Source Coding with Mismatched Distortion Measures

Urs Niesen, Devavrat Shah, and Gregory Wornell

***Abstract*— We consider the problem of lossy source coding with a mismatched distortion measure. That is, we investigate what distortion guarantees can be made with respect to distortion measure $\tilde{\rho}$, for a source code designed such that it achieves distortion less than D with respect to distortion measure ρ. We find a single-letter characterization of this mismatch distortion. We then study properties of this quantity and derive asymptotically tight bounds on it. These results give insight into the robustness of lossy source coding with respect to modelling errors in the distortion measure. They also provide guidelines on how to choose a good tractable approximation of an intractable distortion measure.**

I. Introduction

A. Problem Formulation

Let the source alphabet $\mathcal{X}$ and the reconstruction alphabet $\mathcal{Y}$ be (not necessarily finite) sets, and let $\{X_n\}_{n\geq 1}$ be i.i.d. random variables with distribution $P \in \mathcal{P}(\mathcal{X})$. Given two single-letter distortion measures $\rho, \tilde{\rho}$, i.e., functions $\mathcal{X}\times\mathcal{Y} \to \mathbb{R}_+$, let $\rho_n, \tilde{\rho}_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}_+$ be defined by

$$\rho_n(x^n, y^n) \triangleq \frac{1}{n}\sum_{i=1}^{n} \rho(x_i, y_i) \tag{1}$$

and analogously for $\tilde{\rho}_n$.

Assume we have access to an oracle that, when queried, produces a source code f_n (i.e., a mapping $f_n : \mathcal{X}^n \to \mathcal{Y}^n$) such that[1]

$$\frac{1}{n}\log|f_n(\mathcal{X}^n)| \leq R$$
$$\mathbb{E}\rho_n(X^n, f_n(X^n)) \leq D.$$

What guarantees can we make a priori (i.e., before querying the oracle) about $\mathbb{E}\tilde{\rho}_n(X^n, f_n(X^n))$?

This problem has the following operational significance. Let a source code of rate at most R and with expected distortion according to ρ of at most D be given. Assume instead of using this source code with respect to ρ, we decide to use it with respect to $\tilde{\rho}$. Such a situation occurs if constructing a source code for $\tilde{\rho}$ is not feasible or if $\tilde{\rho}$ is not fully known when constructing the source code. As an example, for an image compression problem, $\tilde{\rho}$ is determined by the human visual system, and any model ρ of it can necessarily be only an approximation of it. An answer to the above question allows thus to analyze the robustness of the coding scheme to modeling errors in the distortion measure.

This work was supported in part by NSF under Grant No. CCF-0515109, and by HP through the MIT/HP Alliance.

The authors are with the Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA 02139, USA. {uniesen,devavrat,gww}@mit.edu

[1] $|f_n(\mathcal{X}^n)|$ denotes the cardinality of the range of the function f_n.

B. Related Work

The question of mismatched distortion measures in source coding has previously been considered in [1], [2], [3], [4], and [5]. In these works the mismatch is only with respect to the encoding part of the source code, whereas at least the decoder is matched to the proper distortion measure. This differs from the setup here, where the mismatch is with respect to both, the encoder and the decoder. We comment on the precise differences in the following paragraphs.

In [1] a partial order among distortion measures is defined such that $\rho \geq \tilde{\rho}$ if for every source code (consisting of an encoder $g_n : \mathcal{X}^n \to \{1,\ldots,\exp(nR)\}$ and a decoder $\phi_n : \{1,\ldots,\exp(nR)\} \to \mathcal{Y}^n$) satisfying $\mathbb{E}\rho_n(X^n, \phi_n(g_n(X^n))) \leq D$ there exists a second decoder $\tilde{\phi}_n$ satisfying $\mathbb{E}\tilde{\rho}_n(X^n, \tilde{\phi}_n(g_n(X^n))) \leq D$. Thus, in this setup, the encoder g_n is designed for a mismatched distortion measure ρ, whereas the decoder $\tilde{\phi}_n$ is matched to the distortion measure $\tilde{\rho}$.

In [2] the following problem is considered. Let $\mathcal{C} \subset \mathcal{Y}^n$ and $g_n : \mathcal{X}^n \to \mathcal{C}$ be an optimal encoder with respect to ρ. Find $\mathcal{C}$ and decoder $\tilde{\phi}_n : \mathcal{C} \to \mathcal{Y}^n$ such that $\mathbb{E}\tilde{\rho}_n(X^n, \tilde{\phi}_n(g_n(X^n)))$ is minimized. Again, the mismatch is only with respect to the encoder g_n, whereas the decoder as well as the codebook $\mathcal{C}$ are matched to the distortion measure $\tilde{\rho}$.

In [3] the problem of finding an encoder $g_n : \mathcal{X}^n \to \{1,\ldots,\exp(nR)\}$ such that there exists a decoder $\phi_n : \{1,\ldots,\exp(nR)\} \to \mathcal{Y}$ satisfying $\mathbb{E}\rho_n(X^n, \phi_n(g_n(X^n))) \leq D$ while maximizing $\inf_{\tilde{\phi}_n} \mathbb{E}\tilde{\rho}_n(X^n, \tilde{\phi}_n(g_n(X^n)))$ is investigated. In other words, the goal is to find an encoder that guarantees distortion at most D with respect to ρ, while making sure that this code has maximum possible distortion with respect to $\tilde{\rho}$. As in the previous cases, the mismatch is only with respect to the encoder, the decoder $\tilde{\phi}_n$ is matched to the distortion measure $\tilde{\rho}$.

In [4, Problem 2.2.14] and [5] the problem of lossy source coding with respect to a class of distortion measures is considered: Given a class of distortion measures Γ, we want to find a source code $f_n : \mathcal{X}^n \to \mathcal{Y}^n$ such that $\sup_{\rho\in\Gamma} \mathbb{E}\rho_n(X^n, f_n(X^n))$ is minimized. In other words, f_n is now "matched" to all $\rho \in \Gamma$ simultaneously.

C. Modeling Perceptual Distortion Measures

In this section, we briefly review the typical structure of perceptual distortion measures. This will motivate the results presented in the main text. We focus here on distortion measures for image compression; the structure of perceptual distortion measures for speech, audio, or video compression is similar (see [6] for details on those distortion measures). The discussion here follows [7] and [8].

The typical structure of a perceptual distortion measure for image compression is depicted in Figure 1. Here x and y are the original and reconstructed image respectively, represented, for example, as vector of gray scale values.

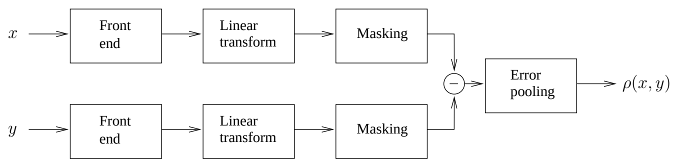

Fig. 1. Typical structure of a perceptual distortion measure. Adapted from [7].

The first block (termed front end), contains conversions from the image format used to physical luminance observed by the human eye and other calibrations. The second block performs a linear transform of the two images, usually decomposing it into a number of spatial frequency bands with different orientations. In the next block, the coefficient of each band is weighted to account for masking effects. The resulting vector of weighted coefficients of the original and reconstructed image are then subtracted. The last block takes this vector of weighted differences and pools it together into one real number. Usually this is done by computing the ℓ_p norm of the difference vector for some $p \geq 1$ or taking some power $r \geq 1$ of that norm. Typical values of p range from 2 to 4.

Formally, the source and reconstruction alphabets are $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$ or $\mathcal{X} = \mathcal{Y} = [0,1]^m$ for some finite m. In the following, we write x, y for elements of general $\mathcal{X}$, $\mathcal{Y}$, and we write $\boldsymbol{x}, \boldsymbol{y}$ if we want to emphasize that $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$ or $\mathcal{X} = \mathcal{Y} = [0,1]^m$. This means that ρ is of the form

$$\begin{aligned}&\rho(\boldsymbol{x}, \boldsymbol{y})\\&= \big\| [v(x_1), \dots, v(x_m)]\boldsymbol{W_x} - [v(y_1), \dots, v(y_m)]\boldsymbol{W_y} \big\|_p^r,\end{aligned}$$

and is sometimes simplified to

$$\begin{aligned}&\rho(\boldsymbol{x}, \boldsymbol{y})\\&= \big\| \big([v(x_1), \dots, v(x_m)] - [v(y_1), \dots, v(y_m)]\big)\boldsymbol{W_x} \big\|_p^r. \quad (2)\end{aligned}$$

$v : \mathbb{R} \to \mathbb{R}$ accounts for the front end, $\boldsymbol{W} : \mathbb{R}^m \to \mathbb{R}^{m \times k}$ accounts for the linear transform and masking. Here (and in the following), we write for $\boldsymbol{a} \in \mathbb{R}^k$ and $p \geq 1$

$$\|\boldsymbol{a}\|_p \triangleq \begin{cases} \big(\sum_{i=1}^k |a_i|^p\big)^{1/p} & \text{if } p < \infty, \\ \max_{1 \leq i \leq k} |a_i| & \text{if } p = \infty. \end{cases}$$

D. Outline of Results

We now discuss several questions that arise when trying to construct and use perceptual distortion measures for source coding. These questions motivate the results presented in this paper, and they are used as examples throughout.

- The choice of r and p for the error pooling seems to vary quite considerably across different perceptual distortion measures for image compression. [9] uses $p = 2, r = 1$, [10] uses $p = 2.4, r = 1$, [11] uses $p = 4, r = 1$, and [12], [13] use $p = 2, r = 2$. It is therefore of interest to know how distortion mismatch in these two parameters affect the performance of the source code. This is discussed in Example 1 (using Theorems 1, 2, 3) and Example 3 (using Theorem 5).
- Given a class of distortion measures Γ, [12] suggests the following approach to find the "best" approximation $\rho \in \Gamma$ to the distortion measure implemented by the human visual system: Simulate the (information theoretically) optimal encoding scheme for all $\rho \in \Gamma$, and determine experimentally (i.e., by showing the original and distorted image to a human) the one yielding the smallest distortion. This optimal distortion measure is then declared to be the best approximation. While this approach yields indeed the best approximation $\rho \in \Gamma$ when used with the *optimal* infinite length source code, it is not clear a priori if this ρ will also yield a good approximation when used with a *suboptimal* source code. Indeed, as we shall see in Example 1, there are situations in which the mismatch for the optimal and (even only slightly) suboptimal source codes are very different. In Example 2 (using Theorem 4), we provide conditions on Γ and the source under which the ρ found with this approach yields also a good approximation when used with good but not optimal source codes. These conditions hold for the model in [12] (with a few additional assumptions, that are implicitly made there). Hence our results provide evidence that the optimal approximation $\rho \in \Gamma$ found in [12] will also be good for practical (and hence necessarily suboptimal) source codes.
- [13] proposes a vector quantizer design procedure for distortion measures of the form

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = w_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{x}\|_2^2, \quad (3)$$

where $w : \mathbb{R}^m \to \mathbb{R}$. Since this is considerably simpler than the standard model (2), the question arises of how to find the $w_{\boldsymbol{x}}$ such that the resulting ρ in (3) is "close" to one of the more complicated form (2). Note that it is not immediately obvious what "close" should mean in this context. Indeed, there are several such notions that are reasonable. In Example 5, we show what properties such a notion should have. The problem posed by [13] discussed above is treated in detail in Example 4 (using Theorem 5) and Example 6 (using Corollaries 6 and 7).
- Essentially all models of perceptual distortion measures contain a number of parameters that are usually chosen to be in "close agreement" with the behavior of the human visual system. Again, it is not clear what "close agreement" should mean here. In Example 7 (using Proposition 8), a simple such measure of closeness is proposed, providing a guideline for how to tune the parameters of a perceptual distortion model to be used for source coding.

E. Organization

The remainder of this paper is organized as follows. In Section II-A, we provide a single-letter characterization of the mismatch distortion. In Section II-B, we investigate properties of the mismatch distortion. Section II-C, considers the problem of finding a good representation of a distortion measure from a class of simpler ones. Section III contains concluding remarks.

Due to space constraints, all results in this paper are presented without proofs.

II. Source Coding with Distortion Mismatch

In this section, we formally introduce the problem of source coding with distortion mismatch. In what follows, we will let the source alphabet $\mathcal{X}$ and the reconstruction alphabet $\mathcal{Y}$ be Polish[2]. We let $\mathcal{B}(\mathcal{X}\times\mathcal{Y})$ be the Borel sets of $\mathcal{X}\times\mathcal{Y}$. By $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, we denote the set of all probability measures on $(\mathcal{X}\times\mathcal{Y}, \mathcal{B}(\mathcal{X}\times\mathcal{Y}))$. For $Q\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, Q_X denotes the $\mathcal{X}$ marginal of Q. For a measurable function $g:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}$, we denote by $\mathbb{E}_Q g(X,Y)$ or $\mathbb{E}_Q g$ the expectation of $g(X,Y)$ with respect to Q. For any $A\in\mathcal{B}(\mathcal{X}\times\mathcal{Y})$, we write $\mathbb{E}_Q(g;A)$ for $\mathbb{E}_Q g\mathbb{1}_A$. $I(Q)$ denotes mutual information (in nats) between the random variables $(X,Y)\sim Q$. Throughout this paper, we restrict attention to single-letter distortion measures, i.e., measurable functions $\rho:\mathcal{X}\times\mathcal{Y}\to\mathbb{R}_+$ with $\rho_n:\mathcal{X}^n\times\mathcal{Y}^n\to\mathbb{R}_+$ as defined in (1). $R_\rho(D)$ and $D_\rho(R)$ denote the rate-distortion and the distortion-rate function for the source $\{X_n\}_{n\geq 1}$ and with respect to the single-letter distortion measure ρ. More precisely, if $\{X_n\}_{n\geq 1}$ is i.i.d. $P\in\mathcal{P}(\mathcal{X})$ then

$$R_\rho(D) \triangleq \inf_{\substack{Q\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}):\\ Q_X=P, \mathbb{E}_Q\rho\leq D}} I(Q),$$

$$D_\rho(R) \triangleq \inf_{\substack{Q\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}):\\ Q_X=P, I(Q)\leq R}} \mathbb{E}_Q\rho.$$

A. Single-Letter Characterization

In this section, we provide a single-letter characterization of the smallest distortion with respect to $\tilde{\rho}$ that can be guaranteed for any source code of rate R designed for distortion D_ρ with respect to ρ.

Define

$$D_{\rho,\tilde{\rho}}(R, D_\rho) \triangleq \sup \mathbb{E}_Q\tilde{\rho}, \tag{4}$$

where the supremum is taken over all $Q\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ such that $Q_X=P$, $\mathbb{E}_Q\rho\leq D_\rho$ and $I(Q)\leq R$. If the set over which this supremum is taken is empty, we define $D_{\rho,\tilde{\rho}}(R,D_\rho)\triangleq -\infty$.

Theorem 1. *Let $\rho,\tilde{\rho}$ be distortion measures with $\mathbb{E}_P\rho(X,y_0)<\infty$ for some $y_0\in\mathcal{Y}$. For every $D_{\tilde{\rho}}<\infty$ such that*

$$0\leq D_{\tilde{\rho}} < \lim_{\delta\downarrow 0} D_{\rho,\tilde{\rho}}(R-\delta, D_\rho-\delta)$$

there exists a sequence of source codes $\{f_n\}_{n\geq 1}$ such that

$$\lim_{n\to\infty}\frac{1}{n}\log|f_n(\mathcal{X}^n)|\leq R,$$
$$\limsup_{n\to\infty}\mathbb{E}\rho_n(X^n, f_n(X^n))\leq D_\rho,$$
$$\liminf_{n\to\infty}\mathbb{E}\tilde{\rho}_n(X^n, f_n(X^n))\geq D_{\tilde{\rho}}.$$

Theorem 2. *If there exists an n and a source code $f_n:\mathcal{X}^n\to\mathcal{Y}^n$ such that*

$$\frac{1}{n}\log|f_n(\mathcal{X}^n)|=R,$$
$$\mathbb{E}\rho_n(X^n, f_n(X^n))\leq D_\rho,$$
$$\mathbb{E}\tilde{\rho}_n(X^n, f_n(X^n))\geq D_{\tilde{\rho}},$$

then[3] $D_{\tilde{\rho}}\leq D_{\rho,\tilde{\rho}}(R+, D_\rho)$ and $D_{\tilde{\rho}}\leq D_{\rho,\tilde{\rho}}(R, D_\rho)$ if $R>R_\rho(D_\rho)$.

Theorems 1 and 2 allow us to make guarantees about the performance of a source code constructed with mismatched distortion measure. Indeed, if $f_n:\mathcal{X}^n\to\mathcal{Y}^n$ is a source code of rate R designed for a distortion measure ρ and distortion level D_ρ, then by Theorem 2, f_n is also a source code for any distortion measure $\tilde{\rho}$ and distortion level $D_{\rho,\tilde{\rho}}(R+,D_\rho)$. Moreover, this is essentially the best guarantee one can make, since by Theorem 1 there exist source codes with same blocklength n and same rate R designed for distortion measure ρ and distortion level D_ρ that result in a distortion level of more than

$$D_{\rho,\tilde{\rho}}(R-\delta(n), D_\rho-\delta(n))-\delta(n)$$

for distortion measure $\tilde{\rho}$ with $\delta(n)\to 0$ as $n\to\infty$.

B. Properties of $D_{\rho,\tilde{\rho}}(R,D_\rho)$

The function $D_{\rho,\tilde{\rho}}(R,D_\rho)$ exhibits the following behavior:

$$D_{\rho,\tilde{\rho}}(R,D_\rho)\in\begin{cases}\{-\infty\} & \text{if } R<R_\rho(D_\rho),\\ \mathbb{R}_+\cup\{\pm\infty\} & \text{if } R=R_\rho(D_\rho),\\ \mathbb{R}_+\cup\{\infty\} & \text{if } R>R_\rho(D_\rho).\end{cases}$$

Moreover, a simple argument shows that $D_{\rho,\tilde{\rho}}(R,D_\rho)$ is concave and increasing in both its arguments, and continuous at all points (R,D_ρ) such that $R>R_\rho(D_\rho)$. $D_{\rho,\tilde{\rho}}(R,D_\rho)$ is necessarily discontinuous at $(R_\rho(D_\rho),D_\rho)$, but could be either left- or right-continuous (as a function of either R or D_ρ). This implies that the function either equals ∞ for all (R,D_ρ) such that $R>R_\rho(D_\rho)$ or is finite on this whole range. The two types of possible behaviors of $D_{\rho,\tilde{\rho}}(R,D_\rho)$ are depicted in Figure 2.

The next three theorems describe the behavior of $D_{\rho,\tilde{\rho}}(R,D_\rho)$ in more detail. Theorem 3 provides conditions under which $D_{\rho,\tilde{\rho}}(R,D_\rho)=\infty$ for all (R,D_ρ) such that $R>R_\rho(D_\rho)$. In these situations, we cannot make any guarantees about the performance of a source code of rate R designed for distortion measure ρ and distortion level D_ρ when used for

[2]i.e., complete, separable, metric spaces (e.g., $\mathbb{R}^m$ or $[0,1]^m$ for some finite m)

[3]For a real valued function g, we write $g(x+)\triangleq\lim_{\delta\downarrow 0}g(x+\delta)$ and $g(x-)\triangleq\lim_{\delta\downarrow 0}g(x-\delta)$ assuming the limits exist.

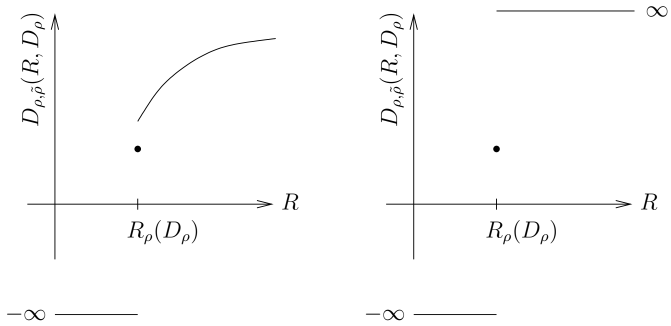

Fig. 2. Possible behaviors of $D_{\rho,\tilde{\rho}}(R, D_\rho)$.

distortion measure $\tilde{\rho}$. Theorem 4 gives sufficient conditions such that $D_{\rho,\tilde{\rho}}(R, D_\rho) \geq 0$ for (R, D_ρ) with $R = R_\rho(D_\rho)$, and conditions for $D_{\rho,\tilde{\rho}}(R, D_\rho)$ to be right-continuous in R at those points. Theorem 5 provides a limiting expression for $D_{\rho,\tilde{\rho}}(R, D_\rho)$ as $R \to \infty$. Since $D_{\rho,\tilde{\rho}}(R, D_\rho)$ is increasing in R, this limiting expression is also a bound on $D_{\rho,\tilde{\rho}}(R, D_\rho)$ for any finite R, and in particular is the best distortion guarantee that is independent of the rate R.

Theorem 3. *If*

i) $0 < R < \infty$
ii) $D_\rho(R) < D_\rho < \infty$
iii) *there exists* $y_0 \in \mathcal{Y}$ *such that* $\mathbb{E}_P \rho(X, y_0) < \infty$
iv) *there exist* $\{A_n\}_{n\geq 1} \subset \mathcal{B}(\mathcal{X})$, $\{y_n^*\}_{n\geq 1} \subset \mathcal{Y}$ *such that*

$$\begin{aligned} \mathbb{E}_P(\rho(X, y_n^*); A_n) < \infty \qquad & \text{for all } n \geq 1, \\ P(A_n) \inf_{x\in A_n} \tilde{\rho}(x, y_n^*) \to \infty \qquad & \text{as } n \to \infty, \\ \sup_{x\in A_n} \rho(x, y_n^*)/\tilde{\rho}(x, y_n^*) \to 0 \qquad & \text{as } n \to \infty \end{aligned}$$

then $D_{\rho,\tilde{\rho}}(R, D_\rho) = \infty$.

Remark. For $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$, the second and third part of assumption iv) are satisfied for example if $\tilde{\rho}(x, y) \to \infty$ and $\rho(x, y)/\tilde{\rho}(x, y) \to 0$ when $\|y - x\|_2 \to \infty$. See also Example 1.

Example 1. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$ for some $m \in \mathbb{N}$, and assume $\rho(x, y) = d(y - x)^r, \tilde{\rho}(x, y) = \tilde{d}(y - x)^{\tilde{r}}$ for norms $d, \tilde{d}$ and $r, \tilde{r} \geq 1$. Let $P \in \mathcal{P}(\mathcal{X})$ such that $\mathbb{E}_P d(X)^{\lceil r\rceil} < \infty$.

Since all norms on a finite dimensional space are equivalent, there exist $a_1, a_2 > 0$ such that

$$a_1 d(z) \leq \tilde{d}(z) \leq a_2 d(z)$$

for all $z \in \mathbb{R}^m$, and thus there exist $b_1, b_2 > 0$ such that

$$b_1 \rho(x, y)^{\tilde{r}/r} \leq \tilde{\rho}(x, y) \leq b_2 \rho(x, y)^{\tilde{r}/r}$$

for all $x \in \mathcal{X}, y \in \mathcal{Y}$.

Case 1: $\tilde{r} \leq r$. Then

$$\begin{aligned} \tilde{\rho}(x, y) &\leq b_2 \rho(x, y)^{\tilde{r}/r} \\ &\leq b_2 \max\{1, \rho(x, y)\} \\ &\leq b_2(1 + \rho(x, y)), \end{aligned}$$

and therefore for all $Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ such that $\mathbb{E}_Q \rho \leq D_\rho < \infty$

$$\mathbb{E}_Q \tilde{\rho} \leq b_2(1 + D_\rho) < \infty.$$

This implies $D_{\rho,\tilde{\rho}}(R, D_\rho) < \infty$ for all $R, D_\rho \in \mathbb{R}_+$.

Case 2: $\tilde{r} > r$. We first show that the conditions of Theorem 3 are satisfied. We have

$$\rho(x, y)/\tilde{\rho}(x, y) \leq b_1 \rho(x, y)^{(r-\tilde{r})/r}$$

for all $x \in \mathcal{X}, y \in \mathcal{Y}$. Let $A \triangleq [-c, c]^m$, and choose c such that $P(A) > 0$. Set $y_n^* \triangleq n\mathbf{1}$, where $\mathbf{1} = (1, \ldots, 1) \in \mathbb{R}^m$. With this

$$\begin{aligned} \sup_{x\in A} \rho(x, y_n^*)/\tilde{\rho}(x, y_n^*) &\leq \sup_{x\in A} b_1 \rho(x, y_n^*)^{(r-\tilde{r})/r} \\ &= \max_{x\in A} b_1 d(y_n^* - x)^{r-\tilde{r}} \to 0 \end{aligned}$$

as $n \to \infty$. Moreover,

$$P(A) \inf_{x\in A} \tilde{\rho}(x, y_n^*) = P(A) \min_{x\in A} \tilde{d}(y_n^* - x)^{\tilde{r}} \to \infty$$

as $n \to \infty$. Finally, with $B_n \triangleq \{x : d(y_n^* - x) \geq 1\}$

$$\begin{aligned} \mathbb{E}_P \rho(X, y_n^*) &\leq P(B_n^c) + \mathbb{E}_P\big(d(y_n^* - X)^r; B_n\big) \\ &\leq 1 + \mathbb{E}_P\big((d(y_n^*) + d(X))^{\lceil r\rceil}; B_n\big) \\ &\leq 1 + \sum_{i=0}^{\lceil r\rceil} \binom{\lceil r\rceil}{i} d(y_n^*)^{\lceil r\rceil - i} \mathbb{E}_P d(X)^i \\ &< \infty, \end{aligned}$$

and hence with $y_0 = 0$, we have $\mathbb{E}_P \rho(X, y_0) < \infty$ and $\mathbb{E}_P(\rho(X, y_n^*); A) < \infty$. Thus applying Theorem 3 with $A_n \triangleq A$ yields $D_{\rho,\tilde{\rho}}(R, D_\rho) = \infty$ for all $0 < R < \infty$ and $D_\rho(R) < D_\rho < \infty$. $\Diamond$

Theorem 3 characterizes the behavior of $D_{\rho,\tilde{\rho}}(R, D_\rho)$ for (R, D_ρ) such that $R > R_\rho(D_\rho)$. The next theorem characterizes the behavior of $D_{\rho,\tilde{\rho}}(R, D_\rho)$ for (R, D_ρ) such that $R = R_\rho(D_\rho)$.

Theorem 4. *Let the distortion measure* ρ *be continuous,* $D_\rho > 0$*. If* $\mathcal{Y}$ *is compact or if there exists compact sets* $K_n \subset \mathcal{X}, M_n \subset \mathcal{Y}$ *such that* $P(K_n) \to 1$ *as* $n \to \infty$ *and*

$$\inf_{x\in K_n, y\in M_n^c} \rho(x, y) \to \infty \tag{5}$$

as $n \to \infty$*. Then* $D_{\rho,\tilde{\rho}}(R_\rho(D_\rho), D_\rho) \geq 0$*, i.e., the set over which we optimize in* (4) *is non-empty.*

If, in addition, $D_{\rho,\tilde{\rho}}(R_\rho(D_\rho) + r, D_\rho) < \infty$ *for some* $r > 0$*,* $\tilde{\rho}$ *is continuous, and there exists* $a > 1$ *and* $c \in \mathbb{R}$ *such that* $\tilde{\rho}^a \leq c + \rho$*, then*

$$D_{\rho,\tilde{\rho}}(R_\rho(D_\rho)+, D_\rho) = D_{\rho,\tilde{\rho}}(R_\rho(D_\rho), D_\rho).$$

Remark. For $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$ (5), is satisfied for example for ρ such that $\rho(x, y) \to \infty$ as $\|y - x\|_2 \to \infty$. Indeed, for $K_n = [-n, n]^m$ and $M_n = [-2n, 2n]^m$,

$$\lim_{n\to\infty} P(K_n) = 1,$$

and

$$\inf_{x\in K_n, y\in M_n^c} \rho(x,y) \geq \min_{x,y:\|y-x\|_2\geq n} \rho(x,y) \to \infty$$

as $n \to \infty$.

Example 2. Given a class of distortion measures Γ, the following approach is suggested in [12] to find the "closest" one to $\tilde{\rho}$ implemented by the human visual system: Determine $D_{\rho,\tilde{\rho}}(R, D_\rho(R))$ for each $\rho \in \Gamma$ and pick a minimizer ρ^*. In situations where a unique distribution Q with $Q_X = P$ achieving $D_\rho(R)$ exists, $D_{\rho,\tilde{\rho}}(R, D_\rho(R))$ can be found empirically by generating samples from Q and having them evaluated by human subjects. The hope is that the distortion measure minimizing $D_{\rho,\tilde{\rho}}(R, D_\rho(R))$ should be a good approximation to $\tilde{\rho}$ also for non-optimal image compression schemes. Formally, this amounts to assuming that $D_{\rho,\tilde{\rho}}(R+r, D_\rho(R))$ is close to $D_{\rho,\tilde{\rho}}(R, D_\rho(R))$ (at least for small r). Hence this approach is only valid, if $D_{\rho,\tilde{\rho}}(R+r, D_\rho(R))$ is right continuous in r at $r = 0$.

Theorem 4 gives conditions under which this is indeed the case. In [12], $\mathcal{X} = \mathcal{Y} = \mathbb{R}_+^m$, and each $\rho \in \Gamma$ is of the form

$$\rho(\boldsymbol{x}, \boldsymbol{y}) = \left\| \left([v(x_1), \dots, v(x_m)] - [v(y_1), \dots, v(y_m)] \right) \boldsymbol{W} \right\|_2^2$$

for some monotonic increasing concave function $v : \mathbb{R}_+ \to \mathbb{R}$ and some matrix $\boldsymbol{W} \in \mathbb{R}^{m\times m}$. In order to apply Theorem 4, we need the additional assumptions that v is continuous at 0, that $v(s) \to \infty$ as $s \to \infty$, that $\boldsymbol{W}^T\boldsymbol{W}$ is positive definite, and (the reasonable assumption) that $\tilde{\rho}$ implemented by the human visual system is continuous and bounded. From Theorem 4, we obtain that under these slightly stronger conditions than in [12], $D_{\rho,\tilde{\rho}}(R+r, D_\rho(R))$ is indeed right continuous at $r = 0$, showing that ρ^* should yield a good approximation to $\tilde{\rho}$ also for compression schemes that are only close to optimal.

We consider the problem of finding an optimal $\rho \in \Gamma$ approximating a given $\tilde{\rho}$ in more detail in Section II-C. $\diamondsuit$

The next theorem provides an upper bound on $D_{\rho,\tilde{\rho}}(R, D_\rho)$, independent of R. This bound is equal to $\lim_{R\to\infty} D_{\rho,\tilde{\rho}}(R, D_\rho)$, and is hence the tightest such bound possible. We shall see in Example 4 that this bound can be quite good for small values of R.

Theorem 5. *If*

i) *$\rho, \tilde{\rho}$ are continuous*
ii) *there exists $y_0 \in \mathcal{Y}$ such that $\mathbb{E}_P \rho(X, y_0) < \infty$*
iii) *$D_\rho(\infty) < D_\rho < \infty$*

then for any $\eta \geq 0$ the expectation

$$\mathbb{E}_P \sup_{y\in\mathcal{Y}} (\tilde{\rho}(X,y) - \eta\rho(X,y))$$

is well defined and

$$D_{\rho,\tilde{\rho}}(\infty, D_\rho) = \min_{\eta\geq 0} \left(\eta D_\rho + E_P \sup_{y\in\mathcal{Y}} (\tilde{\rho}(X,y) - \eta\rho(X,y)) \right).$$

If, moreover, $D_{\rho,\tilde{\rho}}(\infty, D_\rho) < \infty$, then

$$\lim_{R\to\infty} D_{\rho,\tilde{\rho}}(R, D_\rho) = D_{\rho,\tilde{\rho}}(\infty, D_\rho).$$

Example 3. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$, $\rho(x,y) = d(y-x)^r$, $\tilde{\rho}(x,y) = \tilde{d}(y-x)^{\tilde{r}}$ for norms $d, \tilde{d}$, and for $r, \tilde{r} \geq 1$. Let $P \in \mathcal{P}(X)$ be such that $\mathbb{E}_P d(X)^{\lceil r \rceil} < \infty$. With slight abuse of notation, we shall write $\rho(x-y)$ for $\rho(x,y)$ and similar for $\tilde{\rho}$ in this example. Set

$$v^* \in \operatorname*{arg\,max}_{v\in\mathbb{R}^m: d(v)=1} \tilde{d}(v).$$

Since $\tilde{d}$ is continuous and $\{v : d(v) = 1\}$ is compact, at least one such maximizer exists. It is easy to check that

$$\sup_{z\in\mathbb{R}^m} \tilde{\rho}(z) - \eta\rho(z) = \sup_{a\geq 0} a^{\tilde{r}} \tilde{d}(v^*)^{\tilde{r}} - \eta a^r, \tag{6}$$

i.e., the maximizing z is of the form $z^* = av^*$ for some $a \geq 0$.

Case 1: $r < \tilde{r}$. We have seen in Example 1 that then $D_{\rho,\tilde{\rho}}(R, D_\rho) = \infty$ for $R > R_\rho(D_\rho)$.

Case 2: $r = \tilde{r}$. From Theorem 5, we have for $D_\rho(\infty) < D_\rho < \infty$

$$D_{\rho,\tilde{\rho}}(\infty, D_\rho) = \min_{\eta\geq 0} \eta D_\rho + \sup_{z\in\mathbb{R}^m} \tilde{\rho}(z) - \eta\rho(z).$$

Let $z = av^*$ for some $a \geq 0$. Then

$$\tilde{\rho}(z) - \eta\rho(z) = a^r(\tilde{d}(v^*)^r - \eta) \to \infty$$

as $a \to \infty$, provided that $\eta < \tilde{d}(v^*)^r$. On the other hand, if $\eta \geq \tilde{d}(v^*)^r$, then for any $z \in \mathbb{R}^m$ with $d(z) = 1$

$$\tilde{\rho}(z) - \eta\rho(z) \leq \tilde{d}(v^*)^r - \eta \leq 0,$$

and hence $\tilde{\rho}(z) - \eta\rho(z) \leq 0$ for all $z \in \mathbb{R}^m$, with equality for $z = 0$. Therefore the minimizing $\eta \geq 0$ is equal to $\tilde{d}(v^*)^r$ and

$$\lim_{R\to\infty} D_{\rho,\tilde{\rho}}(R, D_\rho) = D_\rho \tilde{d}(v^*)^r.$$

Case 3: $r > \tilde{r}$. Recall that by (6)

$$\sup_{a\geq 0} a^{\tilde{r}} \tilde{d}(v^*)^{\tilde{r}} - \eta a^r = \sup_{z\in\mathbb{R}^m} \tilde{\rho}(z) - \eta\rho(z).$$

The optimal $a^* \geq 0$ maximizing this quantity is

$$a^* = \left(\frac{\tilde{r}}{\eta r} \tilde{d}(v^*)^{\tilde{r}} \right)^{1/(r-\tilde{r})},$$

which by Theorem 5 implies that for $D_\rho(\infty) < D_\rho < \infty$

$$D_{\rho,\tilde{\rho}}(\infty, D_\rho) = \min_{\eta\geq 0} \eta D_\rho + \eta^{-\tilde{r}/(r-\tilde{r})} b \triangleq \min_{\eta\geq 0} g(\eta),$$

where

$$b \triangleq \tilde{d}(v^*)^{\tilde{r}r/(r-\tilde{r})} \left(\left(\frac{\tilde{r}}{r} \right)^{\tilde{r}/(r-\tilde{r})} - \left(\frac{\tilde{r}}{r} \right)^{r/(r-\tilde{r})} \right) > 0.$$

The η^* minimizing g is

$$\eta^* = \left(\frac{r-\tilde{r}}{b\tilde{r}} D_\rho \right)^{(\tilde{r}-r)/r},$$

which finally yields

$$\lim_{R\to\infty} D_{\rho,\tilde{\rho}}(R, D_\rho) = D_\rho^{\tilde{r}/r} \left(\frac{b}{r-\tilde{r}} \right)^{(r-\tilde{r})/r} \tilde{r}^{-\tilde{r}/r} r.$$

For $m = 1$, $r = 2$, $\tilde{r} = 1$, this reduces to

$$\lim_{R\to\infty} D_{\rho,\tilde{\rho}}(R, D_\rho) = \sqrt{D_\rho}.$$

Note that in this case the limiting expression does not depend on the norms d and $\tilde{d}$. $\diamondsuit$

Example 4. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$, $\rho(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{x})^T \boldsymbol{W}_{\boldsymbol{x}} (\boldsymbol{y} - \boldsymbol{x})$, $\tilde{\rho}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{x})^T \widetilde{\boldsymbol{W}}_{\boldsymbol{x}} (\boldsymbol{y} - \boldsymbol{x})$, where $\boldsymbol{W}_{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{W}}_{\boldsymbol{x}}$ are positive definite for P almost every $\boldsymbol{x}$. Let $P \in \mathcal{P}(\mathcal{X})$ such that $\mathbb{E}_P \boldsymbol{X}^T \boldsymbol{W}_{\boldsymbol{X}} \boldsymbol{X} < \infty$.

Hence Theorem 5 yields that for $D_\rho(\infty) < D_\rho < \infty$,

$$\begin{aligned} D_{\rho,\tilde{\rho}}(\infty, D_\rho) &= \min_{\eta \geq 0} \eta D_\rho \\ &\quad + \mathbb{E}_P \sup_{\boldsymbol{y} \in \mathbb{R}^m} (\boldsymbol{y} - \boldsymbol{X})^T (\widetilde{\boldsymbol{W}}_{\boldsymbol{X}} - \eta \boldsymbol{W}_{\boldsymbol{X}})(\boldsymbol{y} - \boldsymbol{X}), \end{aligned}$$

and whenever this quantity is finite then also

$$\lim_{R \to \infty} D_{\rho,\tilde{\rho}}(R, D_\rho) = D_{\rho,\tilde{\rho}}(\infty, D_\rho).$$

If $\widetilde{\boldsymbol{W}}_{\boldsymbol{x}} - \eta \boldsymbol{W}_{\boldsymbol{x}}$ is not negative semidefinite for some $\boldsymbol{x}$, then it has at least one strictly positive eigenvalue $r > 0$ with corresponding eigenvector $\boldsymbol{v}$. Setting $\boldsymbol{y} = \boldsymbol{x} - a\boldsymbol{v}$ yields

$$(\boldsymbol{y} - \boldsymbol{x})^T (\widetilde{\boldsymbol{W}}_{\boldsymbol{x}} - \eta \boldsymbol{W}_{\boldsymbol{x}})(\boldsymbol{y} - \boldsymbol{x}) = a^2 r \boldsymbol{v}^T \boldsymbol{v} \to \infty$$

as $a \to \infty$. Hence η will always be such that $\widetilde{\boldsymbol{W}}_{\boldsymbol{x}} - \eta \boldsymbol{W}_{\boldsymbol{x}}$ is negative semidefinite for P almost every $\boldsymbol{x}$.

In this case

$$\sup_{\boldsymbol{y} \in \mathbb{R}^m} (\boldsymbol{y} - \boldsymbol{x})^T (\widetilde{\boldsymbol{W}}_{\boldsymbol{x}} - \eta \boldsymbol{W}_{\boldsymbol{x}})(\boldsymbol{y} - \boldsymbol{x}) = 0,$$

and we obtain

$$\begin{aligned} &\lim_{R \to \infty} D_{\rho,\tilde{\rho}}(R, D_\rho) \\ &\qquad = D_\rho \inf\{\eta \geq 0 : \widetilde{\boldsymbol{W}}_{\boldsymbol{x}} - \eta \boldsymbol{W}_{\boldsymbol{x}} \leq 0 \ P \text{ a.e.}\}, \end{aligned} \tag{7}$$

where $\widetilde{\boldsymbol{W}}_{\boldsymbol{x}} - \eta \boldsymbol{W}_{\boldsymbol{x}} \leq 0$ means that the matrix on the left hand side is negative semidefinite.

To illustrate that this bound can be fairly tight already for small R, we consider now a special case, for which $D_{\rho,\tilde{\rho}}(R, D_\rho)$ can be calculated analytically. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^2$,

$$\begin{aligned} \rho(\boldsymbol{x}, \boldsymbol{y}) &= (\boldsymbol{y} - \boldsymbol{x})^T \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\boldsymbol{y} - \boldsymbol{x}), \\ \tilde{\rho}(\boldsymbol{x}, \boldsymbol{y}) &= (\boldsymbol{y} - \boldsymbol{x})^T \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} (\boldsymbol{y} - \boldsymbol{x}), \end{aligned}$$

with $a \geq b > 0$, and let $X \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. The asymptotic expression (and upper bound) given by (7) is

$$\lim_{R \to \infty} D_{\rho,\tilde{\rho}}(R, D_\rho) = a D_\rho \tag{8}$$

and on the boundary

$$D_{\rho,\tilde{\rho}}(R_\rho(D_\rho), D_\rho) = \frac{1}{2}(a + b) D_\rho. \tag{9}$$

It can be shown that for $0 < D_\rho \leq 1$

$$\begin{aligned} &D_{\rho,\tilde{\rho}}(R_\rho(D_\rho) + r, D_\rho) \\ &\qquad \leq D_\rho \big((a + b) + \sqrt{1 - \exp(-2r)}(a - b)\big)/2, \end{aligned}$$

This function is plotted in Figure 3. As a quick check, we see that indeed

$$\begin{aligned} \lim_{r \to 0} D_{\rho,\tilde{\rho}}(R_\rho(D_\rho) + r, D_\rho) &= \frac{1}{2} D_\rho (a + b), \\ \lim_{r \to \infty} D_{\rho,\tilde{\rho}}(R_\rho(D_\rho) + r, D_\rho) &= a D_\rho, \end{aligned}$$

which are the values found in (8) and (9).

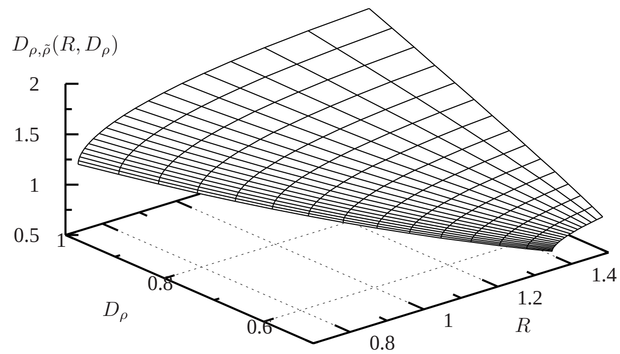

Fig. 3. $D_{\rho,\tilde{\rho}}(R, D_\rho)$ from Example 4 with $a = 2$ and $b = 0.5$.

For $0 < D_\rho \leq 1$, the ratio between the limiting expression as $r \to \infty$ and the value for finite r is independent of D_ρ and given by

$$\begin{aligned} &D_{\rho,\tilde{\rho}}(R_\rho(D_\rho) + r, D_\rho)/D_{\rho,\tilde{\rho}}(\infty, D_\rho) \\ &\qquad = \big((a + b) + \sqrt{1 - \exp(-2r)}(a - b)\big)/2a. \end{aligned}$$

This converges to one quickly as $r \to \infty$, as is shown in Figure 4. Hence in this case the limiting expression found in Theorem 5 is a fairly tight bound even for small values of r. $\diamondsuit$

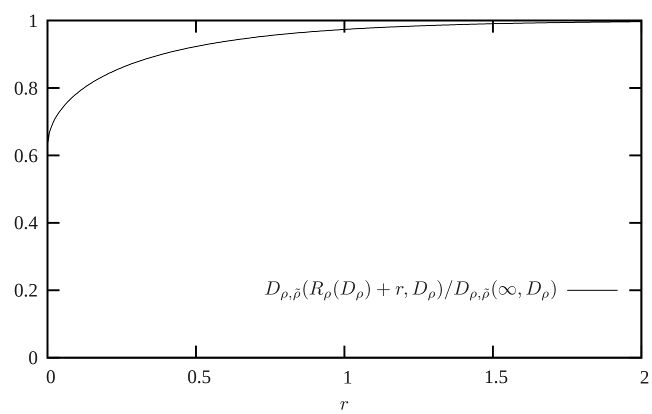

Fig. 4. $D_{\rho,\tilde{\rho}}(R_\rho(D_\rho)+r, D_\rho)/D_{\rho,\tilde{\rho}}(\infty, D_\rho)$ from Example 4 as a function of r with $a = 2$, $b = 0.5$, for all values $0 < D_\rho \leq 1$. Note that for an excess rate of $r = 0.5$, we are already at over 90% of the limiting value, at excess rate of $r = 1$, we are at over 97% of the limiting value.

C. Choosing a "Representative" of a Class of Distortion Measures

Let Γ and $\widetilde{\Gamma}$ denote classes of distortion measures. In this section, we consider the question of how a good "representative" $\rho \in \Gamma$ of $\widetilde{\Gamma}$ can be chosen (in a sense to be made precise).

For rate R, distortion measure ρ, and distortion level D_ρ, define the excess distortion

$$\Delta_\rho(R, D_\rho) \triangleq (D_\rho - D_\rho(R))^+.$$

Consider again the oracle producing source codes as mentioned in the introduction, but assume this time that when queried, we can also supply the oracle with a distortion measure $\rho \in \Gamma$. The oracle then produces a source code f_n such that

$$\frac{1}{n}\log|f_n(\mathcal{X}^n)| \leq R$$
$$\mathbb{E}\rho_n(X^n, f_n(X^n)) \leq D_\rho(R) + \Delta_\rho.$$

Knowing the set of all $\{\Delta_\rho\}_{\rho\in\Gamma}$, and given a $\widetilde{\Gamma}$, how should we choose $\rho \in \Gamma$ to query the oracle with such that f^n will "work well" for all $\tilde{\rho} \in \widetilde{\Gamma}$?

The operational significance of this question follows from the discussion in the introduction. The parameters $\{\Delta_\rho\}_{\rho\in\Gamma}$ allow to account for the difficulty of constructing a source code for distortion measure ρ (see also Example 5 below). Note, however, that there are several reasonable ways in which "work well" in the last paragraph can be defined. We will consider two such definitions in the following.

For rate R, define

$$D_{\Gamma,\widetilde{\Gamma}}(R) \triangleq \inf_{\rho\in\Gamma}\sup_{\tilde{\rho}\in\widetilde{\Gamma}} D_{\rho,\tilde{\rho}}(R, D_\rho(R) + \Delta_\rho),$$
$$\Delta_{\Gamma,\widetilde{\Gamma}}(R) \triangleq \inf_{\rho\in\Gamma}\sup_{\tilde{\rho}\in\widetilde{\Gamma}} \big(D_{\rho,\tilde{\rho}}(R, D_\rho(R) + \Delta_\rho) - D_{\tilde{\rho}}(R)\big).$$

The next example illustrates why introducing $\{\Delta_\rho\}_{\rho\in\Gamma}$ is necessary.

Example 5. Fix a distortion measure $\tilde{\rho}$ and let $\Gamma \triangleq \{a\rho\}_{a\geq 1}$, for some distortion measure ρ. All distortion measures in Γ are equivalent (in the sense that constructing source codes for ρ is as difficult as constructing source codes for any $a\rho$). So we should have that all $a\rho$ represent $\tilde{\rho}$ equally well (in the sense that for appropriately chosen $D_{a\rho}$, $D_{a\rho,\tilde{\rho}}(R, D_{a\rho})$ is the same for all $a \geq 1$). As we will see in a moment, this imposes the introduction of the quantity $\{\Delta_{a\rho}\}_{a\geq 1}$.

For any *fixed* D_ρ, we have

$$D_{a\rho,\tilde{\rho}}(R, D_\rho) = D_{\rho,\tilde{\rho}}(R, D_\rho/a)$$

which goes either to 0 (if $R \geq R_\rho(0)$) or to $-\infty$ as $a \to \infty$. This shows that we should always look at source codes constructed with distortion level relative to $D_{a\rho}(R)$

Assume then we try to minimize $D_{a\rho,\tilde{\rho}}(R, D_{a\rho}(R) + \Delta)$ for some *fixed* $\Delta > 0$. We have

$$D_{a\rho,\tilde{\rho}}(R, D_{a\rho}(R) + \Delta) = D_{\rho,\tilde{\rho}}(R, D_\rho(R) + \Delta/a).$$

Thus, again, the minimum is achieved as $a \to \infty$, irrespective of the choice of $\tilde{\rho}$.

This shows, that we should not choose $\Delta_{a\rho}$ as a constant. The natural choice in this example is $\Delta_{a\rho} = a\Delta$, for which

$$D_{a\rho,\tilde{\rho}}(R, D_{a\rho}(R) + \Delta_{a\rho}) = D_{\rho,\tilde{\rho}}(R, D_\rho(R) + \Delta).$$

$\Diamond$

The following two corollaries of Theorem 1 and 2, respectively, establish the operational meaning of $D_{\Gamma,\widetilde{\Gamma}}(R)$ and $\Delta_{\Gamma,\widetilde{\Gamma}}(R)$.

Corollary 6. *Let $\Gamma, \widetilde{\Gamma}$ be classes of distortion measures such that for all $\rho \in \Gamma$ there exists a $y_0 = y_0(\rho) \in \mathcal{Y}$ satisfying $\mathbb{E}_P\rho(X, y_0) < \infty$. For every $\rho \in \Gamma$ and $D_{\widetilde{\Gamma}}, \Delta_{\widetilde{\Gamma}} < \infty$ such that*

$$0 \leq D_{\widetilde{\Gamma}} < D_{\Gamma,\widetilde{\Gamma}}(R-),$$
$$0 \leq \Delta_{\widetilde{\Gamma}} < \Delta_{\Gamma,\widetilde{\Gamma}}(R-),$$

a) *there exists $\tilde{\rho} \in \widetilde{\Gamma}$ and sequences of source codes $\{f_n\}_{n\geq 1}$ such that*

$$\lim_{n\to\infty}\frac{1}{n}\log|f_n(\mathcal{X}^n)| \leq R,$$
$$\limsup_{n\to\infty}\mathbb{E}\rho_n(X^n, f_n(X^n)) \leq D_\rho(R) + \Delta_\rho,$$
$$\liminf_{n\to\infty}\mathbb{E}\tilde{\rho}_n(X^n, f_n(X^n)) \geq D_{\widetilde{\Gamma}}.$$

b) *there exists $\tilde{\rho} \in \widetilde{\Gamma}$ and sequences of source codes $\{f_n\}_{n\geq 1}$ such that*

$$\lim_{n\to\infty}\frac{1}{n}\log|f_n(\mathcal{X}^n)| \leq R,$$
$$\limsup_{n\to\infty}\mathbb{E}\rho_n(X^n, f_n(X^n)) \leq D_\rho(R) + \Delta_\rho,$$
$$\liminf_{n\to\infty}\big(\mathbb{E}\tilde{\rho}_n(X^n, f_n(X^n)) - D_{\tilde{\rho}}(R)\big) \geq \Delta_{\widetilde{\Gamma}}.$$

Corollary 7. a) *For every $\delta > 0$ there exists $\rho \in \Gamma$ such that if $f_n : \mathcal{X}^n \to \mathcal{Y}^n$ satisfies*

$$\frac{1}{n}\log|f_n(\mathcal{X}^n)| = R,$$
$$\mathbb{E}\rho_n(X^n, f_n(X^n)) \leq D_\rho(R) + \Delta_\rho,$$
$$\sup_{\tilde{\rho}\in\widetilde{\Gamma}}\mathbb{E}\tilde{\rho}_n(X^n, f_n(X^n)) \geq D_{\widetilde{\Gamma}},$$

then $D_{\widetilde{\Gamma}} \leq D_{\Gamma,\widetilde{\Gamma}}(R+) + \delta$.

b) *For every $\delta > 0$ there exists $\rho \in \Gamma$ such that if $\tilde{f}_n : \mathcal{X}^n \to \mathcal{Y}^n$ satisfies*

$$\frac{1}{n}\log|\tilde{f}_n(\mathcal{X}^n)| = R,$$
$$\mathbb{E}\rho_n(X^n, \tilde{f}_n(X^n)) \leq D_\rho(R) + \Delta_\rho,$$
$$\sup_{\tilde{\rho}\in\Gamma}\big(\mathbb{E}\tilde{\rho}_n(X^n, \tilde{f}_n(X^n)) - D_{\tilde{\rho}}(R)\big) \geq \Delta_{\widetilde{\Gamma}},$$

then $\Delta_{\widetilde{\Gamma}} \leq \Delta_{\Gamma,\widetilde{\Gamma}}(R+) + \delta$.

Example 6. Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^m$, $\widetilde{\Gamma} = \{\tilde{\rho}\}$, and

$$\Gamma \triangleq \big\{\rho(\boldsymbol{x}, \boldsymbol{y}) = w_{\boldsymbol{x}}\,\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 : w \in \mathcal{W} \subset \mathcal{X} \to \mathbb{R}_+\big\},$$

Let $P \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be such that $\mathbb{E}_P w_{\boldsymbol{X}}\,\|\boldsymbol{X}\|_2^2 < \infty$ for all $w \in \mathcal{W}$.

In [13], the authors show how vector quantizers can be relatively easily constructed for distortion measures in the class Γ defined here. Given a more sophisticated distortion measure

$\tilde{\rho}$, it is thus of interest to find the "closest" $\rho \in \Gamma$ to $\tilde{\rho}$. In other words, we want to find $D_{\Gamma,\tilde{\Gamma}}(R)$ and a $\rho \in \Gamma$ such that

$$D_{\rho,\tilde{\rho}}(R, D_\rho(R) + \Delta_\rho) \leq D_{\Gamma,\tilde{\Gamma}}(R) + \delta,$$

for some $\delta > 0$.

Computing $D_{\Gamma,\widetilde{\Gamma}}(R)$ could be done numerically; to obtain some insight we will instead minimize $D_{\rho,\tilde{\rho}}(\infty, D_\rho(R)+\Delta_\rho)$. As we have seen this last quantity is usually quite close to $D_{\rho,\tilde{\rho}}(R, D_\rho(R) + \Delta_\rho)$. To be specific, let $\tilde{\rho}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{x})^T \widetilde{\boldsymbol{W}}_{\boldsymbol{x}}(\boldsymbol{y}-\boldsymbol{x})$ for $\widetilde{\boldsymbol{W}}_{\boldsymbol{x}}$ positive definite P almost everywhere. Then from Example 4

$$\begin{aligned} &D_{\rho,\tilde{\rho}}(\infty, D_\rho(R) + \Delta_\rho) \\ &\qquad = (D_\rho(R) + \Delta_\rho) \min\left\{\eta : \widetilde{\boldsymbol{W}}_{\boldsymbol{x}} - \eta w_{\boldsymbol{x}} \boldsymbol{I} \leq 0 \ P \text{ a.e}\right\} \\ &\qquad = (D_\rho(R) + \Delta_\rho) \operatorname*{ess\,sup}_{x \in \mathcal{X}} \lambda_1(\widetilde{\boldsymbol{W}}_{\boldsymbol{x}})/w_{\boldsymbol{x}}, \end{aligned}$$

where $\lambda_1(\widetilde{\boldsymbol{W}}_{\boldsymbol{x}})$ is the largest eigenvalue of $\widetilde{\boldsymbol{W}}_{\boldsymbol{x}}$, and where the essential supremum is with respect to P. $\diamondsuit$

In this last example, we have taken a sophisticated distortion measure $\tilde{\rho}$ and found a good tractable approximation in Γ for it. This approach poses the following question. Even if $\tilde{\rho}$ is a very good model for (say) the human visual system, it will certainly be different from it. In this situation, it is not clear if minimizing $D_{\rho,\tilde{\rho}}(R, D_\rho(R)+\Delta_\rho)$ is meaningful. Indeed, if ρ^* is the distortion measure implemented by the human visual system, we should really be minimizing $D_{\rho,\rho^*}(R, D_\rho(R) + \Delta_\rho)$ instead. The next theorem provides conditions under which $D_{\rho,\tilde{\rho}}(R, D_\rho(R)+\Delta_\rho)$ and $D_{\rho,\rho^*}(R, D_\rho(R)+\Delta_\rho)$ are close and hence the approach of Example 6 is reasonable.

Proposition 8. *Let ρ_1, ρ_2, ρ_3 be continuous distortion measures. Then*

$$\begin{aligned} &D_{\rho_1,\rho_3}(R, D_{\rho_1}) \\ &\qquad \leq D_{\rho_1,\rho_2}(R, D_{\rho_1}) + \mathbb{E}_P(\sup_{y\in\mathcal{Y}} \rho_3(X, y) - \rho_2(X, y)) \end{aligned}$$

and

$$\begin{aligned} &D_{\rho_1,\rho_3}(R, D_{\rho_1}) \\ &\qquad \geq D_{\rho_1,\rho_2}(R, D_{\rho_1}) - \mathbb{E}_P \sup_{y\in\mathcal{Y}} |\rho_3(X, y) - \rho_2(X, y)|. \end{aligned}$$

Example 7. Setting $\rho_1 = \rho_2$, Proposition 8 shows that

$$\left|D_{\rho_2,\rho_3}(R, D_{\rho_1}) - D_{\rho_2}(R)\right| \leq \mathbb{E}_P \sup_{y\in\mathcal{Y}} |\rho_3(X, y) - \rho_2(X, y)|.$$

Thus if

$$\mathbb{E}_P \sup_{y\in\mathcal{Y}} |\rho_3(X, y) - \rho_2(X, y)|$$

is small, then the distortion measures ρ_2 and ρ_3 are almost equivalent (from the point of source coding).

Moreover, if ρ_3 is the actual distortion measure (implemented, e.g., by the human visual system), and ρ_2 is a sophisticated model for it (e.g. $\rho_2(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y}-\boldsymbol{x})^T \widetilde{\boldsymbol{W}}_{\boldsymbol{x}}(\boldsymbol{y}-\boldsymbol{x})$ as in Example 6), then small

$$\mathbb{E}_P \sup_{y\in\mathcal{Y}} |\rho_3(X, y) - \rho_2(X, y)|$$

guarantees that minimizing $D_{\rho_1,\rho_2}(R, D_{\rho_1} + \Delta_{\rho_1})$ over all $\rho_1 \in \Gamma$ (as is done in Example 6) is essentially equivalent to minimizing $D_{\rho_1,\rho_3}(R, D_{\rho_1} + \Delta_{\rho_1})$.

Hence, when constructing a model ρ_2 for the human visual system (implementing ρ_3) to be used for data compression applications, it is reasonable to choose the model parameters such that

$$\mathbb{E}_P \sup_{y\in\mathcal{Y}} |\rho_3(X, y) - \rho_2(X, y)|$$

is minimized. $\diamondsuit$

III. Conclusion

In this paper, we investigated the problem of source coding with mismatched distortion measures. We derived a single-letter characterization $D_{\rho,\tilde{\rho}}(R, D_\rho)$ of the best distortion level with respect to $\tilde{\rho}$ that can be guaranteed for any source code of rate R designed for distortion level D_ρ with respect to ρ. We then looked at properties of $D_{\rho,\tilde{\rho}}(R, D_\rho)$, characterizing its behavior for $R > R_\rho(D_\rho)$ and on the boundary. We also found an asymptotic expression (and upper bound) as $R \to \infty$ for this quantity, that seems to be fairly tight also for small R. This asymptotic expression gives considerable insight into the behavior of $D_{\rho,\tilde{\rho}}(R, D_\rho)$, which we illustrated with several examples. We finally considered the problem of choosing a representative $\rho \in \Gamma$ of $\tilde{\rho}$.

References

[1] J. Stjernvall. Dominance—a relation between distortion measures. *IEEE Transactions on Information Theory*, 29(6):798–807, November 1983.

[2] A. Lapidoth. On the role of mismatch in rate distortion theory. *IEEE Transactions on Information Theory*, 43(1):38–47, January 1997.

[3] H. Yamamoto. A rate-distortion problem for a communication system with a secondary decoder to be hindered. *IEEE Transactions on Information Theory*, 33(4):835–842, July 1988.

[4] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[5] A. Dembo and T. Weissman. The minimax distortion redundancy in noisy source coding. *IEEE Transactions on Information Theory*, 49(11):3020–3030, November 2003.

[6] N. Jayant, J. Johnston, and R. Safranek. Signal compression based on models of human perception. *Proceedings of the IEEE*, 81(10):1385–1422, October 1993.

[7] N. Pappas, R. J. Safranek, and J. Chen. Perceptual criteria for image quality evaluation. In A. C. Bovik, editor, *Handbook of Image and Video Processing*, pages 939–959. Academic Press, 2005.

[8] M. P. Eckert and A. P. Bradley. Perceptual quality metrics applied to still image compression. *Signal Processing*, 70:177–200, November 1998.

[9] H. de Ridder. Minkowski-metrics as a combination rule for digital-image-coding impairments. In *SPIE Conference*, volume 1666, pages 16–26, August 1992.

[10] J. Lubin. The use of psychophysical data and models in the analysis of display system performance. In B. Watson, editor, *Digital Images and Human Vision*, pages 163–178. MIT Press, 1993.

[11] A. B. Watson. DCT quantization matrices visually optimized for individual images. In *SPIE Conference*, volume 1913, pages 202–216, February 1993.

[12] J. L. Mannos and D. J. Sakrison. The effects of a visual fidelity criterion on the encoding of images. *IEEE Transactions on Information Theory*, 20(4):525–536, July 1974.

[13] R. M. Gray, P. C. Cosman, and K. L. Oehler. Incorporating visual factors into vector quantizers for image compression. In B. Watson, editor, *Digital Images and Human Vision*, pages 35–52. MIT Press, 1993.