsilon} \} + \sum_{i \in \mathcal{I}^c} \mathbb{I} \{ x \in S_i \} \right]. \quad x \in \mathbb{R}. \quad (77)$$
This allows us to write

\[
\sum_{i \in I} \int_{B_{i, \epsilon}} \frac{1}{\Lambda_{i, \epsilon}^{\nu}} \Pr(X \in B_{i, \epsilon}) - f_X(x) \|x - \hat{x}_i\|^r \, dx = \sum_{i \in I} \int_{S_i} [f_X^{(A)}(x) - f_X(x)] \|x - \hat{x}_i\|^r \, dx
\]  

(78)

since \( f_X^{(A)}(x) = \frac{1}{\Lambda_{i, \epsilon}^{\nu}} \Pr(X \in B_{i, \epsilon}) \) for \( x \in B_{i, \epsilon}, \ i \in I \) and \( f_X^{(A)}(x) - f_X(x) = 0 \) otherwise.

We next use that \( \|x - \hat{x}_i\|^r \leq \epsilon^r = D/\kappa \) for \( x \in B_{i, \epsilon}, \ i \in I \) and \( f_X^{(A)}(x) - f_X(x) = 0 \) otherwise to upper-bound

\[
\left| \sum_{i} \int_{S_i} [f_X^{(A)}(x) - f_X(x)] \|x - \hat{x}_i\|^r \, dx \right| \leq \frac{D}{\kappa} \int \left| f_X^{(A)}(x) - f_X(x) \right| \, dx.
\]  

(79)

Combining this upper bound with (76) yields

\[
\sum_{i \in I} \Pr(X \in B_{i, \epsilon}) \frac{\Lambda_{i, \epsilon}^{\nu}}{D} \leq V_d^{\nu/d} \left( 1 + \frac{r}{d} \right) \left( 1 + \frac{1}{\kappa} \right) \int \left| f_X^{(A)}(x) - f_X(x) \right| \, dx + \eta^{r/d} V_d^{r/d}. \tag{80}
\]

which in turn yields together with (71) and (72)

\[
\sum_{i} \Pr(X \in B_{i, \epsilon}) \frac{\Lambda_{i, \epsilon}^{\nu}}{D} \leq V_d^{\nu/d} \left( 1 + \frac{r}{d} \right) \left( 1 + \frac{1}{\kappa} \right) \int \left| f_X^{(A)}(x) - f_X(x) \right| \, dx + \eta^{r/d} V_d^{r/d}. \tag{81}
\]

We next show that, for every \( \eta > 0 \),

\[
\limsup_{D \downarrow 0, \ q(\cdot)} \int \left| f_X^{(A)}(x) - f_X(x) \right| \, dx = 0. \tag{82}
\]

It then follows that

\[
\limsup_{D \downarrow 0, \ q(\cdot)} \sum_{i} \Pr(X \in B_{i, \epsilon}) \frac{\Lambda_{i, \epsilon}^{\nu}}{D} \leq V_d^{\nu/d} \left( 1 + \frac{r}{d} \right) + \eta^{r/d} V_d^{r/d}. \tag{83}
\]

which proves Lemma 4 upon letting \( \eta \) tend to zero from above.

It thus remains to prove (82). By definition, \( f_X^{(A)} \) differs from \( f_X \) only when \( x \in B_{i, \epsilon}, \ i \in I \). Since the family of sets \( B_{i, \epsilon}, \ i \in I \) (parametrized by \( D \)) has bounded eccentricity, it follows from Lebesgue’s differentiation theorem that \( f_X^{(A)} \) converges to \( f_X \) almost everywhere as \( D \) (and hence also \( \epsilon \)) tends to zero, which by Scheffe’s lemma then implies (82). However, compared to the standard setting under which Lebesgue’s differentiation theorem is proven, our setting is slightly more complicated, since as \( D \) tends to zero not only the diameters of the sets \( B_{i, \epsilon} \) decay, but also their locations in \( \mathbb{R}^d \) may change. For completeness, we therefore provide all the steps, even though they follow closely the standard proof of the Lebesgue differentiation theorem.

We first note that, since the integral in (82) is bounded above by 2, its supremum is finite and for every \( \nu > 0 \) there exists a sequence of quantizers (parametrized by \( D \)) such that

\[
\lim_{D \downarrow 0} \int \left| f_X^{(A)}(x) - f_X(x) \right| \, dx \geq \limsup_{D \downarrow 0, \ q(\cdot)} \int \left| f_X^{(A)}(x) - f_X(x) \right| \, dx - \nu. \tag{84}
\]

Since the integral in (82) is nonnegative, and since \( \nu > 0 \) is arbitrary, to prove (82) it therefore suffices to show that, for any sequence of quantizers (parametrized by \( D \)),

\[
\lim_{D \downarrow 0} \int \left| f_X^{(A)}(x) - f_X(x) \right| \, dx = 0. \tag{85}
\]

Specifically, we shall show that for any sequence of quantizers (parametrized by \( D \))

\[
\mu \left( \left\{ x \in \mathbb{R}^d : \lim_{D \downarrow 0} \left| f_X^{(A)}(x) - f_X(x) \right| > 2\xi \right\} \right) = 0, \quad \text{for every } \xi > 0 \tag{86}
\]
where \( \mu(\cdot) \) denotes the Lebesgue measure. It then follows that \( f_X^{(A)} \) converges to \( f_X \) almost everywhere as \( D \) tends to zero since
\[
\left\{ x \in \mathbb{R}^d : \lim_{D \downarrow 0} \left| f_X^{(A)}(x) - f_X(x) \right| > 0 \right\} = \bigcup_{i=1}^{\infty} \left\{ x \in \mathbb{R}^d : \lim_{D \downarrow 0} \left| f_X^{(A)}(x) - f_X(x) \right| > \frac{1}{i} \right\}
\]
and the countable union of sets of measure zero has measure zero. By Scheffe’s lemma, almost everywhere convergence of \( f_X^{(A)} \) to \( f_X \) implies (85), which together with (84) proves the desired result (82).

We thus set out to prove (86). By the definition of \( f_X^{(A)} \) and the triangle inequality,
\[
\left| f_X^{(A)}(x) - f_X(x) \right| \leq \sum_{i \in \mathcal{I}} \frac{1}{\Lambda_{i,e}} \Pr(X \in B_{i,e}) - f_X(x) \mathbf{1}\{x \in B_{i,e}\}.
\]  
(88)

We next approximate \( \frac{1}{\Lambda_{i,e}} \Pr(X \in B_{i,e}) \) by replacing \( f_X \) by a continuous function \( g \). Indeed, since \( f_X \) is integrable, for every \( \varepsilon > 0 \) there exists a continuous function \( g \) such that [24, Th. 2.4.14, p. 92]
\[
\int |f_X(x) - g(x)| \, dx \leq \varepsilon.
\]  
(89)

It then follows that, for every \( x \in B_{i,e} \),
\[
\left| \frac{1}{\Lambda_{i,e}} \Pr(X \in B_{i,e}) - f_X(x) \right| = \frac{1}{\Lambda_{i,e}} \int_{B_{i,e}} f_X(y) \, dy - f_X(x)
\leq \frac{1}{\Lambda_{i,e}} \int_{B_{i,e}} |g(y) - g(x)| \, dy \quad + \quad \frac{1}{\Lambda_{i,e}} \int_{B_{i,e}} \left| f_X(y) - g(y) \right| \, dy + \left| f_X(x) - g(x) \right|
\]  
(90)

Let \( B(c, \rho) \triangleq \{ x \in \mathbb{R}^d : \| x - c \| \leq \rho \} \) denote the \( d \)-dimensional ball of radius \( \rho \) centered at \( c \). Note that \( \mu(B(\hat{x}_i, \varepsilon)) = V_d \varepsilon^d \). For every \( x \in B_{i,e} \) and \( i \in \mathcal{I} \), the second term on the RHS of (90) can be upper-bounded by
\[
\frac{1}{\Lambda_{i,e}} \int_{B_{i,e}} \left| f_X(y) - g(y) \right| \, dy \leq \frac{1}{\eta \mu(B(\hat{x}_i, \varepsilon))} \int_{B(\hat{x}_i, \varepsilon)} \left| f_X(y) - g(y) \right| \, dy
\leq \frac{2^d}{\eta} \frac{1}{\mu(B(x, 2\varepsilon))} \int_{B(x, 2\varepsilon)} \left| f_X(y) - g(y) \right| \, dy
\leq \frac{2^d}{\eta} (f_X - g)^*(x)
\]  
(91)

where \( (f_X - g)^* \) denotes the Hardy-Littlewood maximal function for \( f_X - g \), i.e.,
\[
(f_X - g)^*(x) \triangleq \sup_{\rho > 0} \frac{1}{\mu(B(x, \rho))} \int_{B(x, \rho)} \left| f_X(y) - g(y) \right| \, dy, \quad x \in \mathbb{R}^d.
\]  
(92)

In (91), we have used that, for every \( x \in B_{i,e} \) \( i \in \mathcal{I} \), we have \( B_{i,e} \subseteq B(\hat{x}_i, \varepsilon) \subseteq B(x, 2\varepsilon) \) and \( \Lambda_{i,e} \geq \eta \mu(B(\hat{x}_i, \varepsilon)) = 2^{-d} \mu(B(x, 2\varepsilon)) \).

Combining (90) and (91) with (88), we obtain for every \( x \in \mathbb{R}^d \)
\[
\left| f_X^{(A)}(x) - f_X(x) \right| \leq \sum_{i \in \mathcal{I}} \frac{1}{\Lambda_{i,e}} \int_{B_{i,e}} g(y) \, dy - g(x) \mathbf{1}\{x \in B_{i,e}\}
\quad + \quad \sum_{i \in \mathcal{I}} \frac{2^d}{\eta} (f_X - g)^*(x) \mathbf{1}\{x \in B_{i,e}\} + \sum_{i \in \mathcal{I}} \left| f_X(x) - g(x) \right| \mathbf{1}\{x \in B_{i,e}\}
\leq \sum_{i \in \mathcal{I}} \frac{1}{\Lambda_{i,e}} \int_{B_{i,e}} g(y) \, dy - g(x) \mathbf{1}\{x \in B_{i,e}\} + \frac{2^d}{\eta} (f_X - g)^*(x) + \left| f_X(x) - g(x) \right|
\]
(93)
since the sets $B_{i,\epsilon}$ are disjoint. The second and third term on the RHS of (93) are independent of $D$ and $q(\cdot)$. The first term on the RHS of (93) vanishes as $D$ tends to zero for any sequence of quantizers. Indeed, the continuity of $g$ implies that for every $\vartheta > 0$ and $x \in \mathbb{R}^d$ there exists an $\epsilon > 0$ such that

$$|g(y) - g(x)| \leq \vartheta, \quad \text{for } ||y - x|| \leq 2\epsilon. \quad (94)$$

Since $x, y \in B_{i,\epsilon}$ satisfy $||x - y|| \leq 2\epsilon$, it follows that for every $\vartheta > 0$ and $x \in \mathbb{R}^d$ there exists an $\epsilon > 0$ such that

$$\left| \frac{1}{\Lambda_{i,\epsilon}} \int_{B_{i,\epsilon}} g(y) \, dy - g(x) \right| 1 \{x \in B_{i,\epsilon}\} \leq \vartheta 1 \{x \in B_{i,\epsilon}\}, \quad \epsilon \leq \epsilon_0. \quad (95)$$

Using that the sets $B_{i,\epsilon}, i \in \mathcal{I}$ are disjoint, we conclude that for every $\vartheta > 0$ and $x \in \mathbb{R}^d$ there exists an $\epsilon > 0$ such that

$$\sum_{i \in \mathcal{I}} \left| \frac{1}{\Lambda_{i,\epsilon}} \int_{B_{i,\epsilon}} g(y) \, dy - g(x) \right| 1 \{x \in B_{i,\epsilon}\} \leq \vartheta, \quad \epsilon \leq \epsilon_0. \quad (96)$$

Since $\vartheta > 0$ is arbitrary and $\epsilon$ vanishes as $D \to 0$, this implies that for every $x \in \mathbb{R}^d$ and any sequence of quantizers

$$\lim_{D \to 0} \sum_{i \in \mathcal{I}} \left| \frac{1}{\Lambda_{i,\epsilon}} \int_{B_{i,\epsilon}} g(y) \, dy - g(x) \right| 1 \{x \in B_{i,\epsilon}\} = 0. \quad (97)$$

We conclude the proof of Lemma 4 by applying (93) and (97) to upper-bound the measure on the LHS of (86). Indeed, we have

$$\mu \left( \left\{ x \in \mathbb{R}^d : \lim_{D \to 0} \left| f_X^{(A)}(x) - f_X(x) \right| > 2\xi \right\} \right)$$

$$\leq \mu \left( \left\{ x \in \mathbb{R}^d : \frac{2^d}{\eta} (f_X - g)^*(x) + |f_X(x) - g(x)| > 2\xi \right\} \right)$$

$$\leq \mu \left( \left\{ x \in \mathbb{R}^d : \frac{2^d}{\eta} (f_X - g)^*(x) > \xi \right\} \right) + \mu \left( \left\{ x \in \mathbb{R}^d : |f_X(x) - g(x)| > \xi \right\} \right). \quad (98)$$

The first term on the RHS of (98) can be upper-bounded by using the Hardy-Littlewood maximal inequality and (89):

$$\mu \left( \left\{ x \in \mathbb{R}^d : \frac{2^d}{\eta} (f_X - g)^*(x) > \xi \right\} \right) \leq \frac{2^d}{\eta} \int |f_X(x) - g(x)| \, dx \leq \frac{2^d\alpha_d}{\eta \xi} \varepsilon. \quad (99)$$

for some constant $\alpha_d$ that only depends on $d$. Likewise, the second term on the RHS of (98) can be upper-bounded using Chebyshev’s inequality [24, (4.10.7), p. 192] and (89):

$$\mu \left( \left\{ x \in \mathbb{R}^d : |f_X(x) - g(x)| > \xi \right\} \right) \leq \frac{1}{\xi} \int |f_X(x) - g(x)| \, dx \leq \frac{\varepsilon}{\xi}. \quad (100)$$

Combining (99) and (100) with (98), it follows that

$$\mu \left( \left\{ x \in \mathbb{R}^d : \lim_{D \to 0} \left| f_X^{(A)}(x) - f_X(x) \right| > 2\xi \right\} \right) \leq \frac{1 + 2^d\alpha_d/\eta}{\xi} \varepsilon. \quad (101)$$

Since $\varepsilon > 0$ is arbitrary, this proves (86), which in turn demonstrates that $f_X^{(A)}$ converges to $f_X$ almost everywhere as $D$ tends to zero. This was the last step required to prove Lemma 4.

**Acknowledgment**

Stimulating discussions with Albert Guillén i Fàbregas, Tamás Linder, Alfonso Martinez, and Ram Zamir are gratefully acknowledged. The authors further wish to thank Giuseppe Durisi for calling their attention to reference [10].
References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Techn. J.*, vol. 27, pp. 379–423 and 623–656, July and Oct. 1948.

[2] T. Berger, *Rate Distortion Theory: Mathematical Basis for Data Compression*, ser. Electrical Engineering Series. Prentice Hall, 1971.

[3] H. Gish and J. N. Pierce, “Asymptotically efficient quantizing,” *IEEE Trans. Inform. Theory*, vol. 14, no. 5, pp. 676–683, Sept. 1968.

[4] P. L. Zador, “Topics in the asymptotic quantization of continuous random variables,” Bell Laboratories, Tech. Rep., 1966.

[5] Y. N. Linkov, “Evaluation of epsilon entropy of random variables for small epsilon,” *Problemy Peredachi Informatsii (Problems of Inform. Transm.)*, vol. 1, pp. 12–18, 1965.

[6] T. Linder and R. Zamir, “On the asymptotic tightness of the Shannon lower bound,” *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 2026–2031, Nov. 1994.

[7] T. Koch, “The Shannon lower bound is asymptotically tight,” Feb. 2016, subm. to *IEEE Trans. Inform. Theory*. [Online]. Available: http://arxiv.org/abs/1504.08245v4

[8] ——, “A necessary and sufficient condition for the asymptotic tightness of the Shannon lower bound,” in *Proc. IZS*, Zurich, Switzerland, Mar. 2–4, 2016.

[9] R. M. Gray, T. Linder, and J. Li, “A Langrangian formulation of Zador’s entropy-constrained quantization theorem,” *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 695–707, Mar. 2002.

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

[11] A. Rényi, “On the dimension and entropy of probability distributions,” *Acta Mathematica Hungarica*, vol. 10, no. 1–2, Mar. 1959.

[12] T. Kawabata and A. Dembo, “The rate-distortion dimension of sets and measures,” *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1564–1572, Sept. 1994.

[13] T. Linder and K. Zeger, “Asymptotic entropy-constrained performance of tessellating and universal randomized lattice quantization,” *IEEE Trans. Inform. Theory*, vol. 40, no. 2, pp. 575–579, Mar. 1994.

[14] Y. Yamada, S. Tazaki, and R. M. Gray, “Asymptotic performance of block quantizers with difference distortion measures,” *IEEE Trans. Inform. Theory*, vol. 26, pp. 6–14, Jan. 1980.

[15] V. Kostina, “Data compression with low distortion and finite blocklength,” in *Proc. 53rd Allerton Conf. Comm., Contr. and Comp.*, Allerton H., Monticello, Il, Sep. 30 – Oct. 2, 2015.

[16] ——, “Data compression with low distortion and finite blocklength,” Oct. 2015. [Online]. Available: http://arxiv.org/abs/1510.02190

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. John Wiley & Sons, 1991.

[18] P. Billingsley, *Probability and Measure*, 3rd ed., ser. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, 1995.

[19] R. B. Ash, *Information Theory*, ser. Dover Books on Mathematics. Mineola, NY: Courier Dover Publications, 1990.

[20] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
[21] A. György and T. Linder, “On the structure of optimal entropy-constrained scalar quantizers,” IEEE Trans. Inform. Theory, vol. 48, no. 2, pp. 416–427, Feb. 2002.

[22] J. H. Conway and N. J. A. Sloane, “A lower bound on the average error of vector quantizers,” IEEE Trans. Inform. Theory, vol. 31, no. 1, pp. 106–109, Jan. 1985.

[23] P. L. Zador, “Asymptotic quantization error of continuous signals and the quantization dimension,” IEEE Trans. Inform. Theory, vol. 28, no. 2, pp. 139–149, Mar. 1982.

[24] R. B. Ash and C. A. Doléans-Dade, Probability and Measure Theory, 2nd ed. Elsevier/Academic Press, 2000.