On Lossy Compression of Generalized Gaussian Sources

Alex Dytso[1], Ronit Bustin[2], H. Vincent Poor[1], and Shlomo Shamai (Shitz)[2]

Abstract—**This paper considers a problem of lossy compression of generalized Gaussian (GG) sources (i.e., sources with the probability density functions proportional to $\mathrm{e}^{-\frac{|x|^s}{2}}, s > 0$) with an $\ell_r, r > 0$, distortion measure.**

It is shown that an optimal reconstruction distribution always exists and properties of this distribution are studied. In particular, it is shown that if $s \le r - 1$ then an optimal reconstruction must have unbounded support and for $s > r$ an optimal reconstruction must have bounded support. Further, it is shown that Shannon's lower bound is achievable if and only if $r = s \in (0, 1] \cup \{2\}$, or in other words when the GG distribution is self-decomposable. Finally, conditions are shown under which an optimal reconstruction is discrete with finitely many mass points.

I. Introduction

A classical rate-distortion problem, first formulated by Shannon in [1], considers a source with the distribution P_X on $\mathcal{X}$, a reconstruction alphabet $\hat{\mathcal{X}}$, and a distortion measure $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}^+$. One of the crowning achievements of Shannon is an exact expression for the rate-distortion function given by

$$R(D) = \inf_{P_{\hat{X}|X}: \mathbb{E}\left[d(X,\hat{X})\right] \le D} I(X; \hat{X}). \quad (1)$$

For continuous sources the rate-distortion function has been found for Laplace sources with absolute error distortion $d(x, \hat{x}) = |x - \hat{x}|$ [2], exponential sources with absolute error distortion [3], and Gaussian sources with square error distortion $d(x, \hat{x}) = |x - \hat{x}|^2$ [1]. In this paper, we will enlarge this set of known cases by considering the rate-distortion problem for generalized Gaussian (GG) sources.

A. Problem Formulation

We shall refer to X_s with a GG distribution given by the probability density function (pdf)

$$f_{X_s}(x) = \frac{c_s}{\alpha} \mathrm{e}^{-\frac{|x-\mu|^s}{2\alpha^s}}, \quad (2a)$$

$$c_s = \frac{s}{2^{\frac{s+1}{s}} \Gamma\left(\frac{1}{s}\right)}, \ x \in \mathbb{R}, \quad (2b)$$

as $X_s \sim \mathcal{N}_s(\mu, \alpha^s)$. Well-known examples of this family of distributions include: the Laplace distribution for $s = 1$; the Gaussian distribution for $s = 2$; and the uniform distribution on $[-\beta, \beta]$ for $s = \infty$ and $\alpha = \lim_{s\to\infty} \left(\frac{1}{2}\right)^{\frac{1}{s}} \beta$.

[1]A. Dytso and H.V. Poor are with the Department of Electrical Engineering, Princeton University, NJ, Princeton 08544, USA (email: adytso, poor@princeton.edu).

[2]R. Bustin and S. Shamai (Shitz) are with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: bustin@technion.ac.il, sshlomo@ee.technion.ac.il).

The work of A. Dytso and H.V. Poor was supported by the National Science Foundation under Grant CNS-1702808. The work of R. Bustin and S. Shamai has been supported by the European Union's Horizon 2020 Research and Innovation Programme, grant agreement no. 694630.

We consider the rate-distortion problem with a GG source and a distortion measure that corresponds to an ℓ_r-norm

$$d(x, \hat{x}) = |x - \hat{x}|^r, r > 0. \quad (3)$$

Formally, we seek to solve

$$R_{r,s}(\alpha, D) = \inf_{P_{X_s\hat{X}} \in \mathcal{P}_{r,s}(D)} I(X_s; \hat{X}), \quad (4)$$

where $X_s \sim \mathcal{N}_s(\mu, \alpha^s)$ and

$$\mathcal{P}_{r,s}(D) = \left\{ P_{X_s\hat{X}} : \mathbb{E}[|X_s - \hat{X}|^r] \le \frac{2D^r}{r}, \right.$$
$$\left. \text{and } P_{X_s} = \mathcal{N}_s(\mu, \alpha^s) \text{ is the marginal of } P_{X_s\hat{X}} \right\}. \quad (5)$$

B. Why Generalized Gaussian Sources?

The flexible parametric form of the pdf of GG distributions allows for tails that are either heavier than Gaussian ($s < 2$) or lighter than Gaussian ($s > 2$) which makes it an excellent choice for many modeling scenarios. For example, GG distributions appear naturally in a number of quantization problems [4] and [5]. For a detailed survey of applications of GG distributions in engineering, the interested reader is referred to [6].

From an information theoretic perspective the GG distribution is interesting because it maximizes the entropy and Réyni entropy under a s-th absolute moment constraint [7], [8].

Theorem 1. *Let $X \in \mathbb{R}$ such that $\mathbb{E}[|X|^s] \le \frac{2\alpha^s}{s}$. Then,*

$$h(X) \le \frac{1}{s} \log\left(\frac{s\mathrm{e}}{2c_s^s} \cdot \mathbb{E}[|X|^s]\right) \le \frac{1}{s} \log\left(\frac{\alpha^s \mathrm{e}}{c_s^s}\right). \quad (6)$$

The inequality in (6) *is attained with equality if and only if $X \sim \mathcal{N}_s(0, \alpha^s)$.*

Proof: This result can be proved via the method outlined in [7, Chapter 12]. ■

II. On the Existence of An Optimal Reconstruction

We begin our analysis by first showing that the infimum in (4) is attainable.

Theorem 2. *The infimum in* (4) *is achievable. In other words,*

$$\inf_{P_{\hat{X}X_s} \in \mathcal{P}_{r,s}(D)} I(X; \hat{X}) = \min_{P_{\hat{X}X_s} \in \mathcal{P}_{r,s}(D)} I(X; \hat{X}). \quad (7)$$

Proof: The proof of the result is rather involved and is shown in Appendix A. ■

Note that Theorem 2 does not claim that a reconstruction distribution $P_{\hat{X}}$ is unique. However, we conjecture that this is the case and the optimal input distribution is indeed unique.

III. A Characterization of Optimal Reconstruction Distributions

In order to study the support of optimal input distributions we need the following definition.

Definition 1. *A point $x \in \mathbb{R}$ is said to be* a point of increase *of a distribution P_X, if for any open subset $\mathcal{O} \subset \mathbb{R}$ containing x, $P_X(\mathcal{O}) > 0$. We denote the set of points of increase of P_X as $\mathcal{E}(P_X) \subseteq \mathbb{R}$.*

Theorem 3. *Let $X_s \sim \mathcal{N}_s(0, \alpha^s)$ and $(r, s) \in \mathbb{R}_+^2$. For $\hat{X}$ distributed according to $P_{\hat{X}}$ define the following function:*

$$g(\hat{x}; P_{\hat{X}}) = \mathbb{E}_{X_s}\left[\frac{\mathrm{e}^{-\lambda|\hat{x}-X_s|^r}}{h_\lambda(X_s; P_{\hat{X}})}\right], \tag{8a}$$

where for $\lambda > 0$

$$h_\lambda(x; P_{\hat{X}}) = \mathbb{E}_{\hat{X}}\left[\mathrm{e}^{-\lambda|\hat{X}-x|^r}\right]. \tag{8b}$$

Then, $P_{\hat{X}}^\star$ is an optimal reconstruction in (4) if and only if there exists some $\lambda > 0$ such that

$$g(\hat{x}; P_{\hat{X}}^\star) \le 1, \forall \hat{x} \in \mathbb{R}, \tag{8c}$$

$$g(\hat{x}; P_{\hat{X}}^\star) = 1, \forall \hat{x} \in \mathcal{E}\left(P_{\hat{X}}^\star\right). \tag{8d}$$

Proof: The proof follows by setting $d(\hat{x}, x) = |x - \hat{x}|^r$ in the proof giving sufficient and necessary conditions for an arbitrary distortion, $d(\hat{x}, x)$, found in [7, Chapter 10.7] and [9]. ■

IV. When is An Optimal Reconstruction Bounded?

In this section we provide conditions under which the reconstruction distribution $P_{\hat{X}}^\star$ has bounded support.

Theorem 4. *For the optimization problem in (4) an optimal reconstruction $P_{\hat{X}}^\star$ satisfies the following properties:*

- *for $s \le r - 1$ an optimal reconstruction distribution $P_{\hat{X}}^\star$ has unbounded support; and*
- *for $s > r$ an optimal reconstruction distribution $P_{\hat{X}}^\star$ has bounded support.*

Proof: We first prove that an optimal reconstruction $\hat{X}$ must be unbounded for $s \le r - 1$. Towards a contradiction suppose that $\hat{X}$ is bounded. In other words, suppose that there exists some A such that $|\hat{X}| \le A$. Next, we lower bound $g(\hat{x}; P_{\hat{X}})$ as

$$\begin{aligned} g(\hat{x}; P_{\hat{X}}) &= \mathbb{E}_{X_s}\left[\frac{\mathrm{e}^{-\lambda|\hat{x}-X_s|^r}}{h_\lambda(X_s; P_{\hat{X}})}\right] \\ &\ge \mathbb{E}_{X_s}\left[\frac{\mathrm{e}^{-\lambda|\hat{x}-X_s|^r}}{h_\lambda(X; P_{\hat{X}})} \cdot 1_{|X_s| \ge A}\right] \\ &\ge \mathbb{E}_{X_s}\left[\frac{\mathrm{e}^{-\lambda|\hat{x}-X_s|^r}}{\mathrm{e}^{-\lambda|A-X_s|^r}} \cdot 1_{|X_s| \ge A}\right] \qquad (9) \\ &= \frac{c_s}{\alpha}\int_{\mathbb{R}\setminus[-A,A]} \mathrm{e}^{-\lambda|\hat{x}-x|^r + \lambda|A-x|^r - \frac{|x|^s}{2\alpha^s}} dx, \qquad (10) \end{aligned}$$

where the lower bound in (9) follows by using the bound $\mathrm{e}^{-\lambda|\hat{X}-x|^r} \le \mathrm{e}^{-\lambda|A-x|^r}$ for $|\hat{X}| \le A$ and $|x| \ge A$.

Clearly, there exists an $\hat{x}$ such that for $s \le r - 1$ the integral in (10) integrates to infinity. Therefore, there exists an $\hat{x}$ such that the inequality in (8c) does not hold and the reconstruction $\hat{X}$ must be unbounded.

Next we prove that an optimal reconstruction $\hat{X}$ must be bounded for $s > r$. Towards a contradiction suppose that $\hat{X}$ is unbounded and let $\hat{X}'$ and X_s' denote independent copies of $\hat{X}$ and X_s

$$\begin{aligned} & g(\hat{x}; P_{\hat{X}}) \\ &= \mathbb{E}_{X_s}\left[\frac{\mathrm{e}^{-\lambda|\hat{x}-X_s|^r}}{h_\lambda(X_s; P_{\hat{X}})}\right] \\ &\overset{a)}{\le} \mathbb{E}_{X_s}\left[\frac{\mathrm{e}^{-\lambda|\hat{x}-X_s|^r}}{\mathrm{e}^{-\lambda\mathbb{E}_{\hat{X}'}\left[|X_s-\hat{X}'|^r\right]}}\right] \\ &= \mathbb{E}_{X_s}\left[\mathrm{e}^{-\lambda|\hat{x}-X_s|^r + \lambda\mathbb{E}_{\hat{X}'}\left[|X_s-\hat{X}'|^r\right]}\right] \\ &\overset{b)}{\le} \mathbb{E}_{X_s}\left[\mathrm{e}^{-\lambda|\hat{x}-X_s|^r + \lambda\max\left(2^{r-1},1\right)\left(|X|^r + \mathbb{E}_{\hat{X}'}\left[|\hat{X}'|^r\right]\right)}\right] \\ &= \mathrm{e}^{\lambda\max\left(2^{r-1},1\right)\mathbb{E}\left[|\hat{X}'|^r\right]}\mathbb{E}_{X_s}\left[\mathrm{e}^{-\lambda|\hat{x}-X_s|^r + \lambda\max\left(2^{r-1},1\right)|X_s|^r}\right] \\ &= \mathrm{e}^{\lambda\max\left(2^{r-1},1\right)\mathbb{E}\left[|\hat{X}'|^r\right]} \\ &\cdot \int_{\mathbb{R}} \mathrm{e}^{-\lambda|\hat{x}-x|^r + \lambda\max\left(2^{r-1},1\right)|x|^r - \frac{|x|^s}{2\alpha^s}} dx, \end{aligned} \tag{11}$$

where the inequalities follow from: a) applying Jensen's inequality to $h_\lambda(x; P_{\hat{X}}) = \mathbb{E}_{\hat{X}}\left[\mathrm{e}^{-\lambda|\hat{X}-x|^r}\right] \ge \mathrm{e}^{-\lambda\mathbb{E}_{\hat{X}}\left[|\hat{X}-x|^r\right]}$; and b) using the bound $|a + b|^r \le \max(2^{r-1}, 1)(|a|^r + |b|^r)$ for any $r > 0$.

Clearly, the integral in (11) converges for all $\hat{x}$ and λ as long as $s > r$ since $x \mapsto \mathrm{e}^{-\lambda|\hat{x}-x|^r + \lambda\max\left(2^{r-1},1\right)|x|^r - \frac{|x|^s}{2\alpha^s}}$ is a bounded function for $s > r$.

Now since $\hat{X}$ is unbounded there exists a sequence $\{\hat{x}_n\}_{n=1}^\infty \subseteq \mathcal{E}(P_{\hat{X}}^\star)$ such that $\lim_{n\to\infty} \hat{x}_n = \infty$. Applying this sequence to the function $g(\hat{x}; P_{\hat{X}})$ and using the bound in (11) together with the dominated convergence theorem we have that

$$\lim_{n\to\infty} g(\hat{x}_n; P_{\hat{X}}) = 0. \tag{12}$$

Clearly, (12) contradicts the condition in (8d). This implies that for $s > r$ an optimal reconstruction must be bounded. This concludes the proof. ■

The result of Theorem 4 is depicted in Fig. 1.

Remark 1. *Note that in the regime $r - 1 < s < r$ it is not clear whether an optimal reconstruction is bounded or not. In the the next section, we will show that for the case of $r = s \in (0, 1] \cup \{2\}$ the optimal reconstruction is unbounded. Also, in the regime $s < r$ we conjecture that an optimal reconstruction is unbounded. This conjecture is supported by the fact that in the proof of Theorem 4 for the regime $s \le r - 1$ we have made the integral in (10) to be equal to infinity. However, to reach a contradiction it would have sufficed to make the integral in (10) larger than one rather than equal to infinity. Hence, we suspect that with a better bound (10) one might close the gap in the regime when $r - 1 < s < r$.*

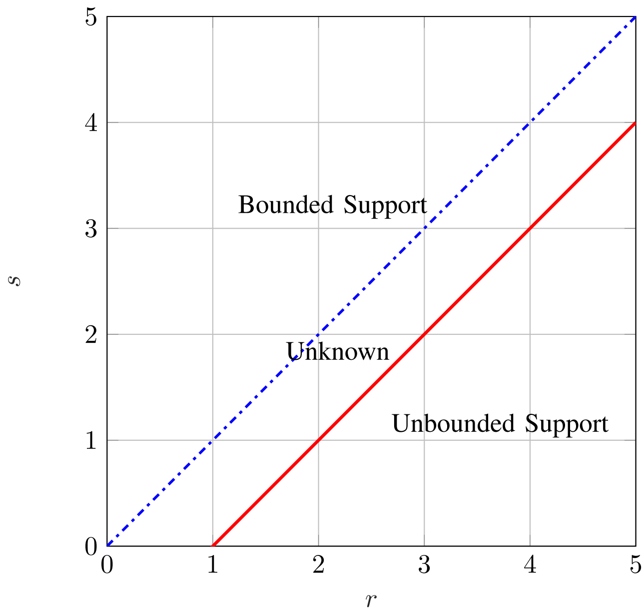

Fig. 1: Depiction of the result in Theorem 4.

V. SHANNON'S LOWER BOUND

Interestingly, in some cases we can give an exact solution for the optimization problem in (4). Specifically, this is possible when X_s has a self-decomposable distribution which is defined as follows.

Definition 2. *A random variable X is* self-decomposable *if for every $\alpha \geq 1$ there exist an independent random $\hat{X}_\alpha$ such that*

$$\alpha X \stackrel{d}{=} \hat{X}_\alpha + X', \tag{13}$$

where X' is an independent copy of X.

In [1] Shannon developed a technique for constructing a lower bound on the rate-distortion function which we now explore in our context. The following result provides a lower bound on the rate-distortion function for arbitrary values of $(r,s) \in \mathbb{R}_+^2$ and shows that this lower bound is tight in cases when the GG distribution is self-decomposable.

Theorem 5. *For $X_s \sim \mathcal{N}_s(0, \alpha^s)$ and $(r,s) \in \mathbb{R}_+^2$*

$$R_{r,s}(\alpha, D) \geq \left[\log\left(\frac{\alpha}{D}\right) + \log\left(\frac{c_r}{c_s} \mathrm{e}^{\frac{1}{s}-\frac{1}{r}}\right)\right]^+. \tag{14}$$

Moreover, the bound in (14) *is achievable if $r = s \in (0,1] \cup \{2\}$.*

Proof: The proof of the lower bounds goes as follows:

$$\begin{aligned}
&I(X_s;\hat{X}) \\
&= h(X_s) - h(X_s|\hat{X}) \\
&= h(X_s) - h(X_s - \hat{X}|\hat{X}) \\
&\stackrel{a)}{\geq} h(X_s) - h(X_s - \hat{X}) && \text{(15a)}\\
&\stackrel{b)}{\geq} \frac{1}{s}\log\left(\frac{\mathbb{E}[|X_s|^s]\cdot s\mathrm{e}}{2c_s^s}\right) - \frac{1}{r}\log\left(\frac{\mathbb{E}[|X_s-\hat{X}|^r]\cdot r\mathrm{e}}{2c_r^r}\right) && \text{(15b)}\\
&\stackrel{c)}{\geq} \log\left(\frac{\alpha}{D}\right) + \log\left(\frac{c_r}{c_s}\mathrm{e}^{\frac{1}{s}-\frac{1}{r}}\right), && \text{(15c)}
\end{aligned}$$

where the inequalities follow from: a) the fact that conditioning reduces entropy; b) the maximum entropy principle from Theorem 1; and c) the distortion constraint $\mathbb{E}\left[|X_s - \hat{X}|^r\right] \leq \frac{2D^r}{r}$.

The inequalities in (15) are tight if there exists a *backward test channel* such that for some random variable $\hat{X}$ we have that

$$X_s = \hat{X} + Z_r, \tag{16}$$

and where $Z_r \sim \mathcal{N}_r(0, D^r)$ and independent of $\hat{X}$. Showing the existence of a test channel in (16) is equivalent to showing that the function

$$\phi_{(s,r,\frac{\alpha}{D})}(t) = \frac{\phi_s(\alpha t)}{\phi_r(Dt)}, \tag{17}$$

is a valid characteristic function of some random variable $\hat{X}$ where $\phi_s(t)$ is characteristic function of X_s and $\phi_r(t)$ is a characteristic function of Z_r.

Observe that in (17) if $r = s$ we are exactly concerned with the self-decomposability property of the GG distribution. Thus, in a sense, when $r \neq s$ (17) is a generalization of the self-decomposability property. A question of whether a GG distribution is self-decomposable was only recently answered in [6]. The following theorem, which looks at the more general case of decomposability, provides a partial answer to when a GG distribution can be additively transformed into another GG distribution, and a complete answer to when a GG distribution is self-decomposable.

Theorem 6. *For $(r,s) \in \mathbb{R}_+^2$ let*

$$\begin{aligned}
\mathbb{S} &= \mathbb{S}_1 \cup \mathbb{S}_2, \\
\mathbb{S}_1 &= \{(r,s) : 2 < s < r\}, \\
\mathbb{S}_2 &= \{(r,s) : s = r \in (0,1] \cup \{2\}\}.
\end{aligned}$$

Then the function $\phi_{(s,r,\alpha)}(t)$ in (17) *has the following properties:*

- *for $(r,s) \in \mathbb{S}_2$, $\phi_{(s,r,\alpha)}(t)$ is a characteristic function (i.e., X_r is self-decomposable for $s = r \in (0,1] \cup \{2\}$);*
- *for $(r,s) \in \mathbb{R}_+^2 \setminus \mathbb{S}$, $\phi_{(s,r,\alpha)}(t)$ is not a characteristic function for any $\alpha \geq 1$; and*
- *for $(r,s) \in \mathbb{S}_1$ and* almost all[1] *$\alpha \geq 1$, $\phi_{(s,r,\alpha)}(t)$ is not a characteristic function.*

The result of Theorem 6 is depicted in Fig. 2.

Note that by Theorem 6 the function in (17) is a valid characteristic function if and only if $r = s \in (0,1] \cup \{2\}$. This concludes the proof. ■

Remark 2. *Note that using an additive backward test channel is not the only way of achieving equalities in (15). However, this is one of the most commonly used techniques and understanding its limitations can be very valuable.*

Note that Theorem 5 characterizes the distribution of an optimal reconstruction through its characteristic function.

[1] In other words, the set of α for which the statement does not hold has Lebesgue measure zero.

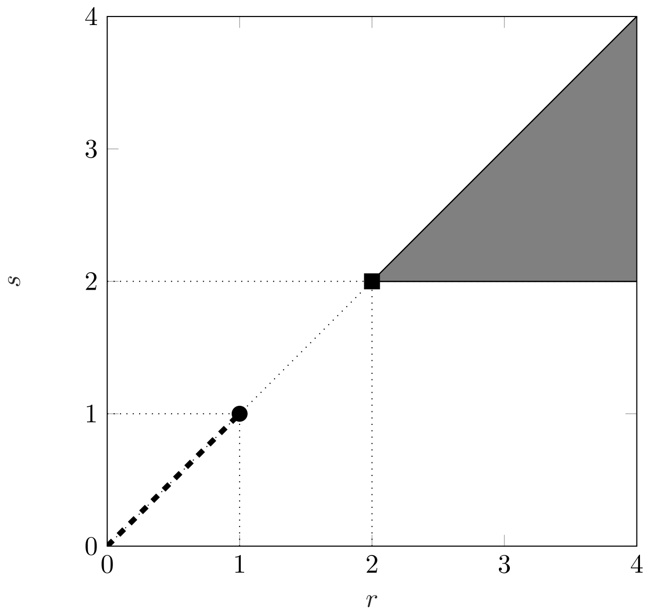

Fig. 2: In the regime $\mathbb{S}_2 = \{(r, s) : 0 < r = s < 1 \text{ or } r = s = 2\}$ (the dashed line) $\phi_{(s,r,\alpha)}(t)$ is self-decomposable. We also emphasize the point $(r, s) = (2, 2)$ (the black square) corresponds to the Gaussian characteristic function, and the point $(r, s) = (1, 1)$ (the black circle) corresponds to the Laplace characteristic function. The regime $\mathbb{S}_1 = \{(r, s) : 2 < s < r\}$ (the gray triangle) is where $\phi_{(s,r,\alpha)}(t)$ is not a characteristic function for almost all $\alpha \geq 1$. The white space is the regime where $\phi_{(s,r,\alpha)}(t)$ is not a characteristic function for all $\alpha \geq 1$.

However, this expression can be difficult to analyze. The following result offers some additional details about an optimal reconstruction. We will also need the following definition.

Definition 3. *A random variable X is infinitely divisible if, for every positive integer n, there exist n independent and identically distributed random variables $X_{n1}, ..., X_{nn}$ such that*

$$X \stackrel{d}{=} X_{n1} + ... + X_{nn}. \tag{18}$$

Proposition 1. *Let $r \in (0, 1] \cup \{2\}$. Then, for every $\alpha \geq 1$ there exists an infinitely divisible random variable $\hat{X}_\alpha$, which is an optimal reconstruction in Theorem 5, such that*

$$\alpha X_r \stackrel{d}{=} \hat{X}_\alpha + X'_r, \tag{19}$$

where X'_r is an independent copy of X_r. Moreover, for $\alpha > 1$ the random variable $\hat{X}_\alpha$ is unbounded.

Proof: The existence of $\hat{X}_\alpha$ and the fact that $\hat{X}_\alpha$ achieves Shannon's lower bound was shown in Theorem 5. The fact that a component $\hat{X}_\alpha$ of a self-decomposable distribution is an infinitely divisible random variable follows from [10, Corollary 5.11.1].

Next we show that this random variable must be unbounded. Assume that $|\hat{X}_\alpha| \leq A$ for some A. Then by the triangle inequality, for all $k \geq 1$

$$\begin{aligned}\mathbb{E}^{\frac{1}{k}}\left[|\alpha X_r|^k\right] &= \mathbb{E}^{\frac{1}{k}}\left[|\hat{X}_\alpha + X'_r|^k\right] \\ &\leq \mathbb{E}^{\frac{1}{k}}\left[|\hat{X}_\alpha|^k\right] + \mathbb{E}^{\frac{1}{k}}\left[|X'_r|^k\right].\end{aligned} \tag{20}$$

Moreover, since $E^{\frac{1}{k}}\left[|X_r|^k\right] = \mathbb{E}^{\frac{1}{k}}\left[|X'_r|^k\right]$ the bound in (20) implies that

$$(\alpha - 1)\mathbb{E}^{\frac{1}{k}}\left[|X_r|^k\right] \leq \mathbb{E}^{\frac{1}{k}}\left[|\hat{X}_\alpha|^k\right] \leq A. \tag{21}$$

However, since X_r is an unbounded random variable we have that $\lim_{k\to\infty} \mathbb{E}^{\frac{1}{k}}\left[|X_r|^k\right] = \infty$ which clearly violates (21). Therefore, $\hat{X}_\alpha$ is not bounded. This concludes the proof. ∎

Remark 3. *Observe that if one can argue that for all $r = s > 2$ the backward test channel is an additive transformation, then the technique used in Proposition 1 can be used to show that an optimal reconstruction is an unbounded random variable for all $r = s > 2$.*

VI. On Discreteness of An Optimal Reconstruction

In this section we show conditions under which an optimal reconstruction is discrete. The following theorem, the proof of which can be found in the extended version of the paper [11], is the main result of this section.

Theorem 7. *For $r < s < 2$ an optimal reconstruction distribution $P^\star_{\hat{X}}$ is discrete with finitely many points.*

Proof: The proof can be found in the extended version of the paper and relies on techniques introduced in [12] and [13]. ∎

Theorem 7 is interesting because it gives an example which shows that an optimal reconstruction of a continuous and unbounded random variable must be discrete with finitely many points and, hence, is also bounded.

VII. Conclusion

The lossy compression of GG sources of order s under ℓ_r distortion has been considered. It has been shown that for all $(r, s) \in \mathbb{R}^+$ an optimal reconstruction exists and properties of this reconstruction have been characterized. In particular, it has been shown that if $s \leq r-1$ then an optimal reconstruction is unbounded, and if $s > r$ then an optimal reconstruction is bounded. Further, it has been shown that for $r = s \in (0, 1] \cup \{2\}$ Shannon's lower bound on the rate-distortion function is achievable by a random variable that is unbounded and infinitely divisible. Finally, conditions under which an optimal reconstruction distribution is discrete with finitely many points have been demonstrated.

As a future direction it would be interesting to consider a remote rate-distortion problem where the goal is to compress a noisy version of the source [14]. Also, it would be interesting to consider dualities between the source and channel coding to address additive channels with GG noises [15].

APPENDIX

To show the existence of an optimizer we demonstrate that the set $\mathcal{P}_{r,s}(D)$ is compact in the topology of weak convergence. This together with the fact that mutual information is lower semicontinuous in the joint distribution $P_{X\hat{X}}$ implies, by the extreme value theorem, that the infimum is attainable. Before showing that $\mathcal{P}_{r,s}(D)$ is compact we present some definitions and results needed for the proof.

Definition 4. *A probability measure P_{XY} defined on $\mathbb{R}^2$ has a* marginal distribution *P_X on $\mathbb{R}$ if and only if for all bounded and continuous functions $f : \mathbb{R} \to \mathbb{R}$*

$$\int f(x) dP_{X,Y}(x, y) = \int f(x) dP_X(x). \tag{22}$$

Next, we show that the set of joint distributions with a fixed marginal distribution is closed in the topology of weak convergence where the notion of weak convergence is defined as follows.

Definition 5. *A sequence of probability measures $\{P_n\}_{n\in\mathbb{N}}$ is said to* converge weakly *to the probability measure P if*

$$\lim_{n\to\infty} \mathbb{E}_{P_n}[\phi(X)] \to \mathbb{E}_P[\phi(X)], \tag{23}$$

for all bounded and continuous functions ϕ.

Another main ingredient comprises linear functionals. The following theorem gives a necessary and sufficient condition for a linear functional to be weakly continuous [16, Lemma 2.1].

Theorem 8. *A linear functional $L : \mathcal{P} \to \mathbb{R}$ is* weakly continuous *on $\mathcal{P}$ if and only if it can be represented as*

$$L(P) = \mathbb{E}_P[\phi(X)], \quad \forall P \in \mathcal{P}$$

for some bounded and continuous function ϕ.

Theorem 9. *For some fixed P_X define*

$$\mathcal{P}_M(P_X) = \{P_{XY} \in \mathcal{P}(\mathbb{R}^2) : P_X \text{ is a marginal of } P_{XY}\}.$$

where $\mathcal{P}(\mathbb{R}^2)$ is the set of all joint distributions on $(X, Y) \in \mathbb{R}^2$. Then $\mathcal{P}_M(P_X)$ is a closed subset of $\mathcal{P}(\mathbb{R}^2)$.

Proof: Let $\mathcal{C}_b$ be a set of continuous and bounded functions on $\mathbb{R}$ and define the following operators:

$$L_1(P_{XY}; f) = \mathbb{E}_{P_{XY}}[f(X)], \tag{24}$$

$$L_2(P_X; f) = \mathbb{E}_{P_X}[f(X)], \tag{25}$$

for $f \in \mathcal{C}_b$. Clearly, $P_{XY} \mapsto L_1$ and $P_X \mapsto L_2$ are weakly continuous linear functionals in view of Theorem 8. Now for a fixed P_X define

$$\mathcal{P}(f; P_X) = \{P_{XY} \in \mathcal{P}(\mathbb{R}^2) : L_1(P_{XY}; f) = L_2(P_X; f)\}, \tag{26}$$

which is closed in view of the fact that $P_{XY} \mapsto L_1$ is weakly continuous. Therefore, using Definition 4 we can write $\mathcal{P}_{\text{M}}(P_X)$ as

$$\mathcal{P}_{\text{M}}(P_X) = \bigcap_{f\in\mathcal{C}_b} \mathcal{P}(f; P_X). \tag{27}$$

Since an arbitrary intersection of closed sets is closed, we have that $\mathcal{P}_{\text{M}}(P_X)$ is closed. This concludes the proof. ∎

Proposition 2. *For fixed $r, s, D > 0$ and $P_{X_s} = \mathcal{N}(0, \alpha^s)$, the set $\mathcal{P}_{r,s}(D)$ (defined in (5)) is a compact set.*

Proof: We note that the proof does not use the fact $P_{X_s} = \mathcal{N}(0, \alpha^s)$. Let

$$\mathcal{P}_{X\hat{X}} = \left\{ P_{X\hat{X}} : \mathbb{E}[|X - \hat{X}|^r] \leq \frac{2}{r} D^r \right\}. \tag{28}$$

From Prohorov's theorem and Markov's inequality it is not difficult to show that the set $\mathcal{P}_{X\hat{X}}$ is compact.

By Theorem 9 we have that $\mathcal{P}_{r,s}(D)$ is a closed subset of a compact set $\mathcal{P}_{X\hat{X}}$. Since an intersection of a compact set with a closed set is compact, we have that $\mathcal{P}_{r,s}(D)$ is compact. This concludes the proof. ∎

Finally, since the mapping $P_{X\hat{X}} \mapsto I(X; \hat{X})$ is lower semicontinuous and by Proposition 2 the set $\mathcal{P}_{r,s}(D)$ is compact, the extreme value theorem asserts that the infimum in (4) is attainable. This concludes the proof.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 379-423, 623-656, Jul.,-Oct. 1948.

[2] H. Si, O. O. Koyluoglu, and S. Vishwanath, "Lossy compression of exponential and laplacian sources using expansion coding," in *Proc. IEEE Int. Symp. Inf. Theory*, 2014, pp. 3052–3056.

[3] S. Verdú, "The exponential distribution in information theory," *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 100–111, 1996.

[4] F. Chen, Z. Gao, and J. Villasenor, "Lattice vector quantization of generalized Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 92–103, 1997.

[5] P. Loyer, J.-M. Moureaux, and M. Antonini, "Lattice codebook enumeration for generalized Gaussian source," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 521–527, 2003.

[6] A. Dytso, R. Bustin, H. V. Poor, and S. S. Shamai, "Analytical properties of generalized Gaussian distributions," www.princeton.edu/ ∼ adytso/papers/GGAnalytic.pdf, 2017.

[7] T. Cover and J. Thomas, *Elements of Information Theory: Second Edition*. Wiley, 2006.

[8] E. Lutwak, D. Yang, and G. Zhang, "Moment-entropy inequalities for a random vector," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1603–1607, April 2007.

[9] I. Csiszár, "On an extremum problem of information theory," *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, no. 1, pp. 57–71, 1974.

[10] E. Lukacs, *Characteristic Functions*. Griffin, 1970.

[11] A. Dytso, R. Bustin, H. V. Poor, and S. S. Shamai, "The generalized Gaussian distribution in information theory: Rate distortion," www.princeton.edu/ ∼ adytso/papers/GGInfoTheoreSourceCoding.pdf, 2017.

[12] A. Dytso, M. Goldenbaum, S. Shamai (Shitz), and H. V. Poor, "Upper and lower bounds on the capacity of amplitude-constrained MIMO channels," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Singapore, 2017, pp. 1–6.

[13] A. Dytso, M. Goldenbaum, H. V. Poor, and S. Shamai (Shitz), "When are discrete channel inputs optimal? – Optimization techniques and some new results," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, 2018.

[14] H. Witsenhausen, "Indirect rate distortion problems," *IEEE Trans. Inf. Theory*, vol. 26, no. 5, pp. 518–521, 1980.

[15] A. Dytso, R. Bustin, H. V. Poor, and S. S. Shitz, "On additive channels with generalized Gaussian noise," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 426–430.

[16] P. J. Huber, *Robust Statistics*. Wiley-Interscience, 1981.