Sampling Distortion Measures

Urs Niesen, Devavrat Shah, and Gregory Wornell

***Abstract*— This paper is motivated by the following problem. Consider an image we want to compress. The appropriate distortion measure with respect to which we compress the image is specified by the human visual system. This distortion measure can only be experimentally determined and our knowledge about it will hence be only partial.**

This leads us to consider the general problem of lossy source coding with partial knowledge about the distortion measure. More precisely, the distortion measure is only known at a number of sampling points. We describe several measures for the loss we incur through the lack of full knowledge of the true distortion measure, each with a different operational meaning. We give an asymptotically tight characterization of this loss in terms of three key parameters: The number of sampling points, the dimensionality of the source, and the smoothness assumptions on the distortion measure.

I. Introduction

Background and Motivation. In most applications of data compression involving a human observer, the distortion measure, with respect to which compression is to be performed, is not known explicitly. Rather, it is the human perception, which defines the relevant distortion measure, and we can only gain partial knowledge of it through experiments. We are hence facing a problem of lossy source coding with partial knowledge about the distortion measure.

Consider first the classic problem of describing a sequence of independent identically distributed random variables $\{X_i\}_{i\geq 1}$ within distortion D, using the (single-letter) distortion measure ρ. The smallest number of bits needed for such a description is given by the rate distortion function (see, e.g., [1])

$$R_\rho(D) \triangleq \inf_{Q:\mathbb{E}(\rho(X,Y))\leq D} I(P,Q),$$

where $P \in \mathcal{P}(\mathcal{X})$ is the marginal distribution of the X_i, $I(P,Q)$ denotes mutual information, and where the minimization is over all conditional distributions $Q \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$.

For a given distortion measure ρ, $R_\rho(D)$ can be computed either analytically or using numerical methods. As we have argued before, ρ might be only partially known. For example, if $\{X_i\}_{i\geq 1}$ represent images that we want to compress then the appropriate ρ is specified by the human visual system. In this case ρ can only be experimentally determined and will never be fully known. In other words, instead of full knowledge of ρ, we only know that ρ belongs to some class of distortion measures.

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

Our goal is to quantify the loss incurred in rate due to the partial knowledge of ρ in such situations. Specifically, we model the human perception by assuming that ρ belongs to a certain class of well-behaved functions, denoted by Γ. Further, in our setup we expect to learn ρ by interaction with humans. Let the class of functions that are in accordance with the first n interactions be Γ_n, with $\Gamma_0 = \Gamma$ and $\Gamma_n \subset \Gamma_{n-1}$ for all $n \geq 1$. If we want to guarantee that the description of $\{X_i\}_{i\geq 1}$ is within a distortion D with respect to (the unknown) $\rho \in \Gamma_n$, we will have to design a source code that operates well simultaneously for all distortion measures in Γ_n. The smallest number of bits per symbol needed for such a description is given by [2], [3]

$$R_{\Gamma_n}(D) = \inf_{Q:\mathbb{E}(\rho(X,Y))\leq D \forall \rho\in\Gamma_n} I(P,Q). \tag{1}$$

Equation (1) holds unconditionally if $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets; it holds under some technical condition for infinite alphabets (see [3] for the details). Given this, there are several ways to measure the loss in terms of rate we incur by the partial knowledge of ρ. As the first criterion, we look at the distance between $R_\Gamma(D)$ and $R_\rho(D)$ for $\rho \in \Gamma$. As a second criterion, we measure the distances between $R_\rho(D)$ for different $\rho \in \Gamma_n$. Our goal is to characterize the behavior of the loss in rate as a function of the number of interactions (or samples) n of ρ we have at our disposition. We obtain tight characterization of the asymptotic behavior of the loss.

We note that the problem of lossy source coding with respect to a class of distortion measures has been investigated in [2] and [3] as mentioned above. A variation of this setup, in which the true distortion measure is unknown at the encoder but known at the decoder, is considered in [4] and [5]. In all of these papers, the set of distortion measures with respect to which we seek universality is fixed. This differs from the problem considered in this paper as here a good sampling procedure is required to be designed; as a consequence the set of possible distortion measures decreases in the number of samples n, and the emphasis here is on the asymptotic behavior in n.

Problem Formulation. Let the source alphabet $\mathcal{X}$ be $[0,1]^m$ and the reconstruction alphabet $\mathcal{Y}$ be some finite set (i.e., $|\mathcal{Y}| < \infty$). The distribution P of the source $\{X_i\}_{i\geq 1}$ is assumed to admit a density p such that $1/M_p \leq p(x) \leq M_p, \forall x \in \mathcal{X}$, for some $0 < M_p < \infty$. Define for a Lebesgue-measurable function f on $\mathcal{X}$ the $L_q(\mathcal{X})$ norm as

$$\|f\|_{L_q(\mathcal{X})} \triangleq \begin{cases} \left(\int_{x\in\mathcal{X}} |f(x)|^q dx\right)^{1/q} & \text{if } 1 \leq q < \infty, \\ \operatorname{ess\,sup}_{x\in\mathcal{X}} |f(x)| & \text{if } q = \infty. \end{cases}$$

$L_q(\mathcal{X})$ is the space of all measurable functions f on $\mathcal{X}$ such that $\|f\|_{L_q(\mathcal{X})}$ is finite. For $\lambda \in \mathbb{Z}_+^m$ let $|\lambda| = \sum_{i=1}^m \lambda_i$ and $\mathbb{Z}_{+j}^m \triangleq \{\lambda \in \mathbb{Z}_+^m : |\lambda| = j\}$. Denote by $\partial^\lambda \triangleq \partial^\lambda / \partial^{\lambda_1} x_1 \cdots \partial^{\lambda_m} x_m$ the (weak) derivative of order $|\lambda|$. The Sobolev space $W_q^l(\mathcal{X})$ is defined as the space of functions $f \in L_q(\mathcal{X})$ such that $\partial^\lambda f$ is defined for all $\lambda \in \mathbb{Z}_+^m$ with $|\lambda| \leq l$, and such that the norm

$$\|f\|_{W_q^l(\mathcal{X})} \triangleq \sum_{j=0}^{l} \sum_{\lambda \in \mathbb{Z}_{+j}^m} \|\partial^\lambda f\|_{L_q(\mathcal{X})}$$

is finite. In this paper, we consider the class of distortion measures

$$\Gamma \triangleq \Gamma^l \triangleq \{\rho : \|\rho(\cdot, y)\|_{W_q^l(\mathcal{X})} \leq K, \\ \|\rho(\cdot, y)\|_{L_\infty(\mathcal{X})} \leq B, \forall y \in \mathcal{Y}\},$$

for some $B, K < \infty$, and with $1 \leq q \leq \infty$, $l \in \mathbb{N}$ and $lq > m$. This last condition ensures that $W_q^l(\mathcal{X})$ is a subspace of the space of continuous functions $C(\mathcal{X})$ (by the Sobolev Embedding Theorem [6, Chapter 3.1]). The above models the fact that there is *inherent smoothness* in the human perception. Note that we assume that all functions in Γ are bounded, i.e., for all $\rho \in \Gamma$, $\rho : \mathcal{X} \times \mathcal{Y} \to [0, B]$, for some $B < \infty$.

Under these assumptions the sufficient condition for (1) to hold is satisfied. By $D_\rho(r)$ and $D_\Gamma(r)$, we denote the distortion rate functions corresponding to $R_\rho(D)$ and $R_\Gamma(D)$, respectively.

Next, we introduce two measures for the width[1] of Γ. Let

$$\Delta(\rho_1, \rho_2, r) \triangleq |D_{\rho_1}(r) - D_{\rho_2}(r)|$$

for $\rho_1, \rho_2 \in \Gamma$, $r \geq 0$. As every $\rho \in \Gamma$ is bounded by B, we have $\Delta(\rho_1, \rho_2, r) \leq B$. With slight abuse of notation, define $\Delta(\Gamma, \rho, r)$ as

$$\Delta(\Gamma, \rho, r) \triangleq D_\Gamma(r) - D_\rho(r)$$

for $\rho \in \Gamma$, $r \geq 0$. Figure 1 illustrates these definitions. Call

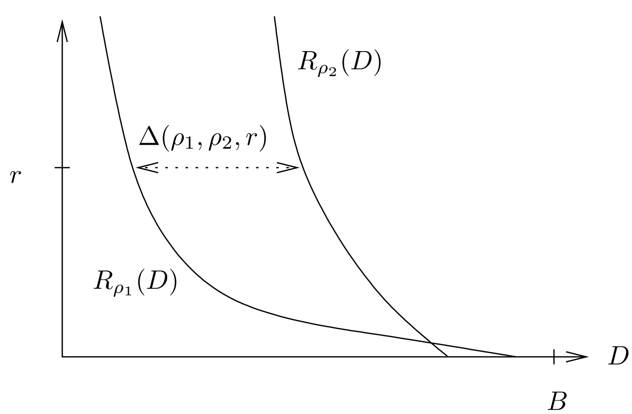

Fig. 1. Distance $\Delta(\rho_1, \rho_2, r)$ between two distortion measures.

[1]The terminology is chosen following the concept of n-widths in approximation theory. For more details about n-widths and the related notion of ϵ-complexity, see for example [7] and [8].

$\mathbf{S}_n = \mathcal{X}^n \times \mathcal{Y}^n$ the set of all possible n sampling points. Given $\mathbf{s}_n = \{(x_i, y_i)\}_{1 \leq i \leq n} \in \mathbf{S}_n$, let $\phi_n : \Gamma \times \mathbf{S}_n \to \mathbb{R}^n$ denote the mapping of $(\rho, \mathbf{s}_n)$ to $\{\rho(x_i, y_i)\}_{1 \leq i \leq n}$. For a $\rho \in \Gamma$, let

$$\Gamma(\rho, \mathbf{s}_n) \triangleq \{\tilde{\rho} \in \Gamma : \phi_n(\tilde{\rho}, \mathbf{s}_n) = \phi_n(\rho, \mathbf{s}_n)\},$$

i.e., $\Gamma(\rho, \mathbf{s}_n)$ is the set of all distortion measures in Γ that coincide with the true distortion measures ρ at all sampling points $\mathbf{s}_n$. Define

$$\varepsilon_1(\Gamma, n) \triangleq \inf_{\mathbf{s}_n \in \mathbf{S}_n} \sup_{\substack{\rho \in \Gamma \\ r \geq 0}} \Delta(\Gamma(\rho, \mathbf{s}_n), \rho, r).$$

Intuitively, $\varepsilon_1(\Gamma, n)$ is the price we pay in terms of distortion, by designing a source code which works simultaneously for all $\rho \in \Gamma$. We define a second measure for the width of Γ

$$\varepsilon_2(\Gamma, n) \triangleq \inf_{\mathbf{s}_n \in \mathbf{S}_n} \sup_{\substack{\rho \in \Gamma \\ \tilde{\rho} \in \Gamma(\rho, \mathbf{s}_n) \\ r \geq 0}} \Delta(\rho, \tilde{\rho}, r).$$

In words, $\varepsilon_2(\Gamma, n)$ measures the price we pay by calculating $R_{\tilde{\rho}}(D)$ with respect to the worst $\tilde{\rho} \in \Gamma$ based on partial knowledge of the actual distortion measure ρ.

From the definitions of $\varepsilon_1(\cdot, \cdot)$ and $\varepsilon_2(\cdot, \cdot)$, it is clear that $\varepsilon_2(\cdot, \cdot) \leq \varepsilon_1(\cdot, \cdot)$. However, as we shall see the asymptotic behavior of both of these losses is exactly the same. The notion of $\varepsilon_1(\cdot, \cdot)$ is operationally more useful. We introduce the definition of $\varepsilon_2(\cdot, \cdot)$ as it will be useful in our proof technique along with the above inequality.

Main Results. We obtain asymptotically tight characterization of the errors $\varepsilon_1(\cdot, \cdot)$ and $\varepsilon_2(\cdot, \cdot)$ as well as the algorithm that achieves this optimal performance.

Recall the definition of Γ^l as

$$\Gamma^l \triangleq \{\rho : \|\rho(\cdot, y)\|_{W_q^l(\mathcal{X})} \leq K, \\ \|\rho(\cdot, y)\|_{L_\infty(\mathcal{X})} \leq B, \forall y \in \mathcal{Y}\},$$

for some $K, B < \infty$, and with $1 \leq q \leq \infty$, $l \in \mathbb{N}$ and $lq > m$.

Theorem 1. *For $i \in \{1, 2\}$ as $n \to \infty$*

$$\varepsilon_i(\Gamma^l, n) = \Theta(n^{-l/m}).$$

Theorem 1 characterizes the asymptotic behavior of the loss $\varepsilon_i(\Gamma^l, n)$ in terms of the dimensionality of the source m, the number of samples n of the distortion measure, and the smoothness assumptions about the class of distortion measures as captured by l. Interestingly, the asymptotic loss is the same for both measures $\varepsilon_2(\Gamma^l, n)$ and $\varepsilon_1(\Gamma^l, n)$. Thus, at least asymptotically, there is no additional penalty to be paid by designing a source code which works simultaneously for all possible distortion measures which agree with the true distortion measure at the sampling points.

By definition, we have assumed that the sampling scheme is non-adaptive, i.e., the choice of the n-th sampling point does not depend on the evaluation of ρ at the first $n-1$ sampling points. It is natural to ask whether an adaptive sampling strategy improves the performance. More formally, let $\mathbf{S}_n^A(\rho)$ be the set of all sampling points of the form

$$\big\{(x_1,y_1),\big(x_2(\rho(x_1,y_1)),y_2(\rho(x_1,y_1))\big),\dots\big\}.$$

Define the corresponding width measures

$$\varepsilon_1^A(\Gamma,n)\triangleq\sup_{\rho\in\Gamma}\inf_{\mathbf{s}_n\in\mathbf{S}_n^A(\rho)}\sup_{r\geq 0}\Delta(\Gamma(\rho,\mathbf{s}_n),\rho,r),$$
$$\varepsilon_2^A(\Gamma,n)\triangleq\sup_{\rho\in\Gamma}\inf_{\mathbf{s}_n\in\mathbf{S}_n^A(\rho)}\sup_{\substack{\tilde\rho\in\Gamma(\rho,\mathbf{s}_n)\\ r\geq 0}}\Delta(\rho,\tilde\rho,r).$$

The following will be an immediate corollary of Theorem 1.

Corollary 2. *For $i\in\{1,2\}$ as $n\to\infty$*

$$\varepsilon_i^A(\Gamma^l,n)=\Theta(n^{-l/m}).$$

Organization. The rest of the paper is organized as follows. In Sections II and III, we present the proof of Theorem 1. Specifically, Section II provides an upper bound on $\varepsilon_1(\cdot,\cdot),\varepsilon_2(\cdot,\cdot)$, while Section III provides a matching lower bound. In Section IV, we present the proof of Corollary 2. Finally, in Section V we present our conclusions.

II. Proof of Theorem 1: Upper Bound

There are several difficulties evaluating $\varepsilon_i(\Gamma,n)$, $i\in\{1,2\}$. First, we have to solve a minimax problem (minimizing over all sampling strategies, maximizing over all distortion measures and rates). Second, $\Delta(\rho_1,\rho_2,r)$, the function over which we optimize, is not given explicitly, but rather as a solution of an optimization problem itself. Moreover, for most input distributions P and distortion measures ρ, the distortion rate function $D_\rho(r)$ cannot be calculated analytically. Instead, we prove a "continuity" property of the distortion rate function in the space of Γ. This, along with a sampling-reconstruction algorithm, leads to the desired result.

Lemma 3. *For some positive constant c_1, $n\geq n_1$, and $i\in\{1,2\}$*

$$\varepsilon_i(\Gamma^l,n)\leq c_1 n^{-l/m}.$$

Proof: From the definition, it is easy to see that $\varepsilon_2(\Gamma^l,n)\leq\varepsilon_1(\Gamma^l,n)$. Hence, it suffices to prove the result for $\varepsilon_1(\Gamma^l,n)$.

To this end, assume there exists a $\mathbf{s}_n\in\mathbf{S}_n$ and a $\delta\geq 0$ such that for all $Q\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, $\rho\in\Gamma^l$, $\tilde\rho\in\Gamma^l(\rho,\mathbf{s}_n)$ we have $|\mathbb{E}(\rho(X,Y))-\mathbb{E}(\tilde\rho(X,Y))|\leq\delta$. Then for all $D\geq 0$

$$\{Q\in\mathcal{P}(\mathcal{Y}|\mathcal{X}):\mathbb{E}(\rho(X,Y))\leq D-\delta\}$$
$$\subset\{Q\in\mathcal{P}(\mathcal{Y}|\mathcal{X}):\mathbb{E}(\tilde\rho(X,Y))\leq D\forall\tilde\rho\in\Gamma^l(\rho,\mathbf{s}_n)\}.$$

Hence for all $\rho\in\Gamma^l$ and all $D\geq 0$

$$\begin{aligned}R_{\Gamma^l(\rho,\mathbf{s}_n)}(D)&=\inf_{Q:\mathbb{E}(\tilde\rho(X,Y))\leq D\forall\tilde\rho\in\Gamma^l(\rho,\mathbf{s}_n)}I(P,Q)\\&\leq\inf_{Q:\mathbb{E}(\rho(X,Y))\leq D-\delta}I(P,Q)\\&=R_\rho(D-\delta).\end{aligned}$$

From this, we get

$$\varepsilon_1(\Gamma^l,n)\leq\sup_{\rho\in\Gamma^l,r\geq 0}\Delta(\Gamma^l(\rho,\mathbf{s}_n),\rho,r)\leq\delta.$$

Hence it suffices to show that such a $\mathbf{s}_n$ and δ exist and to characterize the dependency on n of the latter.

Note that, for any $\mathbf{s}_n\in\mathbf{S}_n$, $Q\in\mathcal{P}(\mathcal{Y}|\mathcal{X})$, $\rho\in\Gamma^l$ and $\tilde\rho\in\Gamma^l(\rho,\mathbf{s}_n)$ we have

$$\begin{aligned}&|\mathbb{E}(\rho(X,Y))-\mathbb{E}(\tilde\rho(X,Y))|\\&\quad\leq\sum_{y\in\mathcal{Y}}\int_{x\in\mathcal{X}}p(x)Q(y|x)|\rho(x,y)-\tilde\rho(x,y)|dx\\&\quad\leq\sum_{y\in\mathcal{Y}}\int_{x\in\mathcal{X}}M_p|\rho(x,y)-\tilde\rho(x,y)|dx\\&\quad\leq|\mathcal{Y}|M_p\big\|\|\rho-\tilde\rho\|_{L_1(\mathcal{X})}\big\|_{L_\infty(\mathcal{Y})}.\end{aligned}\tag{2}$$

Theorem 4.2 in [9] (see also [10, Proposition 5.2, Theorem 6.1]) asserts that for $n\geq n_1$ there exists $\mathbf{s}_n^*\in\mathbf{S}_n$ and $\rho^*\in\Gamma^l(\rho,\mathbf{s}_n^*)$ such that

$$\|\rho^*-\tilde\rho\|_{L_1(\mathcal{X})}\leq\tilde c_1 n^{-l/m}$$

for all $\tilde\rho\in\Gamma^l(\rho,\mathbf{s}_n^*)$ and some positive constant $\tilde c_1$. As $\rho\in\Gamma^l(\rho,\mathbf{s}_n^*)$, we get from this

$$\|\rho-\tilde\rho\|_{L_1(\mathcal{X})}\leq\|\rho-\rho^*\|_{L_1(\mathcal{X})}+\|\rho^*-\tilde\rho\|_{L_1(\mathcal{X})}\leq 2\tilde c_1 n^{-l/m}.$$

With this, we can continue (2) as

$$\begin{aligned}|\mathbb{E}(\rho(X,Y))-\mathbb{E}(\tilde\rho(X,Y))|&\leq 2|\mathcal{Y}|M_p\tilde c_1 n^{-l/m}\\&\triangleq c_1 n^{-l/m}\triangleq\delta,\end{aligned}$$

for $n\geq n_1$. This completes the proof of Lemma 3. ■

The proof of Lemma 3 uses as a central tool an algorithm for the reconstruction of a function from its samples as given in [9] and [10]. This algorithm divides the unit cube $\mathcal{X}$ into smaller cubes of size 2^{-k}. Within each such subcube a number of sampling points are chosen to allow unique interpolation by a polynomial of degree l. k is chosen such that the total number of sampling points is at most n. The resulting reconstruction (called ρ^* in the proof of Lemma 3) is thus a piecewise polynomial.

III. Proof of Theorem 1: Lower Bound

Lemma 4 below provides a lower bound to the loss $\varepsilon_i(\Gamma^l,n), i\in\{1,2\}$ with the same asymptotic behavior as the upper bound in Lemma 3. This shows that the reconstruction ρ^* (as defined in the proof of Lemma 3) of the unknown distortion measure ρ is asymptotically optimal.

Lemma 4. *For some positive constant c_2 and $i\in\{1,2\}$*

$$\varepsilon_i(\Gamma^l,n)\geq c_2 n^{-l/m}.$$

Proof: Since $\varepsilon_1(\Gamma^l,n)\geq\varepsilon_2(\Gamma^l,n)$ it suffices to prove the result for $\varepsilon_2(\Gamma^l,n)$. To this end consider any set of sampling points $\mathbf{s}_n\in\mathbf{S}_n$. Theorem 4.3 in [9] (see also [10, Theorem 6.1]) asserts that for a given $\mathbf{s}_n=\{(x_i,y_i)\}_{1\leq i\leq n}$, there exists a function f satisfying $\|f\|_{W^l_\infty(\mathcal{X})}\leq 1$ and such that $f(x)\geq 0$, $f(x_i)=0$ for all $i\in\{i,\dots,n\}$, and

$$\|f\|_{L_1(\mathcal{X})}\geq\tilde c_2 n^{-l/m}$$

for some constant $\tilde{c}_2$. Let $M \triangleq \min\{B, K/2\}$, and define two functions ρ^* and ρ:

$$\rho^*(x,y) \triangleq M(1 - f(x)), \quad \text{for all } x \in \mathcal{X}, y \in \mathcal{Y},$$
$$\rho(x,y) \triangleq M, \quad \text{for all } x \in \mathcal{X}, y \in \mathcal{Y}.$$

Note that both ρ and ρ^* are elements of Γ^l and that $\rho^*(x_i, y_i) = \rho(x_i, y_i)$ for all $i \in \{1, \ldots, n\}$. We have $D_\rho(r) = M$ and

$$D_{\rho^*}(r) = M\Big(1 - \int_{x \in \mathcal{X}} p(x)|f(x)|dx\Big) \triangleq D^*$$

for all $r \geq 0$. D^* can be upper bounded as

$$\begin{aligned} D^* &\leq M\Big(1 - \frac{1}{M_p}\|f(x)\|_{L_1(\mathcal{X})}\Big) \\ &\leq M\Big(1 - \frac{1}{M_p}\tilde{c}_2 n^{-l/m}\Big) \\ &\triangleq M - c_2 n^{-l/m}. \end{aligned}$$

Putting this together, we get

$$\Delta(\rho, \rho^*, r) = M - D^* \geq c_2 n^{-l/m}.$$

Now, by definition $\rho \in \Gamma^l$ and by construction $\rho^* \in \Gamma^l(\rho, \mathbf{s}_n)$. Hence,

$$\sup_{\substack{\rho \in \Gamma^l \\ \tilde{\rho} \in \Gamma^l(\rho, \mathbf{s}_n) \\ r \geq 0}} \Delta(\rho, \tilde{\rho}, r) \geq c_2 n^{-l/m}. \quad (3)$$

Since (3) holds true for all $\mathbf{s}_n \in \mathbf{S}_n$, we have

$$\varepsilon_2(\Gamma^l, n) = \inf_{\mathbf{s}_n \in \mathbf{S}_n} \sup_{\substack{\rho \in \Gamma^l \\ \tilde{\rho} \in \Gamma^l(\rho, \mathbf{s}_n) \\ r \geq 0}} \Delta(\rho, \tilde{\rho}, r) \geq c_2 n^{-l/m}.$$

This completes the proof of Lemma 4. ∎

The proof of Lemma 4 is based on the construction of a function $f \in W_q^l(\mathcal{X})$ which vanishes at a prescribed number of sampling points as described in [9] and [10]. This construction divides the unit cube $\mathcal{X}$ into $2^m n$ subcubes. As the number of sampling points is n, at least $(2^m - 1)n$ of these subcubes do not contain any sampling point. The function f is constructed by placing a smooth "bump function" ϕ in each empty subcube such that the support of ϕ is entirely within this subcube.

IV. Proof of Corollary 2

We have seen that with non-adaptive sampling the loss $\varepsilon_i(\Gamma^l, n)$ behaves as $n^{-l/m}$ for large n for both $i \in \{1, 2\}$. The next corollary shows that even with adaptive sampling we obtain the same asymptotic behavior. Thus, at least in an asymptotic sense, there is no benefit to be gained from adaptive sampling.

Corollary 5 (Corollary 2)**.** *For $i \in \{1, 2\}$ as $n \to \infty$*

$$\varepsilon_i^A(\Gamma^l, n) = \Theta(n^{-l/m}).$$

Proof: By definition $\varepsilon_i^A(\Gamma^l, n) \leq \varepsilon_i(\Gamma^l, n)$ for $i \in \{1, 2\}$. By Theorem 1, we hence have $\varepsilon_i^A(\Gamma^l, n) = O(n^{-l/m})$ for $i \in \{1, 2\}$ as $n \to \infty$.

As $\varepsilon_2^A(\Gamma^l, n) \leq \varepsilon_1^A(\Gamma^l, n)$, it suffices to show a corresponding lower bound for $\varepsilon_2^A(\Gamma^l, n)$. But this follows now from the proof of Lemma 4. Indeed, if $\rho(x, y) = M$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, the adaptive sampling strategy will result in a set of sampling points $\{(x_i(\rho), y_i(\rho))\}_{1 \leq i \leq n}$. We can now construct a ρ^* which equals ρ at these point, exactly as we have done in the proof of Lemma 4. Thus $\varepsilon_2^A(\Gamma^l, n) = \Omega(n^{-l/m})$ as $n \to \infty$, concluding the proof. ∎

V. Conclusion

In this paper, we have looked at the problem of lossy source coding with partial knowledge about the distortion measure. More precisely, the distortion measure is only known at n sampling points. We have described several measures for the loss we incur through the lack of full knowledge of the true distortion measure, each with a different operational meaning. We have characterized the behavior of this loss in terms of three key parameters: The number of sampling points n, the dimensionality of the source m, and the smoothness assumptions on the distortion measure (quantified by the number l of its derivatives with bounded L_q norm).

The asymptotic behavior for each of the different loss measures considered is $\Theta(n^{-l/m})$ as $n \to \infty$. The fact that these fairly different operational meanings (computation of rate distortion function for the worst case versus construction of universal source code, fixed sampling points versus adaptive sampling) have the same asymptotic behavior, suggests that the $n^{-l/m}$ scaling of the loss is somewhat robust with respect to small changes in the model assumptions.

Acknowledgments

The authors would like to thank Venkat Chandrasekaran, Vivek Goyal, Benjamin Stamm, and Gilbert Strang for helpful discussions.

References

[1] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Prentice Hall, 1971.

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

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

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

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

[6] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, 1978.

[7] A. Pinkus. *n-Widths in Approximation Theory*. Springer Verlag, 1985.

[8] H. Wozniakowski. Information-based complexity. *Annual Review of Computer Science*, 1:319–380, 1986.

[9] S. N. Kudryavtsev. Recovering a function with its derivatives from function values at a given number of samples. *Russian Academy of Sciences Izvestiya Mathematics*, 45(3):505–528, 1995.

[10] S. Heinrich. Random approximation in numerical analysis. In K. D. Bierstedt, A. Pietsch, W. M. Ruess, and D. Vogt, editors, *Functional Analysis*, pages 123–171. Marcel Dekker, 1993.