Universal Single-Shot Sampling Rate Distortion

Sagnik Bhattacharya and Prakash Narayan[†]

***Abstract*—Consider a finite set of multiple sources, described by a random variable with m components. Only $k \leq m$ source components are sampled and jointly compressed in order to reconstruct all the m components under an excess distortion criterion. Sampling can be that of a fixed subset A with $|A| = k$ or randomized over all subsets of size k. In the case of random sampling, the sampler may or may not be aware of the m source components. The compression code consists of an encoder whose input is the realization of the sampler and the sampled source components; the decoder input is solely the encoder output. The combined sampling mechanism and rate distortion code are universal in that they must be devised without exact knowledge of the prevailing source probability distribution. In a Bayesian setting, considering coordinated single-shot sampling and compression, our contributions involve achievability results for the cases of fixed-set, source-independent and source-dependent random sampling.**

***Index Terms*—universal sampling rate distortion, independent random sampler, dependent random sampler**

I. Introduction

Consider a set $\mathcal{M}$ of $m \geq 2$ finite-valued sources described by a random variable (rv) with m components and with a probability mass function (pmf) known only to belong to a given finite family of pmfs. A subset of $k \leq m$ sources is sampled *spatially* and compressed jointly, with the objective of reconstructing all the m sources from the compressed representation within a specified level of distortion. *Universality* requires that a combined sampling procedure and lossy compression code be specified without knowledge of the underlying pmf. How should a randomized sampler optimally sample the sources to form an efficient *single-shot* rate distortion code with the best compression rate for a given distortion level? What are the tradeoffs among sampling mechanisms, estimating the prevailing pmf of the source, code size and distortion level?

The study of coordinated sampling and compression has a long and wide history. Recent relevant contributions in information theoretic settings include lossless source coding of analog sources [20]; compressed sensing subject to error criteria [15]; sub-Nyquist temporal sampling and lossy compression [7]; and rate distortion study for multiple sources with time-shared sampling [12]. In another line of work, the rate distortion function has been characterized when multiple signals from a Gaussian random field are sampled and quantized [14]; centralized and distributed coding schemes for acoustic sensing are studied in [8]; and a Gaussian random field on $[0, 1]$ is reconstructed under a mean-squared error criterion from compressed versions of finitely-many sampled sequences [6]. All these works assume a knowledge of the underlying probability distribution of the signals.

Universal sampling rate distortion theory, where a complete knowledge of the underlying probability distribution is lacking, has been investigated in the framework of classical Bayesian and nonBayesian methods [13]; individual sequence approach [19], [22]; and lossy compression of noisy or remote signals [4], [11], [18]. All these works study asymptotic performance in the limit of increasing code blocklenghts. A single-shot universal rate distortion analysis under an expected log-loss distortion criterion can be found in [16].

Our work constitutes a *single-shot* approach to the concept of universal sampling rate distortion [3] which built on [2]. An ingredient in [2], [3] was the rate distortion function for a *remote* source-receiver model with known probability distributions [1], [5], [21]. Single-shot and finite code blocklength investigations of the remote source-receiver setting with known distributions [10], that are built upon [17] and [9], are pertinent to our work on universal sampling rate distortion and provide useful technical tools.

This work differs materially from the approaches above and, in particular, is distinct from the results in [3]. Our technical contributions are as follows. We develop single-shot achievability results (rather than asymptotic rate distortion tradeoffs) for the sampling schemes of [2] under the excess (distortion) probability criterion (rather than expected distortion). As in [3], here we consider universality that involves a lack of specific knowledge of source pmf within a *finite* family of pmfs. However, whereas the asymptotic analysis in [3] enabled *rate-free* conveyance of consistent pmf estimates by the encoder to the decoder, the single-shot analysis exacts a penalty on coding rate. Unlike in [3], converse results present a difficult challenge, owing to the underlying pmf estimation, and are currently under investigation. An important motivation for our present work is an understanding of single-shot universal sampling rate distortion for random field models.

Our models are described in Section II. The main achievability results, with proof sketches, are presented in Section III. Section IV puts our results in the context of prior work. Section V deals with the proof of Lemma 1 which serves as a key technical tool.

II. Preliminaries

Let $\mathcal{M} = \{1, \ldots, m\}$, $m \geq 2$. Let $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ be a $\mathcal{X}_{\mathcal{M}} = \times_{i=1}^{m} \mathcal{X}_i$-valued rv where each $\mathcal{X}_i$ is a finite set. We shall use the following notation. For $A \subseteq \mathcal{M}$, $A \neq \emptyset$, we denote the rv $X_A = (X_i, i \in A)$ with values in $\times_{i \in A} \mathcal{X}_i$. For

[†]S. Bhattacharya and P. Narayan are with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742, USA. E-mail: {sagnikb, prakash}@umd.edu. This work was supported by the U.S. National Science Foundation under Grant CCF1910497.

$1 \le k \le m$, let $\mathcal{A}_k = \{A : A \subseteq \mathcal{M}, |A| = k\}$ be the set of all k-sized subsets of $\mathcal{M}$. Let $\mathcal{Y}_{\mathcal{M}} = \times_{i=1}^{m} \mathcal{Y}_i$, where $\mathcal{Y}_i$ is a finite reproduction alphabet of X_i. All logs and exps are with respect to the base 2.

Let Θ be a finite set (of parameters) and θ a Θ-valued rv with pmf P_θ of assumed full support. We consider a discrete source $X_{\mathcal{M}}$ with pmf known only to the extent of belonging to a finite family of pmfs $\mathcal{P} = \{P_{X_{\mathcal{M}} \mid \theta=\tau}, \tau \in \Theta\}$ of assumed full support. Two settings are of interest: the pmf P_θ is assumed to be known in a Bayesian formulation, whereas in a nonBayesian formulation, θ is an unknown constant in Θ.

Definition 1. *In a Bayesian setting, a k-independent random sampler (k-IRS) collects a random sample X_S from $X_{\mathcal{M}}$ where S is an $\mathcal{A}_k$-valued rv that is independent of $X_{\mathcal{M}}$ and θ, and has pmf P_S. The output of a k-IRS is (S, X_S). The special case of a k-IRS with P_S = point mass on a fixed $A \in \mathcal{A}_k$ is called a k-fixed set sampler (k-FS). A k-dependent random sampler (k-DRS) is similar except that S can depend on $X_{\mathcal{M}}$ and is conditionally independent of θ given $X_{\mathcal{M}}$ with (conditional) pmf $P_{S \mid X_{\mathcal{M}}}$, and has output (S, X_S). In a nonBayesian setting, S is independent of $X_{\mathcal{M}}$ for a k-IRS but can depend on $X_{\mathcal{M}}$ for a k-DRS.*

Our objective is to reconstruct $X_{\mathcal{M}}$ from a compressed representation of the k-IRS or k-DRS output (S, X_S) under a suitable distortion criterion. We shall restrict ourselves to the Bayesian setting throughout the rest of the paper.

Definition 2. *A code with k-IRS (resp. k-DRS) for the source rv $X_{\mathcal{M}}$ with alphabet $\mathcal{X}_{\mathcal{M}}$ and reproduction alphabet $\mathcal{Y}_{\mathcal{M}}$ is a triple (P_S, f, ϕ) (resp. $(P_{S \mid X_{\mathcal{M}}}, f, \phi)$) where P_S (resp. $P_{S \mid X_{\mathcal{M}}}$) is a k-IRS (resp. k-DRS) as in Definition 1, and (f, ϕ) are a pair of mappings where the encoder f maps the k-IRS (resp. k-DRS) output (S, X_S) into a finite set $\mathcal{J} = \{1, \ldots, J\}$ and the decoder ϕ maps $\mathcal{J}$ into $\mathcal{Y}_{\mathcal{M}}$.*

We are given a finite-valued distortion measure $d : \mathcal{X}_{\mathcal{M}} \times \mathcal{Y}_{\mathcal{M}} \to \mathbb{R}^+ \cup \{0\}$ with

$$\min_{y_{\mathcal{M}} \in \mathcal{Y}_{\mathcal{M}}} d(x_{\mathcal{M}}, y_{\mathcal{M}}) = 0$$

for every $x_{\mathcal{M}} \in \mathcal{X}_{\mathcal{M}}$, and

$$\max_{(x_{\mathcal{M}}, y_{\mathcal{M}})} d(x_{\mathcal{M}}, y_{\mathcal{M}}) = D$$

for some $D > 0$. Then, a code (P_S, f, ϕ) will be required to satisfy the (excess) distortion probability criterion (ϵ, Δ)

$$P\left(d(X_{\mathcal{M}}, \phi(f(X_S))) > \Delta\right) \le \epsilon \quad 0 \le \Delta \le D,\ 0 < \epsilon < 1 \tag{1}$$

where

$$P = \begin{cases} P_\theta P_{X_{\mathcal{M}} \mid \theta} P_S & \text{for a } k\text{-IRS} \\ P_\theta P_{X_{\mathcal{M}} \mid \theta} P_{S \mid X_{\mathcal{M}}} & \text{for a } k\text{-DRS.} \end{cases} \tag{2}$$

We shall not consider here the case where the decoder is provided additional side information regarding the sampled set S.

Definition 3. *A number $R = \log J$ is an achievable k-sample coding rate*[1] *under distortion (ϵ, Δ) if for every $\delta > 0$, there exists a code (f, ϕ) with k-IRS of rate $\le R + \delta$ and satisfying the distortion criterion (ϵ, Δ) in (1). The infimum of such achievable rates is denoted by $R(\epsilon, \Delta)$. We shall refer to $R(\epsilon, \Delta)$ as the single-shot sampling rate distortion function (SSRDf) suppressing the dependence on k. A similar definition holds for the DRS. Note that $R(\epsilon, \Delta)$ depends on θ through P in (1), (2).*

Definition 4. *Given a pmf P_{XYZ} on a finite (product) set $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, the information density of X, Y conditioned on Z is*

$$\iota_{X \wedge Y \mid Z}(x \wedge y \mid Z = z) = \log \frac{P_{Y \mid X,Z}(y \mid x, z)}{P_{Y \mid Z}(y \mid z)},$$

for $x \in \mathcal{X}$, $y \in \mathcal{Y}$, $z \in \mathcal{Z}$, under the assumption that

$$P_{Y \mid Z}(y \mid z) > 0 \quad y \in \mathcal{Y}, z \in \mathcal{Z}.$$

III. Main Results and Sketches of Proofs

For a k-FS with $A \in \mathcal{A}_k$, we note that an encoder with access to X_A cannot distinguish among pmfs in $\mathcal{P}$ (indexed by τ) that have the same $P_{X_A \mid \theta=\tau}$. Clearly, any estimate $\hat{\theta}(X_A)$ of θ will suffer from a significant estimation error. In the same vein, an encoder with access to the output of a k-IRS or k-DRS cannot distinguish among pmfs in $\mathcal{P}$ that have identical k-marginal pmfs $\{P_{X_{A_i} \mid \theta=\tau}, i = 1, \ldots, |\mathcal{A}_k|\}$, with consequent significant errors in estimating θ. We do not delve below into such estimation ambiguities and assume instead that the pmfs in $\mathcal{P}$ are varied enough to enable meaningful estimation. Our achievability proofs rely on such estimation.

Our main Theorems 2, 3 and 4 constitute achievability results for the k-FS, k-IRS and k-DRS, respectively. We first present a key technical lemma that underlies the proofs of the main theorems. This is a conditional version of Corollary 2 in [10].

Lemma 1. *Fix $\gamma > 0$ and $A \in \mathcal{A}_k$. For each $\tau \in \Theta$, pmf*

$$P_{X_{\mathcal{M}} Y_{\mathcal{M}} \mid \theta=\tau} = P_{X_{\mathcal{M}} \mid \theta=\tau} P_{Y_{\mathcal{M}} \mid X_A},$$

and positive integer l_τ, there exists a code (f_τ, ϕ_τ) of rate $\log l_\tau$ that satisfies

$$\begin{aligned} &P\left(d(X_{\mathcal{M}}, \phi_\tau(f_\tau(X_A))) > \Delta \mid \theta = \tau\right) \\ &\quad \le P\left(d(X_{\mathcal{M}}, Y_{\mathcal{M}}) > \Delta \mid \theta = \tau\right) + e^{-\exp(\gamma)} \\ &\quad + P\left(\iota_{X_A \wedge Y_{\mathcal{M}} \mid \theta}(X_A \wedge Y_{\mathcal{M}} \mid \theta) > \log l_\tau - \gamma \mid \theta = \tau\right). \end{aligned}$$

The rate- and pmf-related notions below are needed to state our main theorems.

Definition 5. *For each sampling mechanism, viz. k-FS, k-IRS and k-DRS, Table I denotes the following. For a given $0 < \epsilon < 1$, $\mathcal{E}^{(\cdot)}(\epsilon)$ specifies the set of all τ-dependent thresholds with expected value not exceeding ϵ. For each $\boldsymbol{\epsilon} \in \mathcal{E}^{(\cdot)}(\epsilon)$, $\mathcal{D}^{(\cdot)}(\boldsymbol{\epsilon})$ denotes the set of all rvs $Y_{\mathcal{M}}$ consistent with the form of the joint pmf shown and satisfying the given distortion condition.*

[1]While $R = \log J$ is, in fact, a log code size, we shall conform to the traditional terminology of coding rate.

TABLE I
DISTORTION THRESHOLDS AND JOINT DISTRIBUTIONS

k-FS
$\mathcal{E}^{\mathsf{FS}}(\epsilon) = \{\boldsymbol{\epsilon} = (\epsilon_\tau, \tau \in \Theta) : \sum_\tau P_\theta(\tau)\epsilon_\tau < \epsilon\}$ $\mathcal{D}^{\mathsf{FS}}(\boldsymbol{\epsilon}) = \{Y_\mathcal{M} : P_{\theta X_\mathcal{M} Y_\mathcal{M}} = P_\theta P_{X_\mathcal{M} \mid \theta} P_{Y_\mathcal{M} \mid X_A},$ $P(d(X_\mathcal{M}, Y_\mathcal{M}) > \Delta \mid \theta = \tau) < \epsilon_\tau/2, \tau \in \Theta\}, \boldsymbol{\epsilon} \in \mathcal{E}^{\mathsf{FS}}(\epsilon)$
k-IRS
$\mathcal{E}^{\mathsf{IRS}}(\epsilon) = \{\boldsymbol{\epsilon} = (\epsilon_{s\tau} : \tau \in \Theta, s \in \mathcal{A}_k) : \sum_{s,\tau} P_\theta(\tau) P_S(s) \epsilon_{s\tau} < \epsilon\}$ $\mathcal{D}^{\mathsf{IRS}}(\boldsymbol{\epsilon}) = \{Y_\mathcal{M} : P_{\theta S X_\mathcal{M} X_S Y_\mathcal{M}} = P_\theta P_S P_{X_\mathcal{M} \mid \theta} P_{Y_\mathcal{M} \mid S X_S},$ $P(d(X_\mathcal{M}, Y_\mathcal{M}) > \Delta \mid \theta = \tau) < \epsilon_{s\tau}/2, \tau \in \Theta, A \in \mathcal{A}_k\}, \boldsymbol{\epsilon} \in \mathcal{E}^{\mathsf{IRS}}(\epsilon)$
k-DRS
$\mathcal{E}^{\mathsf{DRS}}(\epsilon) = \{\boldsymbol{\epsilon} = (\epsilon_\tau, \tau \in \Theta) : \sum_\tau P_\theta(\tau)\epsilon_\tau < \epsilon\}$ $\mathcal{D}^{\mathsf{DRS}}(\boldsymbol{\epsilon}) = \{Y_\mathcal{M} : P_{\theta S X_\mathcal{M} X_S Y_\mathcal{M}} = P_\theta P_S P_{X_\mathcal{M} \mid \theta} P_{Y_\mathcal{M} \mid S X_S},$ $P(d(X_\mathcal{M}, Y_\mathcal{M}) > \Delta \mid \theta = \tau) < \epsilon_\tau/2, \tau \in \Theta\}, \boldsymbol{\epsilon} \in \mathcal{E}^{\mathsf{DRS}}(\epsilon)$

Theorem 2 (Achievability for k-FS)**.** *Let $A \in \mathcal{A}$. Fix $\gamma > 0$, $0 < \epsilon < 1$. For $\boldsymbol{\epsilon} = (\epsilon_\tau : \tau \in \Theta) \in \mathcal{E}^{\mathsf{FS}}(\epsilon)$, let $Y_\mathcal{M}$ be an rv in $\mathcal{D}^{\mathsf{FS}}(\boldsymbol{\epsilon})$. Let $L^{\mathsf{FS}}(A, \tau, \Delta, \epsilon_\tau, \gamma)$ be the smallest integer l_τ such that*

$$P(\iota_{X_A \wedge Y_\mathcal{M} \mid \theta}(X_A \wedge Y_\mathcal{M} \mid \theta) > \log l_\tau - \gamma \mid \theta = \tau) < \frac{\epsilon_\tau}{2}$$

Let $\hat{\theta} = \hat{\theta}(X_A)$ be any Bayesian estimate of θ on the basis of X_A. Then

$$R(\Delta) \leq \min_{\boldsymbol{\epsilon} \in \mathcal{E}^{\mathsf{FS}}(\epsilon)} \min_{Y_\mathcal{M} \in \mathcal{D}^{\mathsf{FS}}(\boldsymbol{\epsilon})} \max_{\tau \in \Theta} \log L^{\mathsf{FS}}(A, \tau, \Delta, \epsilon_\tau, \gamma) + \mathcal{O}(\log |\Theta|), \quad 0 \leq \Delta \leq D$$

under the distortion criterion $(\epsilon + P(\hat{\theta} \neq \theta) + e^{-\exp(\gamma)}, \Delta)$.

Sketch of proof. For any $\boldsymbol{\epsilon} \in \mathcal{E}^{\mathsf{FS}}(\epsilon)$, fix an rv $Y_\mathcal{M} \in \mathcal{D}^{\mathsf{FS}}(\boldsymbol{\epsilon})$ such that

$$P_{\theta X_\mathcal{M} Y_\mathcal{M}} = P_\theta P_{X_\mathcal{M} \mid \theta} P_{Y_\mathcal{M} \mid X_A}.$$

For each $\tau \in \Theta$, pick $l_\tau \geq 1$ conditionally i.i.d. codewords $\{Y_{\mathcal{M},\tau,i} : 1 \leq i \leq l_\tau\}$ with pmf $P_{Y_\mathcal{M} \mid \theta = \tau}$, where

$$\begin{aligned} P_{X_\mathcal{M} Y_{\mathcal{M},\tau,1} \ldots Y_{\mathcal{M},\tau,l_\tau} \mid \theta = \tau} &= P_{X_\mathcal{M} \mid \theta = \tau} \prod_{i=1}^{l_\tau} P_{Y_{\mathcal{M},\tau,i} \mid \theta = \tau} \\ &= P_{X_\mathcal{M} \mid \theta = \tau} \left[P_{Y_\mathcal{M} \mid \theta = \tau} \right]^{l_\tau}. \end{aligned}$$

For fixed realizations $\{y_{\mathcal{M},\tau,i} : 1 \leq i \leq l_\tau\}$ of the rvs $\{Y_{\mathcal{M},\tau,i} : 1 \leq i \leq l_\tau\}$, the code is specified as follows. Based on $x_A \in \mathcal{X}_A$, the encoder forms an estimate $\hat{\tau} = \hat{\tau}(x_A)$ of θ and maps x_A to a pair $(i^*, \hat{\tau})$ according to

$$f_{\hat{\tau}}(x_A) = (j^*, \hat{\tau}) \text{ where } i^* = \arg \min_{1 \leq i \leq l_{\hat{\tau}}} \pi(\hat{\tau}, x_A, y_{\mathcal{M},\hat{\tau},i}).$$

The corresponding decoder output is $\phi_{\hat{\tau}}(f_{\hat{\tau}}(x_A)) = y_{\mathcal{M},\hat{\tau},i^*}$. The set $\mathcal{J}$ (see Definition 2) is given by

$$\mathcal{J} = \{(\tau, i_\tau) : 1 \leq i_\tau \leq l_\tau, \tau \in \Theta\}.$$

Then the expected distortion probability is

$$\begin{aligned} & P\left(d(X_\mathcal{M}, \Phi_{\hat{\theta}}(F_{\hat{\theta}}(X_A))) > \Delta\right) \\ &= \mathbb{E}\left[\mathbb{1}\left(d(X_\mathcal{M}, \Phi_{\hat{\theta}}(F_{\hat{\theta}}(X_A))) > \Delta\right) \mathbb{1}\left(\hat{\theta} = \theta\right)\right] \\ &\quad + \mathbb{E}\left[\mathbb{1}\left(d(X_\mathcal{M}, \Phi_{\hat{\theta}}(F_{\hat{\theta}}(X_A))) > \Delta\right) \mathbb{1}\left(\hat{\theta} \neq \theta\right)\right] \\ &\leq P\left(d(X_\mathcal{M}, \Phi_\theta(F_\theta(X_A))) > \Delta\right) + P\left(\hat{\theta} \neq \theta\right) \\ &= \sum_{\tau \in \Theta} P_\theta(\tau) P\left(d(X_\mathcal{M}, \Phi_\theta(F_\theta(X_A))) > \Delta \mid \theta = \tau\right) \\ &\quad + P\left(\hat{\theta} \neq \theta\right). \end{aligned}$$

To prove the claim on the rate in the theorem, we apply Lemma 1 to each term in the summand above. In particular, for a fixed choice of $\boldsymbol{\epsilon} = (\epsilon_\tau, \tau \in \Theta) \in \mathcal{E}^{\mathsf{FS}}(\epsilon)$ and rv $Y_\mathcal{M} \in \mathcal{D}^{\mathsf{FS}}(\boldsymbol{\epsilon})$, the max over $\tau \in \Theta$ and $\mathcal{O}(\log |\Theta|)$ reflect a union of codes over $\tau \in \Theta$. The outer minima indicate optimizations of the error thresholds and pmf of $Y_\mathcal{M}$. □

Theorem 3 (Achievability for k-IRS)**.** *Fix $\gamma > 0$, $0 < \epsilon < 1$. Let $\mathcal{S}^{\mathsf{IRS}}$ be the set of rvs S such that*

$$P_{\theta S X_\mathcal{M}} = P_\theta P_{X_\mathcal{M} \mid \theta} P_S.$$

For $\boldsymbol{\epsilon} = (\epsilon_{s\tau} : \tau \in \Theta, s \in \mathcal{A}_k) \in \mathcal{E}^{\mathsf{IRS}}(\epsilon)$, let $Y_\mathcal{M}$ be an rv in $\mathcal{D}^{\mathsf{IRS}}(\boldsymbol{\epsilon})$. Let $L^{\mathsf{IRS}}(s, \tau, \Delta, \epsilon_{s\tau}, \gamma)$ be the smallest integer l_τ such that

$$P(\iota_{X_s \wedge Y_\mathcal{M} \mid \theta}(X_s \wedge Y_\mathcal{M} \mid \theta) > \log l_\tau - \gamma \mid \theta = \tau) < \frac{\epsilon_{s\tau}}{2}$$

Let $\hat{\theta} = \hat{\theta}(S, X_S)$ be any Bayesian estimate of θ on the basis of (S, X_S). Then

$$\begin{aligned} & R(\Delta) \leq \\ & \min_{S \in \mathcal{S}^{\mathsf{IRS}}} \min_{\boldsymbol{\epsilon} \in \mathcal{E}^{\mathsf{IRS}}(\epsilon)} \min_{Y_\mathcal{M} \in \mathcal{D}^{\mathsf{IRS}}(\boldsymbol{\epsilon})} \max_{\tau \in \Theta} \log L^{\mathsf{IRS}}(s, \tau, \Delta, \epsilon_{s\tau}, \gamma) \\ & \qquad + \mathcal{O}(\log |\Theta|), \quad 0 \leq \Delta \leq D \end{aligned}$$

under the distortion criterion $(\epsilon + P(\hat{\theta} \neq \theta) + e^{-\exp(\gamma)}, \Delta)$.

Sketch of proof. The proof idea of Theorem 2 is modified, using the independence of S from $(\theta, X_\mathcal{M})$ as follows:

$$\begin{aligned} & P\left(d(X_\mathcal{M}, \Phi_{\hat{\theta}}(F_{\hat{\theta}}(S, X_S))) > \Delta\right) \\ &= \sum_{s,\tau} P_S(s) P_\theta(\tau) P\left(d(X_\mathcal{M}, \Phi_{\hat{\theta}}(F_{\hat{\theta}}(s, X_s))) > \Delta \mid \theta = \tau\right). \end{aligned}$$

Applying Theorem 1 to the summand above, note the additional latitude in the choice of $\boldsymbol{\epsilon} = (\epsilon_{s\tau}, s \in \mathcal{A}_k, \tau \in \Theta) \in \mathcal{E}^{\mathsf{IRS}}(\epsilon)$, rv $Y_\mathcal{M} \in \mathcal{D}^{\mathsf{IRS}}(\boldsymbol{\epsilon})$, and P_S. The maximum over $\tau \in \Theta$ and the $\mathcal{O}(\log(|\Theta|))$ term arise as in Theorem 2. □

Theorem 4 (Achievability for k-DRS)**.** *Fix $\gamma > 0$, $\epsilon > 0$. Let $\mathcal{S}^{\mathsf{DRS}}$ be the set of rvs S such that*

$$P_{\theta S X_\mathcal{M}} = P_\theta P_{X_\mathcal{M} \mid \theta} P_{S \mid X_\mathcal{M}}.$$

For $\boldsymbol{\epsilon} = (\epsilon_\tau : \tau \in \Theta) \in \mathcal{E}^{\mathsf{DRS}}(\epsilon)$, let $Y_\mathcal{M}$ be an rv in $\mathcal{D}^{\mathsf{DRS}}(\boldsymbol{\epsilon})$. Let $L^{\mathsf{DRS}}(\tau, \Delta, \epsilon_\tau, \gamma)$ be the smallest integer l_τ such that

$$P(\iota_{S, X_S \wedge Y_\mathcal{M} \mid \theta}(S, X_S \wedge Y_\mathcal{M} \mid \theta) > \log l_\tau - \gamma \mid \theta = \tau) < \frac{\epsilon_\tau}{2}.$$

Let $\hat{\theta} = \hat{\theta}(S, X_S)$ be any Bayesian estimate of θ on the basis of (S, X_S). Then

$$R(\Delta) \leq \min_{S\in\mathcal{S}^{\mathsf{DRS}}} \min_{\boldsymbol{\epsilon}\in\mathcal{E}^{\mathsf{DRS}}(\epsilon)} \min_{Y_{\mathcal{M}}\in\mathcal{D}^{\mathsf{DRS}}(\boldsymbol{\epsilon})} \max_{\tau\in\Theta} \log L^{\mathsf{DRS}}(\tau, \Delta, \epsilon_\tau, \gamma) + \mathcal{O}\left(\log|\Theta|\right), \quad 0 \leq \Delta \leq D$$

under the distortion criterion $(\epsilon + P(\hat{\theta} \neq \theta) + e^{\exp(-\gamma)}, \Delta)$.

Sketch of proof. The proof is along the lines of that of Theorem 2 and the outer minimization over $\mathcal{S}^{\mathsf{DRS}}$ is akin to that in Theorem 3. □

Remark: In the nonBayesian case, the results of Theorems 2, 3 and 4 are modified by changing the outer minima appropriately. Now we cannot average over error thresholds in the absence of an underlying P_θ; instead, maxima must be taken over all pmfs in the family $\mathcal{P}$.

IV. Discussion

In the case $|\mathcal{M}| = 1$, our results particularize to yield a simple single-shot universal rate distortion result (without any sampling) that appears to be new. For arbitrary $\mathcal{M}$, the case $|\Theta| = 1$ and fixed-set sampling corresponds to [10], and gives single-shot versions of the achievability results in [2] for other sampling mechanisms, too.

Our present notion of universality is with respect to a finite number of sources. Involving infinitely many sources is a formidable technical challenge, and a complete answer is not available even in the asymptotic case. We also mention that in a setting that does not involve sampling, and under a log-loss distortion measure, a single shot rate redundancy analysis has been studied in [16].

It is an artifact of single-shot analysis, especially involving operations of sampling and compression, that multiple max/min creep into expressions for achievable rates. In case of a finite-length or asymptotic (in n) analysis (which can be obtained, for instance, by extending our results using techniques in [10]), such optima often resolve themselves through interchanges and actual evaluations of formulae exploiting convexity/concavity of average information quantities.

V. Proof of Lemma 1

The proof of Lemma 1 hinges on the following Lemmas 5 and 6. Lemma 5 generalizes Theorem 3 in [10].

Lemma 5. *Fix $A \in \mathcal{A}_k$. For each $\tau \in \Theta$, pmf $P_{X_{\mathcal{M}}Y_{\mathcal{M}}\,|\,\theta=\tau} = P_{X_{\mathcal{M}}\,|\,\theta=\tau} P_{Y_{\mathcal{M}}\,|\,X_A}$, and positive integer l_τ, there exists a (deterministic) code (f_τ, ϕ_τ) of rate $\log l_\tau$ that satisfies*

$$P\left(d(X_{\mathcal{M}}, \phi_\tau(f_\tau(X_A))) > \Delta \,|\, \theta = \tau\right) \leq \int_0^1 \mathbb{E}\left[\prod_{i=1}^{l_\tau} P\left(\pi(\tau, X_A, Y_{\mathcal{M},\tau,i}) > t \,|\, X_A, \theta\right) \,|\, \theta = \tau\right] dt$$

where

$$\pi(\tau, x_A, y_{\mathcal{M}}) = P\left(d(X_{\mathcal{M}}, y_{\mathcal{M}}) > \Delta \,|\, X_A = x_A, \theta = \tau\right)$$

and

$$P_{X_{\mathcal{M}}Y_{\mathcal{M},\tau,1}\ldots Y_{\mathcal{M},\tau,l_\tau}\,|\,\theta=\tau} = P_{X_{\mathcal{M}}\,|\,\theta=\tau} \prod_{i=1}^{l_\tau} P_{Y_{\mathcal{M},\tau,i}\,|\,\theta=\tau} = P_{X_{\mathcal{M}}\,|\,\theta=\tau} \left[P_{Y_{\mathcal{M}}\,|\,\theta=\tau}\right]^{l_\tau}.$$

Proof. For each $\tau \in \Theta$, pick $l_\tau \geq 1$ conditionally i.i.d. codewords $\{Y_{\mathcal{M},\tau,i} : 1 \leq i \leq l_\tau\}$ with pmf $P_{Y_{\mathcal{M}}\,|\,\theta=\tau}$ where

$$P_{X_{\mathcal{M}}Y_{\mathcal{M},\tau,1}\ldots Y_{\mathcal{M},\tau,l_\tau}\,|\,\theta=\tau} = P_{X_{\mathcal{M}}\,|\,\theta=\tau} \prod_{i=1}^{l_\tau} P_{Y_{\mathcal{M},\tau,i}\,|\,\theta=\tau} = P_{X_{\mathcal{M}}\,|\,\theta=\tau} \left[P_{Y_{\mathcal{M}}\,|\,\theta=\tau}\right]^{l_\tau}.$$

For fixed realizations $\{y_{\mathcal{M},\tau,i} : 1 \leq i \leq l_\tau\}$ of the rvs $\{Y_{\mathcal{M},\tau,i} : 1 \leq i \leq l_\tau\}$, the code is specified as follows. The encoder maps x_A to i^* according to

$$f_\tau(x_A) = i^* \text{ where } i^* = \arg\min_{1\leq i\leq l_\tau} \pi(\tau, x_A, y_{\mathcal{M},\tau,i}).$$

The corresponding decoder output is

$$\phi_\tau(f_\tau(x_A)) = y_{\mathcal{M},\tau,i^*}.$$

For each $\tau \in \theta$ and the random code corresponding to the rvs $\{Y_{\mathcal{M},\tau,i} : 1 \leq i \leq l_\tau\}$, we get

$$P\left(d(X_{\mathcal{M}}, \Phi_\theta(F_\theta(X_A))) > \Delta \,|\, \theta = \tau\right) = \mathbb{E}\left[P\left(d(X_{\mathcal{M}}, \Phi_\theta(F_\theta(X_A))) > \Delta \,|\, F_\theta, \Phi_\theta, X_A, \theta\right) \,|\, \theta = \tau\right]$$

where the inner conditional probability can be written as

$$P\left(d(X_{\mathcal{M}}, \phi_\tau(f_\tau(x_A))) > \Delta \,|\, F_\tau = f_\tau, \Phi_\tau = f_\tau, X_A = x_A, \theta = \tau\right).$$

Since for a fixed $\tau \in \theta$, $f_\tau(\phi_\tau(\cdot))$ is determined by x_A, we get

$$\mathbb{E}\left[P\left(d(X_{\mathcal{M}}, \Phi_\theta(F_\theta(X_A))) > \Delta \,|\, X_A, \theta\right) \,|\, \theta = \tau\right] = \mathbb{E}\left[\pi(\theta, X_A, \Phi_\theta(F_\theta(X_A))) \,|\, \theta = \tau\right]$$

using

$$\pi(\tau, x_A, y_{\mathcal{M}}) = P\left(d(X_{\mathcal{M}}, y_{\mathcal{M}}) > \Delta \,|\, X_A = x_A, \theta = \tau\right).$$

Standard manipulations then yield

$$\mathbb{E}\left[P\left(d(X_{\mathcal{M}}, \Phi_\theta(F_\theta(X_A))) > \Delta \,|\, X_A, \theta\right) \,|\, \theta = \tau\right] = \int_0^1 \mathbb{E}\left[\prod_{i=1}^{l_\theta} P\left(\pi(\theta, X_A, Y_{\mathcal{M},\theta,i}) > t \,|\, X_A, \theta\right) \,|\, \theta = \tau\right] dt$$

from which the existence of a code (f_τ, ϕ_τ) follows as claimed. □

Lemma 6 below is obtained straightforwardly as a conditional version of Lemma 5 in [17].

Lemma 6. *Given $\gamma > 0$, a positive integer M and*

$$\mathcal{S} = \left\{(x, y) \in \mathcal{X} \times \mathcal{Y}; \iota_{X\wedge Y\,|\,Z}(x \wedge y \,|\, Z = z) \leq \log M - \gamma\right\}$$

for rvs X, Y and Z with joint pmf P_{XYZ} and any event $\mathcal{F}$, we have

$$P\left(\bigcap_{m=1}^{M}\{(X,Y_m)\notin\mathcal{F}\}\,|\,X=x,Z=z\right)$$
$$\leq P\left((X,Y)\notin\mathcal{F}\,|\,X=x,Z=z\right)+e^{-\exp(\gamma)}$$
$$+P\left((X,Y)\notin\mathcal{S}\,|\,X=x,Z=z\right)$$

where

$$P_{Y_1Y_2\ldots Y_MXY\,|\,Z=z}=P_{XY\,|\,Z=z}\times\prod_{m=1}^{M}P_{Y_m\,|\,Z=z}$$
$$=P_{XY\,|\,Z=z}\times\left[P_{Y\,|\,Z=z}\right]^M.$$

Finally, we use Lemmas 5 and 6 to prove Lemma 1.

Proof of Lemma 1. By Lemma 5, a code (f_τ,ϕ_τ) of rate $\log l_\tau$ exists that satisfies

$$P\left(d(X_{\mathcal{M}},\phi_\tau(f_\tau(X_A)))>\Delta\,|\,\theta=\tau\right)$$
$$\leq\int_0^1\mathbb{E}\left[\prod_{i=1}^{l_\tau}P\left(\pi(\tau,X_A,Y_{\mathcal{M},\tau,i})>t\,|\,X_A,\theta\right)\,|\,\theta=\tau\right]$$

with π and the rvs $\{Y_{\mathcal{M},\tau,i}:1\leq i\leq l_\tau\}$ defined as in the statement of Lemma 5. Applying Lemma 6 with

$$\mathcal{F}=\{(x_A,y_{\mathcal{M}})\in\mathcal{X}_A\times\mathcal{Y}_{\mathcal{M}}:\pi(\tau,x_A,y_{\mathcal{M}})\leq t\}$$

to the product of the inner conditional probabilities above, and simplifying, we get the required claim. □

References

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

[2] V. P. Boda and P. Narayan, "Sampling rate distortion," *IEEE Trans. Inform. Theory*, vol. 63, no. 1, pp. 563–574, 2017.

[3] ——, "Universal sampling rate distortion," *IEEE Trans. Inform. Theory*, vol. 64, no. 12, pp. 7742–7758, 2018.

[4] A. Dembo and T. Weissman, "The minimax distortion redundancy in noisy source coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 3020–3030, 2003.

[5] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.

[6] A. Kashyap, L. A. Lastras-Montano, C. Xia, and Z. Liu, "Distributed source coding in dense sensor networks," in *Data Compression Conference*. IEEE, 2005, pp. 13–22.

[7] A. Kipnis, A. J. Goldsmith, Y. C. Eldar, and T. Weissman, "Distortion rate function of sub-nyquist sampled gaussian sources," *IEEE Trans. Inform. Theory*, vol. 62, no. 1, pp. 401–429, 2016.

[8] R. L. Konsbruck, E. Telatar, and M. Vetterli, "On sampling and coding for distributed acoustic sensing," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 3198–3214, 2012.

[9] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Trans. Inform. Theory*, vol. 58, no. 6, pp. 3309–3338, 2012.

[10] ——, "Nonasymptotic noisy lossy source coding," *IEEE Trans. Inform. Theory*, vol. 62, no. 11, pp. 6111–6123, 2016.

[11] T. Linder, G. Lugosi, and K. Zeger, "Empirical quantizer design in the presence of source noise or channel noise," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 612–623, 1997.

[12] X. Liu, O. Simeone, and E. Erkip, "Lossy computing of correlated sources with fractional sampling," *IEEE Trans. Comm.*, vol. 61, no. 9, pp. 3685–3696, 2013.

[13] D. Neuhoff, R. Gray, and L. Davisson, "Fixed rate universal block source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. 21, no. 5, pp. 511–523, 1975.

[14] D. L. Neuhoff and S. S. Pradhan, "Information rates of densely sampled gaussian data," in *2011 IEEE International Symposium on Information Theory Proceedings*. IEEE, 2011, pp. 2776–2780.

[15] G. Reeves and M. Gastpar, "The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 3065–3092, 2012.

[16] Y. Shkel, M. Raginsky, and S. Verdú, "Universal lossy compression under logarithmic loss," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 1157–1161.

[17] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2012, pp. 1–8.

[18] T. Weissman, "Universally attainable error exponents for rate-distortion coding of noisy sources," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 1229–1246, 2004.

[19] T. Weissman and N. Merhav, "Universal prediction of individual binary sequences in the presence of noise," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2151–2173, 2001.

[20] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inform. Theory*, vol. 56, no. 8, pp. 3721–3748, 2010.

[21] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *IEICE TRANSACTIONS (1976-1990)*, vol. 63, no. 10, pp. 700–706, 1980.

[22] J. Ziv, "Distortion-rate theory for individual sequences," *IEEE Trans. Inform. Theory*, vol. 26, no. 2, pp. 137–143, 1980.