A converse for lossy source coding in the finite blocklength regime

A Converse for Lossy Source Coding in the Finite Blocklength Regime

Lars Palzer and Roy Timo
Technische Universität München
{lars.palzer, roy.timo}@tum.de

***Abstract*—We present a converse bound for lossy source coding in the finite blocklength regime. The bound is based on d-tilted information, and it combines ideas from two different converse techniques by Kostina and Verdú. When particularised to the binary and Gaussian memoryless sources, the new bound gives slightly tighter results in certain blocklength regimes.**

I. Introduction and Main Results

Kostina and Verdú recently presented two general converse bounds in [1] for the problem of lossy source coding at finite blocklengths. These bounds were respectively based on d-*tilted information* and *binary hypothesis testing* arguments. Matsuta and Uyematsu [2] presented a converse at ISIT'15 that is tighter than Kostina and Verdú's meta-converse, but, unfortunately, this bound is not (yet) numerically computable.

The main purpose of this paper is to report a new converse bound for general sources, and to particularise the bound to the *binary memoryless source* (BMS) with Hamming distortions and the *Gaussian memoryless source* (GMS) with squared-error distortions. The new bound is simply stated and (we believe) quite intuitive. Its proof combines ideas from Kostina and Verdú's d-tilted and meta-converse bounds, and it gives slightly better numerical results for the BMS and GMS at certain blocklengths.

A. Problem Statement & Basic Functions

Our presentation will follow the one-shot paradigm in [1]: We first consider an abstract rate-distortion (RD) problem that consists of compressing and reconstructing a single random variable. We then specialise this one-shot problem setup to the block encoding and decoding of memoryless sources.

Let X be the output of a general source with distribution p_X on an alphabet $\mathcal{X}$. A (possibly stochastic) encoder

$$f : \mathcal{X} \to \mathcal{M} := \{1, 2, \ldots, M\}$$

maps the source output X to an index $T := f(X)$ from which a (possibly stochastic) decoder

$$g : \mathcal{M} \to \hat{\mathcal{X}}$$

outputs $\hat{X} := g(T)$ as its estimate of X. Let

$$\mathsf{d} : \mathcal{X} \times \hat{\mathcal{X}} \to [0, \infty)$$

denote the distortion function.

Definition. An (M, d, ε)-code consists of an encoder f and decoder g, as described above, with $\mathbb{P}\big[\,\mathsf{d}(X, \hat{X}) > d\big] \leq \varepsilon$.

In this paper, the main problem is to find lower bounds on the smallest M for which there exists an (M, d, ε)-code.

The abstract problem formulation above can be specialised to block encoding/decoding of memoryless sources as follows.

Definition. An (n, M, d, ε) code for a memoryless source with distribution $p_{\mathbb{X}} = p_X^n := p_X \times \ldots \times p_X$ putting out strings $\mathbb{X}$ of length n from $\mathcal{X}^n = \mathcal{X} \times \ldots \times \mathcal{X}$ and reconstruction alphabet $\hat{\mathcal{X}}^n$ consists of an encoder $f : \mathcal{X}^n \to \mathcal{M}$ and a decoder $g : \mathcal{M} \to \hat{\mathcal{X}}^n$ satisfying $\mathbb{P}\big[\,\mathsf{d}(\mathbb{X}, \hat{\mathbb{X}}) > d\big] \leq \varepsilon$.

Let

$$R(d) := \inf_{p_{\hat{X}|X} :\ \mathsf{E}[\,\mathsf{d}(X,\hat{X})] \leq d} \mathsf{E}\big[\imath_{X;\hat{X}}(X; \hat{X})\big], \qquad (1)$$

denote the usual *RD function*, where

$$\imath_{X;\hat{X}}(x; \hat{x}) := \log \frac{\mathsf{d}p_{\hat{X}|X=x}}{\mathsf{d}p_{\hat{X}}}(\hat{x}),$$

is the *information density* of $p_{X\hat{X}} = p_{\hat{X}|X} p_X$. As in [1], we make the following two basic assumptions:

A1. The distortion constraint d satisfies $R(d) < \infty$.
A2. The infimum in (1) is achieved by a unique[1] $p^*_{\hat{X}|X}$.

Let $p^*_{\hat{X}}$ denote the $\hat{X}$-marginal on $\hat{\mathcal{X}}$ induced by $p^*_{\hat{X}|X}$ and p_X, and define $\lambda := -R'(d)$ to be the negative slope of the RD function at distortion d. Let

$$\jmath_X(x, d) := \log \frac{1}{\mathsf{E}_{p^*_{\hat{X}}}[\exp(\lambda(d - \mathsf{d}(x, \hat{X}))]},$$

where the expectation is taken with respect to $p^*_{\hat{X}}$. The function $\jmath_X(x, d)$ is called d-*tilted information*, and, intuitively, it corresponds to the number of bits required to represent a particular source realisation x to within distortion d. For example, one can show [3], [4] that $R(d) = \mathsf{E}[\jmath_X(X, d)]$. We now summarise the main results of the paper. These results are proved in Sections II, III and IV.

B. General Sources

We start with a converse for general sources.

Theorem 1. *Any (M, d, ε) code must satisfy*

$$M \geq \sup_{\beta \in \mathbb{R}} \left(\frac{\mathbb{P}\big[\jmath_X(X, d) \geq \beta\big] - \varepsilon}{\sup_{\hat{x} \in \hat{\mathcal{X}}} \mathbb{P}\big[\jmath_X(X, d) \geq \beta,\ \mathsf{d}(X, \hat{x}) \leq d\big]} \right). \qquad (2)$$

[1] We make this assumption for clarity of presentation. As mentioned in [1, Remark 9], it can be relaxed.

C. Binary Memoryless Sources (BMS)

The next corollary specialises Theorem 1 to the special case of a BMS with Hamming distortions. Let $\mathbb{X} = (X_1, X_2, \ldots, X_n)$ be a string of n iid instances of $X \sim \text{Bernoulli}(p)$, and choose the distortion function to be

$$\mathsf{d}(\mathbb{X}, \hat{\mathbb{X}}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{X_i \neq \hat{X}_i\}.$$

Corollary 2 (BMS)**.** *Fix $p \in (0, 1/2)$ and $d \in [0, p)$. Any (n, M, d, ε) code must satisfy*

$$M \geq \max_{0 \leq b \leq n} \left(\frac{\sum_{k=b}^{n} \binom{n}{k} p^k (1-p)^{n-k} - \varepsilon}{\alpha_{n,d,p}(b)} \right), \tag{3}$$

where

$$\alpha_{n,d,p}(b) = \max_{\hat{n}_1} \sum_{k=0}^{\lfloor nd \rfloor} \sum_{l=0}^{k} \binom{\hat{n}_1}{l} \binom{n - \hat{n}_1}{k - l} \cdot p^{\hat{n}_1 + k - 2l} (1-p)^{n - \hat{n}_1 - k + 2l} \mathbb{1}\{\hat{n}_1 + k - 2l \geq b\}$$

and the maximisation is taken over all $\hat{n}_1 \in \mathbb{N}$ satisfying

$$\max\{0, b - \lfloor nd \rfloor\} \leq \hat{n}_1 \leq \min\{n, b + \lfloor nd \rfloor\}.$$

It is worth noting that Corollary 2 does not weaken Theorem 1; that is, the right hand sides of (2) and (3) are equal for the BMS with Hamming distortions.

*R*emark: For $p = 1/2$, $\jmath_{\mathbb{X}}(\mathbb{X}, d)$ does not depend on $\mathbb{X}$ [1, Example 1]. In this case, Theorem 1 coincides with [1, Thm. 20] which is derived from the meta-converse bound.

D. Gaussian Memoryless Sources (GMS)

Now let $\mathbb{X} = (X_1, X_2, \ldots, X_n)$ be a string of n iid instances of $X \sim \mathcal{N}(0, 1)$, and consider the squared-error distortions

$$\mathsf{d}(\mathbb{X}, \hat{\mathbb{X}}) = \frac{1}{n} \sum_{i=1}^{n} |X_i - \hat{X}_i|^2.$$

A slight weakening of Theorem 1 yields the next corollary. Here $f_{\chi_n^2}(\cdot)$ denotes the χ_n^2 probability density function.

Corollary 3 (GMS)**.** *Fix $d \in (0, 1)$. Any (n, M, d, ε) code must satisfy*

$$M \geq \sup_{\gamma \geq nd} \left(\frac{\int_{\gamma}^{\infty} f_{\chi_n^2}(w)\, \mathsf{d}w - \varepsilon}{\frac{1}{2} I_{nd/\gamma}\left(\frac{n-1}{2}, \frac{1}{2}\right) \int_{\gamma}^{\gamma^\star} f_{\chi_n^2}(w)\, \mathsf{d}w} \right), \tag{4}$$

where $I_{(\cdot)}(\cdot, \cdot)$ is the regularized incomplete beta function and

$$\gamma^\star := \left[\frac{2(nd)^{n/2}}{I_{nd/\gamma}\left(\frac{n-1}{2}, \frac{1}{2}\right)} + \gamma^{n/2} \right]^{2/n}. \tag{5}$$

E. Comparions to Existing Bounds

We now compare Theorem 1 and Corollaries 2 and 3 to Kostina and Verdú's d-*tilted information* and *meta-converse* bounds in [1]. Let us first recall the d-tilted information bound.

Theorem KV-1. Any (M, d, ε) code must satisfy [1, Thm. 7]

$$\varepsilon \geq \sup_{\gamma \geq 0} \Big(\mathbb{P}[\jmath_X(X, d) \geq \log M + \gamma] - e^{-\gamma} \Big). \tag{6}$$

To compare Theorem 1 with Theorem KV-1, it is helpful to first rewrite (2) as a lower bound on ε:

$$\varepsilon \geq \sup_{\beta \in \mathbb{R}} \Big(\mathbb{P}\big[\jmath_X(X, d) \geq \beta\big] - M \sup_{\hat{x} \in \hat{\mathcal{X}}} \mathbb{P}\big[\jmath_X(X, d) \geq \beta,\, \mathsf{d}(X, \hat{x}) \leq d\big] \Big). \tag{7}$$

Given the similarities between (6) and (7), one might guess that Theorem KV-1 can be recovered as a special case of Theorem 1 by choosing β appropriately in (7). We now show that this is indeed the case, and, therefore, Theorem KV-1 cannot be stronger than Theorem 1.

Choose $\beta = \log M + \gamma$ and consider the rightmost term in (7). We have[2]

$$\begin{aligned} &M \sup_{\hat{x} \in \hat{\mathcal{X}}} \mathbb{P}\big[\jmath_X(X, d) \geq \log M + \gamma,\, \mathsf{d}(X, \hat{x}) \leq d\big] \\ &\overset{\text{a}}{=} M \sup_{\hat{x} \in \hat{\mathcal{X}}} \mathsf{E}\left[\mathbb{1}\left\{ \frac{1}{M}\, e^{\jmath_X(X,d) - \gamma} \geq 1,\ e^{\lambda(d - \mathsf{d}(X, \hat{x}))} \geq 1 \right\} \right] \\ &\overset{\text{b}}{\leq} M \sup_{\hat{x} \in \hat{\mathcal{X}}} \mathsf{E}\left[\frac{1}{M}\, e^{\jmath_X(X,d) - \gamma}\ \mathbb{1}\big\{ e^{\lambda(d - \mathsf{d}(X, \hat{x}))} \geq 1 \big\} \right] \\ &\overset{\text{c}}{\leq} e^{-\gamma} \sup_{\hat{x} \in \hat{\mathcal{X}}} \mathsf{E}\Big[e^{\jmath_X(X,d) + \lambda(d - \mathsf{d}(X, \hat{x}))} \Big] \\ &\overset{\text{d}}{\leq} e^{-\gamma}, \end{aligned} \tag{8}$$

where (a) follows because the RD function $R(d)$ is non-increasing in d and therefore $\lambda \geq 0$; (b) and (c) follow from Markov's inequality; and (d) applies the next lemma[3].

Lemma 4 (Csiszár)**.** *For all $\hat{x} \in \hat{\mathcal{X}}$ [3, Eq. (1.22)],*

$$\mathsf{E}\Big[e^{\jmath_X(X,d) + \lambda(d - \mathsf{d}(X, \hat{x}))} \Big] \leq 1,$$

with equality for $p_{\hat{X}^\star}$-almost all $\hat{x}$.

The second converse result from Kostina and Verdú that we will consider is based on binary hypothesis testing. Let

$$\beta_\alpha(p, q) = \min_{\substack{p_{W|X} \\ \mathbb{P}[W=1] \geq \alpha}} \mathbb{Q}\big[W = 1\big], \tag{9}$$

denote the optimal performance achievable among all randomised tests $p_{W|X} : \mathcal{X} \to \{0, 1\}$ between probability distributions p and q on $\mathcal{X}$ where 1 indicates that the test chooses p and $\mathbb{Q}[\cdot]$ is the probability of an event if X has distribution q.

Theorem KV-2. Any (M, d, ε) code must satisfy [1, Thm. 8]

$$M \geq \sup_{q_X} \inf_{\hat{x} \in \hat{\mathcal{X}}} \frac{\beta_{1-\varepsilon}(p_X, q_X)}{\mathbb{Q}\big[\mathsf{d}(X, \hat{x}) \leq d\big]}, \tag{10}$$

where the supremum is taken over all distributions on $\mathcal{X}$.

Remark: After submitting this paper, we found that one can derive Theorem 1 by making a (suboptimal) choice for q_X in Theorem KV-2; see the Appendix. This also shows that Theorem KV-2 is never weaker than Theorem KV-1.

[2]The following arguments are based on the proof of Theorem KV-1 in [1].
[3]See also Property 2 in [1, p. 3311].

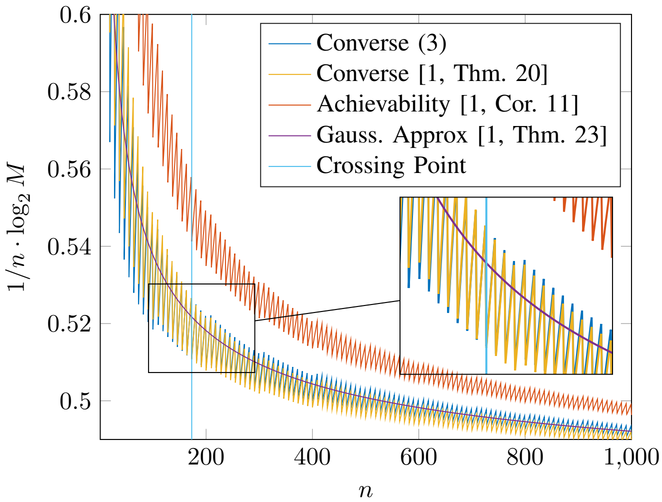

Fig. 1. BMS, $d = 0.11$, $p = 2/5$, $\varepsilon = 10^{-2}$.

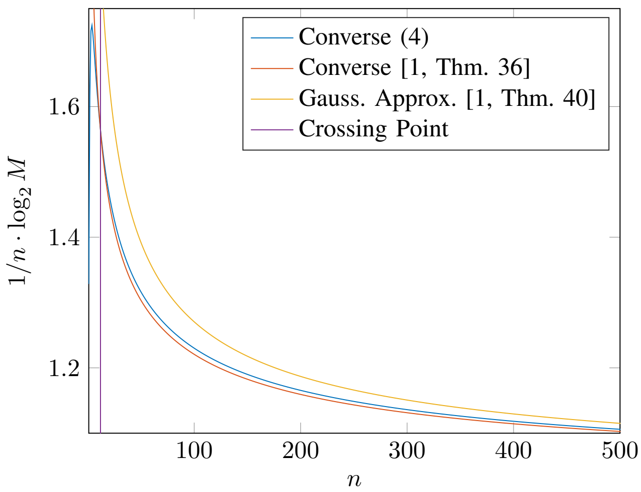

Fig. 2. GMS, $d = 0.25$, $\sigma^2 = 1$, $\varepsilon = 10^{-2}$.

F. Numerical Results

Consider the BMS under Hamming distortions with the following parameters: $p = 2/5$, $d = 0.11$ and $\varepsilon = 10^{-2}$. Figure 1 plots the lower bound (3) from Corollary 2. For comparison, the best converse bound from in [1, Thm. 20] is also plotted. For this particular setup, [1, Thm. 20] is tighter for $n < 173$, but weaker for $n \geq 173$. For completeness, we have also plotted the Gaussian approximation [1, Thm. 23] and an achievability result based on random coding [1, Theorem 10]. There, we chose $p_{\hat{\mathbb{X}}} = p_{\hat{X}}^n$ and set $p_{\hat{X}}(1) = \frac{p-d'}{1-2d'}$ with $d' := \lfloor nd \rfloor / n$, which is slightly better than choosing $p_{\hat{X}}(1) = \frac{p-d}{1-2d}$. Computations with other parameters indicate that the crossing point moves to smaller n when increasing d or ε and to larger n otherwise.

Now consider the GMS under squared error distortions with the following parameters: $d = 0.25$, $\sigma^2 = 1$ and $\varepsilon = 10^{-2}$. Figure 2 plots the bound in (4) and, for comparison, the converse bound [1, Theorem 36], which can be derived from (10). Our result is tighter for $n \geq 12$. We also included the Gaussian approximation [1, Theorem 40]. Here, choosing small values for d shifts the crossing point to larger n whereas varying ε does not seem to have a significant influence.

II. General Sources: Proof of Theorem 1

For ease of notation, we assume that $\mathcal{X}$ and $\hat{\mathcal{X}}$ are finite sets but note that the result applies to general abstract sources.

Let $\beta \in \mathbb{R}$ be arbitrary. In the same manner as the proof of Theorem KV-1 in [1], we start bounding

$$\begin{aligned}\mathbb{P}\big[\jmath_X(X,d) \geq \beta\big] &= \mathbb{P}\big[\jmath_X(X,d) \geq \beta,\ \mathsf{d}(X,\hat{X}) > d\big] \\ &\quad + \mathbb{P}\big[\jmath_X(X,d) \geq \beta,\ \mathsf{d}(X,\hat{X}) \leq d\big] \\ &\leq \varepsilon + \mathbb{P}\big[\jmath_X(X,d) \geq \beta,\ \mathsf{d}(X,\hat{X}) \leq d\big]. \quad (11)\end{aligned}$$

Now consider the second probability of the RHS. Using similar arguments as the proof of Theorem KV-2 in [1],

$$\begin{aligned}&\mathbb{P}\big[\jmath_X(X,d) \geq \beta,\ \mathsf{d}(X,\hat{X}) \leq d\big] \\ &\quad = \sum_{x \in \mathcal{X}} p_X(x) \sum_{t \in \mathcal{M}} \underbrace{p_{T|X}(t|x)}_{\leq 1} \\ &\qquad\qquad \sum_{\hat{x} \in \hat{\mathcal{X}}} p_{\hat{X}|T}(\hat{x}|t)\ \mathbb{1}\{\jmath_X(x,d) \geq \beta,\ \mathsf{d}(x,\hat{x}) \leq d\} \\ &\quad \leq \sum_{t \in \mathcal{M}} \sum_{\hat{x} \in \hat{\mathcal{X}}} p_{\hat{X}|T}(\hat{x}|t) \sum_{x \in \mathcal{X}} p_X(x) \\ &\qquad\qquad\qquad\qquad \mathbb{1}\{\jmath_X(x,d) \geq \beta,\ \mathsf{d}(x,\hat{x}) \leq d\} \\ &\quad = \sum_{t \in \mathcal{M}} \sum_{\hat{x} \in \hat{\mathcal{X}}} p_{\hat{X}|T}(\hat{x}|t) \mathbb{P}\big[\jmath_X(X,d) \geq \beta,\ \mathsf{d}(X,\hat{x}) \leq d\big] \\ &\quad \leq \sum_{t \in \mathcal{M}} \sup_{\hat{x} \in \hat{\mathcal{X}}} \mathbb{P}\big[\jmath_X(X,d) \geq \beta,\ \mathsf{d}(X,\hat{x}) \leq d\big] \\ &\quad = M \sup_{\hat{x} \in \hat{\mathcal{X}}} \mathbb{P}\big[\jmath_X(X,d) \geq \beta,\ \mathsf{d}(X,\hat{x}) \leq d\big]. \qquad (12)\end{aligned}$$

To complete the proof, combine (11) and (12) and take the supremum over β to get (2) or (7). $\square$

III. BMS: Proof of Corollary 2

Fix $p \in (0, 1/2)$, $d \in [0, p)$ and $\beta \in \mathbb{R}$. Let $h_2(\cdot)$ denote the binary entropy function. We have [1, Eqn. (21)]

$$\jmath_{\mathbb{X}}(\mathbb{x}, d) = \mathsf{N}(1|\mathbb{x}) \log \frac{1}{p} + (n - \mathsf{N}(1|\mathbb{x})) \log \frac{1}{1-p} - n h_2(d),$$

where

$$\mathsf{N}(1|\mathbb{x}) := \sum_{k=1}^{n} \mathbb{1}\{x_i = 1\}.$$

Since $p \in (0, 1/2)$, it follows that $p < 1 - p$ and $\jmath_{\mathbb{X}}(\mathbb{x}, d)$ grows linearly in $\mathsf{N}(1|\mathbb{x})$ for fixed n. Let

$$\begin{aligned}b := \min\Big\{ & n' \in \{0, \ldots, n\} : \\ & n' \log \frac{1}{p} + (n - n') \log \frac{1}{1-p} - n h_2(d) \geq \beta \Big\},\end{aligned}$$

and note that

$$\left\{\mathbb{x} \in \mathcal{X}^n : \jmath_{\mathbb{X}}(\mathbb{x}, d) \geq \beta\right\} = \left\{\mathbb{x} \in \mathcal{X}^n : \mathsf{N}(1|\mathbb{x}) \geq b\right\}.$$

Hence,

$$\begin{aligned}\mathbb{P}[\jmath_{\mathbb{X}}(\mathbb{X}, d) \geq \beta] &= \mathbb{P}\left[\mathsf{N}(1|\mathbb{X}) \geq b\right] \\ &= \sum_{k=b}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \qquad (13)\end{aligned}$$

Now consider the denominator of (2). Let $\hat{n}_1 := \mathsf{N}(1|\hat{\mathbb{x}})$. Using Vandermonde's identity, the number of binary sequences in a Hamming ball of size $\lfloor nd \rfloor$ centered at a sequence of Hamming weight $\hat{n}_1$ is given by

$$\sum_{k=0}^{\lfloor nd \rfloor} \binom{n}{k} = \sum_{k=0}^{\lfloor nd \rfloor} \sum_{l=0}^{k} \binom{\hat{n}_1}{l} \binom{n-\hat{n}_1}{k-l},$$

where, in each summand, $\binom{\hat{n}_1}{l}\binom{n-\hat{n}_1}{k-l}$ is the number of sequences of Hamming weight $\hat{n}_1 + k - 2l$. We can thus write

$$\begin{aligned}&\sup_{\hat{\mathbb{x}} \in \hat{\mathcal{X}}^n} \mathbb{P}\left[\jmath_{\mathbb{X}}(\mathbb{X}, d) \geq \beta,\, \mathsf{d}(\mathbb{X}, \hat{\mathbb{x}}) \leq nd\right] \\ &\stackrel{*}{=} \max_{\hat{n}_1} \mathbb{P}\left[\mathsf{N}(1|\mathbb{X}) \geq b,\, \mathsf{d}(\mathbb{X}, \hat{\mathbb{x}}) \leq nd\right] \\ &= \max_{\hat{n}_1} \sum_{\mathbb{x}} p_{\mathbb{X}}(\mathbb{x}) \mathbb{1}\left\{\mathsf{N}(1|\mathbb{x}) \geq b,\, \mathsf{d}(\mathbb{x}, \hat{\mathbb{x}}) \leq nd\right\} \\ &= \max_{\hat{n}_1} \sum_{k=0}^{\lfloor nd \rfloor} \sum_{l=0}^{k} \binom{\hat{n}_1}{l} \binom{n-\hat{n}_1}{k-l} p^{\hat{n}_1+k-2l} \\ &\qquad \cdot (1-p)^{n-\hat{n}_1-k+2l} \mathbb{1}\{\underbrace{\hat{n}_1 + k - 2l}_{=\mathsf{N}(1|\mathbb{x})} \geq b\}, \qquad (14)\end{aligned}$$

where (*) follows since, by symmetry, the probability depends on $\hat{\mathbb{x}}$ only through $\hat{n}_1$.

In fact, we only need to consider $b - \lfloor nd \rfloor \leq \hat{n}_1 \leq b + \lfloor nd \rfloor$ for the maximisation. This is because for $\hat{n}_1 < b - \lfloor nd \rfloor$, we have $\mathbb{1}\{\hat{n}_1 + k - 2l \geq b\} = 0$ for all summands and for $\hat{n}_1 > b + \lfloor nd \rfloor$, $\mathbb{1}\{\hat{n}_1 + k - 2l \geq b\} = 1$ for all summands in which case the sum is monotonically decreasing in $\hat{n}_1$ (we omit the proof of this fact). □

IV. GMS: Proof of Corollary 3

The d-tilted information for the GMS with $d < \sigma^2 = 1$ is given by [1, Example 2]

$$\jmath_{\mathbb{X}}(\mathbb{x}, d) = \frac{n}{2} \log \frac{1}{d} + \frac{\|\mathbb{x}\|^2 - n}{2} \log e,$$

which grows linearly in $\|\mathbb{x}\|^2$. Hence, we can rewrite (2) as

$$M \geq \sup_{\gamma \geq 0} \left(\frac{\mathbb{P}\left[\|\mathbb{X}\|^2 \geq \gamma\right] - \varepsilon}{\sup_{\hat{\mathbb{x}} \in \mathbb{R}^n} \mathbb{P}\left[\|\mathbb{X}\|^2 \geq \gamma,\, \mathsf{d}(\mathbb{X}, \hat{\mathbb{x}}) \leq d\right]} \right). \qquad (15)$$

We will lower bound (15) using a geometric argument for the denominator. By the circular symmetry of the GMS, we only need to consider those $\hat{\mathbb{x}} \in \mathbb{R}^n$ for the supremum that lie on an arbitrary straight line through the origin. Denote

$$\mathcal{A} := \left\{\mathbb{x} \in \mathbb{R}^n : \|\mathbb{x}\|^2 \geq \gamma\right\},$$

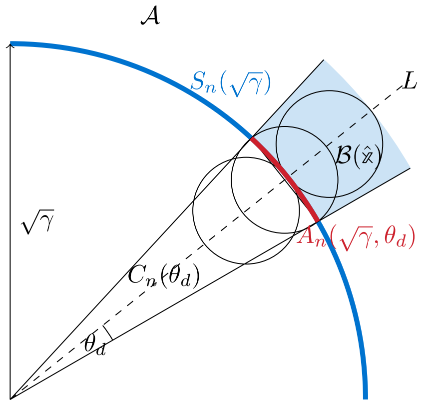

Fig. 3. Intersection of $\mathcal{A}$ with possible distortion balls.

$$\mathcal{B}(\hat{\mathbb{x}}) := \left\{\mathbb{x} \in \mathbb{R}^n : \|\mathbb{x} - \hat{\mathbb{x}}\|^2 \leq nd\right\},$$

and observe that

$$\begin{aligned}&\sup_{\hat{\mathbb{x}} \in \mathbb{R}^n} \mathbb{P}\left[\|\mathbb{X}\|^2 \geq \gamma,\, \mathsf{d}(\mathbb{X}, \hat{\mathbb{x}}) \leq d\right] \\ &\quad = \sup_{\hat{\mathbb{x}} \in \mathbb{R}^n} \mathbb{P}\left[\mathbb{X} \in \mathcal{A} \cap \mathcal{B}(\hat{\mathbb{x}})\right] \\ &\quad = \sup_{\hat{\mathbb{x}} \in L} \mathbb{P}\left[\mathbb{X} \in \mathcal{A} \cap \mathcal{B}(\hat{\mathbb{x}})\right], \qquad (16)\end{aligned}$$

where L denotes the set of points lying on a straight line through the origin, see Figure 3.

Denote the surface area of an n-dimensional sphere of radius r by $S_n(r)$ and the surface area of a n-dimensional spherical cap of radius r and half angle θ by $A_n(r, \theta)$. The following relation holds [5]:

$$A_n(r, \theta) := \frac{1}{2} S_n(r) I_{\sin^2(\theta)}\left(\frac{n-1}{2}, \frac{1}{2}\right),$$

where $I_{(\cdot)}(\cdot, \cdot)$ is the regularized incomplete beta function. Using the law of sines and taking $\gamma \geq nd$, we can determine the half angle θ_d such that $A_n(\sqrt{\gamma}, \theta_d)$ is the largest spherical cap at radius $\sqrt{\gamma}$ contained in some $\mathcal{B}(\hat{\mathbb{x}})$:

$$\theta_d = \sin^{-1} \sqrt{nd/\gamma}.$$

Let $C_n(\theta_d)$ be the n-dimensional infinite cone of half angle θ_d that passes through $A_n(\sqrt{\gamma}, \theta_d)$. Clearly, $\mathcal{A} \cap \mathcal{B}(\hat{\mathbb{x}}) \subset C_n(\theta_d)$, for any $\hat{\mathbb{x}} \in L$. This setup is visualized in Figure 3. Next, denote the volume of $\mathcal{B}(\hat{\mathbb{x}})$ for any $\hat{\mathbb{x}} \in \mathbb{R}^n$ by

$$V_n\left(\sqrt{nd}\right) := \frac{\pi^{n/2}}{\Gamma\left(\frac{n+2}{2}\right)} (nd)^{n/2},$$

with $\Gamma(\cdot)$ being the gamma function. To upper bound (16), we consider the largest probability of any set in $\mathcal{A} \cap C_n(\theta_d)$ (the shaded area in Figure 3) that has the same volume as a distortion ball. We denote this set by

$$\mathcal{K}^{\star} := \mathop{\arg\max}_{\substack{\mathcal{K} \subset \mathcal{A} \cap C_n(\theta_d): \\ \mathrm{Vol}(\mathcal{K}) = V_n(\sqrt{nd})}} \mathbb{P}\left[\mathbb{X} \in \mathcal{K}\right] \qquad (17)$$

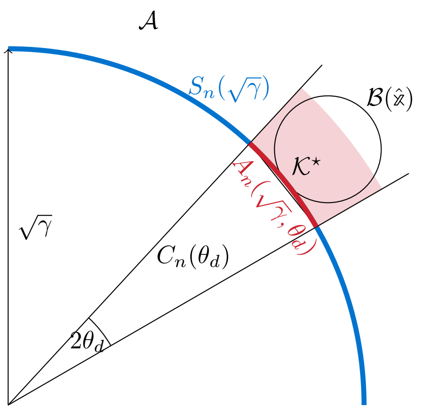

Fig. 4. Geometry of $\mathcal{K}^\star$.

The geometry of the $\arg\max$ problem is depicted in Figure 4. By the circular symmetry, $\mathcal{K}^\star$ is the slice of the cone $C_n(\theta_d)$ that lies on the surface of $S_n(\sqrt{\gamma},\theta_d)$ and has volume $V_n\big(\sqrt{nd}\big)$. More precisely, we can describe $\mathcal{K}^\star$ as the difference between spherical sectors of half angle θ_d whose volumes differ by exactly $V_n(\sqrt{nd})$, see Figure 4. The volume of a hypershperical sector of half angle θ and radius r is given by [5]

$$V_n^{\text{sec}}(r,\theta) := \frac{1}{2}V_n(r)I_{\sin^2(\theta)}\left(\frac{n-1}{2},\frac{1}{2}\right).$$

Now let $\gamma^\star$ be the solution to

$$V_n^{\text{sec}}(\sqrt{\gamma^\star},\theta_d) - V_n^{\text{sec}}(\sqrt{\gamma},\theta_d) = V_n(\sqrt{nd}), \tag{18}$$

which, using $\sin^2(\theta_d) = nd/\gamma$, can be rewritten as (5).

Now, we can use the tools developed in (16)–(18) to bound

$$\begin{aligned}
&\sup_{\hat{\mathbb{x}}\in L}\mathbb{P}\big[\mathbb{X}\in\mathcal{A}\cap\mathcal{B}(\hat{\mathbb{x}})\big] \\
&\quad\le \mathbb{P}\big[\mathbb{X}\in\mathcal{K}^\star\big] && (19)\\
&\quad= \mathbb{P}\Big[\gamma\le\|\mathbb{X}\|^2\le\gamma^\star, \mathbb{X}\in C_n(\theta_d)\Big] && (20)\\
&\quad= \mathbb{P}\big[\gamma\le\|\mathbb{X}\|^2\le\gamma^\star\big]\mathbb{P}\big[\mathbb{X}\in C_n(\theta_d)\big], && (21)\\
&\quad= \mathbb{P}\big[\gamma\le\|\mathbb{X}\|^2\le\gamma^\star\big]\frac{A_n(\sqrt{\gamma},\theta_d)}{S_n(\sqrt{\gamma})} && (22)\\
&\quad= \frac{1}{2}I_{nd/\gamma}\left(\frac{n-1}{2},\frac{1}{2}\right)\int_{\gamma}^{\gamma^\star} f_{\chi^2_n}(w)\,\mathrm{d}w && (23)
\end{aligned}$$

where (19) follows from the definition of $\mathcal{K}^\star$ (17), (20) follows from the definition of $\gamma^\star$ (18), and the geometry of $\mathcal{K}^\star$, (21)–(22) are a result of the circular symmetry of the multivariate Gaussian and $f_{\chi^2_n}(\cdot)$ is the χ^2_n probability density function. Combining (23) and (15) then yields (4).

Acknowledgements

This work was supported by the Alexander von Humboldt Foundation and the German Research Foundation. We would like to thank Rana Ali Amjad for helpful discussions and Victoria Kostina for helpful comments on different drafts.

Appendix

As noted in [1, Rem. 5], Theorem KV-2 can be strengthened by relaxing the requirement that q_X be a probability measure in (9): Instead, we may allow q_X to be any σ-finite measure to obtain the following bound.

Theorem KV-3. Any (M,d,ε)-code must satisfy

$$M \ge \sup_{q_X}\inf_{\hat{x}\in\hat{\mathcal{X}}}\frac{\beta_{1-\varepsilon}(p_X,q_X)}{\mathbb{Q}\big[\mathsf{d}(X,\hat{x})\le d\big]}, \tag{24}$$

where the supremum is taken over all σ-finite measures q_X.

We will now recover Theorem 1 from Theorem KV-3. Choose q_X such that

$$\mathbb{Q}[X\in\mathcal{A}] = \mathbb{P}[X\in\mathcal{A}, \jmath_X(X,d)\ge\beta], \quad \forall\mathcal{A}\subseteq\mathcal{X}.$$

An optimal randomised test between p_X and q_X is

$$p_{W|X}(1|x) := \begin{cases} 1, & \text{if } \jmath_X(x,d)<\beta \\ \dfrac{\mathbb{P}[\jmath_X(X,d)\ge\beta]-\varepsilon}{\mathbb{P}[\jmath_X(X,d)\ge\beta]}, & \text{if } \jmath_X(x,d)\ge\beta, \end{cases}$$

The probability that this test succeeds under p_X is

$$\begin{aligned}
\mathbb{P}[W=1] &= \mathbb{P}[\jmath_X(X,d)<\beta]\ \mathbb{P}[W=1|\jmath_X(X,d)<\beta] \\
&\quad + \mathbb{P}[\jmath_X(X,d)\ge\beta]\ \mathbb{P}[W=1|\jmath_X(X,d)\ge\beta] \\
&= 1-\varepsilon.
\end{aligned}$$

Moreover, the probability that the test fails under q_X is

$$\begin{aligned}
\mathbb{Q}[W=1] &= \mathbb{P}[\jmath_X(X,d)\ge\beta, W=1] \\
&= \mathbb{P}[\jmath_X(X,d)\ge\beta]\cdot\frac{\mathbb{P}[\jmath_X(X,d)\ge\beta]-\varepsilon}{\mathbb{P}[\jmath_X(X,d)\ge\beta]} \\
&= \mathbb{P}[\jmath_X(X,d)\ge\beta]-\varepsilon.
\end{aligned}$$

Substituting q_X into Theorem KV-3 gives

$$M \ge \inf_{\hat{x}\in\hat{\mathcal{X}}}\frac{\mathbb{P}[\jmath_X(X,d)\ge\beta]-\varepsilon}{\mathbb{P}[\jmath_X(X,d)\ge\beta, d(X,\hat{x})\le d]}.$$

Taking the supremum over β gives Theorem 1.

Remark: The above discussion together with that in Section I-E demonstrates that the d-tilted converse in Theorem KV-1 cannot be tighter than the (generalised σ-finite measure) meta converse in Theorem KV-3. To the best of our knowledge, this fact has not been observed in the literature before.

References

[1] V. Kostina, and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, Jun. 2012
[2] T. Matsuta and T. Uyematsu, "Non-asymptotic bounds for fixed-length lossy compression," *Proc IEEE Int. Symp. Inf. Theory*, Hong Kong, Jun. 2015
[3] I. Csiszár, "On an extremum problem of information theory", *Studia Sci. Math. Hungarica*, vol. 9, p. 57–71, 1974
[4] R. Gray, "Entropy and Information Theory," *Springer*, 2nd Edition, 2011
[5] S. Li, "Concise formulas for the area and volume of a hyperspherical cap", *Asian Journal of Math. and Stat.* vol. 4, no. 1, pp. 66–70, 2011