ABSTRACT

Title of dissertation: Sampling Rate Distortion

Vinay Praneeth Boda, Doctor of Philosophy, 2017

Dissertation directed by: Professor Prakash Narayan
Department of Electrical and Computer Engineering
Institute for Systems Research

Consider a memoryless multiple source with m components of which a (possibly randomized) subset of $k \leq m$ components are sampled at each time instant and jointly compressed with the objective of reconstructing a prespecified subset of the m components under a given distortion criterion. The combined sampling and lossy compression mechanisms are to be designed to perform robustly with or without exact knowledge of the underlying joint probability distribution of the source. In this dissertation, we introduce a new framework of *sampling rate distortion* to study the tradeoffs among sampling mechanism, encoder-decoder structure, compression rate and the desired level of accuracy in the reconstruction.

We begin with a discrete memoryless multiple source whose joint probability mass function (pmf) is taken to be *known*. A notion of *sampling rate distortion function* is introduced to study the mentioned tradeoffs, and is characterized first for fixed-set sampling. Next, for independent random sampling performed without the knowledge of the source outputs, it is shown that the sampling rate distortion function is the same whether or not the decoder is informed of the sequence of sampled sets. For memoryless random sampling, with the sampling depending on

the source outputs, it is shown that deterministic sampling, characterized by a conditional point-mass, is optimal and suffices to achieve the sampling rate distortion function.

Building on this, we consider a universal setting where the joint pmf of a discrete memoryless multiple source is known only to belong to a *finite* family of pmfs. In Bayesian and nonBayesian settings, single-letter characterizations are provided for the universal sampling rate distortion function for the fixed-set sampling, independent random sampling and memoryless random sampling. We show that these sampling mechanisms successively improve upon each other: (i) in their ability to enable an associated encoder approximate the underlying joint pmf and (ii) in their ability to choose appropriate subsets of the multiple source for compression by the encoder.

Lastly, we consider a jointly Gaussian multiple memoryless source, to be reconstructed under a mean-squared error distortion criterion, with joint probability distribution function known only to belong to an *uncountable* family of probability density functions (characterized by a convex compact subset in Euclidean space). For fixed-set sampling, we characterize the universal sampling rate distortion function in Bayesian and nonBayesian settings. We also provide optimal reconstruction algorithms, of reduced complexity, which compress and reconstruct the sampled source components first under a modified distortion criterion, and then form MMSE estimates for the unsampled components based on reconstructions of the former.

The questions addressed in this dissertation are motivated by various applications, e.g., dynamic thermal management for multicore processors, in-network computation and satellite imaging.

Sampling Rate Distortion

by

Vinay Praneeth Boda

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professor Prakash Narayan, Chair/Advisor
Professor Alexander Barg
Professor Sennur Ulukus
Professor Piya Pal
Professor Benjamin Kedem

Acknowledgments

This dissertation started with a joint project between Prof. Prakash Narayan and Prof. Ankur Srivastava, as a principled way to studying an important problem in computer architecture. In multicore processors, an increase in transistor density lead to rise in temperatures which in turn caused reliability issues. To avoid this, thermal sensors are placed at various points on the processor to estimate the thermal profile which could be used to schedule processes to prevent overheating. The question was, given that only a limited number of sensors can be placed, where should these be placed and how can the measurements from these be used to estimate the thermal profile of the whole processor? While this was the starting point of this dissertation, over time we posed several other interesting questions which led to important problems and applications in many other domains.

This dissertation could not have been completed without the support of my advisor, friends, family and the generous support from U. S. National Science Foundations under the grants CCF-0917057 and CCF-1319799. I would like to express my deepest gratitude to my advisor Prof. Prakash Narayan, for his guidance and unrelenting patience throughout my Ph. D. I learnt a lot from him (till the very last minute), technically, and otherwise. His advice and suggestions have proved valuable in the long term and prescient every time. I am indebted to Himanshu Tyagi, whose timely inputs have have left an indelible mark on this dissertation. I thank Prashanth L. A. for introducing me to the wonderful world of multi-armed bandits. I would like to thank Zachi Tamo for being a patient friend and from whom

I learned a lot. His questions, antics, support and inputs, made my time at UMD, fun and cherishable. I would like to thank my labmates Pritam Mukherjee, Yi-Peng Wei, Berk Gurakan, Raef Bassily, Jianwei Xie, for the many eye-opening and fun discussions.

My time at UMD would not have been so much cherishable without the fun salsa nights with my friends, Brandon, Jori and Dalia, who also taught me the trade. I would like to thank Bhaskar, Devanarayan, Bharat, Navaneeth, Kapil, Sumit and other innumberable friends who made my time at UMD fun and enjoyable. I've been fortunate to have had some wonderful roommates over the last few years, Krishna, Varun, Raviteja, with whom many a exploits were shared and were of constant support. My friends Rakesh Malladi and Varun Pattabhiraman, inspite of being at far flung locations, were companions in this journey.

I thank my aunt for providing me a home away from home. I would like to thank my parents for their unconditional love and unreasonable support throughout. I cannot express enough gratitude to my brother, Sushanth, for being a silent pillar of support over the years.

Table of Contents

List of Abbreviations and Notations

i.i.d	independent and identically distributed
rv	random variable
pdf	probability distribution function or probability density function
pmf	probability mass function
MSE	mean-squared error
MMSE	minimum mean-squared error

Chapter 1

Introduction

1.1 Motivation and Prior Work

Consider a set $\mathcal{M}$ of m memoryless sources with underlying probability distribution known only to belong to a given family of distributions. At time instants $t = 1, \ldots, n$, possibly different subsets of sources of size $k \leq m$ are sampled "spatially" and compressed *jointly* by a block (source) code with the objective of reconstructing a prespecified subset of the m sources from the compressed representations within a specified level of distortion. The sampling and compression are to be designed in the face of partial information about the underlying probability distribution. Which sources should be sampled and when? How should the sampled sources be compressed, and what are the optimal reconstruction algorithms? The focus of

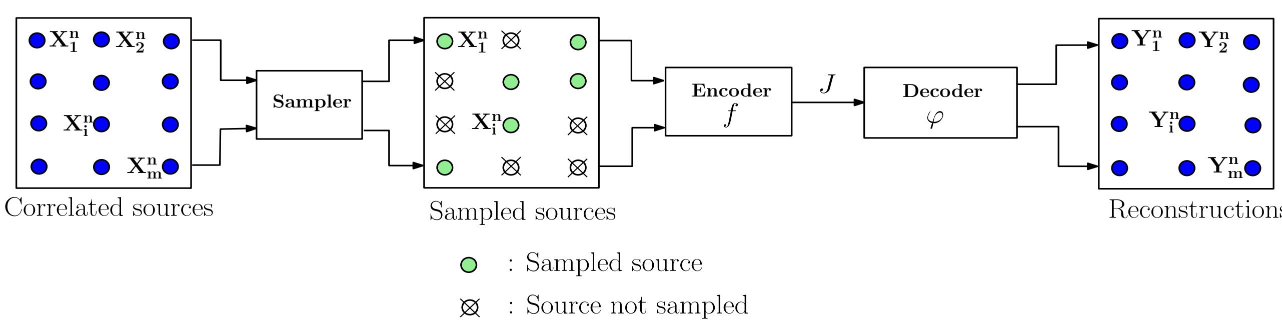

Figure 1.1: Schematic for Sampling Rate Distortion

this dissertation is on answering such questions, where resource constraints, such as bandwidth and computational constraints, restrict the number of sources that can be processed at any time, and furthermore require the sampled sources to be compressed. The sampling and lossy compression mechanisms are to be designed – in a coordinated manner – to be robust in the absence of precise information about the underlying statistics of the sources. In this dissertation we introduce our framework of *sampling rate distortion* for the study of such problems; the processes of sampling, compression and reconstruction of the sources are performed by a *random sampler*, *encoder* and *decoder*, respectively, as shown in Figure 1.1. In such a framework, this dissertation explores the following questions:

i) Which is the structure of the optimal sampling mechanism?

ii) What are the optimal compression and reconstruction algorithms?

iii) What are the tradeoffs – under optimal processing – among sampling procedure, inferential methods for approximating the underlying distribution of the memoryless sources, compression rate and distortion level?

iv) How does a knowledge of the sampling sets at the decoder influence the tradeoff between compression rate and distortion level?

1.1.1 Motivating applications

The questions above are motivated by various applications. An instance arises in hardware "dynamic thermal management," which is the process of controlling surges in the operating temperature of a multicore processor chip during runtime, based on measurements by a limited number of on-chip thermal sensors. Strategic sensor placement and processing are needed to estimate temperatures at grid points over the entire chip; compression is essential to process the continuous measurements in real time.

In a multicore processor, the modes of operation are few and the underlying probability distribution of the temperatures on the processor can (approximately) be estimated *a priori*; however, this is not always possible. In an Internet-of-Things (IoT) setup, multiple networked smart devices/sensors such as thermostats, motion detectors, listening speakers, etc., are often placed in diverse settings and the joint statistics of their measurements are uncertain as they rely on the location and the nature of varying activities at the location. In smart homes, a central hub, e.g., Amazon Echo or Google Home, uses sensor communication to form representations of the signals at various points of the house, for instigating subsequent actions. A large number of sensor measurements makes compression essential. Also, power and bandwidth limitations impose further restrictions on the number of sensors communicating with the hub at any given time, necessitating spatial sampling.

Satellite images are often used to make inferences for tasks which rely on pictures of wide regions of the earth. Many organizations such as Orbital Insight, Descartes Labs and the United Nations often use these images for varied tasks such as analyzing forest cover in the Amazon basin, analyzing traffic patterns in cities and predicting agricultural harvest output. Any particular satellite provides images of only portions of area of interest at any time and the high resolution of satellite images mandate compression for efficient storage and processing; learning and making inference at regions not seen in any image is a particularly important challenge. Similar problems arise in "in-network computation" [1], [2] in which a subset of a network of collocated sensors use only their own measurements to estimate an aggregate function of the entirety of distributed and correlated measurements, e.g., overall average parameter values in environmental monitoring. In such settings, the mechanisms for (spatial) sampling, compression and estimation are collocated, with the latter two being aware of the sampler realizations.

1.1.2 Prior work

The study of problems of combined sampling and compression has a rich and varied history in diverse contexts. Highlights include: classical sampling and processing, rate distortion theory, multiterminal source coding, wavelet-based compression, and compressed sensing, among others. Rate distortion theory [3] rules the compression of a given sampled signal and its reconstruction within a specified distortion level. On the other hand, compressed sensing [4] provides a random linear encoding of nonprobabilistic analog sources marked by a sparse support, with lossless recovery as measured by a block error probability (with respect to the distribution of the encoder). Upon placing the problem of lossless source coding of analog sources in an information theoretic setting, with a probabilistic model for the source that need not be encoded linearly, Rényi dimension is known to determine fundamental performance limits [5] (see also [6, 7]). Several recent studies consider the compressed sensing of a signal with an allowed detection error rate or quantization distortion [8,9]; of multiple signals followed by distributed quantization [10], including a study of scaling laws [11]; or of sub-Nyquist rate sampled signals followed by lossy reconstruction [12]; and rate distortion function for multiple sources with time-shared sampling [13].

Closer to our approach that entails *spatial* sampling, the rate distortion function has been characterized when multiple Gaussian signals from a random field are sampled and quantized (centralized or distributed) in [14]. In a setting of distributed acoustic sensing and reconstruction, centralized as well as distributed coding schemes and sampling lattices are studied in [15]. In [16], a Gaussian random field on the interval $[0, 1]$ and i.i.d. in time, is reconstructed from compressed versions of k sampled sequences under a mean-squared error distortion criterion; the rate distortion function is studied for schemes which reconstruct only the sampled source first and then reconstruct the unsampled source by forming minimum mean-squared error (MMSE) estimates based on the reconstructions for the sampled source. All the sampling problems above assume a knowledge of the underlying distribution.

In the realm of rate distortion theory where a complete knowledge of the source statistics is unknown, there is a rich literature that considers various formulations of universal coding; only a sampling is listed here. Directions include classical Bayesian and nonBayesian methods [17–20]; "individual sequences" studies [21–23]; redundancy in quantization rate or distortion [24–26]; and lossy compression of noisy or remote sources [27–29]. These works propose a variety of distortion measures to investigate universal reconstruction performance.

1.2 Main Contributions

The focus of this dissertation is on studying optimal mechanisms and tradeoffs involved in reconstructing a prespecified subset of components of a memoryless multiple source from the centralized processing of its spatially sampled components. While aspects of this have been explored in some of the works mentioned above, in this dissertation we present our framework of *sampling rate distortion* to study the tradeoffs among sampling mechanism, compression rate and desired level of accuracy in the reconstruction, in an unified manner. Throughout this dissertation, sampling is spatial rather than temporal, unlike in most of the settings above. We introduce new forms of randomized sampling mechanisms that can depend on the observed source realizations. Such randomized sampling mechanisms, albeit of increased complexity, are shown to yield clear gains in performance.

1.2.1 Finite-Valued Multiple Source

We begin our study with the joint compression and reconstruction of a spatially sampled finite-valued memoryless multiple source (referred to as a *discrete* memoryless multiple source in this dissertation) whose underlying probability mass function (pmf) is taken to be known in Chapter 2; in this chapter, we present our framework of sampling rate distortion and introduce several randomized sampling mechanisms. The models and ideas presented in this chapter will form the bedrock of the chapters to follow.

Our formulation involves the notion of a *sampling rate distortion function* (SRDf), which captures the tradeoffs among sampling mechanism, compression rate and desired level of accuracy in the reconstruction. As a basic ingredient, the sampling rate distortion function is characterized first for a fixed sampling set of size $k \leq m$. This characterization is a consequence of prior work by Dobrushin-Tsybakov [30] (see also Berger [3], [31] and Yamamoto-Itoh [32]) on the rate distortion function for a "remote" source-receiver model in which the encoder and receiver lack direct access to the source and decoder outputs, respectively. For the special case of the probability of error distortion criterion, we show that optimal compression and reconstruction can be simplified to a rate distortion code for the sampled source components followed by maximum a posteriori estimation of the remaining source components.

Best fixed-set sampling can be strictly inferior to random sampling. Considering an independent random sampler, in which sampling does not depend on the source realizations and is independent (but not necessarily identically distributed) in time, we show that the corresponding SRDf remains the same regardless of whether or not the decoder is provided information regarding the sequence of sampled sets. This surprising property does not hold for any causal sampler[1], in general. Next, we consider a generalization, namely a memoryless random sampler whose output can depend on the instantaneous source realizations. The associated formula for SRDf is

[1] A casual sampler is one for which sampling at any instant depends only on the current and past DMMS realizations and past actions of the sampler.

used now to study the structural characteristics of the optimal sampler. Specifically, we show that the optimal sampler is characterized by a conditional point-mass; this has the obvious benefit of a reduction in the search space for an optimal sampler. We also show that such a memoryless sampler can outperform strictly a random sampler that lacks access to source realizations. Finally, in a setting in which the decoder is unaware of the sampled sequence, an upper bound for the SRDf is seen to have an optimal conditional point-mass sampler.

In Chapter 3 we consider an universal setting where the underlying pmf of a discrete memoryless multiple source is known only to belong to a *finite* family of pmfs. Building on the concept of an SRDf, in Bayesian and nonBayesian settings we consider an *universal sampling rate distortion function* (USRDf) which captures the inherent tradeoffs among sampling mechanism, approximation of underlying (unknown) pmf, lossy compression rate and distortion level. Our characterizations of USRDf in this chapter build on the study of sampling rate distortion function in Chapter 2. We begin, once again, with fixed-set sampling where the encoder observes the same set of k components of a discrete memoryless multiple source at every time instant. Recognizing that only the k-marginal pmf of the source – corresponding to the sampling set – can be learned by the encoder, the associated USRDf is characterized. In general, allowing randomization in sampling affords two distinct advantages over fixed-set sampling: better approximation of the underlying joint pmf and improved compression performance enabled by sampling different subsets of the source in apposite proportions. An independent random sampler chooses different k-subsets of the multiple source independently of source realizations and independently in time, and can learn all k-marginals of the joint pmf. This reduction in pmf uncertainty (vis-à-vis fixed-set sampling) aids in improving USRDf. Lastly, we consider a memoryless random sampler whose choice of sampling sets can depend on instantaneous source realizations. Surprisingly, this latitude allows the encoder to learn the entire joint pmf, and that too only from the sampling sequence without recourse to the sampled source realizations. Furthermore, we show how the USRDf can be attained by means of a sampling sequence that depends *deterministically* on source realizations thereby reducing overall complexity. Thus, all our achievability proofs in this chapter bring out new ideas for joint source pmf-learning and lossy compression.

1.2.2 Gaussian Multiple Source

In Chapter 4, we consider a jointly Gaussian memoryless multiple source (GMMS) with joint probability density function (pdf) known only to belong to a given family of *uncountable* pdfs (characterized by a convex compact in the Euclidean space). In this chapter, our objective is to reconstruct all the components of a GMMS from the compressed representations of fixed k components under a suitable mean-squared error (MSE) distortion criterion. In Bayesian and nonBayesian settings, we extend the notion of USRDf to a GMMS. We consider first the setting where the underlying pdf of the GMMS is known and, building on the ideas developed in Chapter 2, characterize its SRDf. We also show that a two-step procedure for reconstructing all the components of a GMMS is optimal: the sampled components of the GMMS are reconstructed first under a modified weighted MSE distortion criterion and then MMSE estimates are formed for the unsampled components based on the reconstructions of the former. This is akin to the structure observed in Chapter 2 for a discrete memoryless multiple source reconstructed under the probability of error distortion criterion and in [33] for reconstructing remote Gaussian sources under MSE distortion criterion. Considering a Gaussian memoryless field on an interval sampled at a finite number of points, we characterize its SRDf and illustrate by example the structure of the best fixed-set sampling. Building on the ideas developed for the SRDf, we characterize next the USRDf for a GMMS in the Bayesian and non-Bayesian settings and show that it remains optimal to reconstruct first the sampled components of the GMMS and then form estimates for the unsampled components based on the former, thereby simplifying the reconstruction procedure.

1.3 Organization of Dissertation

In Chapter 2, we introduce our framework of sampling rate distortion and begin with a study of discrete memoryless multiple source with known pmf. The subsequent chapters in this dissertation will build on the models and the framework of sampling rate distortion introduced in this chapter.

Building on the ideas developed in Chapter 2, in Chapter 3 we study a setting where the pmf of the discrete memoryless multiple source is known only to belong to a finite family of pmfs. Here, we discuss procedures to estimate pmf from the sampled source measurements and the optimal tradeoffs arising among sampling mechanism, estimation, compression and reconstruction accuracy. In Chapter 4, we consider a Gaussian memoryless multiple source with joint pdf known only to belong to an uncountable family of pdfs.

In each chapter, we present first our main results followed by proofs of the main results. We present the achievability proofs in an increasing order of sampler complexity; a unified converse proof is presented thereafter. Throughout this dissertation, in the universal setting the emphasis of our achievability proofs is on the Bayesian formulation.

Chapter 2

SRDf for Finite-Valued Multiple Source

2.1 Synopsis

In this chapter, we consider a discrete memoryless multiple source (DMMS) with m components and of known joint *pmf*. A subset of k components of the DMMS are sampled (possibly in a randomized manner) at each time instant and compressed jointly with the objective of reconstructing a subset of DMMS components with indices in an arbitrary but fixed *recovery* set.

Our model is described in Section 2.2. We introduce several sampling mechanisms and the notion of sampling rate distortion function (SRDf) to characterize the tradeoffs among sampling procedure, compression rate and desired level of accuracy in reconstruction. The main results, along with examples, are stated in Section 2.3. Considering the sampling mechanisms introduced in Section 2.2 in an increasing order of complexity, we provide single-letter characterizations for the SRDf. Concomitant improvements in performance thereby become evident. Section 2.4 contains the proofs. Presented first are the achievability proofs that are built successively in the order of increasing complexity of the sampling mechanism. The converse proofs follow in reverse order in a unified manner.

2.2 Model

Let $\mathcal{M} = \{1, \ldots, m\}$ and $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ be a $\mathcal{X}_{\mathcal{M}} = \bigtimes_{i=1}^{m} \mathcal{X}_i$-valued rv where each $\mathcal{X}_i$ is a finite set. It will be convenient to use the following compact notation. For a nonempty set $A \subseteq \mathcal{M}$, we denote by X_A the rv $(X_i, i \in A)$ with values in $\bigtimes_{i \in A} \mathcal{X}_i$, and denote n repetitions of X_A by $X_A^n = (X_i^n, i \in A)$ with values in $\mathcal{X}_A^n = \bigtimes_{i \in A} \mathcal{X}_i^n$, where $X_i^n = (X_{i1}, \ldots, X_{in})$ takes values in the n-fold product space $\mathcal{X}_i^n = \mathcal{X}_i \times \cdots \times \mathcal{X}_i$. For $1 \leq k \leq m$, let $\mathcal{A}_k = \{A : A \subseteq \mathcal{M},\ |A| = k\}$ be the set of all k-sized subsets of $\mathcal{M}$ and let $A^c = \mathcal{M} \setminus A$. All logarithms and exponentiations are with respect to the base 2.

Consider a discrete memoryless multiple source (DMMS) $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ consisting of i.i.d. repetitions of the rv $X_{\mathcal{M}}$ with given pmf $P_{X_{\mathcal{M}}}$ of assumed full support $\mathcal{X}_{\mathcal{M}}$. Let $\mathcal{Y}_{\mathcal{M}} = \bigtimes_{i=1}^{m} \mathcal{Y}_i$, where $\mathcal{Y}_i$ is a finite reproduction alphabet for X_i.

Definition 2.1 *A k-random sampler (k-RS), $1 \leq k \leq m$, collects causally at each $t = 1, \ldots, n$, random samples*[†] *X_{S_t} from $X_{\mathcal{M}t}$, where S_t is a rv with values in $\mathcal{A}_k$ with (conditional) pmf $P_{S_t|X_{\mathcal{M}}^t S^{t-1}}$, with $X_{\mathcal{M}}^t = (X_{\mathcal{M}1}, \ldots, X_{\mathcal{M}t})$ and $S^{t-1} = (S_1, \ldots, S_{t-1})$. Such a k-RS is specified by a (conditional) pmf $P_{S^n|X_{\mathcal{M}}^n}$ with the requirement*

$$P_{S^n|X_{\mathcal{M}}^n} = \prod_{t=1}^{n} P_{S_t|X_{\mathcal{M}}^t S^{t-1}}. \tag{2.1}$$

The output of a k-RS is (S^n, X_S^n) where $X_S^n = (X_{S_1}, \ldots, X_{S_n})$. Successively restric-

[†] With an abuse of notation, we write $X_{S_t t}$ simply as X_{S_t}.

tive choices of a k-RS in (2.1) corresponding to

$$P_{S_t|X^t_{\mathcal{M}}S^{t-1}} = P_{S_t|X_{\mathcal{M}t}}, \quad t = 1, \ldots, n, \tag{2.2}$$

$$P_{S_t|X^t_{\mathcal{M}}S^{t-1}} = P_{S_t}, \qquad t = 1, \ldots, n, \tag{2.3}$$

and, for a given $A \subseteq \mathcal{M}$,

$$P_{S_t|X^t_{\mathcal{M}}S^{t-1}} = \mathbb{1}(S_t = A), \qquad t = 1, \ldots, n \tag{2.4}$$

will be termed the k-memoryless random sampler, k-independent random sampler and the k-fixed-set sampler abbreviated as k-MRS, k-IRS and k-FS, respectively.

Our objective is to reconstruct a subset of DMMS components with indices in an arbitrary but fixed *recovery set* $B \subseteq \mathcal{M}$, namely X^n_B, from a compressed representation of the k-RS output (S^n, X^n_S) under a suitable distortion criterion.

Definition 2.2 *For* $n \geq 1$, *an* n*-length block code with k-RS for a DMMS* $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ *with alphabet* $\mathcal{X}_{\mathcal{M}}$ *and reproduction alphabet* $\mathcal{Y}_B$ *is a triple* $(P_{S^n|X^n_{\mathcal{M}}}, f_n, \varphi_n)$ *where* $P_{S^n|X^n_{\mathcal{M}}}$ *is a k-RS as in* (2.1), *and* (f_n, φ_n) *are a pair of mappings where the encoder* f_n *maps the k-RS output* (S^n, X^n_S) *into some finite set* $\mathcal{J} = \{1, \ldots, J\}$ *and the decoder* φ_n *maps* $\mathcal{J}$ *into* $\mathcal{Y}^n_B$. *We shall use the compact notation* $(P_{S|X_{\mathcal{M}}}, f, \varphi)$, *suppressing* n. *The rate of the code with k-RS* $(P_{S|X_{\mathcal{M}}}, f, \varphi)$ *is* $\frac{1}{n} \log J$.

Remark: An encoder that uses a deterministic estimate of $X^n_{S^c}$ from (S^n, X^n_S) in its operation is subsumed by the definition above.

For a given (single-letter) finite-valued distortion measure $d : \mathcal{X}_B \times \mathcal{Y}_B \to \mathbb{R}^+ \cup \{0\}$, an n-length block code with k-RS $(P_{S|X_{\mathcal{M}}}, f, \varphi)$ will be required to satisfy the expected fidelity criterion (d, Δ), i.e.,

$$\mathbb{E}\left[d\left(X_B^n, \varphi\big(f(S^n, X_S^n)\big)\right)\right] \triangleq \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n} d\left(X_{Bt}, \Big(\varphi\big(f(S^n, X_S^n)\big)\Big)_t\right)\right] \leq \Delta. \quad (2.5)$$

We shall consider also the case where the decoder is informed of the sequence of sampled sets S^n. Denoting such an *informed decoder* by φ_S, the expected fidelity criterion (2.5) will use the augmented $\varphi_S\big(S^n, f(S^n, X_S^n)\big)$ instead of $\varphi\big(f(S^n, X_S^n)\big)$. The earlier decoder (that is not informed of S^n) will be termed an *uninformed decoder.*

Definition 2.3 *A number $R \geq 0$ is an achievable k-sample coding rate at average distortion level Δ if for every $\epsilon > 0$ and sufficiently large n, there exist n-length block codes with k-RS of rate less than $R + \epsilon$ and satisfying the expected fidelity criterion $(d, \Delta + \epsilon)$; and (R, Δ) will be termed an achievable k-sample rate distortion pair. The infimum of such achievable rates is denoted by $R^I(\Delta)$ for an informed decoder, and by $R^U(\Delta)$ for an uninformed decoder. We shall refer to $R^I(\Delta)$ as well as $R^U(\Delta)$ as the sampling rate distortion function (SRDf), suppressing the dependence on k.*

Remarks: (i) An informed decoder obtains information regarding X_B^n in two ways: from the encoder output as well as from S^n (as a k-RS can embed information about

$X^n_{\mathcal{M}}$ in S^n).

(ii) Clearly, $R^I(\Delta) \leq R^U(\Delta)$, and both are nonincreasing in k.

(iii) For a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$, the requirement (2.2) on the sampler renders $\{(X_{\mathcal{M}t}, S_t)\}_{t=1}^{\infty}$ and thereby also $\{(X_{S_t}, S_t)\}_{t=1}^{\infty}$ to be memoryless sequences.

2.3 Sampling Rate Distortion function

We shall state our results in an order of increasing complexity of the sampling mechanism. Single-letter characterizations of the SRDfs in this chapter involve, as an ingredient, a characterization of $R^I(\Delta)$ for a fixed-set sampler, which in turn is based on [30], [3]. For a k-FS, denote the corresponding $R^I(\Delta)$ by $R_A(\Delta)$ (with an abuse of notation).

2.3.1 Fixed-Set Sampling

A fixed-set sampler chooses the same k-sized subset $A \subseteq \mathcal{M}$ of the components of DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ at each time instant. Our first result shows that the fixed-set SRDf $R_A(\Delta)$, in effect, is the (standard) rate distortion function for the DMMS $\{X_{At}\}_{t=1}^{\infty}$ using a modified distortion measure $d_A : \mathcal{X}_A \times \mathcal{Y}_B \to \mathbb{R}^+ \cup \{0\}$ defined by

$$d_A(x_A, y_B) = \mathbb{E}[d(X_B, y_B)|X_A = x_A]. \tag{2.6}$$

Proposition 2.1 *For a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with pmf $P_{X_{\mathcal{M}}}$, the fixed-set SRDf for*

$A \subseteq \mathcal{M}$ *is*

$$R_A(\Delta) = \min_{\substack{X_{A^c} \multimap X_A \multimap Y_B \\ \mathbb{E}[d(X_B, Y_B)] \leq \Delta}} I\big(X_A \wedge Y_B\big) \tag{2.7}$$

for $\Delta_{\min,A} \leq \Delta \leq \Delta_{\max}$, *and equals* 0 *for* $\Delta \geq \Delta_{\max}$, *where*

$$\begin{aligned} \Delta_{\min,A} &= \mathbb{E}\left[\min_{y_B \in \mathcal{Y}_B} d_A(X_A, y_B)\right], \\ \Delta_{\max} &= \min_{y_B \in \mathcal{Y}_B} \mathbb{E}\big[d(X_{\mathcal{M}}, y_B)\big] = \min_{y_B \in \mathcal{Y}_B} \mathbb{E}[d_A(X_A, y_B)]. \end{aligned} \tag{2.8}$$

Corollary 2.1 *With* $A \subseteq B$ *and* $\mathcal{X}_B = \mathcal{Y}_B$, *for the probability of error distortion measure*

$$d(x_B, y_B) = \mathbb{1}(x_B \neq y_B) = 1 - \prod_{i \in B} \mathbb{1}(x_i = y_i), \quad x_B, y_B \in \mathcal{X}_B$$

we have

$$R_A(\Delta) = \begin{cases} \min I\big(X_A \wedge Y_A\big), & \Delta_{\min} \leq \Delta \leq \Delta_{\max} \\ 0, & \Delta \geq \Delta_{\max}, \end{cases} \tag{2.9}$$

where the minimum in (2.9) *is subject to*

$$\mathbb{E}[\alpha(X_A)\mathbb{1}(X_A \neq Y_A)] \leq \Delta - \Delta_{\min} \tag{2.10}$$

with

$$\alpha(x_A) = \max_{\tilde{x}_B \in \mathcal{X}_B} P_{X_B|X_A}(\tilde{x}_B|x_A)$$

and

$$\Delta_{\min} = 1 - \mathbb{E}[\alpha(X_A)], \qquad \Delta_{\max} = 1 - \max_{x_B \in \mathcal{X}_B} P_{X_B}(x_B).$$

Remarks: (i) In (2.10), the overall reconstruction error Δ can be seen to comprise two parts: i) minimum error introduced due the sampling, $\Delta_{\min}$, and ii) the error introduced in compression, $\mathbb{E}[\alpha(X_A)\mathbb{1}(X_A \neq Y_A)]$.

(ii) The minimum in (2.7) exists by virtue of the continuity in $P_{X_{\mathcal{M}}Y_B}$ of $I(X_A \wedge Y_B)$ over the compact set $\{P_{X_{\mathcal{M}}Y_B} : X_{A^c} -\!\!\circ\!\!- X_A -\!\!\circ\!\!- Y_B,\ \mathbb{E}[d(X_B, Y_B)] \leq \Delta\}$.

(iii) The corollary relies on showing that the minimum in (2.7) is attained in this particular instance by a pmf $P_{X_{\mathcal{M}}Y_B}$ under a longer Markov chain

$$X_{A^c} -\!\!\circ\!\!- X_A -\!\!\circ\!\!- Y_A -\!\!\circ\!\!- Y_{B \setminus A}.$$

Interestingly, the achievability proof entails in a first step a mapping of x_A^n in $\mathcal{X}_A^n$ into its codeword y_A^n from which, in a second step, a reconstruction $y_{B\setminus A}^n$ of $x_{B\setminus A}^n$ is obtained as a maximum a posteriori (MAP) estimate.

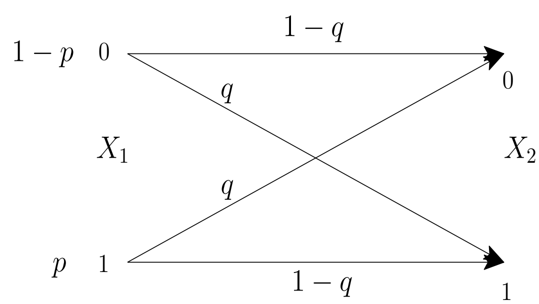

Figure 2.1: BSC (q)

The following example, albeit concocted, shows that for fixed-set sampling with A and recovery set B, a choice of A outside B can be best.

Example 2.1 *Let* $\mathcal{M} = \{1, 2, 3\}$, $B = \{1, 2\}$ *and* $\mathcal{X}_i = \mathcal{Y}_j = \{0, 1\}$, $i = 1, 2, 3; j = 1, 2$. *Consider a DMMS with* $P_{X_1 X_2}$ *as in Figure 2.1 and* $X_3 = X_1 \oplus X_2$ *where* $\oplus$ *denotes addition modulo 2. Here,* $p = 0.5$, $q \leq 0.5$. *For distortion measure* $d(x_B, y_B) \triangleq \mathbb{1}\left((x_1 \oplus x_2) \neq (y_1 \oplus y_2)\right)$, *the SRDf for fixed-set sampling is*

$$R_{\{1\}}(\Delta) = h(0.5) - h\Big(\frac{\Delta - q}{1 - 2q}\Big), \quad q \leq \Delta \leq 0.5,$$

where $h(\cdot)$ *is the binary entropy function. Also,* $R_{\{1\}}(\Delta) = R_{\{2\}}(\Delta)$. *For sampling set* $A = \{3\}$, *the SRDf is*

$$R_{\{3\}}(\Delta) = h(q) - h\left(\Delta\right), \quad 0 \leq \Delta \leq q.$$

Clearly, $R_{\{3\}}(\Delta) \leq R_{\{1\}}(\Delta)$, *with the inequality being strict for suitable values of* Δ.

■

2.3.2 Independent Random Sampling

The k-IRS affords a more capable mechanism than the fixed-set sampler of Proposition 2.1, with the sampling sets possibly varying in time. Surprisingly, our next result shows that the SRDf for a k-IRS, displayed as $R_\imath(\Delta)$, remains the same regardless of whether or not the decoder is provided information regarding the sequence of sampled sets. As seen from its proof in Section 2.4, this is enabled by the lack of dependence of the sampling sequence on the DMMS realization.

Theorem 2.1 *For a k-IRS,*

$$R_\imath^I(\Delta) = R_\imath^U(\Delta) = R_\imath(\Delta) = \min I\big(X_S \wedge Y_B|S\big) \tag{2.11}$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$*, where the minimum is with respect to* $P_{X_{\mathcal{M}}SY_B} = P_{X_{\mathcal{M}}}P_S P_{Y_B|SX_S}$ *and* $\mathbb{E}[d(X_B, Y_B)] \leq \Delta$*, with*

$$\Delta_{\min} = \min_{A \in \mathcal{A}_k} \mathbb{E}\left[\min_{y_B \in \mathcal{Y}_B} d_A(X_A, y_B)\right]$$

and $\Delta_{\max}$ *as in* (2.8)*.*

Clearly, for every $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, $R_\imath^U(\Delta)$ is no smaller than $R_\imath(\Delta)$. In Section 2.4, we provide an achievability proof for $R_\imath(\Delta)$ for a k-IRS with an uninformed decoder and a converse proof for a k-IRS with an informed decoder, thus showing the equivalence of $R_\imath^I(\Delta)$ and $R_\imath^U(\Delta)$. Our achievability proof relies on a "deterministic" implementation of the optimal k-IRS, obviating the need for a decoder to be explicitly informed of the random sampling sets.

A convenient equivalent expression for $R_\imath(\Delta)$ in (2.11) is given by

Proposition 2.2 *For a k-IRS,*

$$R_\imath(\Delta) = \min \sum_{A \in \mathcal{A}_k} P_S(A) R_A(\Delta_A), \quad \Delta_{\min} \leq \Delta \leq \Delta_{\max}, \tag{2.12}$$

where the minimum is with respect to

$$P_S, \ \left\{\Delta_A \geq \Delta_{\min,A}, \ A \in \mathcal{A}_k : \sum_{A \in \mathcal{A}_k} P_S(A)\Delta_A \leq \Delta\right\}. \tag{2.13}$$

Proof. For every $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, in (2.11),

$$\begin{aligned}
&\min_{\substack{P_{X_{\mathcal{M}}} P_S P_{Y_B|SX_S} \\ \mathbb{E}[d(X_B,Y_B)] \leq \Delta}} I(X_S \wedge Y_B|S) \\
&= \min_{\substack{P_{X_{\mathcal{M}}} P_S P_{Y_B|SX_S} \\ \mathbb{E}[d(X_B,Y_B)] \leq \Delta}} \sum_{A \in \mathcal{A}_k} P_S(A) I(X_A \wedge Y_B|S = A) \\
&= \min_{\substack{P_S, \ \Delta_A: \\ \sum_{A \in \mathcal{A}_k} P_S(A)\Delta_A \leq \Delta}} \sum_{A \in \mathcal{A}_k} P_S(A) \min_{\substack{P_{Y_B|S=A,X_A} \\ \mathbb{E}[d(X_B,Y_B)|S=A]=\Delta_A}} I(X_A \wedge Y_B|S = A) \qquad (2.14) \\
&= \min_{\substack{P_S, \ \Delta_A: \\ \sum_{A \in \mathcal{A}_k} P_S(A)\Delta_A \leq \Delta}} \sum_{A \in \mathcal{A}_k} P_S(A) R_A(\Delta_A),
\end{aligned}$$

where $P_{Y_B|S=A,X_A}$ is used to denote $P_{Y_B|S,X_S}(\cdot|A,\cdot)$ for compactness. The validity of (2.14) follows by the introduction of the Δ_As and observing that the order of the minimization does not alter the value of the minimum. The last step obtains upon noting that the value of the inner minimum in (2.14) is the same upon replacing the equality in $\mathbb{E}[d(X_B, Y_B)|S = A] = \Delta_A$ with "$\leq$". ■

Remark: By Proposition 2.2, the SRDf for a k-IRS is the lower convex envelope of the set of SRDfs $\{R_A(\Delta), \ A \in \mathcal{A}_k\}$ and thus is convex in $\Delta \geq \Delta_{\min}$. Furthermore,

$$R_\imath(\Delta) \leq \min_{A \in \mathcal{A}_k} R_A(\Delta).$$

Additionally, a k-IRS can outperform strictly the best fixed-set sampler. For instance, if there is no fixed-set SRDf for any $A \in \mathcal{A}_k$ that is uniformly best for all Δ, then the previous inequality can be strict. This is illustrated by the following example.

Example 2.2 *With $\mathcal{M} = B = \{1, 2\}$, $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y}_2 = \{0, 1\}$, and $\mathcal{Y}_1 = \{0, 1, e\}$, let X_1, X_2 be i.i.d. Bernoulli(0.5) rvs, and*

$$d\big((x_1, x_2), (y_1, y_2)\big) = d_1(x_1, y_1) + d_2(x_2, y_2)$$

with

$$d_1(x_1, y_1) = \begin{cases} 0, & \textit{if } x_1 = y_1 = 0;\ x_1 = y_1 = 1 \\ 1, & \textit{if } x_1 = 0, 1, \quad y_1 = e \\ \infty, & \textit{if } x_1 = 0,\ y_1 = 1;\ x_1 = 1,\ y_1 = 0, \end{cases}$$

$$d_2(x_2, y_2) = \mathbb{1}(x_2 \neq y_2).$$

For $k = 1$,

$$R_{\{1\}}(\Delta) = 1.5 - \Delta, \qquad 0.5 \leq \Delta \leq 1.5,$$

$$R_{\{2\}}(\Delta) = 1 - h(\Delta - 1), \qquad 1 \leq \Delta \leq 1.5$$

whereas

$$R_\imath(\Delta) = \begin{cases} 1.5515 - 1.103\Delta, & 0.5 \le \Delta \le 1.318, \\ 1 - h(\Delta - 1), & 1.318 \le \Delta \le 1.5. \end{cases}$$

Clearly, $R_\imath(\Delta)$ *is strictly smaller than* $\min\left\{R_{\{1\}}(\Delta), R_{\{2\}}(\Delta)\right\}$ *for* $0.5 < \Delta < 1.318$; *see Fig. 2.2. Note that while the distortion measure* d *in Definition 2.2 is taken to be finite-valued, the event* $\{d(X_1, Y_1) = \infty\}$ *above is accommodated by assigning (optimally) zero probability to it.* ■

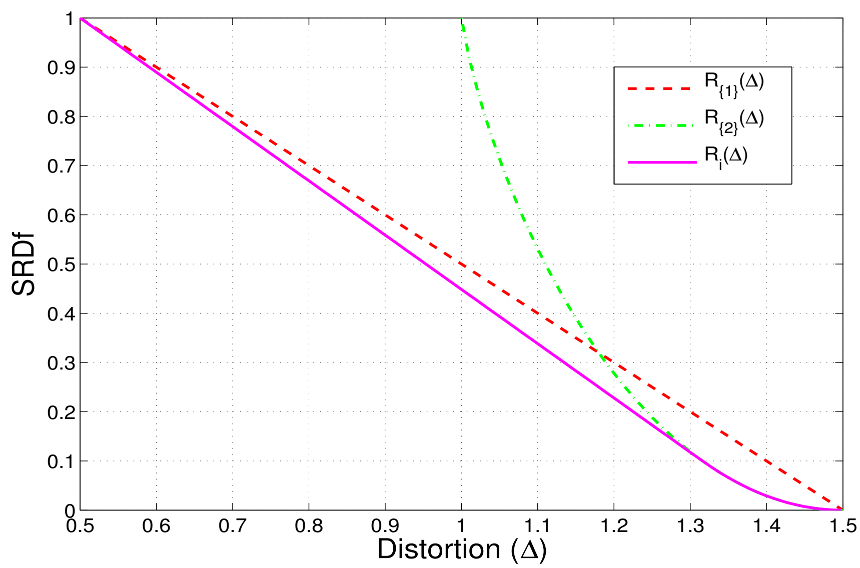

Figure 2.2: SRDfs for k-IRS vs. fixed-set sampler

2.3.3 Memoryless Random Sampling

A k-MRS is more powerful than a k-IRS in that sampling with the former at each time instant can depend on the current DMMS realization. The SRDf for a k-MRS can improve with an informed decoder unlike for a k-IRS.

Theorem 2.2 *For a k-MRS with informed decoder, the SRDf is*

$$R_m^I(\Delta) = \min I\big(X_S \wedge Y_B | S, U\big) \tag{2.15}$$

for $\Delta_{\min} \le \Delta \le \Delta_{\max}$*, where the minimum is with respect to* $P_{UX_{\mathcal{M}}SY_B} = P_U P_{X_{\mathcal{M}}} P_{S|X_{\mathcal{M}}U} P_{Y_B|SX_SU}$ *and* $\mathbb{E}[d(X_B, Y_B)] \le \Delta$*, with*

$$\Delta_{\min} = \min_{P_{S|X_{\mathcal{M}}}} \mathbb{E}\left[\min_{P_{Y_B|SX_S}} \mathbb{E}\big[d(X_B, Y_B)|S, X_S\big]\right], \tag{2.16}$$

$$\Delta_{\max} = \min_{P_{S|X_{\mathcal{M}}}} \mathbb{E}\left[\min_{y_B \in \mathcal{Y}_B} \mathbb{E}\big[d(X_B, y_B)|S\big]\right], \tag{2.17}$$

and U *being a* $\mathcal{U}$*-valued rv with* $|\mathcal{U}| \le 3$.

Remark: Analogously as in Proposition 2.2, the SRDf $R_m^I(\Delta)$ can be expressed as

$$R_m^I(\Delta) = \min_{\substack{P_U,\ \Delta_u: \\ \sum_u P_U(u)\Delta_u \le \Delta}} \sum_u P_U(u) \min_{\substack{P_{S|X_{\mathcal{M}},U=u}, P_{Y_B|SX_S,U=u} \\ \mathbb{E}[d(X_B,Y_B)|U=u]=\Delta_u}} I(X_S \wedge Y_B | S, U = u) \tag{2.18}$$

and thereby equals a lower convex envelope of functions of Δ.

The optimal sampler that attains the SRDf in Theorem 2.2 has a simple structure. It is easy to see that each of $\Delta_{\min}$ and $\Delta_{\max}$ in (2.16) and (2.17), respectively, is attained by a sampler for which $P_{S|X_{\mathcal{M}}}$ takes the form of a conditional point-mass. Such conditionally deterministic samplers (defined below), in fact, are optimal for every distortion level $\Delta_{\min} \le \Delta \le \Delta_{\max}$ and will depend on Δ, in general.

Definition 2.4 *For a mapping* $w : \mathcal{X}_{\mathcal{M}} \times \mathcal{U} \rightarrow \mathcal{A}_k$*, a deterministic sampler is specified in terms of a conditional point-mass pmf*

$$\delta_{w(x_{\mathcal{M}},u)}(s) \triangleq \begin{cases} 1, & s = w(x_{\mathcal{M}}, u) \\ 0, & \textit{otherwise}, \end{cases} \tag{2.19}$$

for $(x_{\mathcal{M}}, u) \in \mathcal{X}_{\mathcal{M}} \times \mathcal{U},\ s \in \mathcal{A}_k$.

The following reduction of Theorem 2.2 shows the optimality of deterministic samplers for a k-MRS which will be seen to play a material role in the achievability proof of Theorem 2.2.

Theorem 2.3 *For a k-MRS with informed decoder,*

$$R_m^I(\Delta) = \min I\big(X_S \wedge Y_B | S, U\big) \tag{2.20}$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$*, with* $\Delta_{\min}$ *and* $\Delta_{\max}$ *as in* (2.16) *and* (2.17)*, respectively, where the minimum is with respect to* $P_{UX_{\mathcal{M}}SY_B}$ *of the form* $P_U P_{X_{\mathcal{M}}} \delta_{w(\cdot)} P_{Y_B|SX_SU}$ *with* $\mathbb{E}[d(X_B, Y_B)] \leq \Delta$*, where the (time-sharing) rv* U *takes values in* $\mathcal{U}$ *with* $|\mathcal{U}| \leq 3$.

Proof: See Appendix A.

The structure of the optimal sampler in Theorem 2.3 implies that the search space for minimization now can be reduced to the corner points of the simplexes of the conditional pmfs $P_{S|X_{\mathcal{M}}U}(\cdot|x_{\mathcal{M}}, u),\ (x_{\mathcal{M}}, u) \in \mathcal{X}_{\mathcal{M}} \times \mathcal{U}$. The SRDf in (2.20) is thus the lower convex envelope of the SRDfs for deterministic samplers. In general, time-sharing between such samplers will be seen to achieve the best compression rate for a given distortion level.

While a k-FS and a k-IRS cannot depend on the DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$, a k-MRS can; this enables a k-MRS to embed information about the DMMS in the sampled sequence S^n. For $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, a optimal k-MRS makes a tradeoff between conveying rate-free information to the decoder via S^n and via in the compressed representation of the encoder output. At $\Delta_{\max}$ however, the k-MRS conveys information to the decoder via the sequence of sampling sets alone.

Finally, for a k-MRS with uninformed decoder, we provide an upper bound for the SRDf $R_m^U(\Delta)$.

Theorem 2.4 *For a k-MRS with uninformed decoder,*

$$R_m^U(\Delta) \leq \min I\big(S, X_S \wedge Y_B\big) \tag{2.21}$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, *where the minimum is with respect to* $P_{X_{\mathcal{M}}SY_B} = P_{X_{\mathcal{M}}} P_{S|X_{\mathcal{M}}} P_{Y_B|SX_S}$ *and* $\mathbb{E}[d(X_B, Y_B)] \leq \Delta$, *with* $\Delta_{\min}$ *and* $\Delta_{\max}$ *being as in (2.16) and* (2.8).

Remark: Clearly, when (S, X_S) in (2.21) determines $X_{\mathcal{M}}$, we have $R_m^U(\Delta) = R(\Delta) =$ the (standard) rate distortion function for the DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$.

The (achievability) proof of Theorem 2.4 is along the lines of Proposition 2.1. The lack of a converse is due to the inability to prove or disprove the convexity of the right-side of (2.21) in Δ. (Convexity would imply equality in (2.21).) The optimal sampler can, however, be shown to be a deterministic sampler (2.19) along the lines of Theorem 2.3. Note that the same deterministic sampler need not be the best in (2.15) and (2.21).

Strong forms of the k-MRS and k-IRS are obtained by allowing time-dependence in sampling. Specifically, (2.2) and (2.3) can be strengthened, respectively, to

$$P_{S_t|X^t_{\mathcal{M}}S^{t-1}} = P_{S_t|X_{\mathcal{M}t}S^{t-1}} \tag{2.22}$$

and

$$P_{S_t|X^t_{\mathcal{M}}S^{t-1}} = P_{S_t|S^{t-1}}. \tag{2.23}$$

Surprisingly, this does not improve SRDf for the k-MRS (with decoder informed) or the k-IRS.

Proposition 2.3 *For a strong k-MRS in* (2.22) *and a strong k-IRS in* (2.23)*, the corresponding SRDfs $R^I_{ms}(\Delta)$ and $R_{is}(\Delta)$ equal the right-sides of* (2.15) *and* (2.11)*, respectively.*

Finally, standard properties of the SRDf for the fixed-set sampler, k-IRS and k-MRS with informed decoder are summarized in the following

Lemma 2.1 *For a fixed $P_{X_{\mathcal{M}}}$, the right-sides of* (2.7), (2.11) *and* (2.15) *are finite-valued, nonincreasing, convex, continuous functions of Δ.*

We close this section with an example showing that (i) the SRDf for a k-MRS with informed decoder can be strictly smaller than that of a k-IRS; and (ii) furthermore, unlike for a k-IRS, a k-MRS with informed decoder can outperform strictly that with an uninformed decoder, uniformly for all feasible distortion values.

Example 2.3 *With $\mathcal{M} = B = \{1,2\}$ and $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1\}$, consider a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with $P_{X_1X_2}$ represented by a virtual binary symmetric channel (BSC) shown in Figure 2.1. Fix $p \leq 0.5$ and $q = 0.5$, i.e., X_1 and X_2 are independent. Let d correspond to the probability of error criterion, i.e., $d(x_{\mathcal{M}}, y_{\mathcal{M}}) = \mathbb{1}(x_{\mathcal{M}} \neq y_{\mathcal{M}})$.*

(i) Considering a k-MRS, $k = 1$, with informed decoder, we obtain by Theorem 2.3 that $\Delta_{\min} = 0$, $\Delta_{\max} = p$, and the deterministic sampler

$$P_{S|X_{\mathcal{M}}}(s|x_{\mathcal{M}}) = \begin{cases} 1, & s = 1,\ x_{\mathcal{M}} = 00 \ \textit{or}\ 11 \\ 1, & s = 2,\ x_{\mathcal{M}} = 01 \ \textit{or}\ 10 \\ 0, & \textit{otherwise} \end{cases} \tag{2.24}$$

is uniformly optimal for all $0 \leq \Delta \leq p$, and

$$R_m^I(\Delta) = h(p) - h(\Delta), \quad 0 \leq \Delta \leq p.$$

To obtain $R_\imath(\Delta)$, the SRDfs for fixed-set samplers (2.7) *are*

$$R_{\{1\}}(\Delta) = h(p) - h\left(2\Delta - 1\right), \quad \tfrac{1}{2} \leq \Delta \leq \frac{1+p}{2},$$

and

$$R_{\{2\}}(\Delta) = h\left(\frac{1}{2}\right) - h\left(\frac{\Delta - p}{1-p}\right), \quad p \le \Delta \le \frac{1+p}{2}.$$

Since $R_{\{2\}}(\Delta) \le R_{\{1\}}(\Delta)$ *for all* Δ*, it is a simple exercise to show that*

$$R_\imath(\Delta) = R_{\{2\}}(\Delta).$$

Clearly, $R^I_m(\Delta) \le R_\imath(\Delta)$*, with* $\Delta_{\max}$ *for the former being* $\Delta_{\min}$ *for the latter, as shown in Figure 2.3.*

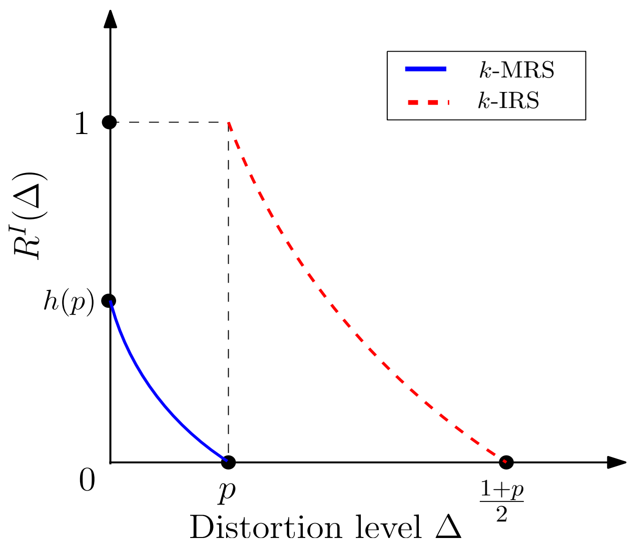

Figure 2.3: SRDf for k-MRS vs. k-IRS

(ii) The conditional pmf $P_{S|X_\mathcal{M}}$ *in* (2.24) *represents a* 1-1 *map between the values of* $X_\mathcal{M}$ *and* (S, X_S)*, and can be seen also to be the optimal choice in the right-side of* (2.21) *for all* $0 \le \Delta \le \frac{1+p}{2}$*. The remaining minimization in* (2.21)*, with respect to* $P_{Y_B|SX_S}$*, renders the right-side to be convex in* Δ*. Consequently, as observed in the passage following Theorem 2.4, the bound in* (2.21) *is tight. For* $p = 0.1$*, the*

resulting values of $R_m^I(\Delta)$ *and* $R_m^U(\Delta)$ *are plotted in Figure 2.4, and of* $R_m^U(\Delta)$ *and* $R_\imath(\Delta)$ *in Figure 2.5. Figure 2.4 illustrates the benefit of decoder information for a k-MRS, while Figure 2.5 shows the compression gain achieved by providing source knowledge to the sampler.*

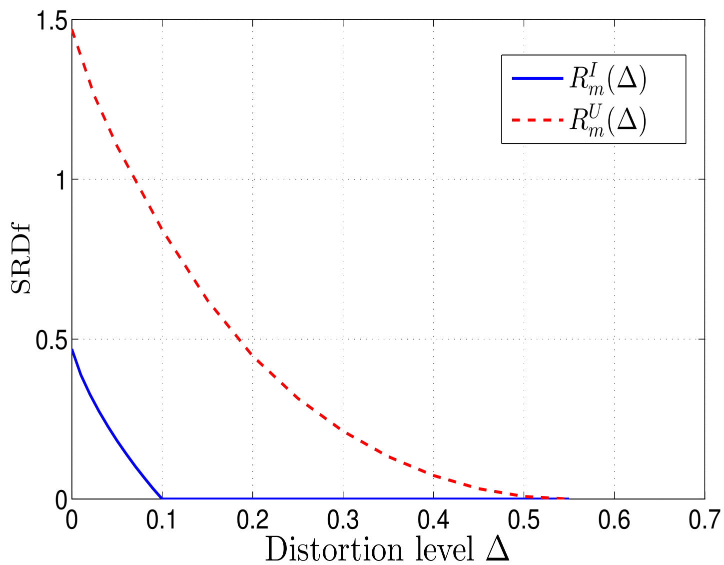

Figure 2.4: SRDf for k-MRS

2.4 Proofs of Main Results

2.4.1 Achievability Proofs

Our achievability proofs successively build upon each other in the order: fixed-set sampler, k-IRS and k-MRS. The achievability proof of Proposition 2.1 for a fixed-set sampler forms a basic building block for subsequent application. Relying on this, the SRDf for a k-IRS is shown to be achieved in Theorem 2.1 without the decoder being informed of the sequence of sampled sets. Next, for a k-MRS with informed decoder, we prove first Theorem 2.3 which shows that the optimal sampler

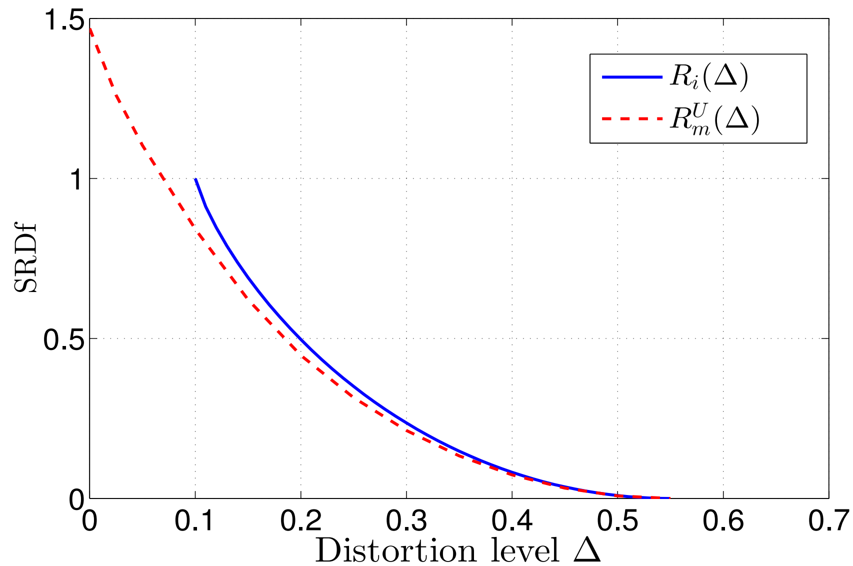

Figure 2.5: $R_m^U(\Delta)$ and $R_\imath(\Delta)$

is conditionally deterministic sampler, i.e., the corresponding $P_{S|X_\mathcal{M}}$ is a point-mass. This structure enables an achievability proof of Theorem 2.2 which builds on that of Proposition 2.1. Lastly, for a k-MRS with uninformed decoder, the achievability proof of Theorem 2.4 rests on the preceding proofs.

Proposition 2.1: The achievability proof below can be obtained directly from [30], but is given here for completeness. Observe first that

$$\begin{aligned}
\Delta_{\min,A} &= \min_{X_{A^c} \multimap\!\!- X_A \multimap\!\!- Y_B} \mathbb{E}[d(X_B, Y_B)] \\
&= \min_{X_{A^c} \multimap\!\!- X_A \multimap\!\!- Y_B} \mathbb{E}\big[\mathbb{E}[d(X_B, Y_B)|X_A]\big] \\
&= \min_{P_{X_A Y_B}} \mathbb{E}[d_A(X_A, Y_B)] \qquad \text{by (2.6) and since } X_{A^c} -\!\!\circ\!\!- X_A -\!\!\circ\!\!- Y_B \\
&= \mathbb{E}\Big[\min_{y_B \in \mathcal{Y}_B} d_A(X_A, y_B)\Big]
\end{aligned}$$

and

$$\Delta_{\max} = \min_{\substack{X_{A^c} -\!\circ\!- X_A -\!\circ\!- Y_B \\ P_{X_A Y_B} = P_{X_A} P_{Y_B}}} \mathbb{E}[d(X_B, Y_B)] = \min_{P_{X_{\mathcal{M}}} P_{Y_B}} \mathbb{E}[d(X_B, Y_B)]$$

$$= \min_{y_B \in \mathcal{Y}_B} \mathbb{E}[d(X_B, y_B)].$$

Next, note that for every $\Delta_{\min,A} \leq \Delta \leq \Delta_{\max}$,

$$\min_{\substack{X_{A^c} -\!\circ\!- X_A -\!\circ\!- Y_B \\ \mathbb{E}[d(X_B, Y_B)] \leq \Delta}} I(X_A \wedge Y_B) \quad = \min_{\mathbb{E}[d_A(X_A, Y_B)] \leq \Delta} I(X_A \wedge Y_B).$$

Clearly every feasible $P_{X_{\mathcal{M}} Y_B} = P_{X_{A^c} X_A Y_B}$ on the left-side above gives a feasible $P_{X_A Y_B}$ on the right-side. Similarly every feasible $P_{X_A Y_B}$ on the right-side leads to a feasible $P_{X_{\mathcal{M}} Y_B}$ on the left-side of the form $P_{X_{\mathcal{M}} Y_B} = P_{X_{A^c}|X_A} P_{X_A Y_B}$.

Given $\epsilon > 0$, consider a (standard) rate distortion code (f, φ) for the DMMS $\{X_{At}\}_{t=1}^{\infty}$ with distortion measure d_A, of rate $\frac{1}{n} \log |f| \leq R_A(\Delta) + \epsilon$ and with expected distortion $\mathbb{E}\left[d_A\big(X_A^n, \varphi(f(X_A^n))\big)\right] \leq \Delta + \epsilon$ for all $n \geq N_A(\epsilon)$, say. Here, $|f|$ denotes the cardinality of the range space of the encoder f.

The code (f, φ) also satisfies

$$\begin{aligned}
\mathbb{E}[d(X_B^n, Y_B^n)] &= \frac{1}{n} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{E}\Big[d\Big(X_{Bt}, \big(\varphi(f(X_A^n))\big)_t\Big)\Big| X_A^n\Big]\right] \\
&= \frac{1}{n} \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{E}\Big[d\Big(X_{Bt}, \big(\varphi(f(X_A^n))\big)_t\Big)\Big| X_{At}\Big]\right] \\
&= \mathbb{E}\left[d_A\Big(X_A^n, \varphi\big(f(X_A^n)\big)\Big)\right] \\
&\leq \Delta + \epsilon,
\end{aligned}$$

thereby yielding achievability in the proposition.

Turning to the corollary, where $A \subseteq B$, for every $P_{X_{\mathcal{M}}Y_B}$ satisfying the constraints in (2.7), consider the pmf $Q_{X_{\mathcal{M}}Y_B}$ defined for $x_{\mathcal{M}} \in \mathcal{X}_{\mathcal{M}},\ y_B \in \mathcal{X}_B$,

$$Q_{X_{\mathcal{M}}Y_B}(x_{\mathcal{M}}, y_B) \triangleq P_{X_{\mathcal{M}}Y_A}(x_{\mathcal{M}}, y_A)\mathbb{1}\big(y_{B\setminus A} = MAP\ (y_A)\big), \tag{2.25}$$

where

$$MAP\ (y_A) = \underset{\tilde{y}_{B\setminus A} \in \mathcal{Y}_{B\setminus A}}{\arg\max}\ P_{X_{B\setminus A}|X_A}(\tilde{y}_{B\setminus A}|y_A) \tag{2.26}$$

is the maximum a posteriori estimate of $y_{B\setminus A}$ given y_A according to $P_{X_{B\setminus A}|X_A}$. Observe that $Q_{X_{\mathcal{M}}Y_B}$ satisfies

$$Q_{X_{A^c}} \multimap Q_{X_A} \multimap Q_{Y_A} \multimap Q_{Y_{B\setminus A}}$$

and

$$\begin{aligned}
\mathbb{E}_P[d(X_B, Y_B)] &= P(X_B \neq Y_B) \\
&= P(X_A \neq Y_A) + P(X_A = Y_A)P(X_{B\setminus A} \neq Y_{B\setminus A}|X_A = Y_A) \\
&= Q(X_A \neq Y_A) + Q(X_A = Y_A)P(X_{B\setminus A} \neq Y_{B\setminus A}|X_A = Y_A) \\
&\geq Q(X_A \neq Y_A) + Q(X_A = Y_A)Q(X_{B\setminus A} \neq Y_{B\setminus A}|X_A = Y_A) \\
&= Q(X_B \neq Y_B) = \mathbb{E}_Q[d(X_B, Y_B)],
\end{aligned}$$

where the inequality is by (2.25), (2.26) and the optimality of the MAP estimator. Also, it is readily checked that

$$\mathbb{E}_Q[d(X_B, Y_B)] \ = 1 - \mathbb{E}[\alpha(X_A)] + \mathbb{E}[\alpha(X_A)\mathbb{1}(X_A \neq Y_A)].$$

Furthermore,

$$\begin{aligned} I_Q(X_A \wedge Y_B) &= I_Q(X_A \wedge Y_A) \\ &= I_P(X_A \wedge Y_A) \\ &\leq I_P(X_A \wedge Y_B). \end{aligned} \tag{2.27}$$

Putting together (2.25) - (2.27) and comparing with (2.7) establishes the corollary.

It is interesting to note that the form of (2.7)

$$\min_{\substack{X_{A^c} -\circ- X_A -\circ- Y_B \\ P(X_B \neq Y_B) \leq \Delta}} I(X_A \wedge Y_B) = \min_{\mathbb{E}[\alpha(X_A)\mathbb{1}(X_A \neq Y_A)] \leq \Delta - (1 - \mathbb{E}[\alpha(X_A)])} I(X_A \wedge Y_A)$$

leads to a simpler and direct proof of achievability of the corollary. Specifically, for a given Δ, first x_A^n is mapped into (only) its corresponding codeword y_A^n but under a modified distortion measure $\tilde{d}(x_A, y_A) \triangleq \alpha(x_A)\mathbb{1}(x_A \neq y_A)$ and a corresponding reduced threshold as indicated by (2.10). Next, the codewords y_A^n serve as sufficient statistics from which (the unsampled) $x_{B\setminus A}^n$ is reconstructed as $y_{B\setminus A}^n = MAP\ (y_A^n)$ under $P_{X_{B\setminus A}^n|X_A^n}$; the corresponding estimation error coincides with the reduction in the threshold.

■

Theorem 2.1: The equivalent expression for $R_\imath(\Delta)$ given by Proposition 2.2 suggests an achievability scheme using a concatenation of fixed-set sampling rate distortion codes from Proposition 2.1. Let P_S and $\{\Delta_A,\ A \in \mathcal{A}_k\}$ yield the minimum in Proposition 2.2. A sequence of sampling sets S^n are constructed *a priori* with $S_t = A$ repeatedly for approximately $nP_S(A)$ time instants, for each A in $\mathcal{A}_k$. Correspondingly, sampling rate distortion codes of blocklength $\cong nP_S(A)$ – with distortion $\cong \Delta_A$ and of rate $\cong R_A(\Delta_A)$ – are concatenated. This predetermined selection of sampling sets does not require the decoder to be additionally informed.

For a fixed $\Delta_{\min} \le \Delta \le \Delta_{\max}$, let P_S and $\{\Delta_A,\ A \in \mathcal{A}_k\}$ attain the minimum in (2.12). Fix $\epsilon > 0$ and $0 < \epsilon' < \epsilon$. Order (in any manner) the elements of $\mathcal{A}_k$ as $A_i,\ i \in \mathcal{M}_k \triangleq \{1, \ldots, M_k\}$, with $M_k = \binom{m}{k}$. For $i \in \mathcal{M}_k$ and $n \ge 1$, define the "time-sets" ν_{A_i} for $A_i \in \mathcal{A}_k$ as

$$\nu_{A_i} = \left\{ t : \lceil n \sum_{j=1}^{i-1} P_S(A_j) \rceil + 1 \le t \le \lceil n \sum_{j=1}^{i} P_S(A_j) \rceil \right\}.$$

The time-sets cover $\{1, \ldots, n\}$, i.e.,

$$\bigcup_{i \in \mathcal{M}_k} \nu_{A_i} = \{1, \ldots, n\}$$

and satisfy

$$\left| \frac{|\nu_{A_i}|}{n} - P_S(A_i) \right| \le \frac{1}{n}, \quad i \in \mathcal{M}_k.$$

Now, a k-IRS is chosen with a deterministic sampling sequence $S^n = s^n$ according to

$$S_t = s_t = A_i, \quad t \in \nu_{A_i}, \; A_i \in \mathcal{A}_k.$$

By Proposition 2.1, for each A_i in $\mathcal{A}_k$, there exists a code (f_{A_i}, φ_{A_i}), $f_{A_i} : \mathcal{X}_{A_i}^{\nu_{A_i}} \to \{1, \ldots, J_{A_i}\}$ and $\varphi_{A_i} : \{1, \ldots, J_{A_i}\} \to \mathcal{Y}_B^{\nu_{A_i}}$ of rate $\frac{1}{|\nu_{A_i}|} \log J_{A_i} \leq R_{A_i}(\Delta_{A_i}) + \frac{\epsilon'}{2}$ and with

$$\begin{aligned} \mathbb{E}\Big[d\Big(X_B^{\nu_{A_i}}, \varphi_{A_i}\big(f_{A_i}(X_{A_i}^{\nu_{A_i}})\big)\Big)\Big] &= \mathbb{E}\left[d_{A_i}\left(X_{A_i}^{\nu_{A_i}}, \varphi_{A_i}\big(f_{A_i}(X_{A_i}^{\nu_{A_i}})\big)\right)\right] \\ &\leq \Delta_{A_i} + \frac{\epsilon'}{2} \end{aligned}$$

for all $|\nu_{A_i}| \geq N_{A_i}\left(\frac{\epsilon'}{2}\right)$ (cf. proof of Proposition 2.1).

Consider a (composite) code (f, φ) as follows. For the deterministic sampling scheme defined above, the encoder f consists of a concatenation of encoders defined for $x^n \in \bigtimes_{i=1}^{M_k} \mathcal{X}_{A_i}^{\nu_{A_i}}$ by

$$f(S^n, x^n) = \left(f_{A_1}\left(x_{A_1}^{\nu_{A_1}}\right), \ldots, f_{A_{M_k}}\left(x_{A_{M_k}}^{\nu_{A_{M_k}}}\right)\right),$$

which maps the output of the k-IRS into the set $\mathcal{J} \triangleq \bigtimes_{i=1}^{M_k} \{1, \ldots, J_{A_i}\}$. The decoder φ is given by

$$\varphi(j_1, \ldots, j_{M_k}) \triangleq \left(\varphi_{A_1}(j_1), \ldots, \varphi_{A_{M_k}}(j_{M_k})\right),$$

for $(j_1, \ldots, j_{M_k}) \in \mathcal{J}$, and is aware of the sampling sequence without being informed additionally of it.

The rate of the code is

$$
\begin{aligned}
\frac{1}{n}\log|\mathcal{J}| &= \frac{1}{n}\sum_{i=1}^{M_k}\log J_{A_i} \\
&\leq \frac{1}{n}\left(\sum_{i=1}^{M_k}|\nu_{A_i}|\left(R_{A_i}(\Delta_{A_i})+\frac{\epsilon'}{2}\right)\right) \\
&\leq \sum_{i=1}^{M_k}\left(\left(P_S(A_i)+\frac{1}{n}\right)\left(R_{A_i}(\Delta_{A_i})+\frac{\epsilon'}{2}\right)\right) \\
&\leq \sum_{i=1}^{M_k}P_S(A_i)R_{A_i}(\Delta_{A_i})+\epsilon' < R_{\imath}(\Delta)+\epsilon, \qquad (2.28)
\end{aligned}
$$

where the previous inequality holds for all n large enough. Denoting the decoder output by

$$
Y_B^n \triangleq \varphi\left(f(S^n, X_S^n)\right),
$$

we have that

$$
\begin{aligned}
\mathbb{E}[d(X_B^n, Y_B^n)] &= \mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}d(X_{Bt}, Y_{Bt})\right] \\
&= \frac{1}{n}\sum_{i=1}^{M_k}|\nu_{A_i}|\mathbb{E}\left[d\left(X_B^{\nu_{A_i}}, \varphi_{A_i}\left(f_{A_i}(X_{A_i}^{\nu_{A_i}})\right)\right)\right] \\
&\leq \sum_{i=1}^{M_k}\left(P_S(A_i)+\frac{1}{n}\right)\left(\Delta_{A_i}+\frac{\epsilon'}{2}\right) \\
&= \sum_{i=1}^{M_k}P_S(A_i)\Delta_{A_i}+\frac{M_k}{n}\left(\frac{\epsilon'}{2}+\Delta_{\max}\right)+\frac{\epsilon'}{2} \\
&\leq \Delta+\epsilon \qquad (2.29)
\end{aligned}
$$

by (2.13) and for all n large enough. The proof is completed by noting that (2.28) and (2.29) hold simultaneously for all n large enough. ■

Theorem 2.2: By (2.18), using the result of Theorem 2.3,

$$R_m^I(\Delta) = \min_{\substack{P_U,\ \Delta_u: \\ \sum\limits_u P_U(u)\Delta_u \leq \Delta}} \sum_{u \in \mathcal{U}} P_U(u)\tilde{R}(\Delta_u) \tag{2.30}$$

where

$$\tilde{R}(\Delta_u) = \min_{\substack{P_{X_\mathcal{M}}\delta_{w_u(\cdot)}P_{Y_B|SX_S} \\ \mathbb{E}[d(X_B,Y_B)] \leq \Delta_u}} I(X_S \wedge Y_B|S) \tag{2.31}$$

for $\Delta_{\min} \leq \Delta, \Delta_u \leq \Delta_{\max}$, with the pmf $P_{X_\mathcal{M}}\delta_{w_u(\cdot)}P_{Y_B|SX_S}$ being understood as $P_{X_\mathcal{M}}\delta_{w(\cdot,u)}P_{Y_B|SX_S,U=u}$. To simplify notation, the conditioning on $U = u$ will be suppressed except when needed. It suffices to show the existence of a code of rate $\cong \tilde{R}(\Delta_u)$ with distortion $\mathbb{E}[d(X_B, Y_B)] \overset{\sim}{\leq} \Delta_u$. A concatenation of such codes indexed by $u \in \mathcal{U}$ yields, in effect, suitable time-sharing among them, leading to the achievability of (2.30). By Theorem 2.3, in view of the optimality of deterministic samplers, concatenating fixed-set sampling rate distortion codes for conditional sources $P_{X_\mathcal{M}|S=A}$, $A \in \mathcal{A}_k$, will suffice.

Given any $\Delta_{\min} \leq \Delta_u \leq \Delta_{\max}$, for the minimizer in (2.31), consider the corresponding

$$P_{S|X_\mathcal{M}} = \delta_{w_u(\cdot)}, \quad \Delta_{A_i} \triangleq \mathbb{E}[d(X_B, Y_B)|S = A_i] \text{ and}$$

$$I(X_{A_i} \wedge Y_B | S = A_i), \quad i \in \mathcal{M}_k.$$

The associated $\{(S_t, X_{S_t})\}_{t=1}^{\infty}$ is an i.i.d. sequence (cf. Remark (ii) following Definition 2.3). The sampling sets characterized by the deterministic sampler above and the DMMS realizations $x_{\mathcal{M}}^n$, are denoted as $s^n(x_{\mathcal{M}}^n) \triangleq \big(s(x_{\mathcal{M}1}), \ldots, s(x_{\mathcal{M}n})\big)$, and hence $S^n = s^n(X_{\mathcal{M}}^n)$.

The idea behind the remainder of the proof below for each $U = u$ is the following. We collect all those time instants at which a particular A_i in $\mathcal{A}_k$ is sampled, with the objective of applying a fixed-set sampling rate distortion code. Since the size of this time-set will vary according to $x_{\mathcal{M}}^n$ in $\mathcal{X}_{\mathcal{M}}^n$, the rate of such a code, too, will vary accordingly. However, since we seek fixed rate codes (rather than codes with a desired average rate), we apply fixed-set sampling codes to subsets of predetermined lengths from among typical sampling sequences in $\mathcal{A}_k^n$.

Fix $\epsilon > 0$ and $0 < \epsilon' < \epsilon$. Ordering the elements of $\mathcal{A}_k$ as in the proof of Theorem 2.1, for $n \geq 1$, the sets $\tau_{s^n}(A_i) \triangleq \{t : 1 \leq t \leq n, \ s_t = A_i\}, \ i \in \mathcal{M}_k$, cover $\{1, \ldots, n\}$; denote the set of the first $\max\{\lceil n(P_S(A_i) - \epsilon')\rceil, 0\}$ time instants in $\tau_{s^n}(A_i)$ by ν_{A_i}. For the (typical) set

$$\mathcal{T}_{\epsilon'}^{(n)} \triangleq \left\{ s^n \in \mathcal{A}_k^n : \ \left| \frac{|\tau_{s^n}(A_i)|}{n} - P_S(A_i) \right| \leq \epsilon', \ i \in \mathcal{M}_k \right\},$$

$P\left(S^n \in \mathcal{T}_{\epsilon'}^{(n)}\right) \geq 1 - \frac{\epsilon'}{2}$ for all $n \geq N_1(\epsilon')$, say.

Along the lines of proof of Theorem 2.1, for each DMMS with (conditional) pmf $P_{X_{\mathcal{M}}|S=A_i}$, $i \in \mathcal{M}_k$, there exists a code (f_{A_i}, φ_{A_i}), $f_{A_i} : \mathcal{X}_{A_i}^{\nu_{A_i}} \to \{1, \ldots, J_{A_i}\}$ and $\varphi_{A_i} : \{1, \ldots, J_{A_i}\} \to \mathcal{Y}_B^{\nu_{A_i}}$ of rate $\frac{1}{|\nu_{A_i}|} \log J_{A_i} \leq I(X_{A_i} \wedge Y_B | S = A_i) + \frac{\epsilon'}{2}$ and with

$$\mathbb{E}\left[d\Big(X_B^{\nu_{A_i}}, \varphi_{A_i}\big(f_{A_i}(X_{A_i}^{\nu_{A_i}})\big)\Big) \Big| S^{\nu_{A_i}} = A_i^{\nu_{A_i}} \right] \leq \Delta_{A_i} + \frac{\epsilon'}{2}$$

for all $|\nu_{A_i}| \geq N_{A_i}\left(\frac{\epsilon'}{2}\right)$.

A (composite) code (f, φ_S), with f taking values in $\mathcal{J} \triangleq \bigtimes_{i=1}^{M_k} \{1, \ldots, J_{A_i}\}$ is constructed as follows. The encoder f consists of a concatenation of encoders defined by

$$f\left(s^n(x_{\mathcal{M}}^n); x_{s_1}, \ldots, x_{s_n}\right) = \begin{cases} \Big(f_{A_1}\left(x_{A_1}^{\nu_{A_1}}\right), \ldots, f_{A_{M_k}}\Big(x_{A_{M_k}}^{\nu_{A_{M_k}}}\Big)\Big), & s^n(x_{\mathcal{M}}^n) \in \mathcal{T}_{\epsilon'}^{(n)}, \\ (1, \ldots, 1), & s^n(x_{\mathcal{M}}^n) \notin \mathcal{T}_{\epsilon'}^{(n)}. \end{cases}$$

For $t = 1, \ldots, n$, and $(j_1, \ldots, j_{M_k}) \in \mathcal{J}$, the informed decoder φ_S is given by

$$\Big(\varphi_S\big(s^n, (j_1, \ldots, j_{M_k})\big)\Big)_t = \begin{cases} \left(\varphi_{A_i}\left(x_{A_i}^{\nu_{A_i}}\right)\right)_t, & s^n \in \mathcal{T}_{\epsilon'}^{(n)} \text{ and } t \in \nu_{A_i},\ i \in \mathcal{M}_k, \\ y_1, & \text{otherwise}, \end{cases}$$

where y_1 is a fixed but arbitrary symbol in $\mathcal{Y}_B$.

The rate of the code is

$$\frac{1}{n} \log |\mathcal{J}| = \frac{1}{n} \sum_{i=1}^{M_k} \log J_{A_i}$$

$$
\begin{aligned}
&\le \frac{1}{n}\left(\sum_{i=1}^{M_k} |\nu_{A_i}| \left(I(X_{A_i} \wedge Y_B | S = A_i) + \frac{\epsilon'}{2}\right)\right) \\
&\le \sum_{i=1}^{M_k} P_S(A_i) \left(I(X_{A_i} \wedge Y_B | S = A_i) + \frac{\epsilon'}{2}\right) \\
&= \sum_{i=1}^{M_k} P_S(A_i) I\left(X_{A_i} \wedge Y_B | S = A_i\right) + \frac{\epsilon'}{2} \\
&\le I(X_S \wedge Y_B | S) + \epsilon. \qquad (2.32)
\end{aligned}
$$

Defining $d_{\max} \triangleq \max\limits_{(x_B, y_B) \in \mathcal{X}_B \times \mathcal{Y}_B} d(x_B, y_B)$, and with Y_B^n denoting the output of the decoder, we have

$$
\begin{aligned}
\mathbb{E}[d(X_B^n, Y_B^n)] =\ & \mathbb{E}\left[\mathbb{E}\left[d(X_B^n, Y_B^n) \middle| S^n\right]\right] \\
=\ & \sum_{s^n \in \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) \mathbb{E}\left[d(X_B^n, Y_B^n) \middle| S^n = s^n\right] + \sum_{s^n \notin \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) \mathbb{E}\left[d(X_B^n, Y_B^n) \middle| S^n = s^n\right] \\
\le\ & \sum_{s^n \in \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) \sum_{i=1}^{M_k} \frac{|\nu_{A_i}|}{n} \mathbb{E}\left[d\left(X_B^{\nu_{A_i}}, \varphi_{A_i}\left(f_{A_i}(X_{A_i}^{\nu_{A_i}})\right)\right) \middle| S^{\nu_{A_i}} = A_i^{\nu_{A_i}}\right] \\
& + \frac{1}{n} \sum_{s^n \in \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) \sum_{i=1}^{M_k} \sum_{t \in \tau_{s^n}(A_i) \setminus \nu_{A_i}} \mathbb{E}\left[d(X_{Bt}, Y_{Bt}) | S^n = s^n\right] + \sum_{s^n \notin \mathcal{T}_{\epsilon'}^{(n)}} P_{S^n}(s^n) d_{\max} \\
\le\ & \sum_{i=1}^{M_k} P_S(A_i) \left(\Delta_{A_i} + \frac{\epsilon'}{2}\right) + \left(1 - \frac{\sum_{i=1}^{M_k} |\nu_{A_i}|}{n}\right) d_{\max} + \frac{\epsilon'}{2} d_{\max} \\
\le\ & \Delta_u + \frac{\epsilon'}{2} + (2 M_k \epsilon') \, d_{\max} + \frac{\epsilon'}{2} d_{\max} \\
<\ & \Delta_u + \epsilon \qquad (2.33)
\end{aligned}
$$

for all n large enough. The proof is completed by noting that for n large enough (2.32) and (2.33) hold simultaneously and time-sharing between the codes corresponding to $U = u,\ u \in \mathcal{U}$, completes the proof. ■

Theorem 2.4: The proof is similar to that of Proposition 2.1 with the i.i.d. sequence $\{X_{At}\}_{t=1}^{\infty}$ replaced by the i.i.d. sequence $\{S_t, X_{S_t}\}_{t=1}^{\infty}$ with joint pmf P_{SX_S} obtained from (2.21) and a modified distortion measure $\tilde{d}\left((s, x_s), y_B\right) \triangleq \mathbb{E}\big[d(X_B, y_B)|S = s, X_S = x_s\big]$. The details, identical to those in the achievability proof of Proposition 2.1, are omitted. ■

2.4.2 Unified Converse Proof

Separate converse proofs can be provided for Proposition 2.1 and Proposition 2.3. However, in order to highlight the underlying ideas economically, we develop the proofs in a unified manner. Specifically, in contrast with the achievability proofs above, our converse proofs are presented in the order of weakening power of the sampler, viz., k-MRS, k-IRS and fixed-set sampler. We begin with the proof of Lemma 2.1 followed by pertinent technical results before turning to Proposition 2.1 and Proposition 2.3.

Lemma 2.1: We need to prove only that the right-sides of (2.7), (2.11) and (2.15) are convex and continuous, since they are evidently finite-valued and nonincreasing in Δ. The convexity of the right-side of (2.7) on $[\Delta_{\min,A}, \Delta_{\max}]$ is a standard consequence of the convexity of

$$I(X_A \wedge Y_B) = I\left(P_{X_A}, P_{Y_B|X_A}\right)$$

in $P_{Y_B|X_A}$ and the convexity of the constraint set in (2.7). The convexity of the right-sides of (2.11) and (2.15) is immediate by the remarks following Proposition 2.2 and Theorem 2.2, and their continuity for $\Delta > \Delta_{\min}$ is a consequence. Continuity at $\Delta = \Delta_{\min}$ in (2.7), (2.11) and (2.15) holds, for instance, as in ([34], Lemma 7.2). ■

Lemma 2.2 *Let the finite-valued rvs C^n, D^n, E^n, F^n be such that (C_t, D_t), $t = 1, \ldots, n$, are mutually independent and satisfy*

$$D^n \multimap\!\!- C^n, E^n \multimap\!\!- F^n \tag{2.34}$$

and

$$E_t \multimap\!\!- C_t, D_t, E^{t-1} \multimap\!\!- C^{n\backslash t}, D^{n\backslash t}, \quad t = 1, \ldots, n, \tag{2.35}$$

where $C^{n\backslash t} = C^n \setminus C_t$. Then, the following hold for $t = 1, \ldots, n$:

$$I(C^t, D^t, E^t \wedge C^n_{t+1}, D^n_{t+1}) = 0 \ ; \tag{2.36}$$

$$C_t, D_t \multimap\!\!- E^t \multimap\!\!- C^{n\backslash t}, D^{n\backslash t}, E^n_{t+1} \ ; \tag{2.37}$$

and

$$D_t \multimap\!\!- C_t, E^t \multimap\!\!- F_t. \tag{2.38}$$

Proof: First, (2.36) is true by the following simple observation: for $t = 1, \ldots, n$,

$$\begin{aligned} & I(C^t, D^t, E^t \wedge C_{t+1}^n, D_{t+1}^n) \\ & \quad = I(C^t, D^t \wedge C_{t+1}^n, D_{t+1}^n) + I(E^t \wedge C_{t+1}^n, D_{t+1}^n | C^t, D^t) \\ & \quad = 0 \end{aligned} \tag{2.39}$$

where the first term in the sum above is zero by the mutual independence of (C_t, D_t), $t = 1, \ldots, n$, and the second term equals zero by (2.35). Next, the claim (2.36) and the Markov property (2.35) imply that for $t = 1, \ldots, n$,

$$\begin{aligned} & I(C_t, D_t, E_t \wedge C^{t-1}, D^{t-1} | E^{t-1}) \\ & \quad = I(C_t, D_t \wedge C^{t-1}, D^{t-1} | E^{t-1}) + I(E_t \wedge C^{t-1}, D^{t-1} | C_t, D_t, E^{t-1}) \\ & \quad = 0. \end{aligned} \tag{2.40}$$

The claim (2.37) now follows, since

$$\begin{aligned} & I(C_t, D_t \wedge C^{n \backslash t}, D^{n \backslash t}, E_{t+1}^n | E^t) \\ & \quad = I(C_t, D_t \wedge C^{t-1}, D^{t-1} | E^t) + I(C_t, D_t \wedge C_{t+1}^n, D_{t+1}^n | C^{t-1}, D^{t-1}, E^t) \\ & \quad\quad + I(C_t, D_t \wedge E_{t+1}^n | C^{n \backslash t}, D^{n \backslash t}, E^t) \\ & \quad = 0 \end{aligned}$$

where the first term in the sum above is zero by (2.40), and the latter two terms are zero by (2.39) and (2.35), respectively.

Now using (2.34),

$$
\begin{aligned}
0 &= I(D^n \wedge F^n | C^n, E^n) \\
&= \sum_{t=1}^{n} I(D_t \wedge F^n | D^{t-1}, C^n, E^n) \\
&= \sum_{t=1}^{n} I(D_t \wedge F^n | D^{t-1}, C_t, C^{n\backslash t}, E^t, E_{t+1}^n) \\
&= \sum_{t=1}^{n} \Big[I(D_t \wedge D^{t-1}, C^{n\backslash t}, E_{t+1}^n, F^n | C_t, E^t) - I(D_t \wedge D^{t-1}, C^{n\backslash t}, E_{t+1}^n | C_t, E^t) \Big] \\
&= \sum_{t=1}^{n} I(D_t \wedge D^{t-1}, C^{n\backslash t}, E_{t+1}^n, F^n | C_t, E^t) \quad \text{by (2.37)} \\
&\geq \sum_{t=1}^{n} I(D_t \wedge D^{t-1}, C^{n\backslash t}, E_{t+1}^n, F_t | C_t, E^t) \\
&\geq \sum_{t=1}^{n} I(D_t \wedge F_t | C_t, E^t),
\end{aligned}
$$

so that the claim (2.38) follows. ■

We now prove Proposition 2.3 which, in effect, implies the converse proofs for Theorem 2.2, Theorem 2.1 and Proposition 2.1. Specifically, a converse is fashioned for $R_{ms}^I(\Delta)$, with those for $R_{\imath s}(\Delta)$ and $R_A(\Delta)$ emerging along the way.

Let $\big(\{P_{S_t|X_{\mathcal{M}t}S^{t-1}}\}_{t=1}^n, f, \varphi_S\big)$ be an n-length strong k-MRS block code with decoder output $Y_B^n = \varphi_S\big(S^n, f(S^n, X_S^n)\big)$ and satisfying $\mathbb{E}\left[d\big(X_B^n, Y_B^n\big)\right] \leq \Delta$. The hypothesis of Lemma 2.2 with $C^n = X_S^n,\ D^n = X_{S^c}^n,\ E^n = S^n$ and $F^n = Y_B^n$ is met since

$$X^n_{S^c} -\!\!\circ\!\!- S^n, X^n_S -\!\!\circ\!\!- Y^n_B$$

and by (2.22),

$$P_{S_t|X^n_{\mathcal{M}}S^{t-1}} = P_{S_t|X_{\mathcal{M}t}S^{t-1}}.$$

Then by Lemma 2.2, for $t = 1, \ldots, n$,

$$I(S^{t-1} \wedge X_{\mathcal{M}t}) = 0, \tag{2.41}$$

$$X_{\mathcal{M}t} -\!\!\circ\!\!- S^t -\!\!\circ\!\!- X^{n\backslash t}_{\mathcal{M}}, X^{n\backslash t}_S, S^n_{t+1} \tag{2.42}$$

and

$$X_{S^c_t} -\!\!\circ\!\!- S^t, X_{S_t} -\!\!\circ\!\!- Y_{Bt}. \tag{2.43}$$

The rate R of the code satisfies

$$\begin{aligned}
nR &= \log|f| \geq H\big(f(S^n, X^n_S)\big) \\
&\geq H\Big(\varphi_S\big(S^n, f(S^n, X^n_S)\big)|S^n\Big) = H\big(Y^n_B|S^n\big) \\
&= H\big(Y^n_B|S^n\big) - H\big(Y^n_B|S^n, X^n_S\big)
\end{aligned}$$

$$= I\big(X_S^n \wedge Y_B^n | S^n\big)$$
$$= \sum_{t=1}^{n} \big(H(X_{S_t}|S^t, S_{t+1}^n, X_S^{t-1}) - H(X_{S_t}|S^t, S_{t+1}^n, X_S^{t-1}, Y_B^n)\big)$$
$$\geq \sum_{t=1}^{n} \big(H(X_{S_t}|S^t) - H(X_{S_t}|S^t, Y_{Bt})\big) \tag{2.44}$$
$$= \sum_{t=1}^{n} I(X_{S_t} \wedge Y_{Bt}|S^t) \tag{2.45}$$

where (2.44) follows from (2.42). Denote $\mathbb{E}[d(X_{Bt}, Y_{Bt})]$ by Δ_t.

For the strong k-MRS code above, in (2.45) using (2.41) and (2.43), we get

$$I(X_{S_t} \wedge Y_{Bt}|S^t) \geq \min I(X_{S_t} \wedge Y_{Bt}|S_t, S^{t-1}) \tag{2.46}$$
$$\geq \min_{\substack{P_{U_t} P_{X_{\mathcal{M}t}} P_{S_t|X_{\mathcal{M}t}U_t} P_{Y_{Bt}|S_t X_{S_t} U_t} \\ \mathbb{E}[d(X_{\mathcal{M}t}, Y_{Bt})] \leq \Delta_t}} I(X_{S_t} \wedge Y_{Bt}|S_t, U_t), \tag{2.47}$$

where the minimum in (2.46) is with respect to $P_{X_{\mathcal{M}t} S^t Y_{Bt}} = P_{S^{t-1}} P_{X_{\mathcal{M}t}} P_{S_t|X_{\mathcal{M}t} S^{t-1}} P_{Y_{Bt}|S_t X_{S_t} S^{t-1}}$ and $\mathbb{E}[d(X_{Bt}, Y_{Bt})] = \Delta_t$, and where U_t is a rv taking values in a set of cardinality $|\mathcal{A}_k|^{t-1}$. The existence of the minima in (2.46) and (2.47) comes from the continuity of the conditional mutual information terms over compact sets of pmfs.

By the Carathéodory Theorem [35], every point in the convex hull of the set

$$\mathcal{C} = \Big\{ \big(\mathbb{E}[d(X_B, Y_B)],\, I(X_S \wedge Y_B|S)\big) : \quad X_{\mathcal{M}} -\!\!\circ\!\!- S, X_S -\!\!\circ\!\!- Y_B \Big\} \subset \mathbb{R}^2$$

can be represented as a convex combination of at most three points in $\mathcal{C}$. Hence, to describe every element in the set

$$\Big\{\big(\mathbb{E}[d(X_{Bt},Y_{Bt})],\ I(X_{S_t}\wedge Y_{Bt}|S_t,U_t)\big):\ P_{U_tX_{\mathcal{M}t}S_tY_{Bt}} = P_{U_t}P_{X_{\mathcal{M}t}}P_{S_t|X_{\mathcal{M}t}U_t}P_{Y_{Bt}|S_tX_{S_t}U_t}\Big\},$$

it suffices to consider a rv U_t with support of size three. (For $t=1$, this assertion is straightforward.) Consequently, the right-side of (2.47) equals $R_m^I(\Delta_t)$ (cf. (2.15)). Using the convexity of $R_m^I(\Delta)$ in Δ, we get from (2.45) that

$$\begin{aligned} nR &\geq \sum_{t=1}^{n} R_m^I(\Delta_t) \\ &\geq nR_m^I\left(\frac{1}{n}\sum_{t=1}^{n}\Delta_t\right) \\ &\geq nR_m^I(\Delta), \end{aligned} \tag{2.48}$$

i.e., $R \geq R_m^I(\Delta),\ \Delta \geq \Delta_{\min}$, thereby completing the converse proof for a strong k-MRS and Theorem 2.2.

Next, an n-length strong k-IRS code and fixed-set sampler code can be viewed as restrictions of the strong k-MRS code above. Specifically, the strong k-IRS and fixed-set sampler respectively entail replacing $P_{S_t|X_{\mathcal{M}t}S^{t-1}}$ by $P_{S_t|S^{t-1}}$ and $P_{S_t} = \mathbb{1}(S_t = A)$. Counterparts of (2.46) and (2.47) hold with the mentioned replacements. For a strong k-IRS, upon replacing $P_{S_t|X_{\mathcal{M}t}S^{t-1}}$ with $P_{S_t|S^{t-1}}$, we observe that the right-side of (2.47), viz.

$$\min_{\substack{P_{U_t}P_{X_{\mathcal{M}t}}P_{S_t|U_t}P_{Y_{Bt}|S_tX_{S_t}U_t} \\ \mathbb{E}[d(X_{Bt},Y_{Bt})]\leq\Delta_t}} I(X_{S_t}\wedge Y_{Bt}|S_t,U_t)$$

is now the lower convex envelope of the SRDf for a k-IRS, already convex in distortion, and hence, equals $R_\imath(\Delta_t)$ itself. Thus, (2.47) becomes

$$\begin{aligned} I(X_{S_t} \wedge Y_{Bt}|S^t) \geq & \min_{\substack{P_{X_{\mathcal{M}t}} P_{S_t} P_{Y_{Bt}|S_t X_{S_t}} \\ \mathbb{E}[d(X_{Bt}, Y_{Bt})] \leq \Delta_t}} I(X_{S_t} \wedge Y_{Bt}|S_t) \\ = & \ R_\imath(\Delta_t). \end{aligned} \tag{2.49}$$

Combining (2.45) and (2.49), we get along the lines of (2.48) that $R \geq R_\imath(\Delta)$, $\Delta \geq \Delta_{\min}$, which gives the converse proof for a strong k-IRS and Theorem 2.1.

In a manner analogous to a strong k-IRS, for a fixed-set sampler the convexity of $R_A(\Delta)$ in Δ implies that the counterpart of the right-side of (2.47), with $P_{S_t|X_{\mathcal{M}t}U_t}$ replaced by $\mathbb{1}(S_t = A)$, simplifies to $R_A(\Delta_t)$. As in (2.48), it follows that $R \geq R_A(\Delta)$, $\Delta \geq \Delta_{\min,A}$, which gives the converse for Proposition 2.1.

■

2.5 Discussion

Our current formulation requires that a prespecified subset of DMMS components be reconstructed from the compressed representations of the output of the sampler. If the reconstruction procedure were restricted to be a two-step procedure (with reduced complexity) – wherein the sampled DMMS components are reconstructed first based on which reconstructions for the unsampled components are formed, what would be the resulting SRDf? In Corollary 2.1, for the probability of error distortion measure, such a two-step procedure was seen to be optimal. Which other distortion measures allow for the DMMS reconstruction to be performed optimally in such steps?

In this chapter two particular instances of random samplers with "memory," a strong k-IRS and a strong k-MRS were seen to not improve over their memoryless counterparts in terms of the SRDf. This can be attributed to the fact that the source under consideration is memoryless, the underlying pmf of the DMMS is known and at any time instant $t \geq 1$, the sampled subset S_t is allowed to depend only on a subset of the output of the sampler from previous time instants (S^{t-1} in the case of the strong k-IRS and k-MRS). Does a k-RS of the form

$$P_{S_t|X_{\mathcal{M}}^t S^{t-1}} = P_{S_t|X_{\mathcal{M}(t-1)} S^{t-1}} \qquad t = 1, \ldots$$

improve over a k-IRS in terms of the SRDf? Does a k-RS of the form

$$P_{S_t|X_{\mathcal{M}}^t S^{t-1}} = P_{S_t|X_{\mathcal{M}t} X_{\mathcal{M}(t-1)} S^{t-1}}, \qquad t = 1, \ldots$$

improve over a k-MRS in terms of the SRDf? These questions can be the first step towards a better understanding the SRDf for a k-RS with memory.

Chapter 3

Universal Sampling Rate Distortion: Finite-Valued Multiple Source

3.1 Synopsis

In this chapter, we consider a DMMS with m components whose *joint pmf* is known only to belong to a given finite family of pmfs. A subset of k components of the DMMS are sampled (possibly in a randomized manner) at each time instant and compressed jointly with the objective of reconstructing a prespecified subset of the m components under a suitable distortion criterion.

In Section 3.2, we describe our model for universality and introduce the notion of a universal sampling rate distortion function (USRDf) to study the tradeoffs among sampling mechanisms, estimation of underlying pmf, compression rate and desired level of accuracy in reconstruction. When the underlying pmf of the DMMS is not known precisely, the sampling mechanism plays the role of: (i) aiding in the estimating of the underlying pmf and (ii) sampling appropriate components of the DMMS to enable optimal compression by the encoder. In Section 3.3, considering the sampling mechanisms introduced in Chapter 2, we provide single-letter characterizations for the USRDf and show how each sampler enables improvements over the previous one in both the roles mentioned above. In Section 3.4, we present the achievability proofs in an increasing complexity of the sampler complexity, with an emphasis on the Bayesian setting. A unified converse proof is presented thereafter.

3.2 Model

Let Θ be a finite set (of parameters) and θ a Θ-valued rv with pmf μ_θ of assumed full support. For a finite-valued rv $X_{\mathcal{M}}$ (as in Chapter 2) we consider a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ consisting of i.i.d. repetitions of the rv $X_{\mathcal{M}}$ with pmf known only to the extent of belonging to a finite family of pmfs $\mathcal{P} = \{P_{X_{\mathcal{M}}|\theta=\tau},\ \tau \in \Theta\}$ of assumed full support. As in Chapter 2, $\mathcal{Y}_{\mathcal{M}}$ is a finite reproduction alphabet for $X_{\mathcal{M}}$. Two settings are studied: in a Bayesian formulation, the pmf μ_θ is taken to be *known* while in a nonBayesian formulation θ is an *unknown constant* in Θ. A k-RS, unaware of the underlying pmf of the DMMS is defined below, along the lines of Definition 2.1.

Definition 3.1 *In the Bayesian setting, a k-random sampler (k-RS),* $1 \leq k \leq m$, *collects causally at each $t = 1, \ldots, n$, random samples X_{S_t} from $X_{\mathcal{M}t}$, where S_t is a rv with values in $\mathcal{A}_k$ with (conditional) pmf $P_{S_t|X^t_{\mathcal{M}}S^{t-1}}$, Such a k-RS is specified by a (conditional) pmf $P_{S^n|X^n_{\mathcal{M}}\theta}$ with the requirement*

$$P_{S^n|X^n_{\mathcal{M}}\theta} = P_{S^n|X^n_{\mathcal{M}}} = \prod_{t=1}^{n} P_{S_t|X^t_{\mathcal{M}}S^{t-1}}. \tag{3.1}$$

In the nonBayesian setting, the first equality above is redundant.

The output of a k-RS is (S^n, X_S^n) *where* $X_S^n = (X_{S_1}, \ldots, X_{S_n})$. *We consider successively restrictive choices of a k-RS in (the right-side of)* (3.1), *namely k-MRS, k-IRS and k-FS as in* (2.2), (2.3) *and* (2.4), *respectively.*

Our objective is to reconstruct a subset of DMMS components with indices in an arbitrary but fixed *recovery set* $B \subseteq \mathcal{M}$, namely X_B^n, from a compressed representation of the k-RS output (S^n, X_S^n), under a suitable distortion criterion.

For $n \geq 1$, an n-length block code with a k-RS for a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with alphabet $\mathcal{X}_{\mathcal{M}}$ and reproduction alphabet $\mathcal{Y}_B$ is $(P_{S^n|X_{\mathcal{M}}^n}, f_n, \varphi_n)$, as in Definition 2.2, with the distinction that now the code is formed without exact knowledge of the underlying pmf. Note that an encoder that operates by forming first an explicit estimate of θ from (S^n, X_S^n) is subsumed by this definition.

Remark: Throughout this chapter we restrict ourselves to an informed decoder. This assumption is meaningful for a k-IRS and k-MRS. For a k-IRS, it will be shown to be not needed.

For a given (single-letter) finite-valued distortion measure $d : \mathcal{X}_B \times \mathcal{Y}_B \rightarrow \mathbb{R}^+ \cup \{0\}$, an n-length block code with k-RS $(P_{S|X_{\mathcal{M}}}, f, \varphi_S)$ will be required to satisfy one of the following distortion criteria (d, Δ) depending on the setting.

(i) Bayesian: The *expected* distortion criterion is

$$
\begin{aligned}
\mathbb{E}\Big[d\Big(X_B^n, \varphi_S\big(S^n, f(S^n, X_S^n)\big)\Big)\Big] &\triangleq \mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} d\Big(X_{Bt}, \Big(\varphi_S\big(S^n, f(S^n, X_S^n)\big)\Big)_t\Big)\Big] \\
&= \sum_{\tau\in\Theta} \mu_\theta(\tau)\mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} d\Big(X_{Bt}, \Big(\varphi_S\big(S^n, f(S^n, X_S^n)\big)\Big)_t\Big)\Big|\theta = \tau\Big] \\
&\leq \Delta.
\end{aligned}
\tag{3.2}
$$

(ii) NonBayesian: The *peak* distortion criterion is

$$
\max_{\tau\in\Theta} \mathbb{E}\Big[d\Big(X_B^n, \varphi_S\big(S^n, f(S^n, X_S^n)\big)\Big)\big|\theta = \tau\Big] \leq \Delta, \tag{3.3}
$$

where the "conditional" expectation denotes, in fact, $\mathbb{E}_{P_{X_{\mathcal{M}}^n S^n|\theta=\tau}} = \mathbb{E}_{P_{X_{\mathcal{M}}^n|\theta=\tau} P_{S^n|X_{\mathcal{M}}^n}}$.

Definition 3.2 *A number $R \geq 0$ is an achievable universal k-RS coding rate at distortion level Δ if for every $\epsilon > 0$ and sufficiently large n, there exist n-length block codes with k-RS of rate less than $R + \epsilon$ and satisfying the distortion criterion $(d, \Delta+\epsilon)$ in* (3.2) *or* (3.3) *above; and (R, Δ) will be termed an achievable universal k-RS rate distortion pair under the expected or peak distortion criterion. The infimum of such achievable rates is denoted by $R_A(\Delta)$, $R_\imath^I(\Delta)$ and $R_m^I(\Delta)$ for a k-FS, k-IRS and k-MRS, respectively. We shall refer to $R_A(\Delta)$, $R_\imath^I(\Delta)$ as well as $R_m^I(\Delta)$ as the universal sampling rate distortion function (USRDf), suppressing the dependence on k.*

Remarks: (i) Clearly, the USRDf under (3.2) will be no larger than that under (3.3).

(ii) For $|\Theta| = 1$, the pmf of the rv $X_{\mathcal{M}}$ is, in effect, known and the distortion criteria and the setting above reduce to that in Chapter 2 and the USRDf is simply a SRDf. Hence, here onwards, we take $|\Theta| > 1$.

3.3 Universal Sampling Rate Distortion function

Throughout this chapter, a salient theme that recurs is this: an encoder without prior knowledge of θ and with access to only k instantaneously sampled components of the DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ can form only a limited estimate of θ. The quality of said estimate improves steadily from k-FS to k-IRS to k-MRS.

3.3.1 Fixed-Set Sampling

Consider first fixed-set sampling with $A \subseteq \mathcal{M}$ in (2.4). An encoder f with access to X_A^n cannot distinguish among pmfs in $\mathcal{P}$ (indexed by τ) that have the same $P_{X_A|\theta=\tau}$. Accordingly, let Θ_1 be a partition of Θ comprising "ambiguity" atoms, with each such atom consisting of τs with *identical marginal pmfs* $P_{X_A|\theta=\tau}$. Indexing the elements of Θ_1 by τ_1, let θ_1 be a Θ_1-valued rv with pmf μ_{θ_1} induced by μ_θ. For each $\tau_1 \in \Theta_1$, let $\Lambda(\tau_1)$ be the collection of τs in the atom of Θ_1 indexed by τ_1. In the Bayesian setting, clearly

$$P_{X_A|\theta_1=\tau_1} = P_{X_A|\theta=\tau}, \quad \tau \in \Lambda(\tau_1).$$

In the nonBayesian setting, in order to retain the same notation, we choose $P_{X_A|\theta_1=\tau_1}$ to be the right-side above.

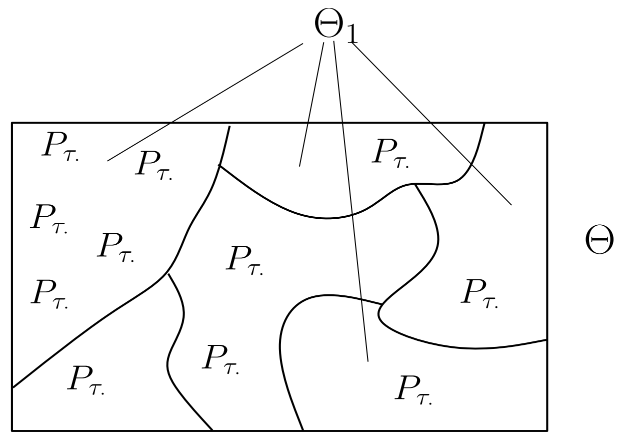

Figure 3.1: Ambiguity atoms for k-FS

When the pmf of the DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ is *known*, say $P_{X_{\mathcal{M}}}$ – corresponding to $|\Theta| = 1$ – we recall from Chapter 2, that the (U)SRDf for fixed $A \subseteq \mathcal{M}$ is

$$R_A(\Delta) = \min_{\substack{X_{\mathcal{M}} \multimap X_A \multimap Y_B \\ \mathbb{E}[d(X_B, Y_B)] \leq \Delta}} I(X_A \wedge Y_B), \qquad \Delta_{\min} \leq \Delta \leq \Delta_{\max}, \tag{3.4}$$

with

$$\Delta_{\min} = \mathbb{E}\big[\min_{y_B \in \mathcal{Y}_B} \mathbb{E}[d(X_B, y_B)|X_A]\big], \quad \Delta_{\max} = \min_{y_B \in \mathcal{Y}_B} \big[\mathbb{E}[d(X_B, y_B)|X_A]\big],$$

which can be interpreted as the (standard) rate distortion function for the DMMS $\{X_{At}\}_{t=1}^{\infty}$ using a modified distortion measure $\tilde{d}$ defined by

$$\tilde{d}(x_A, y_B) = \mathbb{E}[d(X_B, y_B)|X_A = x_A].$$

This fact will serve as a stepping stone to our analysis of USRDf for a k-random sampler. In the Bayesian setting, we consider a modified distortion measure d_{τ_1},

$\tau_1 \in \Theta_1$, given by

$$d_{\tau_1}(x_A, y_B) \triangleq \mathbb{E}[d(X_B, y_B)|X_A = x_A,\ \theta_1 = \tau_1]; \tag{3.5}$$

the set of (constrained) pmfs

$$\kappa_A^{\mathcal{B}}(\delta, \tau_1) \triangleq \{P_{\theta X_{\mathcal{M}} Y_B} :\ \theta, X_{\mathcal{M}} -\!\!\circ\!\!- \theta_1, X_A -\!\!\circ\!\!- Y_B,\ \mathbb{E}[d_{\tau_1}(X_A, Y_B)|\theta_1 = \tau_1] \leq \delta\}, \tag{3.6}$$

and the (minimized) conditional mutual information

$$\rho_A^{\mathcal{B}}(\delta, \tau_1) \triangleq \min_{\kappa_A^{\mathcal{B}}(\delta,\tau_1)} I(X_A \wedge Y_B|\theta_1 = \tau_1) \tag{3.7}$$

which is akin to (3.4) and will play a basal role. In the nonBayesian setting, the counterparts of (3.6) and (3.7) are

$$\kappa_A^{n\mathcal{B}}(\delta, \tau_1) \triangleq \{P_{X_{\mathcal{M}} Y_B|\theta=\tau} = P_{X_{\mathcal{M}}|\theta=\tau} P_{Y_B|X_A, \theta_1=\tau_1} :\ \mathbb{E}[d(X_B, Y_B)|\theta = \tau] \leq \delta,\ \tau \in \Lambda(\tau_1)\} \tag{3.8}$$

and

$$\rho_A^{n\mathcal{B}}(\delta, \tau_1) \triangleq \min_{\kappa_A^{n\mathcal{B}}(\delta,\tau_1)} I(X_A \wedge Y_B|\theta_1 = \tau_1). \tag{3.9}$$

Remarks: (i) The minima in (3.7) and (3.9) exist as those of convex functions over convex, compact sets.

(ii) Clearly, the minimum in (3.9) under pmf-wise constraints (3.8) can be no smaller than that in (3.7) under pmf-averaged constraints (3.6).

(iii) It is seen in a standard manner that $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ in (3.7) and $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ in (3.9) are convex and continuous in δ.

Our first main result states that the USRDf at distortion level Δ for fixed-set sampling in the Bayesian setting is a minmax of quantities in (3.7), where the maximum is over ambiguity atoms τ_1 in Θ_1, while the minimum is over distortion thresholds $\delta = \Delta_{\tau_1},\ \tau_1 \in \Theta_1$ whose mean does not exceed Δ. On the other hand, in the nonBayesian setting, the USRDf at distortion level Δ is a maximum over ambiguity atoms of quantities in (3.9) with $\delta = \Delta$, and hence is no smaller than its Bayesian counterpart.

Theorem 3.1 *The Bayesian USRDf for fixed $A \subseteq \mathcal{M}$ is*

$$R_A(\Delta) = \min_{\substack{\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}] \leq \Delta}} \max_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) \tag{3.10}$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, *where*

$$\Delta_{\min} = \mathbb{E}\Big[\mathbb{E}\big[\min_{y_B \in \mathcal{Y}_B} d_{\theta_1}(X_A, y_B)|\theta_1\big]\Big] = \mathbb{E}\big[\min_{y_B \in \mathcal{Y}_B} d_{\theta_1}(X_A, y_B)\big],$$

$$\Delta_{\max} = \mathbb{E}\big[\min_{y_B \in \mathcal{Y}_B} \mathbb{E}[d_{\theta_1}(X_A, y_B)|\theta_1]\big].$$

The nonBayesian USRDf is

$$R_A(\Delta) = \max_{\tau_1 \in \Theta_1} \rho_A^{n\mathcal{B}}(\Delta, \tau_1), \quad \Delta_{\min} \leq \Delta \leq \Delta_{\max} \tag{3.11}$$

where

$$\Delta_{\min} = \max_{\tau_1 \in \Theta_1} \min_{P_{Y_B|X_A,\theta_1=\tau_1} = P_{Y_B|X_{\mathcal{M}},\theta=\tau}} \max_{\tau \in \Lambda(\tau_1)} \mathbb{E}[d(X_B, Y_B)|\theta = \tau]$$

and

$$\Delta_{\max} = \max_{\tau_1 \in \Theta_1} \min_{y_B \in \mathcal{Y}_B} \max_{\tau \in \Lambda(\tau_1)} \mathbb{E}[d(X_B, y_B)|\theta = \tau].$$

Remarks: (i) In fact, the minimizing pmf $P_{Y_B|X_A\theta_1}$ in $\Delta_{\min}$ is a conditional point-mass.

(ii) We note that for a given distortion level Δ, the set $\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1 : \sum_{\tau_1 \in \Theta_1} \mu_{\theta_1}(\tau_1)\Delta_{\tau_1} \leq \Delta\}$ is a convex, compact set in $\mathbb{R}^{|\Theta_1|}$. Next, observing that

$$\max_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1)$$

is a convex function of $\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\}$, the minimum in (3.10) exists as that of a convex function over a convex, compact set.

(iii) The minimizing $\{\Delta^*_{\tau_1},\ \tau_1 \in \Theta_1\}$ in (3.10) is characterized by the following special property: for a given $\Delta_{\min} \le \Delta \le \Delta_{\max}$, for each $\tau_1 \in \Theta_1$, either

$$\rho_A^{\mathcal{B}}(\Delta^*_{\tau_1}, \tau_1) \equiv \max_{\tilde{\tau}_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta^*_{\tilde{\tau}_1}, \tilde{\tau}_1) \tag{3.12}$$

where the right-side does not depend on τ_1, or

$$\Delta^*_{\tau_1} = \mathbb{E}[\min_{y_B \in \mathcal{Y}_B} d_{\tau_1}(X_A, y_B) | \theta_1 = \tau_1].$$

By a standard argument in convex optimization, if $\{\Delta^*_{\tau_1},\ \tau_1 \in \Theta_1\}$ does not satisfy the property above, then a small perturbation decreases the maximum in (3.12) leading to a contradiction.

(iv) The $\Delta_{\min}$ and $\Delta_{\max}$ for the Bayesian and the nonBayesian settings can be different.

Example 3.1 *For the probability of error distortion measure*

$$d(x_B, y_B) = \mathbb{1}(x_B \neq y_B), \quad x_B, y_B \in \mathcal{X}_B = \mathcal{Y}_B,$$

the Bayesian USRDf for fixed-set sampling with $A \subseteq B$ *in* (3.10) *simplifies with* (3.7) *becoming*

$$\rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) = \min_{\mathbb{E}[\alpha_{\tau_1}(X_A)\mathbb{1}(X_A \neq Y_A)|\theta_1 = \tau_1] \le \Delta_{\tau_1} - (1 - \mathbb{E}[\alpha_{\tau_1}(X_A)|\theta_1 = \tau_1])} I(X_A \wedge Y_A | \theta_1 = \tau_1) \tag{3.13}$$

where

$$\alpha_{\tau_1}(x_A) = \max_{\tilde{x} \in \mathcal{X}_B} P_{X_B|X_A\theta_1}(\tilde{x}|x_A, \tau_1) \tag{3.14}$$

is the MAP estimate of X_B on the basis of $X_A = x_A$ under pmf $P_{X_{\mathcal{M}}|\theta_1=\tau_1}$.

The proof of (3.13), (3.14) *is along the lines of that of Corollary 2.1 under the pmf $P_{X_{\mathcal{M}}|\theta_1=\tau_1}$ (rather than $P_{X_{\mathcal{M}}}$ as in Corollary 2.1), and so is not repeated here. Furthermore,*

$$\Delta_{\min} = 1 - \mathbb{E}[\alpha_{\theta_1}(X_A)] \quad \textit{and} \quad \Delta_{\max} = 1 - \mathbb{E}\big[\max_{x_B \in \mathcal{X}_B} P_{X_B|\theta_1}(x_B|\theta_1)\big].$$

The form of the Bayesian USRDf in (3.13) *suggests a simple achievability scheme comprising two steps. Using a MAP or maximum likelihood (ML) estimate $\widehat{\tau}_1$ of θ_1 on the basis of $X_A^n = x_A^n$, the first step entails a lossy reconstruction of x_A^n by its codeword y_A^n, under pmf $P_{X_{\mathcal{M}}|\theta_1=\widehat{\tau}_1}$ and for a modified distortion measure*

$$\tilde{d}_{\widehat{\tau}_1}(x_A, y_A) \triangleq \alpha_{\widehat{\tau}_1}(x_A)\mathbb{1}(x_A \neq y_A)$$

with a corresponding reduced threshold

$$\Delta_{\widehat{\tau}_1} - (1 - \mathbb{E}[\alpha_{\widehat{\tau}_1}(X_A)|\theta_1 = \widehat{\tau}_1]).$$

This is followed by a second step of reconstructing x_B^n from the output y_A^n of the *previous step as a MAP estimate*

$$y_B^n = \arg\max_{y^n \in \mathcal{Y}_B^n} P_{X_B|X_A\theta_1}(y^n|y_A^n, \widehat{\tau}_1);$$

the corresponding probability of estimation error coincides with the mentioned reduction $1 - \mathbb{E}[\alpha_{\widehat{\tau}_1}(X_A)|\theta_1 = \widehat{\tau}_1]$ *in the threshold.*

In the nonBayesian setting, the USRDf in (3.11), (3.9) *simplifies with*

$$\rho_A^{n\mathcal{B}}(\Delta, \tau_1) = \min_{\substack{P_{Y_A|X_A,\theta_1=\tau_1} P_{Y_{B\setminus A}|Y_A,\theta_1=\tau_1} = P_{Y_B|X_{\mathcal{M}},\theta=\tau} \\ \mathbb{E}[\mathbb{1}(X_B \neq Y_B)|\theta=\tau] \leq \Delta, \quad \tau \in \Lambda(\tau_1)}} I(X_A \wedge Y_A|\theta_1 = \tau_1), \quad (3.15)$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, *where*

$$\Delta_{\min} = \max_{\tau_1 \in \Theta_1} \min_{P_{Y_B|X_A,\theta_1=\tau_1}} \max_{\tau \in \Lambda(\tau_1)} \big(1 - P(X_B = Y_B|\theta = \tau)\big)$$

and

$$\Delta_{\max} = \max_{\tau_1 \in \Theta_1} \min_{y_B \in \mathcal{Y}_B} \max_{\tau \in \Lambda(\tau_1)} \big(1 - P_{X_B|\theta}(y_B|\tau)\big).$$

This leads to the following achievability scheme. With $\widehat{\tau}_1$ *as the ML estimate of* θ_1 *formed from* $X_A^n = x_A^n$, *first* x_A^n *is reconstructed as* y_A^n *according to* $P_{Y_A|X_A,\theta_1=\widehat{\tau}_1}$ *resulting from the minimization in* (3.15). *This is followed by the reconstruction of* x_B^n *from* y_A^n *by means of the estimate*

$$y_B^n = \arg\max_{y^n \in \mathcal{Y}_B^n} P_{Y_B|Y_A\theta_1}(y^n|y_A^n, \widehat{\tau}_1)$$

under pmf $P_{Y_B|Y_A\theta_1}$ *which, too, is obtained from the minimization in* (3.15).

■

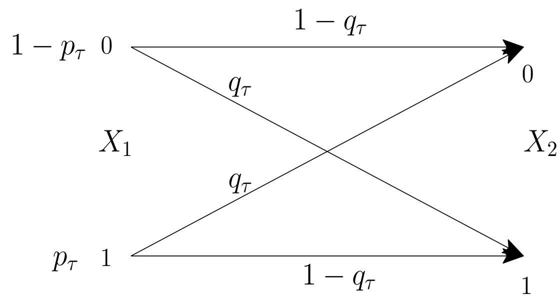

Figure 3.2: Virtual BSC (q_τ)

Example 3.2 *Let* $\mathcal{M} = \{1,2\}$ *and* $\mathcal{X}_1 = \mathcal{X}_2 = \{0,1\}$, *consider a DMMS with* $P_{X_1X_2|\theta=\tau}$ *represented by a virtual binary symmetric channel (BSC) shown in Figure 3.2, where* $p_\tau, q_\tau \leq 0.5,\ \tau \in \Theta$, *where* Θ *is a given finite set. For* $A = \{1\},\ B = \{1,2\}$, *and the probability of error distortion measure of Example 3.1 , the Bayesian USRDf reduces to*

$$R_{\{1\}}(\Delta) = \min_{\substack{\{\Delta_{\tau_1},\ \tau_1\in\Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}]\leq\Delta}} \max_{\tau_1\in\Theta_1} \Big(h(p_{\tau_1}) - h\Big(\frac{\Delta_{\tau_1} - q_{\tau_1}}{1-q_{\tau_1}}\Big)\Big),$$

for $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, *where*

$$\Delta_{\min} = \mathbb{E}[q_{\theta_1}], \quad \Delta_{\max} = \mathbb{E}[p_{\theta_1} + q_{\theta_1} - p_{\theta_1}q_{\theta_1}];$$

and $q_{\tau_1} = P_{X_2|X_1\theta_1}(0|1,\tau_1),\ \tau_1 \in \Theta_1$; *and the nonBayesian USRDf is*

$$R_{\{1\}}(\Delta) = \max_{\tau_1\in\Theta_1} \Big(h(p_{\tau_1}) - \min_{\tau\in\Lambda(\tau_1)} h\Big(\frac{\Delta - q_\tau}{1-q_\tau}\Big)\Big)$$

with

$$\Delta_{\min} = \max_{\tau \in \Theta} q_\tau \quad \textit{and} \quad \Delta_{\max} = \max_{\tau \in \Theta} (p_\tau + q_\tau - p_\tau q_\tau).$$

■

3.3.2 Independent Random Sampling

Turning to a k-IRS in (2.3), the freedom now given to the sampler to rove over all k-sized subsets in $\mathcal{A}_k$ engenders a partition Θ_2 of Θ_1 (and hence a finer partition of Θ) with smaller ambiguity atoms. Let $A_1, \ldots, A_{|\mathcal{A}_k|}$, where $|\mathcal{A}_k| = \binom{m}{k}$, be any fixed ordering of $\mathcal{A}_k$. Let Θ_2 be a partition of Θ consisting of ambiguity atoms, with each atom formed by τs with *identical (ordered) collections of marginal pmfs* $\left(P_{X_{A_i}|\theta=\tau},\ i = 1, \ldots, |\mathcal{A}_k|\right)$.

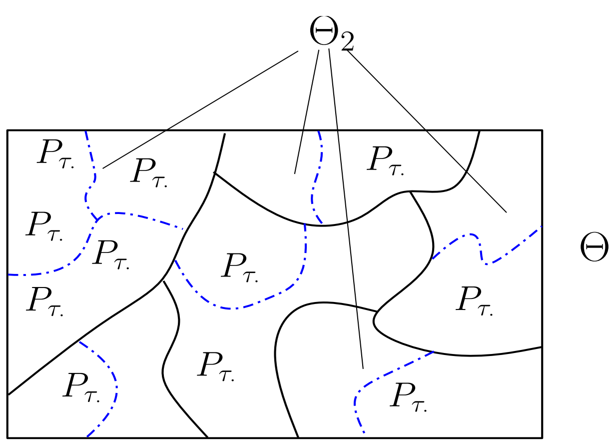

Figure 3.3: Ambiguity atoms for k-IRS: Θ_2 is a refinement of Θ_1

Clearly, Θ_2 is a refinement of Θ_1 (for any A_i). Indexing the elements of Θ_2 by τ_2, let θ_2 be a Θ_2-valued rv with pmf μ_{θ_2} derived from μ_θ. For each τ_2 in Θ_2, let

$\Lambda(\tau_2)$ be the collection of τs in the atom indexed by τ_2. In analogy with (3.7) and (3.9), we define counterparts in the Bayesian and nonBayesian settings as

$$\rho_\imath^{\mathcal{B}}(\delta, P_S, \tau_2) \triangleq \min_{\kappa_\imath^{\mathcal{B}}(\delta, P_S, \tau_2)} I(X_S \wedge Y_B | S, \theta_2 = \tau_2); \tag{3.16}$$

$$\rho_\imath^{n\mathcal{B}}(\delta, P_S, \tau_2) \triangleq \min_{\kappa_\imath^{n\mathcal{B}}(\delta, P_S, \tau_2)} I(X_S \wedge Y_B | S, \theta_2 = \tau_2), \tag{3.17}$$

where d_{τ_2} is defined as in (3.5) with $\theta_2 = \tau_2$ replacing $\theta_1 = \tau_1$, and

$$\begin{aligned}\kappa_\imath^{\mathcal{B}}(\delta, P_S, \tau_2) \triangleq \big\{ & P_{\theta X_{\mathcal{M}} S Y_B} = \mu_\theta P_{X_{\mathcal{M}}|\theta} P_S P_{Y_B|S X_S \theta_2} : \\ & \sum_{A \in \mathcal{A}_k} P_S(A) \mathbb{E}[d_{\tau_2}(X_A, Y_B) | S = A, \theta_2 = \tau_2] \leq \delta \big\},\end{aligned}$$

$$\begin{aligned}\kappa_\imath^{n\mathcal{B}}(\delta, P_S, \tau_2) \triangleq \big\{ & P_{X_{\mathcal{M}} S Y_B|\theta=\tau} = P_{X_{\mathcal{M}}|\theta=\tau} P_S P_{Y_B|S X_S, \theta_2=\tau_2} : \\ & \sum_{A \in \mathcal{A}_k} P_S(A) \mathbb{E}[d(X_B, Y_B) | S = A, \theta = \tau] \leq \delta, \ \tau \in \Lambda(\tau_2) \big\}.\end{aligned}$$

Theorem 3.2 *The Bayesian USRDf for a k-IRS is*

$$R_\imath^I(\Delta) = \min_{\substack{P_S, \ \{\Delta_{\tau_2}, \ \tau_2 \in \Theta_2\} \\ \mathbb{E}[\Delta_{\theta_2}] \leq \Delta}} \max_{\tau_2 \in \Theta_2} \rho_\imath^{\mathcal{B}}(\Delta_{\tau_2}, P_S, \tau_2), \qquad \Delta_{\min} \leq \Delta \leq \Delta_{\max}, \tag{3.18}$$

where

$$\Delta_{\min} = \min_{A \in \mathcal{A}_k} \mathbb{E}\big[\mathbb{E}[\min_{y_B \in \mathcal{Y}_B} d_{\theta_2}(X_A, y_B)|\theta_2]\big] \text{ and } \Delta_{\max} = \min_{A \in \mathcal{A}_k} \mathbb{E}\big[\min_{y_B \in \mathcal{Y}_B} \mathbb{E}[d_{\theta_2}(X_A, y_B)|\theta_2]\big].$$

The nonBayesian USRDf is

$$R_\imath^I(\Delta) = \min_{P_S} \max_{\tau_2 \in \Theta_2} \rho_\imath^{n\mathcal{B}}(\Delta, P_S, \tau_2), \qquad \Delta_{\min} \leq \Delta \leq \Delta_{\max}, \tag{3.19}$$

for

$$\Delta_{\min} = \min_{P_S} \max_{\tau_2 \in \Theta_2} \sum_{A \in \mathcal{A}_k} P_S(A) \min_{P_{Y_B|SX_S,\theta_2=\tau_2} = P_{Y_B|SX_{\mathcal{M}},\theta=\tau}} \max_{\tau \in \Lambda(\tau_2)} \mathbb{E}[d(X_B, Y_B)|S = A, \theta = \tau]$$

and

$$\Delta_{\max} = \max_{\tau_2 \in \Theta_2} \min_{y_B \in \mathcal{Y}_B} \max_{\tau \in \Lambda(\tau_2)} \mathbb{E}[d(X_B, y_B)|\theta = \tau].$$

Remarks: (i) In Bayesian and nonBayesian settings, the USRDf remains unchanged upon restricting the decoder to be uninformed (as in Definition 2.3), i.e., $R_\imath^U(\Delta) = R_\imath^I(\Delta)$. This is shown by means of the achievability proof of the theorem above in Section 3.4. Hence, hereafter the USRDf for a k-IRS is denoted simply by $R_\imath(\Delta)$.

(ii) For a k-IRS we restrict ourselves to the interesting case of $k < |B|$, for otherwise it would suffice to choose $S_t = B,\ t = 1, \ldots, n$.

(iii) Akin to a k-FS, the optimizing $P_S,\ \{\Delta^*_{\tau_2},\ \tau_2 \in \Theta_2\}$ in (3.18) has the following special property: for a given $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, for each $\tau_2 \in \Theta_2$, either

$$\rho_\imath^{\mathcal{B}}(\Delta_{\tau_2}^*, P_S, \tau_2) = \max_{\tilde{\tau}_2 \in \Theta_2} \rho_\imath^{\mathcal{B}}(\Delta_{\tilde{\tau}_2}^*, P_S, \tilde{\tau}_2)$$

or

$$\Delta_{\tau_2}^* = \sum_{A \in \mathcal{A}_k} P_S(A) \mathbb{E}[\min_{y_B \in \mathcal{Y}_B} d_{\tau_2}(X_A, y_B) | \theta_2 = \tau_2].$$

(iv) In general, a k-IRS will outperform a k-FS in two ways. First, the former enables a better approximation of θ in the form of θ_2 whereas the latter estimates $\theta_1 = \theta_1(\theta_2)$. Second, random sampling enables a "time-sharing" over various fixed-set samplers, that can outperform strictly the best fixed-set choice. Both these advantages of a k-IRS over fixed-set sampling are illustrated in Examples 3.3 and 3.4.

Example 3.3 *This example illustrates that a k-IRS can perform strictly better than the best k-FS. For $\mathcal{M} = B = \{1, 2\}$, and $\mathcal{X}_i = \mathcal{Y}_i = \{0, 1\}$, $i = 1, 2$, consider a DMMS with $P_{X_1 X_2 | \theta = \tau} = P_{X_1 | \theta = \tau} P_{X_2 | \theta = \tau}$ where*

$$P_{X_1|\theta}(0|\tau) = 1 - p_\tau, \quad P_{X_2|\theta}(0|\tau) = 1 - q_\tau, \ \tau \in \Theta,$$

and $0 < p_\tau,\ q_\tau < 0.5$. Under the distortion measure $d(x_B, y_B) = \mathbb{1}(x_1 \neq y_1) + \mathbb{1}(x_2 \neq$

y_2), for a k-FS with $k = 1$, the Bayesian USRDf for sampling set $A = \{1\}$ is

$$R_{\{1\}}(\Delta) = \min_{\substack{\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}] \leq \Delta}} \max_{\tau_1 \in \Theta_1} \Big(h(p_{\tau_1}) - h\big(\Delta_{\tau_1} - q_{\tau_1}\big) \Big), \quad \mathbb{E}[q_\theta] \leq \Delta \leq \mathbb{E}[p_\theta + q_\theta]$$

where $q_{\tau_1} = \mathbb{E}[q_\theta | \theta_1 = \tau_1]$, and the nonBayesian USRDf is

$$R_{\{1\}}(\Delta) = \max_{\tau_1 \in \Theta_1} \left(h(p_{\tau_1}) - \min_{\tau \in \Lambda(\tau_1)} h\big(\Delta - q_\tau\big) \right), \quad \max_{\tau \in \Theta} q_\tau \leq \Delta \leq \max_{\tau \in \Theta} (p_\tau + q_\tau).$$

Turning to a k-IRS with $k = 1$, clearly, $\Theta_2 = \Theta$. For a k-IRS the Bayesian USRDf is

$$R_\imath(\Delta) = \min_{\substack{P_S,\ \{\Delta_\tau,\ \tau \in \Theta\} \\ \mathbb{E}[\Delta_\theta] \leq \Delta}} \max_{\tau \in \Theta} \min_{\substack{\Delta_{1\tau},\ \Delta_{2\tau} \\ P_S(\{1\})\Delta_{1\tau} + P_S(\{2\})\Delta_{2\tau} \leq \Delta_\tau}} I, \quad \min\{\mathbb{E}[p_\theta], \mathbb{E}[q_\theta]\} \leq \Delta \leq \mathbb{E}[p_\theta + q_\theta]$$

and the nonBayesian USRDf is

$$R_\imath(\Delta) = \min_{P_S} \max_{\tau \in \Theta} \min_{\substack{\Delta_{1\tau},\ \Delta_{2\tau} \\ P_S(\{1\})\Delta_{1\tau} + P_S(\{2\})\Delta_{2\tau} \leq \Delta}} I, \tag{3.20}$$

for $\min_{0 \leq \alpha \leq 1} \max_{\tau \in \Theta} (\alpha p_\tau + (1-\alpha) q_\tau) \leq \Delta \leq \max_{\tau \in \Theta} (p_\tau + q_\tau)$, and where I equals

$$P_S(\{1\})\big(h(p_\tau) - h\big(\Delta_{1\tau} - q_\tau\big)\big) + P_S(\{2\})\big(h(q_\tau) - h\big(\Delta_{2\tau} - p_\tau\big)\big).$$

An analytical comparison of the USRDfs shows the strict superiority of the k-IRS over the k-FS, as seen – for instance – by the lower values of $\Delta_{\min}$ for the former.

■

Example 3.4 *In Example 3.3, assume that*

$$p_\tau \geq q_\tau, \quad \tau \in \Theta.$$

For a k-FS with $k = 1$, the nonBayesian USRDf is

$$\begin{aligned} R_{\{1\}}(\Delta) &= \max_{\tau_1 \in \Theta_1} \left(h(p_{\tau_1}) - \min_{\tau \in \Lambda(\tau_1)} h\big(\Delta - q_\tau\big) \right), \\ R_{\{2\}}(\Delta) &= \max_{\tau_1 \in \Theta_1} \left(h(q_{\tau_1}) - \min_{\tau \in \Lambda(\tau_1)} h\big(\Delta - p_\tau\big) \right). \end{aligned} \tag{3.21}$$

Now, observe that for each $\tau \in \Theta$

$$h(p_\tau) - h(\delta - q_\tau) \leq h(q_\tau) - h(\delta - p_\tau)$$

holds for $p_\tau \leq \delta \leq p_\tau + q_\tau$. Thus, for a k-IRS with $k = 1$, the nonBayesian USRDf in (3.20) *simplifies to*

$$R_\imath(\Delta) = \max_{\tau \in \Theta} h(p_\tau) - h(\Delta - q_\tau)$$

which is strictly smaller than the USRDf for the better k-FS in (3.21)*. The superior performance of the k-IRS is enabled by its ability to estimate simultaneously both $P_{X_1|\theta}$ and $P_{X_2|\theta}$ (and thereby $P_{X_1X_2|\theta}$); a k-FS can estimate only one of $P_{X_1|\theta}$ or $P_{X_2|\theta}$.* ■

3.3.3 Memoryless Random Sampling

Lastly, for a k-MRS in (2.2), the ability of the sampler to depend instantaneously on the current realization of the DMMS enables an encoder with access to the sampler output to distinguish among all the pmfs in $\mathcal{P}$. Accordingly, for a k-MRS, Θ itself serves as the counterpart of the partitions Θ_1 (for a k-FS) and Θ_2 for a k-IRS. For a rv U with fixed pmf P_U on some finite set $\mathcal{U}$, and for fixed $P_{S|X_{\mathcal{M}}U}$, we define the counterparts of (3.16) and (3.17) as

$$\rho_m^{\mathcal{B}}(\delta, P_U, P_{S|X_{\mathcal{M}}U}, \tau) \triangleq \min_{\kappa_m^{\mathcal{B}}(\delta, P_U, P_{S|X_{\mathcal{M}}U}, \tau)} I(X_S \wedge Y_B | S, U, \theta = \tau), \tag{3.22}$$

and

$$\rho_m^{n\mathcal{B}}(\delta, P_U, P_{S|X_{\mathcal{M}}U}, \tau) \triangleq \min_{\kappa_m^{n\mathcal{B}}(\delta, P_U, P_{S|X_{\mathcal{M}}U}, \tau)} I(X_S \wedge Y_B | S, U, \theta = \tau), \tag{3.23}$$

where the minimization in (3.22) and (3.23), in effect, is with respect to $P_{Y_B|SX_SU\theta}$ and the sets of (constrained) pmfs are

$$\kappa_m^{\mathcal{B}}(\delta, P_U, P_{S|X_{\mathcal{M}}U}, \tau) \triangleq \{P_{\theta U X_{\mathcal{M}} S Y_B} = \mu_\theta P_U P_{X_{\mathcal{M}}|\theta} P_{S|X_{\mathcal{M}}U} P_{Y_B|SX_SU\theta} : \mathbb{E}[d(X_B, Y_B)|\theta = \tau] \leq \delta\},$$

and

$$\kappa_m^{n\mathcal{B}}(\delta, P_U, P_{S|X_{\mathcal{M}}U}, \tau) \triangleq \{P_{UX_{\mathcal{M}}SY_B|\theta=\tau} = P_U P_{X_{\mathcal{M}}|\theta=\tau} P_{S|X_{\mathcal{M}}U} P_{Y_B|SX_SU,\theta=\tau} :$$
$$\mathbb{E}[d(X_B, Y_B)|\theta = \tau] \leq \delta\}.$$

Here, U plays the role of a "time-sharing" rv, as will be seen below.

Theorem 3.3 *For a k-MRS, the Bayesian USRDf is*

$$R_m^I(\Delta) = \min_{\substack{P_U, P_{S|X_{\mathcal{M}}U}, \{\Delta_\tau,\ \tau\in\Theta\} \\ \mathbb{E}[\Delta_\theta]\leq\Delta}} \max_{\tau\in\Theta} \rho_m^{\mathcal{B}}(\Delta_\tau, P_U, P_{S|X_{\mathcal{M}}U}, \tau), \quad \Delta_{\min} \leq \Delta \leq \Delta_{\max} \tag{3.24}$$

where

$$\begin{aligned} \Delta_{\min} &= \min_{P_{S|X_{\mathcal{M}}}} \mathbb{E}\Big[\min_{y_B\in\mathcal{Y}_B} \mathbb{E}\big[d\big(X_B, y_B\big)\big|S, X_S, \theta\big]\Big] \text{ and} \\ \Delta_{\max} &= \min_{P_{S|X_{\mathcal{M}}}} \mathbb{E}\Big[\min_{y_B\in\mathcal{Y}_B} \mathbb{E}\big[d\big(X_B, y_B\big)\big|S, \theta\big]\Big]. \end{aligned} \tag{3.25}$$

The nonBayesian USRDf is

$$R_m^I(\Delta) = \min_{P_U, P_{S|X_{\mathcal{M}}U}} \max_{\tau\in\Theta} \rho_m^{n\mathcal{B}}(\Delta, P_U, P_{S|X_{\mathcal{M}}U}, \tau), \quad \Delta_{\min} \leq \Delta \leq \Delta_{\max}, \tag{3.26}$$

where

$$\Delta_{\min} = \min_{P_{S|X_{\mathcal{M}}}} \max_{\tau\in\Theta} \mathbb{E}\Big[\min_{y_B\in\mathcal{Y}_B} \mathbb{E}\big[d(X_B, y_B)|S, X_S, \theta=\tau\big]\Big|\theta=\tau\Big] \tag{3.27}$$

and

$$\Delta_{\max} = \min_{P_{S|X_{\mathcal{M}}}} \max_{\tau\in\Theta} \sum_{A_i\in\mathcal{A}_k} P_{S|\theta}(A_i|\tau) \min_{y_B\in\mathcal{Y}_B} \mathbb{E}\big[d(X_B, y_B)|S=A_i, \theta=\tau\big]. \tag{3.28}$$

It suffices to take $|\mathcal{U}| \leq 2|\Theta| + 1$.

In (3.25), (3.27) and (3.28), it is readily seen that conditionally deterministic samplers (2.19) attain the minima in $\Delta_{\min}$ and $\Delta_{\max}$. In Chapter 2, among the class of memoryless random samplers, deterministic samplers were seen to be optimal for every feasible distortion level. In the universal setting, too, such deterministic samplers are seen to be optimal for every $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$.

Theorem 3.3 is equivalent to

Proposition 3.1 *For a k-MRS, the Bayesian USRDf is*

$$R_m^I(\Delta) = \min_{\substack{P_U,\ \delta_w,\ \{\Delta_\tau,\ \tau\in\Theta\} \\ \mathbb{E}[\Delta_\theta]\leq\Delta}} \max_{\tau\in\Theta} \rho_m^{\mathcal{B}}(\Delta_\tau, P_U, \delta_w, \tau), \quad \Delta_{\min} \leq \Delta \leq \Delta_{\max} \tag{3.29}$$

with $\Delta_{\min}$ *and* $\Delta_{\max}$ *as in* (3.25)*, and the nonBayesian USRDf is*

$$R_m^I(\Delta) = \min_{P_U,\ \delta_w} \max_{\tau\in\Theta} \rho_m^{n\mathcal{B}}(\Delta, P_U, \delta_w, \tau), \quad \Delta_{\min} \leq \Delta \leq \Delta_{\max} \tag{3.30}$$

with $\Delta_{\min}$ *and* $\Delta_{\max}$ *as in* (3.27) *and* (3.28)*, respectively. It suffices if* $|\mathcal{U}| \leq 2|\Theta|+1$.

Remark: The proposition above and Theorem 2.3 involve a similar set of techniques. In Appendix A, we provide a unified proof for the theorem above and Theorem 2.3.

The achievability proof of Theorem 3.3, by dint of Proposition 3.1, will use a deterministic sampler based on the minimizing w from (3.29) or (3.30).

Example 3.5 *This example compares the USRDfs for a k-MRS and a k-IRS and is an adaptation of Example 3.2 above (and also of Example 2.3). Consider Example 3.2 with* $q_\tau = 0.5$ *for every* $\tau \in \Theta$*, whereby* $P_{X_1X_2|\theta=\tau} = P_{X_1|\theta=\tau}P_{X_2|\theta=\tau}$*. Clearly,*

$\Theta_2 = \Theta$. *For a k-IRS, the Bayesian USRDf is*

$$R_\imath(\Delta) = \min_{\substack{\{\Delta_\tau, \tau\in\Theta\} \\ \mathbb{E}[\Delta_\theta]\leq \Delta}} \max_{\tau\in\Theta} \Big(h(0.5) - h\Big(\frac{\Delta_\tau - p_\tau}{1-p_\tau}\Big)\Big)$$
$$= h(0.5) - h\Big(\frac{\Delta - p}{1-p}\Big)$$

for $0 \leq \Delta \leq p$, *where* $p = \mathbb{E}[p_\theta]$, *and the nonBayesian USRDf is*

$$R_\imath(\Delta) = h(0.5) - \min_{\tau\in\Theta} h\Big(\frac{\Delta - p_\tau}{1-p_\tau}\Big), \quad 0 \leq \Delta \leq \max_{\tau\in\Theta} p_\tau.$$

For a k-MRS, in $\rho_m^{\mathcal{B}}(\delta, P_U, P_{S|X_\mathcal{M}U}, \tau)$ *as well as* $\rho_m^{n\mathcal{B}}(\delta, P_U, P_{S|X_\mathcal{M}U}, \tau)$, P_U = *a point-mass and*

$$P_{S|X_\mathcal{M}U}(s|x_\mathcal{M}, u) = P_{S|X_\mathcal{M}}(s|x_\mathcal{M}) = \begin{cases} 1, & s = 1,\ x_\mathcal{M} = 00 \ \mathit{or}\ 11 \\ 1, & s = 2,\ x_\mathcal{M} = 01 \ \mathit{or}\ 10 \\ 0, & \mathit{otherwise} \end{cases}$$

are uniformly optimal for all $0 \leq \delta \leq p_\tau$ *and for all* $\tau \in \Theta$. *Then, the Bayesian USRDf is*

$$R_m^I(\Delta) = \min_{\substack{\{\Delta_\tau, \tau\in\Theta\} \\ \mathbb{E}[\Delta_\theta]\leq \Delta}} \max_{\tau\in\Theta} \big(h(p_\tau) - h(\Delta_\tau)\big), \quad 0 \leq \Delta \leq p$$

and the nonBayesian USRDf is

$$R_m^I(\Delta) = \max_{\tau\in\Theta} h(p_\tau) - h(\Delta), \quad 0 \leq \Delta \leq \max_{\tau\in\Theta} p_\tau.$$

Clearly, in both the Bayesian and nonBayesian settings $R_m^I(\Delta) < R_\imath(\Delta)$.

■

In closing this section, standard properties of the USRDf for the fixed-set sampler, k-IRS and k-MRS in the Bayesian and nonBayesian settings are summarized below, with the proof provided in Appendix D.1.

Lemma 3.1 *The right-sides of* (3.10), (3.11), (3.18), (3.19), (3.24) *and* (3.26) *are finite-valued, decreasing, convex, continuous functions of* $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$.

3.4 Proofs of Main Results

3.4.1 Achievability Proofs

Our achievability proofs emphasize the Bayesian setting. Counterpart proofs in the nonBayesian setting use similar sets of ideas, and so we limit ourselves to pointing out only the distinctions between these and their Bayesian brethren. In the Bayesian setting, the achievability proofs successively build upon each other according to increasing complexity of the sampler, and are presented in the order: fixed-set sampler, k-IRS and k-MRS.

A common theme in the achievability proofs for a k-FS, a k-IRS and a k-MRS involves forming estimates $\widehat{\tau}_1$ of the underlying τ_1 in Θ_1, $\widehat{\tau}_2$ of τ_2 in Θ_2 and $\widehat{\tau}$ of τ in Θ, respectively. The assumed finiteness of Θ enables $\widehat{\tau}_1$ or $\widehat{\tau}_2$ to be conveyed rate-free to the decoder. Codes for achieving USRDf at a prescribed distortion level Δ are chosen from among fixed-set sampling rate distortion codes for τ_1s in Θ_1

or from among IRS codes for τ_2s in Θ_2 or from among MRS codes for τs in Θ. Such codes, in the Bayesian setting, correspond to appropriate distortion thresholds that, in effect, average to yield a distortion level Δ; in the nonBayesian setting, a suitable "worst-case" distortion must not exceed Δ. A chosen code corresponds to an estimate $\widehat{\tau}_1$, $\widehat{\tau}_2$ or $\widehat{\tau}$.

A mainstay of our achievability proofs is the existence of sampling rate distortion codes with fixed-set sampling for a DMMS with *known* pmf, as in Proposition 2.1.

Theorem 3.1: Considering first the Bayesian setting, observe that

$$
\begin{aligned}
\Delta_{\min} &= \min_{\theta, X_{\mathcal{M}} -\!\!\circ\!\!- \theta_1, X_A -\!\!\circ\!\!- Y_B} \mathbb{E}[d(X_B, Y_B)] \\
&= \min_{\theta, X_{\mathcal{M}} -\!\!\circ\!\!- \theta_1, X_A -\!\!\circ\!\!- Y_B} \mathbb{E}[\mathbb{E}[d(X_B, Y_B)|X_A, \theta_1]] \\
&= \min_{\theta, X_{\mathcal{M}} -\!\!\circ\!\!- \theta_1, X_A -\!\!\circ\!\!- Y_B} \mathbb{E}[d_{\theta_1}(X_A, Y_B)] \quad \text{by (3.5)} \\
&= \mathbb{E}[\mathbb{E}[\min_{y_B \in \mathcal{Y}_B} d_{\theta_1}(X_A, y_B)|\theta_1]]
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta_{\max} &= \min_{\substack{\theta, X_{\mathcal{M}} -\!\!\circ\!\!- \theta_1, X_A -\!\!\circ\!\!- Y_B \\ P_{X_A Y_B|\theta_1=\tau_1} = P_{X_A|\theta_1=\tau_1} P_{Y_B|\theta_1=\tau_1}, \tau_1 \in \Theta_1}} \mathbb{E}[d(X_B, Y_B)] \\
&= \mathbb{E}\Big[\min_{P_{X_A Y_B|\theta_1} = P_{X_A|\theta_1} P_{Y_B|\theta_1}} \mathbb{E}[d_{\theta_1}(X_A, Y_B)|\theta_1] \Big] \\
&= \mathbb{E}\Big[\min_{y_B \in \mathcal{Y}_B} \mathbb{E}[d_{\theta_1}(X_A, y_B)|\theta_1] \Big].
\end{aligned}
$$

Now, consider a partition Θ_1 of Θ as in Section 3.3. Based on the sampler output X_A^n, the encoder forms an ML estimate of θ_1 as

$$\widehat{\tau}_{1,n} = \widehat{\tau}_{1,n}(X_A^n) \triangleq \arg\max_{\tau_1 \in \Theta_1} P_{X_A^n|\theta_1}(X_A^n|\tau_1).$$

For each τ_1 in Θ_1, observe that $\{X_{At}\}_{t=1}^{\infty}$ is a DMMS with pmf $P_{\tau_1} \triangleq P_{X_A|\theta_1=\tau_1}$. The sequence of ML estimates $\{\widehat{\tau}_{1,n}\}_n$ converges in P_{τ_1}-probability to τ_1, so that for every $\epsilon > 0$ and τ_1 in Θ_1, there exists an $N_1(\epsilon, \tau_1)$ such that

$$P_{\tau_1}(\widehat{\tau}_{1,n} \neq \tau_1) = P_{\tau_1}(\widehat{\tau}_{1,n}(X_A^n) \neq \tau_1) \leq \frac{\epsilon}{2d_{\max}}, \quad n \geq N_1(\epsilon, \tau_1),$$

where $d_{\max} = \max\limits_{x_B \in \mathcal{X}_B,\ y_B \in \mathcal{Y}_B} d(x_B, y_B)$. By the finiteness of Θ_1, there exists an $N(\epsilon)$ such that simultaneously for all $\tau_1 \in \Theta_1$,

$$P_{\tau_1}(\widehat{\tau}_{1,n} \neq \tau_1) \leq \frac{\epsilon}{2d_{\max}}, \quad n \geq N(\epsilon)$$

and consequently

$$P(\widehat{\tau}_{1,n} \neq \theta_1) = \sum_{\tau_1 \in \Theta_1} \mu_{\theta_1}(\tau_1) P_{\tau_1}(\widehat{\tau}_{1,n} \neq \tau_1) \leq \frac{\epsilon}{2d_{\max}}, \quad n \geq N(\epsilon). \qquad (3.31)$$

For a fixed $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, let $\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\}$ yield the minimum in (3.10). For each τ_1 in Θ_1, for the DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with pmf $P_{X_{\mathcal{M}}|\theta_1=\tau_1}$ and distortion measure d_{τ_1}, there exists as in the Proof of Proposition 2.1 – with $P_{X_{\mathcal{M}}} = P_{X_{\mathcal{M}}|\theta_1=\tau_1}$

and $d_A = d_{\tau_1}$ – a fixed-set sampling rate distortion code $(f_{\tau_1}, \varphi_{\tau_1})$, $f_{\tau_1} : \mathcal{X}_A^n \to \{1, \ldots, J\}$ and $\varphi_{\tau_1} : \{1, \ldots, J\} \to \mathcal{Y}_B^n$ of rate $\frac{1}{n} \log J \le \max_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) + \frac{\epsilon}{2} = R_A(\Delta) + \frac{\epsilon}{2}$ and with expected distortion

$$\mathbb{E}[d_{\tau_1}(X_A^n, \varphi_{\tau_1}(f_{\tau_1}(X_A^n)))|\theta_1 = \tau_1] \le \Delta_{\tau_1} + \frac{\epsilon}{2}$$

for all $n \ge N_2(\epsilon, \tau_1)$.

A code (f, φ), with f taking values in $\mathcal{J} \triangleq \{1, \ldots, |\Theta_1|\} \times \{1, \ldots, J\}$ is constructed as follows. Order (in any manner) the elements of Θ_1. The encoder f, dictated by the estimate $\widehat{\tau}_{1,n}$, is

$$f(x_A^n) \triangleq (\widehat{\tau}_{1,n}(x_A^n), f_{\widehat{\tau}_{1,n}}(x_A^n)), \quad x_A^n \in \mathcal{X}_A^n.$$

The decoder is

$$\varphi(\widehat{\tau}_{1,n}, j) \triangleq \varphi_{\widehat{\tau}_{1,n}}(j), \quad (\widehat{\tau}_{1,n}, j) \in \mathcal{J}.$$

The rate of the code is

$$\frac{1}{n} \log |\mathcal{J}| = \frac{1}{n} \log |\Theta_1| + \frac{1}{n} \log J \le R_A(\Delta) + \epsilon, \tag{3.32}$$

for all n large enough, by the finiteness of Θ_1.

The code (f, φ) is seen to satisfy

$$
\begin{aligned}
\mathbb{E}[d(X_B^n, \varphi(f(X_A^n)))] &\le \mathbb{E}[\mathbb{1}(\widehat{\tau}_{1,n} = \theta_1) d(X_B^n, \varphi_{\widehat{\tau}_{1,n}}(f_{\widehat{\tau}_{1,n}}(X_A^n)))] + P(\widehat{\tau}_{1,n} \ne \theta_1) d_{\max} \\
&= \mathbb{E}[\mathbb{1}(\widehat{\tau}_{1,n} = \theta_1) d(X_B^n, \varphi_{\theta_1}(f_{\theta_1}(X_A^n)))] + P(\widehat{\tau}_{1,n} \ne \theta_1) d_{\max} \\
&\le \mathbb{E}\big[d(X_B^n, \varphi_{\theta_1}(f_{\theta_1}(X_A^n)))\big] + P(\widehat{\tau}_{1,n} \ne \theta_1) d_{\max}. \qquad (3.33)
\end{aligned}
$$

The first term on the right-side of (3.33) is

$$
\begin{aligned}
&\mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} d\big(X_{Bt}, (\varphi_{\theta_1}(f_{\theta_1}(X_A^n)))_t\big)\Big] \\
&= \mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\big[d(X_{Bt}, (\varphi_{\theta_1}(f_{\theta_1}(X_A^n)))_t) | X_A^n, \theta\big]\Big] \\
&= \mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\big[d(X_{Bt}, (\varphi_{\theta_1}(f_{\theta_1}(X_A^n)))_t) | X_{At}, \theta\big]\Big], \qquad \text{since } P_{X_{\mathcal{M}}^n|\theta} = \prod_{t=1}^{n} P_{X_{\mathcal{M}t}|\theta} \\
&= \mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\big[d(X_{Bt}, (\varphi_{\theta_1}(f_{\theta_1}(X_A^n)))_t) | X_{At}, \theta_1\big]\Big], \qquad \text{since } \theta -\!\circ\!- \theta_1 -\!\circ\!- X_A^n \\
&= \mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n} d_{\theta_1}\big(X_{At}, (\varphi_{\theta_1}(f_{\theta_1}(X_A^n)))_t\big)\Big], \qquad \text{by (3.5)} \\
&= \mathbb{E}[d_{\theta_1}(X_A^n, \varphi_{\theta_1}(f_{\theta_1}(X_A^n)))]. \qquad (3.34)
\end{aligned}
$$

Combining (3.33) and (3.34),

$$
\begin{aligned}
\mathbb{E}[d(X_B^n, \varphi(f(X_A^n)))] &\le \mathbb{E}[d_{\theta_1}(X_A^n, \varphi_{\theta_1}(f_{\theta_1}(X_A^n)))] + P(\widehat{\tau}_{1,n} \ne \theta_1) d_{\max} \\
&\le \mathbb{E}\left[\Delta_{\theta_1}\right] + \epsilon \le \Delta + \epsilon, \qquad (3.35)
\end{aligned}
$$

by (3.31) for all n large enough. Finally, we note that (3.32) and (3.35) hold simultaneously for all n large enough.

In the nonBayesian setting, the achievability proof follows by adapting the steps above with the following differences. For each τ_1 in Θ_1, a fixed-set sampling rate distortion code $(f_{\tau_1}, \varphi_{\tau_1})$ is chosen now with expected distortion $\mathbb{E}[d(X_B^n, \varphi_{\tau_1}(f_{\tau_1}(X_A^n)))|\theta = \tau] \leq \Delta + \frac{\epsilon}{2}$ for every τ in $\Lambda(\tau_1)$ and of rate $\frac{1}{n}\log|f_{\tau_1}| \leq R_A(\Delta) + \frac{\epsilon}{2}$, where $R_A(\Delta)$ is the nonBayesian USRDf for a fixed-set sampler.

■

Theorem 3.2: In the Bayesian setting, for a given $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, consider the P_S, $\{\Delta_{\tau_2},\ \tau_2 \in \Theta_2\}$ that attain the (outer) minimum in (3.18). For the corresponding minimizing $P_{Y_B|SX_S\theta_2}$ in (3.18) (by way of (3.16))

$$\max_{\tau_2\in\Theta_2} \rho_i^{\mathcal{B}}(\Delta_{\tau_2}, P_S, \tau_2) = \max_{\tau_2\in\Theta_2} \sum_{A_i\in\mathcal{A}_k} P_S(A_i) I(X_{A_i} \wedge Y_B | S = A_i, \theta_2 = \tau_2) \quad (3.36)$$

and let

$$\Delta_{A_i,\tau_2} \triangleq \mathbb{E}[d(X_B, Y_B)|S = A_i, \theta_2 = \tau_2], \quad A_i \in \mathcal{A}_k,\ \tau_2 \in \Theta_2.$$

The second expression in (3.36) suggests an achievability scheme using an IRS code (as in the proof of Theorem 2.1) governed by θ_2. Our achievability proof comprises two phases. In the first phase, an estimate $\widehat{\tau}_2$ of θ_2 is formed based on the output of a k-IRS that chooses each A_i in $\mathcal{A}_k$ repeatedly for N time instants. The second phase, of length n, entails choosing each $S_t = A_i$ repeatedly for $\approx nP_S(A_i)$ time instants and an IRS code governed by $\widehat{\tau}_2$ of expected distortion

$$\sum_i P_S(A_i)\Delta_{A_i,\widehat{\tau}_2}$$

is applied to the output of the sampler. This predetermined selection of sampling sets obviates the need for the decoder to be additionally informed.

Denote $|\mathcal{A}_k|$ by $M_k = \binom{m}{k}$. Fix $\epsilon > 0$ and $0 < \epsilon' < \epsilon$. In the first phase, a k-IRS is chosen to sample each $A_i \in \mathcal{A}_k$ over disjoint time-sets μ_i of length N. The union of the time-sets $\mu_i,\ i \in \mathcal{M}_k \triangleq \{1, \ldots, M_k\}$ is denoted by $\mu \triangleq \{1, \ldots, M_k N\}$. Based on the sampler output, an ML estimate $\widehat{\tau}_{2,N} = \widehat{\tau}_{2,N}(S^\mu, X_S^\mu)$ of θ_2 is formed with

$$P(\widehat{\tau}_{2,N} \neq \theta_2) \leq \frac{\epsilon'}{2d_{\max}}, \tag{3.37}$$

for $N \geq N_{\epsilon'}$, say.

In the second phase, we denote the next set of n time instants, i.e., $\{M_k N + 1, \ldots, M_k N + n\}$ simply by $\nu \triangleq \{1, \ldots, n\}$. Further, for each i in $\mathcal{M}_k$, define the time-sets $\nu_{A_i} \subset \nu$, made up of consecutive time instants, as

$$\nu_{A_i} = \Big\{t : \lceil n \sum_{j=1}^{i-1} P_S(A_j)\rceil + 1 \leq t \leq \lceil n \sum_{j=1}^{i} P_S(A_j)\rceil \Big\},$$

and note that the union of ν_{A_i}s is ν, and

$$\left|\frac{|\nu_{A_i}|}{n} - P_S(A_i)\right| \leq \frac{1}{n}, \quad i \in \mathcal{M}_k.$$

In this phase, the k-IRS is now chosen (deterministically) as follows:

$$S_t = s_t = A_i,\ t \in \nu_{A_i},\ i \in \mathcal{M}_k.$$

For each DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with pmf $P_{X_{\mathcal{M}}|\theta_2=\tau_2}$, $\tau_2 \in \Theta_2$, and for each A_i in $\mathcal{A}_k$ and its corresponding distortion measure d_{τ_2}, there exists as in the proof of Proposition 2.1 – with $P_{X_{\mathcal{M}}} = P_{X_{\mathcal{M}}|\theta_2=\tau_2}$ and $d_A = d_{\tau_2}$ – a fixed-set sampling rate distortion code $(f_{A_i}^{\tau_2}, \varphi_{A_i}^{\tau_2})$, $f_{A_i}^{\tau_2} : \mathcal{X}_{A_i}^{\nu_{A_i}} \rightarrow \{1, \ldots, J_{A_i}^{\tau_2}\}$ and $\varphi_{A_i}^{\tau_2} : \{1, \ldots, J_{A_i}^{\tau_2}\} \rightarrow \mathcal{Y}_B^{\nu_{A_i}}$ of rate $\frac{1}{|\nu_{A_i}|} \log J_{A_i}^{\tau_2} \leq I(X_{A_i} \wedge Y_B|S = A_i, \theta_2 = \tau_2) + \frac{\epsilon'}{4}$ (cf. (3.36)) and with

$$\mathbb{E}\Big[d_{\tau_2}\big(X_{A_i}^{\nu_{A_i}}, \varphi_{A_i}^{\tau_2}(f_{A_i}^{\tau_2}(X_{A_i}^{\nu_{A_i}}))\big)\big|\theta_2 = \tau_2\Big] \leq \Delta_{A_i,\tau_2} + \frac{\epsilon'}{2},$$

for all $|\nu_{A_i}| \geq N_{A_i}(\epsilon', \tau_2)$. Note that

$$\sum_{\tau_2 \in \Theta_2} \mu_{\theta_2}(\tau_2) \sum_{i=1}^{M_k} P_S(A_i)\Delta_{A_i,\tau_2} \leq \Delta$$

and

$$\sum_{i=1}^{M_k} P_S(A_i) I(X_{A_i} \wedge Y_B|S = A_i, \theta_2 = \tau_2) \leq R_\iota(\Delta)$$

for every τ_2 in Θ_2.

Consider a (composite) code (f, φ_S) as follows. Denote $n' \triangleq |\mu| + |\nu| = M_k N + n$, and the encoder f consisting of a concatenation of encoders is defined by

$$f(s^{n'}, x^{n'}) \triangleq \left(\widehat{\tau}_{2,N}, f_{A_1}^{\widehat{\tau}_{2,N}}(x_{A_1}^{\nu_{A_1}}), \ldots, f_{A_{M_k}}^{\widehat{\tau}_{2,N}}(x_{A_{M_k}}^{\nu_{A_{M_k}}})\right).$$

The decoder φ_S, which is aware of the predetermined sequence of sampling sets, is defined by

$$\varphi_S(s^{n'}, \widehat{\tau}_{2,N}, j_1, \ldots, j_{M_k}) = \varphi_S(\widehat{\tau}_{2,N}, j_1, \ldots, j_{M_k}) \triangleq \big(\underbrace{y_1, \ldots, y_1}_{\text{first phase}}, \underbrace{\varphi_{A_1}^{\widehat{\tau}_{2,N}}(j_1), \ldots, \varphi_{A_{M_k}}^{\widehat{\tau}_{2,N}}(j_{M_k})}_{\text{second phase}}\big),$$

for each encoder output $(\widehat{\tau}_{2,N}, j_1, \ldots, j_{M_k})$, where $y_1 \in \mathcal{Y}_{\mathcal{M}}$ is an arbitrary symbol. Clearly, $|\Theta_2| \times \max\limits_{\tau_2 \in \Theta_2} \prod\limits_{i=1}^{M_k} J_{A_i}^{\tau_2}$ indices would suffice to describe all possible encoder outputs.

The rate of the code is

$$\begin{aligned}
\frac{1}{n'} \log|\Theta_2| + \max_{\tau_2 \in \Theta_2} \frac{1}{n'} \sum_{i=1}^{M_k} \log J_{A_i}^{\tau_2} &\leq \max_{\tau_2 \in \Theta_2} \sum_{i=1}^{M_k} \frac{|\nu_{A_i}|}{n} \frac{1}{|\nu_{A_i}|} \log J_{A_i}^{\tau_2} + \frac{1}{n'} \log|\Theta_2| \\
&\leq \max_{\tau_2 \in \Theta_2} \sum_{i=1}^{M_k} \left(P_S(A_i) + \frac{1}{n}\right)\left(I(X_{A_i} \wedge Y_B | S = A_i, \theta_2 = \tau_2) + \frac{\epsilon'}{4}\right) \\
&\quad + \frac{1}{n'} \log|\Theta_2|
\end{aligned}$$

$$\leq \max_{\tau_2 \in \Theta_2} \sum_{i=1}^{M_k} P_S(A_i) I(X_{A_i} \wedge Y_B | S = A_i, \theta_2 = \tau_2) + \epsilon'$$

$$< R_\imath(\Delta) + \epsilon, \tag{3.38}$$

where the previous inequality holds for all n large enough. Denoting the output of the decoder by $Y_B^{n'} \triangleq \varphi_S(f(S^{n'}, X_S^{n'}))$

$$\mathbb{E}[d(X_B^{n'}, Y_B^{n'})] = \frac{1}{n'} \mathbb{E}\Big[\sum_{t \in \mu} d(X_{Bt}, Y_{Bt}) + \sum_{t \in \nu} \big(\mathbb{1}(\widehat{\tau}_{2,N} \neq \theta_2) d(X_{Bt}, Y_{Bt}) \\ + \mathbb{1}(\widehat{\tau}_{2,N} = \theta_2) d(X_{Bt}, Y_{Bt})\big)\Big]. \tag{3.39}$$

The first two terms on the right-side of (3.39) are

$$\mathbb{E}\Big[\frac{1}{n'} \sum_{t \in \mu} d(X_{Bt}, Y_{Bt}) + \frac{\mathbb{1}(\widehat{\tau}_{2,N} \neq \theta_2)}{n'} \sum_{t \in \nu} d(X_{Bt}, Y_{Bt})\Big] \leq \frac{M_k N d_{\max}}{n'} + \frac{\epsilon'}{2}, \tag{3.40}$$

by (3.37) for N large enough, and the last term on the right-side of (3.39) is

$$\begin{aligned}
&\mathbb{E}\Big[\frac{\mathbb{1}(\widehat{\tau}_{2,N} = \theta_2)}{n'} \sum_{t \in \nu} d(X_{Bt}, Y_{Bt})\Big] \\
&\qquad \leq \sum_{i=1}^{M_k} \frac{|\nu_{A_i}|}{n} \mathbb{E}\Big[\mathbb{1}(\widehat{\tau}_{2,N} = \theta_2) d\big(X_B^{\nu_{A_i}}, \varphi_{A_i}^{\widehat{\tau}_{2,N}}(f_{A_i}^{\widehat{\tau}_{2,N}}(X_{A_i}^{\nu_{A_i}}))\big)\Big] \\
&\qquad \leq \sum_{i=1}^{M_k} \frac{|\nu_{A_i}|}{n} \mathbb{E}\big[d\big(X_B^{\nu_{A_i}}, \varphi_{A_i}^{\theta_2}(f_{A_i}^{\theta_2}(X_{A_i}^{\nu_{A_i}}))\big)\big] \\
&\qquad = \sum_{i=1}^{M_k} \frac{|\nu_{A_i}|}{n} \mathbb{E}\big[d_{\theta_2}\big(X_A^{\nu_{A_i}}, \varphi_{A_i}^{\theta_2}(f_{A_i}^{\theta_2}(X_{A_i}^{\nu_{A_i}}))\big)\big] \\
&\qquad \leq \sum_{i=1}^{M_k} \Big(P_S(A_i) + \frac{1}{n}\Big) \mathbb{E}\big[\Delta_{A_i, \theta_2} + \frac{\epsilon'}{2}\big]
\end{aligned}$$

$$\leq \Delta + \frac{\epsilon'}{2} + \frac{1}{n}\sum_{i=1}^{M_k}\mathbb{E}[\Delta_{A_i,\theta_2}] + \frac{M_k}{n}\frac{\epsilon'}{2}. \tag{3.41}$$

From (3.39)-(3.41), we have

$$\mathbb{E}[d(X_B^{n'}, Y_B^{n'})] \leq \Delta + \epsilon, \tag{3.42}$$

for n and N large enough. Finally, we note that (3.38) and (3.42) hold simultaneously for all n and N large enough. Remark (i) following Theorem 3.2 is now immediate by the choice of codes with "uninformed" decoder in the proof above.

For the nonBayesian setting, achievability follows by adapting the proof above in a manner similar to that for a k-FS in Theorem 3.1.

■

Theorem 3.3: The achievability proof relies on the deterministic sampler justified by Proposition 3.1. In the Bayesian setting, for a given $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, let $P_U,\ P_{S|X_{\mathcal{M}}U} = \delta_w,\ \{\Delta_\tau,\ \tau \in \Theta\}$ attain the minimum in (3.29). For the corresponding minimizing $P_{Y_B|SX_SU\theta}$ in (3.22), the right-side of (3.29) is

$$\max_{\tau\in\Theta} \rho_m^{\mathcal{B}}(\Delta_\tau, P_U, \delta_w, \tau) = \max_{\tau\in\Theta}\sum_{u\in\mathcal{U}} P_U(u) I(X_S \wedge Y_B|S, U = u, \theta = \tau) \tag{3.43}$$

and we set

$$\Delta_{A_i,u,\tau} \triangleq \mathbb{E}[d(X_B, Y_B)|S = A_i, U = u, \theta = \tau], \quad A_i \in \mathcal{A}_k,\ \tau \in \Theta,\ u \in \mathcal{U}.$$

Our achievability proof uses a k-MRS in two distinct modes. First, a deterministic k-MRS is chosen so as to form an estimate $\widehat{\tau}$ of θ from the sampler output. Next, for each $U = u$, a suitable deterministic k-MRS is chosen in accordance with $w(x_{\mathcal{M}}, u)$, and an MRS code (as in the proof of Theorem 2.2) governed by $\widehat{\tau}$ of expected distortion

$$\overset{\sim}{\leq} \sum_{A_i} P_{S|U\theta}(A_i|u, \widehat{\tau})\Delta_{A_i,u,\widehat{\tau}}$$

is applied to the sampler output. Concatenation of such codes corresponding to various $u \in \mathcal{U}$ yields, in effect, time-sharing that serves to achieve (3.43). To simplify the notation, the conditioning on $U = u$ will be suppressed except when needed.

Fix $\epsilon > 0$ and $0 < \epsilon' < \epsilon$.

(i) We devise a deterministic k-MRS on a time-set μ, based on whose output an estimate $\widehat{\tau}_N = \widehat{\tau}_N(S^\mu, X_S^\mu) = \widehat{\tau}_N(S^\mu)$ of θ is formed with

$$P(\widehat{\tau}_N \neq \theta) \leq \frac{\epsilon'}{4d_{\max}}, \tag{3.44}$$

for $N \geq N_{\epsilon'}$. The estimate $\widehat{\tau}_N$ is formed from *only the sampling sequence* S^μ and thus is available to the encoder as well as the decoder. The k-MRS is chosen on the time-set μ, to signal the occurrences of each $x \in \mathcal{X}_{\mathcal{M}}$ to the encoder and decoder through S^μ above; for each $x \in \mathcal{X}_{\mathcal{M}}$, a distinct $A \in \mathcal{A}_k$ is chosen. If $|\mathcal{A}_k| \geq |\mathcal{X}_{\mathcal{M}}|$, a trivial one-to-one mapping from $\mathcal{X}_{\mathcal{M}}$ to $\mathcal{A}_k$ enables S^μ to determine $X_{\mathcal{M}}^\mu$, where S^μ is of length N, say. Then $\widehat{\tau}_N$ is taken to be the ML estimate of θ based on $X_{\mathcal{M}}^\mu$, which satisfies (3.44).

When $|\mathcal{A}_k| < |\mathcal{X}_{\mathcal{M}}|$, a k-MRS is chosen attuned variously to disjoint subsets of $\mathcal{X}_{\mathcal{M}}$, of size $|\mathcal{A}_k| - 1$, on corresponding disjoint time-sets μ_l of length N, $l = 1, \ldots, \left\lceil \frac{|\mathcal{X}_{\mathcal{M}}|}{|\mathcal{A}_k|-1} \right\rceil$, as follows. In each μ_l, the k-MRS signals the occurrence (or not) of $X_{\mathcal{M}t} = x$ in the l^{th}-subset of $\mathcal{X}_{\mathcal{M}}$ in a (deterministic) manner by choosing $|\mathcal{A}_k| - 1$ distinct sampling sets in $\mathcal{A}_k$; the nonoccurrence of symbols from this l^{th}-subset of $\mathcal{X}_{\mathcal{M}}$ is indicated by the remaining (dummy) sampling set in $\mathcal{A}_k$. We denote $\bigcup\limits_l \mu_l$ by μ. Finally, $\widehat{\tau}_N$ is taken as the ML estimate of θ based on the sampling sequence S^μ of length $\left\lceil \frac{|\mathcal{X}_{\mathcal{M}}|}{|\mathcal{A}_k|-1} \right\rceil N = N'$, say.

(ii) Next, for each $U = u$, a k-MRS is chosen according to $P_{S|X_{\mathcal{M}},U=u} = \delta_{w(\cdot,u)}$ for n time instants. Then, for a DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with pmf $P_{X_{\mathcal{M}}|\theta=\widehat{\tau}_N}$ an MRS code comprising a concatenation of fixed-set sampling rate distortion codes corresponding to the A_is in $\mathcal{A}_k$ is applied to the sampler output.

Denote the set of n time instants $\{N'+1, \ldots, N'+n\}$ simply by $\gamma \triangleq \{1, \ldots, n\}$. Define time-sets $\gamma_{S^n}(A_i) \triangleq \{t : 1 \leq t \leq n,\ S_t = A_i\},\ i \in \mathcal{M}_k$, and note that $\gamma_{S^n}(A_i)$s cover γ, i.e.,

$$\gamma = \bigcup_{A_i \in \mathcal{A}_k} \gamma_{S^n}(A_i).$$

Denote the set of the first $\max\{\lceil (nP_{S|\theta}(A_i|\widehat{\tau}_N)) - \epsilon' \rceil, 0\}$ time instants in each $\gamma_{S^n}(A_i)$ by ν_{A_i} (suppressing the dependence on $\widehat{\tau}_N$). Defining the (typical) set for each τ in

Θ

$$\mathcal{T}^{(n)}(\epsilon',\tau) \triangleq \left\{ s^n \in \mathcal{A}_k^n : \left| \frac{|\gamma_{s^n}(A_i)|}{n} - P_{S|\theta}(A_i|\tau) \right| \leq \epsilon',\ i \in \mathcal{M}_k \right\},$$

we have that

$$P(S^\gamma \notin \mathcal{T}^{(n)}(\epsilon',\widehat{\tau}_N)) = P(S^\gamma \notin \mathcal{T}^{(n)}(\epsilon',\widehat{\tau}_N),\ \widehat{\tau}_N = \theta) + P(S^\gamma \notin \mathcal{T}^{(n)}(\epsilon',\widehat{\tau}_N),\ \widehat{\tau}_N \neq \theta)$$

$$\leq \frac{\epsilon'}{2d_{\max}} \tag{3.45}$$

for all n large enough.

As in the Proof of Proposition 2.1, for *each* DMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with pmf $P_{X_{\mathcal{M}}|S=A_i,\theta=\tau}$, $i \in \mathcal{M}_k$, $\tau \in \Theta$, there exists a code $(f_{A_i}^\tau, \varphi_{A_i}^\tau)$, $f_{A_i}^\tau : \mathcal{X}_{A_i}^{\nu_{A_i}} \to \{1,\ldots,J_{A_i}^\tau\}$ and $\varphi_{A_i}^\tau : \{1,\ldots,J_{A_i}^\tau\} \to \mathcal{Y}_B^{\nu_{A_i}}$ of rate

$$\frac{1}{|\nu_{A_i}|} \log J_{A_i}^\tau \leq I(X_{A_i} \wedge Y_B | S = A_i, \theta = \tau) + \frac{\epsilon'}{2} \tag{3.46}$$

and with

$$\mathbb{E}\left[d\big(X_B^{\nu_{A_i}}, \varphi_{A_i}^\tau(f_{A_i}^\tau(X_{A_i}^{\nu_{A_i}}))\big) \big| S^{\nu_{A_i}} = A_i^{\nu_{A_i}}, \theta = \tau \right] \leq \Delta_{A_i,\tau} + \frac{\epsilon'}{4} \tag{3.47}$$

for all $|\nu_{A_i}| \geq N_{A_i}(\epsilon',\tau)$. Such codes are considered for each $U = u$.

Consider a (composite) code (f, φ_S) as follows. Denoting $N' + n$ by n', an encoder f consisting of a concatenation of encoders is defined as

$$f(s^{n'}, x_s^{n'}) \triangleq \begin{cases} \left(f_{A_1}^{\widehat{\tau}_N}(x_{A_1}^{\nu_{A_1}}), \ldots, f_{A_{M_k}}^{\widehat{\tau}_N}(x_{A_{M_k}}^{\nu_{A_{M_k}}})\right), & s^\gamma \in \mathcal{T}^{(n)}(\epsilon', \widehat{\tau}_N) \\ (1, \ldots, 1), & s^\gamma \notin \mathcal{T}^{(n)}(\epsilon', \widehat{\tau}_N). \end{cases}$$

For $t = 1, \ldots, n'$, and each encoder output $(j_1, \ldots, j_{M_k})$, the decoder φ_S, which can recover the estimate $\widehat{\tau}_N$ from its knowledge of the sampling sequence $S^{n'} = s^{n'}$, is given by

$$\left(\varphi_S(s^{n'}, j_1, \ldots, j_{M_k})\right)_t \triangleq \begin{cases} \left(\varphi_{A_i}^{\widehat{\tau}_N}(j_i)\right)_t, & s^\gamma \in \mathcal{T}^{(n)}(\epsilon', \widehat{\tau}_N) \text{ and } t \in \nu_{A_i},\ i \in \mathcal{M}_k \\ y_1, & \text{otherwise}, \end{cases}$$

where y_1 is a fixed but arbitrary symbol in $\mathcal{Y}_{\mathcal{M}}$.

Finally, for N and n large enough, the codes (f, φ_S) corresponding to each $U = u$ are concatenated so as to effect the time-sharing prescribed by P_U, in a standard manner. We claim that the rate of the resulting code is

$$\begin{aligned} &\overset{\sim}{\leq} \max_{\tau \in \Theta} \sum_{u \in \mathcal{U}} P_U(u) \sum_{A_i \in \mathcal{A}_k} P_{S|U\theta}(A_i|u, \tau) I(X_{A_i} \wedge Y_B | S = A_i, U = u, \theta = \tau) + \epsilon' \\ &\overset{\sim}{\leq} R_m^I(\Delta) + \epsilon, \end{aligned}$$

using (3.46) and the expected distortion is

$$\overset{\sim}{\leq} \mathbb{E}[\Delta_{S,U,\theta}] + \epsilon$$

$$\stackrel{\sim}{\leq} \Delta + \epsilon,$$

from (3.44), (3.45), (3.47) and the definition of $\Delta_{A_i,u,\tau}$. The proof of the claim above is in Appendix C.1. ■

3.4.2 Unified Converse Proof

In contrast with the achievability proofs, we present a unified converse proof for Theorems 3.3, 3.2 and 3.1 according to successive weakening of the sampler, viz. k-MRS, k-IRS and fixed-set sampler. We begin with the technical Lemma 3.2 that is used subsequently in the converse proof.

Lemma 3.2 *Let finite-valued rvs C, D^n, E^n, F^n, be such that (D_t, E_t), $t = 1, \ldots, n$, are conditionally mutually independent given C, i.e.,*

$$P_{D^n E^n|C} = \prod_{t=1}^{n} P_{D_t E_t|C} \tag{3.48}$$

and satisfy

$$C, D^n \multimap E^n \multimap F^n. \tag{3.49}$$

For any function $g(C)$ of C, such that

$$E^n \multimap g(C) \multimap C \quad \text{and} \quad P_{E^n|g(C)} = \prod_{t=1}^{n} P_{E_t|g(C)}, \tag{3.50}$$

it holds that

$$C, D_t -\!\!\circ\!\!- g(C), E_t -\!\!\circ\!\!- F_t, \quad t = 1, \ldots, n. \tag{3.51}$$

Proof: First, from (3.49), we have

$$\begin{aligned} 0 = I(C, D^n \wedge F^n | E^n) &= I\big(C \wedge F^n | E^n\big) + I\big(D^n \wedge F^n | E^n, C\big) \\ &= I\big(C, g(C) \wedge F^n | E^n\big) + I\big(D^n \wedge F^n | E^n, C\big) \\ &\geq I(C \wedge F^n | E^n, g(C)) + I(D^n \wedge F^n | E^n, C). \end{aligned} \tag{3.52}$$

Now, the second term on the right-side of (3.52) is

$$\begin{aligned} 0 = I(D^n \wedge F^n | E^n, C) &= H(D^n | E^n, C) - H(D^n | E^n, F^n, C) \\ &= \sum_{t=1}^{n} \big(H(D_t | E_t, C) - H(D_t | D^{t-1}, E^n, F^n, C)\big), \text{ by (3.48)} \\ &\geq \sum_{t=1}^{n} \big(H(D_t | E_t, C) - H(D_t | E_t, F_t, C)\big) \\ &= \sum_{t=1}^{n} I(D_t \wedge F_t | E_t, C). \end{aligned} \tag{3.53}$$

Next, the first part of (3.50) along with (3.52) implies that

$$\begin{aligned} 0 &= I\big(C \wedge E^n | g(C)\big) + I\big(C \wedge F^n | E^n, g(C)\big) \\ &= I\big(C \wedge E^n, F^n | g(C)\big), \end{aligned}$$

and hence

$$I\big(C \wedge E_t, F_t | g(C)\big) = 0, \quad t = 1, \ldots, n. \tag{3.54}$$

Now, by (3.53) and (3.54), for $t = 1, \ldots, n$,

$$I\big(C, D_t \wedge F_t | E_t, g(C)\big) = I\big(C \wedge F_t | E_t, g(C)\big) + I\big(D_t \wedge F_t | E_t, C\big) = 0,$$

which is the claim (3.51).

■

Converse: In the Bayesian setting, we provide first a converse proof for Theorem 3.3, which is then refashioned to give converse proofs for Theorems 3.2 and 3.1.

Let $(\{P_{S_t|X_{\mathcal{M}t}\theta} = P_{S_t|X_{\mathcal{M}t}}\}_{t=1}^{\infty}, f, \varphi_S)$ be an n-length k-MRS block code of rate R and with decoder output $Y_B^n = \varphi_S(S^n, f(S^n, X_S^n))$ satisfying $\mathbb{E}[d(X_B^n, Y_B^n)] \leq \Delta$. The hypothesis of Lemma 3.2 is met with $C = \theta,\ D^n = X_{\mathcal{M}}^n,\ E^n = (S^n, X_S^n),\ F^n = Y_B^n$ and $g(\theta) = \theta$, since

$$P_{X_{\mathcal{M}}^n S^n|\theta} = P_{X_{\mathcal{M}}^n|\theta} P_{S^n|X_{\mathcal{M}}^n} = \prod_{t=1}^{n} P_{X_{\mathcal{M}t}|\theta} P_{S_t|X_{\mathcal{M}t}} = \prod_{t=1}^{n} P_{X_{\mathcal{M}t}S_t|\theta}, \tag{3.55}$$

while

$$\theta, X_{\mathcal{M}}^n \multimap S^n, X_S^n \multimap Y_B^n$$

holds by code construction. Also, (3.55) implies, upon summing over all realizations of $X_{S^c}^n$, that

$$P_{S^n X_S^n|\theta} = \prod_{t=1}^{n} P_{S_t X_{S_t}|\theta}. \tag{3.56}$$

Then the claim of the lemma implies that

$$\theta, X_{\mathcal{M}t} -\!\circ\!- \theta, S_t, X_{S_t} -\!\circ\!- Y_{Bt}, \quad t = 1, \dots, n. \tag{3.57}$$

Let Δ_τ denote $\mathbb{E}[d(X_B^n, Y_B^n)|\theta = \tau] = \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}[d(X_{Bt}, Y_{Bt})|\theta = \tau]$ for each τ in Θ and note that $\mathbb{E}[\Delta_\theta] \leq \Delta$. For every τ in Θ, the following holds:

$$\begin{aligned}
R = \frac{1}{n}\log|f| &\geq \frac{1}{n}H(f(S^n, X_S^n)|\theta = \tau) \geq \frac{1}{n}H(f(S^n, X_S^n)|S^n, \theta = \tau) \\
&\geq \frac{1}{n}H(\varphi_S(S^n, f(S^n, X_S^n))|S^n, \theta = \tau) = \frac{1}{n}H(Y_B^n|S^n, \theta = \tau) \\
&= \frac{1}{n}I(X_S^n \wedge Y_B^n|S^n, \theta = \tau) \\
&= \frac{1}{n}\sum_{t=1}^{n}\big(H(X_{S_t}|S^n, X_S^{t-1}, \theta = \tau) - H(X_{S_t}|S^n, X_S^{t-1}, Y_B^n, \theta = \tau)\big) \\
&\geq \frac{1}{n}\sum_{t=1}^{n}\big(H(X_{S_t}|S^n, X_S^{t-1}, \theta = \tau) - H(X_{S_t}|S_t, Y_{Bt}, \theta = \tau)\big) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(H(X_{S_t}|S_t, \theta = \tau) - H(X_{S_t}|S_t, Y_{Bt}, \theta = \tau)\right), \qquad \text{by (3.56)} \\
&= \frac{1}{n}\sum_{t=1}^{n} I(X_{S_t} \wedge Y_{Bt}|S_t, \theta = \tau).
\end{aligned} \tag{3.58}$$

By (3.57),

$$\Big(\big(\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[d(X_{Bt},Y_{Bt})|\theta=\tau],\frac{1}{n}\sum_{t=1}^{n}I(X_{S_t}\wedge Y_{Bt}|S_t,\theta=\tau)\big),\tau\in\Theta\Big)$$

lies in the convex hull of

$$\mathcal{C}\triangleq\Big\{\big((\mathbb{E}[d(X_B,Y_B)|\theta=\tau],I(X_S\wedge Y_B|S,\theta=\tau)),\tau\in\Theta\big):$$
$$P_{\theta X_{\mathcal{M}}SY_B}=\mu_\theta P_{X_{\mathcal{M}}|\theta}P_{S|X_{\mathcal{M}}}P_{Y_B|SX_S\theta}\Big\}\ \subset\mathbb{R}^{2|\Theta|}.$$

By the Carathéodory Theorem [35], every point in the convex hull of $\mathcal{C}$ can be represented as a convex combination of at most $2|\Theta|+1$ elements in $\mathcal{C}$. The corresponding pmfs are indexed by the values of a rv U with

$$P_{U\theta X_{\mathcal{M}}SY_B}=P_U\mu_\theta P_{X_{\mathcal{M}}|\theta}P_{S|X_{\mathcal{M}}U}P_{Y_B|SX_S\theta U},\tag{3.59}$$

where the pmf of U has support of size $\leq 2|\Theta|+1$. Then, in a standard manner, (3.58) leads to

$$R\geq\min_{\substack{P_{Y_B|SX_SU,\theta=\tau}\\ \mathbb{E}[d(X_B,Y_B)|\theta=\tau]\leq\Delta_\tau}}I(X_S\wedge Y_B|S,U,\theta=\tau)\tag{3.60}$$

$$=\rho_m^{\mathcal{B}}(\Delta_\tau,P_U,P_{S|X_{\mathcal{M}}U},\tau).\tag{3.61}$$

Now, (3.61) holds for every $\tau \in \Theta$, and hence

$$\begin{aligned} R &\geq \max_{\tau \in \Theta} \rho_m^{\mathcal{B}}(\Delta_\tau, P_U, P_{S|X_{\mathcal{M}}U}, \tau) \\ &\geq \min_{\substack{P_U, P_{S|X_{\mathcal{M}}U}, \{\Delta_\tau,\ \tau \in \Theta\} \\ \mathbb{E}[\Delta_\theta] \leq \Delta}} \max_{\tau \in \Theta} \rho_m^{\mathcal{B}}(\Delta_\tau, P_U, P_{S|X_{\mathcal{M}}U}, \tau) \\ &= R_m^I(\Delta) \end{aligned} \tag{3.62}$$

for $\Delta \geq \Delta_{\min}$.

Turning next to Theorems 3.2 and 3.1, an n-length k-IRS code or a fixed-set sampling block code can be viewed as restrictions of a k-MRS code. Specifically, in Theorem 3.2, for a k-IRS code of rate R with $P_{S_t},\ g(\theta) = \theta_2$ instead of $P_{S_t|X_{\mathcal{M}t}},\ g(\theta) = \theta$ (for a k-MRS), the hypothesis of Lemma 3.2 holds. Denote $\mathbb{E}[d(X_B^n, Y_B^n)|\theta_2 = \tau_2]$ by $\Delta_{\tau_2},\ \tau_2 \in \Theta_2$. Then, the pmfs in (3.59) satisfy

$$P_{U\theta X_{\mathcal{M}} S Y_B} = P_U \mu_\theta P_{X_{\mathcal{M}}|\theta} P_{S|U} P_{Y_B|S X_S \theta U}. \tag{3.63}$$

The counterpart of (3.60) is

$$\begin{aligned} R &\geq \min_{\substack{P_{Y_B|SX_SU,\theta_2=\tau_2} \\ \mathbb{E}[d(X_B,Y_B)|\theta_2=\tau_2] \leq \Delta_{\tau_2}}} I(X_S \wedge Y_B | S, U, \theta_2 = \tau_2) \\ &= \min_{\substack{P_{Y_B|SX_SU,\theta_2=\tau_2} \\ \mathbb{E}[d(X_B,Y_B)|\theta_2=\tau_2] \leq \Delta_{\tau_2}}} \sum_{A,u} P_S(A) P_{U|S}(u|A) I(X_A \wedge Y_B | S = A, U = u, \theta_2 = \tau_2), \end{aligned}$$

noting from (3.63) that $P_{U|S,\theta_2} = P_{U|S}$. Using the convexity of the mutual information terms above with respect to $P_{Y_B|SX_SU\theta_2}$, we get

$$
\begin{aligned}
R &\geq \min_{\substack{P_{Y_B|SX_SU,\theta_2=\tau_2} \\ \mathbb{E}[d(X_B,Y_B)|\theta_2=\tau_2]\leq\Delta_{\tau_2}}} \sum_A P_S(A) I(X_A \wedge Y_B | S = A, \theta_2 = \tau_2) \\
&= \rho_\imath^{\mathcal{B}}(\Delta_{\tau_2}, P_S, \tau_2). \qquad (3.64)
\end{aligned}
$$

Since (3.64) holds for every $\tau_2 \in \Theta_2$

$$
\begin{aligned}
R &\geq \max_{\tau_2\in\Theta_2} \rho_\imath^{\mathcal{B}}(\Delta_{\tau_2}, P_S, \tau_2) \\
&\geq \min_{\substack{P_S,\{\Delta_{\tau_2},\ \tau_2\in\Theta_2\} \\ \mathbb{E}[\Delta_{\theta_2}]\leq\Delta}} \max_{\tau_2\in\Theta_2} \rho_\imath^{\mathcal{B}}(\Delta_{\tau_2}, P_S, \tau_2) \\
&= R_\imath(\Delta),
\end{aligned}
$$

i.e., $R \geq R_\imath(\Delta),\ \Delta \geq \Delta_{\min}$, completing the converse proof of Theorem 3.2.

In a manner analogous to a k-IRS, in Theorem 3.1 for a fixed-set sampler the hypothesis of Lemma 3.2 holds with $P_{S_t} = \mathbb{1}(S_t = A),\ g(\theta) = \theta_1$. Defining $\Delta_{\tau_1} \triangleq \mathbb{E}[d(X_B^n, Y_B^n)|\theta_1 = \tau_1],\ \tau_1 \in \Theta_1$, the counterpart of the right-side of (3.62) reduces to $\max_{\tau_1\in\Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1)$. It then follows that

$$
R \geq \min_{\substack{\{\Delta_{\tau_1},\ \tau_1\in\Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}]\leq\Delta}} \max_{\tau_1\in\Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1), \qquad \Delta \geq \Delta_{\min}
$$

providing the converse proof for Theorem 3.1.

In the nonBayesian setting, the analog of Lemma 3.2 is obtained similarly with $C = c,\ g(C) = g(c)$, and (3.48)–(3.51) expressed in terms of appropriate conditional pmfs. The converse proofs for a k-MRS, k-IRS and k-FS are obtained as above but by excluding the outer minimizations over $\{\Delta_\tau, \tau \in \Theta\}$, $\{\Delta_{\tau_2}, \tau_2 \in \Theta_2\}$ and $\{\Delta_{\tau_1}, \tau_1 \in \Theta_1\}$, respectively.

■

3.5 Discussion

Our formulation of universality requires optimum sampling rate distortion performance when the "true" underlying pmf of the DMMS belongs to a finite family $\mathcal{P} = \{P_{X_{\mathcal{M}}|\theta=\tau},\ \tau \in \Theta\}$. The assumed finiteness of Θ affords two benefits in addition to mathematical ease: (i) simple proofs of estimator consistency uniformly over Θ_1, Θ_2 or Θ; and (ii) rate-free conveyance of corresponding estimates $\widehat{\tau}_1$, $\widehat{\tau}_2$ or $\widehat{\tau}$ to the decoder. While general extensions to the case when Θ is an infinite set (countable or uncountable) remain open, in Chapter 4 an extension to the case where Θ is uncountable is studied for a Gaussian memoryless multiple source.

Unlike for a k-IRS, the assumption in a k-MRS that the decoder is informed of the sampling sequence S^n plays an important role. Specifically, embedded information regarding $X_{\mathcal{M}}^n$ is conveyed implicitly to the decoder through S^n. Also, as a side-benefit, the decoder can replicate the estimate of θ formed by the encoder based on S^n alone, obviating the need for explicitly transmitting it. However, if the decoder were denied a knowledge of S^n, what is the USRDf? This question, too, remains unanswered.

Underlying our achievability proofs of Theorems 3.2 and 3.3 for a k-IRS and

k-MRS, are schemes for distribution-estimation based on (S^n, X_S^n). A distinguishing feature from classical estimation settings is the additional degree of (spatial) freedom in the choice of the sampling sequence S^n. This motivates questions of the following genre: How should S^n, consisting of (possibly different) k-sized subsets, be chosen to form "best" estimates of the underlying joint pmf? How does the degree of the allowed dependence of S^n on $X_{\mathcal{M}}^n$ affect estimator performance? For instance, our choice of sampling sequence and estimation procedure in the achievability proof of Theorem 3.3 is a simple starting point. How must we devise *efficient* sampling mechanisms to exploit an implicit embedding of DMMS realization in the sampler output? These questions are of independent interest in statistical learning theory.

Chapter 4

Gaussian Sampling Rate Distortion

4.1 Synopsis

In this chapter, we consider a Gaussian memoryless multiple source (GMMS) with m components with *joint pdf* known only to belong to a given convex compact family of *uncountable* pdfs. A *fixed* subset of k components of the GMMS are sampled at each time instant and compressed jointly with the objective of reconstructing all the components of the GMMS under a mean-squared error distortion criterion.

In Section 4.2, we describe our model for universality and define the notion of a universal sampling rate distortion function (USRDf) along the lines of Chapter 3, to study the tradeoffs among sampling, estimation of underlying pdf, compression rate and desired level of accuracy in reconstruction. In Section 4.3, we characterize first the USRDf for a GMMS when the pdf of the GMMS is known, i.e., its SRDf. Building on this, a single-letter characterization is provided for the USRDf in Bayesian and nonBayesian universal settings. Throughout this chapter, our results will highlight the structure of our optimal modular reconstruction mechanisms, wherein the overall reconstruction is performed in two steps - estimates for the sampled components are formed first, based on which the unsampled components are reconstructed. In Section 4.4, we present first the achievability proofs, with the achievability proof for known pdf setting serving as a stepping stone for the universal setting. A unified converse proof is presented thereafter.

4.2 Model

Denote $\mathcal{M} = \{1, \ldots, m\}$ and let

$$X_{\mathcal{M}} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}$$

be a $\mathbb{R}^m$-valued zero-mean (jointly) Gaussian random vector with a positive-definite covariance matrix. For a nonempty set $A \subseteq \mathcal{M}$ with $|A| = k$, we denote by X_A the random (column) vector $(X_i,\ i \in A)^T$, with values in $\mathbb{R}^k$. Denote n repetitions of X_A, with values in $\mathbb{R}^{nk}$, by $X_A^n = (X_i^n,\ i \in A)^T$. Each $X_i^n = (X_{i1}, \ldots, X_{in})^T,\ i \in A$, takes values in $\mathbb{R}^n$. Let $\mathbb{R}^m$ be the reproduction alphabet for $X_{\mathcal{M}}$.

Let $\Theta = \{\mathbf{\Sigma}_{\mathcal{M}\tau}\}_\tau$[†] be a set of $m \times m$-positive-definite matrices, and assume Θ to be convex and compact in the Euclidean topology on $\mathbb{R}^{m \times m}$. For instance, for

[†] Θ is a collection of covariance matrices indexed by τ. By an abuse of notation, we shall use τ to refer to the covariance matrix $\mathbf{\Sigma}_{\mathcal{M}\tau}$ itself.

$m = 2$,

$$\Theta = \left\{ \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad c_1 \le \sigma_1^2, \sigma_2^2 \le c_2,\ -d_1 \le r \le d_1 \right\},$$

with $0 < c_1 \le c_2$ and $0 \le d_1 < 1$. Hereafter, all covariance matrices under consideration will be taken as being positive-definite without explicit mention. We assume θ to be a Θ-valued rv with a pdf ν_θ that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m^2}$. We assume θ to be a Θ-valued rv with a pdf ν_θ that is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{m^2}$. Assume

$$\nu_\theta(\tau) > 0, \quad \tau \in \Theta,$$

and that $\nu_\theta(\tau)$ is continuous in τ. We consider a jointly Gaussian memoryless multiple source (GMMS) $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ consisting of i.i.d. repetitions of the rv $X_{\mathcal{M}}$ with pdf known only to the extent of belonging to the family of pdfs $\mathcal{P} = \{\nu_{X_{\mathcal{M}}|\theta=\tau} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathcal{M}\tau})^{\ddagger}, \tau \in \Theta\}$. Two settings are studied: in a Bayesian formulation, the pdf ν_θ is taken to be *known*, while in a nonBayesian formulation θ is an *unknown constant* in Θ.

Remark: Note that in contrast to Chapter 3, the pdf of the underlying GMMS is known only to lie in an *uncountable* family of pdfs.

In this chapter we focus on a k-fixed-set sampler as in (2.4) which, for a fixed $A \subseteq \mathcal{M}$ with $|A| = k$, samples X_{At} from $X_{\mathcal{M}t}$ for $t \ge 1$. The output of the k-FS,

‡Throughout this chapter, $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ is used to denote the pdf of a Gaussian rv with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$.

in effect, is simply $\{X_{At}\}_{t=1}^{\infty}$. For a k-FS and $n \geq 1$, an n-length block code is as in Definition 2.2 with a GMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ and k-FS instead of a DMMS and k-RS, respectively.

In this chapter, our objective is to reconstruct *all* the components of a GMMS from the compressed representations of the sampled GMMS components under a suitable distortion criterion with (single-letter) mean-squared error (MSE) distortion measure

$$||x_{\mathcal{M}} - y_{\mathcal{M}}||^2 = \sum_{i=1}^{m} (x_i - y_i)^2, \qquad x_{\mathcal{M}}, y_{\mathcal{M}} \in \mathbb{R}^m.$$

For threshold $\Delta \geq 0$, an n-length block code (f, φ) with k-FS will be required to satisfy one of the following $(|| \cdot ||^2, \Delta)$ distortion criterion depending on the setting.

(i) *Bayesian*: The *expected* distortion criterion is

$$\begin{aligned} \mathbb{E}\Big[\Big|\Big|X_{\mathcal{M}}^n - \varphi\big(f(X_A^n)\big)\Big|\Big|^2\Big] &\triangleq \mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n}\Big|\Big|X_{\mathcal{M}t} - \Big(\varphi\big(f(X_A^n)\big)\Big)_t\Big|\Big|^2\Big] \\ &= \mathbb{E}\Big[\mathbb{E}\Big[\frac{1}{n}\sum_{t=1}^{n}\Big|\Big|X_{\mathcal{M}t} - \Big(\varphi\big(f(X_A^n)\big)\Big)_t\Big|\Big|^2\Big|\theta\Big]\Big] \qquad (4.1) \\ &\leq \Delta. \end{aligned}$$

(ii) *NonBayesian*: The *peak* distortion criterion is

$$\sup_{\tau \in \Theta} \mathbb{E}\Big[\Big|\Big|X_{\mathcal{M}}^n - \varphi\big(f(X_A^n)\big)\Big|\Big|^2\Big|\theta = \tau\Big] \leq \Delta, \qquad (4.2)$$

where $\mathbb{E}[\cdot|\theta = \tau]$ denotes $\mathbb{E}_{\nu_{X_{\mathcal{M}}^n|\theta=\tau}}[\cdot]$.

For a GMMS, an achievable universal k-sample coding rate at distortion level Δ and *universal sampling rate distortion function* (USRDf) is defined along the lines of Definition 3.2, but with distortion criterion (4.1) and (4.2) instead of (3.2) and (3.3), respectively. For $|\Theta| = 1$, the USRDf is simply referred to as the sampling rate distortion function (SRDf).

Remarks: (i) The USRDf under (4.1) is no larger than that under (4.2).

(ii) When $|\Theta| = 1$, the underlying pdf of the GMMS is, in effect, known.

Below, we recall (Chapter 1, [36]) the definition of mutual information between two random variables.

Definition 4.1 *For real-valued rvs X and Y with a joint probability distribution μ_{XY}, the mutual information between the rvs X and Y is given by*

$$I(X \wedge Y) = \begin{cases} \mathbb{E}_{\mu_{XY}}\left[\log \frac{d\mu_{XY}}{d\mu_X \times d\mu_Y}(X,Y)\right], & \text{if } \mu_{XY} \ll \mu_X \times \mu_Y \\ \infty, & \text{otherwise}, \end{cases}$$

where $\mu_{XY} \ll \mu_X \times \mu_Y$ denotes that μ_{XY} is absolutely continuous with respect to $\mu_X \times \mu_Y$ and $\frac{d\mu_{XY}}{d\mu_X \times d\mu_Y}$ is the Radon-Nikodym derivative of μ_{XY} with respect to $\mu_X \times \mu_Y$.

4.3 Gaussian Sampling Rate Distortion function

We begin with a setting where the pdf of the GMMS is known and provide a (single-letter) characterization for the SRDf. Next, in a brief detour, we introduce an extension of GMMS, namely a Gaussian memoryless field (GMF) and show how the ideas developed for a GMMS can be used to characterize the SRDf for a GMF. Finally, building on the SRDf for a GMMS, a (single-letter) characterization of the USRDf is provided for a GMMS in the Bayesian and nonBayesian settings.

Throughout this chapter, a recurring structural property of our achievability proofs is this: it is optimal to reconstruct the *sampled* GMMS components first under a *(modified) weighted* MSE criterion with *reduced threshold* and then form deterministic (MMSE) estimates of the unsampled components based on the reconstruction of the former.

Before we present our first result, we recall that for a GMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with pdf $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathcal{M}})$ reconstructed under the MSE distortion criterion, the *standard* rate distortion function (RDf) is

$$\begin{aligned} R(\Delta) &= \min_{\substack{\mu_{X_{\mathcal{M}}Y_{\mathcal{M}}} \ll \mu_{X_{\mathcal{M}}} \times \mu_{Y_{\mathcal{M}}} \\ \mathbb{E}[||X_{\mathcal{M}} - Y_{\mathcal{M}}||^2] \leq \Delta}} I(X_{\mathcal{M}} \wedge Y_{\mathcal{M}}), \qquad 0 < \Delta \leq \sum_{i=1}^{m} \mathbb{E}[X_i^2] \qquad (4.3) \\ &= \frac{1}{2} \sum_{i=1}^{m} \left(\log \tfrac{\lambda_i}{\alpha} \right)^+, \qquad 0 < \Delta \leq \sum_{i=1}^{m} \mathbb{E}[X_i^2], \end{aligned}$$

where λ_is are the eigenvalues of $\mathbf{\Sigma}_{\mathcal{M}}$, and α is chosen to satisfy $\sum_{i=1}^{m} \min(\alpha, \lambda_i) = \Delta$.

4.3.1 Known Distribution

Starting with $|\Theta| = 1$, for a GMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with (known) pdf $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathcal{M}})$, our first result shows that the fixed-set SRDf $R_A(\Delta)$ for a GMMS is, in effect, the RDf of a GMMS $\{X_{At}\}_{t=1}^{\infty}$ with a weighted MSE distortion measure d_A and a reduced threshold; here $d_A : \mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}^+ \cup \{0\}$ is given by

$$d_A(x_A, y_A) \triangleq (x_A - y_A)^T \mathbf{G}_A (x_A - y_A), \qquad x_A,\ y_A \in \mathbb{R}^k$$

with

$$\mathbf{G}_A = \mathbf{I} + \mathbf{\Sigma}_A^{-1} \mathbf{\Sigma}_{AA^c} \mathbf{\Sigma}_{AA^c}^T \mathbf{\Sigma}_A^{-1}, \tag{4.4}$$

where $\mathbf{\Sigma}_{AA^c} = \mathbb{E}[X_A X_{A^c}^T]$.

Observe that the modified distortion measure d_A above depends on the rv corresponding to the sampled GMMS components (in contrast to the modified distortion measures (2.6), (3.5)). A similar structure was observed in Corollary 2.1, for the SRDf for a DMMS with the probability of error distortion measure.

Theorem 4.1 *For a GMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with pdf $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\mathcal{M}})$ and fixed $A \subseteq \mathcal{M}$, the SRDf is*

$$\begin{aligned} R_A(\Delta) &= \min_{\substack{\mu_{X_A Y_A} \ll \mu_{X_A} \times \mu_{Y_A} \\ \mathbb{E}[d_A(X_A, Y_A)] \le \Delta - \Delta_{\min,A}}} I\big(X_A \wedge Y_A\big), && \Delta_{\min,A} < \Delta \le \Delta_{\max} \quad (4.5) \\ &= \frac{1}{2} \sum_{i=1}^{k} \left(\log \tfrac{\lambda_i}{\alpha}\right)^+, && \Delta_{\min,A} < \Delta \le \Delta_{\max} \quad (4.6) \end{aligned}$$

where

$$\Delta_{\min,A} = \sum_{i \in A^c} \left(\mathbb{E}[X_i^2] - \mathbb{E}[X_i X_A^T] \mathbf{\Sigma}_A^{-1} \mathbb{E}[X_A X_i]\right), \quad \Delta_{\max} = \sum_{i \in \mathcal{M}} \mathbb{E}[X_i^2]$$

and λ_i*s are the eigenvalues of* $\mathbf{G}_A \mathbf{\Sigma}_A$*, and* α *is chosen to satisfy* $\sum_{i=1}^{k} \min(\alpha, \lambda_i) = \Delta - \Delta_{\min,A}$.

Comparing (4.5) with (4.3), it can be seen that (4.5) is, in effect, the RDf for a GMMS with weighted MSE distortion measure. In contrast to the RDf (4.3), in (4.5) the minimization involves only X_A (and not $X_{\mathcal{M}}$) under a weighted MSE criterion with reduced threshold level. For $k = m$, i.e., $A = \mathcal{M}$, however this reduces to the RDf (4.3). Also, for every feasible distortion level the SRDf for any $A \subset \mathcal{M}$ is no smaller than that with $A = \mathcal{M}$.

In Section 4.4, the achievability proof of the theorem above involves reconstructing the sampled components of the GMMS first, and then forming MMSE estimates for the unsampled components based on the former. Accordingly, in (4.5), the MSE in the reconstruction of the entire GMMS is captured jointly by the weighted MSE (with weight-matrix $\mathbf{G}_A$) in the reconstructions of the sampled components and the minimum distortion $\Delta_{\min,A}$.

The form of the SRDf in (4.5) suggests a modular reconstruction scheme wherein only the sampled GMMS is reconstructed first and then based on it, the unsampled GMMS is reconstructed. Also, observing that (4.5) is equivalent to the standard rate distortion function of a GMMS with a weighted MSE distortion measure enables us to provide an analytic expression for the SRDf using the standard reverse water-filling solution (4.6) [36]. An instance of this is shown in the example below.

Observing that (4.5) is equivalent to the RDf of a GMMS with a weighted

MSE distortion measure enables us to provide an analytic expression for the SRDf using the standard reverse water-filling solution (4.6) [36]. An instance of this is shown in the example below.

Example 4.1 *For a GMMS with a k-FS with $k = 1$, this example illustrates the effect of the choice of the sampling set on SRDf. Consider a GMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with covariance matrix $\mathbf{\Sigma}_{\mathcal{M}}$ given by*

$$\mathbf{\Sigma}_{\mathcal{M}} = \begin{pmatrix} \sigma_1^2 & r_{12}\sigma_1\sigma_2 & \cdots & r_{1m}\sigma_1\sigma_m \\ r_{21}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & r_{2m}\sigma_2\sigma_m \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1}\sigma_1\sigma_m & r_{m2}\sigma_2\sigma_m & \cdots & \sigma_m^2 \end{pmatrix},$$

where $r_{ij} = r_{ji},\ 1 \leq i, j \leq m$. For $A = \{j\},\ j = 1, \ldots, m$, we have

$$\mathbf{G}_{\{j\}}\mathbf{\Sigma}_{\{j\}} = \big(1 + \sum_{i \neq j} \frac{r_{ij}^2\sigma_i^2}{\sigma_j^2}\big)\sigma_j^2 = \sigma_j^2 + \sum_{i \neq j} r_{ij}^2\sigma_i^2$$

and hence from (4.6), the SRDf is

$$\begin{aligned} R_{\{j\}}(\Delta) &= \frac{1}{2}\log\left(\frac{\sigma_j^2 + \sum\limits_{i \neq j} r_{ij}^2\sigma_i^2}{\Delta - \Delta_{\min,\{j\}}}\right) \\ &= \frac{1}{2}\log\left(\frac{\sum\limits_{i=1}^{m}\sigma_i^2 - \Delta_{\min,\{j\}}}{\Delta - \Delta_{\min,\{j\}}}\right) \end{aligned}$$

for $\Delta_{\min,\{j\}} < \Delta \leq \sum\limits_{i=1}^{m}\sigma_i^2$, where $\Delta_{\min,\{j\}} = \sum\limits_{i \neq j}\sigma_i^2(1 - r_{ij}^2)$. Observe that every SRDf

$\rho_{\{j\}}(\Delta)$ *is a monotonically increasing function of* $\Delta_{\min,\{j\}}$ *and that the SRDfs are translations of each other and hence decrease at the same rate. Thus, the SRDf with the smallest* $\Delta_{\min,\{j\}}$ *is uniformly best among all fixed-set SRDfs. For* $k > 2$ *however, there may not be any* $A \subset \mathcal{M},\ |A| = k$, *whose fixed-set SRDf is uniformly best for all distortion levels.* ■

Before turning to the USRDf for a GMMS, the ideas involved in Theorem 4.1 are used to study sampling and lossy compression of a Gaussian field which affords greater flexibility in the choice of sampling set. While Gaussian fields have been studied extensively under different formulations, we consider a Gaussian memoryless field (GMF) as in [16], which is described next. In lieu of $\mathcal{M}$ and Gaussian rv $X_{\mathcal{M}}$ in Section 4.2, consider $I = [0,1] \subset \mathbb{R}$ and let $X_I = \{X_u,\ u \in I\}$ be a $\mathbb{R}^I \triangleq \{\mathbb{R},\ u \in I\}$-valued zero-mean Gaussian process† with a bounded covariance function $r(s_1, s_2) = \mathbb{E}[X_{s_1} X_{s_2}],\ s_1, s_2 \in I$, such that, for any finite $C \subset I$

$$\mathbb{E}[X_C X_C^T]$$

is a positive-definite matrix and

$$\int\limits_I \int\limits_I |r(u,v)| \, du \, dv < \infty.$$

A GMF‡ $\{X_{It}\}_{t=1}^{\infty}$ consists of i.i.d. repetitions of X_I. We consider a GMF sampled

†A Gaussian process on an interval $[0,1]$ means that any finite collection of rvs $(X_{s_1}, \ldots, X_{s_l}),\ s_i \in [0,1],\ i \in \{1, \ldots, l\},\ l \in \mathbb{N}$, are jointly Gaussian.

‡Extensive studies of memoryless repetitions of a Gaussian process exist, cf. [16], [14], under various terminologies.

finitely by a k-FS at $A \subset I$, with $|A| = k$, and with a reconstruction alphabet $\mathbb{R}^I$.

For a GMF with fixed-set sampler and MSE distortion measure

$$||x_I - y_I||^2 = \int\limits_I (x_u - y_u)^2 \, du, \qquad x_I, y_I \in \mathbb{R}^I, \tag{4.7}$$

the sampling rate distortion function is defined as in Definitions 2.2 and 2.3 with the decoder φ characterized by a collection of mappings $\varphi = \{\varphi_u\}_{u \in I}$ with

$$\varphi_u : \{1, \ldots, J\} \to \mathbb{R}^n, \quad u \in I.$$

Analogous to a GMMS, for a GMF sampled at $A = \{a_1, \ldots, a_k\}$, $0 \leq a_i \leq 1, \; i = 1, \ldots, k$, our next result shows that the SRDf is, in effect, the standard rate distortion function of a GMMS $\{X_{At}\}_{t=1}^{\infty}$ with a weighted MSE distortion measure with weight-matrix given by

$$\mathbf{G}_{A,I} = \mathbf{\Sigma}_A^{-1} \Big(\int\limits_I \mathbb{E}[X_A X_u] \mathbb{E}[X_u X_A^T] \, du \Big) \mathbf{\Sigma}_A^{-1}, \tag{4.8}$$

with $\int$ connoting element-wise integration. Note that for every $0 \leq s_1, s_2 \leq 1$, the integral

$$\int\limits_I r(u, s_1) r(u, s_2) \, du$$

exists and hence (4.8) is well-defined.

Proposition 4.1 *For a GMF $\{X_{It}\}_{t=1}^{\infty}$ with $A \subset I$, the SRDf is*

$$R_A(\Delta) = \min_{\substack{\mu_{X_A Y_A} \ll \mu_{X_A} \times \mu_{Y_A} \\ \mathbb{E}[(X_A - Y_A)^T \mathbf{G}_{A,I}(X_A - Y_A)] \leq \Delta - \Delta_{\min,A}}} I\big(X_A \wedge Y_A\big), \quad \Delta_{\min,A} < \Delta \leq \Delta_{\max} \quad (4.9)$$

$$= \frac{1}{2}\sum_{i=1}^{k}\Big(\log\tfrac{\lambda_i}{\alpha}\Big)^+, \qquad \Delta_{\min,A} < \Delta \leq \Delta_{\max} \quad (4.10)$$

where

$$\Delta_{\min,A} = \int_I \big(\mathbb{E}[X_u^2] - \mathbb{E}[X_u X_A^T]\mathbf{\Sigma}_A^{-1}\mathbb{E}[X_A X_u]\big)\, du \quad \text{and} \quad \Delta_{\max} = \int_I \mathbb{E}[X_u^2]\, du,$$

and λ_is are the eigenvalues of $\mathbf{G}_{A,I}\mathbf{\Sigma}_A$, and α satisfies $\sum_{i=1}^{k} \min(\alpha, \lambda_i) = \Delta - \Delta_{\min,A}$.

The SRDf for a GMF (4.9) and its equivalent form (4.10) can be seen as counterparts of (4.5) and (4.6), with (4.10) being the reverse water-filling solution for (4.9). As before, the expression (4.9) is the RDf of a GMMS with a weighted MSE distortion measure. In Section 4.4, an achievability proof for the proposition above is provided by adapting the ideas developed for Theorem 4.1; a converse proof for the proposition is provided involving a set of techniques different from the converse proof provided for Theorem 4.1.

In contrast to a GMMS with a discrete set $\mathcal{M}$, for a GMF, I being an interval affords greater flexibility in the choice of the sampling set allowing for a better understanding of the structural properties of the "best" sampling set. In contrast to Example 4.1 in the example below, considering a GMF with a stationary Gauss-Markov process, we show the structure of the optimal set for minimum distortion for $k > 2$ as well. In general, the optimal sampling set is a function of the threshold Δ.

Example 4.2 *Consider a GMF with a zero-mean, stationary Gauss-Markov process X_I over $I = [0,1]$ with covariance function*

$$r(s,u) = p^{|s-u|}, \quad 0 \le s, u \le 1,$$

and $0 < p < 1$. Note that the correlation between any two points in the interval depends only on the distance between them. For the Gauss-Markov process X_I, for any $0 \le u_1 < u_2 < \cdots < u_l \le 1,\ l > 2$, it holds that

$$X_{u_1} -\!\circ\!- X_{u_2} -\!\circ\!- \cdots -\!\circ\!- X_{u_l}. \tag{4.11}$$

For a k-FS with $k = 1$ and $A = \{a\},\ 0 \le a \le 1$,

$$\mathbf{G}_{\{a\},I} = 1 - \Delta_{\min,\{a\}}$$

and $\mathbb{E}[X_a^2] = 1$. In (4.10), the eigenvalue λ_1 is $\mathbf{G}_{\{a\},I}\mathbf{\Sigma}_{\{a\}} = 1 - \Delta_{\min,\{a\}}$ itself and hence, the SRDf is

$$R_{\{a\}}(\Delta) = \frac{1}{2}\log\frac{1 - \Delta_{\min,\{a\}}}{\Delta - \Delta_{\min,\{a\}}}$$

for $\Delta_{\min,\{a\}} < \Delta \leq 1$, *where*

$$\begin{aligned}\Delta_{\min,\{a\}} &= \int_0^a \Big(\mathbb{E}[X_u^2] - \frac{\mathbb{E}^2[X_u X_a]}{\mathbb{E}[X_a^2]}\Big)\, du + \int_a^1 \Big(\mathbb{E}[X_u^2] - \frac{\mathbb{E}^2[X_u X_a]}{\mathbb{E}[X_a^2]}\Big)\, du \\ &= \int_0^a \big(1 - p^{2(a-u)}\big)\, du + \int_a^1 \big(1 - p^{2(u-a)}\big)\, du \\ &= 1 - \frac{p^{2a} - 1 + p^{2(1-a)} - 1}{\ln p}.\end{aligned}$$

Note that the SRDf $R_{\{a\}}(\Delta)$ *is a monotonically increasing function of* $\Delta_{\min,\{a\}}$, *which in turn is a monotonically increasing function of* $|a - 0.5|$. *Thus,* $R_{\{0.5\}}(\Delta)$ *is uniformly best among all SRDfs* $R_{\{a\}}(\Delta)$, $0 \leq a \leq 1$, *for all distortion levels. Now, for a k-FS with* $k > 2$ *and* $A = \{a_1 = 0, a_2, \ldots, a_{k-1}, a_k = 1\}$, *with* $a_i \leq a_{i+1}$, $i = 1, \ldots, k-1$, *the minimum distortion* $\Delta_{\min,A}$ *admits a simple form*

$$\Delta_{\min,A} = 1 - \sum_{i=1}^{k-1} \gamma(a_{i+1} - a_i),$$

where $\gamma(a_{i+1} - a_i)$ *is according to*

$$\gamma(a) \triangleq \frac{1}{1 - p^{2a}} \Big(\frac{p^{2a}(1 - 2a \log p) - 1}{\log p}\Big), \quad 0 < a < 1.$$

The minimum reconstruction error $\Delta_{\min,A}$ *is the "sum" of the minimum error in reconstructing each segment* $[a_i, a_{i+1}]$ *of the GMF. Now, the Markov property of the field* (4.11) *implies that the minimum distortion in reproducing each* X_i, $i \in I$, *is determined by its nearest sampled points on the GMF and hence the minimum distortion in reconstructing each segment* $[a_i, a_{i+1}]$ *of the GMF is independent of the* *location of sampling points other than a_i, a_{i+1} and is given by*

$$(a_{i+1} - a_i) - \gamma(a_{i+1} - a_i).$$

The stationarity of the field means that this minimum distortion depends on the length $|a_{i+1} - a_i|$ alone. Observing that $\gamma(a)$ is a concave function of a over $(0, 1]$, $\Delta_{\min,A}$ above is seen to be minimized when $a_{i+1} - a_i = \frac{1}{k-1}$, $i = 1, \ldots, k-1$, i.e., when the sampling points are spaced uniformly. However, such a placement is not optimal for all distortion levels.

$\blacksquare$

4.3.2 Universal Setting

Turning to the universal setting with a GMMS, consider a set $\Theta_1 = \{\mathbf{\Sigma}_{A\tau},\ \tau \in \Theta\} \subset \mathbb{R}^{k^2}$ with $\tau_1 \in \Theta_1$ indexing the members of Θ_1, i.e., $\Theta_1 = \{\mathbf{\Sigma}_{A\tau_1}\}_{\tau_1}$ † . An encoder f associated with a k-FS observes X_A^n alone and cannot distinguish among jointly Gaussian pdfs in $\mathcal{P}$ that have the same marginal pdf $\nu_{X_A|\theta=\tau}$. Accordingly (and akin to [40]), consider a partition of Θ comprising "ambiguity" atoms, with each atom of the partition comprising τs with identical $\nu_{X_A|\theta=\tau}$, i.e., identical $\mathbf{\Sigma}_{A\tau}$ and for each $\tau_1 \in \Theta_1$, $\Lambda(\tau_1)$ is the collection of τs in the ambiguity atom indexed by τ_1, i.e.,

$$\mathbf{\Sigma}_{A\tau_1} \triangleq \mathbf{\Sigma}_{A\tau}, \quad \tau \in \Lambda(\tau_1).$$

† Θ_1 is the collection of covariance matrices $\mathbf{\Sigma}_{A\tau}$ indexed by τ_1 and by an abuse τ_1 will also be used to refer to $\mathbf{\Sigma}_{A\tau_1}$ itself.

Let θ_1 be a Θ_1-valued rv induced by θ. It is easy to see that Θ_1 and $\Lambda(\tau_1),\ \tau_1 \in \Theta_1$, are convex, compact subsets of $\mathbb{R}^{k^2}$ and the rv θ_1 admits a pdf ν_{θ_1} induced by ν_θ. In the Bayesian setting,

$$\nu_{X_A|\theta_1=\tau_1} = \nu_{X_A|\theta=\tau} = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{A\tau_1}), \qquad \tau \in \Lambda(\tau_1).$$

In the nonBayesian setting, in order to retain the same notation, we choose $\nu_{X_A|\theta_1=\tau_1}$ to be the right-side above.

Our characterization of the USRDf builds on the structure of the SRDf for a GMMS. Accordingly, in the Bayesian setting, consider the set of (constrained) probability measures

$$\kappa_A^{\mathcal{B}}(\delta,\tau_1) \triangleq \{\mu_{\theta X_{\mathcal{M}} Y_{\mathcal{M}}} :\ \theta, X_{\mathcal{M}} -\!\circ\!- \theta_1, X_A -\!\circ\!- Y_{\mathcal{M}},\ \ \mu_{X_A Y_{\mathcal{M}}|\theta_1=\tau_1} \ll \mu_{X_A|\theta_1=\tau_1} \times \mu_{Y_{\mathcal{M}}|\theta_1=\tau_1},$$

$$\mathbb{E}[||X_{\mathcal{M}} - Y_{\mathcal{M}}||^2|\theta_1 = \tau_1] \leq \delta\}$$

and (constraint) minimized mutual information

$$\rho_A^{\mathcal{B}}(\delta,\tau_1) \triangleq \min_{\kappa_A^{\mathcal{B}}(\delta,\tau_1)} I(X_A \wedge Y_{\mathcal{M}}|\theta_1 = \tau_1). \tag{4.12}$$

Correspondingly, in the nonBayesian setting, consider

$$\kappa_A^{n\mathcal{B}}(\delta,\tau_1) \triangleq \{\mu_{X_\mathcal{M}Y_\mathcal{M}|\theta=\tau}:\ \mu_{Y_\mathcal{M}|X_\mathcal{M},\theta=\tau}=\mu_{Y_\mathcal{M}|X_A,\theta_1=\tau_1},\ \mu_{X_AY_\mathcal{M}|\theta_1=\tau_1}\ll\mu_{X_A|\theta_1=\tau_1}\times\mu_{Y_\mathcal{M}|\theta_1=\tau_1},$$

$$\mathbb{E}[||X_\mathcal{M}-Y_\mathcal{M}||^2|\theta=\tau]\leq\delta,\ \tau\in\Lambda(\tau_1)\}$$

and

$$\rho_A^{n\mathcal{B}}(\delta,\tau_1) \triangleq \inf_{\kappa_A^{n\mathcal{B}}(\delta,\tau_1)} I(X_A\wedge Y_\mathcal{M}|\theta_1=\tau_1). \tag{4.13}$$

Remark: In (4.12) and (4.13), the minimization is with respect to the conditional measure $\mu_{Y_\mathcal{M}|X_A,\theta_1=\tau_1}$.

The minimized conditional mutual informations above will be a key ingredient in the characterization of USRDf. First, we show in the proposition below that (4.12) and (4.13) admit simpler forms involving rvs corresponding to the sampled components of the GMMS and their reconstruction alone. In the Bayesian setting, for each $\tau_1\in\Theta_1$, the mentioned simpler form involves a weighted MSE distortion measure $d_{A\tau_1}$ with weight-matrix $\mathbf{G}_{A,\tau_1}$, defined as in (4.4) with $\mathbf{\Sigma}_{AA^c}$ replaced by $\mathbb{E}[X_AX_{A^c}^T|\theta_1=\tau_1]$ and

$$d_{A\tau_1}(x_A,y_A)\triangleq(x_A-y_A)^T\mathbf{G}_{A,\tau_1}(x_A-y_A),\qquad x_A,y_A\in\mathbb{R}^k.$$

In the Bayesian setting, the modified distortion measure $d_{A\tau_1}$ plays a role similar to that of d_A.

Remark: Clearly, $\rho_A^{n\mathcal{B}}(\delta,\tau_1)$ is a nonincreasing function of $\delta > \Delta_{\min,A,\tau_1}$. Convexity of $\rho_A^{n\mathcal{B}}(\delta,\tau_1)$ can be shown as in [37], and convexity implies the continuity of $\rho_A^{n\mathcal{B}}(\delta,\tau_1)$. Now, to show the convexity, pick any $\delta_1,\ \delta_2 > \Delta_{\min,A,\tau_1}$ and $\epsilon > 0$. For $i = 1, 2$, let $\mu^i \in \kappa_A^{n\mathcal{B}}(\delta_i,\tau_1)$ be such that

$$I_{\mu^i}(X_A \wedge Y_{\mathcal{M}}|\theta_1 = \tau_1) \leq \rho_A^{n\mathcal{B}}(\delta_i) + \epsilon.$$

For $\alpha > 0$, by the standard convexity arguments, it can be seen that $\alpha\mu^1+(1-\alpha)\mu^2 \in \kappa_A^{n\mathcal{B}}(\alpha\delta_1 + (1-\alpha)\delta_2, \tau_1)$ and

$$I_{\alpha\mu^1+(1-\alpha)\mu^2}(X_A \wedge Y_{\mathcal{M}}|\theta_1 = \tau_1) \leq \alpha\rho_A^{n\mathcal{B}}(\delta_1) + (1-\alpha)\rho_A^{n\mathcal{B}}(\delta_2) + \epsilon. \tag{4.14}$$

Since (4.14) holds for any $\epsilon > 0$, in the limit, we have

$$\rho_A^{n\mathcal{B}}(\alpha\delta_1 + (1-\alpha)\delta_2) \leq \alpha\rho_A^{n\mathcal{B}}(\delta_1) + (1-\alpha)\rho_A^{n\mathcal{B}}(\delta_2).$$

Proposition 4.2 *For each $\tau_1 \in \Theta_1$, in the Bayesian setting*

$$\rho_A^{\mathcal{B}}(\delta,\tau_1) = \min_{\substack{\mu_{X_A Y_A|\theta_1=\tau_1} \ll \mu_{X_A|\theta_1=\tau_1} \times \mu_{Y_A|\theta_1=\tau_1} \\ \mathbb{E}[d_{A\tau_1}(X_A,Y_A)|\theta_1=\tau_1] \leq \delta - \Delta_{\min,A,\tau_1}}} I\big(X_A \wedge Y_A|\theta_1 = \tau_1\big) \tag{4.15}$$

for $\delta > \Delta_{\min,A,\tau_1}$*, where*

$$\Delta_{\min,A,\tau_1} = \mathbb{E}\big[\mathbb{E}\big[\min_{y_{A^c}\in\mathbb{R}^{m-k}} \sum_{i\in A^c}(X_i - y_i)^2|X_A, \theta_1 = \tau_1\big]\big|\theta_1 = \tau_1\big].$$

For each $\tau_1 \in \Theta_1$*, in the nonBayesian setting*

$$\rho_A^{n\mathcal{B}}(\delta, \tau_1) = \inf_{\mathbb{E}[||X_{\mathcal{M}}-Y_{\mathcal{M}}||^2|\theta=\tau]\leq\delta,\ \tau\in\Lambda(\tau_1)} I\big(X_A \wedge Y_A|\theta_1 = \tau_1\big), \quad \delta > \Delta_{\min,A,\tau_1}, \tag{4.16}$$

where the infimum in (4.16) *is over* $\mu_{Y_{\mathcal{M}}|X_{\mathcal{M}},\theta=\tau}$*, such that*

$$\mu_{Y_{\mathcal{M}}|X_{\mathcal{M}},\theta=\tau} = \mu_{Y_A|X_A,\theta_1=\tau_1} \times \mu_{Y_{A^c}|Y_A,\theta_1=\tau_1},\ \ \tau \in \Lambda(\tau_1), \quad \textit{and}$$

$$\mu_{X_AY_A|\theta_1=\tau_1} \ll \mu_{X_A|\theta_1=\tau_1} \times \mu_{Y_A|\theta_1=\tau_1}$$

and

$$\Delta_{\min,A,\tau_1} = \inf_{\mu_{Y_{A^c}|X_A,\theta=\tau}=\mu_{Y_{A^c}|X_A,\theta_1=\tau_1}} \max_{\tau\in\Lambda(\tau_1)} \sum_{i\in A^c} \mathbb{E}[(X_i - Y_i)^2|\theta = \tau].$$

Remark: From (4.15), notice that $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ is, in effect, the rate distortion function for a GMMS with pdf $\nu_{X_A|\theta_1=\tau_1}$ and weighted MSE distortion measure. Hence, the minimum in (4.15) and ergo that in (4.12) exist and the standard properties of a rate distortion function are applicable to $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ as well, i.e., $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ is a convex, nonincreasing, continuous function of $\delta > \Delta_{\min,A,\tau_1}$.

Theorem 4.2 *For a GMMS $\{X_{\mathcal{M}t}\}_{t=1}^{\infty}$ with fixed $A \subseteq \mathcal{M}$, the Bayesian USRDf is*

$$R_A(\Delta) = \min_{\substack{\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}] \le \Delta}} \max_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) \tag{4.17}$$

for $\Delta_{\min,A} < \Delta \le \Delta_{\max}$, where

$$\Delta_{\min,A} = \mathbb{E}\big[\mathbb{E}\big[\min_{y_{A^c} \in \mathbb{R}^{m-k}} \sum_{i \in A^c} (X_i - y_i)^2 | X_A, \theta_1\big]\big] \quad \textit{and} \quad \Delta_{\max} = \sum_{i \in \mathcal{M}} \mathbb{E}[X_i^2].$$

The nonBayesian USRDf is

$$R_A(\Delta) = \max_{\tau_1 \in \Theta_1} \rho_A^{n\mathcal{B}}(\Delta, \tau_1) \tag{4.18}$$

for $\Delta_{\min,A} < \Delta \le \Delta_{\max}$, where

$$\Delta_{\min,A} = \sup_{\tau_1 \in \Theta_1} \inf_{\mu_{Y_{A^c}|X_A,\theta=\tau} = \mu_{Y_{A^c}|X_A,\theta_1=\tau_1}} \max_{\tau \in \Lambda(\tau_1)} \sum_{i \in A^c} \mathbb{E}[(X_i - Y_i)^2 | \theta = \tau] \quad \textit{and}$$
$$\Delta_{\max} = \max_{\tau \in \Theta} \sum_{i=1}^{m} \mathbb{E}[X_i^2 | \theta = \tau].$$

Remark: In Appendix C.2 a simple proof (using contradiction arguments) is provided to show the existence of $\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\}$, with Δ_{τ_1} being continuous in τ_1, that attains the minimum and maximum in (4.17).

Notice that $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ and $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ are reminiscent of the SRDf for a GMMS and, in fact, reduce to the SRDf for a GMMS with $\nu_{X_{\mathcal{M}}|\theta=\tau}$ for $\tau \in \Lambda(\tau_1)$ when $\Lambda(\tau_1)$ is a singleton. Thus, the equivalent forms (4.15) and (4.16) can be seen as counterparts of (4.5). Additionally, in Section 4.4, we show that $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ and $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ are continuous in $\tau_1 \in \Theta_1$.

The Bayesian USRDf with an outer minimization over $\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\}$ can be strictly smaller than its nonBayesian counterpart. An illustration of the comparison of the Bayesian and nonBayesian USRDfs is provided in the example below.

Example 4.3 *For $\mathcal{M} = \{1, 2\}$ and fixed $\sigma^2 > 0,\ r_{\min} > 0$ and $r_{\max} < 1$, consider a GMMS with pdf in Θ, where each $\Theta = \{\mathbf{\Sigma}_{\mathcal{M}\tau}\}_\tau$, where each $\mathbf{\Sigma}_{\mathcal{M}\tau}$ is given by*

$$\mathbf{\Sigma}_{\mathcal{M}\tau} = \begin{pmatrix} \sigma^2 & r_\tau \sigma^2 \\ r_\tau \sigma^2 & \sigma^2 \end{pmatrix}$$

for $r_{\min} \leq r_\tau \leq r_{\max},\ \tau \in \Theta$. Let θ be a Θ-valued rv with pdf ν_θ continuous on Θ. For a k-FS with $k = 1$, for both $A = \{1\}$ and $A = \{2\}$, Θ_1 is a singleton. Hence, in the Bayesian setting, the minimum and maximum in (4.17) are vacuous. For $A = \{1\},\ \{2\}$, in the Bayesian setting we have

$$\mathbf{G}_{A,\tau_1} = 1 + \mathbb{E}^2[r_\theta],$$

$$\Delta_{\min,A,\tau_1} = \sigma^2(1 - \mathbb{E}^2[r_\theta]),$$

and (4.17) now yields the Bayesian USRDf to be

$$R_{\{1\}}(\Delta) = R_{\{2\}}(\Delta) = \frac{1}{2} \log \frac{\sigma^2(1 + \mathbb{E}^2[r_\theta])}{\Delta - \sigma^2(1 - \mathbb{E}^2[r_\theta])},\ \sigma^2(1 - \mathbb{E}^2[r_\theta]) \leq \Delta \leq 2\sigma^2. \quad (4.19)$$

Evaluating (4.18), *the nonBayesian USRDf is*

$$R_{\{1\}}(\Delta) = R_{\{2\}}(\Delta) = \frac{1}{2}\log\frac{\sigma^2(1+r_{\min}^2)}{\Delta - \sigma^2(1-r_{\min}^2)}, \quad \sigma^2(1-r_{\min}^2) \leq \Delta \leq 2\sigma^2. \tag{4.20}$$

A simple comparison of (4.19) *and* (4.20) *shows that the nonBayesian USRDf is strictly larger than its Bayesian counterpart. Also, it is seen from* (4.19) *and* (4.20) *above that when* $r_\tau > 0$ *for all* $\tau \in \Theta$, *the average correlation,* $\mathbb{E}[r_\theta]$, *and the smallest correlation,* $r_{\min}$, *play similar roles in the expressions for Bayesian and nonBayesian USRDf, respectively.*

■

Lastly, the standard properties of the SRDf and the USRDf for GMMS and GMF with fixed-set samplers are summarized in the lemma below, with the proof provided in Appendix D.2.

Lemma 4.1 *The right-sides of* (4.5), (4.9), (4.17) *and* (4.18) *are finite-valued, decreasing, convex, continuous functions of* $\Delta_{\min,A} < \Delta \leq \Delta_{\max}$.

4.4 Proofs of Main Results

4.4.1 Achievability Proofs

We present first the achievability proof of Theorem 4.1 where the sampled components of the GMMS are reconstructed first with a weighted MSE distortion measure under a reduced threshold, and then MMSE estimates are formed for the unsampled components based on the former. An achievability proof for Proposition 4.1 is along similar lines. Building on this, we present next an achievability proof for Theorem 4.2 with an emphasis on the Bayesian setting. All our achievability proofs emphasize the modular structure of the reconstruction mechanism, which allows GMMS reconstruction to be performed in two steps.

Theorem 4.1:First, observe that

$$
\begin{aligned}
\Delta_{\min,A} &= \min_{X_{A^c} \,-\!\circ\!-\, X_A \,-\!\circ\!-\, Y_{\mathcal{M}}} \mathbb{E}[||X_{\mathcal{M}} - Y_{\mathcal{M}}||^2] \\
&= \min_{X_{A^c} \,-\!\circ\!-\, X_A \,-\!\circ\!-\, Y_{A^c}} \sum_{i \in A^c} \mathbb{E}[(X_i - Y_i)^2] \qquad \text{with } Y_i = X_i,\ i \in A \\
&= \sum_{i \in A^c} \mathbb{E}[(X_i - \mathbb{E}[X_i|X_A])^2] \\
&= \sum_{i \in A^c} \left(\mathbb{E}[X_i^2] - \mathbb{E}[X_i X_A^T]\mathbf{\Sigma}_A^{-1}\mathbb{E}[X_A X_i]\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta_{\max} &= \min_{\substack{X_{A^c} \,-\!\circ\!-\, X_A \,-\!\circ\!-\, Y_{\mathcal{M}} \\ X_A \perp\!\!\!\perp Y_{\mathcal{M}}}} \mathbb{E}[||X_{\mathcal{M}} - Y_{\mathcal{M}}||^2] \\
&= \min_{y_{\mathcal{M}}} \mathbb{E}[||X_{\mathcal{M}} - y_{\mathcal{M}}||^2] \\
&= \sum_{i=1}^{m} \mathbb{E}[X_i^2],
\end{aligned}
$$

where $\mathbf{\Sigma}_A^{-1}$ exists by the assumed positive-definiteness of $\mathbf{\Sigma}_{\mathcal{M}}$.

Given $\epsilon > 0$, for the GMMS $\{X_{At}\}_{t=1}^{\infty}$ with pdf $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_A)$ and weighted MSE distortion measure d_A, consider a (standard) rate distortion code (f_A, φ_A), f_A :

$\mathbb{R}^{nk} \to \{1, \dots, J\}$ and $\varphi_A : \{1, \dots, J\} \to \mathbb{R}^{nk}$ of rate $\frac{1}{n} \log J \le R_A(\Delta) + \epsilon$ and with

$$\mathbb{E}[d_A(X_A^n, Y_A^n)] \le \Delta - \Delta_{\min,A} + \epsilon,$$

for $n \ge N_\epsilon$, say.

A code (f, φ) is devised as follows. The encoder f is chosen to be f_A, i.e.,

$$f(x_A^n) \triangleq f_A(x_A^n), \quad x_A^n \in \mathbb{R}^{nk}$$

and the decoder φ is given by

$$\varphi(j) \triangleq \big(\varphi_A(j),\ \mathbb{E}[X_{A^c}^n | X_A^n = \varphi_A(j)]\big), \quad j \in \{1, \dots, J\}.$$

The rate of the code (f, φ) is

$$\frac{1}{n} \log J \le R_A(\Delta) + \epsilon.$$

Denote the output of the decoder $\varphi(f(X_A^n))$ by $Y_{\mathcal{M}}^n = (Y_A^n, Y_{A^c}^n)$. Then, $Y_{A^c}^n = \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A Y_A^n$ and by the standard properties of an MMSE estimate, for $t = 1, \dots, n$, it holds that

$$(X_{A^ct} - \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1} X_{At}) \perp\!\!\!\perp X_{At}. \tag{4.21}$$

The code (f, φ) has expected distortion

$$\begin{aligned}
\mathbb{E}[||X^n_{\mathcal{M}} - Y^n_{\mathcal{M}}||^2] &= \mathbb{E}[||X^n_A - Y^n_A||^2] + \mathbb{E}[||X^n_{A^c} - Y^n_{A^c}||^2] &(4.22)\\
&= \mathbb{E}[||X^n_A - Y^n_A||^2] + \mathbb{E}[||X^n_{A^c} - \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y^n_A||^2]\\
&= \mathbb{E}[||X^n_A - Y^n_A||^2]\\
&\quad + \mathbb{E}[||X^n_{A^c} - \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}X^n_A + \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}X^n_A - \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y^n_A||^2]\\
&= \mathbb{E}[||X^n_A - Y^n_A||^2] + \mathbb{E}[||X^n_{A^c} - \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}X^n_A||^2]\\
&\quad + \mathbb{E}[||\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}X^n_A - \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y^n_A||^2] &(4.23)\\
&= \Delta_{\min,A} + \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[(X_{At} - Y_{At})^T(I + \mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{AA^c}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1})(X_{At} - Y_{At})]\\
& &(4.24)\\
&= \Delta_{\min,A} + \mathbb{E}[d_A(X^n_A, Y^n_A)]\\
&\le \Delta + \epsilon, &(4.25)
\end{aligned}$$

where (4.23) is by the orthogonality principle of MMSE estimates (4.21) and since for $t = 1, \ldots, n$,

$$\begin{aligned}
&\mathbb{E}[(X_{A^ct} - \mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}X_{At})^T\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}]\\
&\qquad = \mathbb{E}[X^T_{A^ct}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}] - \mathbb{E}[X^T_{At}\mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{AA^c}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}]\\
&\qquad = \mathbb{E}\big[\mathbb{E}[X^T_{A^ct}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}|X^n_A]\big] - \mathbb{E}[X^T_{At}\mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{AA^c}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}]\\
&\qquad = \mathbb{E}\big[\mathbb{E}[X^T_{A^ct}|X^n_A]\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}\big] - \mathbb{E}[X^T_{At}\mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{AA^c}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}]\\
&\qquad = \mathbb{E}[X^T_{At}\mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{AA^c}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}] - \mathbb{E}[X^T_{At}\mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{AA^c}\mathbf{\Sigma}_{A^cA}\mathbf{\Sigma}_A^{-1}Y_{At}]
\end{aligned}$$

$$= 0.$$

■

Proposition 4.1: The achievability proof of Proposition 4.1 is along the lines of Theorem 4.1. For a given $\Delta_{\min} < \Delta \le \Delta_{\max}$ and $\epsilon > 0$, for the GMMS $\{X_{At}\}_{t=1}^{\infty}$ with weighted MSE distortion measure

$$d_A(x_A, y_A) \triangleq (x_A - y_A)^T \mathbf{G}_{A,I}(x_A - y_A), \quad x_A, y_A \in \mathbb{R}^k,$$

consider a rate distortion code (f_A, φ_A), $f_A : \mathbb{R}^{nk} \to \{1, \dots, J\}$ and $\varphi_A : \{1, \dots, J\} \to \mathbb{R}^{nk}$ of rate $\frac{1}{n} \log J \le R_A(\Delta) + \epsilon$ and with

$$\mathbb{E}[d_A(X_A^n, Y_A^n)] \le \Delta - \Delta_{\min,A} + \epsilon,$$

for $n \ge N_\epsilon$.

A code (f, φ) is then constructed as follows. The encoder f is chosen to be

$$f(x_A^n) = f_A(x_A^n), \qquad x_A^n \in \mathbb{R}^{nk}.$$

The output of decoder φ, corresponding to each $u \in I$, is given by

$$(\varphi(j))_u = \mathbb{E}[X_u^n | X_A^n = \varphi_A(j)], \qquad j \in \{1, \dots, J\}.$$

Denoting the output of the decoder $\varphi(f(X_A^n))$ by Y_I^n, for $u \in I,\ t = 1, \ldots, n$,

$$Y_{ut} = \mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}Y_{At},$$

where $\mathbf{\Sigma}_{\{u\}A} = \mathbb{E}[X_u X_A^T]$ and $\mathbf{\Sigma}_A = \mathbb{E}[X_A X_A^T]$. The rate of the code (f, φ) is

$$\frac{1}{n}\log J \leq R_A(\Delta) + \epsilon.$$

The code (f, φ) has expected distortion

$$\begin{aligned}
\mathbb{E}[||X_I^n - Y_I^n||^2] &= \int_I \mathbb{E}\big[||X_u^n - Y_u^n||^2\big]\, du \\
&= \int_I \mathbb{E}[||X_u^n - \mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}Y_A^n||^2]\, du \\
&= \int_I \mathbb{E}\big[||X_u^n - \mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}X_A^n + \mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}X_A^n - \mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}Y_A^n||^2\big]\, du \\
&= \int_I \mathbb{E}\big[||X_u^n - \mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}X_A^n||^2\big] + \mathbb{E}\big[||\mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}X_A^n - \mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}Y_A^n||^2\big]\, du \qquad (4.26) \\
&= \Delta_{\min,A} + \int_I \mathbb{E}\big[||(X_A^n - Y_A^n)^T\mathbf{\Sigma}_A^{-1}\mathbf{\Sigma}_{\{u\}A}^T\mathbf{\Sigma}_{\{u\}A}\mathbf{\Sigma}_A^{-1}(X_A^n - Y_A^n)||^2\big]\, du \\
&= \Delta_{\min,A} + \mathbb{E}\big[||(X_A^n - Y_A^n)^T\mathbf{G}_{A,I}(X_A^n - Y_A^n)||^2\big] \qquad \text{using (4.8)} \\
&\leq \Delta + \epsilon,
\end{aligned}$$

where (4.26) is by the orthogonality principle of the MMSE estimates as in (4.23), (4.24).

■

Before we present the achievability proof of Theorem 4.2, we present pertinent technical results. We state first a standard technical result, a Vitali covering lemma (Theorem 17.1 in [38]), without proof. For any $\epsilon > 0$, this lemma guarantees the existence of a finite number of nonoverlapping Euclidean "balls" of radius $\leq \epsilon$ such that the Lebesgue measure of points in Θ_1 not covered by the Euclidean balls is $\leq \epsilon$. In the achievability proof of Theorem 4.2, the centers of such balls will be used to approximate Θ_1 and (approximately) estimate θ_1. For $\tau_1 \in \Theta_1$, let $B_{\tau_1,\epsilon} \subset \mathbb{R}^{k^2}$ denote a standard Euclidean ℓ_2-ball with center τ_1 and radius ϵ.

Lemma 4.2 *For every $\epsilon > 0$, there exists an $N_\epsilon > 0$ and a finite disjoint collection of balls $\{B_{\tau_{1,i},\epsilon_i}\}_{i=1}^{N_\epsilon}$ such that $\max\limits_i \epsilon_i \leq \epsilon$ and*

$$\mu\big(\Theta_1 \setminus \bigcup_i B_{\tau_{1,i},\epsilon_i}\big) < \epsilon, \tag{4.27}$$

where μ is the Lebesgue measure on $\mathbb{R}^{k^2}$ and $\setminus$ is the standard set difference.

Remarks: i) The lemma above relies on Θ_1 being a compact subset of $\mathbb{R}^{k^2}$.

ii) For $\epsilon > 0$ and $\{B_{\tau_{1,i},\epsilon_i}\}_{i=1}^{N_\epsilon}$ as in the lemma above, let $\Theta_{1,\epsilon} \subset \Theta_1$ be the collection of "centers" $\{\tau_{1,i}\}_{i=1}^{N_\epsilon}$.

While the lemma above is pertinent to the Bayesian and nonBayesian parts of Theorem 4.2, Lemmas 4.3 and 4.4 below are pertinent to the Bayesian and non-Bayesian settings respectively.

Lemma 4.3 *In the Bayesian setting, for every $x_{\mathcal{M}} \in \mathbb{R}^m$,*

$$\nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tau_1) \tag{4.28}$$

is continuous in τ_1. For any code (f, φ), $f : \mathbb{R}^{nk} \to \{1, \ldots, J\}$, $\varphi : \{1, \ldots, J\} \to \mathbb{R}^{nm}$,

$$\mathbb{E}\big[||X^n_{\mathcal{M}} - \varphi(f(X^n_A))||^2 \big| \theta_1 = \tau_1\big]$$

is continuous in τ_1.

Proof: See Appendix B.1. ■

Remarks: (i) Since Θ_1 is a compact set, for every $x_{\mathcal{M}} \in \mathbb{R}^m$, the pdf $\nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tau_1)$ and $\mathbb{E}\big[||X^n_{\mathcal{M}} - \varphi(f(X^n_A))||^2 \big| \theta_1 = \tau_1\big]$ are, in fact, uniformly continuous in τ_1. Thus, for every $x_{\mathcal{M}} \in \mathbb{R}^m$ and $\epsilon > 0$, there exists a $\delta > 0$ such that for $\tau_{1,1},\ \tau_{1,2} \in \Theta_1$ with $||\tau_{1,1} - \tau_{1,2}|| \leq \delta$, it holds that

$$|\nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tau_{1,1}) - \nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tau_{1,2})| \leq \epsilon,$$

and

$$\Big|\mathbb{E}\big[||X^n_{\mathcal{M}} - \varphi(f(X^n_A))||^2 \big| \theta_1 = \tau_{1,1}\big] - \mathbb{E}\big[||X^n_{\mathcal{M}} - \varphi(f(X^n_A))||^2 \big| \theta_1 = \tau_{1,2}\big]\Big| \leq \epsilon.$$

(ii) The claim (4.28) implies that

$$\mathbb{E}[X_A X_A^T|\theta_1 = \tau_1] \quad \text{and} \quad \mathbb{E}[X_A X_{A^c}^T|\theta_1 = \tau_1]$$

are continuous in τ_1 and hence,

$$\mathbf{G}_{A,\tau_1} = \mathbf{I}+(\mathbb{E}[X_A X_A^T|\theta_1 = \tau_1])^{-1}\mathbb{E}[X_A X_{A^c}^T|\theta_1 = \tau_1]\mathbb{E}[X_{A^c} X_A^T|\theta_1 = \tau_1](\mathbb{E}[X_A X_A^T|\theta_1 = \tau_1])^{-1}$$

is continuous in τ_1. Thus, from (4.15), for every $\delta > \Delta_{\min,A,\tau_1}$, $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ is continuous in τ_1.

The following lemma implies that if $\tau_{1,1}, \tau_{1,2} \in \Theta_1$ are "close," then there exist $\hat{\tau}$ and $\check{\tau}$ in the ambiguity atoms of $\tau_{1,1}$ and $\tau_{1,2}$, respectively, which too are "close."

Lemma 4.4 *For every $\epsilon > 0$, there exists a $\delta > 0$ such that for every $\tau_{1,1}, \tau_{1,2} \in \Theta_1$ with $||\tau_{1,1} - \tau_{1,2}|| \leq \delta$, the following holds*

$$\min_{\hat{\tau}\in\Lambda(\tau_{1,1}),\ \check{\tau}\in\Lambda(\tau_{1,2})} ||\hat{\tau} - \check{\tau}|| \leq \epsilon. \tag{4.29}$$

Proof: See Appendix B.2. ■

Theorem 4.2: Consider Θ_1 as in Section 4.3. Based on the output of the fixed-set sampler X_A^n, the encoder forms a *maximum-likelihood* (ML) estimate for the covariance-matrix $\mathbf{\Sigma}_{A\tau_1}$ as

$$\widehat{\theta}_{1,n} = \widehat{\theta}_{1,n}(X_A^n) = \frac{1}{n}\sum_{t=1}^{n} X_{At} X_{At}^T.$$

Observe that $\{X_{At}\}_{t=1}^{\infty}$ is a GMMS with pdf $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{A\tau_1})$ and $\nu_{X_A|\theta_1=\tau_1}$ is continuous in τ_1. Compactness of Θ_1, the boundedness and continuity of $\nu_{X_A|\theta_1=\tau_1}$ in τ_1 imply, by the uniform law of large numbers [39], that

$$\widehat{\theta}_{1,n} \xrightarrow{a.s.} \tau_1$$

under $\nu_{X_A|\theta_1=\tau_1}$, and that for every $\epsilon_1 > 0$, there exists a δ and N_{ϵ_1} such that for every $\tau_1 \in \Theta_1$

$$P_{\tau_1}\Big(||\tau_1 - \widehat{\theta}_{1,n}|| > \delta\Big) \leq \epsilon_1, \quad n \geq N_{\epsilon_1}. \tag{4.30}$$

Now, considering a subset $\Theta_{1,\delta}$ of Θ_1 as in the remark following Lemma 4.2, define $\tilde{\theta}_{1,n}$ as

$$\tilde{\theta}_{1,n} \triangleq \arg\min_{\check{\tau}_1 \in \Theta_{1,\delta}} ||\widehat{\theta}_{1,n} - \check{\tau}_1||. \tag{4.31}$$

Fixing $\epsilon > 0$ and $0 < \epsilon_1 < \epsilon$, from (4.30), (4.31) and Lemma 4.2, it follows that there exists a δ and N_{ϵ_1} such that

$$P\Big(||\theta_1 - \tilde{\theta}_{1,n}|| > 2\delta\Big) \leq \epsilon_1, \quad n \geq N_{\epsilon_1}. \tag{4.32}$$

Notice that while $\hat{\theta}_{1,n}$ may lie outside Θ_1, $\tilde{\theta}_{1,n}$ is an estimate of θ_1 that takes values in a finite subset of Θ_1. The estimate $\tilde{\theta}_{1,n}$ (of θ_1) will be used in the next part of the proof to select sampling rate distortion codes.

For a fixed $\Delta_{\min,A} < \Delta \leq \Delta_{\max}$, let $\{\Delta_{\tau_1}, \tau_1 \in \Theta_1\}$ be such that it attains the minimum in (4.17) and Δ_{τ_1} is continuous in τ_1 (see the remark below Theorem 4.2). Recall that for each $\tau_1 \in \Theta_{1,\delta}$, $\rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1)$ is, in effect, the RDf for a GMMS $\{X_{At}\}_{t=1}^{\infty}$ with pdf $\nu_{X_A|\theta_1=\tau_1}$ under a weighted MSE distortion measure $d_{A\tau_1}$. Thus, for each $\tau_1 \in \Theta_{1,\delta}$, there exists a (standard) rate distortion code $(f_{\tau_1}, \varphi_{\tau_1})$, $f_{\tau_1} : \mathbb{R}^{nk} \to \{1, \ldots, J\}$ and $\varphi_{\tau_1} : \{1, \ldots, J\} \to \mathbb{R}^{nk}$ of rate $\frac{1}{n} \log J \leq \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) + \epsilon_1 \leq R_A(\Delta) + \epsilon_1$ and with

$$\mathbb{E}[d_{A\tau_1}(X_A^n, \varphi_{\tau_1}(f_{\tau_1}(X_A^n)))|\theta_1 = \tau_1] \leq \Delta_{\tau_1} - \Delta_{\min,A,\tau_1} + \epsilon_1$$

for all $n \geq N_{\epsilon_1}$.

Now, consider a code (f, φ) with f taking values in $\mathcal{J} \triangleq \{1, \ldots, |\Theta_{1,\delta}|\} \times \{1, \ldots, J\}$ as follows. Order (in any manner) the elements of $\Theta_{1,\delta}$. The encoder f, dictated by the estimate $\tilde{\theta}_{1,n}$, is given by

$$f(x_A^n) \triangleq \left(\tilde{\theta}_{1,n}(x_A^n), f_{\tilde{\theta}_{1,n}}(x_A^n)\right), \quad x_A^n \in \mathbb{R}^{nk}$$

and the decoder φ is given by

$$\varphi(\tau_1, j) \triangleq \left(\varphi_{\tau_1}(j), \mathbb{E}[X_{A^c}^n | X_A^n = \varphi_{\tau_1}(j), \theta_1 = \tau_1]\right), \quad (\tau_1, j) \in \mathcal{J}. \qquad (4.33)$$

By the finiteness of $\Theta_{1,\delta}$, the rate of the code (f, φ) is

$$\begin{aligned}\frac{1}{n}\log|\mathcal{J}| &= \frac{1}{n}\log|\Theta_{1,\delta}| + \frac{1}{n}\log J \\ &\leq R_A(\Delta) + 2\epsilon_1,\end{aligned}$$

for n large enough. Denoting the output of the decoder by $Y_{\mathcal{M}}^n$ with $Y_A^n = \varphi_{\tau_1}(j)$ and $Y_{A^c}^n = \mathbb{E}[X_{A^c}^n | X_A^n = \varphi_{\tau_1}(j), \theta_1 = \tau_1]$, we have that

$$\begin{aligned}\mathbb{E}[||X_{\mathcal{M}}^n - Y_{\mathcal{M}}^n||^2] &= \mathbb{E}[\mathbb{1}(||\tilde{\theta}_{1,n} - \theta_1|| \leq 2\delta)||X_{\mathcal{M}}^n - Y_{\mathcal{M}}^n||^2] \\ &\quad + \mathbb{E}[\mathbb{1}(||\tilde{\theta}_{1,n} - \theta_1|| > 2\delta)||X_{\mathcal{M}}^n - Y_{\mathcal{M}}^n||^2] \qquad (4.34)\end{aligned}$$

Using Lemma 4.3, it is shown in Appendix C.3 that the first term in the right-side of (4.34) is

$$\mathbb{E}[\mathbb{1}(||\tilde{\theta}_{1,n} - \theta_1|| \leq 2\delta)||X_{\mathcal{M}}^n - Y_{\mathcal{M}}^n||^2] \leq \Delta + 4\epsilon_1. \qquad (4.35)$$

Next, we show that the second term in the right-side of (4.34) is "small." First, note that the finiteness of $\Theta_{1,\delta}$ implies the existence of an M_1 such that, for $t = 1, \ldots, n$,

$$|(\varphi_{\tau_1}(f_{\tau_1}(x_A^n)))_{i,t}| \leq M_1, \ i \in A, \ \tau_1 \in \Theta_{1,\delta}, \quad x_A^n \in \mathbb{R}^{nk}$$

and hence, from (4.33), there exists an $M_2 > 0$ such that, for $t = 1, \ldots, n$,

$$|(\varphi(f(x_A^n)))_{i,t}| \leq M_2, \quad i \in \mathcal{M}, \ x_A^n \in \mathbb{R}^{nk}. \qquad (4.36)$$

For $i \in \mathcal{M}$, from (4.36), Cauchy-Schwarz inequality, and the fact that $\mathbb{E}[X_i^2]$ is bounded, there exists an M such that

$$\mathbb{E}[(X_{it} - Y_{it})^4] \leq M, \qquad t = 1, \ldots, n. \tag{4.37}$$

Now, the second term on the right-side of (4.34),

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}(||\tilde{\theta}_{1,n} - \theta_1|| > 2\delta)||X_{\mathcal{M}}^n - Y_{\mathcal{M}}^n||^2] \\
&\qquad = \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m} \mathbb{E}\big[\mathbb{1}(||\tilde{\theta}_{1,n} - \theta_1|| > 2\delta)\big(X_{it} - Y_{it}\big)^2\big] \\
&\qquad \leq \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m} \sqrt{\mathbb{E}\big[\mathbb{1}^2(||\tilde{\theta}_{1,n} - \theta_1|| > 2\delta)\big]\mathbb{E}\big[\big(X_{it} - Y_{it}\big)^4\big]} \qquad (4.38) \\
&\qquad \leq \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{m} \sqrt{\epsilon_1 M} \qquad\qquad \text{from (4.32) and (4.37)} \\
&\qquad \leq \sqrt{\epsilon_1 M} m, \qquad\qquad (4.39)
\end{aligned}$$

where (4.38) is by the Cauchy-Schwarz inequality. From (4.35) and (4.39), we get

$$\begin{aligned}
\mathbb{E}\big[||X_{\mathcal{M}}^n - Y_{\mathcal{M}}^n||^2\big] &\leq \Delta + 4\epsilon_1 + m\sqrt{\epsilon_1 M} \\
&\leq \Delta + \epsilon,
\end{aligned}$$

for ϵ_1 small enough.

In the nonBayesian setting, as a first step, Lemma 4.4 is used to show that $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ is a continuous function of τ_1. Then, the maximum in (4.18) is seen to exist as a continuous function over a compact set attains its supremum. Next, the achievability proof follows by adapting the steps above with the following differences. For each $\tau_1 \in \Theta_{1,\delta}$, sampling rate distortion codes $(f_{\tau_1}, \varphi_{\tau_1})$, $f_{\tau_1} : \mathbb{R}^{nk} \to \{1, \ldots, J\}$, $\varphi_{\tau_1} : \{1, \ldots, J\} \to \mathbb{R}^{nm}$ are chosen to satisfy

$$\mathbb{E}\big[||X^n_{\mathcal{M}} - \varphi_{\tau_1}(f_{\tau_1}(X^n_A))||^2|\theta = \tau\big] \leq \Delta, \quad \tau \in \Lambda(\tau_1),$$

with rate $\frac{1}{n}\log||f_{\tau_1}|| \leq R_A(\Delta)+\epsilon$, where $R_A(\Delta)$ is the nonBayesian USRDf. A code (f, φ) with f taking values in $\mathcal{J} = \{1, \ldots, |\Theta_{1,\delta}|\} \times \{1, \ldots, J\}$ is constructed based on the codes $(f_{\tau_1}, \varphi_{\tau_1})$ as before. While counterparts of (4.35) and (4.39) can be shown for each $\tau_1 \in \Theta_1$ using a similar set of ideas, a key distinction in the analysis is that Lemma 4.4 is used in lieu of Lemma 4.3 to show that for $\tau \in \Lambda(\tau_1), \tau_1 \in \Theta_1$,

$$\mathbb{E}\big[\mathbb{1}(||\tilde{\theta}_{1,n} - \tau_1|| \leq 2\epsilon_1)||X^n_{\mathcal{M}} - \varphi(\tilde{\theta}_{1,n}, f_{\tilde{\theta}_{1,n}}(X^n_A))||^2\big|\theta = \tau\big] \leq \Delta + \epsilon_1,$$

the counterpart of (4.35). ■

4.4.2 Converse Proofs

In contrast to the achievability proofs, we present a converse proof for Theorem 4.2 first, with an emphasis on the Bayesian setting; this is then adapted to Theorem 4.1. Prior to this, we prove the equivalence of expressions in (4.40), that will be pertinent to Theorem 4.1. Building on this, we show the equivalence of the simplified forms for $\rho^{\mathcal{B}}_A(\delta, \tau_1)$ and $\rho^{n\mathcal{B}}_A(\delta, \tau_1)$ in Proposition 4.2. Next, we shall present a technical lemma. These will be used subsequently in the unified converse proof for Theorems

4.1 and 4.2. The converse proof for Proposition 4.1 uses an approach that does not rely on Lemma 4.5 and is presented last.

Equivalence for Theorem 4.1: The following equality will be relevant in the proof of converse for Theorem 4.1:

$$\min_{\substack{X_{A^c} -\!\circ\!- X_A -\!\circ\!- Y_{\mathcal{M}} \\ \mu_{X_A Y_A} \ll \mu_{X_A} \times \mu_{Y_A} \\ \mathbb{E}[||X_{\mathcal{M}} - Y_{\mathcal{M}}||^2] \leq \Delta}} I\big(X_A \wedge Y_A\big) = \min_{\substack{\mu_{X_A Y_A} \ll \mu_{X_A} \times \mu_{Y_A} \\ \mathbb{E}[d_A(X_A, Y_A)] \leq \Delta - \Delta_{\min, A}}} I\big(X_A \wedge Y_A\big). \tag{4.40}$$

For any pair of rvs $X_{\mathcal{M}}, Y_{\mathcal{M}}$ satisfying the constraints on the left-side of (4.40), consider

$$\widehat{Y}_{\mathcal{M}} \triangleq \mathbb{E}[X_{\mathcal{M}} | Y_{\mathcal{M}}]. \tag{4.41}$$

Now,

$$\begin{aligned} \widehat{Y}_{A^c} &= \mathbb{E}[X_{A^c} | Y_{\mathcal{M}}] = \mathbb{E}[\mathbb{E}[X_{A^c} | X_A, Y_{\mathcal{M}}] | Y_{\mathcal{M}}] = \mathbb{E}[\mathbb{E}[X_{A^c} | X_A] | Y_{\mathcal{M}}] \\ &= \mathbb{E}[\mathbf{\Sigma}_{A^c A} \mathbf{\Sigma}_A^{-1} X_A | Y_{\mathcal{M}}] = \mathbf{\Sigma}_{A^c A} \mathbf{\Sigma}_A^{-1} \widehat{Y}_A. \end{aligned} \tag{4.42}$$

By the optimality of the MMSE estimate,

$$\mathbb{E}[||X_{\mathcal{M}} - \widehat{Y}_{\mathcal{M}}||^2] \leq \mathbb{E}[||X_{\mathcal{M}} - Y_{\mathcal{M}}||^2] \leq \Delta. \tag{4.43}$$

It is readily checked (along the lines of (4.22)-(4.25)) that

$$\mathbb{E}[||X_{\mathcal{M}} - \widehat{Y}_{\mathcal{M}}||^2] = \mathbb{E}[d_A(X_A, \widehat{Y}_A)] + \Delta_{\min,A}. \tag{4.44}$$

Putting together (4.41)-(4.44), completes the proof of (4.40).

■

Proposition 4.2: The proof of (4.15) and (4.16) is the along the lines of proof of (4.40), with the distinction that in the nonBayesian setting, $\widehat{Y}_A$ is chosen to satisfy the orthogonality principle and $\widehat{Y}_{A^c}$ is chosen to be a linear function of $\widehat{Y}_A$.

■

The following technical lemma is the counterpart of Lemma 6 in [40].

Lemma 4.5 *In the Bayesian setting, for any n-length k-FS code (f, φ) with $f : \mathbb{R}^{nk} \to \{1, \ldots, J\}$, $\varphi : \{1, \ldots, J\} \to \mathbb{R}^{nm}$, for $t = 1, \ldots, n$, denoting $\varphi(f(X_A^n))$ by $Y_{\mathcal{M}}^n$, it holds that*

$$\theta, X_{A^ct} -\!\!\circ\!\!- \theta_1, X_{At} -\!\!\circ\!\!- Y_{\mathcal{M}t}. \tag{4.45}$$

Proof: First, note that

$$\theta, X_{A^c}^n -\!\!\circ\!\!- X_A^n -\!\!\circ\!\!- Y_{\mathcal{M}}^n \tag{4.46}$$

holds by code construction. From (4.46) (and since $Y_{\mathcal{M}}^n$ above is a finite-valued rv), we have

$$
\begin{aligned}
0 = I\big(\theta, X_{A^c}^n \wedge Y_{\mathcal{M}}^n | X_A^n\big) &= I\big(\theta \wedge Y_{\mathcal{M}}^n | X_A^n\big) + I\big(X_{A^c}^n \wedge Y_{\mathcal{M}}^n | X_A^n, \theta\big) \\
&= I\big(\theta, \theta_1 \wedge Y_{\mathcal{M}}^n | X_A^n\big) + I\big(X_{A^c}^n \wedge Y_{\mathcal{M}}^n | X_A^n, \theta\big) \qquad (4.47) \\
&\geq I\big(\theta \wedge Y_{\mathcal{M}}^n | X_A^n, \theta_1\big) + I\big(X_{A^c}^n \wedge Y_{\mathcal{M}}^n | X_A^n, \theta\big), \qquad (4.48)
\end{aligned}
$$

where (4.47) is since θ_1 is a function of θ. Now, the second term on the right-side of (4.48) is

$$
\begin{aligned}
0 = I(X_{A^c}^n \wedge Y_{\mathcal{M}}^n | X_A^n, \theta) &= \sum_{t=1}^n I\big(X_{A^c t} \wedge Y_{\mathcal{M}}^n | X_{A^c}^{t-1}, X_A^n, \theta\big) \\
&= \sum_{t=1}^n \Big(I\big(X_{A^c t} \wedge X_{A^c}^{t-1}, X_A^{n\setminus t}, Y_{\mathcal{M}}^n | X_{At}, \theta\big) \\
&\qquad\qquad - I\big(X_{A^c t} \wedge X_{A^c}^{t-1} X_A^{n\setminus t} | X_{At}, \theta\big)\Big) \\
&= \sum_{t=1}^n I\big(X_{A^c t} \wedge X_{A^c}^{t-1}, X_A^{n\setminus t}, Y_{\mathcal{M}}^n | X_{At}, \theta\big) \qquad (4.49) \\
&\geq \sum_{t=1}^n I\big(X_{A^c t} \wedge Y_{\mathcal{M}t} | X_{At}, \theta\big), \qquad (4.50)
\end{aligned}
$$

where (4.49) is since $\nu_{X_{\mathcal{M}}^n|\theta} = \prod_{t=1}^n \nu_{X_{\mathcal{M}t}|\theta}$. Next, (4.48) and the fact

$$\theta \multimap\!\!- \theta_1 \multimap\!\!- X_A^n$$

imply

$$0 = I\big(\theta \wedge X_A^n | \theta_1\big) + I\big(\theta \wedge Y_{\mathcal{M}}^n | X_A^n, \theta_1\big)$$

$$= I\big(\theta \wedge X_A^n, Y_{\mathcal{M}}^n | \theta_1\big)$$

and hence, for $t = 1, \ldots, n$,

$$I(\theta \wedge X_{At}, Y_{\mathcal{M}t} | \theta_1) = 0. \tag{4.51}$$

Now, by (4.50) and (4.51), for $t = 1, \ldots, n$,

$$I(\theta \wedge Y_{\mathcal{M}t} | X_{At}, \theta_1) + I(X_{A^c t} \wedge Y_{\mathcal{M}t} | X_{At}, \theta) = I(\theta, X_{A^c t} \wedge Y_{\mathcal{M}t} | X_{At}, \theta_1) = 0,$$

hence, the claim of the lemma (4.45).

■

Converse: We provide first a converse proof for the Bayesian setting in Theorem 4.2, which is then refashioned to provide converse proofs for the nonBayesian setting and Theorem 4.1.

Let (f, φ) be an n-length k-FS code of rate R and with decoder output $Y_{\mathcal{M}}^n = \varphi(f(X_A^n))$ satisfying $\mathbb{E}[||X_{\mathcal{M}}^n - Y_{\mathcal{M}}^n||^2] \leq \Delta$. By lemma 4.5, for $t = 1, \ldots, n$, we have

$$\theta, X_{A^c t} -\!\!\circ\!\!- \theta_1, X_{At} -\!\!\circ\!\!- Y_{\mathcal{M}t}. \tag{4.52}$$

For $t = 1, \ldots, n$, and $\tau_1 \in \Theta_1$, let $\Delta_{\tau_1, t}$ denote $\mathbb{E}[||X_{\mathcal{M}t} - Y_{\mathcal{M}t}||^2 | \theta_1 = \tau_1]$ and $\Delta_{\tau_1} \triangleq \frac{1}{n} \sum\limits_{t=1}^{n} \mathbb{E}[||X_{\mathcal{M}t} - Y_{\mathcal{M}t}||^2 | \theta_1 = \tau_1]$. Along the lines of proof of Theorem 9.6.1

in [41], for every $\tau_1 \in \Theta_1$,

$$
\begin{aligned}
R = \frac{1}{n}\log|f| &\geq \frac{1}{n} H(f(X_A^n)|\theta_1 = \tau_1) \\
&\geq \frac{1}{n} H(Y_A^n|\theta_1 = \tau_1) \\
&= \frac{1}{n} I(X_A^n \wedge Y_A^n|\theta_1 = \tau_1) \\
&= \frac{1}{n}\sum_{t=1}^{n}\Big(I\big(X_{At} \wedge X_A^{t-1}, Y_A^n|\theta_1 = \tau_1\big) - I\big(X_{At} \wedge X_A^{t-1}|\theta_1 = \tau_1\big)\Big) \\
&= \frac{1}{n}\sum_{t=1}^{n} I(X_{At} \wedge X_A^{t-1}, Y_A^n|\theta_1 = \tau_1) \qquad\qquad \text{since} \quad \nu_{X_A^n|\theta_1} = \prod_{t=1}^{n}\nu_{X_{At}|\theta_1} \\
&\geq \frac{1}{n}\sum_{t=1}^{n} I(X_{At} \wedge Y_{At}|\theta_1 = \tau_1) \\
&\geq \frac{1}{n}\sum_{t=1}^{n} \min_{\substack{\theta, X_{A^c t} -\circ- \theta_1, X_{At} -\circ- Y_{\mathcal{M}t} \\ \mu_{X_{At}Y_{At}|\theta_1=\tau_1} \ll \mu_{X_{At}|\theta_1=\tau_1} \times \mu_{Y_{At}|\theta_1=\tau_1} \\ \mathbb{E}[||X_{\mathcal{M}t} - Y_{\mathcal{M}t}||^2|\theta_1=\tau_1] \leq \Delta_{\tau_1,t}}} I(X_{At} \wedge Y_{At}|\theta_1 = \tau_1) \qquad \text{by (4.52)} \\
&= \frac{1}{n}\sum_{t=1}^{n} \min_{\substack{\mu_{X_{At}Y_{At}|\theta_1=\tau_1} \ll \mu_{X_{At}|\theta_1=\tau_1} \times \mu_{Y_{At}|\theta_1=\tau_1} \\ \mathbb{E}[d_{A\tau_1}(X_{At}, Y_{At})|\theta_1=\tau_1] \leq \Delta_{\tau_1,t} - \Delta_{\min,A,\tau_1}}} I(X_{At} \wedge Y_{At}|\theta_1 = \tau_1) \quad \text{by Proposition 4.2} \\
&= \frac{1}{n}\sum_{t=1}^{n} \rho_A^{\mathcal{B}}(\Delta_{\tau_1,t}, \tau_1) \\
&\geq \rho_A^{\mathcal{B}}\Big(\frac{1}{n}\sum_{t=1}^{n}\Delta_{\tau_1,t}, \tau_1\Big) \geq \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1). \qquad (4.53)
\end{aligned}
$$

Now, (4.53) holds for every $\tau_1 \in \Theta_1$, hence

$$R \geq \sup_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1)$$

$$\geq \inf_{\substack{\{\Delta_{\tau_1},\ \tau_1\in\Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}]\leq\Delta}} \sup_{\tau_1\in\Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) \tag{4.54}$$

$$= R_A(\Delta)$$

for $\Delta > \Delta_{\min,A}$.

In the nonBayesian setting, the analog of Lemma 4.5 is obtained similarly with $\theta = \tau,\ \theta_1 = \tau_1$ and (4.46), (4.45) replaced with appropriate conditional measures. The proof of the converse is along the lines of the proof above, but with $\rho_A^{n\mathcal{B}}(\Delta, \tau_1)$ in place of $\rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1)$, and without the outer minimization with respect to $\{\Delta_{\tau_1}, \tau_1 \in \Theta_1\}$.

The converse proof for Theorem 4.1 obtains immediately from the Bayesian setting with the following changes: Θ_1 and $\Lambda(\tau_1),\ \tau_1 \in \Theta_1$, are taken to be singletons (rendering the infimum and supremum in (4.54) superfluous) and (4.40) is used in place of Proposition 4.2.

■

The converse proof for Proposition 4.1 involves an approach which does not rely on Lemma 4.5 and is presented next.

Converse proof for Proposition 4.1: Let (f, φ) be an n-length k-FS code of code R with $\mathbb{E}[||X_I^n - \varphi(f(X_A^n))||^2] \leq \Delta$. For $u \in I$ and $t = 1, \ldots, n$, define

$$\begin{aligned}\widehat{Y}_{ut} &= \mathbb{E}[X_{ut}|f(X_A^n)] \\ &= \mathbb{E}\big[\mathbb{E}[X_{ut}|X_A^n, f(X_A^n)]|f(X_A^n)\big]\end{aligned}$$

$$= \mathbb{E}\big[\mathbb{E}[X_{ut}|X_A^n]|f(X_A^n)\big]$$

$$= \mathbb{E}\big[\mathbb{E}[X_{ut}|X_{At}]|f(X_A^n)\big], \text{ since } X_{At}, X_{ut} \perp\!\!\!\perp X_A^{n\setminus t}, X_u^{n\setminus t}$$

$$= \mathbb{E}[X_{\{u\}A}]\mathbf{\Sigma}_A^{-1}\mathbb{E}[X_{At}|f(X_A^n)].$$

Notice that for $u \in I \setminus A$,

$$\widehat{Y}_{ut} = \mathbb{E}[X_{\{u\}A}]\mathbf{\Sigma}_A^{-1}\widehat{Y}_{At}, \qquad t = 1, \ldots, n.$$

By the optimality of the MMSE estimate

$$\Delta \geq \mathbb{E}[||X_I^n - \varphi(f(X_A^n))||^2] \geq \mathbb{E}[||X_I^n - \widehat{Y}_I^n||^2] = \mathbb{E}[(X_A^n - \widehat{Y}_A^n)^T\mathbf{G}_{A,I}(X_A^n - \widehat{Y}_A^n)] + \Delta_{\min,A}. \tag{4.55}$$

The equality in (4.55) can be seen to hold along the lines of (4.22)-(4.25). Now,

$$\begin{aligned}
R &= \frac{1}{n}\log|f| \geq \frac{1}{n}H(f(X_A^n)) \\
&= \frac{1}{n}I(X_A^n \wedge f(X_A^n)) \\
&\geq \min_{\substack{f,\varphi \\ \mathbb{E}[||X_I^n - \varphi(f(X_A^n))||^2] \leq \Delta}} \frac{1}{n}I(X_A^n \wedge f(X_A^n)) \\
&\geq \min_{\substack{\mu_{X_A^n Y_A^n} \ll \mu_{X_A^n} \times \mu_{Y_A^n} \\ \mathbb{E}[(X_A^n - Y_A^n)^T\mathbf{G}_{A,I}(X_A^n - Y_A^n)] \leq \Delta - \Delta_{\min,A}}} \frac{1}{n}I(X_A^n \wedge Y_A^n) \qquad \text{by (4.55)}
\end{aligned}$$

$$
\begin{aligned}
&= \min_{\substack{\mu_{X_A^n Y_A^n} \ll \mu_{X_A^n} \times \mu_{Y_A^n} \\ \mathbb{E}[(X_A^n - Y_A^n)^T \mathbf{G}_{A,I} (X_A^n - Y_A^n)] \le \Delta - \Delta_{\min,A}}} \frac{1}{n} \sum_{t=1}^{n} \Big(I(X_{At} \wedge X_A^{t-1}, Y_A^n) - I(X_{At} \wedge X_A^{t-1}) \Big) \\
&= \min_{\substack{\mu_{X_A^n Y_A^n} \ll \mu_{X_A^n} \times \mu_{Y_A^n} \\ \mathbb{E}[(X_A^n - Y_A^n)^T \mathbf{G}_{A,I} (X_A^n - Y_A^n)] \le \Delta - \Delta_{\min,A}}} \frac{1}{n} \sum_{t=1}^{n} I(X_{At} \wedge X_A^{t-1}, Y_A^n) \qquad \text{since } X_{At} \perp\!\!\!\perp X_A^{n \setminus t} \\
&\ge \min_{\substack{\mu_{X_A^n Y_A^n} \ll \mu_{X_A^n} \times \mu_{Y_A^n} \\ \mathbb{E}[(X_A^n - Y_A^n)^T \mathbf{G}_{A,I} (X_A^n - Y_A^n)] \le \Delta - \Delta_{\min,A}}} \frac{1}{n} \sum_{t=1}^{n} I(X_{At} \wedge Y_{At}) \\
&\ge \min_{\substack{\{\Delta_t,\ 1 \le t \le n\} \\ \frac{1}{n} \sum_{t=1}^{n} \Delta_t \le \Delta}} \frac{1}{n} \sum_{t=1}^{n} \min_{\substack{\mu_{X_{At} Y_{At}} \ll \mu_{X_{At}} \times \mu_{Y_{At}} \\ \mathbb{E}[(X_{At} - Y_{At})^T \mathbf{G}_{A,I} (X_{At} - Y_{At})] \le \Delta_t - \Delta_{\min,A}}} I(X_{At} \wedge Y_{At}) \qquad (4.56) \\
&= \min_{\substack{\{\Delta_t,\ 1 \le t \le n\} \\ \frac{1}{n} \sum_{t=1}^{n} \Delta_t \le \Delta}} \frac{1}{n} \sum_{t=1}^{n} \rho_A(\Delta_t) \\
&= R_A(\Delta),
\end{aligned}
$$

where (4.56) is since

$$
\mu_{X_A^n Y_A^n} \ll \mu_{X_A^n} \times \mu_{Y_A^n} \Rightarrow \mu_{X_{At} Y_{At}} \ll \mu_{X_{At}} \times \mu_{Y_{At}}, \qquad t = 1, \ldots, n. \quad (4.57)
$$

The claim (4.57) is easy to see by contradiction. Consider any real-valued rvs Z_1, Z_2, Z_3 with probability distribution $\mu_{Z_1 Z_2 Z_3} \ll \mu_{Z_1} \times \mu_{Z_2} \times \mu_{Z_3}$. Suppose, if possible, $\mu_{Z_1 Z_2}$ is not absolutely continuous with respect to $\mu_{Z_1} \times \mu_{Z_2}$, i.e., there exist $E_1, E_2 \in \mathcal{B}(\mathbb{R})$ such that

$$
\mu_{Z_1}(E_1) \times \mu_{Z_2}(E_2) = 0 \quad \text{and} \quad \mu_{Z_1 Z_2}(E_1 \times E_2) \neq 0. \quad (4.58)
$$

Considering a $E = E_1 \times E_2 \times \mathcal{B}(\mathbb{R})$, by (4.58) we have $(\mu_{Z_1} \times \mu_{Z_2} \times \mu_{Z_3})(E) = \mu_{Z_1}(E_1) \times \mu_{Z_2}(E_2) \times \mu_{Z_3}(\mathbb{R}) = 0$ but

$$\mu_{Z_1 Z_2 Z_3}(E) \neq 0,$$

since $\mu_{Z_1 Z_2}(E_1 \times E_2) \neq 0$, a contradiction, since $\mu_{Z_1 Z_2 Z_3} \ll \mu_{Z_1} \times \mu_{Z_2} \times \mu_{Z_3}$. ■

Note that a converse proof for Theorem 4.1 can be provided along the lines of the converse proof for Proposition 4.1. However, we prefer the current manner of presentation which provides for unity of ideas.

4.5 Discussion

In this chapter, we restricted our attention to a fixed-set sampler and our formulation of universality in this chapter the underlying pdf of a GMMS was known only to belong to a *convex, compact* set of pdfs $\mathcal{P}$. General extensions to a GMMS sampled by a k-random sampler for any arbitrary $\mathcal{P}$ (countable or uncountable) remain open.

In Section 4.3, our study involved a brief detour to Gaussian random field on an interval. As illustrated by Example 4.2, this allowed for a better understanding of the structure of the optimal fixed-set sampling. Aspects of sampling and compression of Markov random fields and Gaussian random fields have been the subject of extensive studies. It will be interesting to explore the questions posed in this dissertation in the context of these fields.

In Chapter 2, for a DMMS with known pmf, it was seen that a k-IRS and a k-MRS with memory did not improve the SRDf. In an universal setting, adaptive samplers such as

$$P_{S_t|X_{\mathcal{M}}^t S^{t-1}} = P_{S_t|S^{t-1} X_S^{t-1}}, \quad t = 1, \ldots$$

can be expected to be strictly superior to their counterparts with memory. This can easily be seen, for instance, in terms of the improvement in $\Delta_{\min}$. Such samplers allow for the design of sequential decision-making algorithms which can adaptively identify the optimal sampling sets in an "online" manner. For any $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$, the optimal sequential decision-making sampling algorithms, in its limit, can be expected to approximate the optimal k-IRS. An initial study of the design of sequential decision-making algorithms for achieving minimum expected distortion, in a multi-armed bandit setup, is already underway.

Appendix A: Proof of Theorem 2.3 and Proposition 3.1

We begin with the proof of Proposition 3.1 and show that the proof of Theorem 2.3 is obtained from the proof of Proposition 3.1 as a special case, with $|\Theta| = 1$.

First, for Proposition 3.1, for the Bayesian setting, by Theorem 3.3, the claim entails showing that

$$\min_{\substack{P_U, P_{S|X_{\mathcal{M}}U}, \{\Delta_\tau,\ \tau\in\Theta\} \\ \mathbb{E}[\Delta_\theta]\le\Delta}} \max_{\tau\in\Theta} \min_{\substack{P_{Y_B|SX_SU,\theta=\tau} \\ \mathbb{E}[d(X_B,Y_B)|\theta=\tau]\le\Delta_\tau}} I(X_S \wedge Y_B|S,U,\theta=\tau) \tag{A.1}$$

$$= \min_{\substack{P_U, \delta_w, \{\Delta_\tau,\ \tau\in\Theta\} \\ \mathbb{E}[\Delta_\theta]\le\Delta}} \max_{\tau\in\Theta} \min_{\substack{P_{Y_B|SX_SU,\theta=\tau} \\ \mathbb{E}[d(X_B,Y_B)|\theta=\tau]\le\Delta_\tau}} I(X_S \wedge Y_B|S,U,\theta=\tau), \tag{A.2}$$

for $\Delta_{\min} \le \Delta \le \Delta_{\max}$. Denote the expressions in (A.1) and (A.2) by $q(\Delta)$ and $r(\Delta)$, respectively. Now, from the conditional version of Topsøe's identity [34, Lemma 8.5], observe that $q(\Delta)$ equals

$$\min_{\substack{P_U, P_{S|X_{\mathcal{M}}U}, \\ \{\Delta_\tau,\tau\in\Theta\} \\ \mathbb{E}[\Delta_\theta]\le\Delta}} \max_{\tau\in\Theta} \min_{\substack{P_{Y_B|SX_SU,\theta=\tau} \\ \mathbb{E}[d(X_B,Y_B)|\theta=\tau]\le\Delta_\tau}} \min_{Q_{Y_B|SU,\theta=\tau}} D\left(P_{Y_B|SX_SU,\theta=\tau} \middle|\middle| Q_{Y_B|SU,\theta=\tau} \middle| P_{SX_SU|\theta=\tau}\right). \tag{A.3}$$

Note that the inner max and min can be interchanged in (A.3). Denoting

$D\left(P_{Y_B|SX_SU,\theta=\tau}\middle|\middle|Q_{Y_B|SU,\theta=\tau}\middle|P_{SX_SU|\theta=\tau}\right)$ by $D_\tau,\ \tau\in\Theta$, we write (A.3) as

$$\begin{aligned}
&\min_{\substack{P_U,P_{S|X_{\mathcal{M}}U},\{\Delta_\tau,\ \tau\in\Theta\}\\ \mathbb{E}[\Delta_\theta]\leq\Delta}}\ \min_{\substack{P_{Y_B|SX_SU\theta},Q_{Y_B|SU,\theta=\tau}\\ \mathbb{E}[d(X_B,Y_B)|\theta=\tau]\leq\Delta_\tau,\ \tau\in\Theta}}\ \max_{\tau\in\Theta} D_\tau\\
&=\min_{\substack{P_U,P_{S|X_{\mathcal{M}}U},P_{Y_B|SX_SU\theta},Q_{Y_B|SU,\theta=\tau}\\ \mathbb{E}[d(X_B,Y_B)]\leq\Delta}}\ \max_{\tau\in\Theta} D_\tau\\
&=\min_{\substack{t,P_U,P_{S|X_{\mathcal{M}}U},P_{Y_B|SX_SU\theta},Q_{Y_B|SU,\theta=\tau}\\ D_\tau\leq t,\ \tau\in\Theta\\ \mathbb{E}[d(X_B,Y_B)]\leq\Delta}} t,
\end{aligned}\tag{A.4}$$

which is the epigraph form. Also, $r(\Delta)$ can be expressed in a similar manner. Based on (A.4), we define $G_q(\alpha,\{\lambda_\tau,\tau\in\Theta\})$ and $G_r(\alpha,\{\lambda_\tau,\tau\in\Theta\})$ in terms of the Lagrangians of $q(\Delta)$ and $r(\Delta)$, respectively, in a standard way.

Specifically, $G_q(\alpha,\{\lambda_\tau,\ \tau\in\Theta\})$

$$\begin{aligned}
&=\min_{\substack{t,P_U,P_{S|X_{\mathcal{M}}U}\\ P_{Y_B|SX_SU\theta},Q_{Y_B|SU\theta}}} t+\sum_{\tau\in\Theta}\lambda_\tau(D_\tau-t)+\alpha\mathbb{E}\left[d(X_B,Y_B)\right]\\
&=\min_{\substack{t,P_U,P_{S|X_{\mathcal{M}}U}\\ P_{Y_B|SX_SU\theta},Q_{Y_B|SU\theta}}} t(1-\sum_{\tau\in\Theta}\lambda_\tau)+\sum_{\tau\in\Theta}\lambda_\tau D_\tau+\alpha\mathbb{E}\left[d(X_B,Y_B)\right]\\
&=\begin{cases}\displaystyle\min_{\substack{P_U,P_{S|X_{\mathcal{M}}U}\\ P_{Y_B|SX_SU\theta},Q_{Y_B|SU\theta}}} \sum_{\tau\in\Theta}\lambda_\tau D_\tau+\alpha\mathbb{E}\left[d(X_B,Y_B)\right], & \text{if } \sum_{\tau\in\Theta}\lambda_\tau=1\\
-\infty, & \text{otherwise.}\end{cases}
\end{aligned}\tag{A.5}$$

Let $P_\tau\triangleq P_{X_{\mathcal{M}}|\theta=\tau}$. When $\sum_{\tau\in\Theta}\lambda_\tau=1$, from (A.5), $G_q(\alpha,\{\lambda_\tau,\ \tau\in\Theta\})$ equals

$$\min_{\substack{P_U,Q_{Y_B|SU\theta},\\ P_{Y_B|SX_SU\theta}}}\sum_{u,x_{\mathcal{M}}}P_U(u)\min_{P_{S|X_{\mathcal{M}}U}}\sum_{s\in\mathcal{A}_k}P_{S|X_{\mathcal{M}}U}(s|x_{\mathcal{M}},u)\times$$

$$\Bigg(\mathbb{E}\Big[\sum_{\tau\in\Theta}\lambda_\tau P_\tau(x_{\mathcal{M}})\log\frac{P_{Y_B|SX_SU\theta}(Y_B|s,x_s,u,\tau)}{Q_{Y_B|SU\theta}(Y_B|s,u,\tau)}$$

$$+\alpha\sum_{\tau\in\Theta}\mu_\theta(\tau)P_\tau(x_{\mathcal{M}})d(x_B,Y_B)\Big|S=s,X_S=x_s,U=u,\theta=\tau\Big]\Bigg),$$

where the expectation above is with respect to $P_{Y_B|S=s,X_S=x_s,U=u,\theta=\tau}$. Noting that the term $\Big(\cdots\Big)$ above is a function of $s,x_{\mathcal{M}},u$, we get

$$G_q(\alpha,\{\lambda_\tau,\ \tau\in\Theta\})$$

$$=\min_{\substack{P_U,Q_{Y_B|SU\theta}\\P_{Y_B|SX_SU\theta}}}\sum_{u,x_{\mathcal{M}}}P_U(u)\min_{s\in\mathcal{A}_k}\Bigg(\mathbb{E}\Big[\sum_{\tau\in\Theta}\lambda_\tau P_\tau(x_{\mathcal{M}})\log\frac{P_{Y_B|SX_SU\theta}(Y_B|s,x_s,u,\tau)}{Q_{Y_B|SU\theta}(Y_B|s,u,\tau)}$$

$$+\alpha\sum_{\tau\in\Theta}\mu_\theta(\tau)P_\tau(x_{\mathcal{M}})d(x_B,Y_B)\Big|S=s,X_S=x_s,U=u,\theta=\tau\Big]\Bigg)$$

$$=\min_{\substack{P_U,Q_{Y_B|SU\theta}\\P_{Y_B|SX_SU\theta}}}\sum_{u,x_{\mathcal{M}}}P_U(u)\ \min_{\delta_{w(\cdot,\cdot)}}\sum_{s\in\mathcal{A}_k}\delta_{w(x_{\mathcal{M}},u)}(s)\times$$

$$\Bigg(\mathbb{E}\Big[\sum_{\tau\in\Theta}\lambda_\tau P_\tau(x_{\mathcal{M}})\log\frac{P_{Y_B|SX_SU\theta}(Y_B|s,x_s,u,\tau)}{Q_{Y_B|SU\theta}(Y_B|s,u,\tau)}$$

$$+\alpha\sum_{\tau\in\Theta}\mu_\theta(\tau)P_\tau(x_{\mathcal{M}})d(x_B,Y_B)\Big|S=s,X_S=x_s,U=u,\theta=\tau\Big]\Bigg)$$

$$=\min_{\substack{P_U,Q_{Y_B|SU\theta}\\P_{Y_B|SX_SU\theta},\delta_w}}\sum_{\tau\in\Theta}\lambda_\tau D\left(P_{Y_B|SX_SU,\theta=\tau}\middle|\middle|Q_{Y_B|SU,\theta=\tau}\middle|P_{SX_SU|\theta=\tau}\right)+\alpha\mathbb{E}\left[d(X_B,Y_B)\right]$$

$$=G_r(\alpha,\{\lambda_\tau,\ \tau\in\Theta\}).$$

Since $q(\Delta)$ and $r(\Delta)$ are convex in Δ, they can be expressed in terms of their respective Lagrangians as

$$\begin{aligned} q(\Delta) &= \max_{\alpha\geq 0,\ \{\lambda_\tau\geq 0,\ \tau\in\Theta\}} G_q(\alpha, \{\lambda_\tau, \tau\in\Theta\}) - \alpha\Delta \ \text{ and} \\ r(\Delta) &= \max_{\alpha\geq 0,\ \{\lambda_\tau\geq 0,\ \tau\in\Theta\}} G_r(\alpha, \{\lambda_\tau, \tau\in\Theta\}) - \alpha\Delta. \end{aligned} \tag{A.6}$$

Thus,

$$\begin{aligned} q(\Delta) &= \max_{\alpha\geq 0,\ \{\lambda_\tau\geq 0,\ \tau\in\Theta\}} G_q(\alpha, \{\lambda_\tau,\ \tau\in\Theta\}) - \alpha\Delta \\ &= \max_{\substack{\alpha\geq 0,\ \{\lambda_\tau\geq 0,\ \tau\in\Theta\} \\ \sum_{\tau\in\Theta}\lambda_\tau=1}} G_q(\alpha, \{\lambda_\tau,\ \tau\in\Theta\}) - \alpha\Delta \\ &= \max_{\substack{\alpha\geq 0,\ \{\lambda_\tau\geq 0,\ \tau\in\Theta\} \\ \sum_{\tau\in\Theta}\lambda_\tau=1}} G_r(\alpha, \{\lambda_\tau,\ \tau\in\Theta\}) - \alpha\Delta \\ &= r(\Delta), \end{aligned}$$

upon observing that the maxima in (A.6) are attained when $\sum_{\tau\in\Theta}\lambda_\tau = 1$. The nonBayesian setting is shown using a similar set of ideas.

In the nonBayesian setting, the proof of (3.30) can be provided along the lines of proof above with the distinction that the minima in (A.1), (A.2) are now over $P_{UX_\mathcal{M}SY_B|\theta=\tau}$ such that $\mathbb{E}[d(X_B, Y_B)|\theta = \tau] \leq \Delta,\ \tau\in\Theta$. This distinction is then maintained throughout the proof.

The proof of Theorem 2.3 can be obtained from the steps above for $|\Theta| = 1$ and with $P_{X_\mathcal{M}}$, the pmf of $X_\mathcal{M}$, in place of $P_{X_\mathcal{M}|\theta=\tau}$.

Appendix B: Proofs of Technical Lemmas in Gaussian Sampling Rate Distortion

B.1 Proof of Lemma 4.3

Recall that the elements of the compact sets Θ and Θ_1 are indexed by τ and τ_1, which take values in $\mathbb{R}^{m^2}$ and $\mathbb{R}^{k^2}$ respectively. Now, every $\tau \in \Theta$ can be seen as $\tau = (\tau_1, \tau_2)$ with τ_2 taking values in Θ_2, a bounded subset of $\mathbb{R}^{m^2-k^2}$. A continuous function over a compact set is uniformly continuous, hence, for every $x_{\mathcal{M}} \in \mathbb{R}^m$,

$$\nu_{X_{\mathcal{M}}|\theta}(x_{\mathcal{M}}|\tau_1, \tau_2) \quad \text{and} \quad \nu_\theta(\tau_1, \tau_2)$$

are uniformly continuous in (τ_1, τ_2). Furthermore, as a function of τ_2, $\nu_{X_{\mathcal{M}}|\theta}(x_{\mathcal{M}}|\tau_1, \tau_2)$ and $\nu_\theta(\tau_1, \tau_2)$ are bounded functions over bounded set Θ_2 and hence so is $\nu_{X_{\mathcal{M}}|\theta}(x_{\mathcal{M}}|\tau_1, \tau_2)\nu_\theta(\tau_1, \tau_2)$. By the Bounded Convergence Theorem, for every $x_{\mathcal{M}} \in \mathbb{R}^m$ and $\tau_1 \in \Theta_1$

$$\lim_{\tilde{\tau}_1 \to \tau_1} \nu_{\theta_1}(\tilde{\tau}_1) = \lim_{\tilde{\tau}_1 \to \tau_1} \int_{\Theta_2} \nu_\theta(\tilde{\tau}_1, \tau_2)\, d\tau_2 = \int_{\Theta_2} \nu_\theta(\tau_1, \tau_2)\, d\tau_2 = \nu_{\theta_1}(\tau_1) \tag{B.1}$$

and

$$\lim_{\tilde{\tau}_1 \to \tau_1} \int_{\Theta_2} \nu_{X_{\mathcal{M}}|\theta}(x_{\mathcal{M}}|\tilde{\tau}_1, \tau_2)\nu_\theta(\tau_1, \tau_2)\, d\tau_2 = \int_{\Theta_2} \nu_{X_{\mathcal{M}}|\theta}(x_{\mathcal{M}}|\tau_1, \tau_2)\nu_\theta(\tau_1, \tau_2)\, d\tau_2, \quad \text{(B.2)}$$

and thus from (B.1) and (B.2),

$$\lim_{\tilde{\tau}_1 \to \tau_1} \nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tilde{\tau}_1) = \nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tau_1).$$

Continuity of $\nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tau_1)$ in τ_1, implies that for $i = 1, \ldots, m$, and $t = 1, \ldots, n$,

$$\mathbb{E}\big[\big(X_i - (\varphi(f(X_A^n)))_{i,t}\big)^2 | \theta_1 = \tau_1\big]$$

is continuous in τ_1. The continuity of

$$\mathbb{E}\big[||X_{\mathcal{M}}^n - \varphi(f(X_A^n))||^2 \big| \theta_1 = \tau_1\big] \quad \text{(B.3)}$$

in τ_1 is now immediate. Since Θ_1 is a compact set, (B.3) is uniformly continuous in τ_1.

■

B.2 Proof of Lemma 4.4

First, observe that for every $\tau_1 \in \Theta_1$, $\Lambda(\tau_1)$ is a convex, compact set. Now, the minimum in (4.29) exists as that of a continuous function over a compact set. It is seen in a standard manner that the convexity of Θ and Θ_1 imply the convexity of

$$g(\tau_{1,1}, \tau_{1,2}) \triangleq \min_{\hat{\tau} \in \Lambda(\tau_{1,2})} \min_{\check{\tau} \in \Lambda(\tau_{1,1})} ||\hat{\tau} - \check{\tau}||$$

in $(\tau_{1,1}, \tau_{1,2})$. Consequently, $g(\tau_{1,1}, \tau_{1,2})$ is continuous in $(\tau_{1,1}, \tau_{1,2})$. Define

$$D(\delta) \triangleq \max_{\substack{\tau_{1,1}, \tau_{1,2} \in \Theta_1 \\ ||\tau_{1,1} - \tau_{1,2}|| \leq \delta}} g(\tau_{1,1}, \tau_{1,2}).$$

Clearly, $D(0) = 0$ and $D(\delta)$ is a continuous nondecreasing function of δ (Chapter 20, [42]).

Now, we prove the lemma by contradiction. Suppose if possible, there exists an $\epsilon > 0$ such that for every $\delta > 0$ there exist $\tau_{1,1,\delta}, \tau_{1,2,\delta} \in \Theta_1$ with $||\tau_{1,1,\delta} - \tau_{1,2,\delta}|| \leq \delta$ and

$$g(\tau_{1,1,\delta}, \tau_{1,2,\delta}) > \epsilon.$$

Then,

$$0 = D(0) = \lim_{\delta \to 0} D(\delta) = \lim_{\delta \to 0} \max_{\tau_{1,1,\delta},\ \tau_{1,2,\delta}:\ ||\tau_{1,1,\delta} - \tau_{1,2,\delta}|| \leq \delta} g(\tau_{1,1,\delta}, \tau_{1,2,\delta}) \geq \epsilon,$$

a contradiction. Hence, the lemma.

$\blacksquare$

Appendix C: Proof of Key Claims

C.1 Proof of Claim in Achievability Proof of Theorem 3.3

For the code formed by concatenating (f, φ) for each $u \in \mathcal{U}$, the rate is

$$\overset{\sim}{\leq} \max_{\tau\in\Theta} \sum_{u\in\mathcal{U}} P_U(u) \frac{1}{n'} \sum_{i=1}^{M_k} \log J_{A_i}^{u,\tau}$$

$$\leq \max_{\tau\in\Theta} \sum_{u\in\mathcal{U}} P_U(u) \Big(\sum_{i=1}^{M_k} \frac{|\nu_{A_i}^{u,\tau}|}{n} \frac{1}{|\nu_{A_i}^{u,\tau}|} \log J_{A_i}^{u,\tau} \Big)$$

$$\leq \max_{\tau\in\Theta} \sum_{u\in\mathcal{U}} P_U(u) \Big(\sum_{i=1}^{M_k} P_{S|U\theta}(A_i|u,\tau) \Big(I(X_{A_i} \wedge Y_B | S = A_i, U = u, \theta = \tau) + \frac{\epsilon'}{2} \Big) \Big), \text{by (3.46)}$$

$$\leq \max_{\tau\in\Theta} \sum_{u\in\mathcal{U}} P_U(u)\ I(X_S \wedge Y_B | S, U = u, \theta = \tau) + \epsilon'$$

$$\leq R_m(\Delta) + \epsilon,$$

for all n large enough.

For each $U = u$, let $\Delta_u \triangleq \sum_{\tau\in\Theta,\ A_i\in\mathcal{A}_k} \mu_\theta(\tau) P_{S|U\theta}(A_i|u,\tau)\Delta_{A_i,u,\tau}$. Denoting the output of the decoder by $Y_B^{n'}$, we get

$$\begin{aligned}\mathbb{E}[d(X_B^{n'}, Y_B^{n'})] &\leq P(\widehat{\tau}_N \neq \theta) d_{\max} + \mathbb{E}[\mathbb{1}(\widehat{\tau}_N = \theta) d(X_B^{n'}, Y_B^{n'})] \\ &\leq P(\widehat{\tau}_N \neq \theta) d_{\max} + P(S^\gamma \notin \mathcal{T}^{(n)}(\epsilon', \widehat{\tau}_N)) d_{\max}\end{aligned}$$

$$+\,\mathbb{E}\big[\mathbb{E}[\mathbb{1}(\widehat{\tau}_N=\theta, S^\gamma\in\mathcal{T}^{(n)}(\epsilon',\widehat{\tau}_N))d(X_B^{n'},Y_B^{n'})|S^\gamma,\theta]\big] \quad \text{(C.1)}$$

$$\le \mathbb{E}[\Delta_{S,U,\theta}|U=u]+\epsilon \quad \text{(C.2)}$$

$$=\Delta_u+\epsilon$$

for all n,N large enough, where the previous inequality is shown below. Then, expected distortion for the code formed by concatenating (f,φ) for each $u\in\mathcal{U}$, is

$$\stackrel{\sim}{\le}\mathbb{E}[\Delta_U]+\epsilon\le\Delta+\epsilon.$$

It remains to show (C.2). Now, (C.2) follows from the following: In (C.1), for each $\tau\in\Theta$ and $s^n\in\mathcal{T}^{(n)}(\epsilon',\widehat{\tau}_N)$,

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}(\widehat{\tau}_N=\theta)d(X_B^{n'},Y_B^{n'})|S^\gamma=s^n,\theta=\tau]\\
&\quad=\mathbb{E}\Big[\frac{\mathbb{1}(\widehat{\tau}_N=\theta)}{n'}\sum_{t\in\mu}d(X_{Bt},Y_{Bt})+\frac{\mathbb{1}(\widehat{\tau}_N=\theta)}{n'}\sum_{t\in\gamma}d(X_{Bt},Y_{Bt})\big|S^\gamma=s^n,\theta=\tau\Big]\\
&\quad\le\frac{N'}{n'}d_{\max}+\frac{1}{n}\mathbb{E}\Big[\sum_{i=1}^{M_k}\sum_{t\in\gamma_{s^n}(A_i)\setminus\nu_{A_i}}d(X_{Bt},Y_{Bt})|S^\gamma=s^n,\theta=\tau\Big]\\
&\qquad+\sum_{i=1}^{M_k}\mathbb{E}\Big[\frac{|\nu_{A_i}|}{n}\mathbb{1}(\widehat{\tau}_N=\theta)d(X_B^{\nu_{A_i}},\varphi_{A_i}^\theta(f_{A_i}^\theta(X_{A_i}^{\nu_{A_i}})))\big|S^{\nu_{A_i}}=A_i^{\nu_{A_i}},\theta=\tau\Big]\\
&\quad\le\frac{N'}{n'}d_{\max}+M_k\epsilon' d_{\max}+\sum_{i=1}^{M_k}P_{S|U\theta}(A_i|u,\tau)\Big(\Delta_{A_i,u,\tau}+\frac{\epsilon'}{4}\Big),\qquad\text{by (3.47)}\\
&\quad\le\mathbb{E}[\Delta_{S,U,\theta}|U=u,\theta=\tau]+M_k\epsilon' d_{\max}+\frac{N'}{n'}d_{\max}+\frac{\epsilon'}{4}\\
&\quad\le\mathbb{E}[\Delta_{S,U,\theta}|U=u,\theta=\tau]+\epsilon,
\end{aligned}$$

for all n large enough and ϵ' chosen appropriately. ■

C.2 Proof of Existence of Minimum and Maximum in (4.17)

For every $\tau_1 \in \Theta_1$, recall that $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ is, in effect, a rate distortion function, hence its inverse $D_A^{\mathcal{B}}(R, \tau_1)$ is well defined over $R \geq 0$. Continuity of $\nu_{X_{\mathcal{M}}|\theta_1}(x_{\mathcal{M}}|\tau_1)$ in τ_1 for every $x_{\mathcal{M}} \in \mathbb{R}^m$ implies the continuity of $D_A^{\mathcal{B}}(R, \tau_1)$ in τ_1.

We now show the existence of the minimum and maximum on the right-side of (4.17)

$$\inf_{\substack{\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}] \leq \Delta}} \sup_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) = \min_{\substack{\{\Delta_{\tau_1},\ \tau_1 \in \Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}] \leq \Delta}} \max_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1). \tag{C.3}$$

Denote the left-side of (C.3) by r and choose

$$\Delta_{\tau_1}^* = D_A^{\mathcal{B}}(r, \tau_1), \qquad \tau_1 \in \Theta_1.$$

The continuity of $D_A^{\mathcal{B}}(r, \tau_1)$ in τ_1 implies the continuity of $\Delta_{\tau_1}^*$ in τ_1 and hence $\mathbb{E}[\Delta_{\theta_1}^*]$ exists. A simple proof of contradiction can be used to show that $\mathbb{E}[\Delta_{\theta_1}^*] \leq \Delta$. Thus, $\{\Delta_{\tau_1}^*,\ \tau_1 \in \Theta_1\}$ satisfies the constraint on the left-side of (C.3) and for every $\tau_1 \in \Theta_1$, $\rho_A^{\mathcal{B}}(\Delta_{\tau_1}^*, \tau_1) = r$, with

$$\sup_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}^*, \tau_1) = r$$

and hence (C.3) holds. ■

C.3 Proof of Claim in Achievability Proof of Theorem 4.2

Noting that $\tilde{\theta}_{1,n}(X_A^n)$ is a deterministic function of X_A^n, for $\tau_1 \in \Theta_1$ and $\tau_{1,1} \in \Theta_{1,\delta}$ with $||\tau_1 - \tau_{1,1}|| \leq 2\delta$ and $P_{\tilde{\theta}_{1,n}|\theta_1}(\tau_{1,1}|\tau_1) > 0$,

$$\begin{aligned}
&\mathbb{E}\big[\big|\big|X_{\mathcal{M}}^n - \varphi(\tau_{1,1}, f_{\tau_{1,1}}(X_A^n))\big|\big|^2\big|\theta_1 = \tau_1, \tilde{\theta}_{1,n} = \tau_{1,1}\big] \\
&= \frac{1}{P_{\tilde{\theta}_{1,n}|\theta_1}(\tau_{1,1}|\tau_1)}\mathbb{E}\big[\mathbb{1}\big(\tilde{\theta}_{1,n}(X_A^n) = \tau_{1,1}\big)\big|\big|X_{\mathcal{M}}^n - \varphi\big(\tau_{1,1}, f_{\tau_{1,1}}(X_A^n)\big)\big|\big|^2\big|\theta_1 = \tau_1\big] \qquad \text{(C.4)} \\
&\leq \frac{1}{P_{\tilde{\theta}_{1,n}|\theta_1}(\tau_{1,1}|\tau_1)}\mathbb{E}\big[\big|\big|X_{\mathcal{M}}^n - \varphi\big(\tau_{1,1}, f_{\tau_{1,1}}(X_A^n)\big)\big|\big|^2\big|\theta_1 = \tau_1\big] \\
&\leq \frac{1}{P_{\tilde{\theta}_{1,n}|\theta_1}(\tau_{1,1}|\tau_1)}\Big(\mathbb{E}\big[\big|\big|X_{\mathcal{M}}^n - \varphi\big(\tau_{1,1}, f_{\tau_{1,1}}(X_A^n)\big)\big|\big|^2\big|\theta_1 = \tau_{1,1}\big] + \epsilon_1\Big) \text{ by Lemma 4.3}
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{P_{\tilde{\theta}_{1,n}|\theta_1}(\tau_{1,1}|\tau_1)}\big(\Delta_{\tau_{1,1}} + 2\epsilon_1\big) \qquad \text{(C.5)} \\
&\leq \frac{1}{P_{\tilde{\theta}_{1,n}|\theta_1}(\tau_{1,1}|\tau_1)}\big(\Delta_{\tau_1} + 3\epsilon_1\big) \qquad \text{(C.6)}
\end{aligned}$$

where

(i) (C.4) is since $\tilde{\theta}_{1,n}(X_A^n)$ is a deterministic function of X_A^n;

(ii) it is seen along the lines of the achievability proof of Theorem 4.1 that

$$\mathbb{E}\big[||X_{\mathcal{M}}^n - \varphi(\tau_{1,1}, f_{\tau_{1,1}}(X_A^n))||^2\big|\theta_1 = \tau_{1,1}\big] \leq \Delta_{\tau_{1,1}} + \epsilon_1,$$

and hence (C.5) is obtained;

(iii) Δ_{τ_1} is continuous in τ_1 over the compact set Θ_1, hence, Δ_{τ_1} is in fact uniformly continuous in τ_1; (C.6) now follows.

From (C.6), the first term on the right-side of (4.34) is

$$\begin{aligned}
&\mathbb{E}[\mathbb{1}(||\tilde{\theta}_{1,n} - \theta_1|| \leq 2\delta)||X_{\mathcal{M}}^n - \varphi(f(X_A^n))||^2] \\
&\qquad \leq \sum_{\tilde{\tau}_1 \in \Theta_{1,\delta}} \mathbb{E}[\mathbb{1}(||\theta_1 - \tilde{\tau}_1|| \leq 2\delta)(\Delta_{\theta_1} + 3\epsilon_1)] \\
&\qquad \leq \mathbb{E}[\Delta_{\theta_1}] + 3\epsilon_1 \qquad\qquad \text{by (4.27)} \\
&\qquad \leq \Delta + 3\epsilon_1.
\end{aligned}$$

■

Appendix D: Standard Properties of SRDf and USRDf

D.1 Proof of Lemma 3.1

Clearly, for each $\tau_1 \in \Theta_1$, $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ and $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ are finite-valued and, hence, so are the right-sides of (3.10) and (3.11). Also, they are also nonincreasing in Δ. The convexity of the right-sides of (3.10) and (3.11) follows from the convexity of $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ and $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ in δ along with a standard argument shown below; continuity for $\Delta > \Delta_{\min}$ is a consequence. Continuity at $\Delta_{\min}$ holds, for instance, as in ([34], Lemma 7.2). The claimed properties of the right-sides of (3.18), (3.19), (3.24) and (3.26) follow in a similar manner.

The convexity of the right-side of (3.10) can be shown explicitly as follows. Let $\tau_1(1)$ and $\tau_1(2)$ attain the maximum in (3.10) at $\Delta = \Delta_1$ and $\Delta = \Delta_2$, respectively, where $\Delta_1 < \Delta_2$. The corresponding minimizing $\{\Delta_{\tau_1}, \tau_1 \in \Theta_1\}$ are denoted by $\{\Delta^1_{\tau_1},\ \tau_1 \in \Theta_1\}$ and $\{\Delta^2_{\tau_1},\ \tau_1 \in \Theta_1\}$, respectively. For any $0 < \alpha < 1$, for $i = 1, \ldots, |\Theta_1|$

$$\begin{aligned}
\alpha R_A(\Delta_1) + (1-\alpha) R_A(\Delta_2) &= \alpha \rho_A^{\mathcal{B}}(\Delta^1_{\tau_1(1)}, \tau_1(1)) + (1-\alpha)\rho_A^{\mathcal{B}}(\Delta^2_{\tau_1(2)}, \tau_1(2)) \\
&\geq \alpha \rho_A^{\mathcal{B}}(\Delta^1_{\tau_1(i)}, \tau_1(i)) + (1-\alpha)\rho_A^{\mathcal{B}}(\Delta^2_{\tau_1(i)}, \tau_1(i))
\end{aligned}$$

$$\geq \rho_A^{\mathcal{B}}(\alpha\Delta^1_{\tau_1(i)} + (1-\alpha)\Delta^2_{\tau_1(i)}, \tau_1(i)), \tag{D.1}$$

where the inequality above follows by Remark (iii) preceding Theorem 3.1. Now, (D.2) holds for every $i = 1, \ldots, |\Theta_1|$, hence

$$\begin{aligned}
\alpha R_A(\Delta_1) + (1-\alpha) R_A(\Delta_2) &\geq \max_i \rho_A^{\mathcal{B}}(\alpha\Delta^1_{\tau_1(i)} + (1-\alpha)\Delta^2_{\tau_1(i)}, \tau_1(i)) \\
&\geq \min_{\substack{\{\Delta_{\tau_1}, \tau_1 \in \Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}] \leq \alpha\Delta_1 + (1-\alpha)\Delta_2}} \max_{\tau_1 \in \Theta_1} \rho_A^{\mathcal{B}}(\Delta_{\tau_1}, \tau_1) \\
&= R_A(\alpha\Delta_1 + (1-\alpha)\Delta_2).
\end{aligned}$$

■

D.2 Proof of Lemma 4.1

The right-sides of (4.5) and (4.9) are, in effect, the standard rate distortion function for GMMS with weighted MSE distortion criterion, and hence are finite-valued, decreasing, convex, continuous functions of $\Delta > \Delta_{\min,A}$ and $\Delta > \Delta_{\min,A,\tau_1}$, respectively.

The right-sides of (4.17) and (4.18) are clearly nonincreasing functions of Δ. Convexity of the right-sides of (4.17) and (4.18) follows from the convexity of $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ and $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ using standard arguments; continuity for $\Delta > \Delta_{\min,A,\tau_1}$ is a consequence. Finite-valuedness of (4.17) and (4.18) follows from the finite-valuedness of $\rho_A^{\mathcal{B}}(\delta, \tau_1)$ and $\rho_A^{n\mathcal{B}}(\delta, \tau_1)$ for $\delta > \Delta_{\min,A,\tau_1}$, respectively.

The convexity of the right-side of (4.17) can be shown explicitly as follows. Let

$\tau_1(1)$ and $\tau_1(2)$ attain the maximum in (4.17) at $\Delta = \Delta_1$ and $\Delta = \Delta_2$, respectively, where $\Delta_1 < \Delta_2$. For $\Delta_1, \Delta_2 > \Delta_{\min,A}$, let $\{\Delta^1_{\tau_1},\ \tau_1 \in \Theta_1\}$ and $\{\Delta^2_{\tau_1},\ \tau_1 \in \Theta_1\}$, attain the minimum in (4.17), respectively and are as in Appendix C.2. For any $0 < \alpha < 1$, and $\tilde{\tau}_1 \in \Theta_1$,

$$
\begin{aligned}
\alpha R_A(\Delta_1) + (1-\alpha) R_A(\Delta_2) &= \alpha \rho^{\mathcal{B}}_A(\Delta^1_{\tilde{\tau}_1}, \tilde{\tau}_1) + (1-\alpha)\rho^{\mathcal{B}}_A(\Delta^2_{\tilde{\tau}_1}, \tilde{\tau}_1) \\
&\geq \rho^{\mathcal{B}}_A(\alpha \Delta^1_{\tilde{\tau}_1} + (1-\alpha)\Delta^2_{\tilde{\tau}_1}, \tilde{\tau}_1),
\end{aligned}
\tag{D.2}
$$

by the convexity of $\rho^{\mathcal{B}}_A(\delta, \tilde{\tau}_1)$ in δ. Now, (D.2) holds for every $\tilde{\tau}_1 \in \Theta_1$, hence

$$
\begin{aligned}
\alpha R_A(\Delta_1) + (1-\alpha) R_A(\Delta_2) &\geq \sup_{\tilde{\tau}_1 \in \Theta_1} \rho^{\mathcal{B}}_A(\alpha \Delta^1_{\tilde{\tau}_1} + (1-\alpha)\Delta^2_{\tilde{\tau}_1}, \tilde{\tau}_1) \\
&\geq \inf_{\substack{\{\Delta_{\tau_1}, \tau_1 \in \Theta_1\} \\ \mathbb{E}[\Delta_{\theta_1}] \leq \alpha\Delta_1 + (1-\alpha)\Delta_2}} \sup_{\tau_1 \in \Theta_1} \rho^{\mathcal{B}}_A(\Delta_{\tau_1}, \tau_1) \\
&= R_A(\alpha \Delta_1 + (1-\alpha)\Delta_2).
\end{aligned}
$$

$\blacksquare$

Bibliography

[1] A. Giridhar and P. R. Kumar. Computing and communicating functions over sensor networks. *IEEE Journal on Selected Areas in Communications*, 23(4):755–764, April 2005.

[2] N. M. Freris, H. Kowshik, and P. R. Kumar. Fundamentals of large sensor networks: Connectivity, capacity, clocks, and computation. *Proc. IEEE*, 98(11):1828–1846, November 2010.

[3] T. Berger. *Rate distortion theory: A mathematical basis for data compression*. Prentice-Hall, 1971.

[4] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[5] Y. Wu and S. Verdú. Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Transactions on Information Theory*, 56(8):3721–3748, August 2010.

[6] T. Kawabata and A. Dembo. The rate-distortion dimension of sets and measures. *IEEE Transactions on Information Theory*, 40(5):1564–1572, December 1994.

[7] Y. Wu and S. Verdú. Optimal phase transitions in compressed sensing. *IEEE Transactions on Information Theory*, 58(10):6241–6263, October 2012.

[8] G. Reeves and M. Gastpar. The sampling rate-distortion tradeoff for sparsity pattern recovery in compressed sensing. *IEEE Transactions on Information Theory*, 58(5):3065–3092, May 2012.

[9] C. Weidmann and M. Vetterli. Rate distortion behavior of sparse sources. *IEEE Transactions on Information Theory*, 58(8):4969–4992, August 2012.

[10] J. Z. Sun and V. K. Goyal. Intersensor collaboration in distributed quantization networks. *IEEE Transactions on Communications*, 61(9):3931–3942, September 2013.

[11] P. Ishwar, A. Kumar, and K. Ramchandran. Distributed sampling for dense sensor networks: A "bit-conservation principle". *Preprint available at www.eecs.berkeley.edu/∼animesh/jsac03ishwartetal.pdf.*

[12] A. Kipnis, A. J. Goldsmith, Y. C. Eldar, and T. Weissman. Distortion rate function of sub-nyquist sampled gaussian sources. *IEEE Transactions on Information Theory*, 62(1):401–429, January 2016.

[13] X. Liu, O. Simeone, and E. Erkip. Lossy computing of correlated sources with fractional sampling. *IEEE Transactions on Communications*, 61(9):3685–3696, August 2013.

[14] D. L. Neuhoff and S. S. Pradhan. Information rates of densely sampled gaussian data. *Proc. IEEE International Symposium on Information Theory (ISIT), 2011*, pages 2776–2780, July 31-Aug. 5, 2011.

[15] R. L. Konsbruck, E. Telatar, and M. Vetterli. On sampling and coding for distributed acoustic sensing. *IEEE Transactions on Information Theory*, 58(5):3198–3214, May 2012.

[16] A. Kashyap, L. A. Lastras-Montano, C. Xia, and Z. Liu. Distributed source coding in dense sensor networks. *Proc. Data Compression Conference, 2005*, pages 13–22, March 29-31, 2005.

[17] J. Ziv. Coding of sources with unknown statistics–ii: Distortion relative to a fidelity criterion. *IEEE Transactions on Information Theory*, 18(3):389–394, May 1972.

[18] D. Neuhoff, R. Gray, and L. Davisson. Fixed rate universal block source coding with a fidelity criterion. *IEEE Transactions on Information Theory*, 21(5):511–523, September 1975.

[19] D. Neuhoff and P. Shields. Fixed-rate universal codes for markov sources. *IEEE Transactions on Information Theory*, 24(3):360–367, May 1978.

[20] J. Rissanen. Universal coding, information, prediction, and estimation. *IEEE Transactions on Information Theory*, 30(4):629–636, July 1984.

[21] J. Ziv. Distortion-rate theory for individual sequences. *IEEE Transactions on Information Theory*, 26(2):137–143, March 1980.

[22] T. Weissman and N. Merhav. Universal prediction of individual binary sequences in the presence of noise. *IEEE Transactions on Information Theory*, 47(6):2151–2173, September 2001.

[23] T. Weissman and N. Merhav. On limited-delay lossy coding and filtering of individual sequences. *IEEE Transactions on Information Theory*, 48(3):721–733, March 2002.

[24] T. Linder, G. Lugosi, and K. Zeger. Fixed-rate universal lossy source coding and rates of convergence for memoryless sources. *IEEE Transactions on Information Theory*, 41(3):665–676, May 1995.

[25] T. Linder. On the training distortion of vector quantizers. *IEEE Transactions on Information Theory*, 46(4):1617–1623, July 2000.

[26] T. Linder. Learning-theoretic methods in vector quantization. *Principles of nonparametric learning*, pages 163–210, 2002.

[27] T. Linder, G. Lugosi, and K. Zeger. Empirical quantizer design in the presence of source noise or channel noise. *IEEE Transactions on Information Theory*, 43(2):612–623, March 1997.

[28] T. Weissman. Universally attainable error exponents for rate-distortion coding of noisy sources. *IEEE Transactions on Information Theory*, 50(6):1229–1246, June 2004.

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

[30] R. Dobrushin and B. Tsybakov. Information transmission with additional noise. *IRE Transactions on Information Theory*, 8(5):293–304, September 1962.

[31] T. Berger. The information theory approach to communications. *Multiterminal Source Coding, Springer, Berlin*, 1978.

[32] H. Yamamoto and K. Itoh. Source coding theory for multiterminal communication systems with a remote source. *IEICE TRANSACTIONS (1976-1990)*, 63(10):700–706, 1980.

[33] J. Wolf and J. Ziv. Transmission of noisy information to a noisy receiver with minimum distortion. *IEEE Transactions on Information Theory*, 16(4):406–411, July 1970.

[34] I. Csiszár and J. Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.

[35] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[36] S. Ihara. *Information theory for continuous systems*, volume 2. World Scientific, 1993.

[37] D. J. Sakrison. The rate distortion function for a class of sources. *Information and Control*, 15(2):165–195, August 1969.

[38] E. DiBenedetto. *Real analysis*. Springer, 2002.

[39] B. C. Levy. *Principles of signal detection and parameter estimation.* Springer Science & Business Media, 2008.

[40] V. P. Boda and P. Narayan. Universal sampling rate distortion. *Submitted to IEEE Transactions on Information Theory; arXiv preprint arXiv:1706.07409*, 2017.

[41] R. G. Gallager. *Information theory and reliable communication*, volume 2. Springer, 1968.

[42] K. A. Ross. *Elementary analysis.* Springer-Verlag, New York, 1980.