TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Nachrichtentechnik

Rate-Distortion Analysis of Sparse Sources and Compressed Sensing with Scalar Quantization

Lars Palzer

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor–Ingenieurs

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Holger Boche
Prüfer der Dissertation:	1. Prof. Dr. sc. techn. Gerhard Kramer
	2. Prof. Sundeep Rangan, Ph.D.
	3. Prof. Dr. Massimo Fornasier

Die Dissertation wurde am 22.05.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 07.10.2019 angenommen.

To Hanna

Acknowledgements

This thesis was written during my time as a research assistant at the Institute for Communications Engineering (LNT) at Technische Universität München (TUM). I would like to thank my supervisor Gerhard Kramer for accepting me into is research group and giving me the opportunity to conduct research on various topics.

I am indebted to my two main collaborators: Roy Timo and Johannes Maly. It has been a pleasure to work with you and learn from you.

The friendly and lively atmosphere at LNT has made this journey much easier than it might have been otherwise. I would like to thank all my colleagues at LNT, LÜT and COD, many of whom have become close friends, for the time we shared.

Most importantly, I wish to thank Hanna for her love, support, and constant encouragement.

München, May 2019 Lars Palzer

Contents

Zusammenfassung

Diese Arbeit befasst sich mit der digitalen Komprimierung dünnbesetzter Signale aus zwei Perspektiven: informationstheoretische Grenzen und Algorithmen basierend auf dem Prinzip des Compressed Sensing.

Für die informationstheoretische Untersuchung wird eine Rate-Distortion-Funktion mit individuellen Gütekriterien für verschiedene Teile des Signals sowohl für endliche als auch für unendliche Blocklängen betrachtet. Für dünnbesetzte Signal leiten wir eine untere Schranke der Rate-Distortion-Funktion her, welche diese für kleine Verzerrungen exakt bestimmt. Zudem werden dezentrale Signale mit gemeinsamer dünnbesetzter Struktur untersucht und die Menge der erreichbaren Raten wird für kleine Verzerrungen genau charakterisiert.

Im Bereich des Compressed Sensing mit Skalarquantisierung werden zwei verschiedene Probleme untersucht. Zunächst werden Algorithmen basierend auf Bayesian Approximate Message Passing angewandt und die Beziehung zwischen Gesamtbitrate der quantisierten Messungen und Verzerrung für ein oder mehrere Signale untersucht. Hier wird bestimmt, wie weit die Gesamtbitrate durch eine verlustfreie Komprimierung der quantisierten Messungen verringert werden kann. Zudem wird dezentralisiertes Compressed Sensing mit Ein-Bit-Quantisierung untersucht und eine Schranke für den maximalen Rekonstruktionsfehler bewiesen. Die Ergebnisse zeigen, dass für dezentrale Messungen von Signalen mit gemeinsamer dünnbesetzter Struktur die Messraten im Vergleich zu klassischem Compressed Sensing deutlich reduziert werden können.

Abstract

This thesis studies digital compression of sparse signals from two points of view: information-theoretic limits and compressed sensing algorithms.

For the information-theoretic limits, a rate-distortion function with letter-based distortion constraints is proposed and investigated in the infinite and finite block length regimes. For a single sparse source, a converse bound is derived that is tight for low distortions. A distributed compression problem for sparse sources is then studied via inner and outer bounds on the rate-distortion region. The region is accurately characterized for low distortions.

For the compressed sensing algorithms, two different problems with scalar quantization are considered. First, Bayesian approximate message passing algorithms are applied to single and multi-terminal settings to study the rate-distortion trade-offs for different quantizer depths. It is shown how much lossless compression of the quantized measurements improves the trade-offs. Second, uniform approximation guarantees are derived for distributed one-bit compressed sensing. The results show that distributed sensing can significantly reduce the required number of measurements for jointly sparse signals.

1

Introduction

Natural signals of interest usually possess *structure*. In fact, it is difficult for a human to process signals that are truly random as such signals are usually perceived as *noise*. Consider, for example, the image of a zebra in Figure 1.1(a).

(a) Original image [vdM].

(b) Reconstruction using only 1% of the discrete cosine transform coefficients.

Figure 1.1.: An image and its reconstruction from a heavily compressed version.

In many applications such as a digital camera, this structure is exploited *after* sensing the signal. A typical compression algorithm (such as JPEG) *transforms* the digital image into a suitable basis using, e.g., a *wavelet* or *discrete cosine* transform. The signal is now *sparse* in this basis which means that most of its coefficients are close to zero and very few carry most of the "energy". The compression algorithm can then discard the *insignificant*

coefficients and only store the locations and values of the *significant* samples. The image in Figure 1.1, for example, reconstructs the original image from only the largest 1% of its *discrete cosine transform* coefficients. This saves a lot of memory as compared to storing the brightness values at every pixel.

However, even this process is wasteful since the camera first acquired a large amount of measurements (pixel brightness values) and then throws away most of them in order to store the information more efficiently. It would be better to only acquire the information that is actually needed. The field of *Compressed Sensing (CS)* has evolved from the insight that efficiently acquiring information is possible for many classes of structured signals [CRT06b, CRT06a, Don06]. Practically, this means that one can build a measurement system that takes only a few measurements but then reconstruct an image with a high resolution - even if the locations of the significant samples are unknown. This insight has led to significant improvements for applications where taking measurements is costly or slow, such as *Magnetic Resonance Imaging* [LDSP08].

This thesis studies the theory of digital compression of sparse signals from different points of view. Consider the very basic system model in Figure 1.2. The signal, such as a sound snippet or an image, is mapped into a finite number of bits by an *encoder* which, in the above example, includes the camera acquisition system as well as the digital compression algorithm. The bits are then stored digitally and can later be used by the *decoder* to create the *signal reconstruction*. Usually, this encoding/decoding process causes the reconstructed signal to be a distorted version of the original signal, and it is natural to expect that allowing the encoder to use more bits should also lower the distortion that occurs after the reconstruction.

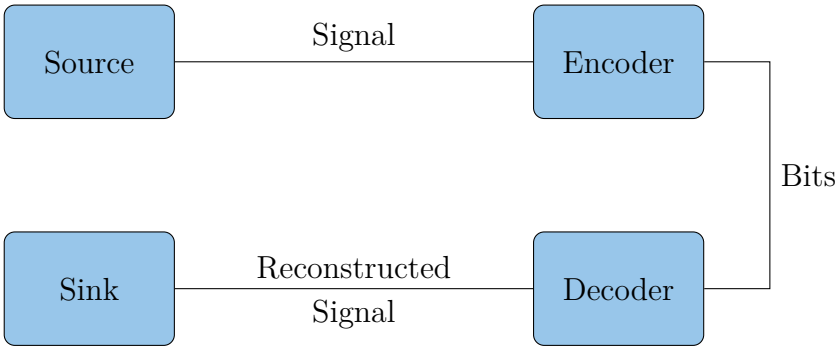


Figure 1.2.: Basic system model for digital source coding.

There are many different questions that can be posed with respect to this basic system model. This thesis focuses on investigating variants of the following questions.

Q1) For a given fidelity criterion, what is the smallest number of bits needed among *any encoder/decoder pair*?

Q2) For a *specific encoder structure* and a given bit rate, what is the smallest error achievable for the *best decoder*?

Q3) Suppose we have the answer to Q2, can one find a decoder with a *tolerable computational complexity* achieving a similar performance as the optimal decoder?

Q4) For a specific encoder/decoder structure, what is the *trade-off* between bit rate and distortion?

Q1 is typically studied in the field of *Information Theory*, a field that was founded by the groundbraking work of Shannon [Sha48]. In [Sha48, Sha59], he formalized Q1 in a probabilistic setting and provided a general solution. For an overview of classical results in information-theoretic source coding, see [Ber71, BG98]. Part I of this thesis focuses on an information-theoretic study of sparse sources.
CS emerged from the observation that a combination of a linear encoder and a decoder based on convex optimization techniques allows to drastically reduce the required sampling rates compared to classical systems [CW08]. The theory of CS thus focuses on (generalized) linear models for the encoder and aims at providing answers to the questions Q2 – Q4. Parts II and III of this thesis investigate different settings of Quantized Compressed Sensing (QCS).

A detailed outline is given below.

1.1. Outline & Contribution

▷ **Chapter 2** provides a brief introduction into the information theory of digital compression, CS, and the relevant literature that this thesis is built upon.

Part I studies aspects of sparse sources from an information-theoretic point of view, adding partial answers to Q1 for single and multi-terminal sparse sources.

▷ **Chapter 3** argues that in certain applications, it is sensible to impose several distortion constraints on a signal and average each distortion function separately over parts of the signal. In this spirit, we study compression for letter-based fidelity measures first in the limit of large signal dimension and then for finite block lengths. As a byproduct, we develop a new converse result for finite length lossy compression. We evaluate our results for the binary memoryless and the Gaussian memoryless sources.

▷ In **Chapter 4**, we study the Rate-Distortion (RD) function of Bernoulli-Spike sources, a popular probabilistic model for sparse signals. We first derive a converse result for sources with a separate distortion constraint for the nonzero elements and the zero elements and then extend this result to the classic case of a single squared error distortion measure. We then show that this converse result is asymptotically tight in the small distortion regime.

▷ **Chapter 5** extends the studies of Chapter 4 to distributed source coding with two terminals, where two correlated Bernoulli-Gaussian spike sources are encoded separately, but reconstructed together. For this purpose, we derive several inner

and outer bounds and determine the achievable rate region in the limit of small distortions at both terminals.

Part II focuses on Bayesian CS and investigates RD trade-offs for QCS systems, touching on Q2-Q4 for the setting of CS with scalar quantization.

- ▷ **Chapter 6** first reviews Generalized Approximate Message Passing (GAMP), a powerful signal reconstruction algorithm, for the setting of CS with scalar quantization. We then numerically study the RD trade-off for different quantizer depths. We further extend this algorithm to distributed QCS with two signals and compare its performance to the results from Chapter 5. Finally, we study the RD trade-off when compressing the quantized measurements.

Part III investigates distributed CS with one-bit quantization. Here, we provide insight into Q2-Q3 by analyzing the worst case error for a distributed QCS system and a low complexity decoder.

- ▷ In **Chapter 7**, we study a setting where many jointly sparse signals are observed at different terminals and reconstructed together from their one-bit measurements via hard thresholding. We provide uniform recovery guarantees for all jointly sparse signals and show the necessary number of measurements can significantly be reduced compared to the setting of just one signal.

Parts of the work presented in Chapters 3 – 4 are published in [PT16a, PT16b, PT16c] and are based on joint work with Roy Timo. A part of the work presented in Chapter 6 is based on joint work with Rami Ezzine and appeared in a student research internship report [Ezz18]. The results presented in Chapter 7 are joint work with Johannes Maly and have been published in [MP19].

1.2. Notation

Below, we give a brief overview on notation. Here and throughout the thesis, we use $:=$ whenever a new quantity is introduced.

Sets

- ▷ We abbreviate $[n] := \{1, ..., n\}$. The set of real and natural numbers is denoted $\mathbb{R}$ and $\mathbb{N}$, respectively. Subsets of the real or natural numbers are usually written with calligraphic font, such as $\mathcal{B} \subset \mathbb{R}$. The cardinality of a set is denoted with $\#$, e.g., $\#([n]) = n$.

Vectors, Matrices and Norms:

- ▷ We denote column vectors by sans serif font and scalars with regular fonts. The length of a vector should be clear from the context. The elements of a vector are indicated with square brackets. Random vectors and scalars are written in uppercase

fonts and realizations in lowercase. Thus, the random vector $\mathsf{Z} \in \mathbb{R}^n$ has a realization z with elements $z[1], \ldots, z[n]$. A matrix $\mathbf{A}$ is denoted with bold font and it should be clear whether it is random or fixed. Id_n represents the n-dimensional identity matrix and $\mathsf{0}$ is the all zeros vector.

- ▷ We write the p-norms of vectors as $\|\cdot\|_p$. Note that the Frobenius norm of matrices $\|\cdot\|_F$ corresponds to the ℓ_2-norm of the vectorization. We use the matrix norm $\|\mathbf{Z}\|_{2,1}$ to represent the sum of the ℓ_2-norms of the columns of $\mathbf{Z}$ and, by abuse of notation, we write $\|\mathsf{z}\|_{2,1} = \|\mathbf{Z}\|_{2,1}$ if $\mathsf{z} = \mathrm{vec}(\mathbf{Z})$ is the vectorized representation of a matrix $\mathbf{Z}$.

Probability and Expectation

- ▷ The probability of an event $\mathcal{A}$ is denoted by $\Pr[\mathcal{A}]$ and the probability of an event $\mathcal{A}_1$ conditioned on an event $\mathcal{A}_2$ is denoted by $\Pr[\mathcal{A}_1 \,|\, \mathcal{A}_2]$. The indicator function of an event $\mathcal{A}$ is written $\mathbb{1}_{\mathcal{A}}$.
- ▷ We will consider random variables that are either discrete, continuous, or mixed discrete-continuous. The probability distribution of a random variable X is usually denoted by P_X and it should be clear from the context whether P_X is discrete, continuous or mixed. The cumulative distribution function of X is denoted by F_X.
- ▷ The expectation of a random variable Z is denoted by $\mathsf{E}[Z]$ and its expectation conditioned on a second random variable Z' is $\mathsf{E}[Z \,|\, Z']$. We sometimes write $\mathsf{E}_Z[XZ]$ to stress that an expectation is with respect to the distribution P_Z. The expectation of a random variable Z is denoted by $\mathsf{Var}[Z]$ and its variance conditioned on an event $\mathcal{A}$ is $\mathsf{Var}[Z|\mathcal{A}]$.
- ▷ The support of a probability distribution P_X is denoted by $\mathrm{supp}(P_X)$. The Gaussian distribution with mean m and covariance matrix $\mathbf{C}$ is denoted by $\mathcal{N}(\mathsf{m}, \mathbf{C})$. δ_0 denotes the probability mass function (PMF) that has probability one at zero.

Information Measures

- ▷ We use classic information measures and their conditional versions in the standard way (see, e.g. [CT06b, PW17]). To introduce the notation, the unconditional quantities are defined below. All logarithms are to the base 2.
- ▷ If X has a PMF P_X on $\mathcal{X}$, we denote its *entropy* by

$$H(X) \coloneqq \sum_{x \in \mathrm{supp}(P_X)} P_X(x) \log \frac{1}{P_X(x)}. \tag{1.1}$$

  For a binary random variable with bias p, we denote its entropy by $H_2(p)$.

- ▷ If X has a probability density function (PDF) P_X on $\mathcal{X}$, we denote its *differential entropy* by

$$h(X) \coloneqq \int_{x \in \mathrm{supp}(P_X)} P_X(x) \log \frac{1}{P_X(x)} \mathrm{d}x. \tag{1.2}$$

▷ For a random variable X with distribution P_X and a second random variable Y induced by $P_{Y|X}$, the *mutual information* between X and Y is given by

$$I(X;Y) \coloneqq \mathsf{E}_{XY}\left[\log \frac{\mathrm{d}P_{Y|X}}{\mathrm{d}P_Y}\right], \tag{1.3}$$

where $\mathrm{d}P_{Y|X}/\mathrm{d}P_Y$ is the Radon-Nikodym derivative of $P_{Y|X}$ with respect to P_X (which is equal to $P_{Y|X}/P_X$ if both X and Y are either discrete or continuous). The *information density* is given by

$$\imath_{X;Y}(x;y) \coloneqq \log \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(x,y) \tag{1.4}$$

for some $(x,y) \in \mathcal{X} \times \mathcal{Y}$.

Complexity:

▷ We use the notation $g(n) = \mathcal{O}(f(n))$ to state that $\lim_{n\to\infty} g(n)/f(n) = c$ for some constant $c \in (0,\infty)$.

▷ The notations $\approx$, $\lesssim$, and $\gtrsim$ are used to denote $=$, $\leq$, and $\geq$ up to multiplicative constants.

2

Preliminaries

2.1. Information-Theoretic Digital Compression

In [Sha48, Sha59], Shannon considered a probabilistic variant of our basic system model, see Figure 2.1. In this model (and throughout this thesis), the signal X is modeled as an n-dimensional vector. Shannon assumed that the signal elements $X[1], X[2], \ldots, X[n]$ are independent and identically distributed (iid) according to some distribution P_X on the set $\mathcal{X}$ and the reconstruction variables $Y[1], Y[2], \ldots, Y[n]$ take values in a (possibly different) set $\mathcal{Y}$.

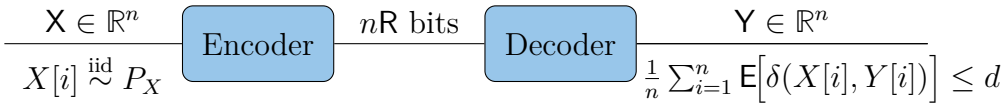


Figure 2.1.: Shannon's system model for digital source coding.

Let $\delta : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a *distortion function*. For vectors, Shannon chose the distortion measure Δ to be

$$\Delta(\mathsf{x}, \mathsf{y}) = \frac{1}{n} \sum_{i=1}^{n} \delta(x[i], y[i]) \tag{2.1}$$

which is *separable*, i.e,

$$\Delta(\mathsf{x}, \mathsf{y}) = \frac{1}{n} \sum_{i=1}^{n} \Delta(x[i], y[i]). \tag{2.2}$$

He then chose the overall fidelity criterion that the system is required to obey to be the average expected distortion. An important observation by Shannon was that every

encoder/decoder pair induces a series of *test channels* $P_{Y[i]|X[i]}$ that determine the expected distortion:

$$\frac{1}{n}\sum_{i=1}^{n}\mathsf{E}[\delta(X[i],Y[i])] = \int\int \delta(x,y)\left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{d}P_{Y[i]|X[i]}\right)\mathrm{d}P_X. \tag{2.3}$$

He then showed that it suffices to find *one good test channel* $P_{Y|X}$ and generate a code book of reconstruction signals by choosing them iid at random according to the induced marginal distribution P_Y. If enough reconstruction signals are drawn at random, the code book will contain good reconstruction signals for a subset of the source sequences that carries most of the probability. This powerful proof technique is called *random coding*.

More precisely, Shannon showed that in the limit of large signal dimensions, i.e., $n \to \infty$, the answer to Q1 is given by the *Rate-Distortion (RD) function*

$$\mathsf{R}(d) = \inf_{P_{Y|X}\in\mathcal{P}(X,d)} I(X;Y) \tag{2.4}$$

where $\mathcal{P}(X,d)$ is the set of $P_{Y|X}$ that satisfy $\mathsf{E}[\delta(X,Y)] \leq d$. The RD function thus provides the smallest number of bits per source symbol needed to compress an iid signal with known distribution subject to an average expected distortion constraint. This result holds true also if the fidelity criterion is the $\Pr[\Delta(\mathsf{X},\mathsf{Y}) > d] \leq \varepsilon$ for a separable Δ and any $\varepsilon > 0$ and has been extended to cases beyond the iid source model [Ber71].

While this result provides the fundamental limit of lossy source coding for this probabilistic source and distortion model, it leaves open the question of how to construct a computationally feasible encoder/decoder pair. Implementing the random coding strategy would require to store all generated reconstruction signals and then finding the closest one for every signal, which has a complexity growing exponentially in the signal dimension.

The RD function is explicitly known only for very few sources and distortion measures such as the Gaussian source with squared error distortion or the binary source with Hamming distortion [Sha59]. A suitable lower bound for sources with a PDF is the *Shannon Lower Bound* [Sha59] which is known to be tight in the low-distortion limit [LZ94, Koc16].

A method of numerically solving the minimization in (2.4) is the *Blahut-Arimoto* algorithm [Bla72], [Ari72], which is given in Algorithm 2.1. This is useful to provide precise numerical computations of the RD function, yet it provides little insight into the general behavior of the RD function or how a good coding scheme might work.

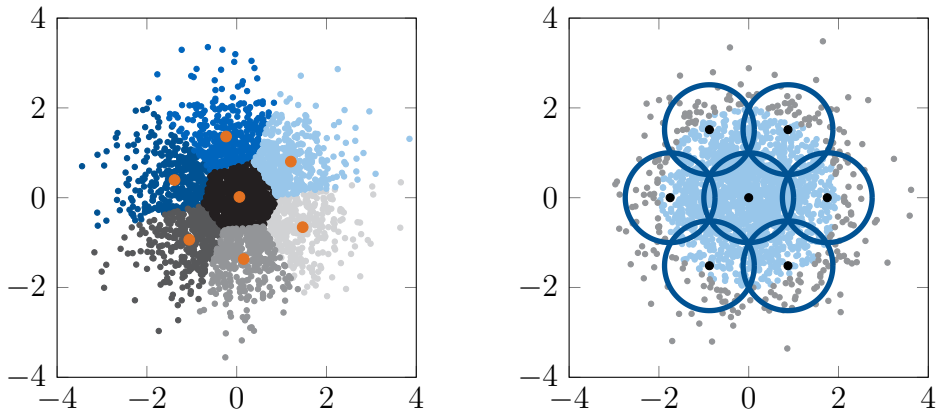
Recently, there has been a growing interest in investigating the performance of lossy source coding with an excess distortion criterion at finite block lengths [KV12, IK11, Kos17, GW19]. While the two distortion criteria

$$\mathsf{E}[\Delta(\mathsf{X},\mathsf{Y})] \leq d \qquad \text{and} \qquad \Pr[\Delta(\mathsf{X},\mathsf{Y}) > d] \leq \varepsilon \tag{2.5}$$

yield the same coding rates for any $\varepsilon > 0$ whenever $n \to \infty$, this is not true at finite block lengths. Figure 2.2 illustrates the geometry of these two different problems in small dimensions.

Algorithm 2.1 Blahut-Arimoto Algorithm for $\mathsf{R}(d)$ [Bla72, Ari72]

Choose $\mathcal{X}_\mathrm{d}$ and $\mathcal{Y}_\mathrm{d}$ as discretizations of $\mathcal{X}$ and $\mathcal{Y}$, compute discretized P_X^d, choose initial reconstruction distribution $P_Y^{(0)}$ on $\mathcal{Y}_\mathrm{d}$, Lagrange multiplier $s > 0$ and a target precision $\varepsilon > 0$.
$t \leftarrow 0$
repeat
  $t \leftarrow t + 1$
  $c^{(t)}(y) \leftarrow \sum_{x \in \mathcal{X}_\mathrm{d}} P_X^\mathrm{d}(x) \frac{\exp(-s\delta(x,y))}{\sum_{y' \in \mathcal{Y}_\mathrm{d}} P_Y^{(t)}(y') \exp(-s\delta(x,y'))}$
  $P_Y^{(t)}(y) \leftarrow P_Y^{(t-1)}(y) c^{(t)}(y)$
  $T_\mathrm{UB}^{(t)} \leftarrow \sum_{y \in \mathcal{Y}_\mathrm{d}} P_Y^{(t)} \log c^{(t)}(y)$
  $T_\mathrm{LB}^{(t)} \leftarrow \max_{y \in \mathcal{Y}_\mathrm{d}} \log c^{(t)}(y)$
until $T_\mathrm{UB}^{(t)} - T_\mathrm{LB}^{(t)} < \varepsilon$
$P_{Y|X}(y|x) \leftarrow \frac{P_Y^{(t)}(y) \exp(-s\delta(x,y))}{\sum_{y' \in \mathcal{Y}_\mathrm{d}} P_Y^{(t)}(y') \exp(-s\delta(x,y'))}$
  $d \leftarrow \sum_{x \in \mathcal{X}_\mathrm{d}} \sum_{y \in \mathcal{Y}_\mathrm{d}} P_X^\mathrm{d}(x) P_{Y|X}(y|x) \delta(x,y)$
  $\mathsf{R}(d) \leftarrow \sum_{x \in \mathcal{X}_\mathrm{d}} \sum_{y \in \mathcal{Y}_\mathrm{d}} P_X^\mathrm{d}(x) P_{Y|X}(y|x) \log \frac{P_{Y|X}(y|x)}{P_Y^{(t)}(y)}$
return $P_{Y|X}$, d, $\mathsf{R}(d)$



(a) Average distortion setting. (b) Excess distortion setting.

Figure 2.2.: Geometry of vector quantization in two dimensions. On the left, $\mathcal{X}$ is partitioned into cells of different size. Individual cells may be large since a greater distortion can be outweighed by a small probability. On the right, we focus on a set that carries a total probability of $1 - \varepsilon$ (light blue). In this set, all points are within a distance of d of the reconstruction points (black). The distortion of a signal point and its reconstruction is irrelevant as long as it is below d.

In the excess distortion setting, Kostina and Verdú [KV12] derived various bounds for the optimal performance of an excess distortion code at a fixed block length and excess distortion probability. In addition (see also [IK11]), they showed that for a fixed block length, the gap to the RD function is on the order of $\mathcal{O}(n^{-1/2})$.

2.2. Compressed Sensing

The field of CS emerged from the following basic question: Given some unknown and high-dimensional signal $\mathsf{x} \in \mathbb{R}^n$, what is the smallest number m of linear measurements

$$\mathsf{z} = \mathbf{A}\mathsf{x} \tag{2.6}$$

needed to uniquely determine x, where $\mathbf{A} \in \mathbb{R}^{m \times n}$. From basic linear algebra we require $m \geq n$ in general.

Motivated by the discussion in Chapter 1, suppose that x is s-sparse, i.e., x has at most $s \leq n$ entries that are not equal to zero, and let $\operatorname{supp}(\mathsf{x}) = \{i \in [n] \colon x[i] \neq 0\} \subset [n]$ be the support of x. Knowing $\operatorname{supp}(\mathsf{x})$, only $m \geq s$ measurements suffice to uniquely identify x from z (assuming that every choice of s different columns of $\mathbf{A}$ is linearly independent). If $s \ll n$, this yields a considerable improvement in the number of measurements. In practice, however, the support of x is unknown.

The seminal works [CRT06b, CRT06a, Don06] lay foundations of CS by showing that if the matrix $\mathbf{A}$ is suitably chosen, then one can uniquely identify *all* s-sparse x from a number of linear measurements that is much smaller than the signal dimension n. Moreover, efficient recovery is possible by means of convex optimization. A sufficient condition for $\mathbf{A}$ to allow this is the so-called *Restricted Isometry Property (RIP)*. A linear operator $\mathbf{A}$ satisfies the Restricted Isometry Property (RIP) of order s with RIP-constant $\delta \in (0, 1)$ if

$$(1 - \delta)\|\mathsf{x}\|_2 \leq \|\mathbf{A}\mathsf{x}\|_2 \leq (1 + \delta)\|\mathsf{x}\|_2 \tag{2.7}$$

for all s-sparse x, i.e., $\mathbf{A}$ embeds the set of s-sparse n-dimensional vectors almost isometrically into $\mathbb{R}^m$ (see [CRT06b], [CT06a]). Though no deterministic construction of RIP matrices has been discovered so far for less than $m = \mathcal{O}(s^2)$ measurements (cf. [FR13, Ch 6]), several classes of randomly generated matrices satisfy the RIP with exceedingly high probability if

$$m \geq Cs \log\left(\frac{en}{s}\right) \tag{2.8}$$

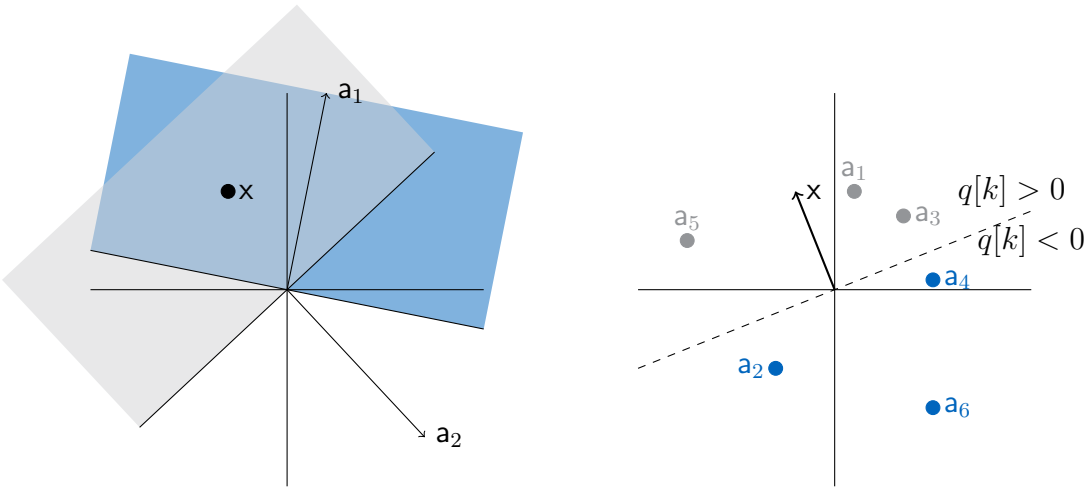
where $C > 0$ is a constant independent of s, m, and n (see [CT06a], [RV08]). Hence, up to the log-factor, $\mathcal{O}(s)$ measurements suffice to capture all information in the n-dimensional signal x. Moreover, the bound (2.8) is robust to noise on the measurements.

However, there are more difficulties to overcome than just noise. In particular, real-valued measurements $z[i] \in \mathbb{R}$ cannot not be stored with infinite precision on a digital

system. The idealistic measurement model presented in (2.6) should be extended by a quantizer Q that maps the real-valued measurement vector $\mathbf{A}\mathsf{x}$ to a finite alphabet. To stick to the common paradigm of low-complexity measurements in CS, we consider *scalar quantizers* that quantize the measurements individually. The extreme case is to choose Q as the sign function acting componentwise on $\mathbf{A}\mathsf{x}$ leading to the *one-bit CS* model first studied in [BB08]

$$\mathsf{q} = \operatorname{sign}(\mathbf{A}\mathsf{x}) \tag{2.9}$$

i.e., $q[k]$ is 1 if $\langle \mathsf{a}_k, \mathsf{x} \rangle \geq 0$ and -1 if $\langle \mathsf{a}_k, \mathsf{x} \rangle < 0$, where a_k is the transposed k-th row of $\mathbf{A}$. One-bit sensing is of great interest for applications because single bit measurement devices are cheap to produce and use. From a geometric point of view, this single bit expresses on which side of the hyperplane H_{a_k} (defined by the normal vector a_k) the signal x lies. A different interpretation is that x is a linear classifier in the signal space that clusters the measurements into two categories. These two interpretations are shown in Figure 2.3.



(a) Half spaces. (b) Classifier.

Figure 2.3.: Geometry of one-bit CS.

Note that the operation (2.9) is blind to scaling and we can only hope to approximate the direction of x (this issue can be tackled by, e.g., adding a random *dither* to the measurements before quantization and thus shifting the hyperplanes away from the origin, see [KSW16, BFN+17, DM18]).

It turns out that for one-bit quantization, a bound similar to (2.8) defines a sufficient number of measurements to *approximate* all s-sparse x of unit norm:

$$m \geq C\delta^{-\alpha} s \log\left(\frac{en}{s}\right). \tag{2.10}$$

In this case, approximating means that one cannot recover x from z exactly but one can bound the *worst case error* (in, e.g., ℓ_2 norm) of certain reconstruction algorithms. The

difference between the required measurements for (2.6) and (2.9) lies in the approximation quality captured by $\delta^{-\alpha}$: the expected worst-case error δ is much better (possibly zero if there is no noise) with unquantized CS. For practical purposes, it is, of course, desirable to have $\alpha > 0$ as small as possible in order to achieve a small m for a fixed accuracy $0 < \delta < 1$.

One suitable reconstruction method for one-bit measurements (2.9) is the linear program [PV13a]

$$\min_{\mathsf{y}\in\mathbb{R}^n} \|\mathsf{y}\|_1 \quad \text{subject to} \quad \operatorname{sign}(\mathbf{A}\mathsf{y}) = \mathsf{q} \quad \text{and} \quad \|\mathbf{A}\mathsf{y}\|_1 = m \tag{2.11}$$

and another method is a hard thresholding procedure [Fou16]

$$\mathsf{y} = \mathbb{H}_s\left(\mathbf{A}^{\mathsf{T}}\mathsf{q}\right) \tag{2.12}$$

where $\mathbb{H}_s$ only keeps the s largest coefficients. More elaborate iterative hard thresholding methods are analyzed in, e.g., [JDV13].

It is worth to stress that these models and results are different from classic information-theoretic approaches to source coding in several ways:

- ▷ No prior knowledge about the stochastic distribution of x is assumed. The only assumption is that x lies in some predefined signal set, such as all s-sparse signals.

- ▷ Once the measurement matrix $\mathbf{A}$ is fixed, we would ideally like to recover or approximate *all signals* x. In contrast to that, information theoretic-studies often require reconstruction with *high probability* or with *small average error*.

- ▷ A specific decoder structure is investigated in order to ensure a computationally tractable reconstruction procedure.

2.3. Probabilistic Compressed Sensing

In probabilistic (or Bayesian) CS, the setting is more similar to the information theoretic approach. In particular, the signal is modeled as stochastic and its distribution may be known to the decoder. Further, the fidelity criterion is often taken as the expected error (e.g., Mean Squared Error (MSE)).

The fundamental limits of analog probabilistic CS have been investigated information-theoretically by Wu and Verdú [WV12b]. For linear encoders such as (2.6), they considered continuous and Lipschitz-continuous decoders to ensure a certain robustness to measurement noise. Further, they allowed the decoder to use the probability distribution P_{X} of the stochastic signal X and demanded to recover the signal with probability $1 - \varepsilon$ for an arbitrary $\varepsilon > 0$. They found that the fundamental limit of the number of samples m is

the *information dimension* which, for a real-valued random variable X, is given by

$$d(X) = \lim_{k\to\infty} \frac{H(\lfloor kX \rfloor)}{\log k}. \tag{2.13}$$

As an example, if the elements of X are iid with distribution P_X and

$$P_X = (1-p)\delta_0 + pP_{\text{cont}} \tag{2.14}$$

where δ_0 is the point mass at zero and P_{cont} is a continuous probability distribution, then

$$m > s \tag{2.15}$$

measurements suffice in order to reconstruct the signal X with high probability from m linear measurements. These results also extend to the case of noisy measurements.

A powerful reconstruction algorithm that can exploit the prior distribution of the signal is *Approximate Message Passing (AMP)* [DMM09]. AMP comes with many benefits such as much smaller computational complexity than convex optimization methods and tools to asymptotically analyze its performance [JM13]. While AMP does not achieve the information theoretic limit (2.15) for the popular choice of dense iid Gaussian matrices [KMS$^+$12a, KMS$^+$12b], AMP combined with *spatially coupled* sensing matrices closely approaches this fundamental limit [KMS$^+$12a, DJM13, BSK15].

An important extension of AMP applies to *Generalized Linear Models (GLMs)* [Ran11, SRF16] of the form:

$$Z_k = P_{\text{out}}\Big(\frac{1}{\sqrt{n}}\langle \mathsf{A}_k, \mathsf{X}\rangle\Big), \qquad 1 \le k \le m \tag{2.16}$$

where A_k is the transposed kth row of $\mathbf{A}$ and P_{out} represents an *output channel* that may include noise, but can also represent a deterministic function. Taking P_{out} as a quantization function, the GLM (2.16) includes the setting of QCS, which was investigated from this point of view in [KGR12]. Recently, Barbier et al. [BKM$^+$19] derived the fundamental limits of information and estimation in GLMs, which we will heavily make use of in Chapter 6.

Part I.

Rate-Distortion Theory for Multiple Constraints and Sparse Sources

3

Compression for Letter-Based Fidelity Measures

Typically, RD theory has focused on the compression of sources with respect to a single average distortion or excess distortion constraint. In applications such as CS, however, the important information is manifested in a few significant samples of the signal while most samples are either zero or close to zero and can be discarded [WV12a]. In this case it seems reasonable to impose a distortion constraint that is averaged over only those significant samples and another constraint that is averaged over the insignificant samples. Motivated by this observation, we define the letter-based RD function in Section 3.1 and investigate it in the limit of larger block lengths. In Section 3.1, we review existing finite block length bounds and extend those to the setting of our new multiple distortion measures. Section 3.3 then applies the results from the previous sections to the case of a binary memoryless source.

This chapter is based on joint work with Roy Timo. Part of the results presented in this chapter have been published in [PT16a, PT16b].

3.1. Infinite Block Length

Not all data is created equal. In network security, a packet's header is often more important than its body; in image compression, a wavelet transform concentrates useful information in a fraction of its coefficients; and in fraud detection, abnormal credit card transactions made overseas are more important than ones made at your local shops. Thus, there is often a need to identify and separately process particular data events that are 'more important' for the end application. This chapter is motivated by such situations, and, to this end, we consider the problem of lossy compression with multiple letter-based distortion constraints.

Consider a *memoryless source* that creates the output sequence X with each $X[i]$, $1 \leq i \leq n$, being distributed according to P_X on some alphabet $\mathcal{X}$. A *lossy source code* consists

of a pair

$$f : \mathcal{X}^n \to \{1, 2, \dots M\} \qquad \text{and} \qquad g : \{1, 2, \dots M\} \to \mathcal{Y}^n \tag{3.1}$$

where the *encoder* f maps the source sequence to an index $T \coloneqq f(\mathsf{X})$ and the *decoder* g maps this index to a reconstruction sequence $\mathsf{Y} \coloneqq g(T)$. Let $\delta : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ be a *distortion function*, and let $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_L\}$ with $L < \infty$ be a partition of $\mathcal{X}$ such that $\Pr[X \in \mathcal{I}_\ell] > 0$ for all $\ell \in [L]$. For every $\mathcal{I}_\ell \in \mathcal{I}$, we define two different n-letter distortion measures:

(i) Normalization based on actual occurrences:

$$\Delta_\ell^{\mathrm{a}}(\mathsf{x}, \mathsf{y}) = \begin{cases} \frac{1}{\mathsf{N}(\mathcal{I}_\ell|\mathsf{x})} \sum_{i=1}^{n} \mathbb{1}_{\{x[i] \in \mathcal{I}_\ell\}} \delta(x[i], y[i]), & \text{if } \mathsf{N}(\mathcal{I}_\ell|\mathsf{x}) \geq 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

where $\mathsf{N}(\mathcal{I}_\ell|\mathsf{x}) \coloneqq \sum_{i=1}^{n} \mathbb{1}_{\{x[i] \in \mathcal{I}_\ell\}}$. Normalizing the distortion this way averages the distortion for each ℓ properly, irrespective of how many source symbols fall into $\mathcal{I}_\ell$. This *fair normalization* comes at the cost of an unseparable distortion measure, that is, we have $\Delta_\ell^{\mathrm{a}}(\mathsf{x}, \mathsf{y}) \neq \frac{1}{n} \sum_{i=1}^{n} \Delta_\ell^{\mathrm{a}}(x[i], y[i])$.

(ii) Normalization based on expected occurrences:

$$\Delta_\ell^{\mathrm{e}}(\mathsf{x}, \mathsf{y}) = \frac{1}{n \Pr[X \in \mathcal{I}_\ell]} \sum_{i=1}^{n} \mathbb{1}_{\{x[i] \in \mathcal{I}_\ell\}} \delta(x[i], y[i]). \tag{3.3}$$

This distortion measure is separable. The fixed normalization, however, means that statistical variations of the number of symbols observed in every $\mathcal{I}_\ell$ change the allowed average distortion for those symbols.

Given some $\varepsilon > 0$ and a vector $\mathsf{d} = (d_1, \dots, d_L)$ of distortion constraints with respective distortion measures Δ^{a} or Δ^{e}, we define a lossy source code with respect to an excess distortion or average distortion constraint as follows.

Definition 3.1 (Excess-Distortion Code)**.** An $(n, M, \mathsf{d}, \varepsilon, \Delta)$ code for a memoryless source with distribution P_X outputting $\mathsf{X} \in \mathcal{X}^n$ consists of an encoder $f : \mathcal{X}^n \to \{1, \dots, M\}$ and a decoder $g : \{1, \dots, M\} \to \mathcal{Y}^n$ satisfying

$$\Pr\Big[\bigcup_{\ell \in [L]} \Big\{ \Delta_\ell\big(\mathsf{X}, g(f(\mathsf{X}))\big) > d_\ell \Big\} \Big] \leq \varepsilon. \tag{3.4}$$

Accordingly, we define the smallest codebook size for a set of parameters as

(i) $M_{\mathrm{a}}^\star(n, \mathsf{d}, \varepsilon) = \min\{M : \exists\, (n, \mathsf{d}, M, \varepsilon, \Delta^{\mathrm{a}}) \text{ code}\}$,

(ii) $M_{\mathrm{e}}^\star(n, \mathsf{d}, \varepsilon) = \min\{M : \exists\, (n, \mathsf{d}, M, \varepsilon, \Delta^{\mathrm{e}}) \text{ code}\}$.

Definition 3.2 (Expected-Distortion Code)**.** An $(n, M, \mathsf{d}, \Delta)$ code for a memoryless source with distribution P_X outputting $\mathsf{X} \in \mathcal{X}^n$ consists of an encoder $f : \mathcal{X}^n \to \{1, \dots, M\}$ and a decoder $g : \{1, \dots, M\} \to \mathcal{Y}^n$ satisfying

$$\mathsf{E}\Big[\Delta_\ell\Big(\mathsf{X}, g(f(\mathsf{X}))\Big)\Big] \le d_\ell \tag{3.5}$$

for all $\ell \in [L]$. Accordingly, we define the smallest codebook size for a set of parameters as

(i) $\overline{M}^\star_{\mathrm{a}}(n, \mathsf{d}) = \min\{M : \exists\, (n, \mathsf{d}, M, \Delta^{\mathrm{a}})\text{ code}\}$

(ii) $\overline{M}^\star_{\mathrm{e}}(n, \mathsf{d}) = \min\{M : \exists\, (n, \mathsf{d}, M, \Delta^{\mathrm{e}})\text{ code}\}$.

Definition 3.3 (Letter-Based RD Function)**.** We define the *Letter-Based Rate-Distortion (RDL) function* as

$$\mathsf{R}_{\mathsf{L}}(\mathsf{d}) = \inf_{P_{Y|X} \in \mathcal{P}(X, \mathsf{d})} I(X;Y)\,, \tag{3.6}$$

where $\mathcal{P}(X, \mathsf{d})$ is the set of all conditional probability distributions from $\mathcal{X}$ to $\mathcal{Y}$ that satisfy

$$\mathsf{E}[\delta(X,Y)\,|\,X \in \mathcal{I}_\ell] \le d_\ell \quad \text{for} \quad 1 \le \ell \le L. \tag{3.7}$$

The RDL function inherits several useful properties from the usual RD function:

- $0 \le \mathsf{R}_{\mathsf{L}}(\mathsf{d}) \le H(X)$.
- R_{L} is non-increasing.
- R_{L} is convex in d.
- R_{L} is continuous on $[0, \infty)^\ell$.

Further, we can recover the usual RD function from the RDL function via

$$\mathsf{R}(d) = \min_{\mathsf{d}:\sum_{\ell=1}^{L} \Pr[X \in \mathcal{I}_\ell]\, d_\ell \le d} \mathsf{R}_{\mathsf{L}}(\mathsf{d}). \tag{3.8}$$

Before presenting the main theorem of this section, we make the assumption that our distortion measure does not grow too fast (see [PW17, Ch. 25.3]).

Assumption 3.1. There is some $p > 1$ such that $d_p < \infty$, where

$$d_p \coloneqq \sup_{n \ge 1} \inf_{\mathsf{y} \in \mathcal{Y}^n} \max_{\ell \in [L]} \mathsf{E}[\Delta^{\mathrm{a}}_\ell(\mathsf{X}, \mathsf{y})^p]^{1/p}\,. \tag{3.9}$$

This assumption is slightly stronger than requiring a finite average distortion at zero rate ($p = 1$), but it is not too restrictive. It will help us to control the distortions via Hölder's inequality.

Next, we state the main theorem of this section, which characterizes the smallest achievable coding rate in the limit of large block lengths.

Theorem 3.1 (Asymptotic RDL Coding Rate)**.** Let $\mathsf{X} \overset{\text{iid}}{\sim} P_X$, $L < \infty$ and suppose that there is some $p > 1$ such that Assumption 3.1 is satisfied. Then, we have

$$\mathsf{R}_{\mathsf{L}}(\mathsf{d}) = \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log M_{\mathrm{e}}^{\star}(n, \mathsf{d}, \varepsilon) = \limsup_{n \to \infty} \frac{1}{n} \log \overline{M}_{\mathrm{e}}^{\star}(n, \mathsf{d}) \tag{3.10}$$

$$= \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log M_{\mathrm{a}}^{\star}(n, \mathsf{d}, \varepsilon) = \limsup_{n \to \infty} \frac{1}{n} \log \overline{M}_{\mathrm{a}}^{\star}(n, \mathsf{d}) \tag{3.11}$$

We remark that Pinkston briefly discussed the extension of classical RD theory to multiple separable distortion constraints in [Pin67, Sec. 2.6]. By rescaling the distortion function Δ^{e}, equality in (3.10) can be deduced. However, the same does not apply to (3.11) since the function Δ^{a} is not a separable distortion measure. We present a proof of Theorem 3.1 that is based on the proof of the standard RD theorem in [PW17, Chap. 25-26], with some additional effort to control the distortions.

3.1.1. Proof of the Coding Theorem (Theorem 3.1)

We prove Theorem 3.1 using a sequence of three lemmas. The first lemma shows that for an expected distortion constraint, the rate is lower bounded by R_{L}.

Lemma 3.2 (Converse for Average Distortion)**.** Suppose that Assumption 3.1 is satisfied. Then, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log \overline{M}_{\mathrm{a}}^{\star}(n, \mathsf{d}) \geq \mathsf{R}_{\mathsf{L}}(\mathsf{d}). \tag{3.12}$$

Next, we show that we can build a good code for an expected distortion constraint from a good code for an excess distortion constraint.

Lemma 3.3 (Excess to Average Distortion)**.** Suppose that Assumption 3.1 is satisfied. Then, we have

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log M_{\mathrm{a}}^{\star}(n, \mathsf{d}, \varepsilon) \geq \limsup_{n \to \infty} \frac{1}{n} \log \overline{M}_{\mathrm{a}}^{\star}(n, \mathsf{d}). \tag{3.13}$$

Finally, we show achievability.

Lemma 3.4 (Achievability for Excess Distortion)**.**

$$\mathsf{R}_{\mathsf{L}}(\mathsf{d}) \geq \lim_{\varepsilon \downarrow 0} \limsup_{n \to \infty} \frac{1}{n} \log M_{\mathrm{a}}^{\star}(n, \mathsf{d}, \varepsilon). \tag{3.14}$$

Proof of Theorem 3.1. Lemma 3.2 – 3.4 together imply that

$$\lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}\frac{1}{n}\log M^{\star}_{\mathrm{a}}(n,\mathsf{d},\varepsilon)=\limsup_{n\to\infty}\frac{1}{n}\log \overline{M}^{\star}_{\mathrm{a}}(n,\mathsf{d})=\mathsf{R}_{\mathsf{L}}(\mathsf{d}) \tag{3.15}$$

which establishes Theorem 3.1. ■

The remainder of this section is dedicated to proving Lemmas 3.2, 3.3 and 3.4.

Proof of Lemma 3.2: Converse for Expected Distortion

For the memoryless source P_X, suppose we are given a sequence of optimal RDL-codes. Observe that the encoder, decoder and the number of code words depend on n, but for simplicity we refer to (f,g) and the code book size M without an index n. For every n, (f,g) satisfies

$$\mathsf{E}[\Delta^{\mathrm{a}}_{\ell}(\mathsf{X},g(f(\mathsf{X}))]\leq d_{\ell},\quad \text{for } \ell\in[L]. \tag{3.16}$$

Let $\mathsf{Y}=g(f(\mathsf{X}))$. We start with the usual converse steps

$$\frac{1}{n}\log M\geq\frac{1}{n}H(f(\mathsf{X}))=\frac{1}{n}I(\mathsf{X};f(\mathsf{X}))\geq\frac{1}{n}I(\mathsf{X};\mathsf{Y})\geq\frac{1}{n}\sum_{i=1}^{n}I(X[i];Y[i])\,. \tag{3.17}$$

Now, for every $i\in[n]$ and $\ell\in[L]$, we define the average conditional distortion achieved by the coding scheme as

$$d_{\ell}[i]=\mathsf{E}[\delta(X[i],Y[i])\,|\,X[i]\in\mathcal{I}_{\ell}] \tag{3.18}$$

and recall that $\mathcal{P}(X,\mathsf{d})$ is the set of $P_{Y|X}$ that satisfy the distortion constraints

$$\mathsf{E}[\delta(X,Y)\,|\,X\in\mathcal{I}_{\ell}]\leq d_{\ell}\quad\text{for}\quad\ell\in[L]. \tag{3.19}$$

We proceed from (3.17) with

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}I(X[i];Y[i])&\geq\frac{1}{n}\sum_{i=1}^{n}\inf_{P_{Y|X}\in\mathcal{P}(X,(d_1[i],\ldots,d_L[i]))}I(X[i];Y[i])\\&=\frac{1}{n}\sum_{i=1}^{n}\mathsf{R}_{\mathsf{L}}\Big(d_1[i],\ldots,d_L[i]\Big)\\&\geq\mathsf{R}_{\mathsf{L}}\Big(\frac{1}{n}\sum_{i=1}^{n}d_1[i],\ldots,\frac{1}{n}\sum_{i=1}^{n}d_L[i]\Big),\end{aligned} \tag{3.20}$$

where the last inequality follows since R_{L} is convex due to the convexity of mutual information [CT06b, Thm. 2.7.4]. To complete the converse, we show that $\mathsf{E}[\Delta^{\mathrm{a}}_{\ell}(\mathsf{X},\mathsf{Y})]$ concentrates around $\frac{1}{n}\sum_{i=1}^{n}d_{\ell}[i]$ (which is trivial for a separable distortion measure). De-

fine the ε-letter typical set

$$\mathcal{T}_\varepsilon \coloneqq \left\{ \mathsf{x} \in \mathcal{X}^n : \left| \frac{1}{n} \mathsf{N}(\mathcal{I}_\ell | \mathsf{x}) - \Pr[X \in \mathcal{I}_\ell] \right| \leq \varepsilon \Pr[X \in \mathcal{I}_\ell], \text{ for } \ell \in [L] \right\} \tag{3.21}$$

and note that Hoeffding's inequality [Ver18, Thm. 2.2.6] and the union bound imply

$$\Pr[\mathsf{X} \notin \mathcal{T}_\varepsilon] \leq 2 \cdot L \cdot e^{-2n(\varepsilon \min_\ell \Pr[X \in \mathcal{I}_\ell])^2} \overset{n\to\infty}{\longrightarrow} 0. \tag{3.22}$$

To proceed, we use the following lemma which establishes that the distortion measured by Δ^{e} for the atypical sequences of a good code is small.

Lemma 3.5. Under Assumption 3.1, every sequence of optimal encoder/decoder pairs satisfies

$$\lim_{n\to\infty} \max_{\ell \in [L]} \mathsf{E}\Big[\mathbb{1}_{\{\mathsf{x} \notin \mathcal{T}_\varepsilon\}} \Delta^{\mathrm{e}}_\ell(\mathsf{X}, \mathsf{Y}) \Big] = 0. \tag{3.23}$$

Proof. Suppose that (3.23) does not hold for some $\ell \in [L]$, i.e., the right hand side (RHS) is larger than zero. We construct a simple modification to the coding scheme that uses the same rate and achieves a smaller error and thus a better RD trade-off. We define the new RD code (f^*, g^*) via:

$$f^*(\mathsf{x}) = \begin{cases} f(\mathsf{x}), & \mathsf{x} \in \mathcal{T}_\varepsilon \\ M+1, & \mathsf{x} \notin \mathcal{T}_\varepsilon \end{cases} \qquad \text{and} \qquad g^*(t) = \begin{cases} g(t), & 1 \leq t \leq M \\ \mathsf{y}_0, & t = M+1 \end{cases} \tag{3.24}$$

where $\mathsf{y}_0 = (y_0, \ldots, y_0)$ satisfies

$$\mathsf{E}\Big[\Delta^{\mathrm{a}}_\ell(\mathsf{X}, \mathsf{y}_0)^p \Big]^{1/p} \leq d_p < \infty \tag{3.25}$$

for all $\ell \in [L]$ as guaranteed by Assumption 3.1. After this modification, the asymptotic rate $\lim_{n\to\infty} \frac{1}{n} \log(M+1) = \lim_{n\to\infty} \frac{1}{n} \log M$ does not change. For $\ell \in [L]$, the expected distortion for the atypical sequences of new our code is

$$\begin{aligned} \mathsf{E}\Big[\mathbb{1}_{\{\mathsf{x} \notin \mathcal{T}_\varepsilon\}} \Delta^{\mathrm{e}}_\ell(\mathsf{X}, \mathsf{Y}) \Big] &= \mathsf{E}\Big[\mathbb{1}_{\{\mathsf{x} \notin \mathcal{T}_\varepsilon\}} \Delta^{\mathrm{e}}_\ell(\mathsf{X}, \mathsf{y}_0) \Big] \\ &\overset{\mathrm{a}}{\leq} \frac{1}{\Pr[X \in \mathcal{I}_\ell]} \mathsf{E}\Big[\mathbb{1}_{\{\mathsf{x} \notin \mathcal{T}_\varepsilon\}} \Delta^{\mathrm{a}}_\ell(\mathsf{X}, \mathsf{y}_0) \Big] \\ &\overset{\mathrm{b}}{\leq} \frac{1}{\Pr[X \in \mathcal{I}_\ell]} \mathsf{E}\Big[\Delta^{\mathrm{a}}_\ell(\mathsf{X}, \mathsf{y}_0)^p \Big]^{1/p} \Pr[\mathsf{X} \notin \mathcal{T}_\varepsilon]^{1-1/p} \\ &\overset{\mathrm{c}}{\leq} \frac{d_p}{\Pr[X \in \mathcal{I}_\ell]} \cdot \Pr[\mathsf{X} \notin \mathcal{T}_\varepsilon]^{1-1/p} \\ &\overset{n\to\infty}{\longrightarrow} 0 \end{aligned} \tag{3.26}$$

where we used the fact that $\Delta^{\mathrm{e}}_\ell(\mathsf{x}, \mathsf{y}) \leq \Delta^{\mathrm{a}}_\ell(\mathsf{x}, \mathsf{y}) / \Pr[X \in \mathcal{I}_\ell]$ in (a), Hölder's inequality in

(b), and (3.25) in (c). Thus, if (3.23) does not hold, we can construct a coding scheme with the same asymptotic rate and a lower average distortion. We conclude that (3.23) must hold for an optimal sequence of codes. ■

Now, we can bound

$$\begin{aligned}
&\mathsf{E}[\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})] \\
&\quad= \mathsf{E}\left[\frac{1}{\mathsf{N}(\mathcal{I}_\ell|\mathsf{X})}\sum_{i=1}^{n}\mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])\right] \\
&\quad\geq \mathsf{E}\left[\mathbb{1}_{\{\mathsf{x}\in\mathcal{T}_\varepsilon\}}\frac{1}{\mathsf{N}(\mathcal{I}_\ell|\mathsf{X})}\sum_{i=1}^{n}\mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])\right] \\
&\quad\geq \frac{1}{n\Pr[X\in\mathcal{I}_\ell](1+\varepsilon)}\mathsf{E}\left[\mathbb{1}_{\{\mathsf{x}\in\mathcal{T}_\varepsilon\}}\sum_{i=1}^{n}\mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])\right] \\
&\quad= \frac{1}{n\Pr[X\in\mathcal{I}_\ell](1+\varepsilon)}\Bigg(\sum_{i=1}^{n}\mathsf{E}\Big[\mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])\Big] \\
&\qquad\qquad - \sum_{i=1}^{n}\mathsf{E}\Big[\mathbb{1}_{\{\mathsf{x}\notin\mathcal{T}_\varepsilon\}}\mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])\Big]\Bigg) \\
&\quad= \frac{1}{n(1+\varepsilon)}\sum_{i=1}^{n}d_\ell[i] - \frac{1}{1+\varepsilon}\mathsf{E}\Big[\mathbb{1}_{\{\mathsf{x}\notin\mathcal{T}_\varepsilon\}}\Delta_\ell^{\mathrm{e}}(\mathsf{X},\mathsf{Y})\Big] \\
&\quad\overset{\mathrm{a}}{\geq} \frac{1}{n(1+\varepsilon)}\sum_{i=1}^{n}d_\ell[i] - \varepsilon'
\end{aligned} \tag{3.27}$$

where we used Lemma 3.5 for an arbitrary $\varepsilon' > 0$ and large enough n in (a). Combining (3.27) and (3.16), we see that

$$\frac{1}{n}\sum_{i=1}^{n}d_\ell[i] \leq (1+\varepsilon)\Big(\mathsf{E}[\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})]+\varepsilon'\Big) = (1+\varepsilon)(d_\ell+\varepsilon') \tag{3.28}$$

for all $\ell \in [L]$. Since R_L is non-increasing and continuous, we can let $\varepsilon, \varepsilon' \searrow 0$ as $n \to \infty$ to conclude that

$$\mathsf{R}_\mathsf{L}\Big(\frac{1}{n}\sum_{i=1}^{n}d_1[i],\ldots,\frac{1}{n}\sum_{i=1}^{n}d_L[i]\Big) \geq \mathsf{R}_\mathsf{L}\Big((1+\varepsilon)(\mathsf{d}+\varepsilon')\Big) \overset{n\to\infty}{\longrightarrow} \mathsf{R}_\mathsf{L}(\mathsf{d}), \tag{3.29}$$

which completes the converse proof. ■

Excess to Average Distortion

Lemma 3.3 is a consequence of the following lemma which is adapted from [PW17, Thm 25.5].

Lemma 3.6. Fix $\varepsilon > 0$ and a distortion tuple d. Suppose that Assumption 3.1 is valid and that we have an RDL code (f, g) of size M satisfying

$$\Pr\left[\bigcup_{\ell\in[L]} \{\Delta_\ell^{\mathrm{a}}(\mathsf{X}, g(f(\mathsf{X}))) > d_\ell\}\right] \le \varepsilon. \tag{3.30}$$

Then we can find a new RDL code $(\hat{f}, \hat{g})$ of size $M+1$ whose average distortion satisfies

$$\mathsf{E}\big[\Delta_\ell^{\mathrm{a}}\left(\mathsf{X}, \hat{g}(\hat{f}(\mathsf{X}))\right)\big] \le d_\ell + d_p \cdot \varepsilon^{1-1/p} \tag{3.31}$$

for all $\ell \in [L]$.

Proof. We define the new RDL code $(\hat{f}, \hat{g})$ via:

$$\hat{f}(\mathsf{x}) = \begin{cases} f(\mathsf{x}), & \Delta_\ell^{\mathrm{a}}(\mathsf{x}, g(f(\mathsf{x}))) \le d_\ell \text{ for all } \ell \in [L] \\ M+1, & \text{otherwise} \end{cases} \tag{3.32}$$

$$g^*(t) = \begin{cases} g(t), & 1 \le t \le M \\ \mathsf{y}_0, & t = M+1. \end{cases} \tag{3.33}$$

Let $\mathsf{Y} = g(f(\mathsf{X}))$ and $\hat{\mathsf{Y}} = \hat{g}(\hat{f}(\mathsf{X}))$. The distortion can be bounded for all $\ell \in [L]$:

$$\begin{aligned} \mathsf{E}\big[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \hat{\mathsf{Y}})\big] &= \mathsf{E}\big[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \hat{\mathsf{Y}})\big(\mathbb{1}_{\{\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})\le d_\ell\}} + \mathbb{1}_{\{\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})> d_\ell\}}\big)\big] \\ &= \Pr\big[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \mathsf{Y}) \le d_\ell\big]\mathsf{E}\big[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \hat{\mathsf{Y}})\big|\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \mathsf{Y}) \le d_\ell\big] + \mathsf{E}\big[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \mathsf{y}_0)\mathbb{1}_{\{\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})> d_\ell\}}\big] \\ &\overset{\mathrm{a}}{\le} d_\ell + \mathsf{E}\big[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \mathsf{y}_0)^p\big]^{1/p} \Pr[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \mathsf{Y}) > d_\ell]^{1-1/p} \\ &\overset{\mathrm{b}}{\le} d_\ell + d_p\, \varepsilon^{1-1/p} \end{aligned} \tag{3.34}$$

where (a) follows from Hölder's inequality and (b) from Assumption 3.1 and the excess distortion probability (3.30). ∎

To prove Lemma 3.3, observe that applying Lemma 3.6 to a sequence of optimal codes with sizes $M_{\mathrm{a}}^\star(n, \mathsf{d}, \varepsilon)$ gives

$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \mathsf{E}\big[\Delta_\ell^{\mathrm{a}}(\mathsf{X}, \hat{\mathsf{Y}})\big] \le d_\ell \tag{3.35}$$

for all $\ell \in [L]$ and thus

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log \overline{M}_{\mathrm{a}}^\star(n, \mathsf{d}) &\le \lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log(M_{\mathrm{a}}^\star(n, \mathsf{d}, \varepsilon) + 1) \\ &= \lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log M_{\mathrm{a}}^\star(n, \mathsf{d}, \varepsilon) \end{aligned} \tag{3.36}$$

which concludes the proof. ■.

Achievability for Excess Distortion

To prove the achievability result of Lemma 3.4, we start from the following theorem.

Theorem 3.7 (Theorem 26.4 in [PW17])**.** For all $P_{W|U}$ and $\gamma > 0$, there exists a code $U \to T \to g(T)$ with $T \in [M]$ and

$$\Pr[\Delta(U, g(T)) > d] \leq e^{-M/\gamma} + \Pr[\imath_{X;Y}(U;W) > \log\gamma] + \Pr[\Delta(U,W) > d]. \tag{3.37}$$

An inspection of the proof of Theorem 3.7 reveals that we can replace the event $\{\Delta(\cdot,\cdot) > d\}$ by $\bigcup_{\ell\in[L]}\{\Delta_\ell^{\mathrm{a}}(\cdot,\cdot) > d_\ell\}$ and similarly on the RHS. Fix a small $\varepsilon' > 0$ and choose $P_{Y|X}$ such that $\mathsf{E}[\delta(X,Y)\,|\,X \in \mathcal{I}_\ell] = d_\ell - \varepsilon'$ for all $\ell \in [L]$ with $d_\ell > 0$ and $\mathsf{E}[\delta(X,Y)\,|\,X \in \mathcal{I}_\ell] = 0$ if $d_\ell = 0$. We choose $(U,W) = (\mathsf{X},\mathsf{Y})$ with

$$\begin{aligned} P_U &= P_{\mathsf{X}} \\ P_{W|U} &= \prod_{i=1}^{n} P_{Y|X} \\ \log M &= n(I(X;Y) + 2\varepsilon') \\ \log\gamma &= n(I(X;Y) + \varepsilon') \\ \{\Delta(U,W) > d\} &= \bigcup_{\ell\in[L]} \{\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y}) > d_\ell\} \end{aligned} \tag{3.38}$$

for a fixed $\ell \in [L]$ and apply the union bound over our L distortion constraints to get

$$\begin{aligned} &\Pr\left[\bigcup_{\ell\in[L]} \left\{\Delta_\ell^{\mathrm{a}}\big(\mathsf{X}, g(f(\mathsf{X}))\big) > d_\ell\right\}\right] \\ &\qquad\leq e^{-M/\gamma} + \Pr[\imath_{X;Y}(\mathsf{X};\mathsf{Y}) > \log\gamma] + \sum_{\ell=1}^{L} \Pr[\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y}) > d_\ell]. \end{aligned} \tag{3.39}$$

To prove Lemma 3.4, we show that each of the three summands on the RHS of (3.37) goes to zero as $n \to \infty$. For the first one, we have

$$e^{-M/\gamma} = e^{-e^{n(I(X;Y)+2\varepsilon')-n(I(X;Y)+\varepsilon')}} = e^{-e^{n\varepsilon'}} \overset{n\to\infty}{\longrightarrow} 0 \tag{3.40}$$

for any $\varepsilon' > 0$. Next, since the pair $(X[i], Y[i])$ is iid, the weak law of large numbers tells us that

$$\Pr[\imath_{X;Y}(\mathsf{X};\mathsf{Y}) > \log\gamma] = \Pr\left[\frac{1}{n}\sum_{i=1}^{n} \imath_{X;Y}(X[i];Y[i]) > I(X;Y) + \varepsilon'\right] \overset{n\to\infty}{\longrightarrow} 0. \tag{3.41}$$

For the third summand in (3.37), first note that if $d_\ell = 0$, then $\mathsf{E}[\delta(X,Y)\,|\,X \in \mathcal{I}_\ell] = 0$

and thus $\Pr[\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})>0]=0$ as well. For $d_\ell>0$, we define the ε''-letter typical set (with ε'' to be specified later) as follows;

$$\mathcal{T}_{\varepsilon''}(\ell) \coloneqq \left\{\mathsf{x}\in\mathcal{X}^n : \left|\frac{1}{n}\mathsf{N}(\mathcal{I}_\ell|\mathsf{x}) - \Pr[X\in\mathcal{I}_\ell]\right| \le \varepsilon \Pr[X\in\mathcal{I}_\ell]\right\}. \tag{3.42}$$

By Hoeffding's inequality [Ver18, Thm. 2.2.6], we have $\Pr[\mathsf{X}\notin\mathcal{T}_{\varepsilon''}(\ell)]\le 2e^{-n(\varepsilon''\Pr[X\in\mathcal{I}_\ell])^2}$. The probability of exceeding the ℓth distortion constraint is then

$$\begin{aligned}
&\Pr[\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})>d_\ell]\\
&\quad= \Pr\left[\frac{1}{\mathsf{N}(\mathcal{I}_\ell|\mathsf{X})}\sum_{i=1}^n \mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])>d_\ell\right]\\
&\quad= \Pr\left[\frac{1}{\mathsf{N}(\mathcal{I}_\ell|\mathsf{X})}\sum_{i=1}^n \mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])>d_\ell, \mathsf{X}\in\mathcal{T}_{\varepsilon''}(\ell)\right]\\
&\qquad+\Pr[\mathsf{X}\notin\mathcal{T}_{\varepsilon''}(\ell)]\Pr\left[\frac{1}{\mathsf{N}(\mathcal{I}_\ell|\mathsf{X})}\sum_{i=1}^n \mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])>d_\ell\middle|\mathsf{X}\notin\mathcal{T}_{\varepsilon''}(\ell)\right]\\
&\quad\le \Pr\left[\frac{1}{n\Pr[X\in\mathcal{I}_\ell]}\sum_{i=1}^n \mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])>d_\ell(1-\varepsilon''), \mathsf{X}\in\mathcal{T}_{\varepsilon''}(\ell)\right]+\Pr[\mathsf{X}\notin\mathcal{T}_{\varepsilon''}(\ell)]\\
&\quad\le \Pr\left[\frac{1}{n\Pr[X\in\mathcal{I}_\ell]}\sum_{i=1}^n \mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])>d_\ell(1-\varepsilon'')\right]+2e^{-n(\varepsilon''\Pr[X\in\mathcal{I}_\ell])^2}.
\end{aligned}\tag{3.43}$$

Now note that

$$\mathsf{E}\left[\frac{1}{n\Pr[X\in\mathcal{I}_\ell]}\sum_{i=1}^n \mathbb{1}_{\{X[i]\in\mathcal{I}_\ell\}}\delta(X[i],Y[i])\right] = \frac{1}{n}\sum_{i=1}^n \mathsf{E}\Big[\delta(X[i],Y[i])\Big|X[i]\in\mathcal{I}_\ell\Big] = d_\ell-\varepsilon'. \tag{3.44}$$

Choosing $\varepsilon''=\varepsilon'/(2d_\ell)$, we ensure that the first probability in (3.43) approaches zero by the weak law of large numbers. Combining this with (3.40) and (3.41), we conclude that

$$\Pr\left[\bigcup_{\ell\in[L]}\{\Delta_\ell^{\mathrm{a}}(\mathsf{X},\mathsf{Y})>d_\ell\}\right]\xrightarrow{n\to\infty}0 \tag{3.45}$$

for all $P_{Y|X}$ with $\mathsf{E}[\delta(X,Y)\,|\,X\in\mathcal{I}_\ell]\le\max(d_\ell-\varepsilon',0)$ and

$$\limsup_{n\to\infty}\frac{1}{n}\log M_{\mathrm{a}}^\star(n,\mathsf{d},\varepsilon)\le I(X;Y)\,. \tag{3.46}$$

Optimizing over $P_{Y|X}$, letting $\varepsilon'\to 0$ as $n\to\infty$, and using the continuity of R_L completes the proof.

3.2. Finite Block Length

This section builds on the work by Kostina and Verdú [KV12], who derived general converse results that, together with Shannon's random coding technique, tightly characterize the finite length RD behavior of memoryless sources with respect to an excess distortion criterion. In this section, we first review the concept of d-tilted information, which plays a central role in finite length lossy compression. We then review existing bounds and add to this a new converse result which establishes an ordering between previous results. Finally, we evaluate our new converse for the binary memoryless and Gaussian sources with a single distortion constraint. An application to finite length coding rates with multiple distortion constraints follows in Chapter 3.3.

3.2.1. d-tilted Information

In [KV12], the d-*tilted information* is a key random variable to evaluate the performance of finite-length lossy compression. Based on the statistical variations of this quantity, one can derive bounds and approximations of the optimal coding rates at finite block length. For $(x, y) \in \mathcal{X} \times \mathcal{Y}$, recall that

$$\imath_{X;Y}(x;y) = \log \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(x,y) \tag{3.47}$$

is the *information density*, which is a key quantity of interest in finite-length almost lossless compression [Kos13]. Csiszár [Csi74] revisits Shannon's [Sha59] usual RD optimization problem (2.4) with one distortion constraint d for a single random variable and derives conditions that the optimal joint distribution $P_{XY}^\star$ must satisfy. It is known (see also [Gra11, Ch. 9.5] for a more accessible treatment) that the RD function can be written as

$$\begin{aligned}
\mathsf{R}(d) &= \inf_{\substack{P_{Y|X}:\\ \mathsf{E}[\delta(X,Y)]\le d}} \mathsf{E}[\imath_{X;Y}(X;Y)] \\
&= \max_{s\ge 0} \left(\inf_{P_{Y|X}} \mathsf{E}\Big[\imath_{X;Y}(X;Y) + s\delta(X,Y)\Big] - sd \right) && (3.48)\\
&= \max_{s\ge 0} \left(\inf_{P_Y} \mathsf{E}_X\Big[-\log \mathsf{E}_Y\Big[e^{-s\delta(X,Y)}\Big]\Big] - sd \right) && (3.49)
\end{aligned}$$

where the minimization in (3.48) is unconstrained over all $P_{Y|X}$ and the minimization in (3.49) is over all Y-marginals on $\mathcal{Y}$ with both expectations being unconditional. For a fixed d, the maximizing value is given by $s^\star := \frac{\partial \mathsf{R}(\hat{d})}{\partial \hat{d}}\Big|_{\hat{d}=d}$. We will assume that the RD function is achieved by a unique reconstruction variable $Y^\star$. This assumption is not essential (see [KV12, Sec. V], [Csi74]), but it simplifies the presentation. Using $Y^\star$, the

d-*tilted information* is defined as [KV12]

$$\jmath_X(x,d) \coloneqq -\log \mathsf{E}_{Y^\star}\Big[e^{s^\star(d-\delta(x,Y^\star))}\Big] \tag{3.50}$$

and (3.49) can be written as

$$\mathsf{R}(d) = \mathsf{E}_X\Big[\jmath_X(X,d)\Big]. \tag{3.51}$$

We next derive a similar characterization for the RDL function (3.6). Following along the same lines as [Csi74, Gra11], we introduce the vector s of length L with nonnegative elements to write

$$\begin{aligned}\mathsf{R}_\mathsf{L}(\mathsf{d}) &= \inf_{P_{Y|X}\in\mathcal{P}(X,\mathsf{d})} \mathsf{E}[\imath_{X;Y}(X;Y)] \\ &\geq \max_{\mathsf{s}\geq 0} \inf_{P_{Y|X}\in\mathcal{P}(X,\mathsf{d})} \left\{\mathsf{E}[\imath_{X;Y}(X;Y)] + \sum_{\ell=1}^{L} s_\ell \mathsf{E}\left[\frac{\mathbb{1}_{\{X\in\mathcal{I}_\ell\}}\delta(X,Y)}{\Pr[X\in\mathcal{I}_\ell]}\right] - \sum_{\ell=1}^{L} s_\ell d_\ell\right\}.\end{aligned} \tag{3.52}$$

For a fixed s, varying d parameterizes a hyperplane $(\mathsf{d}, \varphi(\mathsf{s},\mathsf{d}))$ described by the functional

$$\begin{aligned}\varphi(\mathsf{s},\mathsf{d}) &\coloneqq \inf_{P_{Y|X}\in\mathcal{P}(X,\mathsf{d})} \mathsf{E}\left[\imath_{X;Y}(X;Y) + \sum_{\ell=1}^{L} s_\ell \frac{\mathbb{1}_{\{X\in\mathcal{I}_\ell\}}\delta(X,Y)}{\Pr[X\in\mathcal{I}_\ell]}\right] - \sum_{\ell=1}^{L} s_\ell d_\ell \\ &= c(\mathsf{s}) - \sum_{\ell=1}^{L} s_\ell d_\ell.\end{aligned} \tag{3.53}$$

Now fix an arbitrary $\hat{\mathsf{d}}$. By the convexity of R_L in d, there is a vector $\mathsf{s}^\star = -\nabla \mathsf{R}_\mathsf{L}(\hat{\mathsf{d}})\big|_{\hat{\mathsf{d}}=\mathsf{d}}$ such that $(\mathsf{d}, \varphi(\mathsf{s}^\star,\mathsf{d}))$ passes through $(\hat{\mathsf{d}}, \mathsf{R}_\mathsf{L}(\hat{\mathsf{d}}))$ and has no point above $\mathsf{R}_\mathsf{L}(\mathsf{d})$. This is illustrated in Figure 3.1. With this choice of $\mathsf{s}^\star$, equality holds in (3.49) [Csi74, Cf. Lemma 1.2]. Continuing along the lines of [Gra11, Cor. 9.3], we rewrite (3.49) as

$$\begin{aligned}\mathsf{R}_\mathsf{L}(\mathsf{d}) &= \max_{\mathsf{s}\geq 0} \inf_{P_{Y|X}\in\mathcal{P}(X,\mathsf{d})} \left\{\mathsf{E}_X\left[-\log \mathsf{E}_Y\left[\exp\left(\sum_{\ell=1}^{L} \frac{\mathbb{1}_{\{X\in\mathcal{I}_\ell\}}\delta(X,Y)}{\Pr[X\in\mathcal{I}_\ell]}\right)\right]\right] - \sum_{\ell=1}^{L} s_\ell d_\ell\right\} \\ &= \mathsf{E}[\jmath_X(X,\mathsf{d})]\end{aligned} \tag{3.54}$$

where we define the d-tilted information for multiple constraints via

$$\jmath_X(x,\mathsf{d}) \coloneqq -\log \mathsf{E}_{Y^\star}\left[\exp\Big(-\sum_{\ell=1}^{L} s_\ell^\star \Delta_k^{\mathrm{e}}(X,Y)\Big)\right] - \sum_{\ell=1}^{L} s_\ell^\star d_\ell \tag{3.55}$$

where $Y^\star$ is a reconstruction random variable that achieves the RDL function, and where $s^\star = \frac{\partial \mathsf{R}_\mathsf{L}(\mathsf{d})}{\partial d_\ell}$ for all $\ell \in [L]$.

We later identify a single random variable and its d-tilted information with a string of n letters. To this end, note that the RDL function of X as defined in (3.6) is n times the RDL function of X. Thus, the slope of the n-letter RDL function is $\mathsf{s}^\star n$ and we can

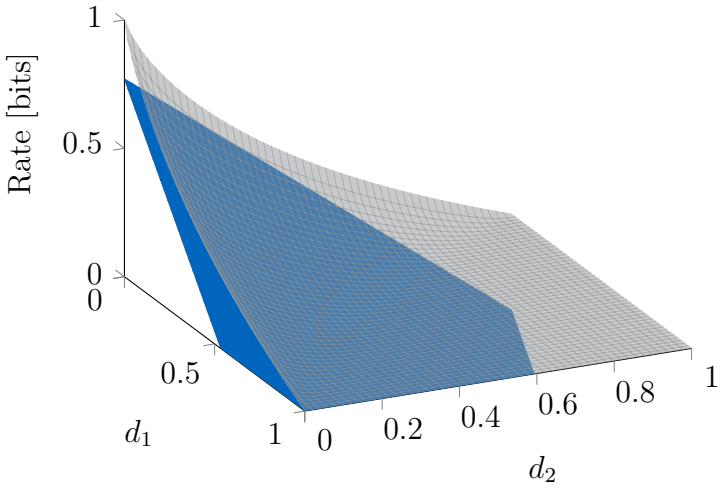


Figure 3.1.: Illustration of $(\mathsf{d}, \mathsf{R}_{\mathsf{L}}(\mathsf{d}))$ (gray) and $(\mathsf{d}, \varphi(\mathsf{s}, \mathsf{d}))$ (blue) for two distortion constraints.

calculate

$$\begin{aligned} \jmath_{\mathsf{X}}(\mathsf{x}, \mathsf{d}) &:= -\log \mathsf{E}_{\mathsf{Y}^\star}\left[\exp\left(-\sum_{\ell=1}^{L} n s_\ell^\star \Delta_\ell^{\mathrm{e}}(\mathsf{x}, \mathsf{Y})\right)\right] - \sum_{\ell=1}^{L} n s_\ell^\star d_\ell \\ &= \sum_{i=1}^{n} \jmath_X(x[i], \mathsf{d}). \end{aligned} \tag{3.56}$$

3.2.2. Previous Finite-Length Bounds

In [KV12], two general converse bounds for the smallest codebook size of a lossy source code with excess distortion probability ε are derived. The first one [KV12, Thm. 7] is rather simple and depends only on the d-tilted information, but also seems to be relatively loose. The second bound [KV12, Thm. 8] is based on a hypothesis testing argument and is more general. Consequently, the latter bound may be tight but needs to be relaxed for computation as it involves the supremization over n-letter (probability) measures.

Before stating the bounds, we make a few remarks.

- ▷ The bounds are derived in a *one-shot* setting, i.e., for a single random variable X. In order to apply them do a memoryless source with block length n, we identify X with an n-letter iid random variable and compute the bound.
- ▷ Since the bounds are one-shot, there are no restrictions with respect to the distortion measure and the results carry over to a non-separable distortion measure.
- ▷ Thus, the extension of the bounds to multiple distortion constraints is straightforward. One simply replaces the requirement of satisfying one distortion constraint by the requirement of satisfying multiple distortion constraints.

For simplicity, we will therefore state the bounds in terms of *distortion balls*:

$$\mathcal{B}(X,\mathsf{d}) = \Big\{ y \in \mathcal{Y} : \begin{smallmatrix}\text{all distortion constraints with}\\ \text{limits given by } \mathsf{d} \text{ are satisfied}\end{smallmatrix} \Big\} \tag{3.57}$$

where the number of constraints and the distortion measure should be clear from the context. To avoid cumbersome notation, we will drop the block length and type of distortion measure and simply refer to an $(M,\mathsf{d},\varepsilon)$-code for now. First, let us restate the converse bounds.

Theorem 3.8 (Kostina & Verdú [KV12])**.** An $(M,\mathsf{d},\varepsilon)$-code satisfies

$$\varepsilon \geq \sup_{\gamma\geq 0} \Big(\Pr[\jmath_X(X,\mathsf{d}) \geq \log M + \gamma] - e^{-\gamma} \Big). \tag{3.58}$$

For the second converse, which is also called *meta-converse* in the literature, let

$$\beta_\alpha(p,q) = \min_{\substack{P_{W|X}:\\ \Pr[W=1]\geq\alpha}} \mathbb{Q}[W=1] \tag{3.59}$$

denote the optimal performance achievable among all randomized tests $P_{W|X}:\mathcal{X}\to\{0,1\}$ between probability distributions P_X an Q_X where $W=1$ indicates that the test chooses P_X, and where $\mathbb{Q}[\cdot]$ is the probability of an event if X has distribution Q_X.

Theorem 3.9 (Kostina & Verdú [KV12])**.** An $(M,\mathsf{d},\varepsilon)$-code satisfies

$$M \geq \sup_{Q_X} \inf_{y\in\mathcal{Y}} \frac{\beta_{1-\varepsilon}(P_X,Q_X)}{\mathbb{Q}[y\in\mathcal{B}(X,\mathsf{d})]} \tag{3.60}$$

where the supremum is over all σ-finite measures (see [KV12, Rem. 5]).

For an achievability result, note that we "simply" need a d-cover for *any* set of probability $1-\varepsilon$ (recall Figure 2.2). Depending on the geometry of $\mathcal{X}$, however, finding precise covering numbers can be difficult. For Euclidean space and MSE (or distance) distortion, for example, one can use the results from [Rog63, VG05] to derive an achievability result, as done for the Gaussian memoryless source in [KV12, Thm. 39].

A general achievability result that has proved to be extremely useful is the finite length version of Shannon's *random coding bound*:

Theorem 3.10 (Shannon [Sha59])**.** For any probability measure P_Y on $\mathcal{Y}$, there is an $(M,\mathsf{d},\varepsilon)$-code satisfying

$$\begin{aligned} \varepsilon &\leq \inf_{P_Y} \mathsf{E}\Big[\big(1 - P_Y(\mathcal{B}(X,\mathsf{d}))\big)^M \Big] \\ &\leq \inf_{P_Y} \mathsf{E}\Big[\exp\big(-M\cdot P_Y(\mathcal{B}(X,\mathsf{d}))\big)\Big] \end{aligned} \tag{3.61}$$

where the second bound is useful for numerical stability, see [KV12, Thm. 9-10].

3.2.3. New Converse Bound

We add to Theorems 3.8 and 3.9 a third converse bound that is computable but tighter than Theorem 3.8. After deriving this bound, we realized that it can also be deduced from [KV12, Thm. 8] which establishes a connection between the two converse bounds in [KV12] and shows that the hypothesis testing bound is always tighter. In hindsight, we can state the following converse bound as a corollary of Theorem 3.9 and present its proof via a specific choice of Q_X as a σ-finite measure.

Corollary 3.11. An $(M, \mathsf{d}, \varepsilon)$ code satisfies

$$M \geq \sup_{\gamma \in \mathbb{R}} \left(\frac{\Pr[\jmath_X(X, \mathsf{d}) \geq \gamma] - \varepsilon}{\sup_{y \in \mathcal{Y}} \Pr\big[\jmath_X(X, \mathsf{d}) \geq \gamma, y \in \mathcal{B}(X, \mathsf{d})\big]} \right). \tag{3.62}$$

Proof. Fix some $\gamma \in \mathbb{R}$. Let $\sigma(X)$ denote the σ-algebra generated by X and choose

$$\mathbb{Q}[X \in \mathcal{F}] = \Pr\big[X \in \mathcal{F}, \jmath_X(X, \mathsf{d}) \geq \gamma\big] \tag{3.63}$$

for all $\mathcal{F} \in \sigma(X)$. An optimal randomized test between P_X and Q_X is

$$P_{W|X}(1|x) = \begin{cases} 1, & \text{if } \jmath_X(X, \mathsf{d}) < \gamma \\ \dfrac{\Pr[\jmath_X(X, \mathsf{d}) \geq \gamma] - \varepsilon}{\Pr[\jmath_X(X, \mathsf{d}) \geq \gamma]}, & \text{if } \jmath_X(X, \mathsf{d}) \geq \gamma. \end{cases} \tag{3.64}$$

The probability that this test succeeds under P_X is

$$\begin{aligned} \Pr[W = 1] &= \Pr[\jmath_X(X, d) < \gamma] \Pr[W = 1 \,|\, \jmath_X(X, d) < \gamma] \\ &\quad + \Pr[\jmath_X(X, d) \geq \gamma] \Pr[W = 1 \,|\, \jmath_X(X, d) \geq \gamma] \\ &= 1 - \varepsilon \end{aligned} \tag{3.65}$$

and the measure of the event $\{W = 1\}$ under Q_X is

$$\begin{aligned} \mathbb{Q}[W = 1] &= \Pr[W = 1, \jmath_X(X, \mathsf{d}) \geq \gamma] \\ &= \Pr[\jmath_X(X, \mathsf{d}) \geq \gamma] \frac{\Pr[\jmath_X(X, \mathsf{d}) \geq \gamma] - \varepsilon}{\Pr[\jmath_X(X, \mathsf{d}) \geq \gamma]} \\ &= \Pr[\jmath_X(X, \mathsf{d}) \geq \gamma] - \varepsilon. \end{aligned} \tag{3.66}$$

Inserting (3.63) and (3.66) into (3.60) and taking the supremum over γ completes the proof. ∎

We next show that Corollary 3.11 gives a better converse bound than Theorem 3.8 for

$\ell = 1$. To this end, it is helpful to rewrite (3.62) as a lower bound on ε:

$$\varepsilon \geq \sup_{\gamma \in \mathbb{R}} \Big(\Pr[\jmath_X(X,d) \geq \gamma] - M \sup_{y \in \mathcal{Y}} \Pr[\jmath_X(X,d) \geq \gamma, \delta(X,y) \leq d] \Big). \tag{3.67}$$

Now choose $\gamma = \log M + \tilde{\gamma}$. Following along the lines of the proof of [KV12, Thm. 7], we have

$$\begin{aligned} & M \sup_{y \in \mathcal{Y}} \Pr[\jmath_X(X,d) \geq \log M + \tilde{\gamma}, \delta(X,y) \leq d] \\ & \quad = M \sup_{y \in \mathcal{Y}} \mathsf{E}\Big[\mathbb{1}_{\left\{ \frac{1}{M} e^{\jmath_X(X,d) - \tilde{\gamma}} \geq 1, e^{s^\star (d - \delta(X,y))} \geq 1 \right\}} \Big] \\ & \quad \leq M \sup_{y \in \mathcal{Y}} \mathsf{E}\Big[\frac{1}{M} e^{\jmath_X(X,d) - \tilde{\gamma}} \mathbb{1}_{\left\{ e^{s^\star (d - \delta(X,y))} \geq 1 \right\}} \Big] \\ & \quad \overset{\text{(i)}}{\leq} e^{-\tilde{\gamma}} \sup_{y \in \mathcal{Y}} \mathsf{E}\Big[e^{\jmath_X(X,d) + s^\star (d - \delta(X,y))} \Big] \\ & \quad \leq e^{-\tilde{\gamma}} \end{aligned} \tag{3.68}$$

where (i) applies the next lemma from [Csi74].

Lemma 3.12 (Eq. (1.22) [Csi74])**.** For all $y \in \mathcal{Y}$, we have

$$\mathsf{E}\Big[e^{\jmath_X(X,d) + s^\star (d - \delta(X,y))} \Big] \leq 1 \tag{3.69}$$

with equality for $P_{Y^\star}$-almost all y.

■

3.2.4. Binary Memoryless Source with Hamming Distortion

In this section, we evaluate Corollary 3.62 for the special case of a Binary Memoryless Source (BMS) with one Hamming distortion constraint. Let X be a string of n iid instances of X with $\Pr[X = 1] = 1 - \Pr[X = 0] = p$, and choose the distortion function as

$$\Delta(\mathsf{x}, \mathsf{y}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{x[i] \neq y[i]\}}.$$

We have the following corollary.

Corollary 3.13 (BMS)**.** Fix $p \in (0, 1/2)$ and $d \in [0, p)$. An (n, M, d, ε) code satisfies

$$M \geq \max_{0 \leq b \leq n} \left(\frac{\sum_{k=b}^{n} \binom{n}{k} p^k (1-p)^{n-k} - \varepsilon}{\alpha_{n,d,p}(b)} \right) \tag{3.70}$$

where

$$\alpha_{n,d,p}(b) = \max_{\hat{n}_1} \sum_{k=0}^{\lfloor nd \rfloor} \sum_{j=0}^{k} \binom{\hat{n}_1}{j} \binom{n-\hat{n}_1}{k-j} \cdot p^{\hat{n}_1+k-2j}(1-p)^{n-\hat{n}_1-k+2j} \mathbb{1}_{\{\hat{n}_1+k-2j \geq b\}} \tag{3.71}$$

and the maximization is taken over all $\hat{n}_1 \in \mathbb{N}$ satisfying

$$\max\left\{0, b - \lfloor nd \rfloor\right\} \leq \hat{n}_1 \leq \min\left\{n, b + \lfloor nd \rfloor\right\}.$$

Note that Corollary 3.13 does not weaken Corollary 3.11; i.e., the RHSs of (3.62) and (3.70) are equal for the BMS with Hamming distortions.

Remark 3.1. For $p = 1/2$, $\jmath_{\mathsf{X}}(\mathsf{x}, d)$ does not depend on x [KV12, Example 1]. In this case, Corollary 3.13 coincides with [KV12, Thm. 20] which is derived from the meta-converse bound.

Figure 3.2 compares our converse with the previously existing bounds. We see that in this example, our bound is slightly better except for very short block lengths.

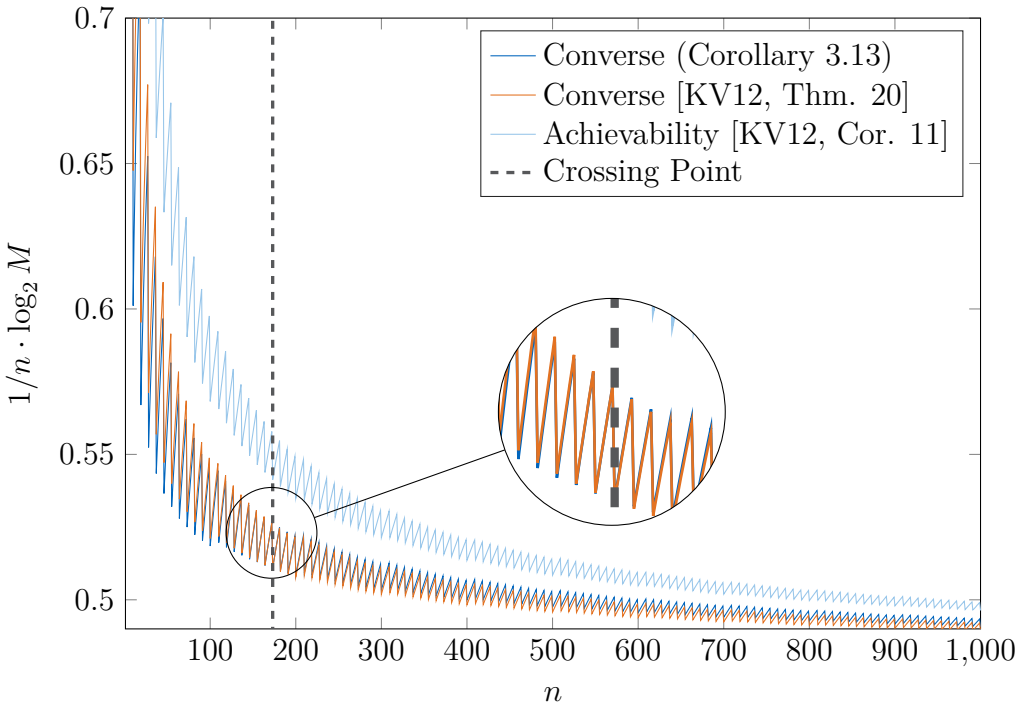


Figure 3.2.: BMS with $p = 2/5$, $d = 0.11$, $\varepsilon = 0.01$.

Proof of Corollary 3.13. Fix $p \in (0, 1/2)$, $d \in [0, p)$ and $\gamma \in \mathbb{R}$. We have [KV12, Eqn. (21)]

$$\jmath_{\mathsf{X}}(\mathsf{x}, d) = \mathsf{N}(1|\mathsf{x}) \log \frac{1}{p} + (n - \mathsf{N}(1|\mathsf{x})) \log \frac{1}{1-p} - nH_2(d)\,.$$

Since $p \in (0, 1/2)$, it follows that $p < 1 - p$ and $\jmath_{\mathsf{X}}(\mathsf{x}, d)$ grows linearly in $\mathsf{N}(1|\mathsf{x})$ for fixed n. Let

$$b \coloneqq \min\left\{n' \in [n] : n' \log \frac{1}{p} + (n - n') \log \frac{1}{1-p} - nH_2(d) \geq \gamma\right\}$$

and note that

$$\Big\{\mathsf{x} \in \mathcal{X}^n \colon \jmath_{\mathsf{X}}(\mathsf{x}, d) \geq \gamma\Big\} = \Big\{\mathsf{x} \in \mathcal{X}^n \colon \mathsf{N}(1|\mathsf{x}) \geq b\Big\}. \tag{3.72}$$

Hence, we have

$$\Pr[\jmath_{\mathsf{X}}(\mathsf{X}, d) \geq \gamma] = \Pr\Big[\mathsf{N}(1|\mathsf{X}) \geq b\Big] = \sum_{k=b}^{n} \binom{n}{k} p^k (1-p)^{n-k}. \tag{3.73}$$

Now consider the denominator of Corollary 3.11. Let $\hat{n}_1 \coloneqq \mathsf{N}(1|\mathsf{y})$. Using Vandermonde's identity, the number of binary sequences in a Hamming ball of size $\lfloor nd \rfloor$ centered at a sequence of Hamming weight $\hat{n}_1$ is given by

$$\sum_{k=0}^{\lfloor nd \rfloor} \binom{n}{k} = \sum_{k=0}^{\lfloor nd \rfloor} \sum_{l=0}^{k} \binom{\hat{n}_1}{l} \binom{n - \hat{n}_1}{k - l} \tag{3.74}$$

where $\binom{\hat{n}_1}{l}\binom{n-\hat{n}_1}{k-l}$ is the number of sequences of Hamming weight $\hat{n}_1 + k - 2l$. We can thus write

$$\begin{aligned}
&\sup_{\mathsf{y} \in \mathcal{Y}^n} \Pr\Big[\jmath_{\mathsf{X}}(\mathsf{X}, d) \geq \gamma, \Delta(\mathsf{X}, \mathsf{y}) \leq nd\Big] \\
&\quad \overset{\text{(i)}}{=} \max_{\hat{n}_1} \Pr\Big[\mathsf{N}(1|\mathsf{X}) \geq b, \Delta(\mathsf{X}, \mathsf{y}) \leq nd\Big] \\
&\quad = \max_{\hat{n}_1} \sum_{\mathsf{x}} \Pr[\mathsf{X} = \mathsf{x}] \mathbb{1}_{\{\mathsf{N}(1|\mathsf{x}) \geq b, \Delta(\mathsf{x}, \mathsf{y}) \leq nd\}} \\
&\quad = \max_{\hat{n}_1} \sum_{k=0}^{\lfloor nd \rfloor} \sum_{l=0}^{k} \binom{\hat{n}_1}{l} \binom{n - \hat{n}_1}{k - l} p^{\hat{n}_1 + k - 2l} \cdot (1-p)^{n - \hat{n}_1 - k + 2l} \mathbb{1}_{\{\hat{n}_1 + k - 2l \geq b\}}
\end{aligned} \tag{3.75}$$

where (i) follows since, by symmetry, the probability depends on y only through $\hat{n}_1$.

To complete the proof, we show that it suffices to consider $b - \lfloor nd \rfloor \leq \hat{n}_1 \leq b + \lfloor nd \rfloor$ for the maximization. The lower bound is immediate because for $\hat{n}_1 < b - \lfloor nd \rfloor$, we have $\mathbb{1}_{\{\hat{n}_1 + k - 2l \geq b\}} = 0$ for all summands. For $\hat{n}_1 > b + \lfloor nd \rfloor$, we have $\mathbb{1}_{\{\hat{n}_1 + k - 2l \geq b\}} = 1$ for all summands, but the sum is monotonically decreasing in $\hat{n}_1$ as we shall show next. Consider two length-n sequences $\mathsf{a}, \mathsf{b} \in \{0,1\}^n$ with $a[j] = 0$, $b[j] = 1$ for some $j \in [n]$ and $a[i] = b[i]$ for all $i \neq j$, i.e., a and b differ at only at the j-th position. Let $\mathcal{B}_H(\mathsf{a}, d)$ be the Hamming ball of radius $\lfloor nd \rfloor$ around a:

$$\mathcal{B}_H(\mathsf{a}, d) \coloneqq \left\{\mathsf{x} \in \{0,1\}^n : \sum_{i=1}^{n} \mathbb{1}_{\{x[i] \neq a[i]\}} \leq \lfloor nd \rfloor\right\}. \tag{3.76}$$

If we can show that

$$\Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{a}, d)\Big] \geq \Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{b}, d)\Big] \tag{3.77}$$

then we can infer by induction that

$$\Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{c}, d)\Big] \geq \Pr\Big[\mathsf{X} \in \mathcal{B}_H(\hat{\mathsf{c}}, d)\Big] \tag{3.78}$$

holds for all $\mathsf{c}, \hat{\mathsf{c}} \in \{0,1\}^n$ with $\mathsf{N}(1|\mathsf{c}) < \mathsf{N}(1|\hat{\mathsf{c}})$, since the probabilities on the left hand side and RHS depend only on $\mathsf{N}(1|\cdot)$ for fixed p, n, d. We first note that

$$\begin{aligned}
&\Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{a}, d)\Big] \\
&\quad = \Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{a}, d) \cap \mathcal{B}_H(\mathsf{b}, d)\Big] + \Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{a}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{b}, d)\Big] \\
&\Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{b}, d)\Big] \\
&\quad = \Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{a}, d) \cap \mathcal{B}_H(\mathsf{b}, d)\Big] + \Pr\Big[\mathsf{X} \in \mathcal{B}_H^{\mathrm{c}}(\mathsf{a}, d) \cap \mathcal{B}_H(\mathsf{b}, d)\Big].
\end{aligned} \tag{3.79}$$

Hence, to prove (3.77) it suffices to show that

$$\Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{a}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{b}, d)\Big] \geq \Pr\Big[\mathsf{X} \in \mathcal{B}_H(\mathsf{b}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{a}, d)\Big]. \tag{3.80}$$

Fix a sequence $\mathsf{x}_0 \in \mathcal{B}_H(\mathsf{a}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{b}, d)$. Since $\mathsf{x}_0 \in \mathcal{B}_H(\mathsf{a}, d)$, but $\mathsf{x}_0 \notin \mathcal{B}_H(\mathsf{b}, d)$, x_0 differs from a at most at d positions and from b at least at $d+1$ positions. Since a and b are the same except at position j, x_0 must be different from b at position j and thus $x_0[j] = 0$. Next, we can define x_1 via

$$\begin{aligned}
&x_1[k] = x_0[k], \quad k \in [n],\, k \neq j \\
&x_1[j] = 1.
\end{aligned} \tag{3.81}$$

Clearly, we have

$$\mathsf{x}_1 \in \mathcal{B}_H(\mathsf{b}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{a}, d) \text{ and } \Pr[\mathsf{X} = \mathsf{x}_0] > \Pr[\mathsf{X} = \mathsf{x}_1] \tag{3.82}$$

since $\mathsf{N}(1|\mathsf{x}_0) < \mathsf{N}(1|\mathsf{x}_1)$ and $p < 1 - p$. Since $\mathsf{x}_0 \in \mathcal{B}_H(\mathsf{a}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{b}, d)$ was arbitrary, we constructed a unique $\mathsf{x}_1 \in \mathcal{B}_H(\mathsf{b}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{a}, d)$ for every such x_0. Furthermore, as any Hamming ball of a fixed radius has the same number of elements, we constructed a one-to-one correspondence between each two elements in $\mathcal{B}_H(\mathsf{a}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{b}, d)$ and $\mathcal{B}_H(\mathsf{b}, d) \cap \mathcal{B}_H^{\mathrm{c}}(\mathsf{a}, d)$ where (3.82) holds for each pair. We can thus conclude that (3.80) holds.

■

3.2.5. Gaussian Memoryless Source with Squared Error Distortion

Let X be a string of n iid instances of $X \sim \mathcal{N}(0,1)$, and consider squared error distortions

$$\Delta(\mathsf{x},\mathsf{y}) = \frac{1}{n}\sum_{i=1}^{n}(x[i]-y[i])^2.$$

A slight weakening of Corollary 3.11 for this setting yields the next corollary for the Gaussian Memoryless Source (GMS). Here $f_{\chi^2_n}(\cdot)$ denotes the χ^2_n PDF.

Corollary 3.14. Fix $d \in (0,1)$. An (n,M,d,ε) code satisfies

$$M \geq \sup_{\gamma \geq nd}\left(\frac{\int_{\gamma}^{\infty} f_{\chi^2_n}(w)\mathrm{d}w - \varepsilon}{\frac{1}{2}I_{nd/\gamma}\left(\frac{n-1}{2},\frac{1}{2}\right)\int_{\gamma}^{\gamma^\star} f_{\chi^2_n}(w)\mathrm{d}w}\right) \tag{3.83}$$

where $I_{(\cdot)}(\cdot,\cdot)$ is the regularized incomplete beta function and

$$\gamma^\star := \left[\frac{2(nd)^{n/2}}{I_{nd/\gamma}\left(\frac{n-1}{2},\frac{1}{2}\right)} + \gamma^{n/2}\right]^{2/n}. \tag{3.84}$$

For a numerical example, let $d = 0.25$, $\sigma^2 = 1$ and $\varepsilon = 10^{-2}$. Figure 3.3 plots the bound in (3.83) and, for comparison, the converse bound [KV12, Theorem 36], which can be derived from the meta converse (Theorem 3.9). Our result is tighter for $n \geq 12$. We also included the Gaussian approximation [KV12, Theorem 40]. Here, choosing small values for d shifts the crossing point to larger n whereas varying ε does not seem to have a significant influence.

Proof of Corollary 3.14. The d-tilted information for the GMS with $d < \sigma^2 = 1$ is given by [KV12, Example 2]

$$\jmath_{\mathsf{X}}(\mathsf{x},d) = \frac{n}{2}\log\frac{1}{d} + \frac{\|\mathsf{x}\|_2^2 - n}{2}\log e$$

which grows linearly in $\|\mathsf{x}\|_2^2$. Hence, we can rewrite (3.62) as

$$M \geq \sup_{\gamma \geq 0}\left(\frac{\Pr\left[\|\mathsf{X}\|_2^2 \geq \gamma\right] - \varepsilon}{\sup\limits_{\mathsf{y}\in\mathbb{R}^n}\Pr\left[\|\mathsf{X}\|_2^2 \geq \gamma, \Delta(\mathsf{X},\mathsf{y}) \leq d\right]}\right). \tag{3.85}$$

We will lower bound (3.85) using a geometric argument for the denominator. By the circular symmetry of the GMS, we need to consider only those $\mathsf{y} \in \mathbb{R}^n$ for the supremum

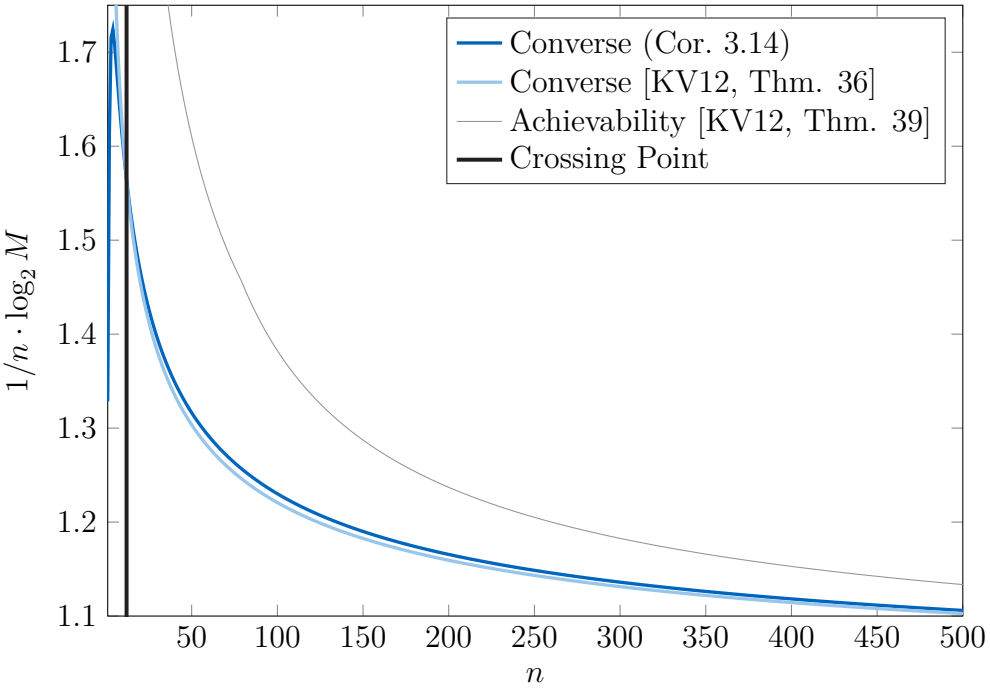


Figure 3.3.: GMS, $d = 0.25$, $\sigma^2 = 1$, $\varepsilon = 0.01$.

that lie on an arbitrary straight line through the origin. Define

$$\begin{aligned} \mathcal{A} &\coloneqq \left\{ \mathsf{x} \in \mathbb{R}^n : \|\mathsf{x}\|_2^2 \geq \gamma \right\} \\ \mathcal{B}_2(\mathsf{y}, d) &\coloneqq \left\{ \mathsf{x} \in \mathbb{R}^n : \|\mathsf{x} - \mathsf{y}\|_2^2 \leq nd \right\} \end{aligned} \tag{3.86}$$

and observe that

$$\begin{aligned} \sup_{\mathsf{y} \in \mathbb{R}^n} \Pr\left[\|\mathsf{X}\|_2^2 \geq \gamma, \Delta(\mathsf{X}, \mathsf{y}) \leq d \right] &= \sup_{\mathsf{y} \in \mathbb{R}^n} \Pr\left[\mathsf{X} \in \mathcal{A} \cap \mathcal{B}_2(\mathsf{y}, d) \right] \\ &= \sup_{\mathsf{y} \in \mathcal{L}} \Pr\left[\mathsf{X} \in \mathcal{A} \cap \mathcal{B}_2(\mathsf{y}, d) \right] \end{aligned} \tag{3.87}$$

where $\mathcal{L}$ denotes the set of points lying on an arbitrary straight line through the origin, see Figure 3.4(a).

Denote the surface area of an n-dimensional sphere of radius r by $S_n(r)$ and the surface area of a n-dimensional spherical cap of radius r and half angle θ by $A_n(r, \theta)$. The following relation holds [Li11]:

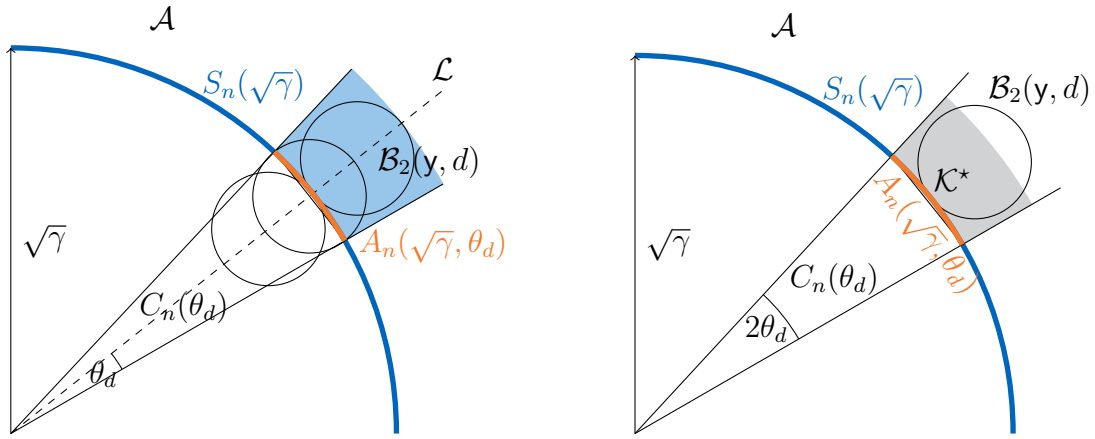
$$A_n(r, \theta) \coloneqq \frac{1}{2} S_n(r) I_{\sin^2(\theta)}\left(\frac{n-1}{2}, \frac{1}{2} \right) \tag{3.88}$$

where $I_{(\cdot)}(\cdot, \cdot)$ is the regularized incomplete beta function. Using the law of sines and taking $\gamma > nd$, we can determine the half angle θ_d such that $A_n(\sqrt{\gamma}, \theta_d)$ is the largest spherical

cap at radius $\sqrt{\gamma}$ contained in some $\mathcal{B}_2(\mathsf{y}, d)$:

$$\theta_d = \sin^{-1}\sqrt{nd/\gamma}. \tag{3.89}$$

Let $C_n(\theta_d)$ be the n-dimensional infinite cone of half angle θ_d that passes through $A_n(\sqrt{\gamma}, \theta_d)$. Clearly, $\mathcal{A} \cap \mathcal{B}_2(\mathsf{y}, d) \subset C_n(\theta_d)$ for any $\mathsf{y} \in \mathcal{L}$. This setup is visualized in Figure 3.4. Next,



(a) Intersection of $\mathcal{A} \cap C_n(\theta_d)$ with possible distortion balls that are centered on the straight line $\mathcal{L}$.

(b) Geometry of $\mathcal{K}^\star$ (gray area).

Figure 3.4.: Illustration of the geometry of the converse bound for the GMS.

denote the volume of $\mathcal{B}_2(\mathsf{y}, d)$ for any $\mathsf{y} \in \mathbb{R}^n$ by

$$V_n\left(\sqrt{nd}\right) := \frac{\pi^{n/2}}{\Gamma\left(\frac{n+2}{2}\right)}(nd)^{n/2} \tag{3.90}$$

where $\Gamma(\cdot)$ is the gamma function. To upper bound (3.87), we consider the largest probability of any set in $\mathcal{A} \cap C_n(\theta_d)$ (the shaded area in Figure 3.4(a)) that has the same volume as a distortion ball. We denote this set by

$$\mathcal{K}^\star := \underset{\substack{\mathcal{K} \subset \mathcal{A} \cap C_n(\theta_d): \\ \mathrm{Vol}(\mathcal{K}) = V_n(\sqrt{nd})}}{\arg\max} \Pr\left[\mathsf{X} \in \mathcal{K}\right]. \tag{3.91}$$

The geometry of the arg max problem is depicted in Figure 3.4(b). By the circular symmetry, $\mathcal{K}^\star$ is the slice of the cone $C_n(\theta_d)$ that lies on the surface of $S_n\left(\sqrt{\gamma}, \theta_d\right)$ and has volume $V_n\left(\sqrt{nd}\right)$. More precisely, we can describe $\mathcal{K}^\star$ as the difference between spherical sectors of half angle θ_d whose volumes differ by exactly $V_n(\sqrt{nd})$, see Figure 3.4(b).

The volume of a hypershperical sector of half angle θ and radius r is given by [Li11]

$$V_n^{\text{sec}}(r,\theta) \coloneqq \frac{1}{2}V_n(r)I_{\sin^2(\theta)}\left(\frac{n-1}{2},\frac{1}{2}\right).$$

Let $\gamma^\star$ be the solution to

$$V_n^{\text{sec}}(\sqrt{\gamma^\star},\theta_d) - V_n^{\text{sec}}(\sqrt{\gamma},\theta_d) = V_n(\sqrt{nd}) \tag{3.92}$$

which, using $\sin^2(\theta_d) = nd/\gamma$, can be rewritten as (3.84). Finally, we can use the tools developed in (3.87)–(3.92) to bound

$$\begin{aligned}
\sup_{\mathsf{y}\in\mathcal{L}} \Pr\Big[\mathsf{X}\in\mathcal{A}\cap\mathcal{B}_2(\mathsf{y},d)\Big] &\overset{\text{(a)}}{\leq} \Pr\Big[\mathsf{X}\in\mathcal{K}^\star\Big] \\
&\overset{\text{(b)}}{=} \Pr\Big[\gamma\leq\|\mathsf{X}\|^2\leq\gamma^\star, \mathsf{X}\in C_n(\theta_d)\Big] \\
&\overset{\text{(c)}}{=} \Pr\Big[\gamma\leq\|\mathsf{X}\|^2\leq\gamma^\star\Big]\Pr\Big[\mathsf{X}\in C_n(\theta_d)\Big], \\
&\overset{\text{(d)}}{=} \Pr\Big[\gamma\leq\|\mathsf{X}\|^2\leq\gamma^\star\Big]\frac{A_n(\sqrt{\gamma},\theta_d)}{S_n(\sqrt{\gamma})} \\
&= \frac{1}{2}I_{nd/\gamma}\left(\frac{n-1}{2},\frac{1}{2}\right)\int_\gamma^{\gamma^\star} f_{\chi^2_n}(w)\mathrm{d}w
\end{aligned} \tag{3.93}$$

where (a) follows from the definition of $\mathcal{K}^\star$ (3.91), (b) follows from the definition of $\gamma^\star$ (3.92) and the geometry of $\mathcal{K}^\star$; and (c)–(d) are a result of the circular symmetry of the multivariate Gaussian. Combining (3.93) and (3.85) yields (3.83).

■

3.3. Binary Memoryless Source with Letter-Based Distortions

This section applies the ideas and results of Section 3.1 and Section 3.2 to a BMS with two individual Hamming distortion constraints. We have $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $P_X(0) = 1 - P_X(1) = p$ and

$$\delta(x,y) = \mathbb{1}_{\{x\neq y\}}. \tag{3.94}$$

We first study the asymptotic RDL function for the BMS with two individual Hamming distortion constraints in Section 3.3.1. Then, Section 3.3.2 uses the tools from Section 3.2 to study finite length bounds for this source for both distortion measures Δ^{a} and Δ^{e}.

3.3.1. Infinite Block Length

Recall the standard RD function of the BMS with a single Hamming distortion constraint.

Proposition 3.15. If $p \leq 1/2$ and $d > 0$ then [CT06b, Thm. 10.3.1]

$$\mathsf{R}(d) = \begin{cases} H_2(p) - H_2(d)\,, & \text{if } 0 \leq d \leq p \\ 0, & \text{otherwise.} \end{cases} \tag{3.95}$$

Now let $\mathsf{R}_\mathsf{L}^{\mathrm{BMS}}(d_0, d_1)$ be the RDL function for the BMS with two Hamming distortion constraints as defined in Def. 3.3, i.e., we choose $L = 2$, $\mathcal{I}_0 = \{0\}$ and $\mathcal{I}_1 = \{1\}$.

Theorem 3.16. Let $q = (1-p)d_0 + p(1-d_1)$. Then, we have

$$\mathsf{R}_\mathsf{L}^{\mathrm{BMS}}(d_0, d_1) = \begin{cases} H_2(q) - (1-p)H_2(d_0) - pH_2(d_1) & \text{if } d_0 + d_1 < 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.96}$$

Figure 3.3.1 plots the RDL function for $p = 0.35$.

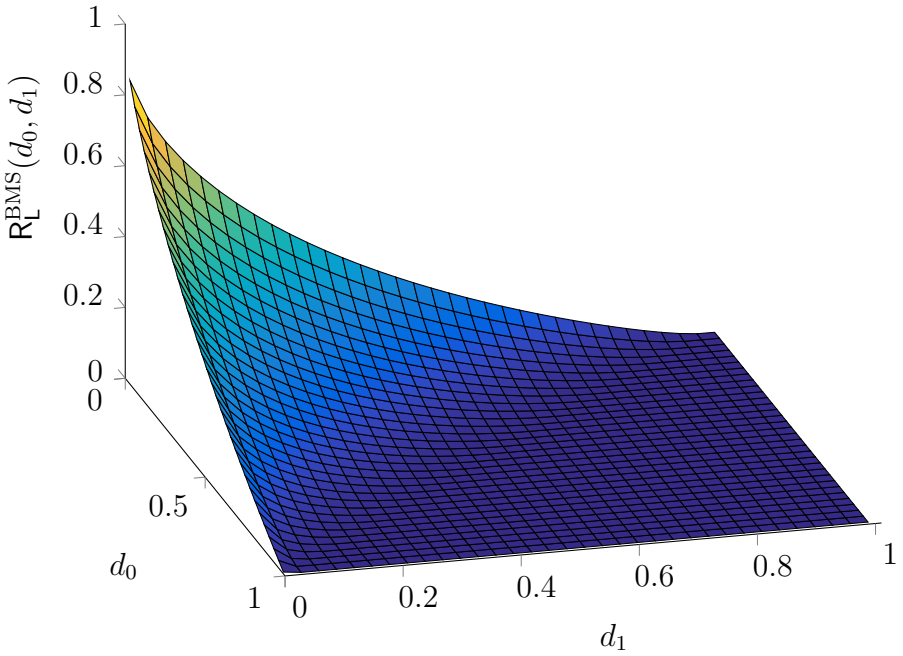
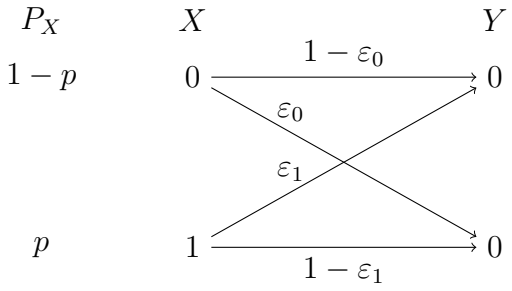


Figure 3.5.: An illustration of $\mathsf{R}_\mathsf{L}^{\mathrm{BMS}}(d_0, d_1)$ with $p = 0.35$ and Hamming distortions.

Proof. We wish to minimize $I(X;Y)$ over all test channels



subject to $0 \leq \varepsilon_\ell \leq \max\{d_\ell, 1\}$ for $\ell = 0, 1$. We can write

$$I(X;Y) = H((1-p)\varepsilon_0 + p(1-\varepsilon_1)) - (1-p)H_2(\varepsilon_0) - pH_2(\varepsilon_1)\,. \tag{3.97}$$

If $d_0 + d_1 \geq 1$, we simply choose $\varepsilon_0 = 1 - \varepsilon_1$ to get $I(X;Y) = 0$. For $d_0 + d_0 < 1$, note that also $\varepsilon_0 + \varepsilon_1 < 1$. Therefore, we have $\varepsilon_0 \leq q \leq 1 - \varepsilon_1$ and $I(X;Y)$ is monotonically decreasing in both ε_0 and ε_1 since

$$\begin{aligned} \frac{\partial I(X;Y)}{\partial \varepsilon_0} &= (1-p)\log\frac{(1-q)\varepsilon_0}{q(1-\varepsilon_0)} > 0 \\ \frac{\partial I(X;Y)}{\partial \varepsilon_1} &= -p\log\frac{(1-q)(1-\varepsilon_1)}{q\varepsilon_1} > 0. \end{aligned} \tag{3.98}$$

The minimum is thus attained by choosing $\varepsilon_0 = d_0$ and $\varepsilon_1 = d_1$ whenever $d_0 + d_1 < 1$. ■

3.3.2. Finite Block Length

Next, we particularize Corollary 3.11 and Theorem 3.10 to the BMS with two Hamming distortion constraints in order to study its finite block length RDL tradeoffs.

For the d-tilted information, we compute

$$\begin{aligned} s_0^\star &= \frac{\partial \mathsf{R}_\mathsf{L}(d_0, d_1)}{\partial d_0} = (1-p)\log\left(\frac{q}{1-q}\frac{1-d_0}{d_0}\right) \\ s_1^\star &= \frac{\partial \mathsf{R}_\mathsf{L}(d_0, d_1)}{\partial d_1} = p\log\left(\frac{1-q}{q}\frac{1-d_1}{d_1}\right) \end{aligned} \tag{3.99}$$

and inserting this into (3.56) yields

$$\begin{aligned} \jmath_\mathsf{X}(\mathsf{x}, \mathsf{d}) = \mathsf{N}(0|\mathsf{x})\log\frac{1-d_0}{1-q} + \mathsf{N}(1|\mathsf{x})\log\frac{1-d_1}{q} \\ - n(1-p)d_0\log\left(\frac{q}{1-q}\frac{1-d_0}{d_0}\right) - npd_1\log\left(\frac{1-q}{q}\frac{1-d_1}{d_1}\right). \end{aligned} \tag{3.100}$$

Note that since $\mathsf{N}(0|\mathsf{x}) = n - \mathsf{N}(1|\mathsf{x})$, the term $\jmath_\mathsf{X}(\mathsf{x}, \mathsf{d})$ is increasing in $\mathsf{N}(1|\mathsf{x})$ if $\frac{1-d_1}{q} > \frac{1-d_0}{1-q}$, it is constant if equality holds, and it is decreasing otherwise. A direct evaluation of

Corollary 3.11 leads to the following result.

Corollary 3.17 (Converse BMS with two Hamming constraints)**.** Fix $p \in (0,1)$ and $\mathsf{d} = (d_0, d_1) \geq 0$ such that $d_0 + d_1 < 1$ and $\frac{1-d_1}{q} > \frac{1-d_0}{1-q}$. An $(n, M, \mathsf{d}, \varepsilon, \Delta)$-code satisfies

$$M \geq \max_{0 \leq b \leq n} \frac{\sum_{k=b}^{n} \binom{n}{k} p^k (1-p)^{n-k} - \varepsilon}{\alpha^{(\cdot)}_{n,\mathsf{d},p}(b)} \tag{3.101}$$

where depending on the choice of the distortion function we have

$$\alpha^{\mathrm{a}}_{n,\mathsf{d},p}(b) = \max_{0 \leq \hat{n}_1 \leq n} \sum_{j=0}^{\lfloor nd_0 \rfloor} \sum_{l=0}^{\lfloor nd_1 \rfloor} \binom{\hat{n}_1}{j} \binom{n-\hat{n}_1}{l} p^{\hat{n}_1+j-l} (1-p)^{n-\hat{n}_1-j+l} \cdot \mathbb{1}_{\{b \leq \hat{n}_1+j-l \leq n\}} \mathbb{1}_{\{j \leq \lfloor (n-\hat{n}_1-j+l)d_0 \rfloor\}} \mathbb{1}_{\{l \leq \lfloor (\hat{n}_1+j-l)d_1 \rfloor\}} \tag{3.102}$$

for Δ^{a} or

$$\alpha^{\mathrm{e}}_{n,\mathsf{d},p}(b) = \max_{0 \leq \hat{n}_1 \leq n} \sum_{j=0}^{\lfloor n(1-p)d_0 \rfloor} \sum_{l=0}^{\lfloor npd_1 \rfloor} \binom{\hat{n}_1}{j} \binom{n-\hat{n}_1}{l} p^{\hat{n}_1+j-l} (1-p)^{n-\hat{n}_1-j+l} \mathbb{1}_{\{b \leq \hat{n}_1+j-l \leq n\}} \tag{3.103}$$

for Δ^{e}.

We would like to compare Corollary 3.17 with a bound derived from the meta-converse. Choosing Q_X as the uniform distribution in (3.60) leads to tight bounds for the BMS with a single average Hamming distortion constraint, see [KV12, Thm. 20]. Making the same choice for two constraints, we get the following numerically simpler converse bound.

Corollary 3.18. Fix $p \in (0,1)$ and $\mathsf{d} = (d_0, d_1)$ such that $d_0 + d_1 < 1$. An $(n, M, \mathsf{d}, \varepsilon, \Delta)$ code satisfies

$$M \geq \frac{\sum_{k=1}^{r^*} \binom{n}{k} + \beta^* \binom{n}{r^*+1}}{\alpha^{(\cdot)}_{n,\mathsf{d},1/2}(0)} \tag{3.104}$$

where $\alpha^{(\cdot)}_{n,\mathsf{d},1/2}(0)$ is given as in Corollary 3.17,

$$r^* := \max \left\{ r : \sum_{k=1}^{r} \binom{n}{k} p^k (1-p)^{n-k} \leq 1 - \varepsilon \right\} \tag{3.105}$$

and $\beta^* \in [0,1)$ is the solution to

$$\sum_{k=1}^{r^*} \binom{n}{k} p^k (1-p)^{n-k} + \beta^* \binom{n}{r^*+1} p^{r^*+1} (1-p)^{n-r^*-1} = 1 - \varepsilon. \tag{3.106}$$

We next apply the random coding bound, Theorem 3.10, to our setting. Here, we choose $P_{\mathsf{Y}} = \prod_{i=1}^{n} P_Y$ to be the product distribution with $P_Y(0) = 1 - P_Y(1) = q$ (which is the RDL achieving reconstruction distribution) for each symbol.

Corollary 3.19 (Random Coding BMS)**.** There exists an $(n, M, \mathsf{d}, \varepsilon, \Delta^{\mathrm{a}})$-code with

$$\varepsilon \leq \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \exp\left(- M \sum_{j=1}^{\lfloor k d_1 \rfloor} \sum_{l=1}^{\lfloor (n-k) d_0 \rfloor} \binom{k}{j} \binom{n-k}{l} q^{k-j+l} (1-q)^{n-k+j-l} \right) \quad (3.107)$$

and an $(n, M, \mathsf{d}, \varepsilon, \Delta^{\mathrm{e}})$-code with

$$\varepsilon \leq \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot \exp\left(- M \sum_{j=1}^{\lfloor n p d_1 \rfloor} \sum_{l=1}^{\lfloor n(1-p) d_0 \rfloor} \binom{k}{j} \binom{n-k}{l} q^{k-j+l} (1-q)^{n-k+j-l} \right). \quad (3.108)$$

Numerical Examples

In Figure 3.6, we evaluate the bounds for the distortion measure Δ^{a} (normalization with $\mathsf{N}(\mathcal{I}_\ell|\mathsf{X})$) for $p = 2/5$, $d_0 = d_1 = 0.11$ and $\varepsilon = 0.01$. Comparing the two converse results, we observe that Corollary 3.17 gives a better bound than Corollary 3.18. The former also seems to vary less for similar block lengths.

A comparison for the distortion measure Δ^{e} (normalization with $n \Pr[X \in \mathcal{I}_\ell]$) with the same parameters is given in Figure 3.7. Observe that the bounds oscillate much more with the block length as compared to using Δ^{a}.

In Figure 3.8, we give an example of a sparse binary source where the zeros are to be reconstructed perfectly and we allow some distortion for the ones. We choose $p = 0.11$, $d_0 = 0$, $d_1 = 0.1$ and $\varepsilon = 0.01$. In this case, we observe that the converse from Corollary 3.18 is useless except for very small block lengths. This is because if we choose Q_Y as the uniform distribution, $\inf_{\mathsf{y} \in \mathcal{Y}} \frac{1}{\mathbb{Q}[\mathsf{y} \in \mathcal{B}_H(\mathsf{X}, \mathsf{d})]}$ is related to the number of elements in the largest distortion ball around some y. This leads to a good bound if all distortion balls are of the same size. In this case, however, the size of the distortion ball greatly varies with the number of ones in X.

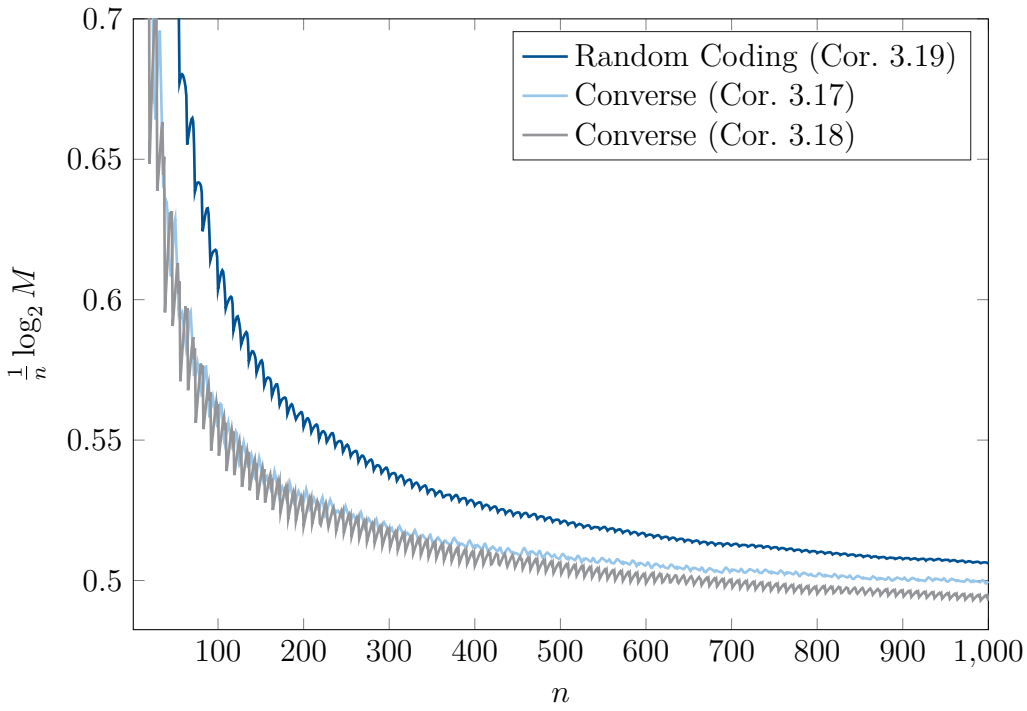


Figure 3.6.: Example with $p = 2/5$, $d_0 = d_1 = 0.11$ and $\varepsilon = 0.01$ for Δ^{a}.

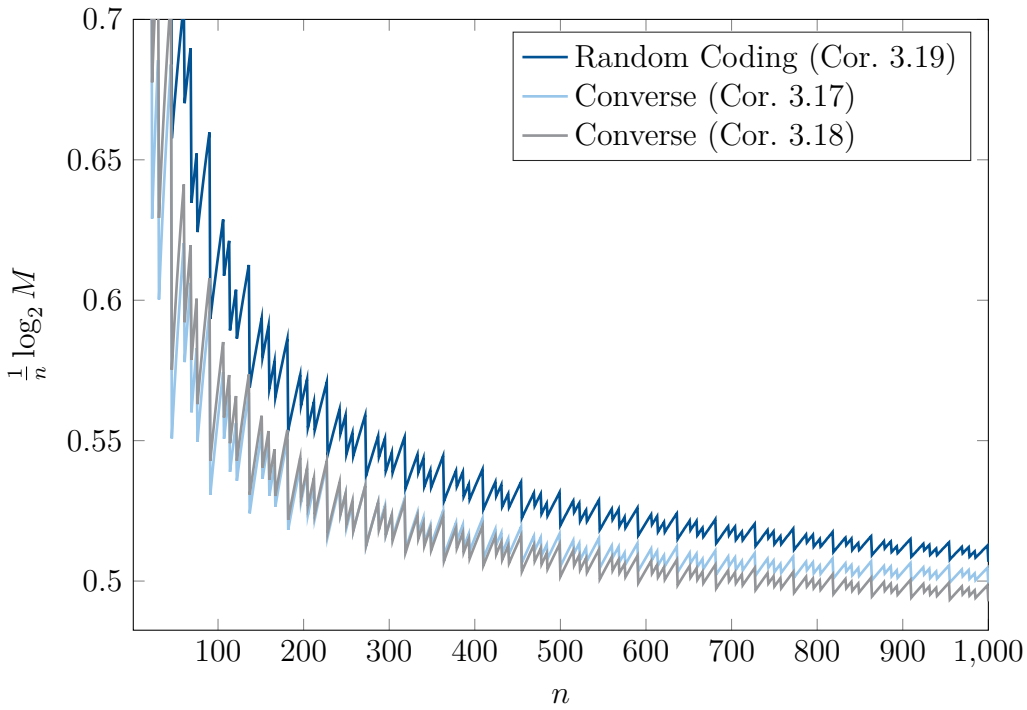


Figure 3.7.: Example with $p = 2/5$, $d_0 = d_1 = 0.11$ and $\varepsilon = 0.01$ for Δ^{e}.

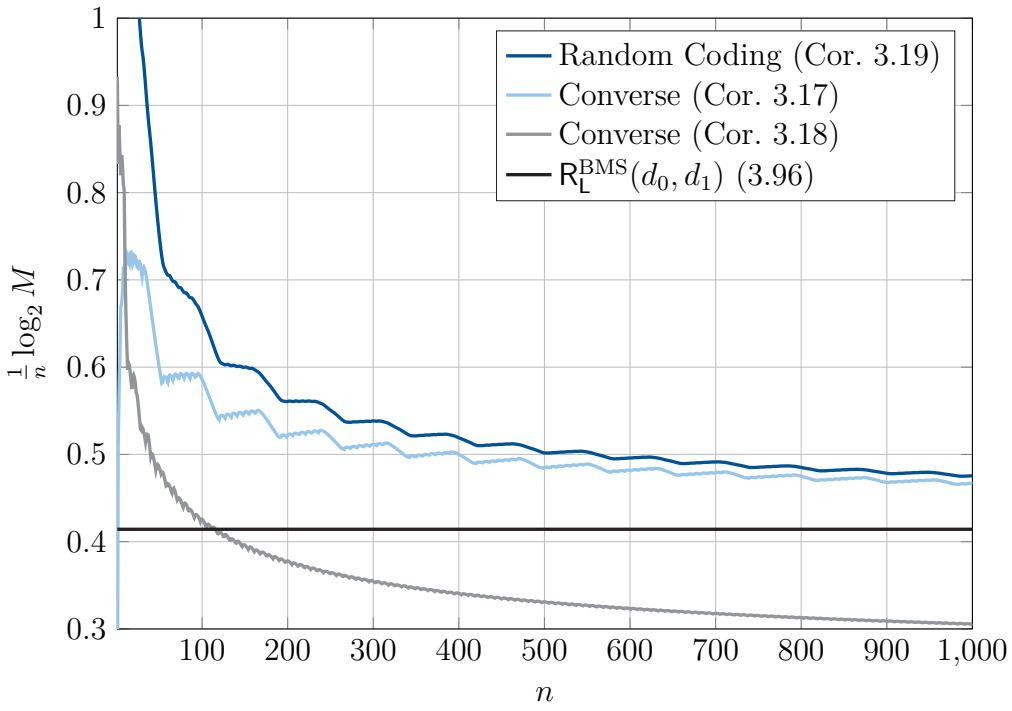


Figure 3.8.: Example with $p = 0.11$, $d_0 = 0$, $d_1 = 0.1$ and $\varepsilon = 0.01$ for Δ^{a}.

4

Bernoulli Spike Sources

In this chapter, we study the (letter-based) RD function of the memoryless *Bernoulli Spike Source (BSS)* in the limit of large block lengths. The BSS emits an iid sequence $X_1, X_2, \ldots$ of real-valued random variables characterized by the relation

$$X = B \cdot Z, \tag{4.1}$$

where B has $\Pr[B = 1] = 1 - \Pr[B = 0] = p$, Z has a PDF P_Z, and B and Z are independent. The probability distribution of X is given by

$$P_X = (1 - p) \cdot \delta_0 + p \cdot P_Z. \tag{4.2}$$

This source model serves as a simple iid model for the sparse sources that are of interest in CS or transform coding [WV12a]. It is often used as a basic probabilistic source model for CS systems, see, e.g., [VS11, WV12b, DJM13, BKM+19]. An example of a string of iid outputs from this source is shown in Figure 4.1.

Previous studies have focused on a single squared error distortion measure. Unfortunately, a closed-form solution of the RD function has, as is the case except for very few

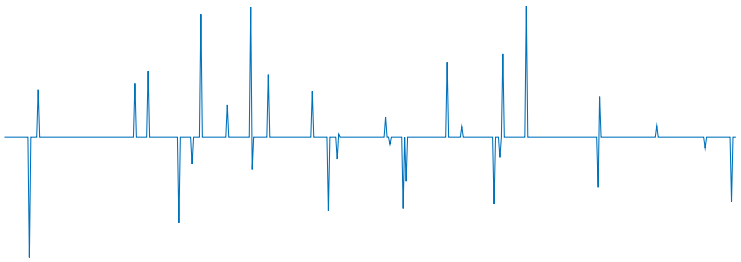


Figure 4.1.: A typical signal sampled from a BSS.

combinations of sources and distortion measures, not been found. For a first insight, two very simple bounds can be made for the RD functions of such sources. Using the *Shannon Lower Bound*, we can lower bound

$$\inf_{\substack{P_{Y|X}:\\ \mathsf{E}\left[(X-Y)^2\right]\leq d}} I(X;Y) \geq \inf_{\substack{P_{Y|X}:\\ \mathsf{E}\left[(X-Y)^2\right]\leq d}} pI(X;Y|\,B=1) \geq p\Big(h(Z)-\frac{1}{2}\log(2\pi ed/p)\Big), \tag{4.3}$$

which corresponds to a coding scheme where a genie provides the value of B as side information. On the other extreme, one could first code the variable B losslessly and then use a good codebook for the continuous part Z of the source. Choosing $Y = Z + N_d$ with $N_d \sim \mathcal{N}(0, d/p)$ independent of Z, we get the upper bound

$$\inf_{\substack{P_{Y|X}\\ \mathsf{E}\left[(X-Y)^2\right]\leq d}} I(X;Y) \leq H(B) + p\Big(h(Z+N_d)-\frac{1}{2}\log(2\pi ed/p)\Big). \tag{4.4}$$

Comparing (4.3) and (4.4), we notice a gap of at least $H(B)$ which can be quite substantial for small p. One might, of course, employ numerical methods such as the Blahut-Arimoto algorithm to compute the rate distortion function for a specific setting. These methods, however, provide little insight into the general RD behavior of sparse signal sources and good coding schemes for them. Precisely this is the goal of previous and our information-theoretic studies via upper and lower bounds as well as asymptotic considerations.

Rosenthal and Binia [RB88] and later György, Linder and Zeger [GLZ99] studied the RD function of such mixed discrete-continuous sources under a squared-error fidelity criterion. Particularized to the BSS model (4.1), their results provide an asymptotically exact ($d \downarrow 0$) approximation $\mathsf{R}_0(d)$ of the RD function

$$\lim_{d\downarrow 0}\Big\{\mathsf{R}_0(d)-\mathsf{R}(d)\Big\} = 0 \tag{4.5}$$

where

$$\mathsf{R}_0(d) \coloneqq H_2(p) + p\Big(h(Z)-\frac{1}{2}\log(2\pi ed/p\Big) \tag{4.6}$$

provided that the continuous random variable Z has finite second moment.

Weidmann and Vetterli [WV12a] studied several classes of *sparse* and *compressible* sources. In particular, they derived an upper bound for the RD function of spike sources that seems to be close to optimal for both high and small distortions. The idea is to distinguish between *significant samples* (i.e., those with a high magnitude) and *insignificant samples*. The insignificant samples are set to zero while the positions of the significant samples are stored and their amplitudes are coded with a Gaussian codebook. This yields the following upper bound.

Theorem 4.1 (Weidmann & Vetterli [WV12a])**.** For a squared error distortion constraint, the RD function of a BSS is upper bounded by

$$R(d) \leq \inf_{\tau \geq 0} \left(H_2(\Pr[|X| > \tau]) + \frac{\Pr[|X| > \tau]}{2} \log \frac{\mathsf{E}\left[X^2 \mathbb{1}_{\{|X|>\tau\}}\right]}{d - \mathsf{E}\left[X^2 \mathbb{1}_{\{|X|\leq\tau\}}\right]} \right). \tag{4.7}$$

It is worth noting that this coding scheme always reconstructs the zero elements perfectly, i.e., all the incurred distortion is on the nonzero elements of X.

Chang [Cha10] derived a lower bound for the RD function of a BSS with a Gaussian Z using a channel coding argument. For small values of d, this bound exhibits a constant gap of $-(1-p)\log(1-p)$ to the best upper bound.

In this chapter, we first consider the RDL function of BSSs with two constraints: Hamming distortion for the *zeros*, squared error distortion for the *nonzeros*, and derive upper and lower bounds for the RDL function.
Further, we leverage the bounds for the RDL function to derive a new lower bound for a single squared error distortion measure. We then show that this bound is asymptotically tight in the small distortion regime, thereby closing the gap left by Chang [Cha10] and, using a result by Koch [Koc16], extending the tightness result (4.5) to the broadest possible class of random variables Z, namely those whose discrete entropy of the integer part is finite.

The work presented in this chapter is based on joint work with Roy Timo and parts of it are published in [PT16c].

4.1. Converse for Two Distortions

Motivated by the previous discussion, we wish to investigate the RD behavior of a BSS with separate distortion constraints for the zeros and the nonzero samples:

▷ Hamming distortion constraint of the *zero event*:

$$\Delta_0^{\mathrm{a}}(\mathsf{x}, \mathsf{y}) = \frac{1}{\mathsf{N}(\{0\}|\mathsf{x})} \sum_{i=1}^{n} \mathbb{1}_{\{x[i]=0\}} \mathbb{1}_{\{y[i]\neq 0\}} \tag{4.8}$$

▷ MSE distortion of the *nonzero event*:

$$\Delta_{\mathsf{S}}^{\mathrm{a}}(\mathsf{x}, \mathsf{y}) = \frac{1}{\mathsf{N}(\{0\}^{\mathrm{c}}|\mathsf{x})} \sum_{i=1}^{n} \mathbb{1}_{\{x[i]\neq 0\}} (x[i] - y[i])^2 \tag{4.9}$$

Thus, the asymptotic letter-based RD function (3.6) is given by

$$\mathsf{R}_{\mathsf{L}}^{\mathrm{BSS}}(d_0, d_{\mathsf{S}}) = \min_{P_{Y|X}} I(X;Y) \qquad \text{subject to} \quad \Pr[Y \neq 0 \,|\, X = 0] \leq d_0 \tag{4.10a}$$

$$\mathsf{E}[(X-Y)^2 \,|\, X \neq 0] \leq d_{\mathsf{S}}. \tag{4.10b}$$

Since Theorem 4.1 provides a coding scheme that has numerically been demonstrated to be very close to the RD function and achieves $\mathsf{E}[\Delta_0^{\mathrm{a}}(\mathsf{X}, \mathsf{Y})] = 0$, we focus on deriving a converse result.

We prove the following lower bound.

Theorem 4.2. For all $0 \leq d_0 \leq 1$ and $d_{\mathsf{S}} \geq 0$, we have

$$\mathsf{R}_{\mathsf{L}}^{\mathrm{BSS}}(d_0, d_{\mathsf{S}}) \geq \mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}(d_0, d_1) + p\Big(h(Z) - \frac{1}{2}\log(2\pi e d_{\mathsf{S}})\Big) \tag{4.11}$$

where d_1 is given by Lemma 4.3 and $\mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}(d_0, d_1)$ is given by (3.96).

Proof. We wish to minimize $I(X;Y)$ over all conditional distributions $P_{Y|X}$ from $\mathcal{X}$ to $\mathcal{Y}$ satisfying (4.10a) and (4.10b). Since B can be computed from X (with probability one), $B - X - Y$ forms a Markov Chain. We write

$$I(X;Y) = I(B,X;Y) = I(B;Y) + pI(X;Y|\,B=1) \tag{4.12}$$

since $I(X;Y|\,B=0) = 0$. Now consider

$$\begin{aligned} I(X;Y|\,B=1) &= h(X|\,B=1) - h(X|\,Y,B=1) \\ &\overset{\mathrm{a}}{\geq} h(Z) - h(X-Y|\,B=1) \\ &\overset{\mathrm{b}}{\geq} h(Z) - \frac{1}{2}\log 2\pi e \mathsf{E}[(Y-Z)^2 \,|\, B=1] \\ &\overset{\mathrm{c}}{\geq} h(Z) - \frac{1}{2}\log 2\pi e d_{\mathsf{S}} \end{aligned} \tag{4.13}$$

where (a) follows from the translation invariance of differential entropy [CT06b, Thm. 8.6.3] and because conditioning does not increase entropy [CT06b, Thm 8.6.1], (b) follows since given $B = 1$, $(X - Y)$ has finite second moment due to the distortion constraint and the maximum entropy property of Gaussian random variables [CT06b, Thm 17.2.3], and (c) uses the distortion constraint (4.10b) again.

To complete the proof, we need to show that $I(B;Y)$ is lower bounded by $\mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}(d_0, d_1)$ for every $P_{Y|X}$ satisfying (4.10a) – (4.10b). To this end, we determine equivalent Hamming distortions d_0 and d_1 such that $I(B;Y)$ can be compared to the RDL function of the binary memoryless source B.

Denote $\hat{B} = \mathbb{1}_{\{Y \neq 0\}}$. From (4.10a), we have

$$d_0 \geq \Pr[Y \neq 0 \,|\, X = 0] = \Pr[\hat{B} = 1 \,|\, B = 0]. \tag{4.14}$$

Now consider (4.10b):

$$\begin{aligned} d_{\mathsf{S}} &\geq \mathsf{E}[(Y-X)^2 \,|\, B=1] \\ &\geq \mathsf{E}[(Y-X)^2 \mathbb{1}_{\{\hat{B}=0\}} \,|\, B=1] \\ &\geq \mathsf{E}\Big[X^2 \mathbb{1}_{\{\hat{B}=0\}} \Big| B=1\Big]. \end{aligned} \tag{4.15}$$

Next, denote the set of conditional distributions satisfying (4.15) by

$$\mathcal{Q}(d_{\mathsf{S}}) \coloneqq \Big\{ P_{Y|X} : \mathsf{E}\Big[X^2 \mathbb{1}_{\{\hat{B}=0\}} \Big| B=1\Big] \leq d_{\mathsf{S}} \Big\} \tag{4.16}$$

and define

$$d_1 \coloneqq \sup_{P_{Y|X} \in \mathcal{Q}(d_{\mathsf{S}})} \Pr[\hat{B}=0 \,|\, B=1]. \tag{4.17}$$

By (4.14) and (4.17), the distribution $P_{Y|X}$ satisfies

$$I(B;Y) \geq I\Big(B;\hat{B}\Big) \geq \min_{\substack{P_{\hat{B}|B}: \\ \mathsf{E}[\mathbb{1}_{\{\hat{B}=1\}} \,|\, B=0] \leq d_0 \\ \mathsf{E}[\mathbb{1}_{\{\hat{B}=0\}} \,|\, B=1] \leq d_1}} I\Big(B;\hat{B}\Big) = \mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}(d_0,d_1)\,. \tag{4.18}$$

To complete the proof, it remains to determine d_1. This is given by the following lemma.

Lemma 4.3. Let Q_{Z^2} be the *quantile function* of Z^2, i.e.,

$$Q_{Z^2}(q) \coloneqq \inf\left\{ w \in \mathbb{R} : q \leq F_{Z^2}(w) \right\}. \tag{4.19}$$

d_1 is the solution to

$$\mathsf{E}\Big[Z^2 \mathbb{1}_{\{Z^2 \leq Q_{Z^2}(d_1)\}}\Big] = d_{\mathsf{S}}. \tag{4.20}$$

Denote the cumulative distribution function of the chi-squared distribution with k degrees of freedom by $F_{\chi_k^2}$ and its inverse by $F_{\chi_k^2}^{-1}$. If Z is a standard Gaussian random variable, then

$$d_1 = F_{\chi_3^2}\Big(F_{\chi_1^2}^{-1}(d_{\mathsf{S}})\Big). \tag{4.21}$$

Lemma 4.3 basically quantifies how much of X can be classified as *insignificant samples* without immediately violating the distortion constraint for the spikes. The proof of Lemma 4.3 is given in Appendix A.1. ■

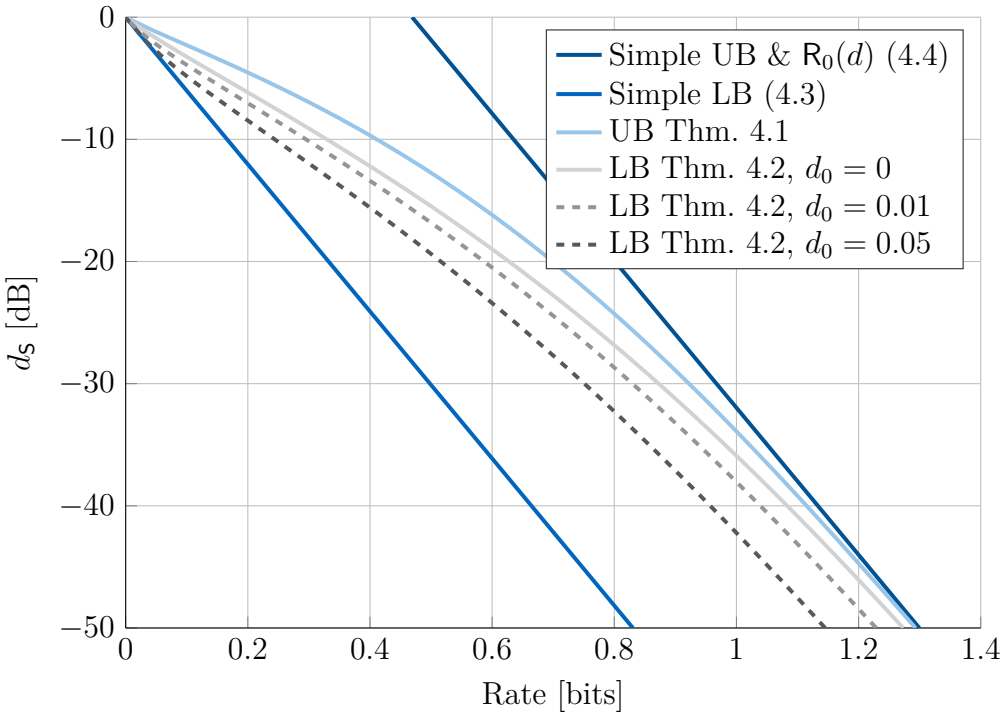


Figure 4.2.: Comparison of upper and lower bounds for $p = 0.1$.

Numerical Example

Figure 4.2 shows an evaluation of the different bounds with $p = 0.1$ for $Z \sim \mathcal{N}(0,1)$. In this case, the upper bound (4.4) can easily be improved to $H_2(p) + \frac{p}{2}\log(1/d_{\mathsf{S}})$ so that it exhibits a constant gap of $H_2(p)$ to the lower bound. The straight lines represent the simple upper and lower bound discussed in (4.3) and (4.4), which leave a gap of $H_2(p) \approx 0.47$ bits that is significant even at low distortions. We see that for $d_0 = 0$, our lower bound from Theorem 4.2 is very close to the upper bound (4.7). An interesting question is whether one can find an achievability scheme that benefits from having nonzero distortion on the zeros.

4.2. Converse for Squared Error Distortions

In this section, we will leverage the results for the RDL function to derive a general converse result for spike sources with just one MSE distortion constraint. The key idea is to split the single squared error distortion constraint into two separate constraints - one for the zeros and one for the nonzeros. We can express the RD function as

$$\mathsf{R}(d) = \inf_{\substack{P_{Y|X}:\\ \mathsf{E}\left[(X-Y)^2\right] \le d}} I(X;Y) = \min_{\substack{d_0', d_{\mathsf{S}}':\\ (1-p)d_0' + p d_{\mathsf{S}}' \le d}} \mathsf{R}_{\mathsf{L}}^{\mathrm{BSS,mse}}(d_0', d_{\mathsf{S}}') \tag{4.22}$$

where we defined the RDL function with two squared error distortion constraints as

$$\mathsf{R}_{\mathsf{L}}^{\mathrm{BSS,mse}}(d_0, d_{\mathsf{S}}) = \min_{P_{Y|X}} I(X;Y) \quad \text{subject to} \quad \begin{aligned} &\mathsf{E}[(X-Y)^2 \,|\, X=0] \le d_0 && \text{(4.23a)}\\ &\mathsf{E}[(X-Y)^2 \,|\, X\neq 0] \le d_{\mathsf{S}}. && \text{(4.23b)}\end{aligned}$$

Introducing these new constraints adds a certain structure we can exploit to investigate the zeros and nonzeros separately in a way similar to Theorem 4.2. Before stating the converse bound for $\mathsf{R}(d)$, we need to introduce a few definitions.

For some $\gamma > 0$ with $\Pr[|Z| > \gamma] > 0$, let

$$W \coloneqq (|Z| - \gamma)^2 \mathbb{1}_{\{|Z| \ge \gamma\}}. \tag{4.24}$$

We are interested in the cumulative distribution function of W conditioned on the event $\mathcal{Z} \coloneqq \{|Z| > \gamma\}$, which is given by

$$F_{W|\mathcal{Z}}(w) = \begin{cases} \dfrac{\Pr\big[|Z| \in (\gamma, \gamma + \sqrt{w}]\big]}{\Pr[|Z| > \gamma]}, & \text{if } w > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.25}$$

and we denote the *quantile function* of W conditioned on $\mathcal{Z}$ by $Q_{W|\mathcal{Z}} : [0,1] \to \mathbb{R}$:

$$Q_{W|\mathcal{Z}}(q) \coloneqq \inf\Big\{w \in \mathbb{R} : q \le F_{W|\mathcal{Z}}(w)\Big\}. \tag{4.26}$$

Further, let

$$g_{W|\mathcal{Z}}(q) \coloneqq \mathsf{E}\Big[W \mathbb{1}_{\{W \le Q_{W|\mathcal{Z}}(q)\}} \Big| \mathcal{Z}, B = 1\Big] \tag{4.27}$$

and denote by $g^-_{W|\mathcal{Z}}$ the inverse of $g_{W|\mathcal{Z}}$. Finally, define $q_\gamma : [0,\infty) \to [0,1]$ to be the (strictly increasing) function given by

$$q_\gamma(d_{\mathsf{S}}) \coloneqq \Pr[\mathcal{Z}, B=1] g^-_{W|\mathcal{Z}}\Big(\frac{d_{\mathsf{S}}}{\Pr[\mathcal{Z}, B=1]}\Big) + \Pr[\mathcal{Z}^{\mathrm{c}}, B=1]. \tag{4.28}$$

We can now state our converse result.

Theorem 4.4. We have $\mathsf{R}(d) \ge \mathsf{R}_{\mathrm{LB}}^{\mathrm{BSS,mse}}(d)$ for all $d > 0$, where

$$\mathsf{R}_{\mathrm{LB}}^{\mathrm{BSS,mse}}(d) \coloneqq \sup_{\gamma \ge 0} \min_{\substack{d_0, d_{\mathsf{S}}: \\ (1-p)d_0 + p d_{\mathsf{S}} \le d}} \Big(\mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}\Big(d_0/\gamma^2, q_\gamma(d_{\mathsf{S}})\Big) + p\Big(h(Z) - \frac{1}{2}\log(2\pi e d_{\mathsf{S}})\Big)\Big). \tag{4.29}$$

The proof of Theorem 4.4 uses similar ideas to the proof of Theorem 4.2, but keeps track of a few more details. We thus defer it to Appendix A.2.

Remark 4.1. One might ask about the usefulness of bounds such as Theorem 4.4, given that numerical methods [Bla72, Ari72] are available that can compute the RD function pre-

cisely. Our answer to this question is threefold. First, such bounds give analytical insight into the problem. We will, e.g., use Theorem 4.4 to show that the approximation (4.6) is asymptotically accurate (as $d \downarrow 0$) for all variables Z for which the *Shannon lower bound* is tight (see [Koc16]). Second, we can leverage this bound in situations in which numerical methods do not exist or are computationally costly, such as the distributed source coding setting discussed in Section 5. Here, there is no known general expression for the optimum coding rates which means that one must resort to upper and lower bounds. Theorem 4.4 can be applied in this case, too. Third, bounds can be useful if they are computationally simpler than existing numerical methods. While Theorem 4.4 itself is not numerically very simple to compute, we remark that it can be relaxed by choosing a specific γ (choosing $\gamma \sim d^{1/3}$ works well for a Gaussian Z) or by relaxing it to Corollary 4.5, which is stated below.

Observe that from $(1-p)d_0+pd_{\mathsf{S}} \leq d$, we may upper bound $d_0 \leq d/(1-p)$ and $d_{\mathsf{S}} \leq d/p$. Since $\mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}$ is non-increasing in d_0 and d_1 and q_γ is strictly increasing, we can immediately relax Theorem 4.4 to the following simpler bound.

Corollary 4.5. For all $d > 0$, we have $\mathsf{R}(d) \geq \mathsf{R}_{\mathrm{LB}}^{\mathrm{BSS,mse}}(d) \geq \tilde{\mathsf{R}}_{\mathrm{LB}}^{\mathrm{BSS,mse}}(d)$, where

$$\tilde{\mathsf{R}}_{\mathrm{LB}}^{\mathrm{BSS,mse}}(d) \coloneqq \sup_{\gamma>0} \mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}\left(\frac{d}{(1-p)\gamma^2}, q_\gamma(d/p)\right) + p\Big(h(Z) - \frac{1}{2}\log(2\pi e d/p)\Big). \tag{4.30}$$

The next result shows that the approximation (4.6) is asymptotically accurate in the small distortion regime.

Theorem 4.6. Let Z be such that $|h(Z)| < \infty$ and $H(\lfloor Z \rfloor) < \infty$. Then, $\mathsf{R}_0(d)$ is asymptotically accurate, that is

$$\lim_{d \downarrow 0} \{\mathsf{R}_0(d) - \mathsf{R}(d)\} = 0. \tag{4.31}$$

Proof of Theorem 4.6. Recall that by Corollary 4.5 and (4.4), we have

$$\begin{aligned}\sup_{\gamma>0} \mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}\left(\frac{d}{(1-p)\gamma^2}, q_\gamma(d/p)\right) + p\Big(h(Z) - \frac{1}{2}\log(2\pi e d/p)\Big) \\ \leq \mathsf{R}(d) \leq H(B) + p\Big(h(Z+N_d) - \frac{1}{2}\log(2\pi e d/p)\Big)\end{aligned} \tag{4.32}$$

where $N_d \sim \mathcal{N}(0, d/p)$ is independent of Z. In [Koc16, Theorem 2], it is shown that

$$\lim_{d \downarrow 0} \big(h(Z) - h(Z+N_d)\big) = 0 \tag{4.33}$$

whenever $|h(Z)| < \infty$ and $H(\lfloor Z \rfloor) < \infty$. Thus, we only need to show that the first term in (4.30) converges to $H_2(p)$ as $d \to 0$. This technical part is done in Appendix A.3. ∎

Numerical Example

Figure 4.3 shows Theorem 4.4 and Corollary 4.5 in comparison with the simple upper and lower bounds (4.3) and (4.4) as well as the threshold coding scheme (4.7) from [WV12a] for $Z \sim \mathcal{N}(0,1)$ and $p = 0.1$. Since Z is Gaussian, the upper bound (4.4) is in fact an achievable rate and equal to $\mathsf{R}_0(d)$. We clearly observe that the gap between the lower bounds and $\mathsf{R}_0(d)$ vanishes as $d \downarrow 0$ as guaranteed by Theorem 4.6.

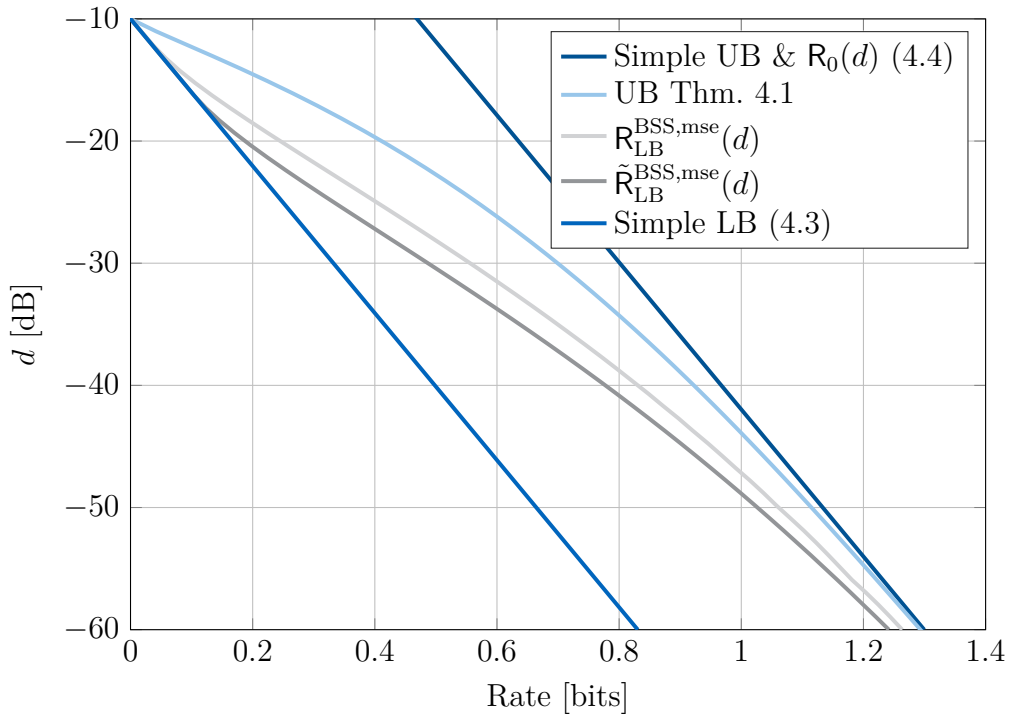


Figure 4.3.: Bernoulli-Gaussian Spike Source with $p = 0.1$.

5

Distributed Bernoulli-Gaussian Spike Source

In this chapter, we derive RD bounds for *distributed source coding* for sparse sources. In this setting, multiple sources are observed and encoded *separately* and decoded *jointly*. Determining the optimal coding rates in lossy distributed source coding has been an open problem for a long time and it is not clear whether a general single-letter solution exists. Still, previous efforts have succeeded in solving several important special cases and deriving inner and outer bounds for the optimal coding rates. Below, we list some important previous works on distributed source coding.

- ▷ Berger [Ber78] and Tung [Tun78] derived general inner and upper bounds. They are, however, known to be different in general (see [GK11, Ch. 12]).

- ▷ For the case of finite-alphabet sources and $d_1 = d_2 = 0$, i.e., the sources are encoded losslessly, the coding rates were determined by Slepian and Wolf [SW73].

- ▷ Berger and Yeung [BY89] solved the case where the source alphabets are finite, $d_1 = 0$, and d_2 is arbitrary.

- ▷ Oohama [Ooh97] solved the case of correlated Gaussian sources with squared error distortion where only one of the two sources is reconstructed with a target distortion.

- ▷ Zamir and Berger [ZB99] considered sources with a PDF in the low distortion limit $d_1, d_2 \to 0$. In particular, they derived a multi-terminal extension of the Shannon lower bound and showed that this is tight in the low distortion limit.

- ▷ Wagner, Tavildar and Viswanath [WTV08] solved the case of correlated Gaussian sources with two squared error distortion constraints.

- ▷ Wagner, Kelly and Altuğ [WKA11] showed that the Berger-Tung inner bound is suboptimal in general and derived an improved inner bound that incorporates *common components*.
- ▷ Courtade and Weissman [CW14] solved the case of discrete sources with logarithmic loss.

5.1. System Model

We consider *Distributed Bernoulli-Gaussian Sources (DBGSs)* where the support is common to both signals. To be more precise, consider the model

$$\begin{aligned} X_1 &= B \cdot Z_1 \\ X_2 &= B \cdot Z_2 \end{aligned} \tag{5.1}$$

where B has $\Pr[B = 1] = 1 - \Pr[B = 0] = p$ and (Z_1, Z_2) are jointly Gaussian with zero mean, variances one and correlation coefficient ρ.

We assume that a distributed source emits iid copies $\{X_1[i], X_2[i]\}_{i=1}^n$ that are encoded *separately* by the encoders

$$f_1 : \mathbb{R}^n \to \left\{1, \dots, 2^{n\mathsf{R}_1}\right\}, \qquad f_2 : \mathbb{R}^n \to \left\{1, \dots, 2^{n\mathsf{R}_2}\right\} \tag{5.2}$$

and decoded *jointly* by the decoders

$$\begin{aligned} g_1 &: \left\{1, \dots, 2^{n\mathsf{R}_1}\right\} \times \left\{1, \dots, 2^{n\mathsf{R}_2}\right\} \to \mathbb{R}^n \\ g_2 &: \left\{1, \dots, 2^{n\mathsf{R}_1}\right\} \times \left\{1, \dots, 2^{n\mathsf{R}_2}\right\} \to \mathbb{R}^n. \end{aligned} \tag{5.3}$$

We further require the reconstructions $(\mathsf{Y}_1, \mathsf{Y}_2)$ to satisfy the distortion constraints

$$\mathsf{E}\Big[\Delta(\mathsf{X}_1, \mathsf{Y}_1)\Big] \le d_1, \qquad \mathsf{E}\Big[\Delta(\mathsf{X}_2, \mathsf{Y}_2)\Big] \le d_2. \tag{5.4}$$

This system model is depicted in Figure 5.1.

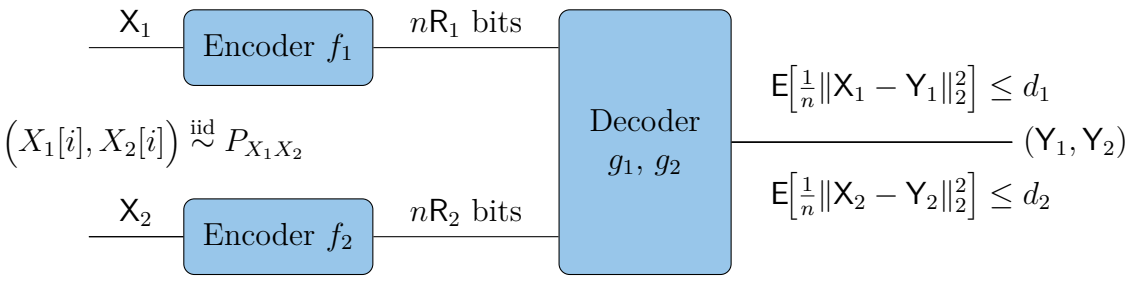


Figure 5.1.: System model for distributed source coding.

We say that a RD quadruple $(\mathsf{R}_1, \mathsf{R}_2, d_1, d_2)$ is *achievable* if there exists a sequence of coding schemes $(f_1^{(n)}, f_2^{(n)}, g_1^{(n)}, g_2^{(n)})$ related to $(\mathsf{R}_1, \mathsf{R}_2)$ as given in (5.2) – (5.3) that satisfies

$$\limsup_{n\to\infty} \frac{1}{n}\mathsf{E}\Big[\|\mathsf{X}_1 - \mathsf{Y}_1\|_2^2\Big] \le d_1 \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n}\mathsf{E}\Big[\|\mathsf{X}_2 - \mathsf{Y}_2\|_2^2\Big] \le d_2. \tag{5.5}$$

As is standard in the literature [GK11], we call the closure of all $(\mathsf{R}_1, \mathsf{R}_2)$ such that $(\mathsf{R}_1, \mathsf{R}_2, d_1, d_2)$ is achievable the *rate region* and denote it by $\mathcal{R}(d_1, d_2)$. Since the exact determination of $\mathcal{R}(d_1, d_2)$ seems out of reach, our main goal of this chapter is to find good *inner* and *outer* bounds for the DBGS.

Inner bounds for $\mathcal{R}(d_1, d_2)$ are derived in Section 5.2 and outer bounds in Section 5.3. Section 5.4 presents numerical evaluations of the bounds in a few example settings.

5.2. Inner Bounds

This section derives inner, i.e., achievability, bounds for the DBGS based on the *Quantize-and-Bin* coding scheme developed by Berger and Tung inner bound. Let us first restate this bound.

Theorem 5.1 (Berger-Tung Inner Bound [Ber78, Tun78])**.** The rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ is achievable with distortion pair (d_1, d_2) if

$$\begin{aligned} \mathsf{R}_1 &\ge I(X_1; U_1 | U_2) \\ \mathsf{R}_2 &\ge I(X_2; U_2 | U_1) \\ \mathsf{R}_1 + \mathsf{R}_2 &\ge I(X_1, X_2; U_1, U_2) \end{aligned} \tag{5.6}$$

for $U_1 - X_1 - X_2 - U_2$ forming a Markov chain and decoding functions g_1 and g_2 satisfying $\mathsf{E}[\Delta(X_j, g_j(U_1, U_2))] \le d_j$ for $j = 1, 2$.

5.2.1. A Simple Inner Bound

We first derive a simple inner bound by first coding B and then optimally coding the two Gaussian random variables Z_1, Z_2. This inner bound extends the coding scheme for a single BSS that led to the bound (4.4) to the distributed setting.

Theorem 5.2. The rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ is an achievable rate pair for distortions (d_1, d_2) if

$$\mathsf{R}_1 \ge \frac{p}{2}\log^+ \frac{1 - \rho^2 + \rho^2 2^{-2[\mathsf{R}_2 - H_2(p)]_+/p}}{d_1/p} \tag{5.7}$$

$$\mathsf{R}_2 \ge \frac{p}{2}\log^+ \frac{1 - \rho^2 + \rho^2 2^{-2[\mathsf{R}_1 - H_2(p)]_+/p}}{d_2/p} \tag{5.8}$$

$$\mathsf{R}_1 + \mathsf{R}_2 \ge H_2(p) + \frac{p}{2}\log^+ \frac{(1-\rho^2)\beta(d_1 d_2/p^2)}{2 d_1 d_2/p^2} \tag{5.9}$$

where

$$\beta(d) \coloneqq 1 + \sqrt{1 + \frac{4\rho^2 d}{(1-\rho^2)^2}} \tag{5.10}$$

and $\log^+(x) \coloneqq \max\{0, \log(x)\}$.

Proof. We choose

$$\begin{aligned} U_1 &= (X_1 + N_1)\mathbb{1}_{\{X_1 \neq 0\}}, & N_1 &\sim \mathcal{N}(0, \sigma_1^2) \\ U_2 &= (X_2 + N_2)\mathbb{1}_{\{X_2 \neq 0\}}, & N_2 &\sim \mathcal{N}(0, \sigma_2^2) \\ g_1 &= \mathsf{E}[X_1 \,|\, U_1, U_2] \\ g_2 &= \mathsf{E}[X_2 \,|\, U_1, U_2] \end{aligned} \tag{5.11}$$

and note that $\mathbb{1}_{\{X_1 \neq 0\}} = \mathbb{1}_{\{X_2 \neq 0\}} = B$ with probability one. That is, (U_1, U_2) is zero whenever (X_1, X_2) are zero and otherwise we use a *distributed Gaussian test channel* which is known to be optimal for distributed jointly Gaussian sources [WTV08]. Evaluating the Berger-Tung bound for this choice yields

$$\begin{aligned} \mathsf{R}_1 &\geq I(X_1; U_1 \,|\, U_2) \\ &= I(X_1; U_1 \,|\, B, U_2) \\ &= pI(X_1; U_1 \,|\, U_2, B = 1) \\ &= pI(Z_1; Z_1 + N_1 \,|\, Z_2 + N_2, B = 1)\,. \end{aligned} \tag{5.12}$$

Similarly, we have

$$\mathsf{R}_2 \geq pI(Z_2; Z_2 + N_2 \,|\, Z_1 + N_1, B = 1)\,. \tag{5.13}$$

For the sum rate, we compute

$$\begin{aligned} \mathsf{R}_1 + \mathsf{R}_2 &\geq I(X_1, X_2; U_1, U_2) \\ &= I(B, X_1, X_2; U_1, U_2) \\ &= I(B; U_1, U_2) + pI(X_1, X_2; U_1, U_2 \,|\, B = 1) \\ &= H_2(p) + p \cdot I(Z_1, Z_2; Z_1 + N_1, Z_2 + N_2 \,|\, B = 1)\,. \end{aligned} \tag{5.14}$$

The mutual information terms in (5.12), (5.13) and (5.14) are similar to the rates for quadratic Gaussian distributed source coding (see the discussion in [GK11, Sec. 12.3]). To derive the rate region, one can follow along the same lines with a few differences.

▷ The distortion constraints are now d_1/p and d_2/p for users one and two since we have

$$\begin{aligned} d_1 &\geq \mathsf{E}\Big[(X_1 - \mathsf{E}[X_1 \,|\, U_1, U_2])^2\Big] = p \cdot \mathsf{E}\Big[(X_1 - \mathsf{E}[X_1 \,|\, U_1, U_2])^2 \Big| B = 1\Big] \\ d_2 &\geq \mathsf{E}\Big[(X_2 - \mathsf{E}[X_2 \,|\, U_1, U_2])^2\Big] = p \cdot \mathsf{E}\Big[(X_2 - \mathsf{E}[X_2 \,|\, U_1, U_2])^2 \Big| B = 1\Big]. \end{aligned} \tag{5.15}$$

▷ We must pay attention to the corner points $(I(X_1;U_1|\,U_2)\,, I(X_2;U_2))$ and $(I(X_1;U_1)\,, I(X_2;U_2|\,U_1))$ because we have

$$\begin{aligned} \mathsf{R}_1 \geq I(X_1;U_1) &= I(X_1,B;U_1) \\ &= I(B;U_1) + p \cdot I(X_1;U_1|\,B=1) \\ &= H_2(p) + p \cdot I(X_1;U_1|\,B=1)\,, \end{aligned} \tag{5.16}$$

which means that $I(X_1;U_1|\,B=1) \leq \frac{[\mathsf{R}_1 - H_2(p)]_+}{p}$, and similarly for R_2.

Continuing the derivations in [GK11, Sec. 12.3.1] using (5.7) - (5.16) completes the proof. ■

5.2.2. A Thresholding Based Inner Bound

We improve Theorem 5.2 by extending the scheme of Weidmann and Vetterli [WV12a, Thm. 5] that is restated above as Theorem 4.1. This scheme is based on distinguishing between *significant* and *insignificant* samples. The insignificant samples with small magnitude are quantized to zero whereas the significant samples are quantized with a Gaussian codebook.

To apply this to the distributed setting, we choose thresholds $\tau_1, \tau_2 > 0$ and let

$$\begin{aligned} U_1 &= (X_1 + N_1)\mathbb{1}_{\{|X_1|>\tau_1\}}, \qquad N_1 \sim \mathcal{N}(0,\sigma_1^2), \\ U_2 &= (X_2 + N_2)\mathbb{1}_{\{|X_2|>\tau_2\}}, \qquad N_2 \sim \mathcal{N}(0,\sigma_2^2). \end{aligned} \tag{5.17}$$

Note that with this scheme, we can have $U_1 = 0$ but $U_2 \neq 0$ and vice versa, as shown in Figure 5.2.

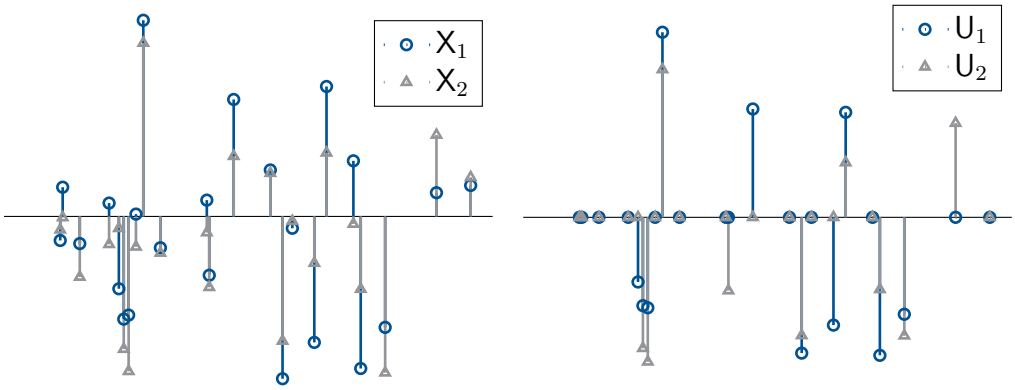


Figure 5.2.: Example of signals X_1, X_2 and auxiliary variables U_1, U_2.

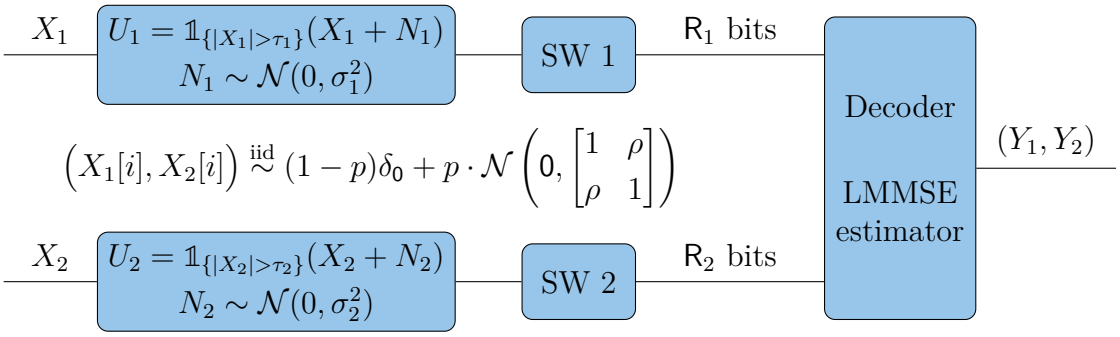


Figure 5.3.: The thresholding based distributed coding scheme.

We define the following events

$$\begin{aligned} \mathcal{U}_{00} &\coloneqq \{U_1 = 0, U_2 = 0\}, & \mathcal{U}_{01} &\coloneqq \{U_1 = 0, U_2 \neq 0\} \\ \mathcal{U}_{10} &\coloneqq \{U_1 \neq 0, U_2 = 0\}, & \mathcal{U}_{11} &\coloneqq \{U_1 \neq 0, U_2 \neq 0\}. \end{aligned} \tag{5.18}$$

To simplify our outer bound, we choose the Linear Minimum Mean Squared Error (LMMSE) decoder in the following four scenarios:

$$g_1(u_1, u_2) = \begin{cases} 0, & \text{if } \mathcal{U}_{00} \text{ occurs} \\ u_1 \frac{\mathsf{E}[X_1 U_1 | \mathcal{U}_{10}]}{\mathsf{E}[U_1^2 | \mathcal{U}_{10}]}, & \text{if } \mathcal{U}_{10} \text{ occurs} \\ u_2 \frac{\mathsf{E}[X_1 U_2 | \mathcal{U}_{01}]}{\mathsf{E}[U_2^2 | \mathcal{U}_{01}]}, & \text{if } \mathcal{U}_{01} \text{ occurs} \\ \mathsf{u}^{\mathrm{T}} \mathbf{C}_{\mathsf{UU}|\mathcal{U}_{11}}^{-1} \mathbf{C}_{\mathsf{U}X_1|\mathcal{U}_{11}}, & \text{if } \mathcal{U}_{11} \text{ occurs.} \end{cases} \tag{5.19}$$

g_2 is chosen similarly. Here, $\mathbf{C}_{UX|\mathcal{U}}$ denotes the covariance matrix of U and X conditioned on the event $\mathcal{U}$. We denote the error of the LMMSE estimator when estimating a random variable X from U conditioned on an event $\mathcal{U}$ by (cf. [Kay93, Ch. 12])

$$\mathsf{lmmse}(X; U|\mathcal{U}) \coloneqq \mathsf{Var}[X|\mathcal{U}] - \mathbf{C}_{XU|\mathcal{U}} \mathbf{C}_{U|\mathcal{U}}^{-1} \mathbf{C}_{YU|\mathcal{U}}. \tag{5.20}$$

The complete coding scheme is sketched in Figure 5.3. A relaxation of the Berger-Tung Inner Bound for this choice yields the following rate region.

Theorem 5.3. Fix some $\tau_1, \tau_2 \geq 0$ and let $\hat{B}_j \coloneqq \mathbb{1}_{\{|X_j|>\tau_j\}}$ for $j = 1, 2$. If the rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ satisfies

$$\mathsf{R}_1 \geq H\big(\hat{B}_1 \big| \hat{B}_2\big) + \frac{\Pr[\mathcal{U}_{10}]}{2} \log\left(1 + \frac{\mathsf{Var}[X_1|\mathcal{U}_{10}]}{\sigma_1^2}\right) + \frac{\Pr[\mathcal{U}_{11}]}{2} \log\left(1 + \frac{\mathsf{lmmse}(X_1; U_2|\mathcal{U}_{11})}{\sigma_2^2}\right) \tag{5.21}$$

$$\mathsf{R}_2 \geq H\big(\hat{B}_2 \big| \hat{B}_1\big) + \frac{\Pr[\mathcal{U}_{01}]}{2} \log\left(1 + \frac{\mathsf{Var}[X_2|\mathcal{U}_{01}]}{\sigma_2^2}\right) + \frac{\Pr[\mathcal{U}_{11}]}{2} \log\left(1 + \frac{\mathsf{lmmse}(X_2; U_1|\mathcal{U}_{11})}{\sigma_1^2}\right) \tag{5.22}$$

$$\mathsf{R}_1 + \mathsf{R}_2 \geq H\big(\hat{B}_1\hat{B}_2\big) + \frac{\Pr[\mathcal{U}_{10}]}{2} \log\left(1 + \frac{\mathsf{Var}[X_1|\mathcal{U}_{10}]}{\sigma_1^2}\right) + \frac{\Pr[\mathcal{U}_{01}]}{2} \log\left(1 + \frac{\mathsf{Var}[X_2|\mathcal{U}_{01}]}{\sigma_2^2}\right) + \frac{\Pr[\mathcal{U}_{11}]}{2} \log\left(\frac{\det \mathbf{C}_{\mathsf{UU}|\mathcal{U}_{11}}}{\sigma_1^2\sigma_2^2}\right) \tag{5.23}$$

and the distortion pair (d_1, d_2) satisfies

$$\begin{aligned} d_1 \geq{}& \Pr[\mathcal{U}_{00}] \cdot \mathsf{E}[X_1^2\,|\mathcal{U}_{00}] + \Pr[\mathcal{U}_{01}] \cdot \mathsf{lmmse}(X_1; U_2|\mathcal{U}_{01}) \\ &+ \Pr[\mathcal{U}_{10}] \cdot \mathsf{lmmse}(X_1; U_1|\mathcal{U}_{10}) + \Pr[\mathcal{U}_{11}] \cdot \mathsf{lmmse}(X_1; U_1U_2|\mathcal{U}_{11}) \end{aligned} \tag{5.24}$$

$$\begin{aligned} d_2 \geq{}& \Pr[\mathcal{U}_{00}] \cdot \mathsf{E}[X_2^2\,|\mathcal{U}_{00}] + \Pr[\mathcal{U}_{01}] \cdot \mathsf{lmmse}(X_2; U_2|\mathcal{U}_{01}) \\ &+ \Pr[\mathcal{U}_{10}] \cdot \mathsf{lmmse}(X_2; U_1|\mathcal{U}_{10}) + \Pr[\mathcal{U}_{11}] \cdot \mathsf{lmmse}(X_2; U_1U_2|\mathcal{U}_{11}) \end{aligned} \tag{5.25}$$

then the rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ is achievable with distortions (d_1, d_2).

Detailed derivations and more specific expressions are presented in Appendix B.1.

5.3. Outer Bounds

We next present outer (or converse) bounds for the setting depicted in Figure 5.1. For better readability, we split our results into bounds for the individual rates and the sum-rate. The longer proofs are deferred to the end of the section. We will make use of the ideas leading to Corollary 4.5, and we define

$$\mathsf{R}_B(d) \coloneqq \sup_{\gamma>0} \mathsf{R}_{\mathsf{L}}^{\mathrm{BMS}}\left(\frac{d}{(1-p)\gamma^2}, q_\gamma(d/p)\right). \tag{5.26}$$

Theorem 5.4 (Sum-Rate Bound)**.** If the rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ is achievable with distortions (d_1, d_2), then it satisfies

$$\mathsf{R}_1 + \mathsf{R}_2 \geq \max_{d\in\{d_1,d_2\}} \mathsf{R}_B(d) + \frac{p}{2}\log\frac{(1-\rho^2)\beta\left(d_1 d_2/p^2\right)}{2d_1 d_2/p^2}. \tag{5.27}$$

This sum-rate bound combines ideas from the single user converses in Section 4.2 and the recently developed converse for the quadratic Gaussian source using a new entropy power inequality [Cou18].

A much simpler sum-rate bound can be stated by allowing the encoders to cooperate, i.e., relaxing the distributed source coding problem to joint source coding. Since joint source coding of two sources is equivalent to simply coding one two-dimensional source, Theorem 5.5 has the advantage that we can numerically evaluate it using the Blahut-Arimoto algorithm.

Theorem 5.5 (Cooperative Sum-Rate Bound)**.** If the rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ is achievable with distortions (d_1, d_2), then it satisfies

$$\mathsf{R}_1 + \mathsf{R}_2 \geq \mathsf{R}_{X_1X_2}(d_1, d_2) \coloneqq \min_{P_{Y_1Y_2|X_1X_2}} I(X_1, X_2; Y_1, Y_2) \tag{5.28}$$

where the minimization is over all *joint test channels* that satisfy

$$\mathsf{E}\Big[(X_1 - Y_1)^2\Big] \leq d_1 \quad \text{and} \quad \mathsf{E}\Big[(X_2 - Y_2)^2\Big] \leq d_2. \tag{5.29}$$

Proof. The RHS of (5.28) is the RD function for the case where $(\mathsf{X}_1, \mathsf{X}_2)$ are encoded *jointly*, i.e., encoders 1 and 2 are allowed to cooperate. ■

The Blahut-Arimoto algorithm for joint source coding is given in Algorithm 5.1. It is worth noting that to numerically compute $\mathsf{R}_{X_1X_2}(d_1, d_2)$, we need to discretize $\mathcal{X}\times\mathcal{X}$ and $\mathcal{Y}\times\mathcal{Y}$ reasonably fine enough. Since we are optimizing over joint test channels $P_{Y_1Y_2|X_1X_2}$, the computational and storage complexity scales at least with the fourth power of the number of discretization points.

To derive rate constraints for the individual rates, we again use the ideas from Section 4.2 and [Cou18].

Algorithm 5.1 Blahut-Arimoto Algorithm for $\mathsf{R}_{X_1X_2}(d_1, d_2)$ [Bla72, Ari72]

Choose $\mathcal{X}_{\mathrm{d}}^2$ and $\mathcal{Y}_{\mathrm{d}}^2$ as discretizations of $\mathcal{X} \times \mathcal{X}$ and $\mathcal{Y} \times \mathcal{Y}$, compute discretized $P_{X_1X_2}^{\mathrm{d}}$, choose initial reconstruction distribution $P_{Y_1Y_2}^{(0)}$ on $\mathcal{Y}_{\mathrm{d}}^2$, Lagrange multipliers $s_1, s_2 > 0$, and a target precision $\varepsilon > 0$.
$t \leftarrow 0$
repeat
$$t \leftarrow t + 1$$
$$c^{(t)}(y_1, y_2) \leftarrow \sum_{(x_1,x_2)\in\mathcal{X}_{\mathrm{d}}^2} P_{X_1X_2}^{\mathrm{d}}(x_1, x_2) \frac{\exp(-s_1\delta(x_1,y_1)-s_2\delta(x_2,y_2))}{\sum_{(y_1',y_2')\in\mathcal{Y}_{\mathrm{d}}^2} P_{Y_1Y_2}^{(t)}(y_1',y_2') \exp(-s_1\delta(x_1,y_1')-s_2\delta(x_2,y_2'))}$$
$$P_{Y_1Y_2}^{(t)}(y_1, y_2) \leftarrow P_{Y_1Y_2}^{(t-1)}(y_1, y_2) c^{(t)}(y_1, y_2)$$
$$T_{\mathrm{UB}}^{(t)} \leftarrow \sum_{(y_1,y_2)\in\mathcal{Y}_{\mathrm{d}}^2} P_{Y_1Y_2}^{(t)} \log c^{(t)}(y_1, y_2)$$
$$T_{\mathrm{LB}}^{(t)} \leftarrow \max_{(y_1,y_2)\in\mathcal{Y}_{\mathrm{d}}^2} \log c^{(t)}(y_1, y_2)$$
until $T_{\mathrm{UB}}^{(t)} - T_{\mathrm{LB}}^{(t)} < \varepsilon$
$$P_{Y_1Y_2|X_1X_2}(y_1, y_2|x_1, x_2) \leftarrow \frac{P_{Y_1Y_2}^{(t)}(y_1,y_2) \exp(-s_1\delta(x_1,y_1)-s_2\delta(x_2,y_2))}{\sum_{(y_1',y_2')\in\mathcal{Y}_{\mathrm{d}}^2} P_{Y}^{(t)}(y_1',y_2') \exp(-s_1\delta(x_1,y_1')-s_2\delta(x_2,y_2'))}$$
$$d_1 \leftarrow \sum_{(x_1,x_2)\in\mathcal{X}_{\mathrm{d}}^2} \sum_{(y_1,y_2)\in\mathcal{Y}_{\mathrm{d}}^2} P_{X_1X_2}^{\mathrm{d}}(x_1, x_2) P_{Y_1Y_2|X_1X_2}(y_1, y_2|x_1, x_2) \delta(x_1, y_1)$$
$$d_2 \leftarrow \sum_{(x_1,x_2)\in\mathcal{X}_{\mathrm{d}}^2} \sum_{(y_1,y_2)\in\mathcal{Y}_{\mathrm{d}}^2} P_{X_1X_2}^{\mathrm{d}}(x_1, x_2) P_{Y_1Y_2|X_1X_2}(y_1, y_2|x_1, x_2) \delta(x_2, y_2)$$
$$\mathsf{R}_{X_1X_2}(d_1, d_2) \leftarrow \sum_{(x_1,x_2)\in\mathcal{X}_{\mathrm{d}}^2} \sum_{(y_1,y_2)\in\mathcal{Y}_{\mathrm{d}}^2} P_{X_1X_2}^{\mathrm{d}}(x_1, x_2) P_{Y_1Y_2|X_1X_2}(y_1, y_2|x_1, x_2) \cdot \log \frac{P_{Y_1Y_2|X_1X_2}(y_1,y_2|x_1,x_2)}{P_{Y_1Y_2}^{(t)}(y_1,y_2)}$$
return $P_{Y_1Y_2|X_1X_2}$, d_1, d_2, $\mathsf{R}_{X_1X_2}(d_1, d_2)$

Theorem 5.6 (Individual Rates)**.** If the rate pair $(\mathsf{R}_1, \mathsf{R}_2)$ is achievable with distortions (d_1, d_2), then it satisfies

$$\mathsf{R}_1 \geq \begin{cases} \mathsf{R}_B(d_1) - \mathsf{R}_2 + \dfrac{p}{2} \log \dfrac{1}{d_1/p}, & \text{if } \mathsf{R}_2 \leq \mathsf{R}_B(d_1) \\ \dfrac{p}{2} \log \dfrac{(1-\rho^2) + \rho^2 2^{-2(\mathsf{R}_2 - \mathsf{R}_B(d_1))/p}}{d_1/p}, & \text{if } \mathsf{R}_2 > \mathsf{R}_B(d_1) \end{cases} \tag{5.30}$$

$$\mathsf{R}_2 \geq \begin{cases} \mathsf{R}_B(d_2) - \mathsf{R}_1 + \dfrac{p}{2} \log \dfrac{1}{d_2/p}, & \text{if } \mathsf{R}_1 \leq \mathsf{R}_B(d_2) \\ \dfrac{p}{2} \log \dfrac{(1-\rho^2) + \rho^2 2^{-2(\mathsf{R}_1 - \mathsf{R}_B(d_2))/p}}{d_2/p}, & \text{if } \mathsf{R}_1 > \mathsf{R}_B(d_2). \end{cases} \tag{5.31}$$

In the remainder of this section, we present proofs of Theorems 5.4 and 5.6.

5.3.1. Proof of the Sum-Rate Bound (Theorem 5.4)

Let $\mathsf{N}(1|\mathsf{B}) = \sum_{k=1}^{n} B[i]$ again be the number of nonzero entries in B. Since the conditional expectation is the optimal estimator with respect to an MSE error criterion, we assume that $\mathsf{Y}_j = \mathsf{E}[\mathsf{X}_j \,|\, \mathsf{U}_1, \mathsf{U}_2]$ for $j = 1, 2$. Define

$$\begin{aligned} d_1[k] &\coloneqq \mathsf{E}\Big[\|\mathsf{X}_1 - \mathsf{E}[\mathsf{X}_1 \,|\, \mathsf{U}_1, \mathsf{U}_2]\|_2^2 \Big| \mathsf{N}(1|\mathsf{B}) = k\Big] \\ d_2[k] &\coloneqq \mathsf{E}\Big[\|\mathsf{X}_2 - \mathsf{E}[\mathsf{X}_2 \,|\, \mathsf{U}_1, \mathsf{U}_2]\|_2^2 \Big| \mathsf{N}(1|\mathsf{B}) = k\Big] \end{aligned} \tag{5.32}$$

and denote by d_1 and d_2 the MSEs at terminals one and two, respectively:

$$d_1 = \sum_{k=1}^{n} d_1[k] \Pr[\mathsf{N}(1|\mathsf{B}) = k], \qquad d_2 = \sum_{k=1}^{n} d_2[k] \Pr[\mathsf{N}(1|\mathsf{B}) = k]. \tag{5.33}$$

We start with the usual converse steps to get

$$\begin{aligned} \mathsf{R}_1 + \mathsf{R}_2 &\geq \frac{1}{n} H(\mathsf{U}_1, \mathsf{U}_2) \\ &= \frac{1}{n} I(\mathsf{X}_1, \mathsf{X}_2; \mathsf{U}_1, \mathsf{U}_2) \\ &= \frac{1}{n} I(\mathsf{X}_1, \mathsf{X}_2, \mathsf{B}; \mathsf{U}_1, \mathsf{U}_2) \\ &= \frac{1}{n} \Big(I(\mathsf{B}; \mathsf{U}_1, \mathsf{U}_2) + I(\mathsf{X}_1, \mathsf{X}_2; \mathsf{U}_1, \mathsf{U}_2 | \, \mathsf{B}) \Big) \\ &= \frac{1}{n} I(\mathsf{B}; \mathsf{U}_1, \mathsf{U}_2) + \frac{1}{n} \sum_{k=1}^{n} \Pr[\mathsf{N}(1|\mathsf{B}) = k] I(\mathsf{X}_1, \mathsf{X}_2; \mathsf{U}_1, \mathsf{U}_2 | \, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k) \,. \end{aligned} \tag{5.34}$$

Let $\hat{B}_1[k] \coloneqq \mathbb{1}_{\{Y_1[k] \neq 0\}}$ and $\hat{B}_2[k] \coloneqq \mathbb{1}_{\{Y_2[k] \neq 0\}}$ and note that both random variables are functions of $(\mathsf{U}_1, \mathsf{U}_2)$. Hence, we have

$$\begin{aligned} I(\mathsf{B}; \mathsf{U}_1, \mathsf{U}_2) &\geq I\Big(\mathsf{B}; \hat{\mathsf{B}}_1, \hat{\mathsf{B}}_2\Big) \\ &\geq \max\left\{ I\Big(\mathsf{B}; \hat{\mathsf{B}}_1\Big), I\Big(\mathsf{B}; \hat{\mathsf{B}}_2\Big) \right\} \\ &\geq n \max\left\{ \min_{\substack{P_{\hat{B}_1|B}: \\ \mathsf{E}[(X_1 - Y_1)^2] \leq d_1}} I\Big(B; \hat{B}_1\Big), \min_{\substack{P_{\hat{B}_2|B}: \\ \mathsf{E}[(X_2 - Y_2)^2] \leq d_2}} I\Big(B; \hat{B}_2\Big) \right\} \\ &\geq n \max_{d \in \{d_1, d_2\}} \mathsf{R}_B(d) \end{aligned} \tag{5.35}$$

where the last step follows from the standard converse steps of the RD theorem [CT06b, p. 317] and Corollary 4.5.

Now consider the second term in (5.34). We follow Courtade's alternative proof of the sum rate for the distributed quadratic Gaussian source coding setting using his recently developed strong entropy power inequality [Cou18]. To this end, define $\mathsf{X}_1^{\mathsf{G}}$ and $\mathsf{X}_2^{\mathsf{G}}$ to be

the nonzero elements of X_1 and X_2, respectively. That is, there are (with probability one) one-to-one mappings from X_1 to $(\mathsf{B}, \mathsf{X}_1^{\mathsf{G}})$ and X_2 to $(\mathsf{B}, \mathsf{X}_2^{\mathsf{G}})$. If $\mathsf{N}(1|\mathsf{B}) = k$, then the average error among the k Gaussians at terminal one is at most

$$\frac{1}{k}\mathsf{E}\Big[\|\mathsf{X}_1^{\mathsf{G}} - \mathsf{E}[\mathsf{X}_1^{\mathsf{G}} \,|\, \mathsf{U}_1, \mathsf{U}_2]\|_2^2 \Big| \mathsf{N}(1|\mathsf{B}) = k\Big] \le d_1[k]\frac{n}{k}. \tag{5.36}$$

Now, note that

$$\begin{aligned}
&\frac{1}{k} I(\mathsf{X}_1; \mathsf{U}_1, \mathsf{U}_2 |\, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k) \\
&= \frac{1}{k} I\Big(\mathsf{X}_1^{\mathsf{G}}; \mathsf{U}_1, \mathsf{U}_2 \Big|\, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k\Big) \\
&\overset{\text{a}}{=} \frac{1}{k}\left(\frac{1}{2}\log(2\pi e)^k - h\Big(\mathsf{X}_1^{\mathsf{G}} \Big|\, \mathsf{U}_1, \mathsf{U}_2, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k\Big)\right) \\
&\overset{\text{b}}{\ge} \frac{1}{k}\left(\frac{1}{2}\log(2\pi e)^k - \frac{1}{2}\log\Big(2\pi e d_1[k] n/k\Big)^k\right) \\
&= \frac{1}{2}\log\frac{k/n}{d_1[k]}
\end{aligned} \tag{5.37}$$

where (a) follows since $\mathsf{X}_1^{\mathsf{G}}$ is a k-dimensional standard Gaussian vector and (b) follows from (5.36) and the maximum entropy property of Gaussian random variables. Similarly, we have

$$\frac{1}{k} I(\mathsf{X}_2; \mathsf{U}_1, \mathsf{U}_2 |\, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k) \ge \frac{1}{2}\log\frac{k/n}{d_2[k]} \tag{5.38}$$

and thus

$$\frac{1}{k}\Big(I(\mathsf{X}_1; \mathsf{U}_1, \mathsf{U}_2 |\, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k) + I(\mathsf{X}_2; \mathsf{U}_1, \mathsf{U}_2 |\, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k)\Big) \ge \frac{1}{2}\log\frac{(k/n)^2}{d_1[k]d_2[k]}. \tag{5.39}$$

Following the same steps that lead to the sum rate bound in Courtade's proof [Cou18, Thm. 6] using (5.36) and (5.39), we see that

$$\frac{1}{k} I(\mathsf{X}_1, \mathsf{X}_2; \mathsf{U}_1, \mathsf{U}_2 |\, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k) \ge \frac{1}{2}\log\frac{(1-\rho^2)\beta\Big(d_1[k]d_2[k](n/k)^2\Big)}{2d_1[k]d_2[k](n/k)^2} \tag{5.40}$$

where β is defined in (5.10). To resolve the average over $\mathsf{N}(1|\mathsf{B})$, define the typical set

$$\mathcal{T}_\varepsilon \coloneqq \Big\{\mathsf{b} \in \{0,1\}^n : (p-\varepsilon)n \le \mathsf{N}(1|\mathsf{b}) \le (p+\varepsilon)n\Big\} \tag{5.41}$$

for some $\varepsilon > 0$ and note that Hoeffding's inequality (see, e.g., [Ver18, Thm. 2.2.6]) ensures

$$\Pr[\mathsf{B} \notin \mathcal{T}_\varepsilon] \le 2e^{-\varepsilon^2 n} =: \delta_n \tag{5.42}$$

which in turn implies

$$\begin{aligned}\frac{1}{n}\mathsf{E}\Big[\|\mathsf{X}_1-\mathsf{E}[\mathsf{X}_1\,|\,\mathsf{U}_1,\mathsf{U}_2]\|_2^2\Big|\mathsf{B}\in\mathcal{T}_\varepsilon\Big]&\le\frac{d_1}{1-\delta_n}\\ \frac{1}{n}\mathsf{E}\Big[\|\mathsf{X}_2-\mathsf{E}[\mathsf{X}_2\,|\,\mathsf{U}_1,\mathsf{U}_2]\|_2^2\Big|\mathsf{B}\in\mathcal{T}_\varepsilon\Big]&\le\frac{d_2}{1-\delta_n}.\end{aligned}\tag{5.43}$$

We can now lower bound

$$\begin{aligned}&\frac{1}{n}\sum_{k=1}^{k}\Pr[\mathsf{N}(1|\mathsf{B})=k]I(\mathsf{X}_1,\mathsf{X}_2;\mathsf{U}_1,\mathsf{U}_2|\,\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)\\
&\ge\sum_{k=\lceil(p-\varepsilon)n\rceil}^{\lfloor(p+\varepsilon)n\rfloor}\Pr[\mathsf{N}(1|\mathsf{B})=k]\frac{k}{n}\frac{1}{k}I(\mathsf{X}_1,\mathsf{X}_2;\mathsf{U}_1,\mathsf{U}_2|\,\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)\\
&\overset{\text{a}}{\ge}\sum_{k=\lceil(p-\varepsilon)n\rceil}^{\lfloor(p+\varepsilon)n\rfloor}\Pr[\mathsf{N}(1|\mathsf{B})=k]\frac{p-\varepsilon}{2}\log\frac{(1-\rho^2)\beta\Big(\frac{d_1[k]d_2[k]}{(n/k)^2}\Big)}{2\frac{d_1[k]d_2[k]}{(n/k)^2}}\\
&\ge(1-\delta_n)\sum_{k=\lceil(p-\varepsilon)n\rceil}^{\lfloor(p+\varepsilon)n\rfloor}\Pr[\mathsf{N}(1|\mathsf{B})=k\,|\,\mathsf{B}\in\mathcal{T}_\varepsilon]\frac{p-\varepsilon}{2}\log\frac{(1-\rho^2)\beta\Big(\frac{d_1[k]d_2[k]}{(p-\varepsilon)^2}\Big)}{2\frac{d_1[k]d_2[k]}{(p-\varepsilon)^2}}\\
&\overset{\text{b}}{\ge}(1-\delta_n)\frac{p-\varepsilon}{2}\log\frac{(1-\rho^2)(1-\delta_n)^2\beta\left(\frac{d_1d_2}{(1-\delta_n)^2(p-\varepsilon)^2}\right)}{2d_1d_2/(p-\varepsilon)^2}\end{aligned}\tag{5.44}$$

where (a) follows from (5.40) and (b) follows from the convexity of $\log(\beta(x)/x)$, Jensen's inequality and (5.43). For $n\to\infty$, $\delta_n\to 0$ by (5.42) and we can combine (5.34), (5.35) and (5.44) to obtain

$$\mathsf{R}_1+\mathsf{R}_2\ge\max_{d\in\{d_1,d_2\}}\mathsf{R}_B(d)+\frac{p}{2}\log\frac{(1-\rho^2)\beta\left(\frac{d_1d_2}{p^2}\right)}{2d_1d_2/p^2}.\tag{5.45}$$

■

Remark 5.1. In (5.35), we get the lower bound $\frac{1}{n}I(\mathsf{B};\mathsf{U}_1,\mathsf{U}_2)\ge\max_{d\in\{d_1,d_2\}}\mathsf{R}_B(d)$ which somehow neglects the different supports of the *significant* samples at both terminals. As a consequence, we will see in Section 5.4 that this bound is loose at larger distortions. Still, this bound is useful to determine the behavior for $d_1,d_2\to 0$, as shown in Section 5.3.3. We remark that if one can find a better lower bound for $I(\mathsf{B};\mathsf{U}_1,\mathsf{U}_2)$, then this could be used to improve Theorem 5.4 (and Theorem 5.6).

5.3.2. Proof of the Bound for Individual Rates (Theorem 5.6)

For the individual rate constraints, we start with

$$\begin{aligned} n\mathsf{R}_1 &\geq H(\mathsf{U}_1) \\ &\geq H(\mathsf{U}_1\,|\mathsf{U}_2\,) \\ &= H(\mathsf{U}_1\,|\mathsf{U}_2\,) - H(\mathsf{U}_1\,|\mathsf{U}_2,\mathsf{X}_1\,) \\ &= I(\mathsf{X}_1;\mathsf{U}_1|\,\mathsf{U}_2) \\ &= I(\,\mathsf{B},\mathsf{X}_1;\mathsf{U}_1|\,\mathsf{U}_2) \\ &= I(\,\mathsf{B};\mathsf{U}_1|\,\mathsf{U}_2) + I(\mathsf{X}_1;\mathsf{U}_1|\,\mathsf{B},\mathsf{U}_2) \\ &= I(\mathsf{B};\mathsf{U}_1,\mathsf{U}_2) - I(\mathsf{B};\mathsf{U}_2) + I(\mathsf{X}_1;\mathsf{U}_1|\,\mathsf{B},\mathsf{U}_2)\,. \end{aligned} \tag{5.46}$$

Using the relation (5.37) in (a), we have for a fixed k that

$$\begin{aligned} 2^{\frac{2}{k}I(\mathsf{X}_1;\mathsf{U}_1|\mathsf{U}_2,\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)-\log\frac{k/n}{d_1[k]}} &\overset{\text{a}}{\geq} 2^{\frac{2}{k}(I(\mathsf{X}_1;\mathsf{U}_1|\mathsf{U}_2,\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)-I(\mathsf{X}_1;\mathsf{U}_1,\mathsf{U}_2|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k))} \\ &= 2^{-\frac{2}{k}I(\mathsf{X}_1;\mathsf{U}_2|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)} \\ &= 2^{-\frac{2}{k}I\left(\mathsf{X}_1^{\mathsf{G}};\mathsf{U}_2\middle|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k\right)}. \end{aligned} \tag{5.47}$$

We lower bound (5.47) by using the conditional version of the strong entropy power inequality [Cou18, Cor. 2]. Note that conditioned on B and $\mathsf{N}(1|\mathsf{B}) = k$, $\mathsf{X}_2^{\mathsf{G}}$ can be written as

$$\mathsf{X}_2^{\mathsf{G}} = \rho\mathsf{X}_1^{\mathsf{G}} + \sqrt{1-\rho^2}\mathsf{Z}, \qquad \mathsf{Z} \sim \mathcal{N}(0, \mathsf{Id}_k). \tag{5.48}$$

Applying [Cou18, Cor. 2], we have

$$\begin{aligned} &2^{\frac{2}{k}\left(h\left(\mathsf{X}_2^{\mathsf{G}}\middle|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k\right)-I\left(\mathsf{X}_1^{\mathsf{G}};\mathsf{U}_2\middle|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k\right)\right)} \\ &\geq 2^{\frac{2}{k}\left(h\left(\rho\mathsf{X}_1^{\mathsf{G}}\middle|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k\right)-I\left(\mathsf{X}_2^{\mathsf{G}};\mathsf{U}_2\middle|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k\right)\right)} + 2^{\frac{2}{k}h\left(\sqrt{1-\rho^2}\mathsf{Z}\middle|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k\right)} \\ &= \rho^2 2^{\frac{2}{k}\left(h\left(\mathsf{X}_1^{\mathsf{G}}\middle|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k\right)-I(\mathsf{X}_2;\mathsf{U}_2|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)\right)} + (1-\rho^2)2^{\frac{2}{k}h(\mathsf{Z}|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)}. \end{aligned} \tag{5.49}$$

Note that $\mathsf{X}_1^{\mathsf{G}}$, $\mathsf{X}_2^{\mathsf{G}}$, and Z have the same conditional distributions and entropies, so (5.47) and (5.49) together show that

$$2^{\frac{2}{k}I(\mathsf{X}_1;\mathsf{U}_1|\mathsf{U}_2,\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)-\log\frac{k/n}{d_1[k]}} \geq \rho^2 2^{-\frac{2}{k}I(\mathsf{X}_2;\mathsf{U}_2|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)} + (1-\rho^2) \tag{5.50}$$

which we can rearrange to become

$$I(\mathsf{X}_1;\mathsf{U}_1|\,\mathsf{U}_2,\mathsf{B},\mathsf{N}(1|\mathsf{B})=k) \geq \frac{1}{2}\log\frac{(1-\rho^2)+\rho^2 2^{-\frac{2}{k}I(\mathsf{X}_2;\mathsf{U}_2|\mathsf{B},\mathsf{N}(1|\mathsf{B})=k)}}{d_1[k]\cdot n/k}. \tag{5.51}$$

We next resolve the average over $\mathsf{N}(1|\mathsf{B})$. Recall the definition of $\mathcal{T}_\varepsilon$ in (5.41) and its

probability bound (5.42). We have

$$
\begin{aligned}
&\frac{1}{n} I(\mathsf{X}_1; \mathsf{U}_1 | \, \mathsf{U}_2, \mathsf{B}) \\
&\geq \frac{1}{n}(1-\delta_n) I(\mathsf{X}_1; \mathsf{U}_1 | \, \mathsf{U}_2, \mathsf{B}, \mathsf{B} \in \mathcal{T}_\varepsilon) \\
&= (1-\delta_n) \frac{1}{n} \sum_{k=\lceil (p-\varepsilon)n \rceil}^{\lfloor (p+\varepsilon)n \rfloor} \Pr[\mathsf{N}(1|\mathsf{B}) = k \,|\, \mathsf{B} \in \mathcal{T}_\varepsilon] I(\mathsf{X}_1; \mathsf{U}_1 | \, \mathsf{U}_2, \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k) \\
&\overset{\text{a}}{\geq} (1-\delta_n) \sum_{k=\lceil (p-\varepsilon)n \rceil}^{\lfloor (p+\varepsilon)n \rfloor} \Pr[\mathsf{N}(1|\mathsf{B}) = k \,|\, \mathsf{B} \in \mathcal{T}_\varepsilon] \\
&\qquad \cdot \frac{k/n}{2} \log \frac{(1-\rho^2) + \rho^2 2^{-\frac{2}{k} I(\mathsf{X}_2; \mathsf{U}_2 | \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k)}}{d_1[k](n/k)} \\
&\geq (1-\delta_n) \sum_{k=\lceil (p-\varepsilon)n \rceil}^{\lfloor (p+\varepsilon)n \rfloor} \Pr[\mathsf{N}(1|\mathsf{B}) = k \,|\, \mathsf{B} \in \mathcal{T}_\varepsilon] \\
&\qquad \cdot \frac{p-\varepsilon}{2} \log \frac{(1-\rho^2) + \rho^2 2^{-\frac{2}{n(p+\varepsilon)} I(\mathsf{X}_2; \mathsf{U}_2 | \mathsf{B}, \mathsf{N}(1|\mathsf{B}) = k)}}{d_1/(p-\varepsilon)} \\
&\overset{\text{b}}{\geq} (1-\delta_n) \frac{p-\varepsilon}{2} \log \frac{(1-\rho^2) + \rho^2 2^{-\frac{2}{n(p+\varepsilon)} I(\mathsf{X}_2; \mathsf{U}_2 | \mathsf{B}, \mathsf{B} \in \mathcal{T}_\varepsilon)}}{d_1/(p-\varepsilon)} \\
&\geq (1-\delta_n) \frac{p-\varepsilon}{2} \log \frac{(1-\rho^2) + \rho^2 2^{-\frac{2}{n} \frac{I(\mathsf{X}_2; \mathsf{U}_2 | \mathsf{B})}{(p+\varepsilon)(1-\delta_n)}}}{d_1/(p-\varepsilon)}.
\end{aligned} \tag{5.52}
$$

Here, we used (5.51) in (a) and applied Jensen's inequality in (b) by using the convexity of $\log((1-\rho^2) + \rho^2 2^{-2x/k})$.
We can now reuse (5.35) for user 1 and insert (5.52) into (5.46) to obtain

$$
\begin{aligned}
\mathsf{R}_1 &\geq \frac{1}{n} \Big(I(\mathsf{B}; \mathsf{U}_1, \mathsf{U}_2) - I(\mathsf{B}; \mathsf{U}_2) + I(\mathsf{X}_1; \mathsf{U}_1 | \, \mathsf{U}_2, \mathsf{B}) \Big) \\
&\geq \Big[\mathsf{R}_B(d_1) - I(\mathsf{B}; \mathsf{U}_2) \Big]_+ + (1-\delta_n) \frac{p-\varepsilon}{2} \log \frac{(1-\rho^2) + \rho^2 2^{-\frac{2}{n} \frac{I(\mathsf{X}_2; \mathsf{U}_2 | \mathsf{B})}{(p-\varepsilon)(1-\delta_n)}}}{d_1/(p+\varepsilon)}
\end{aligned} \tag{5.53}
$$

where we added the positive part $[\cdot]_+$ because $I(\mathsf{B}; \mathsf{U}_1, \mathsf{U}_2) \geq I(\mathsf{B}; \mathsf{U}_2)$. To tackle the remaining mutual information terms above, the rate for user 2 can be estimated via

$$
\frac{1}{n} \mathsf{R}_2 \geq H(\mathsf{U}_2) = H(\mathsf{U}_2) - H(\mathsf{U}_2 | \mathsf{X}_2, \mathsf{B}) = I(\mathsf{U}_2; \mathsf{X}_2 \mathsf{B}) = I(\mathsf{U}_2; \mathsf{B}) + I(\mathsf{U}_2; \mathsf{X}_2 | \, \mathsf{B}) \,. \tag{5.54}
$$

Further, note that as $n \to \infty$, we have $\delta_n \to 0$ for any $\varepsilon > 0$. We can thus reformulate the

rate requirement for terminal 1 and $n \to \infty$ as

$$\mathsf{R}_1 \geq \min_{R'+R''\leq \mathsf{R}_2} \left\{ \left[\mathsf{R}_B(d_1) - R' \right]_+ + \frac{p}{2} \log \frac{(1-\rho^2) + \rho^2 2^{-2R''/p}}{d_1/p} \right\}. \tag{5.55}$$

To resolve the minimization, consider $\mathsf{R}_2 \leq \mathsf{R}_B(d_1)$ and denote the RHS of (5.55) by Λ. In this case, we have $\frac{\partial \Lambda}{\partial R'} = -1$ and

$$\frac{\partial \Lambda}{\partial R''} = -\frac{\rho^2 2^{-2R''/p}}{(1-\rho^2) + \rho^2 2^{-2R''/p}} > -1. \tag{5.56}$$

Thus, as long as $\mathsf{R}_2 \leq \mathsf{R}_B(d_1)$, the minimum is attained by choosing $R' = \mathsf{R}_2$. For $R' \geq \mathsf{R}_B(d_1)$, Λ is constant in R'. Hence, the minimum is attained by choosing $R' = \mathsf{R}_B(d_1)$ and $R'' = \mathsf{R}_2 - \mathsf{R}_B(d_1)$. This completes the proof. ■

5.3.3. Small Distortion Regime

Using the results obtained for the single user setting, we can also determine the rate region for the DBGS in the low distortion limit $d_1, d_2 \searrow 0$. Denote inner bound on the rate region derived in Theorem 5.2 by

$$\mathcal{R}_0(d_1, d_2) \coloneqq \left\{ (\mathsf{R}_1, \mathsf{R}_2) : \begin{array}{r} \mathsf{R}_1 \geq \frac{p}{2} \log^+ \frac{1-\rho^2+\rho^2 2^{-2[\mathsf{R}_2 - H_2(p)]_+/p}}{d_1/p} \\ \mathsf{R}_2 \geq \frac{p}{2} \log^+ \frac{1-\rho^2+\rho^2 2^{-2[\mathsf{R}_1 - H_2(p)]_+/p}}{d_2/p} \\ \mathsf{R}_1 + \mathsf{R}_2 \geq H_2(p) + \frac{p}{2} \log^+ \frac{(1-\rho^2)\beta(d_1,d_2)}{2d_1 d_2/p^2} \end{array} \right\} \tag{5.57}$$

where β is defined in (5.10). Further, denote the outer bound given by Theorem 5.4 and Theorem 5.6 by $\mathcal{R}_{\text{outer}}(d_1, d_2)$. For $\mathsf{R}_1, \mathsf{R}_2 \geq \max_{d\in\{d_1,d_2\}} \mathsf{R}_B(d)$, we have

$$\mathcal{R}_{\text{outer}}(d_1, d_2) \coloneqq \left\{ (\mathsf{R}_1, \mathsf{R}_2) : \begin{array}{r} \mathsf{R}_1 \geq \frac{p}{2} \log \frac{(1-\rho^2)+\rho^2 2^{-2(\mathsf{R}_2 - \mathsf{R}_B(d_1))/p}}{d_1/p} \\ \mathsf{R}_2 \geq \frac{p}{2} \log \frac{(1-\rho^2)+\rho^2 2^{-2(\mathsf{R}_1 - \mathsf{R}_B(d_2))/p}}{d_2/p} \\ \mathsf{R}_1 + \mathsf{R}_2 \geq \max\limits_{d\in\{d_1,d_2\}} \mathsf{R}_B(d) + \frac{p}{2} \log \frac{(1-\rho^2)\beta\left(d_1 d_2/p^2\right)}{2d_1 d_2/p^2} \end{array} \right\}. \tag{5.58}$$

Theorem 5.7. As $d_1, d_2 \searrow 0$, $\mathcal{R}(d_1, d_2)$ and $\mathcal{R}_0(d_1, d_2)$ are asymptotically equal.

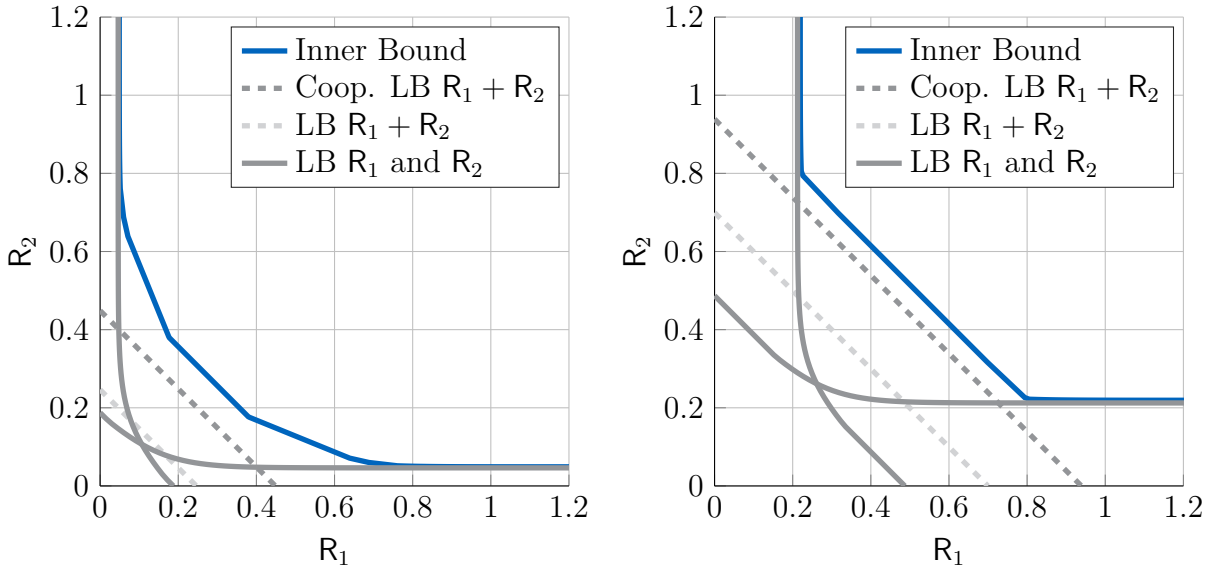
Proof. Since $\mathcal{R}_0(d_1, d_2)$ is an achievable rate region, we have $\mathcal{R}_0(d_1, d_2) \subseteq \mathcal{R}(d_1, d_2)$ for any d_1, d_2. To show that $\mathcal{R}(d_1, d_2) \subseteq \mathcal{R}_0(d_1, d_2)$ as $d_1, d_2 \searrow 0$, note that $\mathcal{R}(d_1, d_2) \subseteq \mathcal{R}_{\text{outer}}(d_1, d_2)$ for any distortion pair. Thus, it suffices to show that $\mathcal{R}_{\text{outer}}(d_1, d_2) \subseteq \mathcal{R}_0(d_1, d_2)$ as $d_1, d_2 \searrow 0$. To this end, we only need to verify that

$$\lim_{d_1,d_2\searrow 0} \max_{d\in\{d_1,d_2\}} \mathsf{R}_B(d) = H_2(p) \tag{5.59}$$

which is done in the proof of Corollary 4.30. We conclude that Theorem 5.7 holds. ■

5.4. Numerical Examples

We present numerical examples for the inner and outer bounds derived in Sections 5.2 and 5.3. Figure 5.4 shows an example with high correlation between the two terminals for distortions of -20 dB and -30 dB at both terminals.



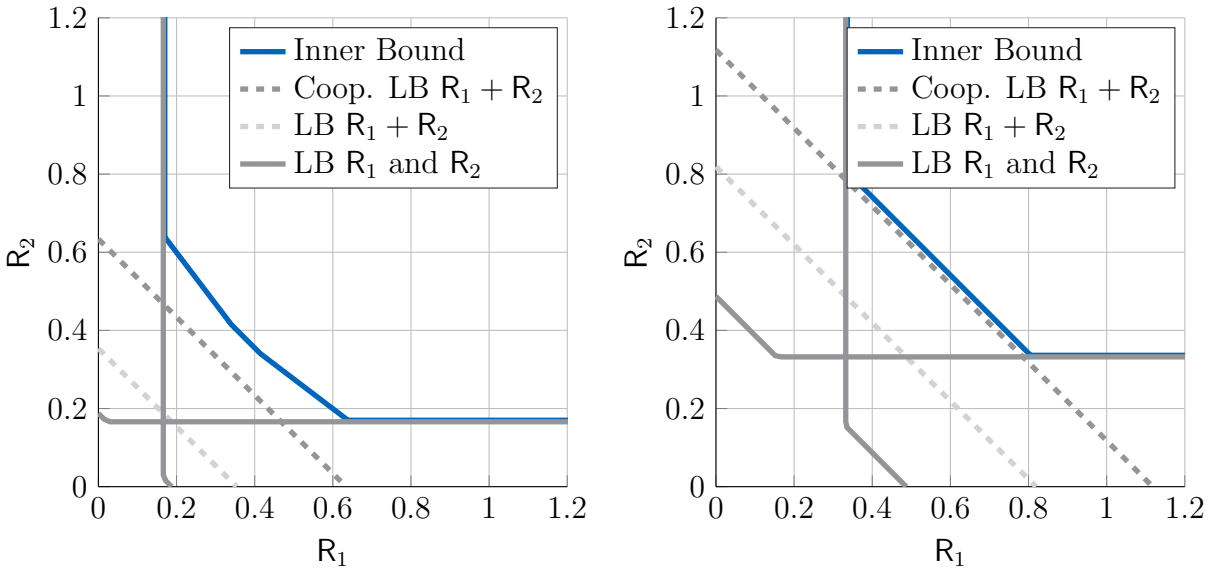
(a) $d_1 = d_2 = -20$ dB.

(b) $d_1 = d_2 = -30$ dB.

Figure 5.4.: Inner bound (Theorem 5.3) and outer bounds from Section 5.3 for the rate region of the DBGS for $p = 0.1$ and $\rho = 0.9$.

In both examples, the cooperative lower bound from Theorem 5.5 (Coop. LB $\mathsf{R}_1 + \mathsf{R}_2$) computed by the Blahut-Arimoto algorithm gives a better bound than Theorem 5.4 (LB $\mathsf{R}_1 + \mathsf{R}_2$). This is likely because the first term in (5.27), $\max_{d \in \{d_1, d_2\}} \mathsf{R}_B(d)$, is a loose lower bound for $\frac{1}{n} I(\mathsf{B}; \mathsf{U}_1, \mathsf{U}_2)$ which captures the amount of information about the position of the spikes that must be stored by the encoders.

As a comparison, Figure 5.5 shows two examples where the Gaussian components are independent, i.e., $\rho = 0$. We see that for $d_1 = d_2 = -20$ dB, the sum rate bound from Theorem 5.4 is only slightly better than what is provided by the individual rate constraints. The cooperative lower bound, however, is close to the achievable rates provided by the distributed threshold coding scheme from Theorem 5.3. For $d_1 = d_2 = -30$ dB, the difference between the inner and outer bounds is already negligible and the rate region is very well approximated by the asymptotic expression given in Theorem 5.7.



(a) $d_1 = d_2 = -20$ dB.

(b) $d_1 = d_2 = -30$ dB.

Figure 5.5.: Inner bound (Theorem 5.3) and outer bounds from Section 5.3 for the rate region of the DBGS for $p = 0.1$ and $\rho = 0$.

Part II.

Bayesian Compressed Sensing

6

Quantized Compressed Sensing with Message Passing Reconstruction

In this chapter, we study Quantized Compressed Sensing (QCS) from a statistical inference point of view. Consider the model

$$Q[k] = \varphi\Big(\frac{1}{\sqrt{n}}\langle \mathsf{A}_k, \mathsf{X}\rangle\Big), \qquad 1 \le k \le m \tag{6.1}$$

where

- ▷ X is ouput by a memoryless source with distribution P_X,
- ▷ A_k is the transposed kth row of $\mathbf{A} \in \mathbb{R}^{m\times n}$ which is a dense measurement matrix with iid $\mathcal{N}(0,1)$ entries, and
- ▷ $\varphi : \mathbb{R} \to \mathcal{Q}$ with $\#(\mathcal{Q}) = 2^b$ is a b-bit quantization function.

We assume that P_X, φ and $\mathbf{A}$ are known to the decoder. Based on this model, we can form the posterior distribution

$$P(\mathsf{x}|\mathsf{Q}, \mathbf{A}) \propto \prod_{i=1}^{n} P_X(x[i]) \prod_{k=1}^{m} \mathbb{1}_{\left\{Q[k]=\varphi\left(\frac{1}{\sqrt{n}}\langle \mathsf{A}_k, \mathsf{x}\rangle\right)\right\}} \tag{6.2}$$

to compute the optimal estimator

$$\mathsf{Y}_{\mathrm{MMSE}} = \arg\min \mathsf{E}\Big[\|\mathsf{X} - \mathsf{Y}\|^2 \Big| \mathsf{Q}, \mathbf{A}\Big] = \mathsf{E}[\mathsf{X} \,|\, \mathsf{Q}, \mathbf{A}] \tag{6.3}$$

with respect to the *Minimum Mean Squared Error (MMSE)*

$$\mathsf{mmse}(\mathsf{X}|\mathsf{Q}, \mathbf{A}) \coloneqq \mathsf{E}\Big[\|\mathsf{X} - \mathsf{E}[\mathsf{X} \,|\, \mathsf{Q}, \mathbf{A}]\|_2^2\Big]. \tag{6.4}$$

Unfortunately, finding the optimal estimator is computationally infeasible unless the dimensions are extremely small. A growing body of recent research, much of which is built on ideas and tools from statistical physics, has been focused on developing computationally feasible estimators that approximate the MMSE estimator (6.3) and investigates the fundamental limits of the optimal estimator [DMM09, BM11, Ran11, JM12, KMS+12a, KMS+12b, JM13, BSK15, BKM+19]. We shall review some of these works below.

Our goal in this chapter is to investigate the RD trade-offs for a QCS system. To this end, we first review the GAMP algorithm and apply it to our QCS setting in Section 6.1. There, we also numerically compare the performance for a Bernoulli-Gaussian source with the RD function. Section 6.2 conducts a similar study for the case of a distributed Bernoulli-Gaussian source. There, we extend the GAMP algorithm to the two-terminal setting and then numerically compare its RD performance to the RD limits using the results from Chapter 5. Finally, Section 6.3 applies the recent theory developed in [BKM+19] to compute achievable *information rates* for QCS systems. The information rates effectively determine the RD performance of a QCS if one applies an optimal lossless compression algorithm to the quantized measurements. We again compare these results to the RD function of a Bernoulli-Gaussian source.

6.1. Bayesian Compressed Sensing via Approximate Message Passing

Approximate Message Passing (AMP) was introduced as a computationally efficient iterative thresholding algorithm for large scale CS problems by Donoho, Maleki and Montanari [DMM09, DMM10a, DMM10b]. Rangan [Ran11] provided an extension to more general signal priors and elementwise output functions and established the term *Generalized Approximate Message Passing (GAMP)* that is widely used. For more details regarding the origin and different variants of AMP algorithms, see [ZK16, Sec. VI.C].

We give a brief sketch of the main ideas behind (G)AMP. For a detailed and accessible derivation of (G)AMP, see, e.g. [ZK16] or [EK12, Ch. 9]. The starting point for the derivation of AMP are the Belief Propagation (BP) equations corresponding to its graphical model, see Figure 6.1. In this graphical model, the square *factor nodes* at the top represent the quantizer φ with the observations Q, whereas the circular *variable nodes* represent the signal components about which the distribution P_X is known as an initial condition for the algorithm. The BP algorithm then iteratively exchanges the available information (called *beliefs*) between the variable nodes and factor nodes. Unfortunately, this exchange of information involves tracking complicated probability measures and is unfeasible for applications such as CS. Loosely speaking, this challenge can be tackled by exploiting the fact that mixtures of many random variables tend to become Gaussian by the central limit theorem. Since a Gaussian distribution is fully specified by its mean and variance, these distributions can easily be tracked. Carefully using the central limit theorem and other approximations, one can then reduce the BP iterations to a sequence of matrix-vector

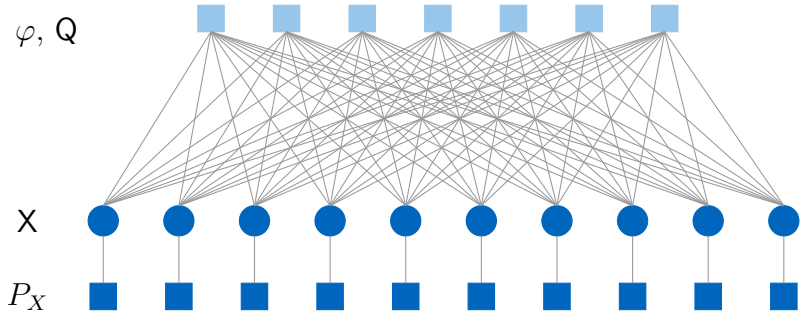


Figure 6.1.: Graphical model for QCS. The light blue *factor nodes* represent the scalar quantizer φ and the observed quantized measurements Q. The dark blue *variable nodes* represent the signal components, each of which has marginal distribution P_X.

multiplications and two scalar inference problems.

We will tailor the GAMP algorithm steps to QCS, as presented in [KGR12]. The first scalar problem is related to the factor nodes. To this end, denote $\mu := \mathsf{E}[X^2]$, let $\varphi : \mathbb{R} \to \{1, \dots, 2^b\}$ be a quantization function, $(V, W) \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$, and consider the quantizer output

$$\tilde{Q} = \varphi\left(\sqrt{\eta} \cdot V + \sqrt{\mu - \eta} \cdot W\right) \tag{6.5}$$

for $\eta \in [0, \mu]$. We interpret $\sqrt{\eta}V$ as side information and are interested in estimating W from the quantized measurement. Define the two functions $g_{P_{\text{out}}} : \mathbb{R} \to \mathbb{R}$ and $h_{P_{\text{out}}} : \mathbb{R} \to \mathbb{R}$ via:

$$g_{P_{\text{out}}}(\tilde{q}, v, \mu - \eta; \varphi) = \frac{1}{\sqrt{\mu - \eta}} \mathsf{E}[W \,|\, \tilde{Q} = \tilde{q}, \sqrt{\eta}V = v] \tag{6.6}$$

$$h_{P_{\text{out}}}(\tilde{q}, v, \mu - \eta; \varphi) = \frac{1}{\mu - \eta}\left(1 - \mathsf{Var}[W | \tilde{Q} = \tilde{q}, \sqrt{\eta}V = v]\right). \tag{6.7}$$

The second inference problem is that of estimating a single $X \sim P_X$ from a measurement corrupted by Gaussian noise

$$\tilde{X} = X + N/\sqrt{\mathsf{snr}} \tag{6.8}$$

where $\mathsf{snr} \geq 0$ and $N \sim \mathcal{N}(0, 1)$ independent of X. Note that $\tilde{X}$ has a PDF irrespective of whether X is discrete, continuous or mixed. We define the two functions $g_{P_X} : \mathbb{R} \to \mathbb{R}$ and $h_{P_X} : \mathbb{R} \to \mathbb{R}$ via

$$\begin{aligned} g_{P_X}(\tilde{x}, \mathsf{snr}) &= \mathsf{E}[X \,|\, \tilde{X} = \tilde{x}] \\ h_{P_X}(\tilde{x}, \mathsf{snr}) &= \mathsf{Var}[X | \tilde{X} = \tilde{x}]. \end{aligned} \tag{6.9}$$

Taking vectors as inputs, the functions $g_{P_{\text{out}}}$, $h_{P_{\text{out}}}$, g_{P_X}, h_{P_X}, and $(\cdot)^{-1}$ are applied component-wise and $\odot$ denotes component-wise multiplication for vectors and matrices. The GAMP algorithm for QCS is given in Algorithm 6.1.

Algorithm 6.1 GAMP for QCS [KGR12]

Initialize:
$\mathsf{y}^0 = \mathsf{E}[\mathsf{X}]$
$\mathsf{v}_x^0 = \mathsf{Var}[\mathsf{X}]$
$\hat{\mathsf{s}}^0 = \mathsf{0}$
for $t = 1, 2, 3, \ldots$ **do**
Factor update:
$\mathsf{v}_p^t = \frac{1}{n}(\mathbf{A} \odot \mathbf{A})\mathsf{v}_x^{t-1}$
$\hat{\mathsf{p}}^t = \frac{1}{\sqrt{n}}\mathbf{A}\mathsf{y}^{t-1} - \mathsf{v}_p^t \odot \hat{\mathsf{s}}^{t-1}$
$\hat{\mathsf{s}}^t = g_{P_{\text{out}}}(\mathsf{q}, \hat{\mathsf{p}}^t, \mathsf{v}_p^t; \varphi)$
$\mathsf{v}_s^t = h_{P_{\text{out}}}(\mathsf{q}, \hat{\mathsf{p}}^t, \mathsf{v}_p^t; \varphi)$
Variable update:
$\mathsf{v}_r^t = \frac{1}{n}(\mathbf{A} \odot \mathbf{A})^{\mathsf{T}}\mathsf{v}_s^t$
$\hat{\mathsf{r}}^t = \mathsf{y}^{t-1} + (\mathsf{v}_r^t)^{-1} \odot \left(\frac{1}{\sqrt{n}}\mathbf{A}^{\mathsf{T}}\hat{\mathsf{s}}^t\right)$
$\mathsf{y}^t = g_{P_X}(\hat{\mathsf{r}}^t, \mathsf{v}_r^t)$
$\mathsf{v}_x^t = h_{P_X}(\hat{\mathsf{r}}^t, \mathsf{v}_r^t)$
end for
return y^t

An important property of (G)AMP algorithms is that their asymptotic performance (as $n, m \to \infty$ with $m/n \to \alpha$) can be predicted via the *State Evolution (SE)*. We define two state variables - one for each scalar inference problem (6.5) and (6.8). The SE then iteratively recomputes the state variables via the functions $h_{P_{\text{out}}}$ and h_{P_X} until convergence. The correctness of SE for GAMP has been proved in [JM13]. The SE procedure for QCS is given in Algorithm 6.2.

Algorithm 6.2 GAMP SE for QCS

Initialize:
$\mu = \mathsf{E}[X^2]$
$\eta_{\text{SE}}^0 = 0$
for $t = 1, 2, 3, \ldots$ **do**
Factor update:
$\mathsf{snr}^t = \alpha \cdot \mathsf{E}_{VY}[h_{P_{\text{out}}}(Q, V, \mu - \eta_{\text{SE}}^t; \varphi)]$
Variable update:
$\eta_{\text{SE}}^t = \mu - \mathsf{E}_{\tilde{X}}\left[h_{P_X}\left(\tilde{X}, \mathsf{snr}^t\right)\right]$
end for
return $\text{MSE} = \mu - \eta_{\text{SE}}^t$

6.1.1. Numerical Example for Bernoulli-Gaussian Signals

As an example, consider the Bernoulli-Gaussian spike source with distribution

$$P_X = (1-p)\cdot\delta_0 + p\cdot\mathcal{N}(0,1). \tag{6.10}$$

For this source, the estimation functions in (6.9) are

$$g_{P_X}(\tilde{x},\mathsf{snr}) = \frac{\tilde{x}}{1+\frac{(1-p)}{p}\sqrt{1+\mathsf{snr}}\exp\left(-\frac{\mathsf{snr}^2\tilde{x}^2}{2(1+\mathsf{snr})}\right)}\cdot\frac{\mathsf{snr}}{1+\mathsf{snr}} \tag{6.11}$$

$$h_{P_X}(\tilde{x},\mathsf{snr}) = \frac{1}{1+\frac{(1-p)}{p}\sqrt{1+\mathsf{snr}}\exp\left(-\frac{\mathsf{snr}^2\tilde{x}^2}{2(1+\mathsf{snr})}\right)}\left(\frac{1}{1+\mathsf{snr}}+\left(\frac{\mathsf{snr}\cdot\tilde{x}}{1+\mathsf{snr}}\right)^2\right) - g_{P_X}(\tilde{x},\mathsf{snr})^2. \tag{6.12}$$

The estimation functions on the quantizer side, Eq. (6.6)-(6.7), are

$$g_{P_{\text{out}}}(\tilde{q},v,\mu-\eta;\varphi) = \frac{1}{\mu-\eta}\Big(\mathsf{E}[Z\,|\,\varphi(Z)=\tilde{q}]-v\Big), \qquad Z\sim\mathcal{N}(v,\mu-\eta) \tag{6.13}$$

$$h_{P_{\text{out}}}(\tilde{q},v,\mu-\eta;\varphi) = \frac{1}{\mu-\eta}\left(1-\frac{\mathsf{Var}[Z|\varphi(Z)=\tilde{q}]}{\mu-\eta}\right), \quad Z\sim\mathcal{N}(v,\mu-\eta). \tag{6.14}$$

Since Z is a truncated Gaussian on the event $\{\varphi(Z)=\tilde{q}\}$ for some $\tilde{q}\in\mathcal{Q}$, the above expectation and variance can easily be calculated numerically in terms of the Gaussian probability and cumulative density functions.

In Figure 6.2, we compare the SE predictions of the asymptotic MSE with the errors empirically observed through simulations for different b and α. Here, we chose P_X to be Bernoulli-Gaussian with $p=0.1$ and the signal length $n=5000$. For $b\geq 2$, we choose φ such that each quantization interval has probability 2^{-b} under the Gaussian measure with mean zero and variance μ. For each b and α we plot the median MSE of 250 experiments.

We further show the critical measurement rate α_{crit}, at which the phase transition to perfect recovery happens for Gaussian matrices with noiseless and unquantized measurements. While the optimal estimator achieves perfect reconstruction for $\alpha > p$ even for Gaussian matrices, this is not the case for AMP [KMS$^+$12a, KMS$^+$12b]. In this case, we can use SE to compute $\alpha_{\text{crit}}\approx 0.21$.

As expected, the MSE decreases with increasing b. Further, there is a sharp decline in the MSE for $\alpha > \alpha_{\text{crit}}$, which matches the phase transition in the limit of infinite quantization rate. We conclude that the SE predictions for these parameters are very accurate. For $\alpha \gg \alpha_{\text{crit}}$, the error decreases slowly in α as we are effectively *oversampling* the signal which is known to yield an error decrease inversely proportional to the sampling rate, see [JLBB13, Thm. 1] and [TV94, GVT98].

Figure 6.3 compares the SE predictions for different b and α but with a fixed bit budget of $b\alpha$ bits per source symbol. This is compared to the RD function of the Bernoulli-Gaussian source with MSE constraint which was computed using the Blahut-Arimoto

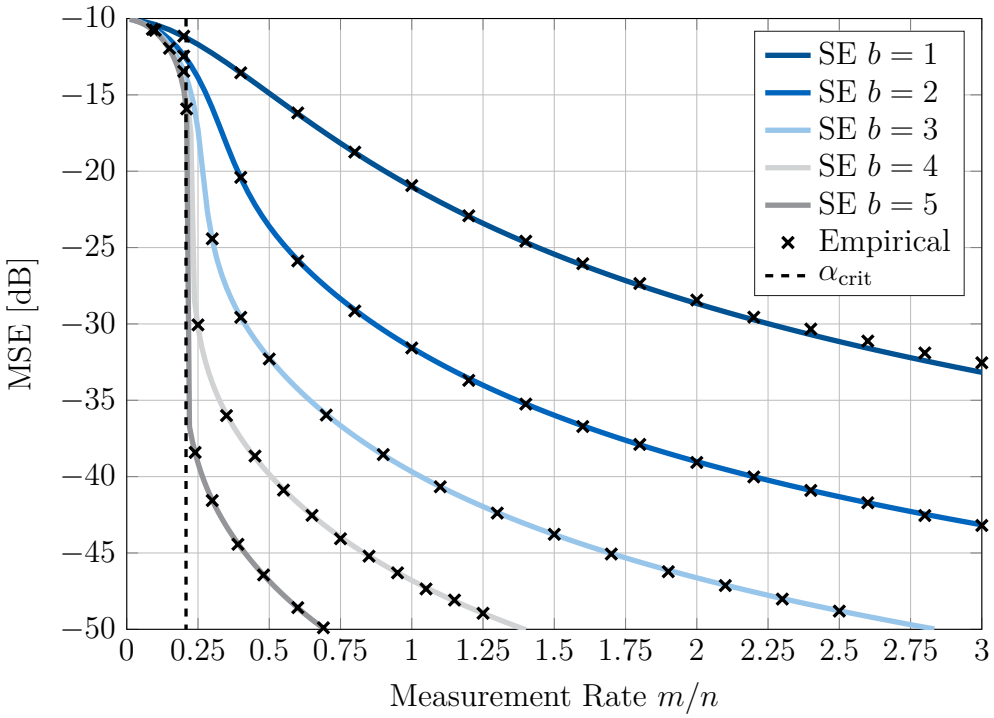


Figure 6.2.: GAMP performance as predicted by SE and empirically observed for a Bernoulli-Gaussian source with $p = 0.1$ with $n = 5000$.

algorithm (Algorithm 2.1). We observe that for a target bit rate R the RD trade-off is best when choosing b to be the largest value under the condition $b\alpha_{\text{crit}} < R$. Further, there is a significant gap between the RD function and the best performing SE graph. To investigate whether we can close this gap, we ask the following to questions.

1) When is GAMP approximately the MMSE estimator? If GAMP is suboptimal for QCS, a better (but possibly much more complicated) reconstruction algorithm would improve the RD trade-off.

2) How well can we compress the quantized measurements? One might guess that for $\alpha \gg \alpha_{\text{crit}}$, the quantized measurements become statistically more dependent. In this case, a lossless compression algorithm improves the RD trade-off.

3) Can we add helpful preprocessing? We mentioned in Chapter 4 that a good vector quantization scheme is to code only some *significant samples* of a sparse signal. We propose a similar preprocessing for QCS.

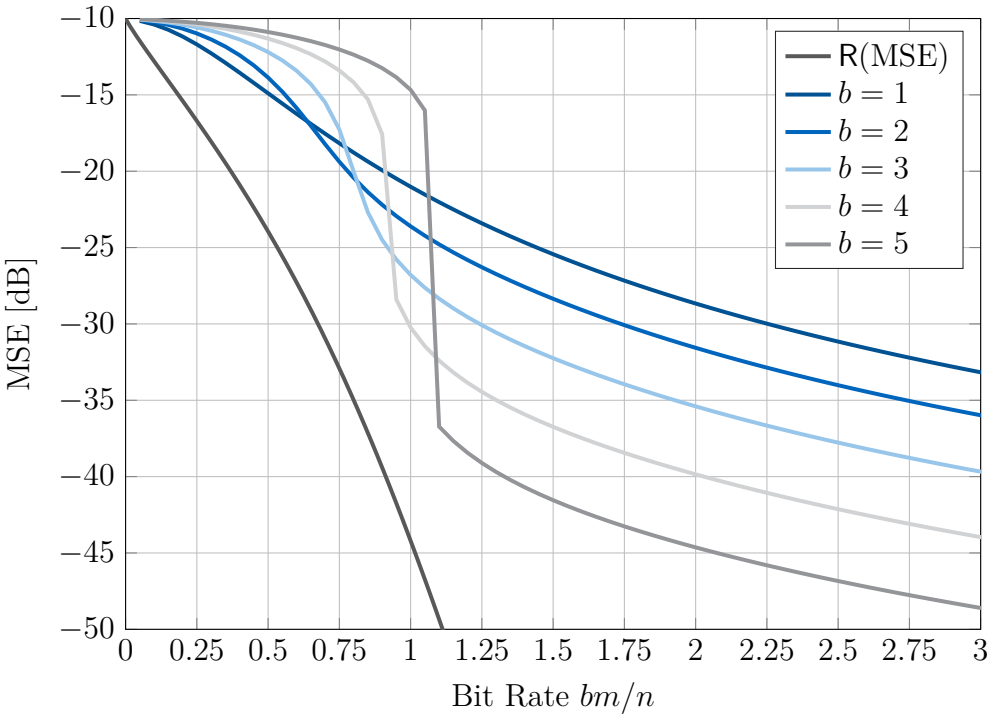


Figure 6.3.: GAMP performance as predicted by SE and empirically observed for a Bernoulli-Gaussian source with $p = 0.1$ with $n = 5000$. R(MSE) is the RD function computed with the Blahut-Arimoto algorithm (Algorithm 2.1).

6.2. Two-Terminal Bayesian Quantized Compressed Sensing

This section is based on joint work with Rami Ezzine and is presented in [Ezz18]. AMP was first extended to a distributed setting in [Hag14a, Hag14b] for unquantized two-terminal CS and termed *Multi-Terminal Approximate Message Passing (MAMP)*. In this section, we combine the GAMP and MAMP algorithms for the distributed problem that we investigated in Chapter 5. Consider the setting depicted in Figure 6.4.

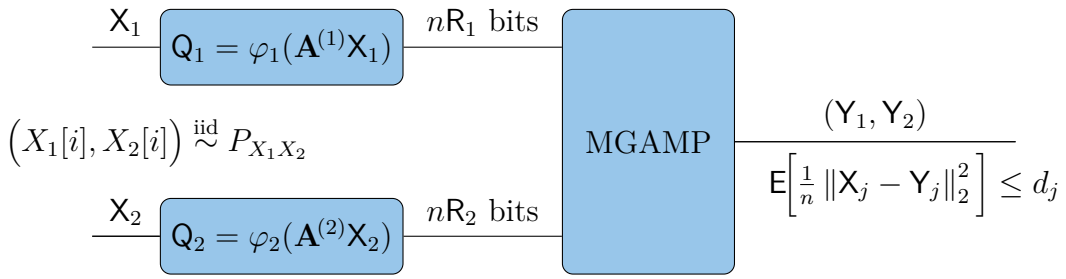


Figure 6.4.: System model for distributed CS with two terminals.

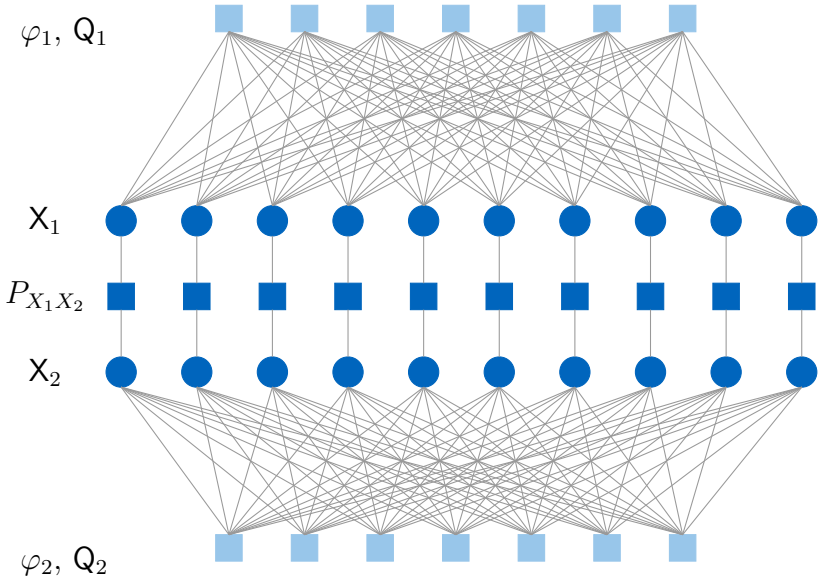


Figure 6.5.: Graphical model for two-terminal QCS. The light blue *factor nodes* represent the scalar quantizers and the observed quantized measurements at each terminal. The dark blue *variable nodes* represent the signal components of the two terminals and the knowledge of their joint distribution.

Formally, we have two generalized linear models

$$\begin{aligned} Q_1[k] &= \varphi_1\Big(\frac{1}{\sqrt{n}}\langle \mathsf{A}_k^{(1)}, \mathsf{X}_1\rangle\Big), \qquad 1 \le k \le m_1 \\ Q_2[k] &= \varphi_2\Big(\frac{1}{\sqrt{n}}\langle \mathsf{A}_k^{(2)}, \mathsf{X}_2\rangle\Big), \qquad 1 \le k \le m_2 \end{aligned} \tag{6.15}$$

where

- ▷ $(\mathsf{X}_1, \mathsf{X}_2)$ are output by a memoryless source with distribution $P_{X_1X_2}$,
- ▷ $\mathbf{A}^{(1)} \in \mathbb{R}^{m_1 \times n}$ and $\mathbf{A}^{(2)} \in \mathbb{R}^{m_2 \times n}$ are the measurement matrices, each with iid $\mathcal{N}(0,1)$ entries, and $\mathsf{A}_k^{(j)}$ is the transposed kth row of $\mathbf{A}^{(j)}$,
- ▷ $\varphi_1 : \mathbb{R} \to \mathcal{Q}_1$ and $\varphi_2 : \mathbb{R} \to \mathcal{Q}_2$ are two quantization functions with b_1 and b_2 bits, respectively.

The graphical model for this setting is depicted in Figure 6.5. We see that the two terminals are connected only via the knowledge of the joint distribution of the two signals. To get the *Multi-Terminal Generalized Approximate Message Passing (MGAMP)* reconstruction algorithm, we combine the GAMP and MAMP steps in an obvious way without giving any

formal derivations. To this end, recall the two scalar channels (6.5) and (6.8). The first channel was related to the quantization of the measurements. As this happens individually in the two terminals, those factor updates are also done individually in the MGAMP algorithm and we an reuse the functions $g_{P_{\text{out}}}$ and $h_{P_{\text{out}}}$ given in (6.6)-(6.7). For the additive noise channel in Eq. (6.8), we now have two parallel noise channels

$$\begin{aligned} \tilde{X}_1 &= X_1 + Z_1/\sqrt{\mathsf{snr}_1} \\ \tilde{X}_2 &= X_2 + Z_2/\sqrt{\mathsf{snr}_2} \end{aligned} \tag{6.16}$$

where $(X_1, X_2) \sim P_{X_1X_2}$ and Z_1 and Z_2 are independent of each other and (X_1, X_2), and each have distribution $\mathcal{N}(0,1)$. Define the functions $g^{(1)}_{P_{X_1X_2}}$, $g^{(2)}_{P_{X_1X_2}}$, $h^{(1)}_{P_{X_1X_2}}$, and $h^{(2)}_{P_{X_1X_2}}$ (all $\mathbb{R}^2 \to \mathbb{R}$) via

$$\begin{aligned} g^{(1)}_{P_{X_1X_2}}(\tilde{x}_1, \tilde{x}_2, \mathsf{snr}_1, \mathsf{snr}_2) &= \mathsf{E}[X_1 \,|\, \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2] \\ g^{(2)}_{P_{X_1X_2}}(\tilde{x}_1, \tilde{x}_2, \mathsf{snr}_1, \mathsf{snr}_2) &= \mathsf{E}[X_2 \,|\, \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2] \\ h^{(1)}_{P_{X_1X_2}}(\tilde{x}_1, \tilde{x}_2, \mathsf{snr}_1, \mathsf{snr}_2) &= \mathsf{Var}[X_1 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2] \\ h^{(2)}_{P_{X_1X_2}}(\tilde{x}_1, \tilde{x}_2, \mathsf{snr}_1, \mathsf{snr}_2) &= \mathsf{Var}[X_2 | \tilde{X}_1 = \tilde{x}_1, \tilde{X}_2 = \tilde{x}_2]. \end{aligned} \tag{6.17}$$

For vectors, these functions are again applied component-wise. The MGAMP algorithm is described more precisely in Algorithm 6.3. Similarly, the behavior of MGAMP can be predicted by its SE, which is given in Algorithm 6.4.

6.2.1. Numerical Example for Distributed Bernoulli-Gaussian Signals

As an example, and to compare with the RD results from Section 5, we perform MGAMP experiments and compute the SE predictions for a distributed Bernoulli-Gaussian spike source

$$P_{X_1X_2} = (1-p) \cdot \delta_0 + p \cdot \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right) \tag{6.18}$$

for some $\rho \in (-1, 1)$. The scalar quantizers φ_1 and φ_2 are again chosen to maximize the entropies of their outputs, i.e., they partition the real line into intervals of equal probability under the Gaussian measure. Let $\tilde{\mathsf{x}} = [\tilde{x}_1, \tilde{x}_2]^\mathsf{T}$. For this source, the estimation functions in (6.9) can be computed to be

$$g^{(1)}_{P_{X_1X_2}}(\tilde{x}_1, \tilde{x}_2, \mathsf{snr}_1, \mathsf{snr}_2) = \frac{1}{1 + \frac{(1-p)}{p} \frac{\mathcal{N}(\tilde{\mathsf{x}}; 0, \Sigma_0)}{\mathcal{N}(\tilde{\mathsf{x}}; 0, \Sigma_1)}} \cdot \begin{bmatrix} 1 & \rho \end{bmatrix} \Sigma_1^{-1} \tilde{\mathsf{x}} \tag{6.19}$$

Algorithm 6.3 MGAMP for QCS

Initialize: for $j = 1, 2$, set

$\mu_j = \mathsf{E}\left[X_j^2\right]$

$\mathsf{y}_j^0 = \mathsf{E}[\mathsf{X}_j]$

$\mathsf{v}_{x_j}^0 = \mathsf{Var}[\mathsf{X}_j]$

$\hat{\mathsf{s}}_j^0 = \mathsf{0}$

for $t = 1, 2, 3, \ldots$ **do**

Factor update: for $j = 1, 2$, set

$\mathsf{v}_{p_j}^t = \frac{1}{n}\left(\mathbf{A}^{(j)} \odot \mathbf{A}^{(j)}\right)\mathsf{v}_{x_j}^{t-1}$

$\hat{\mathsf{p}}^t = \frac{1}{\sqrt{n}}\mathbf{A}^{(j)}\mathsf{y}_j^{t-1} - \mathsf{v}_{p_j}^t \odot \hat{\mathsf{s}}_j^{t-1}$

$\hat{\mathsf{s}}_j^t = g_{P_{\text{out}}}(\mathsf{q}_j, \hat{\mathsf{p}}_j^t, \mathsf{v}_{p_j}^t; \varphi_j)$

$\mathsf{v}_{s_j}^t = h_{P_{\text{out}}}(\mathsf{q}_j, \hat{\mathsf{p}}_j^t, \mathsf{v}_{p_j}^t; \varphi_j)$

Variable update:

Linear step: for $j = 1, 2$, set

$\mathsf{v}_{r_j}^t = \frac{1}{n}\left(\mathbf{A}^{(j)} \odot \mathbf{A}^{(j)}\right)^{\mathsf{T}}\mathsf{v}_{s_j}^t$

$\hat{\mathsf{r}}_j^t = \mathsf{y}_j^{t-1} + (\mathsf{v}_{r_j}^t)^{-1} \odot \left(\frac{1}{\sqrt{n}}\mathbf{A}^{(j)\mathsf{T}}\hat{\mathsf{s}}_j^t\right)$

Nonlinear step: for $j = 1, 2$, set

$\mathsf{y}_j^t = g_{P_{X_1X_2}}^{(j)}(\hat{\mathsf{r}}_1^t, \hat{\mathsf{r}}_2^t, \mathsf{v}_{r_1}^t, \mathsf{v}_{r_2}^t)$

$\mathsf{v}_{x_j}^t = h_{P_{X_1X_2}}^{(j)}(\hat{\mathsf{r}}_1^t, \hat{\mathsf{r}}_2^t, \mathsf{v}_{r_1}^t, \mathsf{v}_{r_2}^t)$

end for

return y_1^t, y_2^t

Algorithm 6.4 MGAMP State Evolution for QCS

Initialize: for $j = 1, 2$, set

$\mu_j = \mathsf{E}\left[X_j^2\right]$

$\eta_j^0 = 0$

for $t = 1, 2, 3, \ldots$ **do**

Factor update: for $j = 1, 2$, set

$\mathsf{snr}_j^t = \alpha_j\, \mathsf{E}_{QY}\left[h_{P_{\text{out}}}(Q, V, \mu_j - \eta_j^t; \varphi_j)\right]$

Variable update: for $j = 1, 2$, set

$\eta_i^t = \mu_j - \mathsf{E}_{\tilde{X}_1\tilde{X}_2}\left[h_{P_{X_1X_2}}^{(j)}\left(\tilde{X}_1, \tilde{X}_2, \mathsf{snr}_1^t, \mathsf{snr}_2^t\right)\right]$

end for

return MSE $\mu_j - \eta_j^t$ for $j = 1, 2$

and

$$h_{P_{X_1X_2}}^{(1)}(\tilde{x}_1, \tilde{x}_2, \mathsf{snr}_1, \mathsf{snr}_2)$$

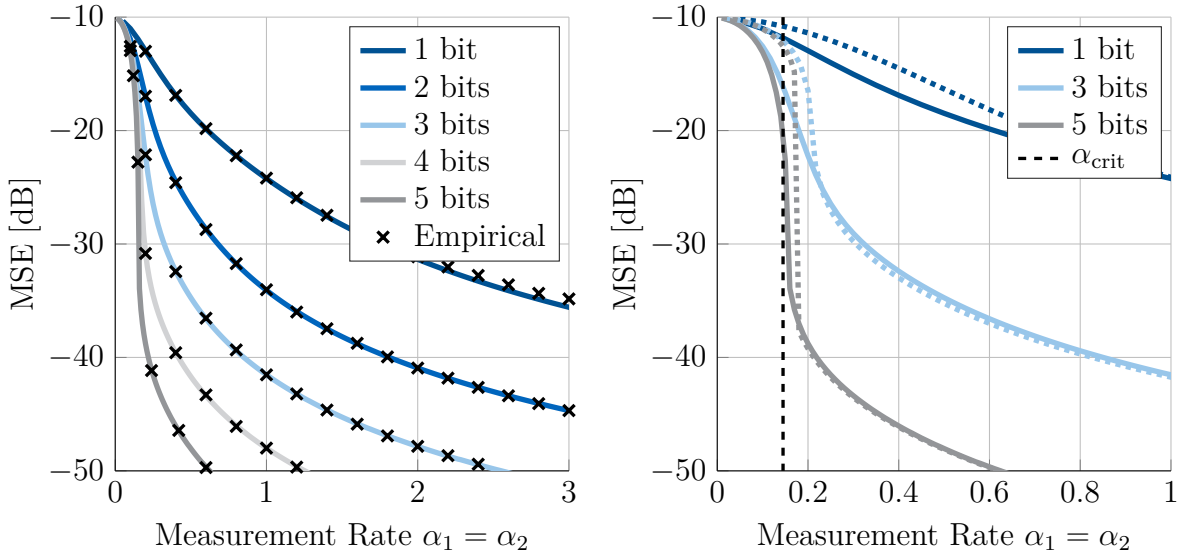
$$= \frac{1 - \begin{bmatrix}1 & \rho\end{bmatrix}\Sigma_1^{-1}\begin{bmatrix}1\\ \rho\end{bmatrix} + \left(\begin{bmatrix}1 & \rho\end{bmatrix}\Sigma_1^{-1}\tilde{\mathsf{x}}\right)^2}{1 + \frac{(1-p)}{p}\frac{\mathcal{N}(\tilde{\mathsf{x}};0,\Sigma_0)}{\mathcal{N}(\tilde{\mathsf{x}};0,\Sigma_1)}} - g_{P_{X_1X_2}}^{(1)}(\tilde{x}_1, \tilde{x}_2, \mathsf{snr}_1, \mathsf{snr}_2)^2 \quad (6.20)$$

where

$$\Sigma_0 = \begin{bmatrix} 1/\mathsf{snr}_1 & 0 \\ 0 & 1/\mathsf{snr}_2 \end{bmatrix} \quad \text{and} \quad \Sigma_1 = \begin{bmatrix} 1+1/\mathsf{snr}_1 & \rho \\ \rho & 1+1/\mathsf{snr}_2 \end{bmatrix}. \tag{6.21}$$

The functions $g^{(2)}_{P_{X_1X_2}}$ and $h^{(2)}_{P_{X_1X_2}}$ are computed similarly. Since the functions $g_{P_{\text{out}}}$ and $h_{P_{\text{out}}}$ depend only on the quantizer and are computed individually in the two terminals, we can reuse (6.13) - (6.14).

For our experiments, we chose the measurement rates and quantizers to be the same at both terminals. Thus, the average MSE is also the same at both terminals. Figure 6.6(a) plots the SE and experimental results for $p = 0.1$, $n = 5000$ and the correlation coefficient $\rho = 0.9$. Observe that SE again accurately predicts the experimental performance. Figure 6.6(b) compares the SE predictions for $\rho = 0.9$ (solid lines) and $\rho = 0$ (dotted lines). Observe that for small measurement rates, a high correlation can be exploited to reduce the estimation error. For larger rates, the performance is nearly identical in both cases. Observe also that the phase transition is reduced to a measurement rate of approximately $\alpha_{\text{crit}} \approx 0.15$ (as compared to $\alpha_{\text{crit}} \approx 0.21$ in Fig. 6.2) at each terminal. This is in line with our results in Chapter 7 where we show that in a distributed setting, exploiting the joint sparsity of several signals helps to reduce the measurement rate to the sparsity ($p = 0.1$ in this case).



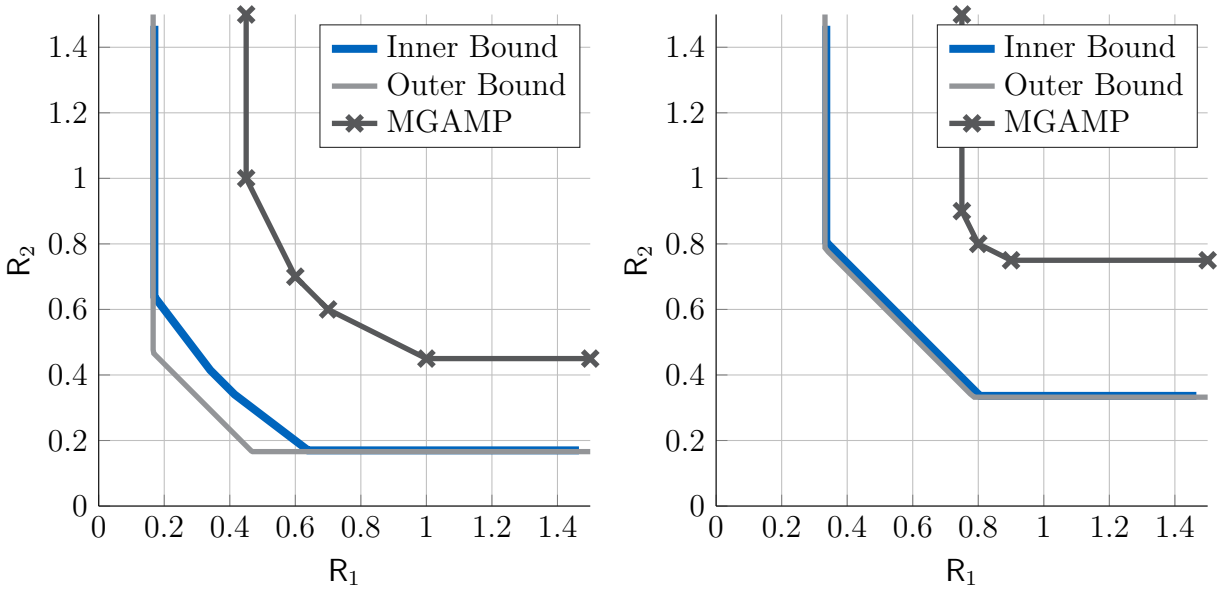
(a) $p = 0.1$, $\rho = 0.9$.

(b) $\rho = 0.9$ corresponds to the solid line and $\rho = 0$ to the dotted lines.

Figure 6.6.: Comparison of SE and empirical performance for MGAMP with equal measurement rates and quantizers at both terminal. The MSE is the same at both terminals.

In a second experiment, we computed the state evolution for many different choices of α

and b in both terminals and collected the achievable rate pairs (in terms of total number of bits stored) for $p = 0.1$, $\rho = 0$, and two distortion levels of -20 dB and -30 dB. Figure 6.7 compares these to the inner and outer bounds for the rate region of the DBGS developed in Chapter 5.



(a) $d_1 = d_2 = -20$ dB. (b) $d_1 = d_2 = -30$ dB.

Figure 6.7.: Comparison of MGAMP with the inner and outer bounds for the rate region of the distributed Bernoulli-Gaussian spike source (see Chapter 5) for $p = 0.1$ and $\rho = 0$.

6.3. Information Rates and Optimal Errors

In this section, we investigate information-theoretic limits for QCS based on the recent work by Barbier et al. [BKM+19]. There, the authors consider the *Generalized Linear Model (GLM)*

$$Z_k \sim P_{\text{out}}\Big(\cdot \,\Big|\, \frac{1}{\sqrt{n}} \langle \mathsf{A}_k, \mathsf{X} \rangle \Big), \qquad 1 \leq k \leq m \tag{6.22}$$

where the entries of X are iid samples from some distribution P_X, $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a dense random matrix with independent entries of unit variance and P_{out} represents an output channel that may include a function, noise, or other effects. This model covers both *estimation* and *learning* problems, where in the first we view X as a *signal* to be recovered from the *measurements* Z and in the second we would like to predict new labels Z_{new} when adding new rows to the matrix $\mathbf{A}$. We focus on the first view point that models QCS.

Recall the QCS model (6.1) with Q being the quantized measurements. One insight from [BKM+19] is that GLMs of the form (6.1) can asymptotically be precisely characterized in terms of the mutual information $\frac{1}{n}I(\mathsf{X};\mathsf{Q}|\,\mathbf{A})$ as well as $\frac{1}{n}\mathsf{mmse}(\mathsf{X}|\mathsf{Q},\mathbf{A})$. Because we have $I(\mathsf{X};\mathsf{Q}|\,\mathbf{A}) = H(\mathsf{Q}\,|\mathbf{A})$ for noiseless QCS, Shannon's source coding theorem [Sha48] tells us that we can losslessly compress the measurements to $\frac{1}{n}H(\mathsf{Q}\,|\mathbf{A})$ bits per signal dimension. This allows us to compare the optimal performance of GLMs with RD bounds that characterize the optimal performance of *any* encoding/decoding scheme.

To state the results from [BKM+19], we return to the two scalar channels (6.5) and (6.8). For the first channel (6.5) associated with the quantizers, we denote the mutual information of Q and W given V by

$$\begin{aligned} I_\varphi(\eta) &\coloneqq I\Big(W; \varphi\Big(\sqrt{\eta}V + \sqrt{\mu-\eta}W\Big)\Big|\, V\Big) \\ &= H\Big(\varphi\Big(\sqrt{\eta}V + \sqrt{\mu-\eta}W\Big)\,|V\Big)\,. \end{aligned} \tag{6.23}$$

For the second channel, the additive noise channel (6.8), we denote the mutual information between X and $\tilde{X}$ by

$$I_{P_X}(\mathsf{snr}) \coloneqq I\Big(X;\tilde{X}\Big) = h(P_{\tilde{X}}) - \frac{1}{2}\log(2\pi e/\mathsf{snr}). \tag{6.24}$$

We next define the *replica-symmetric information* as

$$i_{\mathrm{RS}}(\eta,\mathsf{snr};\mu) \coloneqq I_{P_X}(\mathsf{snr}) + \alpha I_\varphi(\eta) - \frac{\mathsf{snr}(\mu-\eta)}{2} \tag{6.25}$$

for some measurement rate $\alpha > 0$ and denote its *fixed points* by

$$\Gamma \coloneqq \left\{(\eta,\mathsf{snr}) \in [0,\mu]\times[0,\infty] \;\middle|\; \begin{array}{rcl} \eta & = & \mu - 2I'_{P_X}(\mathsf{snr}) \\ \mathsf{snr} & = & 2\alpha I'_\varphi(\eta) \end{array}\right\} \tag{6.26}$$

where $I'_{P_X}(\mathsf{snr})$ and $I'_\varphi(\eta)$ denote the derivatives.

We now state the asymptotic expressions for the average entropy of the quantized measurements as well as the MMSE, tailored to the setting of QCS.

Theorem 6.1 (Barbier et al. [BKM+19])**.** Suppose the following conditions hold:

- ▷ P_X admits a finite third moment and $\#(\mathrm{supp}(P_X)) \geq 2$.
- ▷ The entries of $\mathcal{A}$ are independent random variables with zero mean, unit variance and finite third moment that is bounded by n.
- ▷ $\varphi : \mathbb{R} \to \mathcal{Q}$ quantizes to at least one bit, i.e., $\#(\mathcal{Q}) \geq 2$, and
- ▷ $m/n \to \alpha > 0$ as $m, n \to \infty$.

Then the limits of the entropy and the MMSE are

$$\frac{1}{n}H(\mathsf{Q}\,|\mathbf{A}\,) \xrightarrow{n\to\infty} \inf_{\eta\in[0,\mu]} \sup_{\mathsf{snr}\geq 0} i_{\mathrm{RS}}(\eta,\mathsf{snr};\mu) = \inf_{(\eta,\mathsf{snr})\in\Gamma} i_{\mathrm{RS}}(\eta,\mathsf{snr};\mu) \tag{6.27}$$

$$\frac{1}{n}\mathsf{mmse}(\mathsf{X}|\mathsf{Q},\mathbf{A}) \xrightarrow{n\to\infty} \mu-\eta^\star(\alpha) \tag{6.28}$$

where $\eta^\star(\alpha)$ is the unique minimizer of $i_{\mathrm{RS}}(\eta,\mathsf{snr};\mu)$ in (6.27).

Observe that by Theorem 6.1 and the properties of SE, GAMP achieves the optimal MSE whenever $\lim_{t\to\infty}\eta_{\mathrm{SE}}^t=\eta^\star(\alpha)$. In Sections 6.3.1 and 6.3.2, we evaluate (6.27) and (6.28) for Bernoulli-Gaussian spike sources and compare them to RD bounds.

6.3.1. Bernoulli-Gaussian

To compare the limits of QCS with RD bounds, we consider the Bernoulli-Gaussian spike source

$$P_X=(1-p)\delta_0+p\mathcal{N}(0,1) \tag{6.29}$$

and compute the mutual information for each of the two scalar inference problems (6.24) and (6.23). By (6.8), the PDF of $\tilde{X}$ is given by

$$P_{\tilde{X}}=(1-p)\cdot\mathcal{N}(0,1/\mathsf{snr})+p\cdot\mathcal{N}(0,1+1/\mathsf{snr}) \tag{6.30}$$

and we compute

$$\begin{aligned} I_{P_X}(\mathsf{snr}) &= h(P_{\tilde{X}})-\frac{1}{2}\log(2\pi e/\mathsf{snr}) \\ &= \mathsf{E}\left[-\log\left((1-p)+\frac{p}{\sqrt{1+\mathsf{snr}}}e^{\frac{\mathsf{snr}^2\tilde{X}^2}{2(1+\mathsf{snr})}}\right)\right]+\frac{\mu\cdot\mathsf{snr}}{2}. \end{aligned} \tag{6.31}$$

It remains to compute $I_\varphi(\eta)$, which depends on the quantizer. For a 1-bit quantizer without dithering, i.e., $\varphi(z)=\mathrm{sign}(z)$, let $\tilde{W}\sim\mathcal{N}(0,1)$ independently of V. We have

$$\begin{aligned} I_\varphi(\eta)^{1\mathrm{bit}} &= \mathsf{E}_{V,Q}\left[-\log\int\frac{e^{-w^2/2}}{\sqrt{2\pi}}\mathbb{1}_{\left\{Q=\varphi\left(\sqrt{\eta}V+\sqrt{\mu-\eta}w\right)\right\}}\mathrm{d}w\right] \\ &= \mathsf{E}_{V,\tilde{W}}\left[-\log\int\frac{e^{-w^2/2}}{\sqrt{2\pi}}\mathbb{1}_{\left\{\varphi\left(\sqrt{\eta}V+\sqrt{\mu-\eta}\tilde{W}\right)=\varphi\left(\sqrt{\eta}V+\sqrt{\mu-\eta}w\right)\right\}}\mathrm{d}w\right] \\ &= \mathsf{E}_V\left[H_2\left(\Phi\left(\frac{-\sqrt{\eta}V}{\sqrt{\mu-\eta}}\right)\right)\right] \end{aligned} \tag{6.32}$$

where Φ is the cumulative distribution function of a standard Gaussian random variable. For a general b-bit quantizer, we denote the interval boundaries by $\{\tau_0,\tau_1,\ldots,\tau_{2^b}\}$ and let

$\tau_0 = -\infty$ and $\tau_{2^b} = \infty$. We compute

$$
\begin{aligned}
&I_\varphi(\eta)^{b\,\mathrm{bit}} \\
&= \sum_{\ell=0}^{2^b-1} \mathsf{E}_{V\tilde{W}}\left[-\mathbb{1}_{\{\sqrt{\eta}V+\sqrt{\mu-\eta}\tilde{W}\in[\tau_\ell,\tau_{\ell+1})\}} \log \int \frac{e^{-w^2/2}}{\sqrt{2\pi}} \mathbb{1}_{\{\sqrt{\eta}V+\sqrt{\mu-\eta}w\in[\tau_\ell,\tau_{\ell+1})\}} \mathrm{d}w\right] \qquad (6.33)\\
&= \sum_{\ell=1}^{2^b-1} \mathsf{E}_V\left[-\left(\Phi\left(\frac{\tau_\ell-\sqrt{\eta}V}{\sqrt{\mu-\eta}}\right)-\Phi\left(\frac{\tau_{\ell-1}-\sqrt{\eta}V}{\sqrt{\mu-\eta}}\right)\right)\log\left(\Phi\left(\frac{\tau_\ell-\sqrt{\eta}V}{\sqrt{\mu-\eta}}\right)-\Phi\left(\frac{\tau_{\ell-1}-\sqrt{\eta}V}{\sqrt{\mu-\eta}}\right)\right)\right].
\end{aligned}
$$

In Figure 6.8, we plot the error predicted by the GAMP SE, as well as the MMSE computed from Theorem 6.1. For comparison, we plot the RD function $\mathsf{R}(\mathrm{MSE})$ of the source that was computed with the Blahut-Arimoto algorithm (Algorithm 2.1). The results indicate that GAMP achieves the MMSE for low-resolution QCS (see $b = 1$ and $b = 3$) but has a suboptimal phase for high-resolution QCS when $\alpha \approx \alpha_{\mathrm{crit}}$ (see $b = 5$). The suboptimality should not be surprising as we know (see, e.g. [KMS+12b, KMS+12a]) that AMP is suboptimal for noiseless (i.e., the limit of $b \to \infty$) CS.

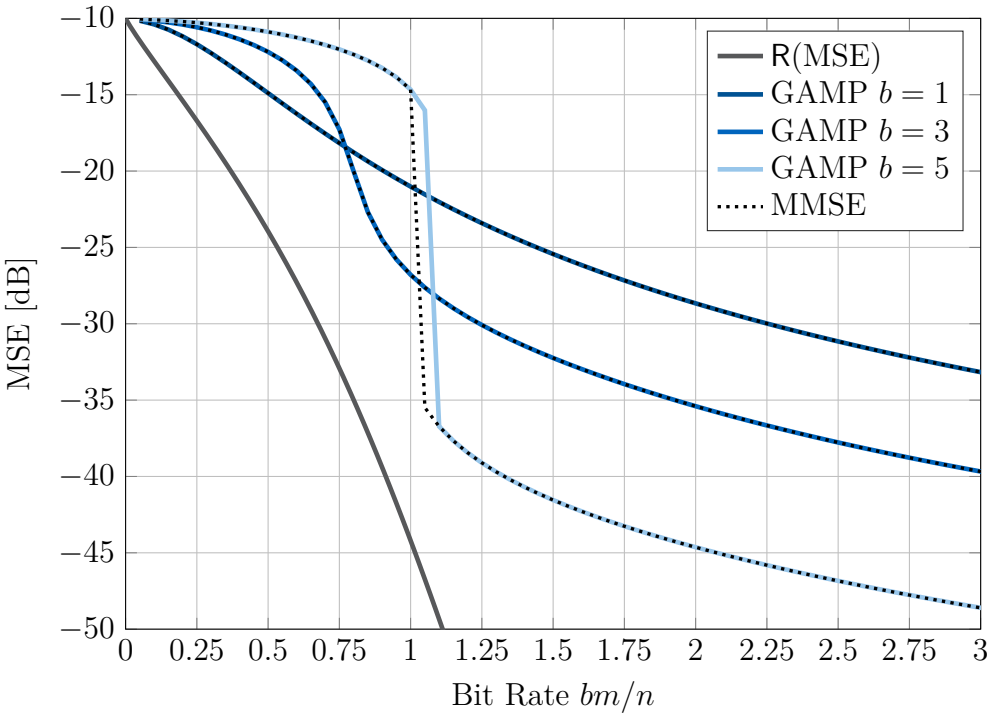


Figure 6.8.: GAMP SE versus MMSE for the Bernoulli-Gaussian source with $p = 0.1$.

Next, Figure 6.9 shows the asymptotic entropy of the quantizer outputs, i.e., the minimum bit rate after optimally compressing the quantized measurements, versus the RD function. First, we observe that optimally compressing the quantizer outputs significantly improves the RD trade-off of QCS. We further see that after compression, one-bit mea-

surements seem to have the best RD trade-off for any fixed bit rate. For higher resolution quantizers, this performance is matched for sufficiently large rates.

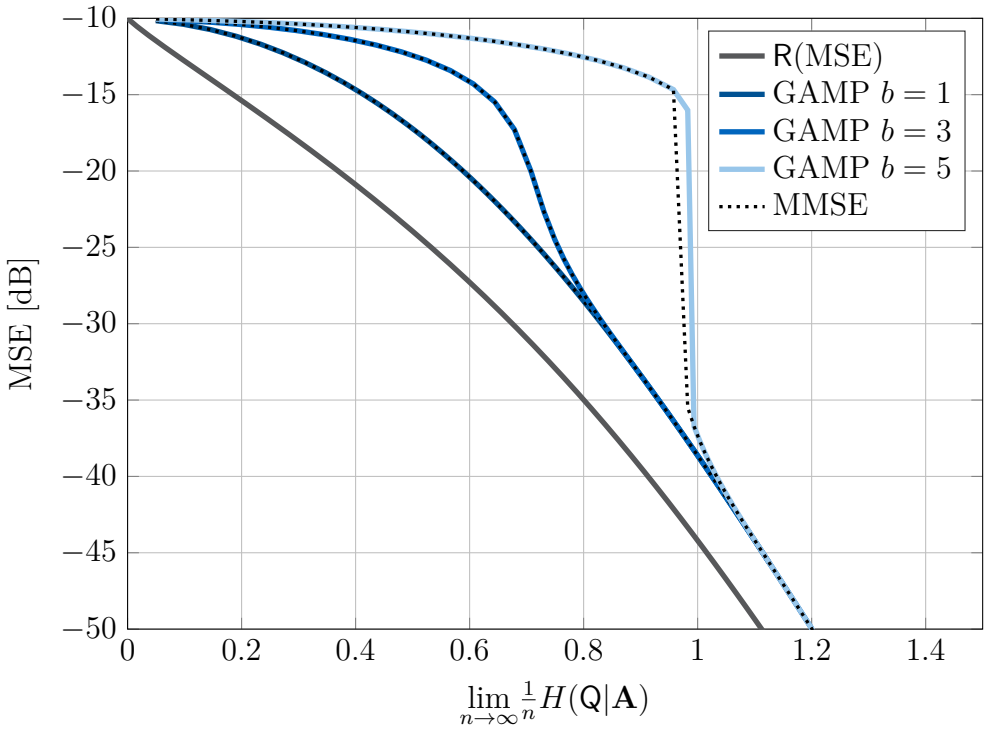


Figure 6.9.: RD performance of QCS with optimal lossless compression.

We remark, however, that finding an optimal compression algorithm for the quantized measurements may not be an easy task. To illustrate this, Figure 6.10 plots the entropy of the measurements versus the measurement rate α for three different quantizers. We observe that up to approximately α_{crit} (the critical measurement rate at which the error starts to drop significantly, cf. Figure 6.2), the measurements have nearly maximum entropy, thanks to the scalar quantizer being chosen to maximize the entropy of a single measurement. As a consequence, up to this point the measurements are nearly independent. Any useful compression algorithm must therefore exploit a dependence between the measurements that is somehow spread over at least $n\alpha_{\text{crit}}$ measurements.

6.3.2. Bernoulli-Gaussian with Thresholding

An important insight from the information-theoretic analysis of spike sources in [WV12a] and Chapter 4 is that one can efficiently code such sources by distinguishing between *significant* and *insignificant* samples. Setting the insignificant samples to zero helps because this reduces the information needed to code their positions and thus leaves more bits to code the values. In a CS setting, this would correspond to adding a filter *before* measuring the signal to zero out small magnitudes and keep larger ones. This creates a sparser signal

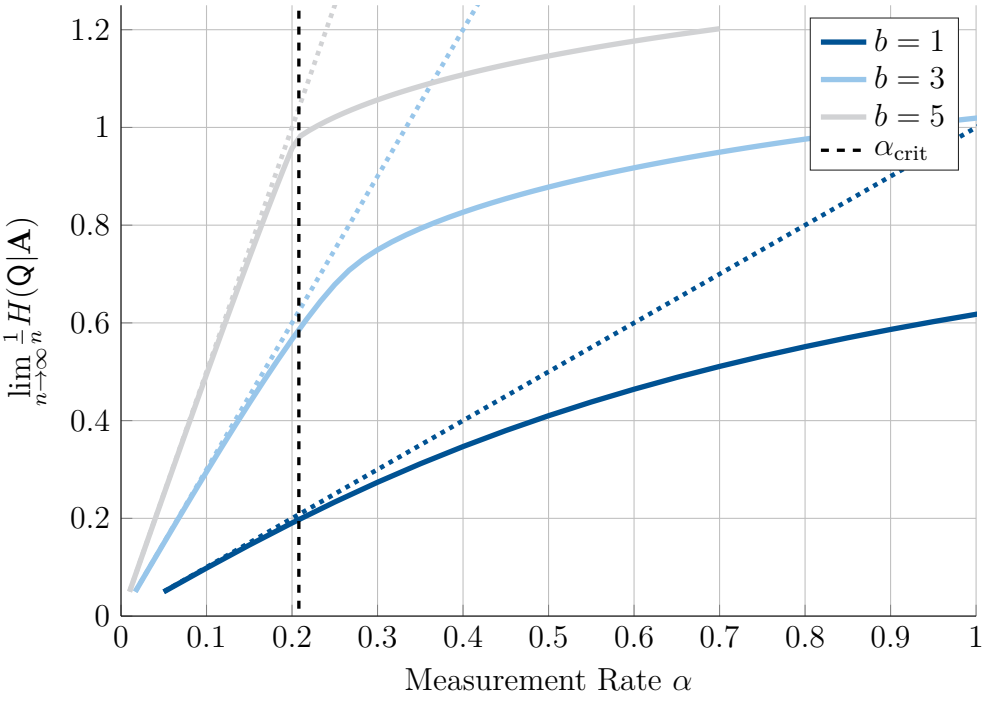


Figure 6.10.: Measurement rate versus entropy of the measurements in QCS for a Bernoulli-Gaussian source. The dotted lines show the uncompressed bit rates αb of the measurements.

that needs fewer measurements to approximate the signal well. This section investigates the RD trade-off when adding such a filter before the measurements.

Let $\mathsf{X} \overset{\text{iid}}{\sim} P_X$ be a Bernoulli-Gaussian source and define

$$X^\dagger[i] = X[i] \cdot \mathbb{1}_{\{|X[i]|>\tau\}}, \qquad 1 \leq i \leq n \tag{6.34}$$

for some threshold $\tau \geq 0$. $\mathsf{X}^\dagger$ is then the input to a QCS system and its estimate (via GAMP or an MMSE estimator) is used as an estimate for X. We will determine the asymptotic limits of $\frac{1}{n} H(\mathsf{Q}\,|\mathbf{A}\,)$ and $\frac{1}{n}\mathsf{mmse}(\mathsf{X}^\dagger|\mathsf{Q},\mathbf{A})$ as given by Theorem 6.1. These two quantities together provide achievable RD pairs for X since

$$\begin{aligned}\mathsf{E}\Big[\|\mathsf{X} - \mathsf{E}[\mathsf{X}^\dagger\,|\,\mathsf{Q},\mathbf{A}]\|^2\Big] &\leq \mathsf{E}\Big[\|\mathsf{X} - \mathsf{X}^\dagger\|^2\Big] + \mathsf{E}\Big[\|\mathsf{X}^\dagger - \mathsf{E}[\mathsf{X}^\dagger\,|\,\mathsf{Q},\mathbf{A}]\|^2\Big] \\ &= \sum_{i=1}^{n} \mathsf{E}\Big[\Big(X[i] \cdot \mathbb{1}_{\{|X[i]|\leq\tau\}}\Big)^2\Big] + \mathsf{mmse}(\mathsf{X}^\dagger|\mathsf{Q},\mathbf{A})\end{aligned} \tag{6.35}$$

and

$$I(\mathsf{X};\mathsf{Q}|\,\mathbf{A}) = I\Big(\mathsf{X}^\dagger;\mathsf{Q}\Big|\,\mathbf{A}\Big) \tag{6.36}$$

because of the Markov chain X–$\mathsf{X}^\dagger$–$\mathsf{Q}|\mathbf{A}$. Let $p^\dagger \coloneqq \Pr[|X| > \tau] = 2p\Phi(-\tau)$. We consider $X^\dagger[i] \overset{\text{iid}}{\sim} P_{X^\dagger}$ as the new source for our QCS system with

$$P_{X^\dagger} = (1-p^\dagger)\cdot\delta_0 + p\cdot\mathcal{N}(0,1)\cdot\mathbb{1}_{\mathcal{T}} \tag{6.37}$$

where $\mathbb{1}_{\mathcal{T}}$ is one for all values with magnitudes larger than τ and zero otherwise. Using [Tur10], the new PDF for $\tilde{X} = X^\dagger + N/\sqrt{\mathsf{snr}}$ is

$$P_{\tilde{X}}(\tilde{x}) = (1-p^\dagger)\cdot\mathcal{N}(\tilde{x};0,1) + \frac{p^\dagger\cdot e^{-\frac{\mathsf{snr}\cdot\tilde{x}^2}{2(1+\mathsf{snr})}}}{\sqrt{2\pi\,\frac{1+\mathsf{snr}}{\mathsf{snr}}}}\,\frac{g_1(\tilde{x})+g_2(\tilde{x})}{2} \tag{6.38}$$

where

$$\begin{aligned} g_1(\tilde{x}) &\coloneqq \frac{1}{1-\Phi(\tau)}\left[1-\Phi\left(\frac{\frac{\mathsf{snr}\cdot\tilde{x}}{1+\mathsf{snr}}+\tau}{1/\sqrt{1+\mathsf{snr}}}\right)\right] \\ g_2(\tilde{x}) &\coloneqq \frac{1}{\Phi(-\tau)}\Phi\left(\frac{\frac{\mathsf{snr}\cdot\tilde{x}}{1+\mathsf{snr}}-\tau}{1/\sqrt{1+\mathsf{snr}}}\right). \end{aligned} \tag{6.39}$$

Since $I_\varphi(q)$ depends only on the quantizer, we can reuse (6.32)-(6.32) and only need to adapt $I_{P_X}(\mathsf{snr})$ to the new source distribution. We have

$$\begin{aligned} I_{P_{X^\dagger}}(\mathsf{snr}) &= h(P_{\tilde{X}}) - \frac{1}{2}\log(2\pi e/\mathsf{snr}) \\ &= \mathsf{E}_{\tilde{X}}\left[-\log\left((1-p^\dagger) + \frac{\tilde{p}}{\sqrt{1+\mathsf{snr}}}e^{\frac{\mathsf{snr}^2\tilde{X}^2}{2(1+\mathsf{snr})}}\,\frac{g_1(\tilde{X})+g_2(\tilde{X})}{2}\right)\right] + \frac{\mu\,\mathsf{snr}}{2}. \end{aligned} \tag{6.40}$$

Similar to Figure 6.8, we plot the bit rate versus the MSE (according to (6.35)) achieved by GAMP and the MMSE for different quantizers in Figure 6.11. For this plot, we computed the error for many different thresholds τ at each rate and chose the lowest overall error among those. Observe that while GAMP still achieves the MMSE performance for $b = 1$, it is suboptimal for a larger range of rates and quantizers compared to the regular Bernoulli-Gaussian source. The RD performance for the MMSE estimator is significantly improved compared to Figure 6.8.

Finally, we compare the RD performance of optimally compressed 1-bit measurements with and without filtering in Figure 6.12. We see that filtering out the samples of low magnitude significantly improves the RD performance at lower rates, but the improvement is reduced at larger rates. It would be interesting to understand this theoretically.

6.3.3. Summary and Discussion

In this chapter, we investigated the RD behavior of (multi terminal) Bayesian QCS with dense random matrices. We started by applying the GAMP algorithm to QCS as in [KGR12].

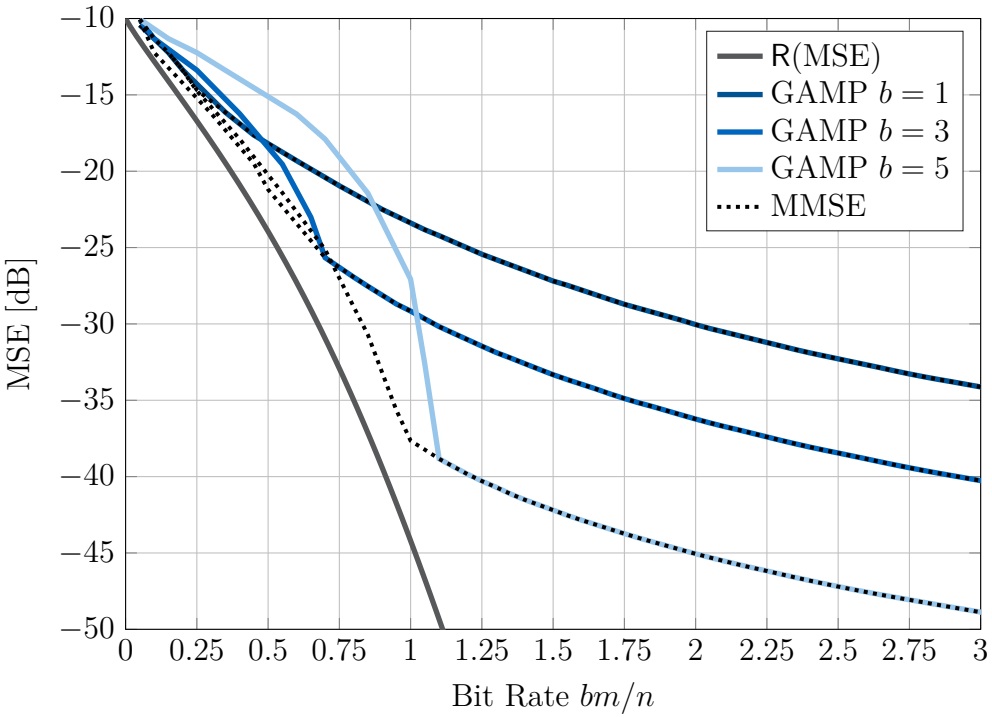


Figure 6.11.: GAMP SE versus MMSE in QCS when filtering the insignificant samples before measuring.

There, however, it is written that *"we are not advocating quantized linear expansions as a compression technique"* due to the unfavorable RD behavior for both small measurement rates (the *undersampled* case) and large measurement rates (the *oversampled* case) as observed in Figure 6.3 and in the two-terminal case in Figure 6.6. This led us to pose the following three questions about possible improvements which we could partially answer for a single terminal with the help of recent results by Barbier et al. [BKM+19].

1) *When is GAMP equal to the MMSE estimator?* For 1-bit QCS, GAMP is optimal with respect to the MSE in all settings the we computed. For larger quantizer depths, GAMP becomes suboptimal around the phase transition for the Bernoulli-Gaussian source and suboptimal for rates up to the phase transition for the truncated Bernoulli-Gaussian source. We remark that while finding the MMSE estimator in such cases may be extremely difficult, there are settings in which structured sensing matrices have shown to be superior to dense matrices with independent entries, see [DJM13, KMS+12a, BSK15].

2) *How well can we compress the quantized measurements?* We computed the entropies of the quantizer outputs in various settings and found the RD performance with optimal lossless compression of the quantizer outputs. This shows that QCS systems with lossless compression can exhibit an excellent RD trade-off.

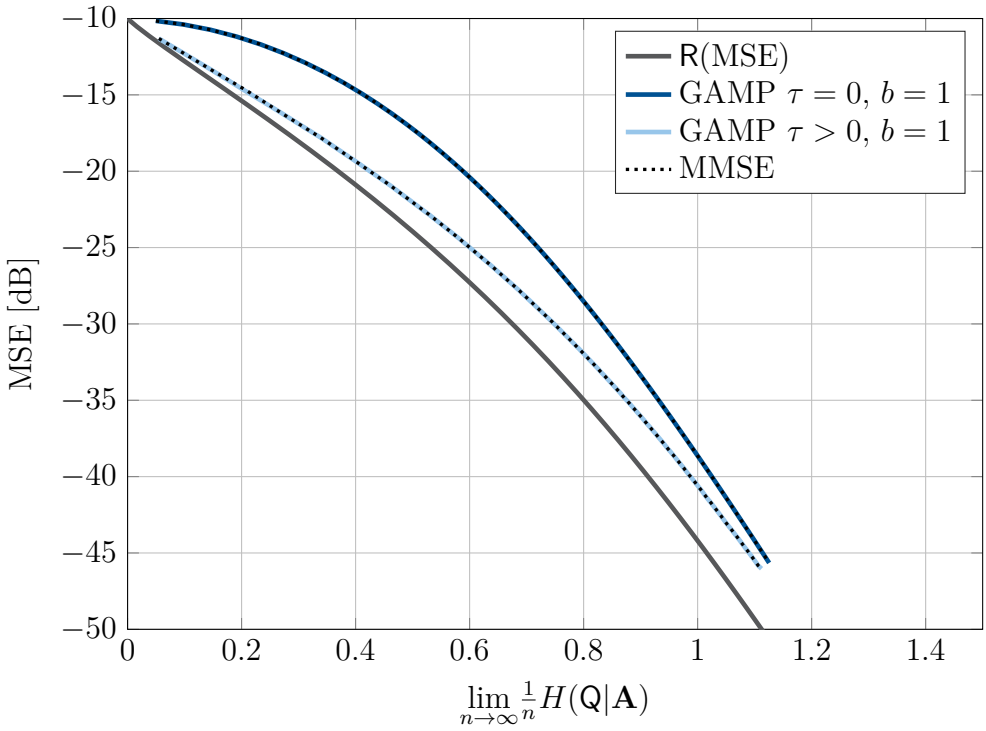


Figure 6.12.: Comparison of the RD performance of one-bit CS with and without filtering with optimal lossless compression.

3) *Can we add some helpful preprocessing?* We proposed to add a filter before measuring the signal that zeros insignificant samples. If physically possible, this can significantly improve the RD performance, especially at low rates.

Part III.
Uniform Approximation in Compressed Sensing

7

Analysis of Hard-Thresholding for Distributed Compressed Sensing with One-Bit Measurements

All results and numerics in this chapter are joint work with Johannes Maly and have been published in [MP19].

In this chapter, we investigate a *Distributed Compressed Sensing* (see [BDW+09], [DSW+05]) setting with one-bit measurements. Consider the setting depicted in Figure 7.1.

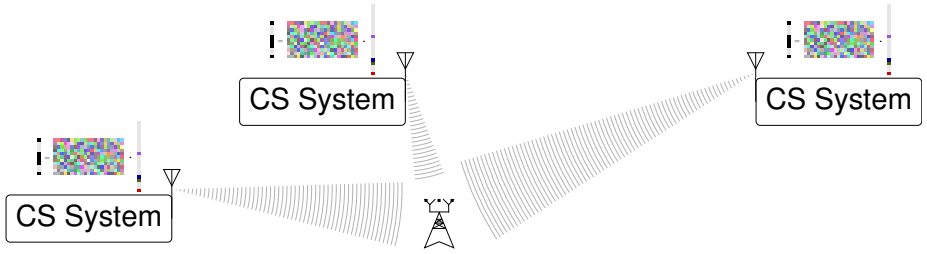


Figure 7.1.: Distributed CS system.

The system consists of several wireless sensor nodes that each measure a signal using QCS techniques. The nodes do not need to reconstruct their signal individually but send their acquired information to a central processing unit that reconstructs all signals *jointly*.

Recall from the introduction in Chapter 2.2 that, knowing that a signal of interest x is s-sparse, all such signals can be reconstructed from

$$m \geq Cs \log\left(\frac{en}{s}\right) \tag{7.1}$$

linear measurements of the form

$$\mathsf{z} = \mathbf{A}\mathsf{x} \tag{7.2}$$

where $C > 0$ is a constant independent of s, m, and n (see [CT06a], [RV08]). This scaling also carries over to one-bit measurements of the form

$$\mathsf{q} = \operatorname{sign}(\mathbf{A}\mathsf{x}) \tag{7.3}$$

with the difference that the scaling $m \geq C\delta^{-\alpha}\, s \log(en/s)$ now includes an accuracy parameter $\delta^{-\alpha}$ that guarantees a certain reconstruction error and may depend on the exact geometry of the signal set and the reconstruction algorithm in use.

The main idea in this chapter is that the log-factor in (7.1) is caused by not knowing the support of x. This is intuitive because if we know the support of x, we can simply reconstruct it using the sub-matrix of $\mathbf{A}$ consisting of only those columns corresponding to the nonzero entries of x. If one must recover several signals $\mathsf{x}_1, ..., \mathsf{x}_L$, $L \in \mathbb{N}$, sharing a common support, it might be possible to reduce the number of measurements per signal from $\mathcal{O}(s \log(en/s))$ to $\mathcal{O}(s)$ by exploiting the joint structure. In theory the improvement seems small, but in practice it can make a notable difference (cf. [SCS14]). Moreover, the common support might appear naturally. For example, a signal that is sparse in the Fourier basis may be measured at different locations, which leads to different attenuations and phase shifts at every node. This can be exploited in imaging applications such as *Magnetic Resonance Imaging* [WZT+14]. Another prominent application is *Multiple-Input and Multiple-Output* communications [RL14].

There are two popular settings for joint recovery from compressed measurements. The first model is called *Multiple Measurement Vectors*. All signals are measured by the same measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ (resp. the same sensor) and the model in (2.6) becomes

$$\mathbf{Z} = \mathbf{A}\mathbf{X} \tag{7.4}$$

where $\mathbf{X} \in \mathbb{R}^{n \times L}$ and $\mathbf{Z} \in \mathbb{R}^{m \times L}$ are matrices containing the signals and their corresponding measurement vectors as columns. As shown in [ER10], for this model one can improve only the average performance as compared to single vector CS. The worst-case analysis shows no improvement.

The second model considers distinct measurement matrices $\mathbf{A}^{(1)}, ..., \mathbf{A}^{(L)} \in \mathbb{R}^{m \times n}$ (resp. distinct sensors) for each signal $\mathsf{x}_l \in \mathbb{R}^n$, $l \in [L]$. Hence, there is a separate measurement process of type (2.8) for each $l \in [L]$ yielding L different $\mathsf{y}_l \in \mathbb{R}^m$. We have

$$\operatorname{vec}(\mathbf{Z}) = \mathbf{A} \cdot \operatorname{vec}(\mathbf{X}) \tag{7.5}$$

where $\mathbf{A} \in \mathbb{R}^{mL \times nL}$ is block diagonal and built from the blocks $\mathbf{A}^{(l)}$, and $\operatorname{vec}(\cdot)$ denotes the vectorization of a matrix. The authors of [EM09] guarantee recovery of jointly sparse signal ensembles $\mathbf{X}$ from measurements of type (7.5) via $\ell_{2,1}$-minimization provided $\mathbf{A}$ satisfies a certain block RIP. A direct connection between the number of measurements to guarantee block RIPs for random matrices and properties of the signal ensembles $\mathbf{X}$ is

presented in [EYRW15]. In particular, the authors show that one can profit from joint structure if the information in $\mathbf{X}$ is spread among multiple signals x_l. For instance, if all x_l but one are zero then one will need $m = \mathcal{O}(s \log(en/s))$ measurements per signal, rendering joint recovery useless. Hence, to obtain meaningful recovery guarantees for distributed CS one needs assumptions beyond a joint support set (see also Remark 7.2 below).

Both extensions of the classical CS model (2.6), namely, one-bit CS and distributed CS, are useful in practice. One might thus try to combine both approaches to reduce the number of measurements in one-bit sensing. The papers [TXY14], [KKWV16], [KGK+19] show promising numerical results, but they do not provide theoretical justification for the improvements.

Contribution

We provide uniform approximation guarantees for distributed CS from one-bit measurements quantifying the influence of the size L of signal ensembles $\mathbf{X}$ on the required number of measurements per signal m. Our analysis considers the second model above, i.e., distinct measurement matrices $\mathbf{A}^{(1)}, ..., \mathbf{A}^{(L)}$ corresponding to distinct sensors. In particular, we show that if the entries of all $\mathbf{A}^{(l)}$ are drawn as iid Gaussian random variables, then the matrix $\mathbf{A}$ will satisfy an $\ell_1/\ell_{2,1}$-RIP on a suitable set of jointly sparse signal ensembles with high probability. We adapt the ideas of [Fou16] to deduce a uniform error bound for recovering appropriate signal ensembles $\mathbf{X}$ from their one-bit measurements $\mathbf{Q}$ by applying one simple hard-thresholding step to $\mathbf{A}^\mathsf{T}\mathrm{vec}(\mathbf{Q})$. We find that $mL \geq Cs(\log(en/s) + L)$ measurements suffice to well-approximate $\mathbf{X}$ with high probability which means, for $L \simeq \log(en/s)$, $\mathcal{O}(s)$ measurements per single signal. This improves the classical CS results for Gaussian measurements of $\mathcal{O}(s \log(en/s))$ (cf. [FR13]). Moreover, we provide numerical evidence matching the experimental results in [TXY14], [KKWV16], [KGK+19].

Outline

This chapter is organized as follows. Section 7.1 introduces our problem in detail, Section 7.2 presents our main results, and Section 7.3 gives proofs. Section 7.4 supports the theory by numerical experiments, and Section 7.5 concludes with a brief summary and outlook on future work.

7.1. Problem Setup

Suppose we are given one-bit measurements $\mathbf{Q} \in \mathbb{R}^{m \times L}$ obtained from L signals $\mathsf{x}_l \in \mathbb{R}^n$, $l \in [L]$, that form the columns of a matrix $\mathbf{X} \in \mathbb{R}^{n \times L}$. For simplicity we write $\mathsf{x} = \mathrm{vec}(\mathbf{X}) = (\mathsf{x}_1^\mathsf{T}, ..., \mathsf{x}_L^\mathsf{T})^\mathsf{T}$ and $\mathsf{q} = \mathrm{vec}(\mathbf{Q}) = (\mathsf{q}_1^\mathsf{T}, ..., \mathsf{q}_L^\mathsf{T})^\mathsf{T}$. The linear measurement process can then be

described by

$$\mathsf{q} = \operatorname{sign}\Big(\mathbf{A}\mathsf{x}\Big) \tag{7.6}$$

where $\mathbf{A} \in \mathbb{R}^{Lm \times Ln}$ is a measurement matrix of the following form: $\mathbf{A}$ is block diagonal and built from the sub-matrices $\mathbf{A}^{(l)} \in \mathbb{R}^{m \times n}$, $l \in [L]$, which have iid Gaussian entries $\mathcal{N}(0,1)$, i.e., we have

$$\mathbf{A} = \begin{pmatrix} \boxed{\mathbf{A}^{(1)}} & & \\ & \ddots & \\ & & \boxed{\mathbf{A}^{(L)}} \end{pmatrix}. \tag{7.7}$$

We denote the i-th column of $(\mathbf{A}^{(l)})^\mathsf{T}$ by $\mathsf{a}_i^{(l)}$, i.e., $\mathsf{a}_i^{(l)}$ is the transposed i-th row of $\mathbf{A}^{(l)}$. Let $\theta > 0$ be an appropriate scaling to be determined later. Denote by $\mathbb{H}_s : \mathbb{R}^{n \times L} \to \mathbb{R}^{n \times L}$ the hard-thresholding operator which, for any $\mathbf{Z} \in \mathbb{R}^{n \times L}$, keeps only the s rows of largest ℓ_2-norm and sets the remaining $n-s$ rows to zero. Inspired by [Fou16], we aim to approximate x by

$$\mathsf{y} = \tilde{\mathbb{H}}_s\left((\theta\mathbf{A})^\mathsf{T}\mathsf{q}\right) \tag{7.8}$$

where $\tilde{\mathbb{H}}_s(\mathsf{z}) = \operatorname{vec}(\mathbb{H}_s(\mathbf{Z}))$, for $\mathsf{z} = \operatorname{vec}(\mathbf{Z})$. We will see that this simple procedure leads to near-optimal approximation guarantees for signal ensembles $\mathbf{X}$ whose signals x_l share a common support and the same magnitude in ℓ_2-norm. We denote the support of a signal ensemble $\mathbf{Z} \in \mathbb{R}^{n \times L}$, i.e., the set of non-zero rows of $\mathbf{Z}$, by $\operatorname{supp}(\mathbf{Z}) \subset [n]$. We define the set $\mathcal{S}_{s,L}$ of admissible signal ensembles

$$\mathcal{S}_{s,L} = \left\{ \mathsf{z} = \operatorname{vec}(\mathbf{Z}) \colon \mathbf{Z} = \begin{pmatrix} | & & | \\ \mathsf{z}_1 & \cdots & \mathsf{z}_L \\ | & & | \end{pmatrix} \in \mathbb{R}^{n \times L}, |\operatorname{supp}(\mathbf{Z})| \leq s, \|\mathsf{z}_l\|_2 = \|\mathsf{z}\|_2/\sqrt{L} \right\}. \tag{7.9}$$

As the simple sign-bit measurements (7.6) are invariant under scaling of the signals and, hence, dismiss any information on signal magnitudes, all we can hope for is approximating the directions of the individual signals. Hence, we can restrict the x_l to have constant norm without loss of generality. Consequently, whenever we use the terms "approximation of signals" or "recovery of signals" we implicitly mean "approximation/recovery of each signal up to the scaling" and restrict the results to signals of fixed norm.

7.2. Main Results

We show that Gaussian measurements of the form (7.7) fulfill under suitable scaling with high probability an $\ell_1/\ell_{2,1}$-*Restricted Isometry Property ($\ell_1/\ell_{2,1}$-RIP)* on

$$\mathcal{K}_{s,L} = \left\{ \mathsf{z} = \mathrm{vec}(\mathbf{Z}) \colon \mathbf{Z} \in \mathbb{R}^{n \times L}, |\mathrm{supp}(\mathbf{Z})| \leq s \right\} \tag{7.10}$$

(a relaxation of $\mathcal{S}_{s,L}$) if $mL \gtrsim s(\log(en/s) + L)$. We further show that all signals $\mathsf{x} \in \mathcal{S}_{s,L}$ can be well approximated from $mL \gtrsim s(\log(en/s)+L)$ one-bit measurements (7.6). Proofs can be found in Section 7.3. We first define what we mean by $\ell_1/\ell_{2,1}$-RIP.

Definition 7.1 ($\ell_1/\ell_{2,1}$-RIP)**.** A matrix $\mathbf{B} \in \mathbb{R}^{Lm \times Ln}$ satisfies the $\ell_1/\ell_{2,1}$-RIP on $\mathcal{K}_{s,L}$ with RIP-constant $\delta \in (0,1)$ if

$$\frac{\|\mathsf{z}\|_{2,1}}{\sqrt{L}} - \delta\|\mathsf{z}\|_2 \leq \|\mathbf{B}\mathsf{z}\|_1 \leq \frac{\|\mathsf{z}\|_{2,1}}{\sqrt{L}} + \delta\|\mathsf{z}\|_2 \tag{7.11}$$

for all $\mathsf{z} \in \mathcal{K}_{s,L}$.

The following lemma provides a sufficient number of measurements for $\theta\mathbf{A}$, with $\mathbf{A}$ given by (7.7), to fulfill the $\ell_1/\ell_{2,1}$-RIP. Its proof is inspired by [PV14, Cor. 2.3].

Lemma 7.1 ($\ell_1/\ell_{2,1}$-RIP)**.** For $\theta = \sqrt{\pi/(2Lm^2)}$ and $mL \gtrsim \delta^{-2}s(\log(en/s) + L)$, the operator $\theta\mathbf{A}$, with $\mathbf{A}$ given by (7.7), has the $\ell_1/\ell_{2,1}$-RIP on $\mathcal{K}_{s,L}$ with RIP-constant δ with probability at least $1 - 2\exp\left(-\delta^2 mL/(4\pi)\right)$.

Remark 7.1. For $L = 1$ this result agrees with known bounds on the sufficient number of measurements to have ℓ_1/ℓ_2-RIPs for random Gaussian matrices with high probability, namely, $m \gtrsim s\log(en/s)$. If $L \geq \log(en/s)$, we have $(\log(en/s) + L)/L \leq 2$ and, hence, Lemma 7.1 requires $m \gtrsim \delta^{-2}s$ for an RIP on signal ensembles in $\mathcal{K}_{s,L}$, i.e., only $\mathcal{O}(s)$ measurements per signal.
In [EYRW15] the authors examined how many measurements suffice for random Gaussian block matrices $\mathbf{A}$ to satisfy classical ℓ_2-RIPs depending on how the information of sparse signals is distributed on the different blocks of $\mathbf{A}$. Lemma 7.1 extends their result to $\ell_1/\ell_{2,1}$-RIPs when all signals have the same support.
As $\|\mathsf{z}\|_{2,1} \leq \sqrt{L}\|\mathsf{z}\|_2$, the upper bound in (7.11) can be replaced by $(1+\delta)\|\mathsf{z}\|_2$. Moreover, if restricted to $\mathcal{S}_{s,L}$ the $\ell_1/\ell_{2,1}$-RIP in (7.11) becomes a full ℓ_1/ℓ_2-RIP, i.e.,

$$(1-\delta)\|\mathsf{x}\|_2 \leq \|\mathbf{B}\mathsf{x}\|_1 \leq (1+\delta)\|\mathsf{x}\|_2 \tag{7.12}$$

as in this case $\|\mathsf{x}\|_{2,1} = \sqrt{L}\|\mathsf{x}\|_2$. This observation suggests that the signal model $\mathcal{S}_{s,L}$ is well-chosen as the signal ensembles in $\mathcal{S}_{s,L}$ when multiplied by block-diagonal Gaussian measurement matrices (induced by the distributed setting) behave like single sparse vectors multiplied by dense Gaussian measurement matrices.

The next theorem is our main result. It guarantees uniform recovery of all signal ensembles $\mathsf{x} \in \mathcal{S}_{s,L}$ by a simple hard-thresholding step. This result generalizes [Fou16, Thm. 8] to joint recovery of signals sharing a common support.

Theorem 7.2. Let $n, m, s > 0$, and let $\mathbf{A}$ be a random $Lm \times Ln$ matrix as defined in (7.7). Set

$$mL \gtrsim \delta^{-2} s(\log(en/s) + L) \tag{7.13}$$

and $\theta = \sqrt{\pi/(2Lm^2)}$. Then with probability at least $1 - 2\exp\left(-\delta^2 mL/(4\pi)\right)$ (over the entries of $\mathbf{A}$), and for all $\mathsf{x} \in \mathcal{S}_{s,L}$ with $\|\mathsf{x}\|_2 = 1$, we have

$$\|\mathsf{x} - \mathsf{y}\|_2 \lesssim \sqrt{\delta} \tag{7.14}$$

where y is defined in (7.8) and δ is the $\ell_1/\ell_{2,1}$-RIP constant of $\theta\mathbf{A}$.

Remark 7.2.
(*i*) As already mentioned in Remark 7.1, the required number of measurements per signal does not depend on $s\log(n/s)$ if $L \geq \log(n/s)$ but only on s, i.e., when recovering several signals that share a common support from sign-measurements collected independently for each single signal, one can significantly reduce the number of measurements.
(*ii*) For unit norm signals $\|\mathsf{x}_l\| = 1$ the error bound (7.14) becomes

$$\|\mathsf{x} - \mathsf{y}\|_2 \lesssim \sqrt{L\delta} \tag{7.15}$$

i.e., the error per single signal x_l is only less than $\sqrt{\delta}$ on average. In the worst case this error concentrates on one signal. However, if the signals all are dense on a shared support set $\mathcal{T} \subset [n]$, the support will be recovered even in this case because a large error on one signal implies less error on the remaining signals. It is not surprising that a dense support of all signals is needed to profit from joint recovery. If only one signal has dense support while the rest have mostly zeros on $\mathcal{T}$, then most of the signals do not carry helpful support information, i.e., joint recovery cannot be expected to improve performance.
(*iii*) At first glance, the proof of Theorem 7.2 hardly differs from the one of [Fou16, Thm. 8]. One first proves an ℓ_1/ℓ_2-RIP for $\theta\mathbf{A}$ and then concludes by a simple computation. However, the model selection $\mathcal{S}_{s,L}$ is crucial and must treat the matrix $\mathbf{A}$ as a whole to reach the sample complexity in (7.13). Consider the following naive approach: If $m \gtrsim \delta^{-2} s\log(en/s)$ for some $\delta > 0$, then for each $l \in [L]$ and Gaussian $\mathbf{A}^{(l)} \in \mathbb{R}^{m\times n}$, and with probability exceeding $1 - C\exp(-c\delta^2 m)$ we have

$$(1-\delta)\|\mathsf{z}\|_2 \leq \frac{\sqrt{2}}{m\sqrt{\pi}}\|\mathbf{A}^{(l)}\mathsf{z}\|_1 \leq (1+\delta)\|\mathsf{z}\|_2 \tag{7.16}$$

for all s-sparse $\mathsf{z} \in \mathbb{R}^n$ (see [Sch06]). Applying the union bound and summing over (7.16)

for $l \in [L]$, with probability at least $1 - C\exp(-c\delta^2 m + \log(L))$, we have

$$(1-\delta)\|\mathsf{x}\|_2 \leq \|(\theta\mathbf{A})\mathsf{x}\|_1 \leq (1+\delta)\|\mathsf{x}\|_2 \tag{7.17}$$

for all $\mathsf{x} \in \mathcal{S}_{s,L}$ and $\theta = \sqrt{2}/(m\sqrt{\pi L})$. Choosing $\delta' = \sqrt{L}\delta$ (to obtain comparable probabilities of success) shows that this leads to a worse sample complexity than (7.13).
(*iv*) The proof of Theorem 7.2 relies on the assumption that $\mathsf{x} \in \mathcal{S}_{s,L}$. As mentioned in Remark 7.1 this assumption corresponds to the equivalence of $\ell_1/\ell_{2,1}$-RIP and ℓ_1/ℓ_2-RIP on $\mathcal{S}_{s,L}$. One can relax the restriction a little. To this end, define for $\varepsilon \in (0,1)$ the set

$$\mathcal{S}_\varepsilon = \left\{ \mathsf{z} = \mathrm{vec}(\mathbf{Z}) : \mathbf{Z} \in \mathbb{R}^{N\times L},\ \mathrm{supp}(\mathbf{Z}) \leq s,\ \|\mathsf{z}_l\|_2 \in \left[\frac{1-\varepsilon}{\sqrt{L}}\|\mathsf{z}\|_2, \frac{1+\varepsilon}{\sqrt{L}}\|\mathsf{z}\|_2 \right] \right\} \tag{7.18}$$

of signal ensembles which differ in norm by a bounded perturbation. Assume that a matrix $\mathbf{B}$ satisfies the $\ell_1/\ell_{2,1}$-RIP on $\mathcal{K}_{s,L}$ with RIP-constant $\delta > 0$. Using $\|\mathsf{x}\|_{2,1} \in [1-\varepsilon, 1+\varepsilon]\sqrt{L}\|\mathsf{x}\|_2$ if $\mathsf{x} \in \mathcal{S}_\varepsilon$, this implies

$$(1-\delta)(1-\varepsilon)\|\mathsf{x}\|_2 \leq \|\mathbf{B}\mathsf{x}\|_1 \leq (1+\delta)(1+\varepsilon)\|\mathsf{x}\|_2. \tag{7.19}$$

To rewrite (7.19) as an ℓ_1/ℓ_2-RIP on $\mathcal{S}_\varepsilon$ for some $\delta' \in (0,1)$, i.e.,

$$(1-\delta')\|\mathsf{x}\|_2 \leq \|\mathbf{B}\mathsf{x}\|_1 \leq (1+\delta')\|\mathsf{x}\|_2 \tag{7.20}$$

for all $\mathsf{x} \in \mathcal{S}_\varepsilon$, it suffices that

$$(1-\delta') \leq (1-\delta)(1-\varepsilon) \tag{7.21}$$

which is equivalent to

$$\varepsilon \leq \frac{\delta' - \delta}{1-\delta}. \tag{7.22}$$

We can upper bound the right-hand side by δ' because it is positive for $\delta < \delta'$ and a decreasing function in δ for $0 \leq \delta < \delta'$. Hence, the more general $\ell_1/\ell_{2,1}$-RIP becomes an ℓ_1/ℓ_2-RIP on $\mathcal{S}_\varepsilon$ only for $\varepsilon \leq \delta'$, meaning that only small perturbations ε are possible if the approximation error in (7.14) is small. However, the assumption that all signals x_l share the same norm is a mild condition in our setting as (2.9) is blind to scaling and norm variations in signal ensembles.

7.3. Proofs of Lemma 7.1 and Theorem 7.2

Our proofs rely on an improved understanding of $\mathcal{K}_{s,L}$ defined in (7.10), and we start by analyzing this set in Section 7.3.1. The proof of Lemma 7.1 can be found in Section 7.3.2 and the proof of Theorem 7.2 is presented in Section 7.3.3.

7.3.1. Properties of $\mathcal{K}_{s,L}$

An important measure of complexity for subsets of $\mathbb{R}^d$ is the so-called *Gaussian width*. This quantity generalizes the notion of linear dimension to arbitrary sets and is a useful tool for estimating the sampling requirements of signal sets in CS.

Definition 7.2 (Gaussian width [PV14, Eq. (1.2)])**.** The Gaussian width of $\mathcal{K} \subset \mathbb{R}^d$ is defined as

$$w(\mathcal{K}) = \mathsf{E}\left[\sup_{\mathsf{z}\in\mathcal{K}} |\langle \mathsf{G}, \mathsf{z}\rangle|\right] \tag{7.23}$$

where $\mathsf{G} \sim \mathcal{N}(\mathsf{0}, \mathrm{Id}_d)$ is a random vector with iid Gaussian entries.

Remark 7.3. Let $\mathcal{B}(\mathsf{0}, 1)$ denote the Euclidean ball of radius 1 centered at $\mathsf{0}$. Examples illustrating the relation between Gaussian width and set complexity are as follows [PV13b]:

(*i*) For $\mathcal{K} = \mathcal{B}(\mathsf{0}, 1) \subset \mathbb{R}^d$ one has $w(\mathcal{K}) \approx \sqrt{d}$.

(*ii*) If the linear dimension of a set $\mathcal{K} \subset \mathcal{B}(\mathsf{0}, 1) \subset \mathbb{R}^d$ is $\dim(\mathcal{K}) = k$, then $w(\mathcal{K}) \approx \sqrt{k}$.

(*iii*) Let $\Sigma_s \subset \mathbb{R}^d$ denote the set of s-sparse vectors. Then $w(\Sigma_s \cap \mathcal{B}(\mathsf{0}, 1)) \approx \sqrt{s \log(ed/s)}$.

Examples (*i*) and (*ii*) show that $w(\mathcal{K})$ provides a consistent extension of the linear dimension to arbitrary sets in $\mathbb{R}^d$. A helpful rule of thumb is $w(\mathcal{K})^2 \sim \dim(\mathcal{K})$, i.e., the complexity of a set corresponds to the squared Gaussian width. However, note that contrary to $\dim(\mathcal{K})$ the Gaussian width scales with $\sup_{\mathsf{z}\in\mathcal{K}} \|\mathsf{z}\|_2$.

The Gaussian width of a set $\mathcal{K}$ is closely related to the covering number $N(\mathcal{K}, \varepsilon)$ via Dudley's and Sudakov's inequalities (cf. [Tal14]). The covering number $N(\mathcal{K}, \varepsilon)$ of a set is defined as the minimal number of ε-balls in ℓ_2-norm (centered in $\mathcal{K}$) one needs to cover $\mathcal{K}$ completely. The cardinality of any ε-net of a set $\mathcal{K}$ provides an upper bound of $N(\mathcal{K}, \varepsilon)$. A subset $\tilde{\mathcal{K}} \subset \mathcal{K}$ is called an ε-net of $\mathcal{K}$ if for any $\mathsf{z} \in \mathcal{K}$ there exists $\tilde{\mathsf{z}} \in \tilde{\mathcal{K}}$ with $\|\mathsf{z} - \tilde{\mathsf{z}}\|_2 \leq \varepsilon$. We obtain a bound on $w(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1))$ by first bounding $N(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1), \varepsilon)$ in Lemma 7.3 and then applying Dudley's inequality in Lemma 7.4.

Lemma 7.3 (Covering Number of $\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1)$)**.** For $\varepsilon \in (0, 1)$ we have

$$\log\left(N(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1), \varepsilon)\right) \leq s \log\left(\frac{en}{s}\right) + sL \log\left(\frac{3}{\varepsilon}\right). \tag{7.24}$$

Proof. As $\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1)$ is the union of $\binom{n}{s}$ unit ℓ_2-balls in $\mathbb{R}^{sL}$ embedded into $\mathbb{R}^{nL}$ and each unit ball can be covered by an ε-net of cardinality at most $(3/\varepsilon)^{sL}$ (see [CP11, Section 3]), we have

$$N\left(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1), \varepsilon\right) \leq \binom{n}{s} \left(\frac{3}{\varepsilon}\right)^{sL} \leq \left(\frac{en}{s}\right)^{s} \left(\frac{3}{\varepsilon}\right)^{sL}. \tag{7.25}$$

■

Lemma 7.3 leads to a bound on $w(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1))$.

Lemma 7.4 (Gaussian width of $\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1)$)**.** We have

$$w(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1)) \lesssim \sqrt{s \left(\log\left(\frac{en}{s}\right) + L \right)}. \tag{7.26}$$

Proof. By [PV13b, Prop. 2.1] one has $w(\mathcal{K}) = \mathsf{E}[\sup_{\mathsf{z} \in \mathcal{K}} \langle \mathsf{G}, \mathsf{z} \rangle]$ for an origin symmetric set $\mathcal{K}$. Hence, we obtain

$$\begin{aligned} w(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1)) &\leq \mathsf{E}\left[\sup_{\mathsf{z} \in \mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0},1)} \langle \mathsf{G}, \mathsf{z} \rangle \right] \\ &\overset{\text{a}}{\leq} 24 \int_0^1 \sqrt{\log\left(N(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1), \varepsilon)\right)}\, d\varepsilon \\ &\overset{\text{b}}{\leq} 24 \sqrt{\int_0^1 1^2 \, d\varepsilon} \cdot \sqrt{\int_0^1 \log\left(N(\mathcal{S}_{s,L} \cap \mathcal{B}(\mathsf{0}, 1), \varepsilon)\right)\, d\varepsilon} \\ &\overset{\text{c}}{\leq} 24 \sqrt{s \left(\log\left(\frac{en}{s}\right) + L(1 + \log 3) \right)} \end{aligned} \tag{7.27}$$

where (a) follows from Dudley's inequality [LT02, Thm. 11.17], (b) from Hölder's inequality and (c) from Lemma 7.3. ■

7.3.2. Proof of the RIP Lemma (Lemma 7.1)

To prove the $\ell_1/\ell_{2,1}$-RIP for $\theta\mathbf{A}$ on the signal set $\mathcal{K}_{s,L}$, we restrict ourselves to $\mathcal{K}_{s,L} \cap \mathbb{S}^{nL-1}$ where $\mathbb{S}^{nL-1}$ denotes the unit sphere in $\mathbb{R}^{nL}$. It suffices to prove (7.11) for all $\mathsf{z} \in \mathcal{K}_{s,L} \cap \mathbb{S}^{nL-1}$, as (7.11) is invariant under scaling of the ℓ_2-norm. The proof hence reduces to a direct application of the following concentration lemma which is a slightly adapted version of [PV14, Lemma 2.1]. For sake of completeness we report its full proof.

Lemma 7.5. Consider a bounded subset $\mathcal{K} \subset \mathbb{R}^{NL}$ and let $\mathsf{a}_i^{(l)} \sim \mathcal{N}(\mathsf{0}, \mathrm{Id}_n)$, $i \in [m]$, $l \in [L]$ be independent Gaussian vectors in $\mathbb{R}^n$. Define

$$Z \coloneqq \sup_{\mathsf{x} \in \mathcal{K}} \left| \sum_{i=1}^{m} \sum_{l=1}^{L} \sqrt{\frac{\pi}{2Lm^2}} \left| \langle \mathsf{A}_i^{(l)}, \mathsf{x}_l \rangle \right| - \frac{1}{\sqrt{L}} \|\mathsf{x}\|_{2,1} \right|. \tag{7.28}$$

Then we have

$$\mathsf{E}[Z] \leq \sqrt{8\pi} \frac{w(\mathcal{K})}{\sqrt{mL}} \tag{7.29}$$

and

$$\Pr\left[Z > \frac{\sqrt{8\pi} w(\mathcal{K})}{\sqrt{mL}} + u \right] \leq 2 \exp\left(-\frac{mLu^2}{\pi d(\mathcal{K})^2} \right) \tag{7.30}$$

where $d(\mathcal{K}) \coloneqq \max_{\mathsf{x}\in\mathcal{K}} \|\mathsf{x}\|_2$.

Proof. Let $G \sim \mathcal{N}(0,1)$ and note that $\mathsf{E}[|G|] = \sqrt{2/\pi}$. We have

$$\mathsf{E}\left[\sum_{i=1}^{m}\sum_{l=1}^{L}\sqrt{\frac{\pi}{2Lm^2}}\left|\langle \mathsf{A}_i^{(l)}, \mathsf{x}_l\rangle\right|\right] = \sum_{i=1}^{m}\sum_{l=1}^{L}\sqrt{\frac{\pi}{2Lm^2}}\mathsf{E}[|G|]\,\|\mathsf{x}_l\|_2 = \frac{\|\mathsf{x}\|_{2,1}}{\sqrt{L}}. \tag{7.31}$$

Now define the random variables $\boldsymbol{\vartheta}_i^{(l)} = \sqrt{\pi/(2Lm^2)}\left|\langle \mathsf{A}_i^{(l)}, \mathsf{x}_l\rangle\right|$, for $i \in [m], l \in [L]$, iid copies $\hat{\boldsymbol{\vartheta}}_i^{(l)}$, and independent Rademacher variables $\varepsilon_{i,l}$, i.e., $\mathsf{P}[\varepsilon_{i,l} = 1] = \mathsf{P}[\varepsilon_{i,l} = -1] = 1/2$. We obtain

$$\begin{aligned}
\mathsf{E}[Z] &= \mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\left(\boldsymbol{\vartheta}_i^{(l)} - \mathsf{E}\left[\boldsymbol{\vartheta}_i^{(l)}\right]\right)\right|\right] \\
&= \mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\left(\boldsymbol{\vartheta}_i^{(l)} - \mathsf{E}\left[\boldsymbol{\vartheta}_i^{(l)}\right]\right) - \mathsf{E}\left[\hat{\boldsymbol{\vartheta}}_i^{(l)} - \mathsf{E}\left[\hat{\boldsymbol{\vartheta}}_i^{(l)}\right]\right]\right|\right] \\
&= \mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\mathsf{E}\left[\boldsymbol{\vartheta}_i^{(l)} - \hat{\boldsymbol{\vartheta}}_i^{(l)}\right]\right|\right] \\
&\overset{\text{a}}{\leq} \mathsf{E}\left[\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\boldsymbol{\vartheta}_i^{(l)} - \hat{\boldsymbol{\vartheta}}_i^{(l)}\right|\right]\right] \\
&= \mathsf{E}\left[\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\varepsilon_{i,l}\left(\boldsymbol{\vartheta}_i^{(l)} - \hat{\boldsymbol{\vartheta}}_i^{(l)}\right)\right|\right]\right] \\
&\overset{\text{b}}{\leq} 2\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\varepsilon_{i,l}\boldsymbol{\vartheta}_i^{(l)}\right|\right] \qquad (7.32) \\
&= 2\sqrt{\frac{\pi}{2Lm^2}}\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\varepsilon_{i,l}\left|\langle \mathsf{A}_i^{(l)}, \mathsf{x}_l\rangle\right|\right|\right] \\
&\overset{\text{c}}{\leq} 4\sqrt{\frac{\pi}{2Lm^2}}\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{i=1}^{m}\sum_{l=1}^{L}\varepsilon_{i,l}\langle \mathsf{A}_i^{(l)}, \mathsf{x}_l\rangle\right|\right] \\
&= 4\sqrt{\frac{\pi}{2Lm^2}}\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\sum_{l=1}^{L}\left\langle\sum_{i=1}^{m}\varepsilon_{i,l}\mathsf{A}_i^{(l)}, \mathsf{x}_l\right\rangle\right|\right] \\
&= 4\sqrt{\frac{\pi}{2Lm^2}}\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\left\langle\sum_{i=1}^{m}\left(\varepsilon_{i,1}(\mathsf{A}_i^{(1)})^{\mathsf{T}}, \ldots, \varepsilon_{i,L}(\mathsf{A}_i^{(L)})^{\mathsf{T}}\right)^{\mathsf{T}}, \mathsf{x}\right\rangle\right|\right] \\
&\overset{\text{d}}{=} 4\sqrt{\frac{\pi}{2Lm^2}}\mathsf{E}\left[\sup_{\mathsf{x}\in\mathcal{K}}\left|\langle\sqrt{m}\mathsf{G}, \mathsf{x}\rangle\right|\right] \qquad (7.33) \\
&= \sqrt{8\pi}\frac{w(\mathcal{K})}{\sqrt{mL}}
\end{aligned}$$

where (a) follows from Jensen's inequality and (b) from the triangle inequality, (c) is a

consequence of [LT02, Thm. 4.12] and in (d) we let $\mathsf{G} \sim \mathcal{N}(0, \mathrm{Id}_{nL})$. To prove the deviation inequality (7.30) we will first show that Z, as defined in (7.28), is Lipschitz continuous in $\mathbf{A}$. Consider two fixed block diagonal matrices $\mathbf{A}, \mathbf{B}$ as in (7.7) and define the operator

$$Z(\mathbf{A}) \coloneqq \sup_{\mathsf{x} \in \mathcal{K}} \left| \sum_{i=1}^{m} \sum_{l=1}^{L} \sqrt{\frac{\pi}{2Lm^2}} \left| \langle \mathsf{a}_i^{(l)}, \mathsf{x}_l \rangle \right| - \frac{\|\mathsf{x}\|_{2,1}}{\sqrt{L}} \right|. \tag{7.34}$$

Then, we have

$$\begin{aligned}
&|Z(\mathbf{A}) - Z(\mathbf{B})| \\
&\quad = \sup_{\mathsf{x} \in \mathcal{K}} \left| \sum_{i=1}^{m} \sum_{l=1}^{L} \sqrt{\frac{\pi}{2Lm^2}} \left| \langle \mathsf{a}_i^{(l)}, \mathsf{x}_l \rangle \right| - \frac{\|\mathsf{x}\|_{2,1}}{\sqrt{L}} \right| - \sup_{\mathsf{x} \in \mathcal{K}} \left| \sum_{i=1}^{m} \sum_{l=1}^{L} \sqrt{\frac{\pi}{2Lm^2}} \left| \langle \mathsf{b}_i^{(l)}, \mathsf{x}_l \rangle \right| - \frac{\|\mathsf{x}\|_{2,1}}{\sqrt{L}} \right| \\
&\quad \leq \sup_{\mathsf{x} \in \mathcal{K}} \left\{ \left| \sum_{i=1}^{m} \sum_{l=1}^{L} \sqrt{\frac{\pi}{2Lm^2}} \left| \langle \mathsf{a}_i^{(l)}, \mathsf{x}_l \rangle \right| - \frac{\|\mathsf{x}\|_{2,1}}{\sqrt{L}} \right| - \left| \sum_{i=1}^{m} \sum_{l=1}^{L} \sqrt{\frac{\pi}{2Lm^2}} \left| \langle \mathsf{b}_i^{(l)}, \mathsf{x}_l \rangle \right| - \frac{\|\mathsf{x}\|_{2,1}}{\sqrt{L}} \right| \right\} \\
&\quad \leq \sup_{\mathsf{x} \in \mathcal{K}} \left| \sum_{i=1}^{m} \sum_{l=1}^{L} \sqrt{\frac{\pi}{2Lm^2}} \left| \langle \mathsf{a}_i^{(l)} - \mathsf{b}_i^{(l)}, \mathsf{x}_l \rangle \right| \right| \\
&\quad \leq \sup_{\mathsf{x} \in \mathcal{K}} \sqrt{\frac{\pi}{2Lm^2}} \sum_{i=1}^{m} \sum_{l=1}^{L} \|\mathsf{a}_i^{(l)} - \mathsf{b}_i^{(l)}\|_2 \|\mathsf{x}_l\|_2 \qquad (7.35) \\
&\quad \leq \sup_{\mathsf{x} \in \mathcal{K}} \sqrt{\frac{\pi}{2Lm^2}} \|\mathbf{A} - \mathbf{B}\|_F \left(\sum_{i=1}^{m} \sum_{l=1}^{L} \|\mathsf{x}_l\|_2^2 \right)^{\frac{1}{2}} \\
&\quad \leq \sqrt{\frac{\pi}{2Lm^2}} \sqrt{m} \|\mathbf{A} - \mathbf{B}\|_F d(\mathcal{K}) \\
&\quad = \frac{d(\mathcal{K})}{\sqrt{mL}} \sqrt{\frac{\pi}{2}} \|\mathbf{A} - \mathbf{B}\|_F.
\end{aligned}$$

Hence, $Z(\cdot)$ is Lipschitz continuous with constant $\frac{d(\mathcal{K})}{\sqrt{mL}}\sqrt{\frac{\pi}{2}}$. Using [LT02, Eq. (1.6)], we see that for our random choice of $\mathbf{A}$, we have

$$\mathsf{P}[|Z - \mathsf{E}[Z]| > u] \leq 2 \exp\left(-\frac{2u^2 mL}{2\pi d(\mathcal{K})^2} \right). \tag{7.36}$$

Thus, using (7.32), we have

$$\Pr\left[Z - \sqrt{8\pi} \frac{w(\mathcal{K})}{\sqrt{mL}} > u \right] \leq \Pr[Z - \mathsf{E}[Z] > u] \leq \Pr[|Z - \mathsf{E}[Z]| > u] \leq 2 \exp\left(-\frac{mLu^2}{\pi d(\mathcal{K})^2} \right) \tag{7.37}$$

which yields the claim. ∎

Proof of Lemma 7.1. The lemma is a direct consequence of Lemmas 7.4 and 7.5. We choose $u = \delta/2$ and $mL \geq 8\pi(\delta/2)^{-2} w(\mathcal{K}_{s,L} \cap \mathcal{B}(0,1))^2$ and note that by Lemma 7.4,

we have $w(\mathcal{K}_{s,L} \cap \mathcal{B}(\mathsf{0},1)) \geq w(\mathcal{K}_{s,L} \cap \mathbb{S}^{nL-1})$. Thus, with probability at least $1 - 2\exp\left(-mL\delta^2/(4\pi)\right)$, we have

$$\left| \sqrt{\frac{\pi}{2Lm^2}} \|\mathbf{A}\mathsf{z}\|_1 - \frac{\|\mathsf{z}\|_{2,1}}{\sqrt{L}} \right| \leq \sqrt{8\pi} \frac{w\left(\mathcal{K}_{s,L} \cap \mathbb{S}^{nL-1}\right)}{\sqrt{mL}} + \frac{\delta}{2} \leq \delta \tag{7.38}$$

for all $\mathsf{z} \in \mathcal{K}_{s,L} \cap \mathbb{S}^{nL-1}$. The statement follows for $\mathsf{z} \in \mathcal{K}_{s,L}$ by multiplying by $\|\mathsf{z}\|_2$ on both sides. ■

7.3.3. Proof of the Main Result (Theorem 7.2)

We denote the set of nonzero rows of $\mathbf{X}$ by $\operatorname{supp}(\mathbf{X}) = \operatorname{supp}(\mathsf{x}) \subset \mathcal{T}$ for some $\mathcal{T} \subset [n]$ with $|\mathcal{T}| \leq s$. For $\mathsf{z} = \operatorname{vec}(\mathbf{Z}) \in \mathbb{R}^{nL}$, let $\mathsf{z}_\mathcal{T} = \operatorname{vec}(\mathbf{Z}_\mathcal{T})$, where $\mathbf{Z}_\mathcal{T}$ is the matrix in which all rows not in $\mathcal{T}$ are zero. The proof of Theorem 7.2 follows the argument of [Fou16, Thm. 8] but relies on the assumption that all signals x_l share a common ℓ_2-norm.

Lemma 7.6. If the operator $\theta\mathbf{A}$ satisfies the $\ell_1/\ell_{2,1}$-RIP on $\mathcal{K}_{s,L}$, then all $\mathsf{x} \in \mathcal{S}_{s,L}$ with $\|\mathsf{x}\|_2 = 1$ satisfy

$$\left\| \left((\theta\mathbf{A})^\mathsf{T} \operatorname{sign}(\mathbf{A}\mathsf{x}) \right)_\mathcal{T} - \mathsf{x} \right\|_2^2 \leq 5\delta. \tag{7.39}$$

Proof. Define $\theta\mathsf{b} = \theta\mathbf{A}^\mathsf{T} \operatorname{sign}(\mathbf{A}\mathsf{x}) \in \mathbb{R}^{nL}$ to be the back-projected quantized measurements. We then have

$$\left\| \left((\theta\mathbf{A})^\mathsf{T} \operatorname{sign}(\mathbf{A}\mathsf{x}) \right)_\mathcal{T} - \mathsf{x} \right\|_2^2 = \|(\theta\mathsf{b})_\mathcal{T}\|_2^2 - 2\langle (\theta\mathsf{b})_\mathcal{T}, \mathsf{x} \rangle + \|\mathsf{x}\|_2^2 \tag{7.40}$$

and

$$\begin{aligned} \|(\theta\mathsf{b})_\mathcal{T}\|_2^2 &= \langle (\theta\mathsf{b})_\mathcal{T}, (\theta\mathsf{b})_\mathcal{T} \rangle = \langle (\theta\mathbf{A})^\mathsf{T} \operatorname{sign}(\mathbf{A}\mathsf{x}), (\theta\mathsf{b})_\mathcal{T} \rangle \\ &= \langle \operatorname{sign}(\mathbf{A}\mathsf{x}), (\theta\mathbf{A})(\theta\mathsf{b})_\mathcal{T} \rangle \leq \|(\theta\mathbf{A})(\theta\mathsf{b})_\mathcal{T}\|_1 \\ &\leq \frac{\|(\theta\mathsf{b})_\mathcal{T}\|_{2,1}}{\sqrt{L}} + \delta \|(\theta\mathsf{b})_\mathcal{T}\|_2 \leq (1+\delta)\|(\theta\mathsf{b})_\mathcal{T}\|_2. \end{aligned} \tag{7.41}$$

Hence, we have $\|(\theta\mathsf{b})_\mathcal{T}\|_2 \leq 1 + \delta$ and

$$\langle (\theta\mathsf{b})_\mathcal{T}, \mathsf{x} \rangle = \langle \operatorname{sign}(\mathbf{A}\mathsf{x}), (\theta\mathbf{A})\mathsf{x} \rangle = \|(\theta\mathbf{A})\mathsf{x}\|_1 \geq \frac{\|\mathsf{x}\|_{2,1}}{\sqrt{L}} - \delta\|\mathsf{x}\|_2 = (1-\delta) \tag{7.42}$$

where we used $\|\mathsf{x}\|_{2,1} = \sqrt{L}\|\mathsf{x}\|_2 = \sqrt{L}$. We conclude that

$$\left\| \left((\theta\mathbf{A})^\mathsf{T} \operatorname{sign}(\mathbf{A}\mathsf{x}) \right)_\mathcal{T} - \mathsf{x} \right\|_2^2 \leq (1+\delta)^2 - 2(1-\delta) + 1 \leq 5\delta. \tag{7.43}$$

■

Algorithm 7.1 : sHT$(\mathsf{y}, \mathbf{A}, s)$

Require: $\mathbf{Q} \in \{-1,1\}^{m\times L}$, $\mathbf{A} \in \mathbb{R}^{mL\times nL}$

1: $\mathsf{y} \leftarrow \tilde{\mathsf{H}}_s(\mathbf{A}^\mathsf{T}\mathrm{vec}(\mathbf{Q}))$ $\{\tilde{\mathsf{H}}_s$ is defined in (7.8)$\}$
2: $\mathbf{Y} \leftarrow \mathbf{reshape}(\mathsf{x}, n, L)$ $\{\mathbf{reshape}(\cdot)$ reverses $\mathrm{vec}(\cdot)\}$
3: **return** $\mathbf{Y}$

Proof of Theorem 7.2. Choose $mL \gtrsim \delta^{-2}2s(\log(en/(2s))+L)$ such that by Lemma 7.1, $\theta\mathbf{A}$ satisfies the $\ell_1/\ell_{2,1}$-RIP on $\mathcal{K}_{2s,L}$ with high probability. Let $\mathcal{T} = \mathrm{supp}(\mathsf{x})$ and $\hat{\mathcal{T}} = \mathrm{supp}(\hat{\mathsf{x}})$ where $\hat{\mathsf{x}} = \tilde{\mathsf{H}}_s((\theta\mathbf{A})^\mathsf{T}\mathsf{y})$. Note that $\hat{\mathsf{x}}$ is also the best s-row approximation of $((\theta\mathbf{A})^\mathsf{T}\mathsf{y})_{\mathcal{T}\cup\hat{\mathcal{T}}}$. Hence, we have

$$\begin{aligned}\|\mathsf{x}-\hat{\mathsf{x}}\|_2 &\le \|((\theta\mathbf{A})^\mathsf{T}\mathsf{y})_{\mathcal{T}\cup\hat{\mathcal{T}}} - \hat{\mathsf{x}}\|_2 + \|((\theta\mathbf{A})^\mathsf{T}\mathsf{y})_{\mathcal{T}\cup\hat{\mathcal{T}}} - \mathsf{x}\|_2 \\ &\le 2\|((\theta\mathbf{A})^\mathsf{T}\mathsf{y})_{\mathcal{T}\cup\hat{\mathcal{T}}} - \mathsf{x}\|_2 \le 2\sqrt{5\delta}\end{aligned} \tag{7.44}$$

where we applied Lemma 7.6 for $\mathcal{K}_{2s,L}$ in the last inequality (note that $|\mathcal{T}\cup\hat{\mathcal{T}}| \le 2s$). ■

7.4. Numerical Experiments

We illustrate numerically the theoretical results of Section 7.2. Recall that we propose to recover an unknown signal ensemble $\mathbf{X} \in \mathbb{R}^{n\times L}$ from its one-bit measurements $\mathbf{Q} \in \{-1,1\}^{m\times L}$ by a single hard-thresholding step which needs the measurements $\mathbf{Q}$, the block diagonal measurement matrix $\mathbf{A}$ and the sparsity level $s = |\mathrm{supp}(\mathbf{X})|$. The simple approximation procedure is presented in Algorithm 7.1. We present two experiments which document the asymptotically linear dependence of $m = \mathcal{O}(s)$ measurements per signal. In both experiments the block diagonal measurement matrix $\mathbf{A}$ has iid Gaussian entries and is scaled by $\theta = \sqrt{\pi/(2Lm^2)}$ as required in Lemma 7.1. Signal ensembles $\mathbf{X} \in \mathbb{R}^{n\times L}$ with $|\mathrm{supp}(\mathbf{X})| = s$ are created by first drawing some support set $\mathcal{T} \subset [n]$ uniformly at random, then drawing the single entries as iid Gaussians of mean 0 and variance 1, and finally rescaling all single signals x_l, $l \in [L]$, to have unit norm.

In the first experiment we approximate 500 randomly drawn signal ensembles $\mathbf{X} \in \mathbb{R}^{n\times L}$ of signal dimension $n = 100$, ensemble size $L = 1, 2, 5, 20$, and support size $s = 5$ from their one-bit measurements $\mathbf{Q} \in \{-1,1\}^{m\times L}$. Figure 7.2 depicts the average approximation error $\|\mathbf{X}-\mathbf{Y}\|_F$ against the measurement rate $\alpha = m/n$. One observes an improvement for larger ensembles. The benchmark decay of order $\mathcal{O}(m^{-1/2})$ indicates that, on average, the algorithm performs better than one might expect from the worst case guarantee of $\mathcal{O}(m^{-1/4})$ in Theorem 7.2.

The second experiment (see Figure 7.3) illustrates the dependence of m and s. We again approximate 500 randomly drawn signal ensembles $\mathbf{X} \in \mathbb{R}^{n\times L}$ of signal dimension $n = 100$ and ensemble size $L = 1, 2, 5, 20$ from their one-bit measurements $\mathbf{Q} \in \mathbb{R}^{m\times L}$. This time

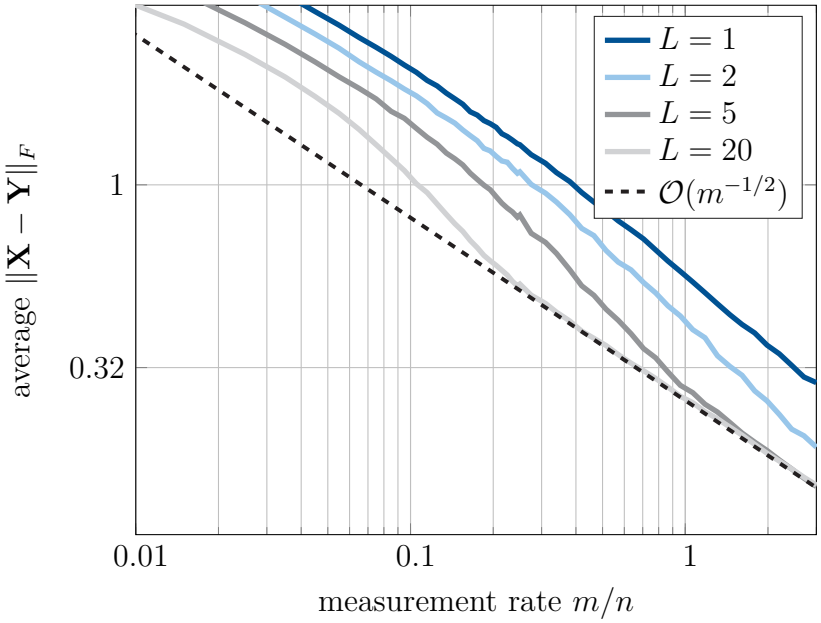


Figure 7.2.: Log-log plot of the simulated error $\|\mathbf{X} - \mathbf{Y}\|_F$ averaged over 500 experiments for $s = 5$ and $n = 100$.

the support size of $\mathbf{X}$ varies from $s = 1$ to $s = 50$ while the measurement rate $\alpha = m/n$ ranges from $\alpha = 0.01$ up to $\alpha = 3$. The average approximation error $\|\mathbf{X} - \mathbf{Y}\|_F$ is plotted in color while a selected error level is highlighted. When comparing the different choices of L, the linear dependence of m on s for $L = 20$ and fixed error levels is clearly visible and different from the $s \ln(en/s)$ behavior for $L = 1$.

The reader might notice that the measurement rate does not behave linearly in the plots $L = 2$ and $L = 5$ for $s/n \geq e^{1-L}$ which corresponds to the case $L \geq \log(en/s)$ for which we claimed $\mathcal{O}(s)$ behavior in Remark 7.1. This is no contradiction of our theory because for the $\mathcal{O}(s)$ argument it suffices to bound $(\log(en/s) + L)/L \leq 2$. In the numerical experiments with fixed L we observe that the transition from $(\log(en/s) + L)/L \approx 2$ for small values of s/n (corresponding to large values of $\log(en/s)$) to $(\log(en/s) + L)/L \approx 1$ for large values of s/n (corresponding to small values of $\log(en/s)$) causes a non-linear shape as long as L is not clearly dominating (cf. $L = 20$).

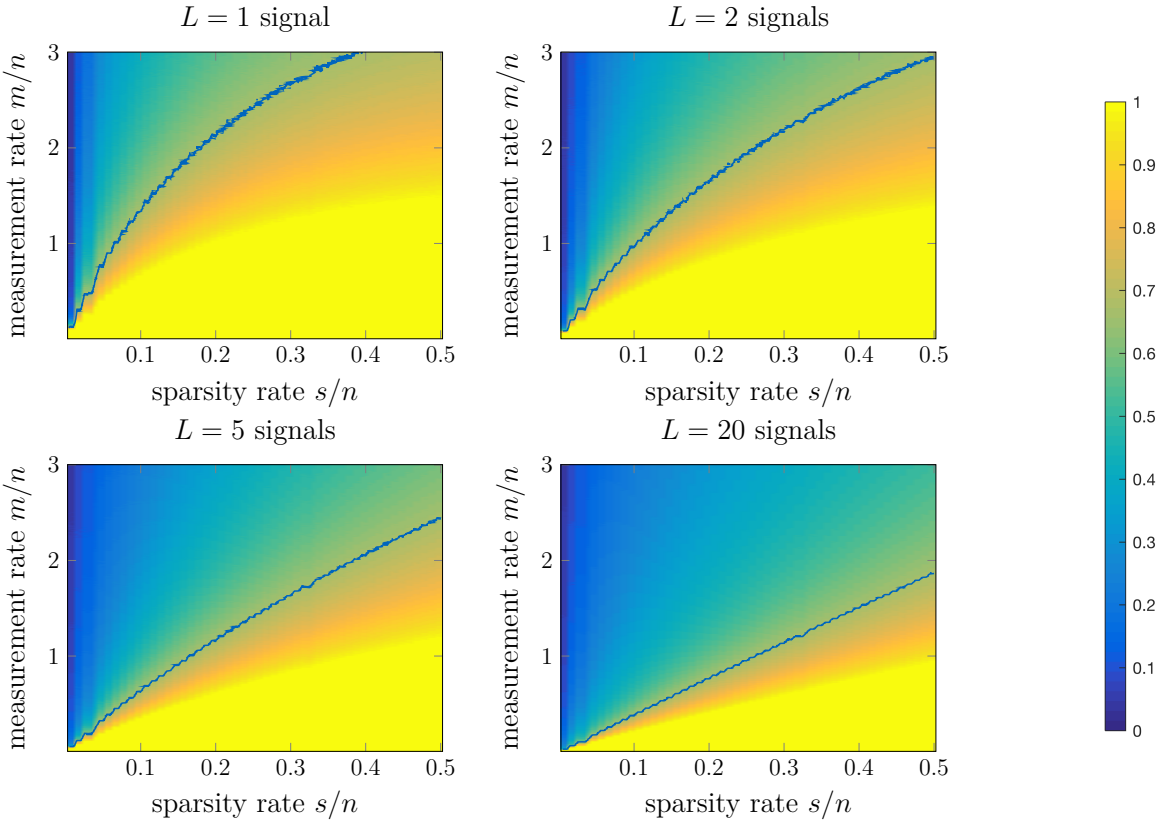


Figure 7.3.: Simulated error $\|\mathbf{X} - \mathbf{Y}\|_F$ averaged over 500 experiments with $n = 100$. The blue contour lines correspond to $\|\mathbf{X} - \mathbf{Y}\|_F = 2/3$.

7.5. Conclusion

We examined how heavily quantized measurements and distributed CS can be combined. We showed that a single hard thresholding step enables uniform joint approximation of several signals sharing a common support for $m = \mathcal{O}(s)$ measurements per signal. We see several possible directions of future research. First, sophisticated alternatives to a single hard-thresholding step have been proposed (see [KKWV16]) which numerically give a smaller approximation error. It would be interesting to extend our theory to these methods. Second, extending our results to noisy measurements would be useful. The proof of Theorem 7.2, however, relies on noiseless measurements to exploit the equivalence of $\langle \operatorname{sign}(\mathbf{A}\mathsf{x}), (\theta\mathbf{A})\mathsf{x}\rangle$ and $\|(\theta\mathbf{A})\mathsf{x}\|_1$ in (7.42). It seems difficult to modify the above proof to tolerate noise on the measurements. Third, our analysis relies on choosing the entries of the measurement matrices as iid Gaussian random variables. Numerical simulations indicate that, for example, iid Rademacher entries provide a similar (average) performance. It would be useful to extend our results to more general classes of measurements. Finally, increasing the quantization depth to multi-bit quantizers is desirable as this should decrease the approximation error (cf. [JDV13, Jac16]) and bridge the wide performance gap between

unquantized measurements and one-bit measurements.

8

Summary and Conclusions

In this dissertation, we investigated the digital compression of structured signals modeled as sparse sources both from information-theoretic and algorithmic points of view.

Part I

In Part I, we focused on the question of finding the smallest coding rates among all possible encoder and decoder pairs for a given average or probabilistic excess distortion criterion. We first studied the RD function with multiple distortion criteria in Chapter 3 both for infinite and finite block lengths. For finite block lengths, we derived a converse bound that provides tight bounds for single and multiple distortions and established a connection between previously known bounds.

In Chapter 4, we studied the RD function for Bernoulli Spike Sources. Since an intuitive upper bound was available in the literature, we focused on deriving a converse bound, first with two constraints and then for the usual case of a single squared error distortion constraint. This converse bound improves on previously known bounds as it captures the correct behavior at small distortions, which we used to characterize the RD function in the limit of small distortions.

In Chapter 5, we studied the RD behavior of the Distributed Bernoulli-Gaussian Spike Source with two terminals, building on the results from Chapter 4. Here, we derived an inner bound based on distinguishing between significant and insignificant samples and outer bounds using the single terminal converse result above. While the inner and outer bounds exhibit a gap at larger distortions, they closely match at small distortion values. We made this observation precise and characterized the achievable rate region exactly in the limit of small distortions at both terminals.

Part II

The second part investigated the RD performance of a specific encoder/decoder structure: QCS combined with approximate message passing reconstruction. To this end, we reviewed the GAMP algorithm for CS with scalar quantization and numerically investigated its RD tradeoff for different measurement rates and quantizer depths for Bernoulli-Gaussian signals. Further, we extended the algorithm to the multi-terminal setting and numerically investigated the RD performance of MGAMP.

A key observation was that the RD tradeoff is best if the measurement rate is just above the critical rate (the rate at which the phase transition happens in noiseless CS). For larger rates, the error decays only slowly in the measurement rate which leads to an increasingly worse tradeoff between the total number of bits and the reconstruction error. Based on this observation, we investigated how this tradeoff changes if one compresses the quantized measurements down to their joint entropy. Using the recent theory about asymptotic properties of GLMs [BMDK17], we determined the asymptotic limit of the measurement entropy and thus the RD behavior of QCS with lossless compression. An interesting observation is that for a given entropy, 1-bit CS seems to outperform QCS with a higher quantizer depth. Of course, finding a compression algorithm that achieves this optimal compression may be difficult since it must exploit the dependence of a large number of measurements. This is an interesting task for future work.

A second question investigated in this chapter was that of optimality of the GAMP algorithm. AMP for dense matrices has been shown to exhibit a suboptimal phase transition for noiseless CS [KMS+12a], so one might expect the same to hold true for QCS. Interestingly, in all performed experiments, GAMP achieves the MMSE whenever we used 1-bit quantization. Only for larger quantizer depths, where a phase transition starts to develop, GAMP becomes suboptimal around this phase transition. An interesting question is whether the *seeded sensing matrices* that have been shown to exhibit an optimal phase transition in the noiseless case can also be applied for this purpose here.

Last, we used the intuition from Part I that an optimal coding scheme should distinguish between significant and insignificant samples. To apply this to our QCS framework, we assumed that before sensing a signal, all small values are set to zero. This filtered signal was then used as the input to our QCS system. For this case, we performed the same investigations as above. We found that in the 1-bit case, GAMP is still optimal at all measurement rates. Further, the RD tradeoff with optimal lossless compression is significantly improved as compared to directly sensing a Bernoulli-Gaussian signal. Of course, an important question is whether this filtering operation can be realized in practice.

Part III

Part III also investigated a specific setting for QCS. Here, we assumed that 1-bit measurements of several signals are taken independently but reconstructed jointly. Further, the sparse signals were assumed to have a common support. In contrast to Part II, our fidelity criterion was the worst case error among all signals and we were mainly interested

in how the measurement rate scales in the ambient dimension and the sparsity of the signals. Our key finding was that compared to the single terminal signal, we can reduce the required number of measurements from $\mathcal{O}(s\log(en/s))$ to $\mathcal{O}(s)$ in the distributed case - an improvement that can be significant in practice. This scaling was also confirmed in numerical simulations.

There are not many related results for distributed QCS available in the literature, which leaves many directions for future work. Our work has several strong assumptions that would be important to relax. First, we assume that our measurements are Gaussian, which is difficult to realize in practice. Recent results for the single terminal setting [DM18] with sub-Gaussian and heavy tailed measurements offer new tools that could be used for this purpose. Similarly, is important to consider noisy measurements and show that the results still hold in a similar way. Another direction is to consider signal sets that have weaker assumptions or nicer properties such as convexity. Finally, it is desirable to extend these results to multi-bit quantization, a task that is far from trivial.

A

Proofs for Chapter 4

A.1. Proof of Lemma 4.3

Recall that

$$d_1 = \sup_{P_{Y|X} \in \mathcal{Q}(d_{\mathsf{S}})} \Pr[\hat{B} = 0 \,|\, B = 1]. \tag{A.1}$$

Since $B = 1$ implies $X = Z$, we can write d_1 as

$$d_1 = \sup_f \mathsf{E}[f(Z)] \tag{A.2}$$

where the supremum is taken over all Borel measureable mappings $f : \mathbb{R} \to [0, 1]$ satisfying $\mathsf{E}[Z^2 f(Z)] \leq d_{\mathsf{S}}$. Intuitively, $f(z)$ is the conditional probability $\Pr[\hat{B} = 0 \,|\, Z = z]$. Moreover, let $W \coloneqq Z^2$ and note that by symmetry, we only need to consider mappings $f : [0, \infty) \to [0, 1]$ such that

$$d_1 = \sup_f \mathsf{E}[f(W)] \qquad \text{subject to} \qquad \mathsf{E}[W f(W)] \leq d_{\mathsf{S}}. \tag{A.3}$$

Let $\mu((a, b]) \coloneqq Q_W(b) - Q_W(a)$, where Q_W is the quantile function of W. Since f is Borel measurable, the expectation $\mathsf{E}[f(W)]$ is given by the Lebesgue integral of f with respect to μ [Hal74, Ch. V]:

$$\mathsf{E}[f(W)] = \sup \left\{ \int s(w) \mathrm{d}\mu(w) \,:\, s \text{ is simple and } s \leq g \right\}. \tag{A.4}$$

Recall that s is simple if it can be written as a linear combination of indicator functions $s(w) = \sum_{k=1}^{K} a_k \mathbb{1}_{\{w \in \mathcal{A}_k\}}$, where the $\mathcal{A}_k$ are disjoint sets and $0 \leq a_1 \leq a_2 \leq \ldots \leq a_K = 1$.

Since f is bounded from below by zero and from above by one, each s can be written as a convex combination of indicator functions of (overlapping) sets $\mathcal{B}_k$ such that $s(w) = \sum_{k=1}^{K} \lambda_k \mathbb{1}_{\{w \in \mathcal{B}_k\}}$, where $\lambda_k = a_k - a_{k-1}$, $a_{k-1} \coloneqq 0$ and $\mathcal{B}_k \coloneqq \bigcup_{k' \geq k} \mathcal{B}_{k'}$.
Now, we define for some s:

$$d_1^\dagger \coloneqq \mathsf{E}[s(W)] = \int s(w) \mathrm{d}\mu(w) = \sum_{k=1}^{K} a_k \mu(\mathcal{A}_k) = \sum_{k=1}^{K} \lambda_k \mu(\mathcal{B}_k) \tag{A.5}$$

and

$$d_{\mathsf{S}}^\dagger \coloneqq \mathsf{E}[W s(W)] = \sum_{k=1}^{K} \lambda_K \mathsf{E}\Big[W \mathbb{1}_{\{W \in \mathcal{B}_k\}}\Big]. \tag{A.6}$$

Next, choose $\gamma = F_W^{-1}(d_1^\dagger)$ and $\tilde{s}(w) \coloneqq \mathbb{1}_{\{w \in [0,\gamma]\}}$, so that $\mathsf{E}[\tilde{s}(W)] = \mu([0,\gamma]) = d_1^\dagger$. Then, we have

$$\begin{aligned} \mathsf{E}[\tilde{s}(W)] &= \sum_{k=1}^{K} \lambda_k \mathsf{E}\Big[W \mathbb{1}_{\{W \in [0,\gamma]\}}\Big] \\ &\overset{\text{a}}{\leq} d_1^\dagger + \gamma \sum_{k=1}^{K} \lambda_k \Big(\mu\Big([0,\gamma] \cap \mathcal{B}_k^{\text{c}}\Big) - \mu\Big([0,\gamma]^{\text{c}} \cap \mathcal{B}_k\Big) \Big) \\ &\overset{\text{b}}{=} d_1^\dagger \end{aligned} \tag{A.7}$$

where (a) follows since

$$\begin{aligned} \mathsf{E}\Big[W \mathbb{1}_{\{W \in [0,\gamma]\}}\Big] &= \mathsf{E}\Big[W \mathbb{1}_{\{W \in [0,\gamma] \cap \mathcal{B}_k\}}\Big] + \mathsf{E}\Big[W \mathbb{1}_{\{W \in [0,\gamma] \cap \mathcal{B}_k^{\text{c}}\}}\Big] \\ &= \mathsf{E}\Big[W \mathbb{1}_{\{W \in \mathcal{B}_k\}}\Big] - \mathsf{E}\Big[W \mathbb{1}_{\{W \in [0,\gamma]^{\text{c}} \cap \mathcal{B}_k\}}\Big] + \mathsf{E}\Big[W \mathbb{1}_{\{W \in [0,\gamma] \cap \mathcal{B}_k^{\text{c}}\}}\Big] \\ &\leq \mathsf{E}\Big[W \mathbb{1}_{\{W \in \mathcal{B}_k\}}\Big] - \gamma \mu\Big([0,\gamma]^{\text{c}} \cap \mathcal{B}_k\Big) + \gamma \mu\Big([0,\gamma] \cap \mathcal{B}_k^{\text{c}}\Big) \end{aligned} \tag{A.8}$$

and (b) follows since

$$\begin{aligned} &\sum_{k=1}^{K} \lambda_k \Big(\mu\Big([0,\gamma] \cap \mathcal{B}_k^{\text{c}}\Big) - \mu\Big([0,\gamma]^{\text{c}} \cap \mathcal{B}_k\Big) \Big) \\ &= \sum_{k=1}^{K} \lambda_k \Big(\mu([0,\gamma]) - \mu\Big([0,\gamma] \cap \mathcal{B}_k\Big) - \Big(\mu\Big(\mathcal{B}_k\Big) - \mu\Big([0,\gamma] \cap \mathcal{B}_k\Big) \Big) \Big) \\ &= \underbrace{\mu([0,\gamma])}_{= d_1^\dagger} - \underbrace{\sum_{k=1}^{K} \lambda_k \mu(\mathcal{B}_k)}_{d_1^\dagger} \\ &= 0. \end{aligned} \tag{A.9}$$

Thus, $\tilde{s}(w) = \mathbb{1}_{\{W \in Q_W^{-1}(d_\mathsf{S})\}}$ yields an upper bound for any simple $s(w)$. Since $\tilde{s}(w)$ is itself simple and satisfies the constraint in (A.3), it achieves the supremum in (A.4).

For the case of $Z \sim \mathcal{N}(0,1)$, note that W is a χ^2-distributed random variable and $Q_W = F_{\chi_1^2}^{-1}$. A direct calculation shows that

$$\begin{aligned}\mathsf{E}\Big[W\mathbb{1}_{\{W \le F_W^{-1}(d_\mathsf{S})\}}\Big] &= \int_0^{F_{\chi_1^2}^{-1}(d_\mathsf{S})} w \frac{1}{2^{1/2}\Gamma(1/2)} w^{1/2-1} e^{-w/2} \mathrm{d}w \\ &\overset{\text{a}}{=} \int_0^{F_{\chi_1^2}^{-1}(d_\mathsf{S})} \frac{1}{2^{3/2}\Gamma(3/2)} w^{3/2-1} e^{-w/2} \mathrm{d}w \\ &= F_{\chi_3^2}\Big(F_{\chi_1^2}^{-1}(d_\mathsf{S})\Big)\end{aligned} \tag{A.10}$$

where (a) follows from the relation $z\Gamma(z) = \Gamma(z+1)$ for the gamma function. ■

A.2. Proof of Theorem 4.4

Proof of Theorem 4.4. To derive (4.22), we start by expanding

$$\begin{aligned}\mathsf{R}(d) &= \inf_{\substack{P_{Y|X}:\\ \mathsf{E}\left[(X-Y)^2\right] \le d}} I(X;Y) \\ &= \inf_{\substack{P_{Y|X}:\\ (1-p)\,\mathsf{E}[(X-Y)^2 \,|\, X=0] \\ \quad +p\,\mathsf{E}[(X-Y)^2 \,|\, X\neq 0] \le d}} I(X;Y) \\ &= \min_{\substack{d_0', d_\mathsf{S}':\\ (1-p)d_0' + p d_\mathsf{S}' \le d}} \inf_{\substack{P_{Y|X}:\\ \mathsf{E}[(X-Y)^2 \,|\, X=0] \le d_0' \\ \mathsf{E}[(X-Y)^2 \,|\, X\neq 0] \le d_\mathsf{S}'}} I(X;Y) \\ &= \min_{\substack{d_0', d_\mathsf{S}':\\ (1-p)d_0' + p d_\mathsf{S}' \le d}} \mathsf{R}_\mathsf{L}^{\mathrm{BSS,mse}}(d_0', d_\mathsf{S}')\end{aligned} \tag{A.11}$$

where $\mathsf{R}_\mathsf{L}^{\mathrm{BSS,mse}}(d_0', d_\mathsf{S}')$ is defined in (4.23a) – (4.23b). Let $P_{Y|X}^\star$ achieve the aforementioned minimum. Fix some $\gamma > 0$ and define $\hat{B} = \mathbb{1}_{\{|Y|>\gamma\}}$. Similar to (4.12) – (4.13), we can write

$$I(X;Y) \ge I\Big(B;\hat{B}\Big) + p\Big(h(Z) - \frac{1}{2}\log(2\pi e d_\mathsf{S}')\Big). \tag{A.12}$$

Next, we lower bound the first term in (A.12) by the RDL function for the binary memoryless source with two Hamming distortion constraints. To this end, we determine appropriate Hamming distortion constraints in a similar manner as in the proof of Theorem 4.2.

From the first constraint (4.23a), we get

$$\begin{aligned} d_0' &\geq \mathsf{E}[(X-Y)^2 \,|\, B=0] \\ &= \mathsf{E}[Y^2 \,|\, B=0] \\ &\geq \mathsf{E}[Y^2 \mathbb{1}_{\{\hat{B}=1\}} \,|\, B=0] \\ &\geq \mathsf{E}[\gamma^2 \mathbb{1}_{\{\hat{B}=1\}} \,|\, B=0] \\ &= \gamma^2 \Pr[\hat{B}=1 \,|\, B=0]. \end{aligned} \tag{A.13}$$

Therefore, the joint distribution of B and $\hat{B}$ induced by $P^{\star}_{Y|X}$ satisfies

$$\Pr[\hat{B}=1 \,|\, B=0] \leq d_0/\gamma^2. \tag{A.14}$$

The second step of the argument is to upper bound $\Pr[\hat{B}=0 \,|\, B=1]$. Using the second constraint (4.23a), we have

$$\begin{aligned} d_{\mathsf{S}}' &\geq \mathsf{E}[(X-Y)^2 \,|\, B=1] \\ &\geq \mathsf{E}\Big[(X-Y)^2 \mathbb{1}_{\{|X|>\gamma, \hat{B}=1\}} \Big| B=1\Big] \\ &\geq \mathsf{E}\Big[(|X|-\gamma)^2 \mathbb{1}_{\{|X|>\gamma, \hat{B}=1\}} \Big| B=1\Big] \\ &\geq \Pr[|X>\gamma \,|\, B=1] \mathsf{E}\Big[(|X|-\gamma)^2 \mathbb{1}_{\{|X|>\gamma, \hat{B}=1\}} \Big| |X|>\gamma, B=1\Big]. \end{aligned} \tag{A.15}$$

Now note that every distribution $P_{Y|X}$ induces a distribution $P_{\hat{B}|B}$ and, similar to (4.16), consider the following set of distributions

$$\mathcal{Q}(d_{\mathsf{S}}') \coloneqq \left\{ P_{Y|X} \,:\, \mathsf{E}\Big[(|X|-\gamma)^2 \mathbb{1}_{\{|X|>\gamma, \hat{B}=1\}} \Big| B=0\Big] \leq d_{\mathsf{S}}' \right\} \tag{A.16}$$

that, by (A.15), contains all distributions satisfying the required constraint (4.23a). Thus, the joint distribution of B and $\hat{B}$ satisfies

$$\begin{aligned} &\Pr[\hat{B}=0 \,|\, B=1] \\ &\quad \leq \sup_{P_{Y|X} \in \mathcal{Q}(d_{\mathsf{S}}')} \Pr[\hat{B}=0 \,|\, B=1] \\ &\quad = \sup_{P_{Y|X} \in \mathcal{Q}(d_{\mathsf{S}}')} \Big(\Pr\Big[\hat{B}=0, |X| \geq \gamma \Big| B=1\Big] + \Pr[\hat{B}=0, |X|<\gamma \,|\, B=1] \Big) \\ &\quad \leq \Pr[|X| \geq \gamma \,|\, B=1] \sup_{P_{Y|X} \in \mathcal{Q}(d_{\mathsf{S}}')} \Pr\Big[\hat{B}=0 \Big| B=1, |X| \geq \gamma\Big] \\ &\qquad + \Pr[|X|<\gamma \,|\, B=1]. \end{aligned} \tag{A.17}$$

Now let $W \coloneqq (|Z| - \gamma)^2 \mathbb{1}_{\{|Z| \geq \gamma\}}$ and $\mathcal{Z} \coloneqq \{|Z| > \gamma\}$. To complete the proof, it remains to show that

$$q^\dagger(d'_{\mathsf{S}}) \coloneqq \sup_{P_{Y|X} \in \mathcal{Q}(d'_{\mathsf{S}})} \Pr[\hat{B} = 0 \,|\, \mathcal{Z}, B = 1] \tag{A.18}$$

is the solution to

$$\frac{d'_{\mathsf{S}}}{\Pr[\mathcal{Z}, B = 1]} = \mathsf{E}\Big[W \mathbb{1}_{\{W \leq Q_{W|\mathcal{Z}}(q)\}} \Big| \mathcal{Z}, B = 1\Big] \tag{A.19}$$

because it is the distortion constraint given by (A.15). But this is implied by Lemma 4.3. ■

A.3. Proof of Theorem 4.6

We show that $\tilde{\mathsf{R}}_{\mathrm{LB}}^{\mathrm{BSS,mse}}(d)$ and therefore also $\mathsf{R}_{\mathrm{LB}}^{\mathrm{BSS,mse}}(d)$ converge to $\mathsf{R}_0(d)$ as the distortion d approaches zero.

To this end, we resolve the supremum in (4.30) by showing that for a good choice of γ, the first term in (4.30) converges to $H_2(p)$ as d tends to zero. We will do this in three steps.

Step 1 (Sequences): Let $\{d_n\}_{n=1}^\infty$ be a sequence of positive real numbers $d_1 > d_2 > \cdots$ with $d_n \overset{n\to\infty}{\longrightarrow} 0$. Further, for every n, let $\mathcal{Z}_n \coloneqq \{|Z| > \gamma_n\}$ and choose $\gamma_n = d_n^{1/4}$. Define W_n and $Q_{W_n|\mathcal{Z}}$ according to (4.24)–(4.26) for each n. For sufficiently large n, we define

$$\tilde{q}_n \coloneqq g^-_{W_n|\mathcal{Z}_n}\left(\frac{d_n}{p \Pr[\mathcal{Z}_n \,|\, B = 1]}\right) \tag{A.20}$$

according to (4.28).

Step 2 (Large n): Fix some arbitrary $\varepsilon > 0$. By construction, there is an n_0 such that for all $n \geq n_0$ the following holds:

$$\varepsilon^2 \geq \frac{d_n}{p \Pr\Big[\mathcal{Z}_n \Big| B = 1\Big]} = g_{W_n|\mathcal{Z}_n}(\tilde{q}_n) = \mathsf{E}\Big[W_n \mathbb{1}_{\{W_n \leq Q_{W_n|\mathcal{Z}_n}(\tilde{q}_n)\}} \Big| \mathcal{Z}_n, B = 1\Big]. \tag{A.21}$$

By Markov's inequality, this implies that for all $n \geq n_0$, we have

$$\varepsilon \geq \Pr\Big[W_n \mathbb{1}_{\{W_n \leq Q_{W_n|\mathcal{Z}_n}(\tilde{q}_n)\}} \geq \varepsilon \Big| \mathcal{Z}_n, B = 1\Big]. \tag{A.22}$$

Resolving the indicator function above, we can rewrite (A.22) as

$$\begin{aligned}
\varepsilon &\geq \Pr\Big[W_n \in \big[\varepsilon, Q_{W_n|\mathcal{Z}_n}(\tilde{q}_n)\big]\Big|\mathcal{Z}_n, B=1\Big] \\
&\overset{\text{a}}{\geq} \Pr\Big[W_n \leq Q_{W_n|\mathcal{Z}_n}(\tilde{q}_n)\Big|\mathcal{Z}_n, B=1\Big] - \Pr\Big[W_n \leq \varepsilon\Big|\mathcal{Z}_n, B=1\Big] \\
&= \tilde{q}_n - \Pr\Big[W_n \leq \varepsilon\Big|\mathcal{Z}_n, B=1\Big] \qquad \text{(A.23)}
\end{aligned}$$

where (a) is an inequality since ε might be larger than $Q_{W_n|\mathcal{Z}_n}(\tilde{q}_n)$.

Step 3 (The limit): Starting with the nonnegativity of $q_{d_n^{1/4}}(d_n)$ as given by (4.28), we have

$$\begin{aligned}
0 &\leq \limsup_{n\to\infty}\ q_{d_n^{1/4}}(d_n) \\
&\overset{\text{a}}{=} \limsup_{n\to\infty} \Big(\Pr[\mathcal{Z}_n \,|\, B=1]\,\tilde{q}_n + \Pr[\mathcal{Z}_n^{\text{c}} \,|\, B=1]\Big) \\
&\overset{\text{b}}{\leq} \limsup_{n\to\infty} \Big(\Pr[\mathcal{Z}_n \,|\, B=1]\Big(\varepsilon + \Pr[W_n \leq \varepsilon \,|\, \mathcal{Z}_n, B=1]\Big) + \Pr[\mathcal{Z}_n^{\text{c}} \,|\, B=1]\Big) \\
&\leq \limsup_{n\to\infty} \Big(\varepsilon + \Pr\Big[\mathcal{Z}_n, W_n \leq \varepsilon\Big|B=1\Big] + \Pr[\mathcal{Z}_n^{\text{c}} \,|\, B=1]\Big) \\
&\overset{\text{c}}{\leq} \limsup_{n\to\infty} \left(\varepsilon + \Pr\Big[|Z| \in (d_n^{1/4}, d_n^{1/4} + \sqrt{\varepsilon}]\Big|B=1\Big] + \Pr\Big[|Z| \leq d_n^{1/4}\Big|B=1\Big]\right) \\
&\leq \limsup_{n\to\infty} \left(\varepsilon + F_{|Z|}\Big(d_n^{1/4} + \sqrt{\varepsilon}\Big) + F_{|Z|}\Big(d_n^{1/4}\Big)\right) \\
&\overset{\text{d}}{=} \varepsilon + F_{|Z|}(\sqrt{\varepsilon}) + F_{|Z|}(0). \qquad \text{(A.24)}
\end{aligned}$$

The steps are justified as follows:

(a) uses the definition of $q_{d_n^{1/4}}$ (4.28) and (A.20).

(b) follows from (A.23).

(c) uses the definition of W_n (4.24) and its distribution function conditioned on $\mathcal{Z}_n$ (4.25).

(d) holds since every cumulative distribution function is right continuous.

Since Z^2 has a PDF and (A.24) holds for any arbitrary $\varepsilon > 0$, we conclude that $\lim_{n\to\infty} q_{d_n^{1/4}}(d_n) \to 0$.

Putting everything together and using the continuity of $\mathsf{R}_{\mathsf{L}}^{\text{BMS}}(\cdot)$ in both arguments, we see that

$$\lim_{n\to\infty} \mathsf{R}_{\mathsf{L}}^{\text{BMS}}\left(\frac{\sqrt{d_n}}{(1-p)}, q_{d_n^{1/4}}(d_n)\right) = \mathsf{R}_{\mathsf{L}}^{\text{BMS}}(0,0) = H_2(p) \qquad \text{(A.25)}$$

which concludes the proof of Theorem 4.4.

B

Proofs for Chapter 5

B.1. Proof of Theorem 5.3

We provide a more detailed derivation of the thresholding-based inner bound.

Proof. We start with (5.17) – (5.19) for U_1, U_2 and the decoders g_1 and g_2. Let $\hat{B}_j \coloneqq \mathbb{1}_{\{|X_j|>\tau_j\}}$ for $j = 1, 2$. The rate bound for user 1 as given by the Berger-Tung inner bound (Theorem 5.1) is

$$
\begin{aligned}
&I(X_1; U_1 | U_2) \\
&\quad = I\Big(X_1; U_1, \hat{B}_1 \Big| U_2\Big) \\
&\quad = I\Big(X_1; \hat{B}_1 \Big| U_2\Big) + I\Big(X_1; U_1 | U_2, \hat{B}_1\Big) \\
&\quad = H\Big(\hat{B}_1 | U_2\Big) - \underbrace{H\Big(\hat{B}_1 | U_2, X_1\Big)}_{=0} + \Pr[\mathcal{U}_{00} \cup \mathcal{U}_{01}] \underbrace{I(X_1; U_1 | U_2, \mathcal{U}_{00} \cup \mathcal{U}_{01})}_{=0} \\
&\qquad + \Pr[\mathcal{U}_{10}] I(X_1; U_1 | U_2, \mathcal{U}_{10}) + \Pr[\mathcal{U}_{11}] I(X_1; U_1 | U_2, \mathcal{U}_{11}) \\
&\quad = H\Big(\hat{B}_1 | U_2\Big) + \Pr[\mathcal{U}_{10}] I(X_1; X_1 + N_1 | \mathcal{U}_{10}) + \Pr[\mathcal{U}_{11}] I(X_1; X_1 + N_1 | U_2, \mathcal{U}_{11}) \,. \qquad \text{(B.1)}
\end{aligned}
$$

The first mutual information in (B.1) can be bounded by

$$
\begin{aligned}
I(X_1; X_1 + N_1 | \mathcal{U}_{10}) &= h(X_1 + N_1 | \mathcal{U}_{10}) - h(X_1 + N_1 | X_1, \mathcal{U}_{10}) \\
&= h(X_1 + N_1 | \mathcal{U}_{10}) - h(N_1) \\
&\overset{\text{a}}{\leq} \frac{1}{2} \log \left(\frac{\mathsf{Var}[X_1 + N_1 | \mathcal{U}_{10}]}{\sigma_1^2} \right) \\
&\overset{\text{b}}{=} \frac{1}{2} \log \left(1 + \frac{\mathsf{Var}[X_1 | \mathcal{U}_{10}]}{\sigma_1^2} \right) \qquad \text{(B.2)}
\end{aligned}
$$

where (a) is due to the maximum entropy property of Gaussian random variables [CT06b, Thm 17.2.3] and (b) holds true because X_1 and N_1 are independent given $\mathcal{U}_{10}$. For the second mutual information in (B.1), we similarly have

$$\begin{aligned} I(X_1; U_1 + N_1 | U_2, \mathcal{U}_{11}) &= h(X_1 + N_1 | U_2, \mathcal{U}_{11}) - h(X_1 + N_1 | X_1, U_2, \mathcal{U}_{11}) \\ &= h(X_1 + N_1 | U_2, \mathcal{U}_{11}) - h(N_1) \\ &\leq \frac{1}{2} \log \left(\frac{\mathsf{Var}[X_1 + N_1 | U_2, \mathcal{U}_{11}]}{\sigma_1^2} \right) \\ &= \frac{1}{2} \log \left(1 + \frac{\mathsf{Var}[X_1 | U_2, \mathcal{U}_{11}]}{\sigma_1^2} \right) \\ &\leq \frac{1}{2} \log \left(1 + \frac{\mathsf{lmmse}(X_1; U_2 | \mathcal{U}_{11})}{\sigma_1^2} \right) \end{aligned} \tag{B.3}$$

where the last line follows because the conditional variance represents the MMSE when estimating X_1 from U_2 which is upper bounded by the LMMSE. This bound is done for numerical convenience. The LMMSE is given by

$$\begin{aligned} \mathsf{lmmse}(X_1; U_2 | \mathcal{U}_{11}) &= \mathsf{Var}[X_1 | \mathcal{U}_{11}] - \frac{\mathsf{E}[X_1 U_1 | \mathcal{U}_{11}]}{\mathsf{Var}[U_2 | \mathcal{U}_{11}]} \\ &= \mathsf{Var}[X_1 | \mathcal{U}_{11}] - \frac{\mathsf{E}[X_1 U_1 | \mathcal{U}_{11}]}{\mathsf{Var}[X_2 | \mathcal{U}_{11}] + \sigma_2^2}. \end{aligned} \tag{B.4}$$

The rate bound for R_2 is similar. For the sum rate, we have

$$\begin{aligned} & I(X_1, X_2; U_1, U_2) \\ &= I\Big(X_1, X_2; U_1, U_2, \hat{B}_1, \hat{B}_2\Big) \\ &= I\Big(X_1, X_2; \hat{B}_1, \hat{B}_2\Big) + I\Big(X_1, X_2; U_1, U_2 | \, \hat{B}_1, \hat{B}_2\Big) \\ &= H\Big(\hat{B}_1, \hat{B}_2\Big) + \Pr[\mathcal{U}_{00}] \cdot \underbrace{I(X_1, X_2; U_1, U_2 | \, \mathcal{U}_{00})}_{=0} + \Pr[\mathcal{U}_{01}] \cdot I(X_1, X_2; U_1, U_2 | \, \mathcal{U}_{01}) \\ &\quad + \Pr[\mathcal{U}_{10}] \cdot I(X_1, X_2; U_1, U_2 | \, \mathcal{U}_{10}) + \Pr[\mathcal{U}_{11}] \cdot I(X_1, X_2; U_1, U_2 | \, \mathcal{U}_{11}) \\ &\overset{\text{a}}{=} H\Big(\hat{B}_1, \hat{B}_2\Big) + \Pr[\mathcal{U}_{01}] \cdot I(X_2; U_2 | \, \mathcal{U}_{01}) \\ &\quad + \Pr[\mathcal{U}_{10}] \cdot I(X_1; U_1 | \, \mathcal{U}_{10}) + \Pr[\mathcal{U}_{11}] \cdot I(X_1, X_2; U_1, U_2 | \, \mathcal{U}_{11}) \\ &\overset{\text{b}}{\leq} H\Big(\hat{B}_1, \hat{B}_2\Big) + \frac{\Pr[\mathcal{U}_{01}]}{2} \log \left(1 + \frac{\mathsf{Var}[X_2 | \mathcal{U}_{01}]}{\sigma_1^2} \right) \\ &\quad + \frac{\Pr[\mathcal{U}_{10}]}{2} \log \left(1 + \frac{\mathsf{Var}[X_1 | \mathcal{U}_{10}]}{\sigma_1^2} \right) + \frac{\Pr[\mathcal{U}_{11}]}{2} \log \left(\frac{\det C_{\mathsf{UU}|\mathcal{U}_{11}}}{\sigma_1^2 \sigma_2^2} \right) \end{aligned} \tag{B.5}$$

where (a) follows from the Markov chain U_1–X_1–X_2–U_2 and (b) uses the upper bound

(B.2) applied to X_1, X_2 and then (X_1, X_2). The covariance matrix $C_{\mathsf{U}\mathsf{U}|\mathcal{U}_{11}}$ is given by

$$C_{\mathsf{U}\mathsf{U}|\mathcal{U}_{11}} = \begin{bmatrix} \mathsf{Var}[U_1|\mathcal{U}_{11}] & \mathsf{E}[U_1U_2\,|\mathcal{U}_{11}] \\ \mathsf{E}[U_1U_2\,|\mathcal{U}_{11}] & \mathsf{Var}[U_2|\mathcal{U}_{11}] \end{bmatrix} = \begin{bmatrix} \mathsf{Var}[X_1|\mathcal{U}_{11}] + \sigma_1^2 & \mathsf{E}[X_1X_2\,|\mathcal{U}_{11}] \\ \mathsf{E}[X_1X_2\,|\mathcal{U}_{11}] & \mathsf{Var}[X_2|\mathcal{U}_{11}] + \sigma_2^2 \end{bmatrix}. \tag{B.6}$$

It remains to determine the distortion of the LMMSE decoders in these four different scenarios. The distortion is then given by

$$\begin{aligned} \mathsf{E}[\delta(X_1, Y_1)] = {} & \Pr[\mathcal{U}_{00}] \cdot \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{00}] + \Pr[\mathcal{U}_{01}] \cdot \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{01}] \\ & + \Pr[\mathcal{U}_{10}] \cdot \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{10}] + \Pr[\mathcal{U}_{11}] \cdot \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{11}] \end{aligned} \tag{B.7}$$

which is, using the properties of the LMMSE estimators [Kay93, Ch. 12], given by

$$\begin{aligned} \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{00}] &= \mathsf{E}[X_1^2\,|\mathcal{U}_{00}] \\ \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{01}] &= \mathsf{E}[X_1^2\,|\mathcal{U}_{01}] - \frac{\mathsf{E}[X_1U_2\,|\mathcal{U}_{01}]^2}{\mathsf{E}[U_2^2\,|\mathcal{U}_{01}]} \\ &= \mathsf{E}[X_1^2\,|\mathcal{U}_{01}] - \frac{\mathsf{E}[X_1X_2\,|\mathcal{U}_{01}]^2}{\mathsf{E}[X_2^2\,|\mathcal{U}_{01}] + \sigma_2^2} \\ \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{10}] &= \mathsf{E}[X_1^2\,|\mathcal{U}_{10}] - \frac{\mathsf{E}[X_1U_1\,|\mathcal{U}_{10}]^2}{\mathsf{E}[U_1^2\,|\mathcal{U}_{10}]} \\ &= \mathsf{E}[X_1^2\,|\mathcal{U}_{10}] - \frac{\mathsf{E}[X_1^2\,|\mathcal{U}_{10}]^2}{\mathsf{E}[X_1^2\,|\mathcal{U}_{10}] + \sigma_1^2} \\ \mathsf{E}[\delta(X_1, Y_1)\,|\mathcal{U}_{11}] &= \mathsf{E}[X_1^2\,|\mathcal{U}_{11}] - C_{X_1\mathsf{U}|\mathcal{U}_{11}} C_{\mathsf{U}\mathsf{U}|\mathcal{U}_{11}}^{-1} C_{\mathsf{U}X_1|\mathcal{U}_{11}} \\ &= \frac{\mathsf{E}[X_1^2\,|\mathcal{U}_{11}]\sigma_1^2\big(\mathsf{E}[X_2^2\,|\mathcal{U}_{11}] + \sigma_2^2\big) - \mathsf{E}[X_1X_2\,|\mathcal{U}_{11}]^2\sigma_1^2}{\big(\mathsf{E}[X_1^2\,|\mathcal{U}_{11}] + \sigma_1^2\big)\big(\mathsf{E}[X_2^2\,|\mathcal{U}_{11}] + \sigma_2^2\big) - \mathsf{E}[X_1X_2\,|\mathcal{U}_{11}]^2}. \end{aligned} \tag{B.8}$$

The rate bound and distortions for user 2 are computed similarly. ■

C

Abbreviations

List of Abbreviations

AMP	Approximate Message Passing
BMS	Binary Memoryless Source
BP	Belief Propagation
BSS	Bernoulli Spike Source
CS	Compressed Sensing
DBGS	Distributed Bernoulli-Gaussian Source
GAMP	Generalized Approximate Message Passing
GLM	Generalized Linear Model
GMS	Gaussian Memoryless Source
iid	independent and identically distributed
JPEG	Joint Photographic Experts Group
LB	Lower Bound
LMMSE	Linear Minimum Mean Squared Error
MAMP	Multi-Terminal Approximate Message Passing
MGAMP	Multi-Terminal Generalized Approximate Message Passing
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
PDF	probability density function
PMF	probability mass function
QCS	Quantized Compressed Sensing
RD	Rate-Distortion
RDL	Letter-Based Rate-Distortion

RHS	right hand side
RIP	Restricted Isometry Property
SE	State Evolution
UB	Upper Bound

Bibliography

[Ari72] Siguru Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Trans. Inf. Theory*, 18(1):14–20, Jan 1972.

[BB08] Petros T. Boufounos and Richard G. Baraniuk. 1-bit compressive sensing. In *Proc. 42nd Ann. Conf. on Inf. Sci. and Sys.*, pages 16–21, Mar 2008.

[BDW+09] Dror Baron, Marco F. Duarte, Michael B. Wakin, Shriram Sarvotham, and Richard G. Baraniuk. Distributed compressive sensing. *arXiv:0901.3403*, 2009.

[Ber71] Toby Berger. *Rate-Distortion Theory: A Mathematical Basis for Data Compression*. Prentice-Hall, 1971.

[Ber78] T. Berger. *Multiterminal Source Coding*, volume 229 of *G. Longo Ed.* Springer New York, 1978.

[BFN+17] Richard G. Baraniuk, Simon Foucart, Deanna Needell, Yaniv Plan, and Mary Wootters. Exponential decay of reconstruction error from binary measurements of sparse signals. *IEEE Trans. Inf. Theory*, 63(6):3368–3384, Jun 2017.

[BG98] Toby Berger and Jerry D. Gibson. Lossy source coding. *IEEE Trans. Inf. Theory*, 44(6):2693–2723, Oct 1998.

[BKM+19] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proc. US Nat. Acad. Sci.*, 116(12):5451–5460, Mar 2019.

[Bla72] Richard Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inf. Theory*, 18(4):460–473, Jul 1972.

[BM11] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory*, 57(2):764–785, Feb 2011.

[BMDK17] Jean Barbier, Nicolas Macris, Mohamad Dia, and Florent Krzakala. Mutual information and optimality of approximate message-passing in random linear estimation. *arXiv:1701.05823*, 2017.

[BSK15] Jean Barbier, Christophe Schülke, and Florent Krzakala. Approximate message-passing with spatially structured operators, with applications to compressed sensing and sparse superposition codes. *J. of Statistics and Mechanics: Theory and Experiment*, 2015(5):P05013, May 2015.

[BY89] Toby Berger and Raymond W. Yeung. Multiterminal source coding with one distortion criterion. *IEEE Trans. Inf. Theory*, 35(2):228–236, Mar 1989.

[Cha10] Cheng Chang. On the rate distortion function of Bernoulli Gaussian sequences. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 66–70, Jun 2010.

[Cou18] Thomas A. Courtade. A strong entropy power inequality. *IEEE Trans. Inf. Theory*, 64(4):2173–2192, Apr 2018.

[CP11] Emmanuel J. Candés and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inf. Theory*, 57(4):2342–2359, Apr 2011.

[CRT06a] Emmanuel Candés, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, Feb 2006.

[CRT06b] Emmanuel Candés, Justin Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. on Pure and Applied Math.*, 59(8):1207–1223, Aug 2006.

[Csi74] I. Csiszár. On an extremum problem of information theory. *Studia Sci. Mathem. Hungarica*, 9:57–71, 1974.

[CT06a] Emmanuel J. Candés and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, Dec 2006.

[CT06b] Thomas M. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, 2006.

[CW08] Emmanuel J. Candés and Michael B. Wakin. An introduction to compressive sampling. *IEEE Sig. Proc. Mag.*, 25(2):21–30, Mar 2008.

[CW14] Thomas A. Courtade and Tsachy Weissman. Multiterminal source coding under logarithmic loss. *IEEE Trans. Inf. Theory*, 60(1):740–761, Jan 2014.

[DJM13] David L. Donoho, Adel Javanmard, and Andrea Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE Trans. Inf. Theory*, 59(11):7434–7464, Nov 2013.

[DM18] Sjoerd Dirksen and Shahar Mendelson. Non-Gaussian hyperplane tesselations and robust one-bit compressed sensing. *arXiv:1805.09409*, Aug 2018.

[DMM09] David L. Donoho, Arian Maleki, and Andrea Montanari. Message passing algorithms for compressed sensing. *Proc. US Nat. Acad. Sci.*, 106(45):18914–18919, Nov 2009.

[DMM10a] David L. Donoho, Arian Maleki, and Andrea Montanari. Message passing algorithms for compressed sensing: I. motivation and construction. In *Proc. IEEE Inf. Theory Workshop*, Jan 2010.

[DMM10b] David L. Donoho, Arian Maleki, and Andrea Montanari. Message passing algorithms for compressed sensing: Ii. analysis and validation. In *Proc. IEEE Inf. Theory Workshop*, Jan 2010.

[Don06] David L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, Apr 2006.

[DSW+05] Marco F. Duarte, Shriram Sarvotham, Michael B. Wakin, Dror Baron, and Richard G. Baraniuk. Joint sparsity models for distributed compressed sensing. In *Proc. Workshop on Sig. Proc. with Adaptive Sparse Structured Representations*, Nov 2005.

[EK12] Yonina C. Eldar and Gitta Kutyniok. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[EM09] Yonina C. Eldar and Moshe Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory*, 55(11):5302–5316, 2009.

[ER10] Yonina C. Eldar and Holger Rauhut. Average case analysis of multichannel sparse recovery using convex relaxation. *IEEE Trans. Inf. Theory*, 56(1):505–519, Jan 2010.

[EYRW15] Armin Eftekhari, Han Lun Yap, Christopher J. Rozell, and Michael B. Wakin. The restricted isometry property for random block diagonal matrices. *Applied and Computational Harmonic Analysis*, 38(1):1–31, 2015.

[Ezz18] Rami Ezzine. Two-terminal quantized compressive sensing with approximate message passing reconstruction. research internship report at Technische Universität München, Jun 2018.

[Fou16] Simon Foucart. Flavors of compressive sensing. In *Proc. Int. Conf. Approx. Theory*, May 2016.

[FR13] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.

[GK11] A. E. Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[GLZ99] A. György, T. Linder, and K Zeger. On the rate-distortion function of random vectors and stationary sources with mixed distributions. *IEEE Trans. Inf. Theory*, 45(6):2110–2115, Sep 1999.

[Gra11] Robert M. Gray. *Entropy and Information Theory*. Springer International Publishing, second edition, 2011.

[GVT98] Vivek K. Goyal, Martin Vetterli, and Nguyen T. Thao. Quantized overcomplete expansions in $\mathbb{R}^N$: Analysis, synthesis, and algorithms. *IEEE Trans. Inf. Theory*, 44(1):16–31, Jan 1998.

[GW19] Chen Gong and Xiaodong Wang. On finite block-length quantization distortion. *IEEE Trans. Inf. Theory*, 65(2):1172–1188, Feb 2019.

[Hag14a] Saeid Haghighatshoar. *Compressed Sensing of Memoryless Sources: A Deterministic Hadamard Construction*. PhD thesis, EPFL, Dec 2014.

[Hag14b] Saeid Haghighatshoar. Multi terminal probabilistic compressed sensing. In *Proc. IEEE Int. Symp. Inf. Theory*, Jul 2014.

[Hal74] Paul R. Halmos. *Measure Theory*. Springer, 1974.

[IK11] Amir Ingber and Yuval Kochman. The dispersion of lossy source coding. In *Proc. Data Proc. Conf.*, pages 53–62, Mar 2011.

[Jac16] Laurent Jacques. Error decay of (almost) consistent signal estimations from quantized gaussian random projections. *IEEE Trans. Inf. Theory*, 62(8):4696–4709, Aug 2016.

[JDV13] Laurent Jacques, Kevin Degraux, and Christophe De Vleeschouwer. Quantized iterative hard thresholding: Bridging 1-bit and high-resolution quantized compressed sensing. In *Proc. 10th Int. Conf. on Sampling Theory and Applications*, Jul 2013.

[JLBB13] Laurent Jacques, Jason Noah Laska, Petros T. Boufounos, and Richard G. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Trans. Inf. Theory*, 59(4):2082–2102, Apr 2013.

[JM12] Adel Javanmard and Andrea Montanari. Subsampling at information theoretically optimal rates. In *Proc. IEEE Int. Symp. Inf. Theory*, pages pp. 2431–2435, 2012.

[JM13] A. Javanmard and A. Montanari. State evolution for general approximate message passing algorithms, with applications to compressed sensing. *Inf. and Inference: A J. of the IMA*, 2(2):115–144, Dec 2013.

[Kay93] Steven M. Kay. *Fundementals of Statistical Signal Processing*. Prentice Hall PTR, 1993.

[KGK+19] Swatantra Kafle, Vipul Gupta, Bhavya Kailkhura, Thakshila Wimalajeewa, and Pramod K. Varshney. Joint sparsity pattern recovery with 1-b compressive sensing in distributed sensor networks. *IEEE Trans. Sig. Proc.*, 5(1):15–30, Mar 2019.

[KGR12] Ulugbek S. Kamilov, Vivek K. Goyal, and Sundeep Rangan. Message-passing de-quantization with applications to compressed sensing. *IEEE Trans. Sig. Proc.*, 60(12):6270–6281, Dec 2012.

[KKWV16] Swatantra Kafle, Bhavya Kailkhura, Thakshila Wimalajeewa, and Pramod K. Varshney. Decentralized joint sparsity pattern recovery using 1-bit compressive sensing. In *Proc. IEEE Global Conf. Sig. and Inf. Proc.*, pages 1354–1358, Dec 2016.

[KMS+12a] Florent Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and Lenka Zdeborová. Statistical-physics-based reconstruction in compressed sensing. *Physical Review X*, 2(2), Apr 2012.

[KMS+12b] Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *J. of Statistics and Mechanics: Theory and Experiment*, 2012(8):P08009, Aug 2012.

[Koc16] Tobias Koch. The Shannon lower bound is asymptotically tight. *IEEE Trans. Inf. Theory*, 62(11):6155–6161, Nov 2016.

[Kos13] Victoria Kostina. *Lossy Data Compression: Nonasymptotic Fundamental Limits*. PhD thesis, Princeton University, Sep 2013.

[Kos17] Victoria Kostina. Data compression with low distortion and finite blocklength. *IEEE Trans. Inf. Theory*, 63(7):4268–4285, Jul 2017.

[KSW16] Karin Knudson, Rayan Saab, and Rachel Ward. One-bit compressive sensing with norm estimation. *IEEE Trans. Inf. Theory*, 62(5):2748–2758, May 2016.

[KV12] V. Kostina and S. Verdú. Fixed-length lossy compression in the finite blocklength regime. *IEEE Trans. Inf. Theory*, 58(6):3309–3338, Jun 2012.

[LDSP08] Michael Lustig, David L. Donoho, Juan M. Santos, and John M. Pauly. Compressed sensing MRI. *IEEE Sig. Proc. Mag.*, 25(2):72–82, Mar 2008.

[Li11] S. Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian J. of Mathem. and Stat.*, 4(1):66–70, 2011.

[LT02] Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces: isoperimetry and processes*. Springer Verlag, 2002.

[LZ94] T. Linder and R. Zamir. On the asymptotic tightntess of the Shannon lower bound. *IEEE Trans. Inf. Theory*, 40(6):2026–2031, Nov 1994.

[MP19] Johannes Maly and Lars Palzer. Analysis of hard-thresholding for distributed compressed sensing with one-bit measurements. *Inf. and Inference: A J. of the IMA*, Apr 2019.

[Ooh97] Yasutada Oohama. Gaussian multiterminal source coding. *IEEE Trans. Inf. Theory*, 43(6):1912–1923, Nov 1997.

[Pin67] John T. Pinkston III. Encoding independent sample information sources. Technical Report 462, MIT Research Laboratory of Electronics, Oct 1967.

[PT16a] Lars Palzer and Roy Timo. A converse for lossy source coding in the finite blocklength regime. In *Proc. Int. Zurich Sem. on Comm.*, Mar 2016.

[PT16b] Lars Palzer and Roy Timo. Fixed-length compression for letter-based fidelity measures in the finite blocklength regime. In *Proc. IEEE Int. Symp. Inf. Theory*, Jul 2016.

[PT16c] Lars Palzer and Roy Timo. A lower bound for the rate-distortion function of spike sources that is asymptotically tight. In *Proc. IEEE Inf. Theory Workshop*, Sep 2016.

[PV13a] Y. Plan and R. Vershynin. One-bit compressed sensing by linear programming. *Comm. Pure Appl. Math.*, 66(8):1275–1297, Aug 2013.

[PV13b] Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Trans. Inf. Theory*, 59(1):482–494, Jan 2013.

[PV14] Yaniv Plan and Roman Vershynin. Dimension reduction by random hyperplane tesselation. *Discrete & Computational Geometry*, 51(2):438–461, Mar 2014.

[PW17] Yury Polyanskiy and Yihong Wu. *Lecture notes on Information Theory*. MIT (6.441), UIUC (ECE 563), Yale (STAT 664), 2017.

[Ran11] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *Proc. IEEE Int. Symp. Inf. Theory*, pages 2168–2172, Aug. 2011.

[RB88] H. Rosenthal and J. Binia. On the epsilon entropy of mixed random variables. *IEEE Trans. Inf. Theory*, 34(5):1110–1114, Sep 1988.

[RL14] Xiongbin Rao and Vincent K. N. Lau. Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems. *IEEE Trans. Sig. Proc.*, 62(12):3261–3271, Jun 2014.

[Rog63] C. A. Rogers. Covering a sphere with spheres. *Mathematika*, 10:157–164, 1963.

[RV08] Mark Rudelson and Roman Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. on Pure and Applied Math.*, 61(8):1025–1045, Aug 2008.

[Sch06] Gideon Schechtman. Two observations regarding embedding subsets of Euclidean spaces in normed spaces. *Advances in Mathematics*, 200(1):125–135, Feb 2006.

[SCS14] Dennis Sundman, Saikat Chatterjee, and Mikael Skoglund. Methods for distributed compressed sensing. *J. Sens. Actuator Netw.*, 3(1):1–25, 2014.

[Sha48] Claude Elwood Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27(3):379–423, Jul 1948.

[Sha59] Claude Elwood Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE Int. Conv. Rec.*, 7:142–163, 1959.

[SRF16] Philip Schniter, Sundeep Rangan, and Alyson K. Fletcher. Vector approximate message passing for the generalized linear model. In *Proc. 50th Asilomar Conf. on Sig., Sys., and Comp.*, pages 1525–1529, Nov 2016.

[SW73] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theory*, IT-19(4):471–480, Jul 1973.

[Tal14] Michel Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems*. Springer Science & Business Media, 2014.

[Tun78] S.-Y. Tung. *Multiterminal Source Coding*. PhD thesis, Cornell University, 1978.

[Tur10] Sebastien Turban. Convolution of a truncated normal and a centered normal variable. Online at http://www.columbia.edu/∼st2511/notes, 2010.

[TV94] Nguyen T. Thao and Martin Vetterli. Reduction of the mse in r-times oversampled r/d conversion from $\mathcal{O}(1/r)$ to $\mathcal{O}(1/r^2)$. *IEEE Trans. Signal Process.*, 42(1):200–203, Jan 1994.

[TXY14] Yuan Tian, Wenbo Xu, and Hongwen Yang. A distributed compressed sensing scheme based on one-bit quantization. In *Proc. 79th IEEE Vehic. Technol. Conf.*, May 2014.

[vdM] Jean van der Meulen.

[Ver18] Roman Vershynin. *High-Dimensional Probability.* Cambridge University Press, 2018.

[VG05] Jean-Louis Verger-Gaugry. Covering a ball with smaller balls in $\mathbb{R}^n$. *Discrete & Computational Geometry*, 33:143–155, 2005.

[VS11] J. P. Vila and Philip Schniter. Expectation-maximization bernoulli-gaussian approximate message passing. In *Proc. 45th Asilomar Conf. on Sig., Sys., and Comp.*, Nov 2011.

[WKA11] Aaron B. Wagner, Benjamin G. Kelly, and Yücel Altuğ. Distributed rate-distortion with common components. *IEEE Trans. Inf. Theory*, 57(7):4035–4057, Jul 2011.

[WTV08] Aaron B. Wagner, Saurabha Tavildar, and Pramod Viswanath. Rate-region of the quadratic Gaussian two-encoder source-coding problem. *IEEE Trans. Inf. Theory*, 54(5):1938–1961, May 2008.

[WV12a] Claudio Weidmann and Martin Vetterli. Rate distortion behavior of sparse sources. *IEEE Trans. Inf. Theory*, 58(8):4969–4992, Aug 2012.

[WV12b] Yihong Wu and Sergio Verdú. Optimal phase transitions in compressed sensing. *IEEE Trans. Inf. Theory*, 58(10):6241–6263, Oct 2012.

[WZT+14] Yin Wu, Yan-Jie Zhu, Qiu-Yang Tang, Chao Zhu, Wei Liu, Rui-Bin Dai, Xin Liu, Ed X. Wu, Leslie Ying, and Dong Liang. Accelerated mr diffusion tensor imaging using distributed compressed sensing. *Magnetic Resonance in Medicine*, 71(2):764–772, 2014.

[ZB99] Ram Zamir and Toby Berger. Multiterminal source coding with high resolution. *IEEE Trans. Inf. Theory*, 45(1):106–117, Jan 1999.

[ZK16] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.