On the asymptotic tightness of the Shannon lower bound

Tamás Linder * Ram Zamir †

December 18, 1997

Abstract

New results are proved on the convergence of the Shannon lower bound (SLB) to the rate distortion function as the distortion decreases to zero. The key convergence result is proved using a fundamental property of informational divergence. As a corollary, it is shown that the SLB is asymptotically tight for norm-based distortions, when the source vector has a finite differential entropy and a finite αth moment for some $\alpha > 0$, with respect to the given norm. Moreover, we derive a theorem of Linkov on the asymptotic tightness of the SLB for general difference distortion measures with more relaxed conditions on the source density. We also show that the SLB relative to a stationary source and single letter difference distortion is asymptotically tight under very weak assumptions on the source distribution.

Key words: rate distortion theory, Shannon lower bound, difference distortion measures, stationary sources

*T. Linder is with the Coordinated Science Laboratory, University of Illinois, Urbana-Champaign, IL 61801. He is on leave from the Dept. of Communications, Technical University of Budapest, 1521 Stoczek u. 2, Budapest, Hungary. (email: linder@tornado.csl.uiuc.edu) His research was supported in part by the National Science Foundation under Grant No. NCR-91-57770.

†R. Zamir is with Tel Aviv University, Dept. of Electrical Eng.-Systems, Tel-Aviv 69978, Israel. (email: zamir@eng.tau.ac.il) His research was supported in part by the Wolfson Research Awards, administrated by the Israeli Academy of Science and Humanities at Tel-Aviv University.

1 Introduction

The *Shannon lower bound* or SLB for difference distortion measures (Shannon [15], Berger [2]) is one of the few tools that make possible the explicit evaluation of rate distortion functions. This lower bound actually achieves the rate distortion function only for some special source distributions. However, Linkov [13] showed (see also [2]), that for vector sources with sufficiently "nice" densities, and distortion measures satisfying some regularity conditions the lower bound is asymptotically, i.e., when the distortion goes to zero, tight. This result makes it possible to estimate the otherwise unknown behavior of the rate distortion function, and has fundamental importance in analyzing the rate and distortion redundancies of high resolution quantizers (see e.g. Gish and Pierce [9], Gersho [8], Yamada *et al.* [17]).

Binia *et al.* [4] gave lower and upper bounds on the squared error ϵ-entropy of second-order stochastic processes in terms of their relative entropy with Gaussian processes. Their result implies, in particular, that the SLB is asymptotically tight for squared distortion for vector sources with densities and finite differential entropy and second moment.

In asymptotic quantization theory often more general source distributions or distortion measures are considered than the ones covered by the above results. Bucklew and Wise [5] established the rate of decrease of the minimum rth power distortion of a k-dimensional vector quantizer with N codevectors, when N grows to infinity. Their only condition on the distribution of the source vector X was that $E\|X\|^{r+\epsilon} < \infty$ for some $\epsilon > 0$. Zamir and Feder [18] determined the precise asymptotics of the entropy rate of randomized lattice quantizers for all sources with densities. With the same conditions Linder and Zeger [12] gave the exact asymptotics of the entropy rate of tessellating vector quantizers of small rth power distortion. The Shannon lower bound is an extremely useful tool to relate these asymptotic distortion and rate formulas to the theoretical rate-distortion limits. However, the existing results are not applicable since they give asymptotically tight bounds only for squared distortion or when the source density is from a restricted class.

In this paper we prove new results on the asymptotic tightness of the SLB for large classes of difference distortion measures and general source densities. Theorem 1 provides a set of very general conditions under which the SLB is asymptotically tight. The proof of this result is based on a fundamental property of informational divergence (relative entropy). Two corollaries of the main result extend the existing results on the asymptotic tightness of the SLB. First we consider distortion measures of the form $\|x-y\|^r$, where $\|\cdot\|$ is an arbitrary norm on $\mathcal{R}^k$ and $r > 0$. Corollary 1 shows that the SLB converges to the rate distortion function, if the source vector X has a finite differential entropy and there exists an $\alpha > 0$, such that $E\|X\|^\alpha < \infty$. This means that we can have a tight asymptotic bound on the rate distortion function even if there exists no quantizer with a finite number of codevectors and of finite distortion, i.e., when $E\|X\|^r = \infty$.

Next we deal with a more general class of difference distortion measures. In Corollary 2 we extend Linkov's result for a less restrictive class of source densities, namely to source densities with finite differential entropy and finite first moment with respect to the given distortion measure. Moreover, the proof given is actually simpler than Linkov's original proof.

In Section 4 the differences between Linkov's theorem and our result are discussed. The fact that we could get rid of Linkov's tail condition on the source density allows us to state in Corollary 3 the asymptotic tightness of the SLB for stationary sources with a finite differential entropy rate relative to a large class of single letter difference distortion measures. The main point is that our conditions do not involve the multidimensional distributions of the stationary source, only the *one dimensional marginal* of the source distribution has to satisfy a moment-type condition.

2 Preliminaries

Let X be an $\mathcal{R}^k$ valued random vector with a k-dimensional probability density f. The distortion measure we consider is a so called *difference distortion measure*, that is, the measure of dissimilarity between the vectors $x, y \in \mathcal{R}^k$ is given as $\rho(x-y)$, where $\rho : \mathcal{R}^k \to [0, \infty)$ is a (Borel) measurable nonnegative function. The *rate distortion* function $R(D)$ of X with respect to ρ is given for $D \geq 0$ by

$$R(D) = \inf\{I(X;Y) : \; E\rho(X-Y) \leq D\}. \tag{1}$$

Here $I(X;Y)$ denotes the mutual information between X and Y, and the infimum is taken over all k-dimensional random vectors Y such that the joint distribution of the pair (X,Y) satisfies $E\rho(X-Y) \leq D$. Note that $R(D)$ is not normalized by the blocklength; it is measured in nats/block. $R(D)$ can be thought of as the rate distortion function of an i.i.d. vector source producing vectors with the distribution of X. Alternatively, when X consists of the first k coordinates of a real source $X_1, X_2, \ldots$, then $(1/k)R(D)$ is called the kth order rate distortion function of the source.

The Shannon lower bound $R_L(D) \leq R(D)$ is given for $D > 0$ by

$$R_L(D) = h(f) + \log a(D) - Ds(D), \tag{2}$$

where $h(f) = -\int f \log f$ is the differential entropy of X, and $a(D) > 0$, $s(D) > 0$ satisfy the following two equations

$$a(D) \int_{\mathcal{R}^k} e^{-s(D)\rho(x)}\, dx \;=\; 1 \tag{3}$$

$$a(D) \int_{\mathcal{R}^k} \rho(x) e^{-s(D)\rho(x)}\, dx \;=\; D. \tag{4}$$

These equations mean that for each $D > 0$ there exists a probability density g_D of the form $g_D(x) = Ae^{-B\rho(x)}$ which satisfies $\int \rho(x) g_D(x)\, dx = D$.

It is easy to check that (2) can be written as

$$R_L(D) = h(f) - h(g_D), \tag{5}$$

where

$$g_D(x) = a(D) e^{-s(D)\rho(x)} \tag{6}$$

is a probability density function by (3). Throughout the text we will use base e logarithm, and $h(Z)$ as well as $h(f_Z)$ will denote the differential entropy of a random vector Z with density f_Z.

The Shannon lower bound can be derived in different ways. In [2] the parametric representation of $R(D)$ is used to obtain (2) for scalar sources. $R_L(D) \le R(D)$ always holds if $h(f) < \infty$ and (3) and (4) have a solution. The generalization to vector sources is straightforward. Gray [10] based the development on Csiszár's general result [6] on the parametric representation of $R(D)$ for abstract alphabets.

An alternative derivation (see e.g. Linkov [13]) results in (5) directly, characterizing g_D as the probability density having the largest differential entropy among the family of probability densities $\{g : \int \rho(x)g(x)\,dx \le D\}$.

Of course, the above form of $R_L(D)$ makes sense only if the integrals in question are finite and the system of equations (3) and (4) has solutions for $a(D)$ and $s(D)$. Linkov [13] established precise conditions on f and ρ under which (2) is valid (see Section 4 for details) and proved via a tedious argument that with some additional tail condition on the source density

$$\lim_{D\to 0} \left(R(D) - R_L(D)\right) = 0. \tag{7}$$

This result is generally referred to as Linkov's theorem. (7) was independently derived by Gerrish and Schultheiss [7] for mean squared distortion when the source vector has a finite second moment and differential entropy, and its density is bounded and continuous.

3 The main result and corollaries

Let $\mathcal{A}$ denote the set of all difference distortion measures ρ which satisfy the following two conditions

(i) The equations

$$a(D)\int_{\mathcal{R}^k} e^{-s(D)\rho(x)}\,dx = 1 \tag{8}$$

$$a(D)\int_{\mathcal{R}^k} \rho(x)e^{-s(D)\rho(x)}\,dx = D. \tag{9}$$

have a unique pair of solutions $(a(D), s(D))$ for all $D > 0$. Moreover, $a(D)$ and $s(D)$ are continuous functions of D.

(ii) If Z_D is a random variable with density $g_D(x) = a(D)e^{-s(D)\rho(x)}$, then $Z_D \to 0$ as $D \to 0$ *in distribution.*

The following theorem is the main result of the paper. It says that the SLB for a difference distortion measure $\rho \in \mathcal{A}$ is asymptotically tight if there exists an appropriate auxiliary distortion measure δ in the set $\mathcal{A}$.

Theorem 1 *Suppose that X has a density, and $h(X) > -\infty$. Let $\rho \in \mathcal{A}$, and let Z_D be a random vector independent of X and having density given by* (6). *If there exists a $\delta \in \mathcal{A}$ with $0 < E\delta(X) < \infty$, such that $\lim_{D\to 0} E\delta(X + Z_D) = E\delta(X)$, then*

$$\lim_{D\to 0} (R(D) - R_L(D)) = 0, \tag{10}$$

where $R(D)$ and $R_L(D)$ are relative to ρ.

Note that the only conditions in the above theorem concerning the source distribution are $h(X) > -\infty$ and $E\delta(X) < \infty$. There are no tail or smoothness conditions on the density of X. Corollaries 1 and 2 will show how the specific choices of δ and ρ result in general classes of distortions and sources for which the SLB converges to the rate distortion function.

If we take $\delta = \rho$ in Theorem 1, then the conditions require that $\rho \in \mathcal{A}$, and $0 < \lim_{D\to 0} E\rho(X + Z_D) = E\rho(X) < \infty$. In Lemma 3 $\lim_{D\to 0} E\rho(X + Z_D) = E\rho(X)$ is shown to hold for a large class of distortions. By a similar proof it is not hard to see that $\lim_{D\to 0} E\delta(X + Z_D) = E\delta(X)$ holds for $\delta, \rho \in \mathcal{A}$, if $\delta(x) \le \rho(x)$, $\delta(x+y) \le c\delta(x) + c\delta(y)$ for some $c > 0$, and δ is bounded in a neighborhood of 0.

The proof of the theorem uses a technique apparently introduced by Binia *et al.* [4] for squared distortion (see also Barron [1]). This technique is based on a fundamental property of *informational divergence*, also called relative entropy. The informational divergence (see Kullback [11]) of two k-dimensional probability densities p and q is given as

$$D(p\|q) = \int_{\mathcal{R}^k} p(x) \log \frac{p(x)}{q(x)}\, dx. \tag{11}$$

When the random vectors X and Y have densities p and q, respectively, then $D(p\|q)$ will be denoted by $D(X\|Y)$. The definition of $D(X\|Y)$ can be extended to X and Y of arbitrary distribution, but in our case X and Y will always have densities. We will need the following lower semicontinuity property of the informational divergence.

Lemma 1 (Csiszár [6]) *Suppose we are given two sequences of random vectors X_n and Y_n, such that $X_n \to X$ and $Y_n \to Y$ in distribution as $n \to \infty$ for some random vectors X and Y. Then*

$$\liminf_{n\to\infty} D(X_n \| Y_n) \geq D(X\|Y). \tag{12}$$

Proof of Theorem 1 Since both $\rho, \delta \in \mathcal{A}$, by condition (i), for all $D > 0$ there exist random vectors Z_D and Y_D with densities

$$g_\rho(x) = a_\rho(D) e^{-s_\rho(D)\rho(x)} \quad \text{and} \quad g_\delta(x) = a_\delta(D) e^{-s_\delta(D)\delta(x)}, \tag{13}$$

respectively, satisfying

$$E\rho(Z_D) = D \quad \text{and} \quad E\delta(Y_D) = D. \tag{14}$$

Let now Z_D be independent of X. Since $E\rho(Z_D) = D$, we have from the definition of $R(D)$ that

$$\begin{aligned} R(D) &\leq I(X; X + Z_D) \\ &= h(X + Z_D) - h(X + Z_D|X) \\ &= h(X + Z_D) - h(Z_D). \end{aligned}$$

On the other hand, by (5) we have

$$R(D) \geq R_L(D) = h(X) - h(Z_D).$$

Thus we can write

$$0 \leq R(D) - R_L(D) \leq h(X + Z_D) - h(X), \tag{15}$$

and the SLB is asymptotically tight if we can prove that

$$\limsup_{D\to 0} h(X + Z_D) \leq h(X). \tag{16}$$

Now define $\Delta(D) = E\delta(X + Z_D)$, and note that $\lim_{D\to 0} \Delta(D) = E\delta(X) < \infty$ by the condition of the theorem. If we put $\Delta(0) = E\delta(X)$, then $\Delta(D)$ is continuous at 0. Next notice that the definition (11) of the informational divergence, the special form (13) of the density of Y_D, and the fact that $E\delta(Y_{\Delta(D)}) = \Delta(D)$ give

$$D(X + Z_D \| Y_{\Delta(D)}) = h(Y_{\Delta(D)}) - h(X + Z_D) \tag{17}$$

and

$$D(X \| Y_{\Delta(0)}) = h(Y_{\Delta(0)}) - h(X). \tag{18}$$

Note that since the informational divergence is nonnegative by Jensen's inequality, (18) implies $h(X) < \infty$. Using (17) and (18) we can write

$$\begin{aligned} h(X + Z_D) - h(X) &= h(X + Z_D) - h(Y_{\Delta(D)}) \\ &\quad + h(Y_{\Delta(D)}) - h(Y_{\Delta(0)}) \\ &\quad + h(Y_{\Delta(0)}) - h(X) \\ &= -D(X + Z_D \| Y_{\Delta(D)}) \\ &\quad + h(Y_{\Delta(D)}) - h(Y_{\Delta(0)}) \\ &\quad + D(X \| Y_{\Delta(0)}). \end{aligned} \tag{19}$$

First consider the term $h(Y_{\Delta(D)}) - h(Y_{\Delta(0)})$. We have that

$$h(Y_{\Delta(D)}) = \Delta(D) s_\delta(\Delta(D)) - \log a_\delta(\Delta(D)).$$

Since $\delta \in \mathcal{A}$, the parameters $s_\delta(\cdot)$ and $a_\delta(\cdot)$ are continuous functions of their arguments. Hence by the continuity of $\Delta(D)$ at 0 we have

$$\lim_{D\to 0} h(Y_{\Delta(D)}) = h(Y_{\Delta(0)}). \tag{20}$$

For the same reason, the density of $Y_{\Delta(D)}$ converges pointwise to the density of $Y_{\Delta(0)}$, which implies that $Y_{\Delta(D)} \to Y_{\Delta(0)}$ in distribution as $D \to 0$ (see Scheffe's theorem in Billingsley [3]). On the other hand, since X and Z_D are independent, and $Z_D \to 0$ in distribution, we have that $X + Z_D \to X$ in distribution as $D \to 0$. Then Lemma 1 gives

$$\limsup_{D\to 0} \Big(D(X \| Y_{\Delta(0)}) - D(X + Z_D \| Y_{\Delta(D)}) \Big) \le 0. \tag{21}$$

But this and (20) imply that by taking the limsup of the right hand side of (19) we get

$$\limsup_{D \to 0} h(X + Z_D) \le h(X).$$

In view of (15) this proves the theorem along with the statement $\lim_{D\to 0} h(X + Z_D) = h(X)$ $\square$

Let us now see an application of the above result. In vector quantization a rather popular class of difference distortion measures is the family of norm based distortions. When the distortion is given by

$$\rho(x - y) = \|x - y\|^r$$

for some norm $\|\cdot\|$ on $\mathcal{R}^k$ and some $r > 0$, the SLB can be evaluated explicitly (see Yamada et al. [17]), resulting in an expression which depends on the particular choice of $\|\cdot\|$ only through the volume V_k of the "unit sphere" $S_k = \{x : \|x\| \le 1\}$. In this case the SLB has the following form

$$R_L(D) = h(X) - \frac{k}{r} + \log\left(\frac{r}{kV_k\Gamma(k/r)}\left(\frac{k}{rD}\right)^{k/r}\right),$$

where $\Gamma(\cdot)$ is the gamma function. The next corollary states that if there exists an $\alpha > 0$ such that $E\|X\|^\alpha < \infty$, and $h(X) > -\infty$, then the SLB with respect to $\rho(x) = \|x\|^r$ is asymptotically tight for all choices of $r > 0$.

Corollary 1 *Let $\rho(x) = \|x\|^r$, where $\|\cdot\|$ is a norm and $r > 0$. Suppose that $h(X) > -\infty$, and there exists an $\alpha > 0$ such that $E\|X\|^\alpha < \infty$. Then*

$$\lim_{D\to 0} (R(D) - R_L(D)) = 0. \tag{22}$$

Proof Put $\rho(x) = \|x\|^r$ and $\delta(x) = \|x\|^\alpha$ in Theorem 1. It was shown in Yamada et al. [17] that for difference distortion measures in the form $\|x\|^p$, $p > 0$, the parameters

$$s(D) = \frac{k}{pD}$$

and

$$a(D) = \frac{p}{kV_k}\frac{s(D)^{k/p}}{\Gamma(k/p)}$$

satisfy (8) and (9) for all $D > 0$. This pair of solutions is clearly unique as the differentiation of the function $\left(\int \|x\|^p e^{-s\|x\|^p} dx\right) \left(\int e^{-s\|x\|^p} dx\right)^{-1}$ with respect to s shows. Furthermore, $s(D)$ and $a(D)$ are clearly continuous in D. Let now W have the density $a(1)e^{-s(1)\|x\|^p}$. Then $E\|W\|^p = 1$. Clearly, $W_D = D^{1/p}W$ has density $a(D)e^{-s(D)\|x\|^p}$, and $E\|W_D\|^p = D$. But $W_D \to 0$ as $D \to 0$ almost surely, and thus also in distribution, and it follows that $\|x\|^p$ satisfies (i) and (ii) for all p. In particular, we have $\rho, \delta \in \mathcal{A}$.

Since $E\delta(X) = E\|X\|^\alpha < \infty$ by our assumption, it only remains to check the condition $E\delta(X + Z_D) \to E\delta(X)$ in Theorem 1. From the argument above we have $Z_D = D^{1/r}Z_1$. Also, we can assume that $\alpha \le \min(1, r)$, since otherwise we can put $\alpha = \min(1, r)$ without violating the condition $E\|X\|^\alpha < \infty$. Since $E\|Z_1\|^r = 1$, this gives $E\|Z_D\|^\alpha < \infty$. But then the triangle inequality and the fact $\alpha \le 1$ imply

$$|E\|X + Z_D\|^\alpha - E\|X\|^\alpha| \le E\|Z_D\|^\alpha = D^{\alpha/r}E\|Z_1\|^\alpha \to 0 \quad \text{as} \quad D \to 0.$$

This proves the corollary. □

Next we deal with more general distortion measures. The class we consider is very similar to that considered by Linkov [13]. However, Theorem 1 will allow us to get rid of of Linkov's restrictive tail condition on the source density, replacing it with the condition that there exists a y^* such that $E\rho(X + y^*) < \infty$. A detailed discussion of the differences between Linkov's and our conditions is given in Section 4.

We consider the class of difference distortion measures that satisfy the following conditions.

(a) $\rho(x) = 0$ iff $x = 0$, $\lim_{\|x\|\to 0} \rho(x) = 0$, and $\lim_{\|x\|\to\infty} \rho(x) = \infty$.

(b) There exists a monotone increasing function $\Psi : [0, \infty) \to [0, \infty)$, and a norm $\|\cdot\|$ on $\mathcal{R}^k$ such that $\Psi(u) > 0$ if $u > 0$ and $\rho(x) \ge \Psi(\|x\|)$ for all $x \in \mathcal{R}^k$.

(c) $\displaystyle\int_{\mathcal{R}^k} e^{-s\rho(x)} dx < \infty$ for all $s > 0$.

(d) There exists a $c > 0$ such that $\rho(x + y) \le c\rho(x) + c\rho(y)$ for all $x, y \in \mathcal{R}^k$.

Corollary 2 *If $h(X) > -\infty$, ρ satisfies* (a)–(d)*, and there exists a $y^* \in \mathcal{R}^k$ such that $E\rho(X - y^*) < \infty$, then*

$$\lim_{D\to 0}(R(D) - R_L(D)) = 0. \tag{23}$$

The following two lemmas will show that the conditions of Theorem 1 are satisfied with the assumptions of the above corollary. The proofs are given in the Appendix.

Lemma 2 *If ρ satisfies conditions* (a)–(c)*, then for each $D > 0$ there is a unique continuous pair of solutions $(a(D), s(D))$ of the defining equations* (3) *and* (4) *of the SLB. Furthermore, if Z_D has density $g_D(x) = a(D)e^{-s(D)\rho(x)}$, then $Z_D \to 0$ as $D \to 0$ in distribution.*

Lemma 3 *Suppose that conditions* (a)–(d) *hold and $E\rho(X) < \infty$, and let Z_D be independent of X. Then*

$$\lim_{D\to 0} E\rho(X + Z_D) = E\rho(X). \tag{24}$$

Proof of Corollary 2 If we put $\delta = \rho$ in Theorem 1, we get $\rho, \delta \in \mathcal{A}$ by Lemma 2. But then by Lemma 3 we have $\lim_{D\to 0} E\delta(X + Z_D) = E\delta(X)$, thus the conditions of Theorem 1 are satisfied. □

4 Discussion: stationary sources

Clearly, Corollaries 1 and 2 are not the only possible applications of Theorem 1. In general, in order to establish the asymptotic tightness of the SLB, first we have to check whether $\rho \in \mathcal{A}$. Then an auxiliary $\delta \in \mathcal{A}$ has to be found for which $E\delta(X + Z_D) \to E\delta(X)$.

Let us now compare Corollary 2 with Linkov's result, the only one available for distortion measures other than the squared distortion. The differences can be summarized as follows

Conditions on the distortion measure Of the conditions of Corollary 2 Linkov required (b) with *Euclidean norm* and the condition that $\rho(x) = 0$ iff $x = 0$ from (a). Moreover he assumed that

(A) $\int \rho^2(x) e^{-s\Psi(x)}\, dx < \infty$.

(B) $\limsup_{x\to 0} \rho(x)\|x\|^{-\nu} < \infty$ for some $\nu > 0$.

Conditions (A) and (B) combined with (b) clearly imply (a). On the other hand, although they are similar, neither (A) nor (c) imply the other. This means, in effect, that while we relaxed the condition about the behavior of ρ for small x's, we imposed the global condition $\rho(x+y) \le c\rho(x) + c\rho(y)$ (Condition (d)), and we allowed arbitrary norm in (b).

Conditions on the source density Linkov assumed $h(X) > -\infty$ and the following tail condition

(C) There exists a nonnegative monotone nonincreasing function ϕ of a real variable such that for some $r > 0$ we have $f(x) \le \phi(\|x\|)$ for all $\|x\| \ge r$ and

$$\left| \int_{\|x\|\ge r} \phi(\|x\|) \log \phi(\|x\|)\, dx \right| < \infty.$$

We replaced this tail condition with the condition that $E\rho(X + y^*) < \infty$ for some y^*. This latter condition is equivalent to assuming that there exists a 1-level quantizer with finite ρ distortion. This is a usual assumption for the proof of the source coding theorem, and is assumed in virtually all works dealing with vector quantization.

One of the advantages of replacing a tail condition with a moment condition becomes apparent when *single letter distortion measures* and *stationary sources* are considered. Let $\rho_k : \mathcal{R}^k \times \mathcal{R}^k \to [0,\infty)$ be given as

$$\rho_k(x^k, y^k) = \frac{1}{k} \sum_{i=1}^{k} \rho(x_i, y_i), \tag{25}$$

where $x^k = (x_1, \ldots, x_k) \in \mathcal{R}^k$, $y^k = (y_1, \ldots, y_k) \in \mathcal{R}^k$, and $\rho : \mathcal{R} \times \mathcal{R} \to [0,\infty)$. The kth order rate distortion function of the stationary source $X_1, X_2, \ldots$ is defined as

$$R_k(D) = \inf \left\{ \frac{1}{k} I(X^k; Y^k) :\ E\rho_k(X^k, Y^k) \le D \right\},$$

where $X^k = (X_1, \ldots, X_k)$ and $Y^k = (Y_1, \ldots, Y_k)$. Note that with the notation of the previous section $X^k = X$ and $R_k(D) = (1/k)R(D)$. The rate distortion function of the stationary source is given as

$$\hat{R}(D) \stackrel{\text{def}}{=} \lim_{k\to\infty} R_k(D)$$

(see e.g. Berger [2] for the proof that the limit exists). Let us suppose that ρ is a difference distortion measure, and for $D > 0$ let $g_D(x) = a(D)e^{-s(D)\rho(x)}$ satisfying

$$\int_{\mathcal{R}} g_D(x)\, dx = 1 \quad \text{and} \qquad \int_{\mathcal{R}} \rho(x) g_D(x)\, dx = D.$$

Then taking the k-fold product $g_{k,D}(x^k) = a(D)^k e^{-ks(D)\rho(x^k)}$, we have

$$\int_{\mathcal{R}^k} g_{k,D}(x^k)\, dx^k = 1 \quad \text{and} \qquad \int_{\mathcal{R}^k} \rho_k(x^k) g_D(x^k)\, dx^k = D.$$

Thus $h(g_{k,D}) = kh(g_D)$, and the Shannon lower bound on $R_k(D)$ gives

$$\begin{aligned} R_k(D) \geq R_{k,L}(D) &\stackrel{\text{def}}{=} \frac{1}{k}\left(h(X^k) - h(g_{k,D})\right) \\ &= \frac{1}{k}h(X^k) - h(g_D), \end{aligned} \tag{26}$$

if X^k has a density, and $h(X^k) < \infty$. Let $h \stackrel{\text{def}}{=} \lim_{k\to\infty}(1/k)h(X^k)$ be the differential entropy rate of the source. Then taking the limit as $k\to\infty$ in (26) implies

$$\hat{R}(D) \geq h - h(g_D) \stackrel{\text{def}}{=} \hat{R}_L(D),$$

and $\hat{R}_L(D)$ is called the *generalized Shannon lower bound* on the rate distortion function of the source (see e.g. [2, p. 132]).

Let $Z_{1,D}, Z_{2,D}, \ldots$ be a sequence of i.i.d. random variables with common density g_D (implying that $Z_D^k = (Z_{1,D}, \ldots, Z_{k,D})$ has density $g_{k,D}$), and let $(Z_{1,D}Z_{2,D}\ldots)$ and $(X_1, X_2, \ldots)$ be independent. Then by the proof of Theorem 1 we have (see eq. (15)) that

$$0 \leq R_k(D) - R_{k,L}(D) \leq \frac{1}{k}h(X^k + Z_D^k) - \frac{1}{k}h(X^k). \tag{27}$$

Since the sequence $X_1 + Z_{1,D}, X_2 + Z_{2,D}, \ldots$ is stationary, we have $(1/n)h(X^n + Z_D^n) \geq \lim_{k\to\infty}(1/k)h(X^k + Z_D^k)$ for all n. Thus by taking the limit as $k\to\infty$ in (27), we get for all $n > 0$

$$0 \leq \hat{R}(D) - \hat{R}_L(D) \leq \lim_{k\to\infty} \frac{1}{k}h(X^k + Z_D^k) - h \leq \frac{1}{n}h(X^n + Z_D^n) - h.$$

Suppose now that for all n

$$\lim_{D \to 0} h(X^n + Z_D^n) = h(X^n). \tag{28}$$

Then

$$\limsup_{D \to 0} \left(\hat{R}(D) - \hat{R}_L(D) \right) \leq \frac{1}{n} h(X^n) - h$$

for all n, which implies

$$\lim_{D \to 0} \left(\hat{R}(D) - \hat{R}_L(D) \right) = 0,$$

i.e., the generalized SLB is asymptotically tight. But if the one dimensional difference distortion measure ρ and X_1 satisfy the conditions of Theorem 1, then clearly ρ_k and $X^k = (X_1, \ldots, X_k)$ will also satisfy these conditions, and (28) holds for all n by the proof of Theorem 1. Thus we have proved

Corollary 3 *Let $\hat{R}(D)$ be the rate distortion function of the stationary source $X_1, X_2, \ldots$, relative to the single letter difference distortion measure ρ. Suppose that the source has differential entropy rate $h > -\infty$, and ρ together with X_1 satisfy the conditions of Theorem 1. Then*

$$\lim_{D \to 0} \left(\hat{R}(D) - \hat{R}_L(D) \right) = 0,$$

where $\hat{R}_L(D) = h - h(g_D)$.

The main merit of Corollary 3 is that the tightness of the SLB can be deduced even if we have no information whatsoever about the higher dimensional distributions of $X_1, X_2, \ldots$, other than $h > -\infty$. On the other hand, Linkov's result implies the tightness of the SLB for stationary sources only when the k-dimensional densities of the $X^k = (X_1, \ldots, X_k)$ satisfy the tail condition (C) for all k large enough. But this is a strict restriction on the process statistics. In fact, it is not hard to construct a stationary Markov source for which (C) is satisfied by the one dimensional marginal, but it is violated by all the higher dimensional densities. Such an example is given in the Appendix.

Appendix

Example We will construct a stationary first-order Markov source which satisfies the conditions of Corollary 3 with the squared distortion, but violates Linkov's tail condition for each dimension $k \geq 2$. Let $B \in \mathcal{R}^2$ be defined as $B = \{(x_1, x_2) \in \mathcal{R}^2 : x_1, x_2 \geq 1, x_1 \geq x_2 - \frac{1}{2}x_2^{-4}, x_2 \geq x_1 - \frac{1}{2}x_1^{-4}\}$. Since $\int_B dx = A < \infty$, we can define the two-dimensional probability density $f(x_1, x_2)$ as

$$f(x_1, x_2) = \begin{cases} A^{-1} & \text{if } (x_1, x_2) \in B \\ 0 & \text{otherwise.} \end{cases}$$

Let $p(x_1) = \int_{\mathcal{R}} f(x_1, x_2)\, dx_2$ and define the conditional density $f(x_2|x_1) = f(x_1, x_2)/p(x_1)$. Then the symmetry of B implies that the first-order Markov source $X_1, X_2, \ldots$ defined by the marginal $p(x_1)$ and the conditional density $f(x_2|x_1)$ is *stationary.* Since $E|X_1|^2 < \infty$ and $h = h(X_2|X_1) = -\int f(x_1, x_2) \log f(x_2|x_1)\, dx_1 dx_2$ is finite, it follows by Corollary 3 that the generalized SLB is tight for this source. On the other hand, for any $x \geq 1$ $k \geq 2$, the joint density $f_k(x_1, \ldots, x_k) = p(x_1) \prod_{i=2}^{k} f(x_i|x_{i-1})$ is greater than A^{-1} in a small enough neighborhood of $x_1^k = (x, \ldots, x)$. It follows that if ϕ is to satisfy $\phi(\|x_1^k\|) \geq f(x_1^k)$ if $\|x_1^k\| \geq r$ for some $r > 0$, then $\phi(x) \geq A^{-1}$ for $x \geq r$. But then Linkov's tail condition is clearly violated since the integral in (C) is infinite.

Proof of Lemma 2 The same statements were proved by Linkov with similar, but not identical conditions given in Section 4. For this reason, the proof is detailed only in steps that are different in Linkov's development.

For $s > 0$ define the function $l(s)$ by

$$l(s) = \int_{\mathcal{R}^k} \rho(x)\gamma(s)e^{-s\rho(x)}\, dx,$$

where $\gamma(s) = \left(\int e^{-s\rho(x)}\, dx\right)^{-1}$. Then $0 < \gamma(s) < \infty$ for all $s > 0$ by (b) and (c), and it is easy to see that (a) and (c) imply $l(s) < \infty$, $s > 0$. The uniqueness and continuity of the solutions $a(D)$ and $s(D)$ (when a solution exists) follows from the fact that $l(s)$ and $\gamma(s)$ are differentiable in s and $\gamma(s) > 0$, $l'(s) < 0$ for all $s > 0$. The existence of these derivatives is shown by using conditions (b) and (c), and the dominated convergence theorem. To see that the solution exists for all $D > 0$, we have to show that

$$\lim_{s \to 0} l(s) = \infty, \tag{29}$$

and

$$\lim_{s \to \infty} l(s) = 0. \tag{30}$$

By Fatou's lemma we have

$$\liminf_{s \to 0} \int e^{-s\rho(x)}\, dx \geq \int \lim_{s \to 0} e^{-s\rho(x)}\, dx = \infty,$$

thus

$$\lim_{s \to 0} \gamma(s) = 0 \tag{31}$$

follows for all x. Since $\lim_{\|x\| \to \infty} \rho(x) = \infty$ by (a), we have that for any $B > 0$ there exists an $r > 0$ such that $\rho(x) \geq B$ if $\|x\| \geq r$. But then

$$\int_{\|x\| \geq r} \rho(x)\gamma(s)e^{-s\rho(x)}\, dx \geq B \int_{\|x\| \geq r} \gamma(s)e^{-s\rho(x)}\, dx \to B \quad \text{as} \quad s \to 0,$$

since $\int_{\|x\| < r} \gamma(s)e^{-s\rho(x)}\, dx = 0$ as $s \to 0$ by (31). Since $B > 0$ can be arbitrarily large, (29) is proved.

To prove (30) notice that $\lim_{s \to \infty} \gamma(s)e^{-s\epsilon} = 0$ for all $\epsilon > 0$, since otherwise $\lim_{\|x\| \to 0} \rho(x) = 0$ would imply $\limsup_{s \to \infty} \int \gamma(s)e^{-s\rho(x)}\, dx = \infty$, a contradiction. But this gives

$$\lim_{s \to \infty} \gamma(s)e^{-s\rho(x)} = 0 \tag{32}$$

for all $x \neq 0$. Thus by $\lim_{\|x\| \to 0} \rho(x) = 0$, we have for $r > 0$ small enough that

$$\lim_{s \to \infty} \int_{\|x\| < r} \rho(x)\gamma(s)e^{-s\rho(x)} = 0. \tag{33}$$

Letting $\rho_r = \inf_{\|x\| \geq r} \rho(x)$ (note that $\rho_r > 0$ by (b)), we have from (32) that for a given $s_0 > 0$ and all $s > s_0$ large enough

$$\gamma(s)e^{-s\rho_r} \leq \gamma(s_0)e^{-s_0\rho_r}.$$

This implies

$$\rho(x)\gamma(s)e^{-s\rho(x)} \leq \rho(x)\gamma(s_0)e^{-s_0\rho(x)} \tag{34}$$

for all x with $\|x\| \geq r$ and s large enough. In view of (32) and (34), the dominated convergence theorem gives

$$\lim_{s\to\infty} \int_{\|x\|\geq r} \rho(x)\gamma(s)e^{-s\rho(x)} = 0,$$

which, combined with (33) proves (30).

The fact that $Z_D \to 0$ in distribution as $D \to 0$ follows from the fact that$E\rho(Z_D) = D$ (see Theorem 4.3.4 in [2]), and is shown by the following simple argument. For all $r > 0$ we have

$$\begin{aligned}
\int_{\|x\|\geq r} g_D(x)\,dx &\leq \int_{\|x\|\geq r} \frac{\rho(x)}{\Psi(x)} g_D(x)\,dx \\
&\leq \frac{1}{\Psi(r)} \int_{\|x\|\geq r} \rho(x) g_D(x)\,dx \\
&\leq \frac{D}{\Psi(r)} \to 0 \quad \text{as} \quad D \to 0.
\end{aligned}$$

□

Proof of Lemma 3 Let q_D be the density of $X + Z_D$. Then $q_D = f * g_D$, where $*$ denotes convolution. We have to prove that

$$\lim_{D\to 0} \int_{\mathcal{R}^k} \rho(x) q_D(x)\,dx = \int_{\mathcal{R}^k} \rho(x) f(x)\,dx.$$

Conditions (a) and (d) imply that $\rho(x)$ is bounded on compact subsets of $\mathcal{R}^k$. In particular, we have $\rho^{(r)} = \sup_{\|x\|\leq r} \rho(x) < \infty$ for all $r > 0$. It is well known (see e.g. [16, Theorem 9.6]) that if for all $r > 0$

$$\lim_{D\to 0} \int_{\|x\|>r} g_D(x)\,dx = 0,$$

then $q_D = f * g_D \to f$ in L_1 norm. Hence,

$$\lim_{D\to 0} \int_{\|x\|\leq r} \rho^{(r)} q_D(x)\,dx = \int_{\|x\|\leq r} \rho^{(r)} f(x)\,dx,$$

and the generalized dominated convergence theorem (see [14]) implies

$$\lim_{D\to 0} \int_{\|x\|\leq r} \rho(x) q_D(x)\,dx = \int_{\|x\|\leq r} \rho(x) f(x)\,dx. \tag{35}$$

It is clear that the lemma will follow from (35) if we can prove that for any given $\epsilon > 0$, there exist $r_0 > 0$ and $D_0 > 0$ such that

$$\int_{\|x\|>r} \rho(x) q_D(x)\, dx < \epsilon \tag{36}$$

for all $r > r_0$ and $D < D_0$. To this end, we can write

$$\begin{aligned}\int_{\|x\|>r} \rho(x) q_D(x)\, dx &= \int_{\mathcal{R}^k} \left(\int_{\|x\|>r} \rho(x) f(x-y)\, dx \right) g_D(y)\, dy \\ &= \int_{\mathcal{R}^k} \Phi(y,r) g_D(y)\, dy, \end{aligned} \tag{37}$$

where

$$\Phi(y,r) = \int_{\|u+y\|>r} \rho(u+y) f(u)\, du. \tag{38}$$

When $\|y\| \le r/2$, the sphere $\{u : \|u\| \le r/2\}$ is contained in the sphere $\{u : \|u+y\| \le r\}$, and we have by condition (d) that

$$\begin{aligned}\Phi(y,r) &\le \int_{\|u\|>r/2} \rho(u+y) f(u)\, du \\ &\le \int_{\|u\|>r/2} c\rho(u) f(u)\, du + c\rho(y). \end{aligned} \tag{39}$$

By $E\rho(X) < \infty$ the first term in (39) is less than $\epsilon/2$ if r is large enough. Thus

$$\int_{\|y\|\le r/2} \Phi(y,r) g_D(y)\, dy \le \frac{\epsilon}{2} + cD. \tag{40}$$

When $\|y\| > r/2$, we have

$$\Phi(y,r) \le \int_{\mathcal{R}^k} \rho(u+y) f(u)\, du \le cE\rho(X) + c\rho(y),$$

and we get

$$\int_{\|y\|>r/2} \Phi(y,r) g_D(y)\, dy \le cE\rho(X) \int_{\|y\|>r/2} g_D(y)\, dy + cD. \tag{41}$$

Since $\lim_{D\to 0} \int_{\|y\|>r/2} g_D(y)\, dy = 0$ for all $r > 0$ by the proof of Lemma 2, both terms on the right hand side of (41) tend to zero as $D \to 0$. Thus (37), (40), and (41) imply (36) and the lemma is proved. □

References

[1] A. R. Barron. Entropy and the central limit theorem. *Annals of Probability*, 14:336–342, 1986.

[2] T. Berger. *Rate Distortion Theory.* Prentice–Hall, Englewood Cliffs, New Jersey, 1971.

[3] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

[4] J. Binia, M. Zakai, and J. Ziv. On the ϵ-entropy and the rate-distortion function of certain non-Gaussian processes. *IEEE Trans. Inform. Theory*, IT-20:514–524, July 1974.

[5] J. A. Bucklew and G. L. Wise. Multidimensional asymptotic quantization theory with rth power distortion measures. *IEEE Trans. Inform. Theory*, IT-28:239–247, Mar. 1982.

[6] I. Csiszár. On an extremum problem of information theory. *Studia Scientiarum Mathematicarum Hungarica*, pages 57–70, 1974.

[7] A. M. Gerrish and P.M Schultheiss. Information rates of non-Gaussian processes. *IEEE Trans. Inform. Theory*, IT-10:265–271, Oct. 1964.

[8] A. Gersho. Asymptotically optimal block quantization. *IEEE Trans. Inform. Theory*, IT-25:373–380, July 1979.

[9] H. Gish and J. N. Pierce. Asymptotically efficient quantizing. *IEEE Trans. Inform. Theory*, IT-14:676–683, Sept. 1968.

[10] R. M. Gray. *Source Coding Theory.* Kluwer, Boston, 1990.

[11] S. Kullback. *Information Theory and Statistics.* Wiley, New York, 1959.

[12] T. Linder, , and K. Zeger. Asymptotic entropy constrained performance of tessellating and universal randomized lattice quantization. to appear in *IEEE Trans. Inform. Theory*, 1994.

[13] Yu. N. Linkov. Evaluation of epsilon entropy of random variables for small epsilon. *Problems of Information Transmission*, 1:12–18, 1965. Translated from Problemy Peredachi Informatsii, Vol. 1, 18-28.

[14] H. Royden. *Real Analysis.* Macmillan, New York, 1968.

[15] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record, Part 4*, pages 142–163, 1959.

[16] R. L. Wheeden and A. Zygmund. *Measure and integral.* Marcel Dekker, New York, 1977.

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

[18] R. Zamir and M. Feder. On universal quantization by randomized uniform/lattice quantizers. *IEEE Trans. Inform. Theory*, IT-38:428–436, Mar. 1992.