Finite-Length Analysis of Lossy Compression via Sparse Regression Codes

Research Thesis

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering

Gal Livny

Submitted to the Senate
of the Technion — Israel Institute of Technology
Iyar 5779 Haifa May 2019

This research thesis was done under the supervision of Prof. Igal Sason in the Faculty of Electrical Engineering.

Acknowledgment

I would like to express my deepest gratitude to my supervisor Prof. Igal Sason, for his guidance and support throughout this research. This thesis would not have been possible without his devotion and patience, and I am truly indebted to him for the time he dedicated to this work.

The generous help of the Technion is gratefully acknowledged.

Contents

Contents - Continued

List of Figures

Abstract

One of the main goals of information theory has been to design practical codes for reliable transmission and lossy compression, approaching the fundamental information-theoretic limits with feasible complexity. Recently, a new type of codes has been suggested for reliable communication over memoryless channels, and for lossy compression of memoryless and stationary sources with continuous alphabets. These codes, named as sparse regression codes (SPARCs) or sparse superposition codes, rely on a coding technique where the codewords are linear combinations of columns of the design matrix of the code. SPARCs were originally developed for communication over the additive white Gaussian noise (AWGN) channel, and they were proved to asymptotically achieve the channel capacity. Subsequently, SPARCs were adapted for lossy compression, and it was shown that they asymptotically attain the rate-distortion function of a Gaussian memoryless source with a computational complexity which grows polynomially in the blocklength of the source output.

The main focus of this thesis is the examination of the performance of SPARCs for lossy compression of memoryless sources, obtained by tightening the existing asymptotic bounds on the probability of excess distortion, and by adapting these bounds to finite blocklengths. Furthermore, an improvement of the encoding scheme for lossy compression with SPARCs is proposed, analyzed, and examined by computer simulations. We also discuss the tradeoff between performance and complexity of SPARCs in the context of lossy compression.

List of Abbreviations and Notations

SPARC	Sparse regression code
AWGN	Additive white Gaussian noise
$\|\cdot\|$	Euclidean norm
$\lvert\cdot\rvert$	Scaled Euclidean norm
$< \cdot, \cdot >$	Inner product
$\Phi(\cdot)$	Gaussian cumulative distribution function
$\phi(\cdot)$	Gaussian probability density function
$Q(\cdot)$	Complementary Gaussian cumulative distribution function
Z_M	Maximum of M i.i.d. standard Gaussian random variables
e_M	Expectation of maximum of M i.i.d. standard Gaussian random variables
F_X	Cumulative distribution function of a random variable X
f_X	Probability density function of a random variable X
$W(\cdot)$	Lambert W function
$\bar{\gamma}(\cdot, \cdot)$	Incomplete Gamma function

Chapter 1

Introduction

1.1 Background on Sparse Regression Codes

From its inception, one of the goals of coding theory has been to design high-performance and low-complexity codes for reliable communication over noisy channels and for lossless or lossy data compression. While it is well known that such codes exist in principle when no system constraints on complexity and delay are imposed [1], the main practical challenge has been to construct codes with low computational and storage complexity for both the encoder and decoder while approaching the information-theoretic fundamental limits of source and channel coding in practice. Starting from the 1990's, several practical codes which asymptotically approach these theoretical limits have been designed such as turbo codes [2], codes defined on sparse graph (e.g., low-density parity-check (LDPC) codes) [3], polar codes [4, 5] and spatially-coupled LDPC codes [6]. These codes, however, have been mainly studied for discrete-input channels and sources with a discrete output alphabet.

There are yet many channel and source models of practical interest with a continuous alphabet. In particular, the additive white Gaussian noise (AWGN) channel and the Gaussian memoryless source have been of special interest (e.g., [1, 7, 8]). A recently developed approach for these cases is the sparse regression code (SPARC) (a.k.a. sparse superposition code) [9].

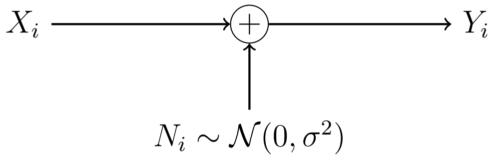

Figure 1.1: The AWGN channel

For the AWGN channel model (Figure 1.1), independent and identically distributed (i.i.d.) samples of an additive Gaussian noise with zero mean and variance σ^2 are added to input symbols which are subject to an input power constraint P. The capacity of the AWGN channel is given by $C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right)$ bits per channel use [1]. The aim of the designer is to reliably transmit information over the channel at rates approaching the channel capacity with a decoding error probability which asymptotically decays to zero when the blocklength n of the code is increased, while keeping the processing delay and encoding/decoding complexity reasonable.

Gaussian codebooks, for which each codeword is an i.i.d. Gaussian random vector, have been proved to achieve the capacity of the AWGN channel. However, these codes are not practical due to the high decoding complexity of un-structured Gaussian codebooks. In practice, the popular approach for communication over the AWGN channel is coded modulation, a method which consists of two separate steps: finite-alphabet coding and modulation; one example is a combination of a standard modulation scheme like Quadrature Amplitude Moderation (QAM) and a known capacity-achieving binary code, such as LDPC codes [10]. While these methods show good empirical results in simulation and practice, it has not been proved that they manage to achieve the capacity of the AWGN channel. Other proposed codes for communication over the AWGN channel are lattice codes; however, they are infeasible for high dimensional lattices [11].

A SPARC code is a type of a structured Gaussian codebook, which manages to achieve similar results in performance with reasonable complexity. A SPARC code has an $n \times N$ design matrix $\mathbf{A}$ with i.i.d. Gaussian entries, where n is the blocklength. The columns of the matrix $\mathbf{A}$ are divided into L sections of M columns each (see Figure 1.2), and hence $N = ML$. A SPARC codeword is a linear combination of L columns of the matrix $\mathbf{A}$, one from each section.

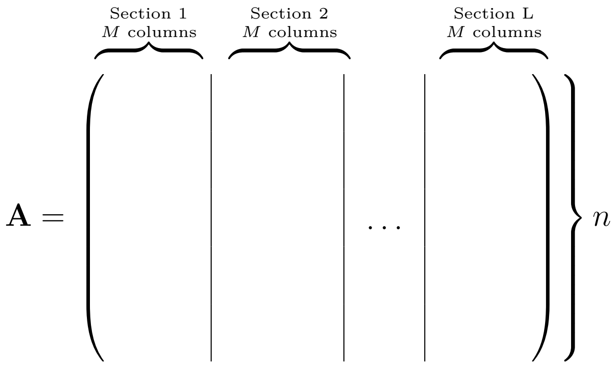

Figure 1.2: $\mathbf{A}$ is an $n \times N$ matrix, divided into L sections with M columns each

For coding over a noisy channel, the entries of the design matrix $\mathbf{A}$ are i.i.d. Gaussian random variables with zero mean and variance equal to $\frac{1}{n}$. A codeword of a SPARC is a linear combination of L columns from $\mathbf{A}$, one from each section; it is obtained by multiplying the design matrix $\mathbf{A}$ by a sparse (column) vector $\beta = \{\beta_1, \beta_2, \ldots, \beta_N\}^\top$ of length N, which contains a single non-zero entry in each of the L blocks of M consecutive entries, i.e., there exists exactly one non-zero β_i for $i \in \{1, \ldots, M\}$, one non-zero β_i for $i \in \{M+1, \ldots, 2M\}$, and finally one non-zero β_i for $i \in \{(L-1)M+1, \ldots, LM\}$. The values of the non-zero entries $\{\beta_j\}$ in each section are determined to accommodate the average power constraint P per channel symbol, and therefore they are set to $\{\sqrt{nP_j}\}_{j=1}^{L}$ with $\sum_{j=1}^{L} P_j = P$. For lossy compression, the entries of the design matrix $\mathbf{A}$ are i.i.d. samples of a standard Gaussian distribution, and the non-zero entries of the sparse vector β are determined to minimize the distortion in the reconstruction, as it is explained in Chapter 3.

Since there are L sections with M columns each, the total number of codewords in the SPARC is M^L. This holds since each such codeword corresponds to one possible choice of the sparse vector β (recall that the values of the non-zero entries in each of the L segments are fixed, and there are M possibilities for choosing the single non-zero entry of β in each segment). Hence, since the number of codewords $\exp(nR)$ is equal to M^L, the code rate R satisfies the equality

$$nR = L \log M \tag{1.1}$$

where, unless mentioned explicitly, the logarithms are expressed on the natural base. For a pair of values of blocklength n and code rate R, there are multiple valid choices for M and L. For example, picking $L = 1$, we have an unstructured Gaussian codebook since the matrix $\mathbf{A}$ is comprised of one big section; in this case, the number of columns in $\mathbf{A}$ grows *exponentially* in n. A more practical choice, which is often used for SPARCs, is $M = L^b$ for some positive constant b; here, (1.1) gives

$$nR = bL \log L, \tag{1.2}$$

$$\Rightarrow L = \exp\left(W\left(\frac{nR}{b}\right)\right), \tag{1.3}$$

where $W(\cdot)$ denotes the Lambert W function [12]; this function, as presented in Section 2.2, is the inverse function of $f(x) = xe^x$ for $x \geq -\frac{1}{e}$. In view of the following asymptotic result for the Lambert W function (see (2.25)):

$$W(x) = \log x - \log\log x + O\left(\frac{\log\log x}{\log x}\right) \tag{1.4}$$

and, since the number of columns in the design matrix $\mathbf{A}$ is equal to $ML = L^{b+1}$, it therefore follows that the number of columns in A grows only *polynomially* in n (more precisely, it scales like $\left(\frac{n}{\log n}\right)^{b+1}$).

Codewords of a SPARC are statistically dependent when they are constructed by linear combinations which contain at least one shared column; this is equivalent to having at least one shared non-zero entry in the sparse vectors which are used to construct the two codewords from the design matrix.

The decision on how to allocate the power among the L non-zero entries of the sparse vector β has a large effect on the performance of the SPARC. Two possible power allocations which have been studied in the literature are the flat power allocation where $P_j = \frac{P}{L}$ for all $j \in \{1, \dots, L\}$, and the exponentially decaying power allocation with $P_j = C\exp(-aj)$ for $a > 0$ and $j \in \{1, \dots, L\}$ (the former is a special case of the latter when $a \downarrow 0$). For both power allocations, it has been proved that there exists a decoder which asymptotically approaches the channel capacity [13, 14].

Channel Coding with Sparse Regression Codes

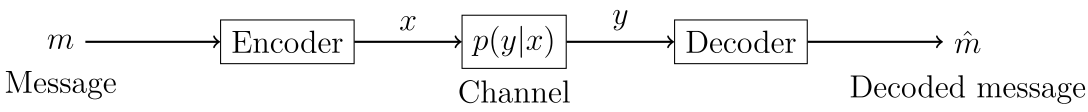

Figure 1.3: A standard channel coding set-up

For communication over the AWGN channel, encoding with SPARCs is done as follows: each input block of $k = nR$ bits is partitioned into L blocks of $\log_2 M$ bits, so that each of these blocks is a binary representation of an integer between 0 and $M - 1$. The non-zero entry in every section of the sparse vector β is at the index of that integer plus one, and the input codeword to be sent through the channel is given by $\mathbf{A}\beta$.

The error probability of a SPARC is measured by its *section error rate*,

$$\mathcal{E}_{\text{sec}} = \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{1}\left\{\hat{\beta}_\ell \neq \beta_\ell\right\}, \tag{1.5}$$

where $\hat{\beta}$ is the decoded version of the vector β, and β_ℓ represents the sub-vector of the sparse vector β in section $\ell \in \{1, \ldots, L\}$. The section error rate denotes the fraction of sections which are decoded erroneously. Assuming a uniform mapping between the input stream to the encoder (and, therefore, an equi-probable distribution for the non-zero index in each section), the bit error rate is approximately equal to one-half the section error rate, i.e., $\mathcal{E}_{\text{ber}} \approx \frac{1}{2}\mathcal{E}_{\text{sec}}$. For a given decoding algorithm, it is of interest to bound the probability $\mathbb{P}(\mathcal{E}_{\text{sec}} > \epsilon)$ as a function of $\epsilon > 0$. In order to improve the decoding error probability of the message, one can use a concatenated code where the SPARC serves as an inner code and a Reed-Solomon (RS) code is the outer code; an RS code of rate $1 - 2\epsilon$, whose symbols over the Galois field are represented by $\log_2 M$ bits, can correct up to a fraction ϵ of the section errors in the SPARC.

Several decoders were proposed and analyzed for SPARCs. The first one, analyzed in [13], is an optimal decoder when the codewords are equally likely, and it is based on the following decision rule:

$$\hat{\beta}_{\text{opt}} = \arg\min_{\beta} \|\mathbf{y} - \mathbf{A}\beta\|^2, \tag{1.6}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ is the output sequence of the channel, and the minimum is taken over all possible sparse vectors β. Joseph and Barron showed in [13] that, at all rates below the channel capacity, the error probability of the optimal decoder decays exponentially in n; however, this decoder is not practical computationally since the minimization in (1.6) is performed over an exponentially growing number of vectors in n.

The *adaptive successive decoder*, which originally appeared in [14], was the first practical decoder to be proposed for SPARCs. The decoding goes as follows: in Step 1, the decoder calculates the inner products between every column of the design matrix $\mathbf{A}$ with the normalized output sequence $\mathbf{y}$, (i.e., $\frac{\mathbf{y}}{\|\mathbf{y}\|}$), and the results are compared to a pre-specified threshold. For each inner product that exceeds this threshold, the corresponding column in $\mathbf{A}$ is picked as part of the solution, and therefore at the end of the step, we have the first estimate for the sparse vector $\hat{\beta}^1$. In step i, for $i > 1$, a residual $\mathbf{R}_i$ is produced according to $\mathbf{R}_i = \mathbf{R}_{i-1} - \mathbf{A}\hat{\beta}^i$ with $\mathbf{R}_0 = \mathbf{y}$. Then, the decoder calculates the inner products between the columns of $\mathbf{A}$ and the normalized residual $\frac{\mathbf{R}_i}{\|\mathbf{R}_i\|}$, and again picks those (from the as yet undecoded sections) who surpass the pre-specified threshold. The number of steps the decoder runs is predetermined, but it may finish beforehand if one column has been already selected from each section, or if no inner product surpasses the threshold in a certain step.

Consider the AWGN channel with average power constraint P, and assume a power allocation of the SPARC which is given by $P_\ell \propto \exp(-\alpha\ell)$ for $\ell \in \{1, \ldots, L\}$ and $\alpha > 0$. The error probability of the adaptive successive decoder was analyzed in [14], providing an upper bound which decays exponentially in L at any rate below the channel capacity. The empirical results at rates close to capacity are, however, quite large for the section error rate when practical blocklengths are used.

The next two suggested decoders for SPARCs are *iterative soft-decision decoders*. Both of them aim to iteratively update the posterior probabilities of each entry of β being the true non-zero in its section. The objective of these decoders is to produce a test statistics vector at each iteration which has the form

$$\mathrm{stat}_i \approx \beta + \tau_i Z_i, \tag{1.7}$$

where Z_i is a standard Gaussian random variable independent of β. The test statistics sequence stat_i is what one would expect to receive at the output of an AWGN channel with noise variance τ_i^2. From the test statistics vector, it is possible to extract the next estimate for β^i, by using the optimal Bayesian estimator,

$$\beta^{i+1} = \mathbb{E}[\beta|\beta + \tau_i Z_i = \mathrm{stat}_i] = \eta_i(\mathrm{stat}_i), \tag{1.8}$$

where η_i denotes the conditional expectation. For index j in section $\ell \in \{1, \ldots, L\}$ of β, we have

$$\eta_{i,j}(s) = \sqrt{nP_\ell}\, \frac{\exp(\sqrt{nP_\ell}\, s_j/\tau_i^2)}{\sum_{k\in\mathrm{sec}_\ell} \exp(\sqrt{nP_\ell}\, s_k/\tau_i^2)}. \tag{1.9}$$

In addition to updating β^i iteratively, we must also update the variance τ_i^2 accordingly, so that it reflects the variance of the difference between β and β^i. Starting with $\tau_0^2 = \sigma^2 + P$, we define

$$\tau_i^2 = \sigma^2 + \frac{1}{n}\,\mathbb{E}\big\|\beta - \eta(\beta + \tau_{i-1}Z_{i-1})\big\|^2, \tag{1.10}$$

with the expectation being over both β and Z_{i-1}. This way, the variance of the Gaussian part of stat_i would consist of two independent terms: one is the inherent Gaussian noise of the channel, and the other derived from the difference between β and its estimate β^i. The recursion function of τ_i can be rewritten as

$$\tau_i^2 = \sigma^2 + P(1 - x(\tau_{i-1})), \tag{1.11}$$

where $x_i \triangleq x(\tau_{i-1})$ is the expected value of a function of standard Gaussian random variables, and has an exact expression in [18, Sec.3]. In [18], it is shown that for rates below the channel capacity and with an exponential power allocation, the recursion of τ_i has a fixed point close to σ^2, i.e. that the gap between β^i and β diminishes.

However, the open question is still how to generate the coveted test statistics sequence stat_i, which should be approximately equal to $\beta + \tau_i Z_i$ in each step. The *adaptive-successive soft-decision decoder* proposes one method to reach this goal. It is based on the following fits: $\text{Fit}_1 \triangleq \mathbf{y}$, $\text{Fit}_2 \triangleq \mathbf{A}\beta^1, \ldots, \text{Fit}_i \triangleq \mathbf{A}\beta^i$. From these fits, we recursively define $\mathbf{G}_i$: set $\mathbf{G}_0 \triangleq \mathbf{y}$, and subsequently define $\mathbf{G}_i$ to be the part of Fit_i that is orthogonal to $\mathbf{G}_0, \mathbf{G}_1, \ldots, \mathbf{G}_{i-1}$. The actual components from which we build the test statistics vector are $\mathcal{Z}_i$, which are defined as

$$\mathcal{Z}_k = \sqrt{n}\,\frac{\mathbf{A}^\top \mathbf{G}_k}{\|\mathbf{G}_k\|} \tag{1.12}$$

for $k \geq 0$. Then, we define

$$\text{stat}_i = \tau_i \sum_{k=0}^{i} \lambda_k \mathcal{Z}_k + \beta^i. \tag{1.13}$$

The coefficients $\{\lambda_k\}$ are chosen so that the demand on the distribution of the test statistics in (1.7) is met.

Choosing suitable values for $\{\lambda_k\}$ is based upon identifying what the distribution of $\{\mathcal{Z}_k\}$ is, which was outlined in [15]. One possible choice, first proposed in [16], is

$$(\lambda_1, \lambda_2, \ldots, \lambda_i) = \left(\frac{1}{\tau_0}, -\sqrt{\frac{1}{\tau_1^2} - \frac{1}{\tau_0^2}}, \ldots, -\sqrt{\frac{1}{\tau_i^2} - \frac{1}{\tau_{i-1}^2}}\right). \tag{1.14}$$

In [16], it is shown that this choice guarantees an approximately correct distribution for stat_i; as a result, it is proved that for every rate below the channel capacity, with the exponentially decaying power allocation, the error probability of this decoder decays exponentially in $n/(\log n)^{2k+1}$, where k is the number of steps the decoder executes.

Another way to obtain a test statistics sequence stat_i which satisfies (1.7) is used by the *approximate message-passing (AMP) decoder*. Originally, the AMP decoding algorithm gives a fast solution to the problem

$$\beta = \arg\min_{\hat{\beta}} \left\{ \|\mathbf{y} - \mathbf{A}\hat{\beta}\|_2^2 + \lambda \|\hat{\beta}\|_1 \right\}, \tag{1.15}$$

for some $\lambda > 0$. This is not quite the problem which the decoder of the SPARC needs to solve, since in (1.6) the minimum is only over possible codewords, i.e. that there is exactly one non-zero entry in each section of β, rather than over all possible vectors. However, if the decoding problem could be described with min-sum like message passing updates, then an AMP decoder could find an approximate solution to it, as per [17]. This was done in [18], with the following set of update rules:

$$\mathbf{r}^i = \mathbf{y} - \mathbf{A}\beta^i + \frac{\mathbf{r}^{i-1}}{\tau_{i-1}^2}\left(P - \frac{\|\beta^i\|^2}{n}\right), \tag{1.16}$$

$$\text{stat}_i = \mathbf{A}^\top \mathbf{r}^i + \beta^i, \tag{1.17}$$

and with β^i and τ_i updating according to (1.8) and (1.10) respectively. In [19] it was proved that the error probability of the AMP decoder decays exponentially in $n/(\log n)^{2T}$, where T is the number of iterations that is required for a successful decoding.

In terms of computational complexity, both soft-decision decoders are similar, and require $O(nN)$ time; in practice, the adaptive successive soft-decision decoder is more costly, as each iteration requires orthogonalization and expensive computation of coefficients. The computational complexity can be reduced by replacing the i.i.d. Gaussian design matrix $\mathbf{A}$ with lines randomly selected from an $N \times N$ Hadamard matrix as suggested in [18]; this reduces the computational complexity of the AMP decoder to $O(ML \log L)$, and greatly improves storage complexity, as the matrix does not have to be saved in the memory.

Some other improvements to the SPARCs have recently been proposed. Spatially coupled SPARCs, in which the design matrix $\mathbf{A}$ is comprised of blocks with different variances, appear to have better empirical results than regular SPARCs [20],[21]. Other suggested techniques include new power allocation routines, using an outer LDPC code, and using an online estimate of the parameter τ_i^2 (see (1.10)) [22].

Lossy Compression with Sparse Regression Codes

As it is mentioned above, SPARCs can be also utilized for lossy compression of sources with continuous alphabet. Specifically, for a memoryless Gaussian source with zero mean and variance σ^2, SPARCs can approach its distortion-rate function with the mean-square error distortion:

$$D(R) = \sigma^2 e^{-2R}. \tag{1.18}$$

For general zero-mean ergodic sources with a fixed variance, SPARCs attain the distortion-rate function of a Gaussian memoryless source in (1.18); this is the best possible feat for universal lossy compression of zero-mean ergodic sources with fixed variance when Gaussian codebooks are utilized [23].

The construction of SPARCs for lossy source compression is similar to channel coding, with codewords of the form $A\beta$ where the design matrix $\mathbf{A}$ is composed of i.i.d. standard Gaussian entries, and the sparse vector β has a single non-zero entry in each section. The only difference in comparison to SPARCs for power-limited channel coding is that the non-zero entries of β

are not subject to satisfy a power constraint, so that they can be chosen arbitrarily to help the source encoder to reduce the distortion.

Similarly to optimal decoding in communication over a noisy channel, optimal encoding for lossy compression is based on the following rule:

$$\hat{\beta}_{opt} = \arg\min_{\beta} \|\mathbf{S} - \mathbf{A}\beta\|^2, \tag{1.19}$$

where $\mathbf{S}$ is a source sequence of length n, and the minimization is carried over all possible sparse vectors β whose non-zero entries in each of the L sections are fixed. Every non-zero index in each section of $\hat{\beta}_{opt}$ is converted into a sequence of $\log_2 M$ bits based on its index in the section, hence the rate of the code is

$$R = \frac{L \log M}{n}. \tag{1.20}$$

The reconstructed approximation of the source $\mathbf{S}$ in the decoder is $\hat{\mathbf{S}} = \mathbf{A}\hat{\beta}_{opt}$. Papers [24] and [25] show that with this optimal encoding scheme, the excess-distortion probability which is given by $\mathbb{P}\left(\frac{1}{n}\|\mathbf{s} - \mathbf{A}\hat{\beta}_{opt}\|^2 > D\right)$ decays exponentially in the source blocklength n, with the *optimal* excess-distortion exponent for memoryless discrete and Gaussian sources. This is in contrast to the error exponent of optimal decoding of SPARCs over the AWGN channel, which is sub-optimal.

Since the optimal encoder is impractical due to the fact that the number of codewords grows exponentially with n, a sub-optimal feasible encoder for lossy compression was designed and suggested in [26]. This proposal is a variant of successive cancellation, in which the non-zero indices in every section of β are picked sequentially: the encoder initializes a residual vector to $\mathbf{R}_0 = \mathbf{S}$, and chooses the non-zero index m_i in each section $i \in \{1 \ldots, L\}$ of $\mathbf{A}$, as the one which maximizes the inner product between the columns of $\mathbf{A}$ and the normalized residual $\frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|}$. The residual is updated by

$$\mathbf{R}_i = \mathbf{R}_{i-1} - c_i \mathbf{A}_{m_i}, \tag{1.21}$$

where

$$c_i = \sqrt{2\sigma^2 \log M \left(1 - \frac{2R}{L}\right)^{i-1}}, \tag{1.22}$$

and is also the non-zero coefficients of β. At the end of the run, the encoded codeword is $\mathbf{A}\beta$. A more detailed description of the encoding algorithm is provided in Chapter 3.

In [26], it is proved that the probability of excess distortion for this feasible encoder exponentially tends to 0 in n. While the results in [26] show that the rate-distortion function is asymptotically achievable for a memoryless Gaussian source, there are no explicit implications for the case where the blocklength of the source is finite; furthermore, even though the bound on the excess distortion decays exponentially to 0, it is rather loose for practical values of n.

1.2 Structure of the Thesis

We outline in the following the structure of the thesis.

- Chapter 2 provides preliminaries, notation, and new related results on Lambert's W function which are relevant to our analysis.

- Chapter 3 proposes a new version of lossy compression with SPARCs, improving the performance of the algorithm in [26], especially at high code rates. This is done by using better approximations than those used in [26].

- Chapter 4 provides an asymptotic analysis of the encoding algorithm, demonstrating that the distortion-rate function of a Gaussian memoryless source is achievable.

- In Chapter 5, an adaptation of the main theorem from [26] is derived for lossy compression of memoryless Gaussian sources with SPARCs of finite blocklength. The new theorem is derived in view of the modifications which follow from the new compression algorithm, and by improving the upper bound on the probability of excess distortion from [26]. It is further shown that the new theorem is applicable for memoryless non-Gaussian sources as well, and that SPARCs can successfully compress any memoryless source with a finite second moment up to the compression rate which corresponds to the rate-distortion function of a Gaussian memoryless source with the same variance.

- In Chapter 6, we find an asymptotic upper bound on the gap to the distortion-rate function which follows from the main result in Theorem 1 for the memoryless Gaussian case. The result scales similarly to [26], but it has a better (smaller) coefficient.

- Chapter 7 contains computer simulations of the new algorithm for lossy compression with SPARCs: first, we compare the performance of our improved algorithm to its original version in [26]; then, we examine the quality of our bound on the probability of excess distortion from Theorem 1 by comparing the bound with computer simulations.

- Chapter 8 summarizes briefly this thesis, and provides some open questions.

Chapter 2

Preliminaries, and New Related Results for the Lambert W Function

Chapter 2 is organized as follows: the first section is dedicated to the basic notation and basic results that are used throughout this thesis. The second section introduces the Lambert W function, and known approximations and bounds for it are detailed. Then, new tighter upper and lower bounds on the Lambert W function are derived in the third section, and they compared numerically to the previously known bounds. In the last section of the chapter, a closed-form approximation for the expected value of the maximum of M i.i.d. standard Gaussian random variables is presented, which is used multiple times in the course of this thesis.

2.1 Preliminaries and Notation

Throughout this document, unless stated otherwise, logarithms are on the natural base.

Notation 1. Let $\mathbf{r}, \mathbf{s} \in \mathbb{R}^n$. The norm, scaled norm, and inner product are given, respectively, by

$$\|\mathbf{r}\| = \left(\sum_{i=1}^{n} r_i^2\right)^{\frac{1}{2}}, \tag{2.1}$$

$$|\mathbf{r}| = \frac{\|\mathbf{r}\|}{\sqrt{n}}, \tag{2.2}$$

$$\langle \mathbf{r}, \mathbf{s} \rangle = \sum_{i=1}^{n} r_i s_i. \tag{2.3}$$

Notation 2. Let Φ be the cumulative distribution function of a standard Gaussian random variable, ϕ the corresponding probability density function, and Q the complementary Gaussian cumulative distribution function, i.e.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,\mathrm{d}t, \tag{2.4}$$

$$\phi(z) = \Phi'(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}, \tag{2.5}$$

$$Q(z) = 1 - \Phi(z) = 1 - Q(-z), \tag{2.6}$$

for all $z \in \mathbb{R}$.

Notation 3. Let $X_1, \dots, X_M$ be i.i.d. standard Gaussian random variables. Let their maximum and its expected value be denoted by

$$Z_M = \max_{i \in \{1,\dots,M\}} X_i, \tag{2.7}$$

$$e_M = \mathbb{E}\left[Z_M\right]. \tag{2.8}$$

The cumulative distribution function of Z_M satisfies

$$\begin{aligned} F_{Z_M}(z) &= \mathbb{P}\left(\max_{i \in \{1,\dots,M\}} X_i \le z\right) && (2.9)\\ &= \mathbb{P}(X_1 \le z)\,\mathbb{P}(X_2 \le z) \cdots \mathbb{P}(X_M \le z) && (2.10)\\ &= \Phi^M(z), && (2.11) \end{aligned}$$

for all $z \in \mathbb{R}$. Consequently, the probability density function of Z_M is given by

$$\begin{aligned} f_{Z_M}(z) = F'_{Z_M}(z) &= M\,\Phi^{M-1}(z)\phi(z) && (2.12)\\ &= M\,Q^{M-1}(-z)\phi(z), && (2.13) \end{aligned}$$

for all $z \in \mathbb{R}$; and thus, the expectation e_M in (2.8) satisfies

$$e_M = \int_{-\infty}^{\infty} z\, f_{Z_M}(z)\,\mathrm{d}z. \tag{2.14}$$

Lemma 1. *Let $X_1, \dots, X_M$ be i.i.d. standard Gaussian random variables, then*

$$e_M \le \sqrt{2\log M}. \tag{2.15}$$

Proof. By invoking Jensen's inequality, it follows from (2.7) and (2.8) that for all $t \in \mathbb{R}$,

$$\exp(t\,e_M) \le \mathbb{E}\left[\exp\left(t\,Z_M\right)\right] \le \sum_{i=1}^{M} \mathbb{E}\left[\exp\left(t\,X_i\right)\right] = M\exp\left(\tfrac{1}{2}\,t^2\right), \tag{2.16}$$

which implies that for all $t > 0$,

$$e_M \leq \frac{\log M}{t} + \frac{t}{2}. \tag{2.17}$$

Minimization of the right side of (2.17) over $t > 0$ yields the optimized value $t = \sqrt{2 \log M}$, leading to the required result. $\square$

Remark. The proof appears in [27, Eq. (A.3)]. This result can be specialized from the maximal inequality in [28, Lemma 2.3] by letting $T = \{1, \ldots, M\}$, $A = \mathbb{R}$ and $\psi(t) = t^2$ for all $t \in \mathbb{R}$.

2.2 The Lambert W Function

The Lambert W function is a set of functions, namely the branches of the inverse relation of the function $f(z) = ze^z$. Hence, the function $W(\cdot)$ satisfies the identity

$$z = W(z)\, e^{W(z)}, \tag{2.18}$$

for all $z \in \mathbb{C}$. Since the function f is not injective (i.e., it is one-to-one), the relation W is multi-valued (except at zero). If the attention is restricted to real-valued W, then the complex variable z is replaced by the real variable x, and the relation is defined only for $x \geq -\frac{1}{e}$, and it is double valued on $(-\frac{1}{e}, 0)$ (see Figure 2.1).

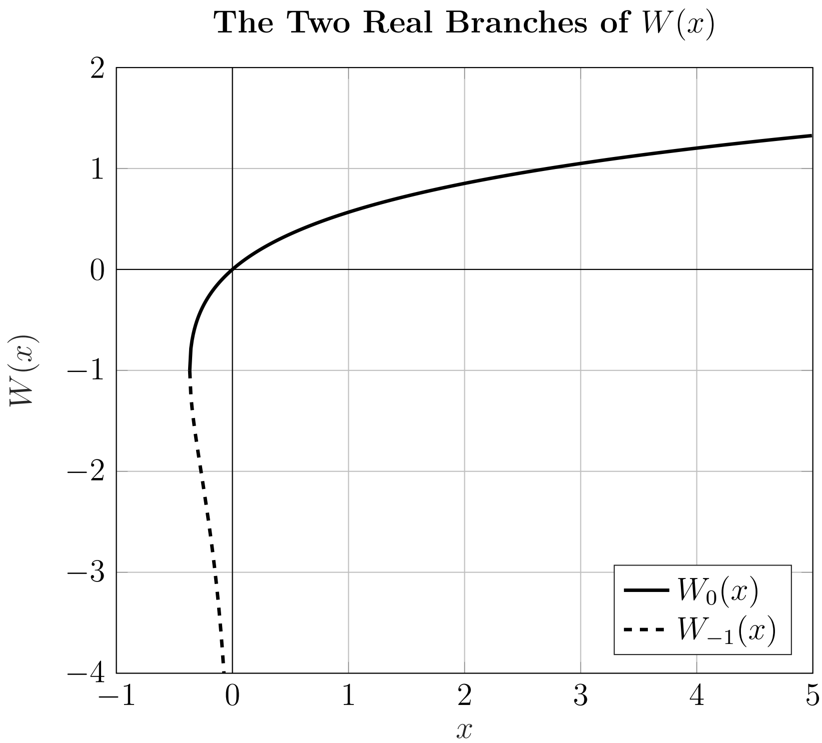

Figure 2.1: The two real branches of the Lambert W function. The solid line is $W_0(x)$, the principal branch; the dashed line is $W_{-1}(x)$.

The additional constraint $W \geq -1$ defines a single-valued function, denoted by $W_0(x)$

where $W_0(0) = 0$ and $W_0(-\frac{1}{e}) = -1$, which refers to the principal branch of the Lambert W function (the solid line in Figure 2.1). The secondary (lower) branch (see the dashed line in Figure 2.1) satisfies $W \geq -1$, and it is denoted by $W_{-1}(x)$, decreasing from $W_{-1}(-\frac{1}{e}) = -1$ to $W_{-1}(0^-) = -\infty$. Unless the branch is not explicitly stated, it refers to the principal branch W_0.

Among its uses, the Lambert W function can be employed to solve exponential equations. For example, the equation $xb^x = a$ has the solution

$$x = \frac{W(a \log b)}{\log b}. \tag{2.19}$$

In [12], many other practical applications of Lambert-W function are detailed, which show how important it is to have good approximations and bounds on this function.

We begin by presenting some known results. Two useful identities that are derived directly from (2.18) are

$$e^{W(x)} = \frac{x}{W(x)}, \quad \text{for } x \neq 0, \tag{2.20}$$

$$\log W(x) = \log x - W(x), \quad \text{for } x > 0. \tag{2.21}$$

The following asymptotic expansion to $W(x)$ was developed in [12] for large values of x,

$$W(x) = \log x - \log\log x + \sum_{k=0}^{\infty}\sum_{m=1}^{\infty} c_{km} (\log\log x)^m (\log x)^{-k-m}, \tag{2.22}$$

with $c_{km} = \frac{(-1)^k}{m!}\left[{k+m \atop k+1}\right]$, and where $\left[{k+m \atop k+1}\right]$ is a Stirling cycle number of the first kind, defined by

$$x(x+1)\dots(x+n-1) := \sum_{k=0}^{n} \left[{n \atop k}\right] x^k. \tag{2.23}$$

This expansion is absolutely convergent, and can be expressed (after rearranging terms) as

$$W(x) = L_1 - L_2 + \frac{L_2}{L_1} + \frac{L_2(-2+L_2)}{2L_1^2} + \frac{L_2(6 - 9L_2 + 2L_2^2)}{6L_1^3} + O\left(\frac{L_2^4}{L_1^4}\right), \tag{2.24}$$

where $L_1 := \log x$ and $L_2 := \log\log x$. In [30], the following upper and lower bounds were proved:

$$\log x - \log\log x + \frac{1}{2}\frac{\log\log x}{\log x} \leq W(x) \leq \log x - \log\log x + \frac{e}{e-1}\frac{\log\log x}{\log x}, \tag{2.25}$$

for every $x \geq e$.

2.3 New Bounds on the Lambert W Function

The next lemma provides new bounds to the Lambert W function, which are tighter than the previously known ones.

Lemma 2. *For $x \geq e$,*

$$s(x) \leq e^{W(x)} \leq t(x), \tag{2.26}$$

with

$$s(x) \triangleq \frac{x}{\log t(x)}, \tag{2.27}$$

$$t(x) \triangleq \frac{x}{\log x - \log\log\left(\frac{x}{v(x)}\right)}, \tag{2.28}$$

$$v(x) \triangleq \log x - \log\left(\log\left(\frac{x}{\log x}\right) - \log\left(1 - \frac{\log\log x}{1+\log x}\right)\right). \tag{2.29}$$

Furthermore, both the upper and lower bounds coincide if and only if $x = e$.

Proof. From [30, Theorem 2.5], for all $x > 1$,

$$W(x) \geq \left(\frac{\log x}{1+\log x}\right)(\log x - \log\log x + 1) \tag{2.30}$$

$$= \log x - \frac{\log x}{1+\log x} \cdot \log\log x, \tag{2.31}$$

with equality if and only if $x = e$. Denote the right side of (2.31) by $f(x)$. The function f is monotonically increasing on $[e, \infty)$ since both $\frac{\log x}{1+\log x}$ and $(\log x - \log\log x + 1)$ are nonnegative increasing functions in this domain. Since $f(e) = 1$, we have $f(x) \geq 1$ for all $x \geq e$. Thus, taking logarithms on both sides of (2.31) gives

$$\log W(x) \geq \log\left(\log x - \frac{\log x}{1+\log x} \cdot \log\log x\right) \tag{2.32}$$

$$= \log\log x + \log\left(1 - \frac{\log\log x}{1+\log x}\right). \tag{2.33}$$

From (2.21),

$$W(x) = \log x - \log W(x) \tag{2.34}$$

$$\leq \log x - \log\log x - \log\left(1 - \frac{\log\log x}{1+\log x}\right) \tag{2.35}$$

$$= \log\left(\frac{x}{\log x}\right) - \log\left(1 - \frac{\log\log x}{1+\log x}\right), \tag{2.36}$$

where (2.35) follows from (2.33). Hence,

$$\log W(x) \leq \log \left(\log \left(\frac{x}{\log x} \right) - \log \left(1 - \frac{\log \log x}{1 + \log x} \right) \right), \tag{2.37}$$

and by invoking (2.21) once again, it follows that

$$\begin{aligned} W(x) &= \log x - \log W(x) && (2.38) \\ &\geq \log x - \log \left(\log \left(\frac{x}{\log x} \right) - \log \left(1 - \frac{\log \log x}{1 + \log x} \right) \right) && (2.39) \\ &= v(x), && (2.40) \end{aligned}$$

with (2.40) is due to the definition of v in (2.29). The function v is positive on $[e, \infty)$, since for $x \geq e$

$$\begin{aligned} v(x) &\geq \log x - \log \left(\log x - \log \left(\frac{1 + \log x - \log \log x}{1 + \log x} \right) \right) && (2.41) \\ &\geq \log x - \log \left(\log x + \log \left(1 + \log x \right) \right) && (2.42) \\ &\geq \log x - \log \left(2 \log x + \log \left(\frac{2}{e} \right) \right) && (2.43) \\ &= \log \left(\frac{x}{2 \log x + \log \left(\frac{2}{e} \right)} \right) && (2.44) \\ &\geq \tfrac{3}{2} - \tfrac{3}{2} \log 2 && (2.45) \\ &\approx 0.4603, && (2.46) \end{aligned}$$

where (2.41) holds since $\log \log x \geq 0$ for $x \geq e$; (2.42) holds since $\log x \geq \log \log x$ for $x \geq e$; (2.43) holds since $\frac{2x}{e} \geq 1 + \log x$ for $x \geq e$; (2.45) is obtained by minimization of the right side of (2.44) for $x \geq e$ where the minimal value is attained at $x = \frac{1}{\sqrt{2}} \, e^{3/2}$. Therefore, for all $x \geq e$,

$$\log W(x) \geq \log v(x), \tag{2.47}$$

and then we continue in a similar fashion where from (2.21)

$$\begin{aligned} W(x) &= \log x - \log W(x) && (2.48) \\ &\leq \log \frac{x}{v(x)}, && (2.49) \end{aligned}$$

where (2.49) follows from (2.47). Again using (2.21),

$$\begin{aligned} W(x) &= \log x - \log W(x) && (2.50) \\ &\geq \log x - \log \log \frac{x}{v(x)} && (2.51) \\ &= \frac{x}{t(x)}, && (2.52) \end{aligned}$$

with $t(x)$ as defined in (2.28). Since

$$\frac{x}{t(x)} = \log x - \log\log\left(\frac{x}{h(x)}\right) \tag{2.53}$$

$$\geq \log x - \log\log\left(\frac{x}{0.46}\right) \tag{2.54}$$

$$\geq \log(0.46e) \geq 0.22, \tag{2.55}$$

where (2.54) is due to (2.45), and (2.55) is found by minimization. Thus,

$$\log W(x) \geq \log\left(\frac{x}{t(x)}\right), \tag{2.56}$$

and using (2.21) one last time,

$$W(x) = \log x - \log W(x) \tag{2.57}$$

$$\leq \log t(x). \tag{2.58}$$

From (2.52) and (2.58) and the identity in (2.20), it follows that for all $x \geq e$,

$$\frac{x}{\log t(x)} \leq e^{W(x)} \leq t(x). \tag{2.59}$$

$\square$

Corollary 1. *For large x,*

$$W(x) \approx \log x - \log\log x + \frac{\log\log x}{\log x}. \tag{2.60}$$

Proof. From Lemma 2 and (2.20),

$$\frac{x}{t(x)} \leq W(x) \leq \log t(x). \tag{2.61}$$

For large x,

$$t(x) \approx \frac{x}{\log x - \log\log(\frac{x}{\log x})}, \tag{2.62}$$

and therefore,

$$\frac{x}{t(x)} \approx \log x - \log\log\left(\frac{x}{\log x}\right) \tag{2.63}$$

$$= \log x - \log(\log x - \log\log x) \tag{2.64}$$

$$= \log x - \log\left(\log x\left(1 - \frac{\log\log x}{\log x}\right)\right) \tag{2.65}$$

$$\approx \log x - \log\log x + \frac{\log\log x}{\log x}, \tag{2.66}$$

where in (2.66), the first-order Taylor approximation $\log(1-x) \approx -x$ was used. Consequently,

$$\log t(x) \approx \log\left(\frac{x}{\log x - \log\log\left(\frac{x}{\log x}\right)}\right) \tag{2.67}$$

$$= \log x - \log\left(\log x - \log\log\left(\frac{x}{\log x}\right)\right) \tag{2.68}$$

$$\approx \log x - \log\left(\log x\left(1 - \frac{\log\log x}{\log x}\right)\right) \tag{2.69}$$

$$\approx \log x - \log\log x + \frac{\log\log x}{\log x}. \tag{2.70}$$

$\square$

Comparison to existing bounds: From the previously known bound in (2.25), we have

$$\frac{x}{\log x}\exp\left(\frac{1}{2}\frac{\log\log x}{\log x}\right) \leq e^{W(x)} \leq \frac{x}{\log x}\exp\left(\frac{e}{e-1}\frac{\log\log x}{\log x}\right). \tag{2.71}$$

Another bound can be derived by using the inequality in (2.25) once more, this time along with the identity in (2.20),

$$\frac{x}{\log x - \log\log x + \frac{e}{e-1}\frac{\log\log x}{\log x}} \leq e^{W(x)} \leq \frac{x}{\log x - \log\log x + \frac{1}{2}\frac{\log\log x}{\log x}}. \tag{2.72}$$

Figure 2.2 shows the ratio between each of the bounds in (2.26), (2.71) and (2.72) and the exact value of $e^{W(x)}$. The new bound, in (2.26), is a significant improvement to the those in (2.71) and (2.72), and is almost identical to $e^{W(x)}$.

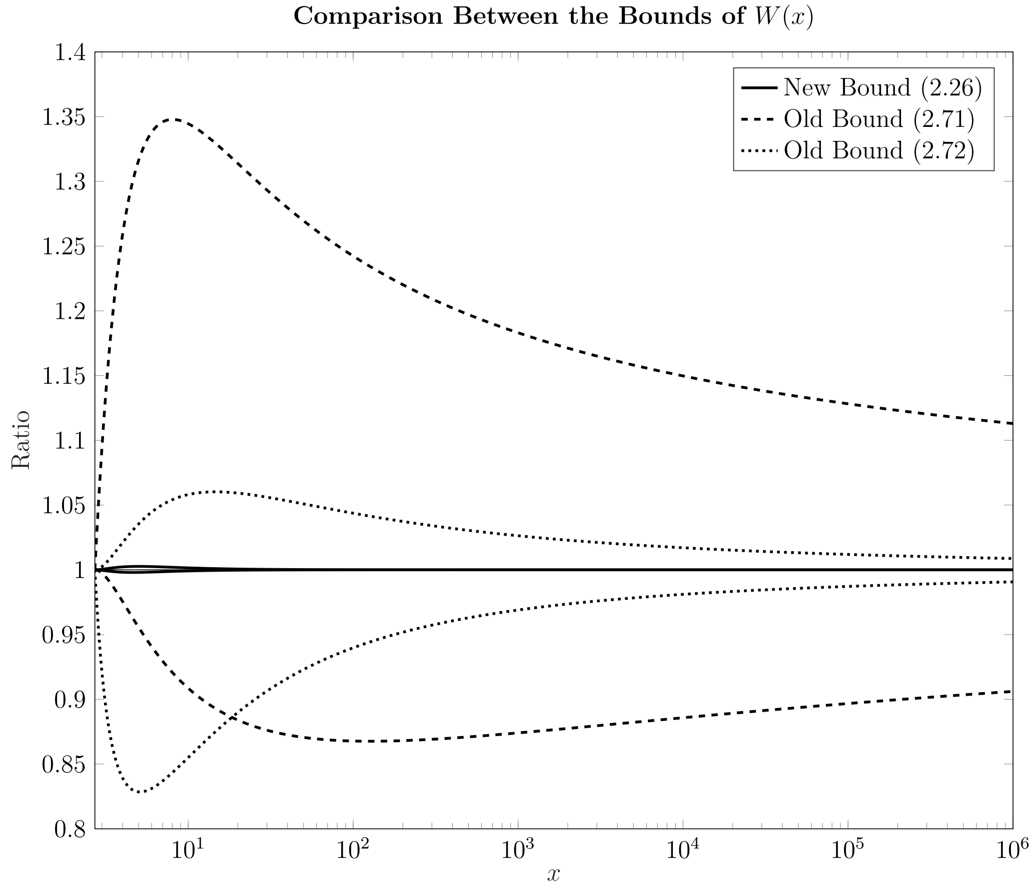

Figure 2.2: Comparison of the upper and lower bounds on the Lambert W function: the solid line is the ratio between the new upper and lower bounds in (2.26) and the exact value of $e^{W(x)}$, the dashed line is the ratio between the bounds in (2.71) and the exact value of $e^{W(x)}$, and the dotted line is the ratio between the bounds in (2.72) and the exact value of $e^{W(x)}$. The new upper and lower bounds are much tighter than the previously reported bounds, and they are almost equal to the exact value of $e^{W(x)}$.

2.4 A Closed-Form Approximation for e_M

The following lemma gives a closed-form asymptotic approximation to e_M as defined in (2.8). It relies on [29], which describes a general method to find the expected value of extreme values of M continuously distributed random variables. As an example, the method is applied to the Gaussian distribution.

Lemma 3. *For large M, the following asymptotic result holds:*

$$e_M = \sqrt{2\log M} - \frac{\log\log M + \log 4\pi - 2\gamma}{2\sqrt{2\log M}} + O\left(\frac{\log\log M}{\log^{1.5} M}\right), \tag{2.73}$$

where γ is the Euler-Mascheroni constant,

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log n\right) \approx 0.577216\,. \tag{2.74}$$

Proof. Let ξ_M be the following random variable,

$$\xi_M \triangleq M(1 - \Phi(Z_M)), \tag{2.75}$$

with Φ and Z_M as defined in (2.4) and (2.7), respectively. The cumulative distribution function of ξ is given by

$$\begin{aligned} F_{\xi_M}(x) &= \mathbb{P}(\xi_M \leq x) && (2.76)\\ &= \mathbb{P}\Big(1 - \frac{x}{M} \leq \Phi(Z_M)\Big) && (2.77)\\ &= \int_{\Phi^{-1}\left(1-\frac{x}{M}\right)}^{\infty} M\Phi^{M-1}(t)\,\phi(t)\,\mathrm{d}t, && (2.78)\end{aligned}$$

where (2.78) follows from (2.12). By applying Leibniz's integral rule, the probability density function is given by

$$\begin{aligned} f_{\xi_M}(x) &= F'_{\xi_M}(x) && (2.79)\\ &= -M\ \Phi^{M-1}\Big(\Phi^{-1}\left(1-\frac{x}{M}\right)\Big)\ \phi\Big(\Phi^{-1}\left(1-\frac{x}{M}\right)\Big)\cdot\frac{\mathrm{d}}{\mathrm{d}x}\Big\{\Phi^{-1}\Big(1-\frac{x}{M}\Big)\Big\} && (2.80)\\ &= -M\left(1-\frac{x}{M}\right)^{M-1}\ \phi\Big(\Phi^{-1}\left(1-\frac{x}{M}\right)\Big)\cdot\left(-\frac{1}{M}\,\frac{1}{\phi\left(\Phi^{-1}\left(1-\frac{x}{M}\right)\right)}\right) && (2.81)\\ &= \left(1-\frac{x}{M}\right)^{M-1}, && (2.82)\end{aligned}$$

for all $x \in [0, M]$. By letting $M \to \infty$, the function $f_{\xi_M}(x)$ converges to

$$\lim_{M\to\infty} f_{\xi_M}(x) = e^{-x}, \quad x \geq 0. \tag{2.83}$$

Since $f_{\xi_M}(x)$ is uniformly bounded for all M on every finite interval of x, the distribution of ξ_M as $M \to \infty$ is equal to the exponential probability density function in (2.83). We continue by expressing Z_M as function of ξ_M. From its definition in (2.75),

$$\xi_M = M\int_{Z_M}^{\infty}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{t^2}{2}}\,\mathrm{d}t. \tag{2.84}$$

Since

$$\mathbb{P}(Z_M \leq 0) = \Phi^M(0) = 2^{-M}, \tag{2.85}$$

it follows that, as $M \to \infty$, the probability that $Z_M > 0$ tends to 1 exponentially in M.

Integrating (2.84) by parts, we have

$$\frac{\sqrt{2\pi}\,\xi_M}{M} = \int_{Z_M}^{\infty} \left(-\frac{1}{t}\right) \cdot \left(-te^{-\frac{t^2}{2}}\right) \mathrm{d}t \tag{2.86}$$

$$= \frac{1}{Z_M}\, e^{-\frac{Z_M^2}{2}} - \int_{Z_M}^{\infty} \frac{1}{t^2}\, e^{-\frac{t^2}{2}}\, \mathrm{d}t. \tag{2.87}$$

The integral in the right side of (2.87) can be upper bounded as follows: let

$$g(x) \triangleq \frac{1}{x^3}\, e^{-\frac{x^2}{2}} - \int_{x}^{\infty} \frac{1}{t^2}\, e^{-\frac{t^2}{2}}\, \mathrm{d}t \tag{2.88}$$

for $x > 0$. Then,

$$g'(x) = -\frac{3}{x^4}\, e^{-\frac{x^2}{2}} - \frac{1}{x^2}\, e^{-\frac{x^2}{2}} + \frac{1}{x^2}\, e^{-\frac{x^2}{2}} \tag{2.89}$$

$$= -\frac{3}{x^4}\, e^{-\frac{x^2}{2}} < 0, \tag{2.90}$$

and

$$\lim_{x \to \infty} g(x) = 0. \tag{2.91}$$

Since the function $g(\cdot)$ is monotonically decreasing on $(0, \infty)$, and it tends to 0 as we let $x \to \infty$, it is nonnegative on $(0, \infty)$, i.e.

$$\int_{x}^{\infty} \frac{1}{t^2} e^{-\frac{t^2}{2}} \leq \frac{1}{x^3} e^{-\frac{x^2}{2}}, \quad x > 0. \tag{2.92}$$

Hence, (2.87) can be rewritten as

$$\frac{\sqrt{2\pi}\,\xi_M}{M} = \frac{1}{Z_M} e^{-\frac{Z_M^2}{2}} \left(1 + O\left(\frac{1}{Z_M^2}\right)\right), \tag{2.93}$$

$$\Rightarrow \frac{M^2}{2\pi \xi_M^2} = Z_M^2 e^{Z_M^2} \left(1 + O\left(\frac{1}{Z_M^2}\right)\right). \tag{2.94}$$

The random variable ξ_M is bounded asymptotically in probability as follows:

$$\liminf_{M \to \infty}\ \mathbb{P}\left(\frac{1}{\log M} \leq \xi_M \leq \log M\right) = \liminf_{M \to \infty} \int_{\frac{1}{\log M}}^{\log M} \left(1 - \frac{x}{M}\right)^{M-1} \mathrm{d}x \tag{2.95}$$

$$= \liminf_{M \to \infty} \int_{1-\frac{\log M}{M}}^{1-\frac{1}{M \log M}} M y^{M-1} \mathrm{d}y \tag{2.96}$$

$$= \liminf_{M \to \infty} \left(1 - \frac{1}{M \log M}\right)^{M} - \left(1 - \frac{\log M}{M}\right)^{M}, \tag{2.97}$$

where (2.95) follows from ξ_M's probability density function in (2.82), and (2.96) comes from substituting $y = 1 - \frac{x}{M}$. From the inequality $\log(1-x) \leq -x$ for $x \in [0,1)$,

$$\limsup_{M\to\infty} \left(1 - \frac{\log M}{M}\right)^M = \limsup_{M\to\infty} \exp\left(M \log\left(1 - \frac{\log M}{M}\right)\right) \tag{2.98}$$

$$\leq \limsup_{M\to\infty} \frac{1}{M} \tag{2.99}$$

$$= 0. \tag{2.100}$$

Similarly, from the inequality $\log(1-x) \geq -x - x^2$ for $x \in [0, \frac{1}{2}]$,

$$\liminf_{M\to\infty} \left(1 - \frac{1}{M \log M}\right)^M = \liminf_{M\to\infty} \exp\left(M \log\left(1 - \frac{1}{M \log M}\right)\right) \tag{2.101}$$

$$\geq \liminf_{M\to\infty} \exp\left(-\frac{1}{\log M} - \frac{1}{M \log^2 M}\right) \tag{2.102}$$

$$= 1. \tag{2.103}$$

Consequently, as $M \to \infty$, ξ_M is upper bounded by $\log M$ and lower bounded by $\frac{1}{\log M}$ with probability that tends to 1, and henceforth we assume that

$$\frac{1}{\log M} \leq \xi_M \leq \log M. \tag{2.104}$$

In view of (2.94), we next use the Lambert W function. Using the asymptotic approximation in Corollary 1,

$$W(x) = \log x - \log\log x + O\left(\frac{\log\log x}{\log x}\right), \tag{2.105}$$

we have

$$\log(M^2) - \log(2\pi\xi_M^2) - \log\log\left(\frac{M^2}{2\pi\xi_M^2}\right) + O\left(\frac{\log\log M}{\log M}\right) = Z_M^2 + O\left(\frac{1}{Z_M^2}\right). \tag{2.106}$$

Therefore,

$$Z_M = \sqrt{2\log M} \cdot \sqrt{1 - \frac{\log 2\pi + 2\log\xi_M + \log\big(2\log M - \log(2\pi\xi_M^2)\big) + O\left(\frac{\log\log M}{\log M}\right)}{2\log M}}, \tag{2.107}$$

and by using another Taylor approximation where $(1-x)^{1/2} \approx 1 - \frac{1}{2}x$ for values of x which are close to zero, it follows that

$$Z_M = \sqrt{2\log M}\left(1 - \frac{\log 2\pi + 2\log\xi_M + \log\bigl(2\log M - \log(2\pi\xi_M^2)\bigr) + O\left(\frac{\log\log M}{\log M}\right)}{4\log M}\right) \tag{2.108}$$

$$= \sqrt{2\log M} - \frac{\log\bigl(2\log M - \log(2\pi\xi_M^2)\bigr) + \log 2\pi}{2\sqrt{2\log M}} - \frac{\log\xi_M}{\sqrt{2\log M}} + O\left(\frac{\log\log M}{\log^{1.5} M}\right). \tag{2.109}$$

Finally, we use a Taylor approximation for $\log(x)$,

$$Z_M = \sqrt{2\log M} - \frac{\log\log M + \log 4\pi}{2\sqrt{2\log M}} - \frac{\log\xi_M}{\sqrt{2\log M}} + O\left(\frac{\log\log M}{\log^{1.5} M}\right). \tag{2.110}$$

Therefore, in view of (2.83), the expected value of Z_M is approximately equal to

$$e_M = \sqrt{2\log M} - \frac{\log\log M + \log 4\pi - 2\gamma}{2\sqrt{2\log M}} + O\left(\frac{\log\log M}{\log^{1.5} M}\right), \tag{2.111}$$

where we rely on the following identity:

$$\gamma = -\int_0^\infty e^{-t}\log t \;\mathrm{d}t. \tag{2.112}$$

□

Two important results stem from Lemma 3. First, a simpler (albeit less accurate) asymptotic approximation of e_M can be acquired,

$$\lim_{M\to\infty} \frac{e_M}{\sqrt{2\log M}} = 1; \tag{2.113}$$

second, we can deduce that for a large enough M,

$$\sqrt{2\log M} - \frac{\log\log M + \log 4\pi}{2\sqrt{2\log M}} \le e_M \le \sqrt{2\log M} - \frac{\log\log M - 2\gamma}{2\sqrt{2\log M}}. \tag{2.114}$$

Remark. In (2.106), the approximation from Corollary 1 was applied rather than the tight bounds of Lemma 2, even though a close approximation of e_M for finite M could be extracted by using it. However, using Lemma 2 is unnecessary for our purpose, since there are approximations of e_M which are sufficiently tight even for small values of M. Figure 2.3 shows a comparison between e_M as calculated by its expression in (2.14), and several approximations:

1. Upper bound: $\sqrt{2\log M}$ (2.115)
2. Approx. 1: $\sqrt{\log\left(\frac{M^2}{2\pi\log\left(\frac{M^2}{2\pi}\right)}\right)}$ (2.116)

3. Approx. 2: $\sqrt{\log\left(\frac{M^2}{2\pi\log\left(\frac{M^2}{2\pi}\right)}\right)}\left(\frac{\gamma}{2\log M}+1\right)$ (2.117)

4. Approx. 3: $(1-\gamma)Q^{-1}\left(\frac{1}{M}\right)+\gamma Q^{-1}\left(\frac{1}{Me}\right)$ (2.118)

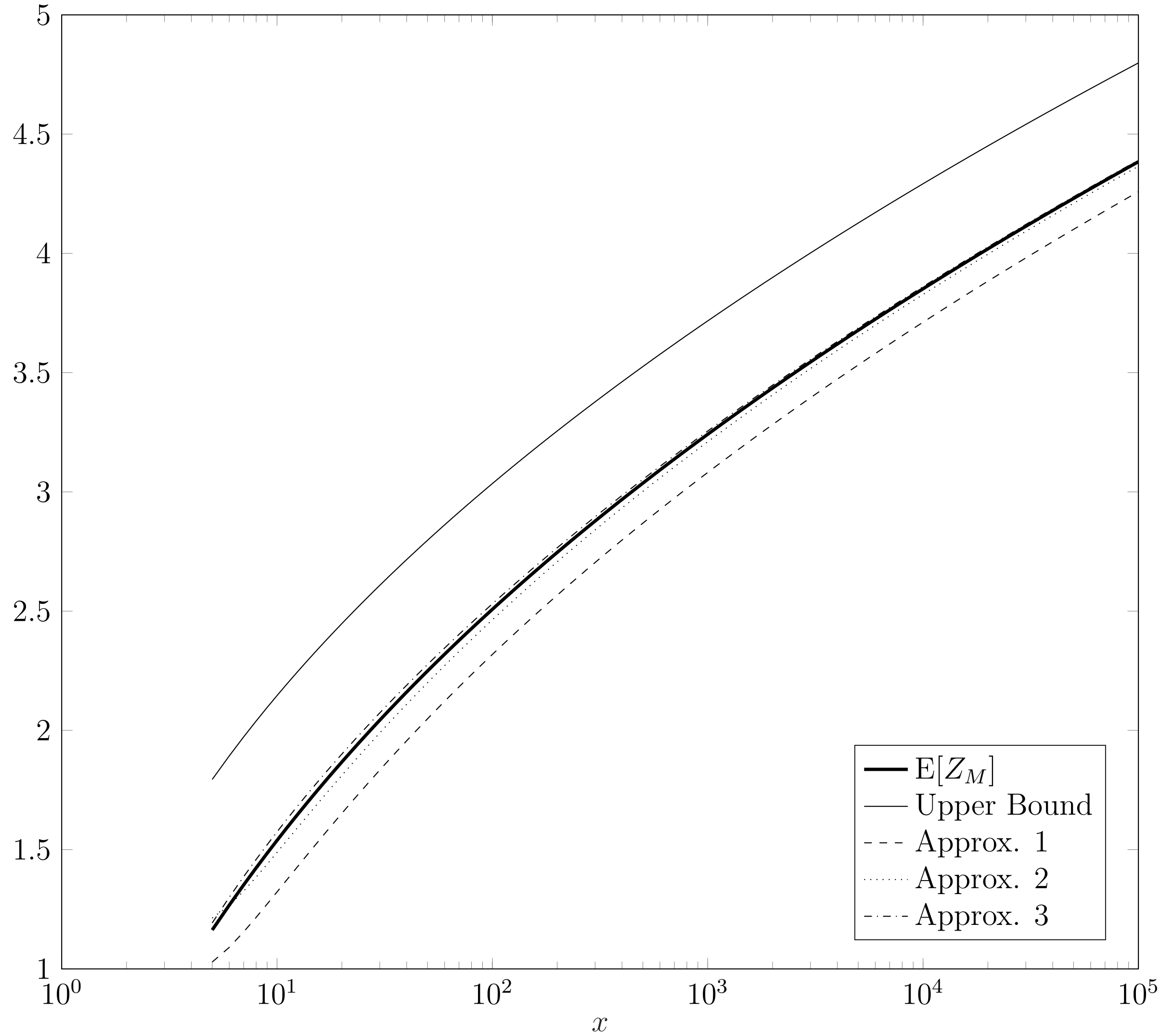

Figure 2.3: A comparison between e_M and its approximations. The bold line is an exact calculation of e_M. The solid line is the upper bound (2.115), The dashed line is approximation 1 (2.116), The dotted line is approximation 2 (2.117), The dash-dot line is approximation 3 (2.118).

Chapter 3

Lossy Compression with the Sparse Regression Codebook

The following lossy compression algorithm is the one which was described in [26]. We then propose a modification to the encoder which improves the performance of the compression, and enables reaching a tighter upper bound on the probability of excess distortion.

3.1 Original Encoding Algorithm

Let $\mathbf{S}$ be an ergodic source, emitting a sequence of length n whose symbols have zero mean and variance σ^2. Let $\mathbf{A}$ be an $n \times N$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries independent of $\mathbf{S}$, and let $N = ML$ with $M > L$ and $M, L \in \mathbb{N}$. Let b satisfy $M = L^b$.

Conceptually, the columns of the matrix $\mathbf{A}$ are divided into L sections with M columns each; a codeword is generated by a linear combination of L columns, one from every section. The linear combination can be thought of as multiplying $\mathbf{A}$ with a sparse vector β of length N, consisting of a single non-zero entry in M consecutive components.

Thus, if the rate R of the code is expressed in nats per source symbol, the number of codewords is $M^L = \exp(nR)$, and therefore

$$nR = L \log M = bL \log L. \tag{3.1}$$

The non-zero entries of β are marked by $\{c_i\}_{i=1}^{L}$ and are given by

$$c_i = \sigma \sqrt{\frac{2R}{L}\left(1 - \frac{2R}{L}\right)^{i-1}}. \tag{3.2}$$

The Algorithm. The codeword $\mathbf{A}\hat{\beta}$, that corresponds to a source sequence $\mathbf{S}$ of length n, is determined in the following way:

1. Set $\mathbf{R}_0 = \mathbf{S}$. (3.3)

2. For $i \in \{1, \dots, L\}$, choose

$$m_i \triangleq \underset{j:(i-1)M<j\leq iM}{\arg\max} \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle, \tag{3.4}$$

where $\mathbf{A}_j$ is the j's column in the matrix $\mathbf{A}$, and define recursively

$$\mathbf{R}_i = \mathbf{R}_{i-1} - c_i \mathbf{A}_{m_i}. \tag{3.5}$$

3. At the end of the L'th step, the sparse vector $\hat{\beta}$ of length N is comprised of the values of $\{c_i\}_{i=1}^L$ in indices $\{m_i\}_{i=1}^L$ respectively, and its entries are zero elsewhere. The codeword after the lossy compression is equal to $\mathbf{A}\hat{\beta}$.

3.2 Modified Encoding Algorithm

The modification in the encoding algorithm from the one described in Section 3.1, is a new set of coefficients $\{c_i\}_{i=1}^L$. Instead of using (3.2), we propose the following,

$$c_i = \frac{\sigma e_M}{\sqrt{n}} \sqrt{\left(1 - \frac{e_M^2}{n}\right)^{i-1}}, \qquad i \in \{1, \dots, L\}, \tag{3.6}$$

with e_M as defined in (2.8). In Chapter 4, the reason for the change is explained more thoroughly, but in essence the improvement is due to the fact that e_M is a more accurate approximation of the maximum of M i.i.d. standard Gaussian random variables than $\sqrt{2\log M}$. Indeed, by replacing the asymptotic approximation $e_M \approx \sqrt{2\log M}$ in (3.6), we have

$$\begin{aligned} c_i &\approx \sigma \sqrt{\frac{2\log M}{n}} \sqrt{\left(1 - \frac{2\log M}{n}\right)^{i-1}} &(3.7)\\ &= \sigma \sqrt{\frac{2R}{L}\left(1 - \frac{2R}{L}\right)^{i-1}}, &(3.8) \end{aligned}$$

which is identical to (3.2). The improvement that this change induces is more significant for higher values of the rate R and lower values of the source blocklength n, as can be seen in the computer simulations in Chapter 7, Figure 7.1.

Chapter 4

Preliminary Asymptotic Analysis

To be self-contained, we provide the following result:

Lemma 4. [26, Lemma 1]. *Let $\{\mathbf{A}_j\}_{j=1}^N$ be N mutually independent random vectors of length n, and suppose that the components of each vector are i.i.d. standard Gaussian random variables. Let $\mathbf{R}$ be a random vector independent of $\{\mathbf{A}_j\}_{j=1}^N$ whose support lies on the n-dimensional unit sphere, i.e.,*

$$\sum_{i=1}^{n} r_i^2 = 1, \tag{4.1}$$

and let

$$T_j = \langle \mathbf{A}_j, \mathbf{R} \rangle \tag{4.2}$$

for every $j \in \{1, \ldots, N\}$. Then, $\{T_j\}_{j=1}^N$ are i.i.d. standard Gaussian random variables which are independent of $\mathbf{R}$.

Proof. The joint probability density function of $\{T_j\}_{j=1}^N$ is given by

$$\begin{aligned} f_{T_1,T_2,\ldots,T_N}(t_1, t_2, \ldots, t_N) &= \int_{\mathbb{R}^n} f_{T_1,T_2,\ldots,T_N,\mathbf{R}}(t_1, t_2, \ldots, t_N, \mathbf{r}) \, \mathrm{d}\mathbf{r} && (4.3) \\ &= \int_{\mathbb{R}^n} f_{T_1,T_2,\ldots,T_N|\mathbf{R}}(t_1, t_2, \ldots, t_N|\mathbf{r}) \, f_{\mathbf{R}}(\mathbf{r}) \, \mathrm{d}\mathbf{r}, && (4.4) \end{aligned}$$

for all $\mathbf{t} = (t_1, \ldots, t_N) \in \mathbb{R}^N$. By assumption, for all $j \in \{1, \ldots, N\}$, $\mathbf{A}_j = (A_{1,j}, \ldots, A_{n,j})^\top$ is a column Gaussian random vector with i.i.d. components of zero mean and a unit variance. Since the vectors $\{\mathbf{A}_j\}_{j=1}^N$ are mutually independent and are also independent of $\mathbf{R}$, the random variables $\{T_j\}_{j=1}^N$ are mutually independent given $\mathbf{R}$. Furthermore, since a linear combination of jointly Gaussian random variables is Gaussian, the random variable T_j is Gaussian conditioned on $\mathbf{R}$, and

$$\mathbb{E}[T_j|\mathbf{R}=\mathbf{r}] = \mathbb{E}\left[\sum_{i=1}^{n} A_{i,j} r_i \middle| \mathbf{R}=\mathbf{r}\right] = \sum_{i=1}^{n} \mathbb{E}[A_{i,j}] r_i = 0, \tag{4.5}$$

$$\mathrm{Var}[T_j|\mathbf{R}=\mathbf{r}] = \mathrm{Var}\left[\sum_{i=1}^{n} A_{i,j} r_i \middle| \mathbf{R}=\mathbf{r}\right] = \sum_{i=1}^{n} \mathrm{Var}[A_{i,j}] r_i^2 = \sum_{i=1}^{n} r_i^2 = 1, \tag{4.6}$$

where (4.6) is due to (4.1), and since $\{A_{i,j}\}$ are independent and of unit variance. Therefore,

$$f_{T_1,T_2,\dots,T_N|\mathbf{R}}(\mathbf{t}|\mathbf{r}) = \prod_{i=1}^{N} \phi(t_i) \tag{4.7}$$

for all $\mathbf{r} \in \mathbb{R}^n$ such that $\|\mathbf{r}\| = 1$, with ϕ as given in (2.5). Substituting (4.7) into (4.4) gives

$$f_{T_1,T_2,\dots,T_N}(\mathbf{t}) = \int_{\mathbb{R}^n} \prod_{i=1}^{N} \phi(t_i)\, f_{\mathbf{R}}(\mathbf{r})\, \mathrm{d}\mathbf{r} = \prod_{i=1}^{N} \phi(t_i). \tag{4.8}$$

The random vector $(T_1, \dots, T_N)$ therefore has i.i.d. standard Gaussian components. □

Let $\mathbf{S}$, $\mathbf{A}$ and $\{\mathbf{R}_i\}_{i=1}^{L}$ be as defined by the encoding algorithm in Section 3.2. If L is large enough, so is n (see (3.1)), and by the ergodicity of $\mathbf{S}$ and the law of large numbers, we get from (2.2) that

$$|\mathbf{S}|^2 = \frac{1}{n}\sum_{i=1}^{n} S_i^2 \approx \mathbb{E}\left[S_1^2\right] = \sigma^2, \tag{4.9}$$

$$|\mathbf{A}_j|^2 = \frac{1}{n}\sum_{i=1}^{n} A_{i,j}^2 \approx \mathbb{E}\left[A_{1,j}^2\right] = 1, \quad \forall j \in \{1, \dots, N\}. \tag{4.10}$$

In addition, Lemma 4 implies that for all $j \in \{M(i-1)+1, \dots, Mi\}$, the inner products $\left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle$ are M i.i.d. standard Gaussian random variables, so for large M,

$$\left\langle \mathbf{A}_{m_i}, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle = \max_{(i-1)M+1 \le j \le iM} \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle \approx e_M \tag{4.11}$$

with e_M as defined in (2.8), and the approximation in the right side of (4.11) follows from the concentration of Z_M around its average e_M [27, (A.7)]. The goal is to find an approximation for the residue after L steps, $|R_L|^2$, and additionally, to show that our choice for $\{c_i\}_{i=1}^{L}$ in (3.6) achieves the minimal residue at each step of the recursion in Section 3.2. From (2.2) and (3.5),

$$|\mathbf{R}_i|^2 = |\mathbf{R}_{i-1}|^2 + c_i^2\, |\mathbf{A}_{m_i}|^2 - \frac{2c_i}{n} \langle \mathbf{A}_{m_i}, \mathbf{R}_{i-1} \rangle \tag{4.12}$$

$$\approx |\mathbf{R}_{i-1}|^2 \left(1 + \frac{c_i^2}{|\mathbf{R}_{i-1}|^2} - \frac{2c_i}{|\mathbf{R}_{i-1}|} \cdot \frac{e_M}{\sqrt{n}}\right), \tag{4.13}$$

where (4.13) holds due to the approximations in (4.10) and (4.11). Let $t_i \triangleq \frac{c_i}{|\mathbf{R}_{i-1}|}$ for all

$i \in \{1, \ldots, L\}$. In order to determine the value of t_i which minimizes the right side of (4.13) - thus approximating the solution of the minimization problem of $|\mathbf{R}_i|$ - we set the derivative of $f(t_i) = t_i^2 - \frac{2t_i e_M}{\sqrt{n}} + 1$ to zero, and get $t_i = \frac{e_M}{\sqrt{n}}$. The minimum of the approximated residue is therefore given by

$$
|\mathbf{R}_i|^2 \approx |\mathbf{R}_{i-1}|^2 \left(1 - \frac{e_M^2}{n}\right), \tag{4.14}
$$

$$
c_i = |\mathbf{R}_{i-1}| \frac{e_M}{\sqrt{n}}. \tag{4.15}
$$

Using the approximation in (4.9), $|\mathbf{R}_0| = |\mathbf{S}| \approx \sigma$, and it follows from (4.14) and (4.15) that

$$
|\mathbf{R}_i|^2 \approx \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^i \tag{4.16}
$$

$$
c_i = \frac{\sigma e_M}{\sqrt{n}} \sqrt{\left(1 - \frac{e_M^2}{n}\right)^{i-1}}, \tag{4.17}
$$

which coincides with $\{c_i\}_{i=1}^L$ in (3.6). After L steps of the recursion,

$$
|\mathbf{R}_L|^2 \approx \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L. \tag{4.18}
$$

The following Proposition shows that for a large enough L, the residue in (4.18) is approximately equal to $\sigma^2 e^{-2R}$, the distortion-rate function of a memoryless Gaussian source.

Proposition 1. *Let $M = L^b$ for some $b > 0$. Let $R > 0$ satisfy* (3.1)*. Then,*

$$
\lim_{L\to\infty} \left(1 - \frac{e_M^2}{n}\right)^L = e^{-2R}, \tag{4.19}
$$

with e_M given in (2.8)*.*

Proof. From (2.114),

$$
\begin{aligned}
\limsup_{L\to\infty} \left(1 - \frac{e_M^2}{n}\right)^L &\leq \limsup_{L\to\infty} \left(1 - \frac{2\log M - \log\log M - \log 4\pi}{n}\right)^L && (4.20)\\
&= \limsup_{L\to\infty} \left(1 - \frac{2R}{L} + \frac{\log\log L + \log 4b\pi}{\frac{b}{R} L \log L}\right)^L && (4.21)\\
&= \limsup_{L\to\infty} \left(1 - \frac{2R}{L}(1 - f(L))\right)^L && (4.22)\\
&= \limsup_{L\to\infty} \ \exp\left(L \log\left(1 - \frac{2R}{L}(1 - f(L))\right)\right) && (4.23)
\end{aligned}
$$

where (4.21) follows from (3.1) and $M = L^b$, and with

$$
f(L) \triangleq \frac{\log\log L + \log 4b\pi}{2b \log L}. \tag{4.24}
$$

From the inequality $\log(1-x) \leq -x$ for $x \in [0,1)$,

$$\limsup_{L\to\infty} \left(1 - \frac{e_M^2}{n}\right)^L \leq \limsup_{L\to\infty} \; e^{-2R+2Rf(L)} \tag{4.25}$$

$$= e^{-2R}. \tag{4.26}$$

For the lower bound we use (2.15) and (3.1) to get

$$\liminf_{L\to\infty} \left(1 - \frac{e_M^2}{n}\right)^L \geq \liminf_{L\to\infty} \left(1 - \frac{2\log M}{n}\right)^L \tag{4.27}$$

$$= \liminf_{L\to\infty} \left(1 - \frac{2R}{L}\right)^L \tag{4.28}$$

$$= e^{-2R}. \tag{4.29}$$

From (4.25)–(4.26) and (4.27)–(4.29), the result in (4.19) follows. □

The asymptotic analysis in Chapter 4 yields the achievability of the Gaussian distortion-rate function. The result in [26] is similar, but it is validated under the approximation $e_M \approx \sqrt{2\log M}$ (see (2.113)), and although asymptotically $\frac{e_M^2}{n} \approx \frac{2R}{L}$, it was not clear whether (4.19) holds without replacing e_M with its approximations.

The change we make in the recursion is taking the sequence of $\{c_i\}_{i=1}^L$ as given in (3.6), without relying on the approximation of e_M by $\sqrt{2\log M}$. This modification has two principal effects: on the one hand, it improves the approximation for (4.11), enabling us to get a tighter bound to the probability of excess distortion; but, on the other hand, it increases the residue after the L'th step, $|\mathbf{R}_L|$. As mentioned above, despite the increase in $|\mathbf{R}_L|$ the distortion-rate function is still approachable, making our choice for $\{c_i\}_{i=1}^L$ a valid one.

Chapter 5

Non-Asymptotic Upper Bound on the Distortion

This chapter derives a probabilistic non-asymptotic upper bound on the distortion of the SPARC from Section 3.2. First, Theorem 1 in Chapter 5.1 bounds the probability of excess distortion by the sum of three separate probabilities; then, in Sections 5.2–5.4 we prove that each of these three probabilities tend to 0 exponentially with n, the length of the source. Theorem 2 summarizes the results of this section, for the case of a memoryless Gaussian source.

5.1 Main Theorem

The following theorem relies on [26, Theorem 1]. However, it has three important modifications: first, the algorithm analyzed by Theorem 1 has been modified in the set of parameters $\{c_i\}$ as described in Section 3.2, in order to improve the performance of the code. Second, it is stated for finite L and M, rather than having just an asymptotic result for large enough M and L; finally, the bound on the probability of excess distortion is substantially tightened in comparison to [26].

Theorem 1. *Let* $\mathbf{S}$ *be an ergodic source sequence of length* n *whose symbols have zero mean and variance* σ^2*. Let* $\delta_0, \delta_1, \delta_2$ *be positive constants such that*

$$\Delta \triangleq \delta_0 + 5R\left(\delta_1 + \delta_2\right) < \frac{1}{2}. \tag{5.1}$$

Let $\mathbf{A}$ *be an* $n \times ML$ *matrix with i.i.d.* $\mathcal{N}(0,1)$ *entries independent of* $\mathbf{S}$*, with* $M > L$ *and* $M, L \in \mathbb{N}$ *satisfying* (3.1)*. Let* b *satisfy* $M = L^b$*. For the SPARC defined by the matrix* $\mathbf{A}$ *and for* $L \geq 10R$*, the encoding algorithm in Section 3.2 produces a codeword* $\mathbf{A}\hat{\beta}$*, for which*

$$\mathbb{P}\left[|\mathbf{S} - \mathbf{A}\hat{\beta}|^2 > \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2\right] < p_0 + p_1 + p_2, \tag{5.2}$$

with e_M as defined in (2.8), *and*

$$w \triangleq 1 + \frac{e_M^2}{2(n - e_M^2)}, \tag{5.3}$$

$$p_0 = \mathbb{P}\left[\left| \frac{|\mathbf{S}|}{\sigma} - 1 \right| > \delta_0 \right], \tag{5.4}$$

$$p_1 = \mathbb{P}\left[\frac{1}{L} \sum_{i=1}^{L} |\gamma_i| > \delta_1 \right], \tag{5.5}$$

$$p_2 = \mathbb{P}\left[\frac{1}{L} \sum_{i=1}^{L} |\epsilon_i| > \delta_2 \right], \tag{5.6}$$

where γ_i and ϵ_i are defined to satisfy

$$\max_{(i-1)M < j \leq iM} \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle = \left\langle \mathbf{A}_{m_i}, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle = e_M(1 + \epsilon_i), \tag{5.7}$$

$$|\mathbf{A}_{m_i}|^2 = 1 + \gamma_i, \tag{5.8}$$

for $i \in \{1, \dots, L\}$.

Proof. We first find an accurate recursion for $|\mathbf{R}_i|$. To that end, we denote the multiplicative deviation of $|\mathbf{R}_i|^2$ from its approximated value in (4.16) by Δ_i, i.e.,

$$|\mathbf{R}_i|^2 = \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^i (1 + \Delta_i)^2, \tag{5.9}$$

with $\Delta_i \geq -1$. According to the recursion in (3.5), and the notation in (2.2),

$$|\mathbf{R}_i|^2 = |\mathbf{R}_{i-1}|^2 + c_i^2 \, |\mathbf{A}_{m_i}|^2 - \frac{2c_i \|\mathbf{R}_{i-1}\|}{n} \left\langle \mathbf{A}_{m_i}, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle \tag{5.10}$$

$$= \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^{i-1} (1 + \Delta_{i-1})^2 + c_i^2 (1 + \gamma_i) - \frac{2c_i \sigma}{\sqrt{n}} \left(1 - \frac{e_M^2}{n}\right)^{\frac{i-1}{2}} (1 + \Delta_{i-1}) \, e_M (1 + \epsilon_i) \tag{5.11}$$

$$= \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^{i-1} \left[(1 + \Delta_{i-1})^2 + \frac{e_M^2}{n} (1 + \gamma_i) - \frac{2e_M^2}{n} (1 + \Delta_{i-1})(1 + \epsilon_i) \right] \tag{5.12}$$

$$= \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^{i} \left[\frac{n \, (1 + \Delta_{i-1})^2}{n - e_M^2} + \frac{e_M^2}{n - e_M^2} (1 + \gamma_i) - \frac{2e_M^2}{n - e_M^2} (1 + \Delta_{i-1})(1 + \epsilon_i) \right] \tag{5.13}$$

$$= \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^{i} \left[(1 + \Delta_{i-1})^2 + \frac{e_M^2}{n - e_M^2} \Big(\Delta_{i-1}^2 + \gamma_i - 2\epsilon_i (1 + \Delta_{i-1}) \Big) \right], \tag{5.14}$$

where (5.11) holds by assembling (5.7)–(5.9); (5.12) follows from (3.6). Thus, we obtain the following recursion for Δ_i:

$$(1+\Delta_i)^2 = (1+\Delta_{i-1})^2 + \left(\frac{e_M^2}{n-e_M^2}\right)\Big(\Delta_{i-1}^2 + \gamma_i - 2\epsilon_i(1+\Delta_{i-1})\Big). \tag{5.15}$$

Let $\mathcal{A}$ be the event which satisfies the following conditions:

1. $\left|\frac{|\mathbf{S}|}{\sigma} - 1\right| \leq \delta_0,$ (5.16)

2. $\frac{1}{L}\sum_{i=1}^{L} |\gamma_i| \leq \delta_1,$ (5.17)

3. $\frac{1}{L}\sum_{i=1}^{L} |\epsilon_i| \leq \delta_2.$ (5.18)

By the union bound, it follows that

$$\mathbb{P}[\mathcal{A}^c] \leq p_0 + p_1 + p_2, \tag{5.19}$$

where p_0, p_1 and p_2 are the probabilities defined in (5.4)–(5.6). We next derive an upper bound for the distortion in (5.9) after the final step conditioned on the event $\mathcal{A}$.

Lemma 5. *For $L \geq 10R$, conditioning on the event $\mathcal{A}$, we have*

$$\Delta_i \geq \Delta_0 - \frac{2e_M^2}{n-e_M^2}\sum_{j=1}^{i}\left(|\gamma_j| + |\epsilon_j|\right), \quad i = 1, \ldots, L. \tag{5.20}$$

Proof: See Appendix 1.

Lemma 6. *For $L \geq 10R$, conditioning on the event $\mathcal{A}$,*

$$|\Delta_i| \leq |\Delta_0| w^i + \frac{2e_M^2}{n-e_M^2}\sum_{j=1}^{i} w^{i-j}(|\gamma_j| + |\epsilon_j|), \quad i \in \{1, \ldots, L\}. \tag{5.21}$$

with w as defined in (5.3).

Proof: See Appendix 2.

Lemma 6 guarantees that conditioning on the event $\mathcal{A}$, for $L \geq 10R$,

$$|\Delta_L| \leq |\Delta_0| w^L + \frac{2e_M^2}{n-e_M^2}\sum_{j=1}^{L} w^{L-j}(|\gamma_j| + |\epsilon_j|) \tag{5.22}$$

$$= w^L\left(|\Delta_0| + \frac{2e_M^2 L}{n-e_M^2}\sum_{j=1}^{L} w^{-j}\frac{|\gamma_j| + |\epsilon_j|}{L}\right). \tag{5.23}$$

Since $w \geq 1$ (see (5.3) and (A.2)), and from (5.16)–(5.18),

$$\begin{aligned}
|\Delta_L| &\leq w^L \left(|\Delta_0| + \frac{2e_M^2 L}{n - e_M^2} \sum_{j=1}^{L} \frac{|\gamma_j| + |\epsilon_j|}{L} \right) && (5.24)\\
&\leq w^L \left(\delta_0 + \frac{2e_M^2 L}{n - e_M^2} \, (\delta_1 + \delta_2) \right), && (5.25)\\
&\leq w^L \left(\delta_0 + 5R(\delta_1 + \delta_2) \right) && (5.26)\\
&= w^L \Delta, && (5.27)
\end{aligned}$$

where in (5.25), $|\Delta_0| \leq \delta_0$ due to (3.3), (5.9) and (5.16), (5.26) holds due to (A.3), and (5.27) holds due to (5.1). By the definition of Δ_L in (5.9),

$$\begin{aligned}
|R_L|^2 &= \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L (1 + \Delta_L)^2 && (5.28)\\
&\leq \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2 && (5.29)
\end{aligned}$$

holds under the conditioning on the event $\mathcal{A}$. Therefore,

$$\begin{aligned}
\mathbb{P}\left[|R_L|^2 > \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2 \right] &\leq \mathbb{P}[\mathcal{A}^c] && (5.30)\\
&\leq p_0 + p_1 + p_2, && (5.31)
\end{aligned}$$

where (5.31) is (5.19). $\square$

Discussion on Theorem 1: Theorem 1 determines that the probability of excess distortion as in the left side of (5.30), is bounded by the sum of the probabilities p_0, p_1 and p_2 as defined in (5.4)–(5.6).

1. By their definition, only p_0 depends on the distribution of the source sequence $\mathbf{S}$. Since we assume that $\mathbf{S}$ is ergodic, p_0 tends asymptotically to 0 by letting $n \to \infty$. In Lemma 7 we show that under a condition, which applies to the Gaussian case among others, p_0 decays to zero exponentially in n.

2. Lemmas 9 and 10 give tight exponential bounds on p_1 and p_2 respectively. Thus, when the conditions of Lemma 7 hold, the probability of excess distortion decays exponentially with n.

3. Asymptotically, the difference between the Gaussian distortion-rate function,

$$D(R) = \sigma^2 e^{-2R}, \tag{5.32}$$

and $\sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2$ (see the left side in (5.30)), can be arbitrarily small, by a proper choice of L and Δ (i.e., L large enough and Δ small enough and close to zero).

This holds since by the definition of w in (5.3),

$$w^L = \left(1 + \frac{e_M^2}{2(n - e_M^2)}\right)^L \tag{5.33}$$

$$\leq \exp\left(\frac{e_M^2 L}{2(n - e_M^2)}\right) \tag{5.34}$$

$$\leq e^{5R/4}, \tag{5.35}$$

where (5.34) follows from the inequality $1 + x \leq e^x$ for $x \in \mathbb{R}$, and (5.35) follows from (A.4). Therefore, from Proposition 1, for every $\epsilon > 0$ there exist $L' \geq 10R$ and $\Delta > 0$ such that

$$\sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2 \leq \sigma^2 e^{-2R}(1 + \epsilon) \tag{5.36}$$

for all $L \geq L'$. Hence, it achieves the distortion-rate function of a memoryless Gaussian source; otherwise, for other memoryless i.i.d. source models, it achieves the distortion-rate function of a Gaussian source with the same variance.

5.2 Exact Expression for p_0 and Exponential Upper Bounds

In this section, Lemma 7 gives an upper bound to p_0 for a general source $\mathbf{S}$ under certain conditions, followed by two examples (an i.i.d. Gaussian source and an i.i.d. Uniform source); in Lemma 8, an accurate expression for p_0 is acquired for the i.i.d. Gaussian case.

Lemma 7. *Let X be a random variable with zero mean and variance σ^2, such that X^2 has a moment-generating function $M_{X^2}(t)$ in a neighborhood of $t = 0$. Let $\mathbf{S}$ be an i.i.d. source sequence of length n, generated according to the probability distribution of X, and let $\delta \in (0, 1)$. Then,*

$$p_0 \triangleq \mathbb{P}\left(\left|\frac{|\mathbf{S}|}{\sigma} - 1\right| > \delta\right) \leq \inf_{t>0}\left\{e^{-n\left[t(1+\delta)^2 - \log M_{X^2}\left(\frac{t}{\sigma^2}\right)\right]}\right\} + \inf_{t>0}\left\{e^{n\left[t(1-\delta)^2 + \log M_{X^2}\left(-\frac{t}{\sigma^2}\right)\right]}\right\}, \tag{5.37}$$

which decays exponentially to 0 *with* n.

Proof. From (2.2),

$$p_0 = \mathbb{P}\left(\frac{\|\mathbf{S}\|^2}{\sigma^2} > n(1 + \delta)^2\right) + \mathbb{P}\left(\frac{\|\mathbf{S}\|^2}{\sigma^2} < n(1 - \delta)^2\right). \tag{5.38}$$

The upper bound follows from the Chernoff bound. Applying it on the first term in the right side of (5.38) yields

$$\mathbb{P}\left(\frac{\|\mathbf{S}\|^2}{\sigma^2} > n(1+\delta)^2\right) \leq \inf_{t>0}\left\{e^{-tn(1+\delta)^2}\prod_{i=1}^{n}\mathbb{E}\left[e^{t\frac{S_i^2}{\sigma^2}}\right]\right\} \tag{5.39}$$

$$= \inf_{t>0}\left\{e^{-tn(1+\delta)^2}\left(\mathbb{E}\left[e^{t\frac{X^2}{\sigma^2}}\right]\right)^n\right\} \tag{5.40}$$

$$= \inf_{t>0}\left\{e^{-tn(1+\delta)^2}M_{X^2}^n\left(\frac{t}{\sigma^2}\right)\right\}. \tag{5.41}$$

Similarly, by applying the Chernoff bound on the second term in (5.38), we have

$$\mathbb{P}\left(\frac{\|\mathbf{S}\|^2}{\sigma^2} < n(1-\delta)^2\right) \leq \inf_{t>0}\left\{e^{tn(1-\delta)^2}\prod_{i=1}^{n}\mathbb{E}\left[e^{-t\frac{S_i^2}{\sigma^2}}\right]\right\} \tag{5.42}$$

$$= \inf_{t>0}\left\{e^{tn(1-\delta)^2}M_{X^2}^n\left(-\frac{t}{\sigma^2}\right)\right\}. \tag{5.43}$$

Combining (5.38), (5.39)–(5.41) and (5.42)–(5.43) yields (5.37). Let $f_1(t)$ and $f_2(t)$ be the following functions,

$$f_1(t) \triangleq t(1+\delta)^2 - \log M_{X^2}\left(\frac{t}{\sigma^2}\right), \tag{5.44}$$

$$f_2(t) \triangleq t(1-\delta)^2 + \log M_{X^2}\left(-\frac{t}{\sigma^2}\right). \tag{5.45}$$

Then,

$$f_1'(t) = (1+\delta)^2 - \frac{M_{X^2}'\left(\frac{t}{\sigma^2}\right)}{\sigma^2 M_{X^2}\left(\frac{t}{\sigma^2}\right)}, \tag{5.46}$$

$$f_2'(t) = (1-\delta)^2 - \frac{M_{X^2}'\left(-\frac{t}{\sigma^2}\right)}{\sigma^2 M_{X^2}\left(-\frac{t}{\sigma^2}\right)}, \tag{5.47}$$

and at $t = 0$,

$$f_1(0) = f_2(0) = 0, \tag{5.48}$$

$$f_1'(0) = (1+\delta)^2 - 1 > 0, \tag{5.49}$$

$$f_2'(0) = (1-\delta)^2 - 1 < 0. \tag{5.50}$$

Therefore, there exists $t_1 > 0$ and $t_2 > 0$ such that

$$f_1(t_1) > 0, \tag{5.51}$$

$$f_2(t_2) < 0. \tag{5.52}$$

From (5.37), (5.44), (5.45), (5.51) and (5.52), p_0 has a bound that is exponentially decreasing with n to 0. $\square$

Example 1 (Memoryless Gaussian Source)**.** If $\mathbf{S}$ is an i.i.d. source sequence of length n, generated according to the Gaussian distribution $\mathcal{N}(0, \sigma^2)$, then

$$M_{X^2}(t) = \left(1 - 2\sigma^2 t\right)^{-\frac{1}{2}}, \tag{5.53}$$

and from (5.37) we have

$$p_0 \leq \inf_{0<t<\frac{1}{2}} \left\{ e^{-n\left[t(1+\delta)^2 + \frac{1}{2}\log(1-2t)\right]} \right\} + \inf_{t>0} \left\{ e^{n\left[t(1-\delta)^2 - \frac{1}{2}\log(1+2t)\right]} \right\}. \tag{5.54}$$

In order to minimize the terms in (5.54), we define

$$f_1(t) \triangleq t(1+\delta)^2 + \frac{1}{2}\log(1-2t), \tag{5.55}$$
$$f_2(t) \triangleq t(1-\delta)^2 - \frac{1}{2}\log(1+2t), \tag{5.56}$$

and set the derivative of f_1 and f_2 to zero to find the optimal t's,

$$t_1^* = \frac{1}{2} - \frac{1}{2(1+\delta)^2}, \tag{5.57}$$
$$t_2^* = \frac{1}{2(1-\delta)^2} - \frac{1}{2} \tag{5.58}$$

Substituting t_1^* and t_2^* into (5.54) yields

$$p_0 \leq e^{-\frac{n}{2}\left(\delta^2 + 2\delta - 2\log(1+\delta)\right)} + e^{-\frac{n}{2}\left(\delta^2 - 2\delta - 2\log(1-\delta)\right)}. \tag{5.59}$$

Example 2 (Memoryless Uniform Source)**.** If $\mathbf{S}$ is an i.i.d. source sequence of length n, generated according to the Uniform distribution over $[-a, a]$ for some $a > 0$, then

$$M_{X^2}(t) = \frac{1}{2a}\int_{-a}^{a} e^{tx^2}\mathrm{d}x = \begin{cases} \frac{\sqrt{\pi}}{2a\sqrt{t}}\operatorname{erfi}\left(a\sqrt{t}\right), & \text{for } t > 0 \\ \frac{\sqrt{\pi}}{2a\sqrt{-t}}\operatorname{erf}\left(a\sqrt{-t}\right), & \text{for } t < 0, \end{cases} \tag{5.60}$$

where $\operatorname{erf}(\cdot)$ is the Gaussian error function and $\operatorname{erfi}(\cdot)$ is the Gaussian imaginary error function,

$$\operatorname{erf}(x) \triangleq \frac{2}{\sqrt{\pi}}\int_0^x e^{-s^2}\mathrm{d}s, \tag{5.61}$$
$$\operatorname{erfi}(x) \triangleq \frac{2}{\sqrt{\pi}}\int_0^x e^{s^2}\mathrm{d}s, \tag{5.62}$$

for $x \geq 0$. Then, from (5.37),

$$\begin{aligned} p_0 \leq & \inf_{t>0}\left\{\exp\left(-n\left(t(1+\delta)^2 - \log\left(\frac{\sqrt{\pi}}{2\sqrt{3t}}\operatorname{erfi}\left(\sqrt{3t}\right)\right)\right)\right)\right\} \\ & + \inf_{t>0}\left\{\exp\left(n\left(t(1-\delta)^2 + \log\left(\frac{\sqrt{\pi}}{2\sqrt{3t}}\operatorname{erf}\left(\sqrt{3t}\right)\right)\right)\right)\right\}. \end{aligned} \tag{5.63}$$

Lemma 8. *Let* $\mathbf{S}$ *be an i.i.d. source sequence of length* n*, generated according to the Gaussian distribution* $\mathcal{N}(0, \sigma^2)$*, and let* $\delta \in (0, 1)$*. Then,*

$$p_0 \triangleq \mathbb{P}\left(\left|\frac{|\mathbf{S}|}{\sigma} - 1\right| > \delta\right) = 1 - \bar{\gamma}\left(\frac{n}{2}, \frac{n(1+\delta)^2}{2}\right) + \bar{\gamma}\left(\frac{n}{2}, \frac{n(1-\delta)^2}{2}\right) \tag{5.64}$$

where $\bar{\gamma}$ *is the incomplete Gamma function,*

$$\bar{\gamma}(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} \mathrm{d}t. \tag{5.65}$$

Proof. Following the assumption of the lemma, $\frac{\|\mathbf{S}\|^2}{\sigma^2}$ is Chi-squared distributed with n degrees of freedom. Continuing from (5.38), we have

$$\begin{aligned}
p_0 &= \int_{n(1+\delta)^2}^{\infty} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \mathrm{d}x + \int_0^{n(1-\delta)^2} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \mathrm{d}x && (5.66)\\
&= 1 - \int_0^{n(1+\delta)^2} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \mathrm{d}x + \int_0^{n(1-\delta)^2} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \mathrm{d}x && (5.67)\\
&= 1 - \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^{n(1+\delta)^2} \frac{1}{2}\left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathrm{d}x + \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^{n(1-\delta)^2} \frac{1}{2}\left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathrm{d}x && (5.68)\\
&= 1 - \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^{\frac{n}{2}(1+\delta)^2} t^{\frac{n}{2}-1} e^{-t} \mathrm{d}t + \frac{1}{\Gamma\left(\frac{n}{2}\right)} \int_0^{\frac{n}{2}(1-\delta)^2} t^{\frac{n}{2}-1} e^{-t} \mathrm{d}t && (5.69)\\
&= 1 - \bar{\gamma}\left(\frac{n}{2}, \frac{n(1+\delta)^2}{2}\right) + \bar{\gamma}\left(\frac{n}{2}, \frac{n(1-\delta)^2}{2}\right). && (5.70)
\end{aligned}$$

□

Table 5.1 presents a brief comparison between the exact expression of p_0 for the Gaussian case in (5.64), the bound in (5.59) and the bound from [26],

$$p_0 \leq 2e^{-\frac{3n\delta^2}{4}}. \tag{5.71}$$

For certain values of n and δ, only the exact expression can be used to evaluate the probability p_0, since the bounds in (5.59) and (5.71) can be greater than 1. For other values, the exact expression is significantly better than both bounds.

$(n,\ \delta)$	Exact p_0 (5.64)	Chernoff Bound (New) (5.59)	Looser Chernoff Bound [26] (5.71)
(10, 0.1)	0.66	> 1	> 1
(1000, 0.1)	$7.86 \cdot 10^{-6}$	$6.19 \cdot 10^{-5}$	$1.1 \cdot 10^{-3}$

Table 5.1: Comparison between the exact value of p_0 and its Chernoff upper bounds.

5.3 An Upper Bound on p_1

Lemma 9. *Let $\{\mathbf{A}_{m_i}\}$ for $i \in \{1, \ldots, L\}$ be the columns of the matrix $\mathbf{A}$ as defined in* (3.4)*, and let γ_i be given by*

$$\gamma_i = |\mathbf{A}_{m_i}|^2 - 1, \quad 1 \le i \le L. \tag{5.72}$$

Then, for all $\delta \in (0, 1)$,

$$p_1 \triangleq \mathbb{P}\left(\frac{1}{L}\sum_{i=1}^{L} |\gamma_i| > \delta\right) \tag{5.73}$$

$$\le \left(\inf_{0<t<\frac{1}{2}} \left\{\left(\frac{e^{t(1-\delta)}}{\sqrt{1+2t}}\right)^n \bar{\gamma}\left(\tfrac{n}{2}, \tfrac{n}{2} + nt\right) + \left(\frac{e^{-t(1+\delta)}}{\sqrt{1-2t}}\right)^n \left[1 - \bar{\gamma}\left(\tfrac{n}{2}, \tfrac{n}{2} - nt\right)\right]\right\}\right)^L \tag{5.74}$$

where $\bar{\gamma}(\cdot, \cdot)$ denotes the incomplete Gamma function in (5.65)*.*

Proof. From (5.72) and (5.73),

$$p_1 = \mathbb{P}\left(\frac{1}{L}\sum_{i=1}^{L} \left||\mathbf{A}_{m_i}|^2 - 1\right| > \delta\right). \tag{5.75}$$

Since the columns in each section of the matrix $\mathbf{A}$ are i.i.d., for an arbitrary sequence of columns $\{\mathbf{A}_{k_i}\}_{i=1}^{L}$ with $k_i \in \{(i-1)M + 1, \ldots, iM\}$, we have

$$p_1 = \mathbb{P}\left(\frac{1}{L}\sum_{i=1}^{L} \left||\mathbf{A}_{k_i}|^2 - 1\right| > \delta\right) \tag{5.76}$$

$$= \mathbb{P}\left(\frac{1}{L}\sum_{i=1}^{L} \left|\frac{1}{n}\sum_{j=1}^{n} A_{j,k_i}^2 - 1\right| > \delta\right), \tag{5.77}$$

and $\left\{\left|\frac{1}{n}\sum_{j=1}^{n} A_{j,k_i}^2 - 1\right|\right\}_{i=1}^{L}$ are i.i.d. random variables. Applying the Chernoff bound to (5.77) yields

$$p_1 \le \inf_{t>0} \left\{ e^{-t\delta L} \left(\mathbb{E}\left[e^{t\left|\frac{1}{n}\sum_{j=1}^{n} A_{j,k_1}^2 - 1\right|}\right]\right)^L \right\}. \tag{5.78}$$

Since $A_{j,k_1} \sim \mathcal{N}(0, 1)$, the random variable $X_1 \triangleq \sum_{j=1}^{n} A_{j,k_1}^2$ is Chi-squared distributed with n degrees of freedom, and

$$\mathbb{E}\left[e^{t\left|\frac{1}{n}X_1 - 1\right|}\right] = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{t\left|\frac{x}{n} - 1\right|} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathrm{d}x \tag{5.79}$$

$$= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \left(e^t \int_0^n e^{-\left(\frac{1}{2}+\frac{t}{n}\right)x} x^{\frac{n}{2}-1} \mathrm{d}x + e^{-t} \int_n^\infty e^{-\left(\frac{1}{2}-\frac{t}{n}\right)x} x^{\frac{n}{2}-1} \mathrm{d}x\right). \tag{5.80}$$

Both integrals in the right side of (5.80) can be expressed in terms of the incomplete Gamma function as follows:

$$\int_0^n e^{-\left(\frac{1}{2}+\frac{t}{n}\right)x} x^{\frac{n}{2}-1} \mathrm{d}x = \int_0^{t+\frac{n}{2}} e^{-s} \left(\frac{s}{\frac{1}{2}+\frac{t}{n}}\right)^{\frac{n}{2}-1} \left(\frac{1}{\frac{1}{2}+\frac{t}{n}}\right) \mathrm{d}s \tag{5.81}$$

$$= \Gamma\left(\frac{n}{2}\right) \left(\frac{2n}{n+2t}\right)^{\frac{n}{2}} \bar{\gamma}\left(\frac{n}{2}, \frac{n}{2}+t\right), \tag{5.82}$$

and, similarly for $t \in (0, \frac{n}{2})$,

$$\int_n^{\infty} e^{-\left(\frac{1}{2}-\frac{t}{n}\right)x} x^{\frac{n}{2}-1} \mathrm{d}x = \int_{\frac{n}{2}-t}^{\infty} e^{-s} \left(\frac{s}{\frac{1}{2}-\frac{t}{n}}\right)^{\frac{n}{2}-1} \left(\frac{1}{\frac{1}{2}-\frac{t}{n}}\right) \mathrm{d}s \tag{5.83}$$

$$= \Gamma\left(\frac{n}{2}\right) \left(\frac{2n}{n-2t}\right)^{\frac{n}{2}} \left[1 - \bar{\gamma}\left(\frac{n}{2}, \frac{n}{2}-t\right)\right]. \tag{5.84}$$

Note that the integral on the left side of (5.83) diverges to $+\infty$ for all $t \geq \frac{n}{2}$. Substituting (5.82) and (5.84) into the right side of (5.80) gives

$$\mathbb{E}\left[e^{t\left|\frac{1}{n}X_1-1\right|}\right] = e^t \left(\frac{n}{n+2t}\right)^{\frac{n}{2}} \bar{\gamma}\left(\tfrac{n}{2}, \tfrac{n}{2}+t\right) + e^{-t} \left(\frac{n}{n-2t}\right)^{\frac{n}{2}} \left[1 - \bar{\gamma}\left(\tfrac{n}{2}, \tfrac{n}{2}-t\right)\right] \tag{5.85}$$

for all $t \in (0, \frac{n}{2})$. Substituting $t \mapsto nt$ in (5.85) and then optimizing numerically (5.78) over the free parameter $t \in (0, \frac{1}{2})$, we get the bound on p_1 in (5.74). □

Remark. A looser but more simple bound than (5.74), not involving the incomplete Gamma function, can be readily derived from (5.74) by relying on the fact that $0 \leq \bar{\gamma}(a, x) \leq 1$ for $a > 0$ and $x \geq 0$. This yields

$$p_1 \leq \left(\inf_{0<t<\frac{1}{2}} \left\{ \left(\frac{e^{t(1-\delta)}}{\sqrt{1+2t}}\right)^n + \left(\frac{e^{-t(1+\delta)}}{\sqrt{1-2t}}\right)^n \right\}\right)^L. \tag{5.86}$$

$(n,\, L,\, M,\, \delta)$	Chernoff Bound (5.74)	Looser Bound (5.86)	Bound in [26, (15)] (5.87)
$(10^2,\, 10,\, 10^2,\, 0.25)$	$1.83 \cdot 10^{-4}$	$3.30 \cdot 10^{-4}$	> 1
$(10^3,\, 10,\, 10^3,\, 0.10)$	$1.57 \cdot 10^{-8}$	$1.83 \cdot 10^{-8}$	> 1

Table 5.2: Comparison between upper bounds on p_1

Table 5.2 compares the upper bounds on p_1 in (5.74) and (5.86) with the bound in [26]:

$$p_1 \leq 2MLe^{-\frac{n\delta^2}{8}}. \tag{5.87}$$

As is illustrated in Table 5.2, the bound on p_1 in (5.86) is fairly tight in comparison to (5.74), whereas its upper bound in (5.87) (see [26, (15)]) is loose and it may even exceed 1.

5.4 An Upper Bound on p_2

Lemma 10. *For $i \in \{1, \ldots, L\}$, let ϵ_i be a random variable which stands for the deviation as defined in* (5.7)*:*

$$\max_{(i-1)M<j\leq iM} \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle = e_M(1+\epsilon_i). \tag{5.88}$$

Then, for all $\delta \in (0,1)$,

$$p_2 \triangleq P\left(\frac{1}{L}\sum_{i=1}^{L} |\epsilon_i| > \delta\right) \tag{5.89}$$

$$\leq \left(\inf_{t>0}\left\{e^{-te_M\delta}\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right\}\right)^L, \tag{5.90}$$

where $f_{Z_M}(\cdot)$ is the probability density function in (2.13)*. Moreover, the infimum in the right side of* (5.90) *is a minimum, which can be restricted to the finite interval $[0, t_M^*]$ with*

$$t_M^* = \frac{1}{e_M}\log\left(\frac{\sqrt{2}M}{\sqrt{\pi}f_{Z_M}(-2e_M)}\right). \tag{5.91}$$

Proof. Define the following random variables for $i \in \{1, \ldots, L\}$:

$$T^{(i)} \triangleq \big(T^{(i)}_{(i-1)M+1}, \ldots, T^{(i)}_{iM}\big), \tag{5.92}$$

with

$$T^{(i)}_j \triangleq \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle, \quad j \in \{(i-1)M+1, \ldots, iM\}. \tag{5.93}$$

Each of the M coordinates in the random vector $T^{(i)}$ is a function of two random vectors: one of the random columns $\mathbf{A}_j$ in the i'th section of the matrix $\mathbf{A}$, and the random vector $\mathbf{R}_{i-1}$. In view of (3.5), it follows by induction that the random vector $\mathbf{R}_{i-1}$ is a function of $\{\mathbf{A}_1, \ldots, \mathbf{A}_{(i-1)M}, \mathbf{R}_0\}$ only, while $\mathbf{A}_j$ for $j \in \{(i-1)M+1, \ldots, iM\}$ is independent of $\{\mathbf{A}_1, \ldots, \mathbf{A}_{(i-1)M}, \mathbf{R}_0\}$; therefore $\mathbf{A}_j$ is independent of $\mathbf{R}_{i-1}$. Furthermore, since $\{T^{(i-1)}, ..., T^{(1)}, \mathbf{R}_0\}$ is a function of $\{\mathbf{A}_1, \ldots, \mathbf{A}_{(i-1)M}, \mathbf{R}_0\}$, then $\mathbf{A}_j$ for $j \in \{(i-1)M+1, \ldots, iM\}$ remains independent of $\mathbf{R}_{i-1}$ conditioned on $\{T^{(i-1)}, ..., T^{(1)}, \mathbf{R}_0\}$. Therefore, according to Lemma 4,

$$F_{T^{(i)}|T^{(i-1)},\ldots,T^{(1)},\mathbf{R}_0} = F_{T^{(i)}} = \prod_{j=(i-1)M+1}^{iM} F_{T^{(i)}_j}, \quad i \in \{1, \ldots, L\}. \tag{5.94}$$

Applying recursively yields

$$F_{\mathbf{R}_0,T^{(1)}} = F_{\mathbf{R}_0}\, F_{T^{(1)}|\mathbf{R}_0} = F_{\mathbf{R}_0}\, F_{T^{(1)}}, \tag{5.95}$$

$$F_{\mathbf{R}_0,T^{(1)},T^{(2)}} = F_{\mathbf{R}_0,T^{(1)}}\, F_{T^{(2)}|T^{(1)},\mathbf{R}_0} = F_{\mathbf{R}_0}\, F_{T^{(1)}}\, F_{T^{(2)}}, \tag{5.96}$$

$$\vdots$$

$$F_{\mathbf{R}_0,T^{(1)},\ldots,T^{(L)}} = F_{\mathbf{R}_0} \prod_{i=1}^{L} F_{T^{(i)}}. \tag{5.97}$$

Hence, from Lemma 4, $\{T^{(i)}\}$ are i.i.d. Gaussian random vectors for $i \in \{1,\ldots,L\}$, and independent of $\mathbf{R}_0$, whose components are i.i.d. standard Gaussian random variables. It follows that the maximums of the random vectors $\{T^{(i)}\}$,

$$V_i = \max_{(i-1)M+1 \le j \le iM} T_j^{(i)}, \quad i \in \{1,\ldots,L\}, \tag{5.98}$$

are i.i.d. random variables which are distributed like Z_M (see (2.7)). Therefore, the deviations $\{\epsilon_i\}_{i=1}^{L}$, given by $\epsilon_i = \frac{V_i}{e_M} - 1$, are also i.i.d. random variables. Applying Chernoff's bound to the right side of (5.89) yields

$$p_2 = P\left(\frac{1}{L}\sum_{i=1}^{L} |\epsilon_i| > \delta\right) \le \inf_{t>0} \left(e^{-t\delta}\, \mathbb{E}\left[e^{t|\epsilon_1|}\right]\right)^L \tag{5.99}$$

$$= \left(\inf_{t>0}\left\{e^{-t\delta}\, \mathbb{E}\left[e^{t\left|\frac{Z_M}{e_M}-1\right|}\right]\right\}\right)^L \tag{5.100}$$

$$= \left(\inf_{t>0}\left\{e^{-te_M\delta}\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\ \mathrm{d}z\right\}\right)^L, \tag{5.101}$$

where (5.101) follows from (5.100) by the mapping $t \mapsto e_M t$. Following (5.101),

$$p_2 \le \exp\left(-L\ \sup_{t>0}\left\{te_M\delta - \log\left(\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\ \mathrm{d}z\right)\right\}\right). \tag{5.102}$$

Define, for $t > 0$,

$$u_{\delta,M}(t) \triangleq \frac{\mathrm{d}}{\mathrm{d}t}\left(\log\left(\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\ \mathrm{d}z\right) - te_M\delta\right) \tag{5.103}$$

$$= \frac{\int_{-\infty}^{\infty} |z-e_M|\, e^{t|z-e_M|} f_{Z_M}(z)\ \mathrm{d}z}{\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\ \mathrm{d}z} - e_M\delta \tag{5.104}$$

$$= \frac{\mathbb{E}\left[|Z_M - e_M|\, e^{t|Z_M-e_M|}\right]}{\mathbb{E}\left[e^{t|Z_M-e_M|}\right]} - e_M\delta, \tag{5.105}$$

so the derivative of $u_{\delta,M}$ is equal to

$$u'_{\delta,M}(t) = \frac{\mathbb{E}\left[(Z_M - e_M)^2 e^{t|Z_M-e_M|}\right]\mathbb{E}\left[e^{t|Z_M-e_M|}\right] - \left(\mathbb{E}\left[|Z_M-e_M|\, e^{t|Z_M-e_M|}\right]\right)^2}{\left(\mathbb{E}\left[e^{t|Z_M-e_M|}\right]\right)^2} \tag{5.106}$$

$$\geq 0, \tag{5.107}$$

where (5.107) follows from the Cauchy-Schwarz inequality; thus, $u_{\delta,M}$ is monotonically increasing. Therefore, the infimum in (5.102) can be obtained numerically by the bisection method in the interval $[0, t^*_M]$ for any $t^*_M > 0$ such that $u_{\delta,M}(t^*_M) \geq 0$.

Let t^*_M be as given in (5.91). From (5.105), since $\delta \in (0,1)$,

$$u_{\delta,M}(t) > \frac{\mathbb{E}\left[|Z_M-e_M|\, e^{t|Z_M-e_M|}\right]}{\mathbb{E}\left[e^{t|Z_M-e_M|}\right]} - e_M, \tag{5.108}$$

so, it is sufficient to show that

$$\mathbb{E}\left[|Z_M-e_M|\, e^{t^*_M|Z_M-e_M|}\right] - e_M\mathbb{E}\left[e^{t^*_M|Z_M-e_M|}\right] \geq 0. \tag{5.109}$$

For $t > 0$,

$$\begin{aligned}
&\mathbb{E}\left[|Z_M-e_M|\, e^{t|Z_M-e_M|}\right] - e_M\,\mathbb{E}\left[e^{t|Z_M-e_M|}\right] \\
&= \int_{-\infty}^{\infty} (|z-e_M| - e_M)\, e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z && (5.110)\\
&= \int_{-\infty}^{e_M} (-z)\, e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z + \int_{e_M}^{\infty} (z-2e_M)\, e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z && (5.111)\\
&\geq \left(\int_{-2e_M}^{-e_M} (-z)\, e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z - \int_0^{e_M} z e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right) + \int_{e_M}^{2e_M} (z-2e_M) e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z, && (5.112)
\end{aligned}$$

where (5.112) holds since the integrand of the left integral in (5.111) is non-negative for $z \in (-\infty, 0]$, and since the integrand of the right integral is non-negative for $z \in [2e_M, \infty)$. From (2.13), note that $f_{Z_M}(z) \leq \frac{M}{\sqrt{2\pi}}$ for all $z \in \mathbb{R}$, and that $f_{Z_M}(\cdot)$ is monotonically increasing on $(-\infty, 0]$. Consequently, we get from (5.110)–(5.112) that

$$\mathbb{E}\left[|Z_M-e_M|\, e^{t|Z_M-e_M|}\right] - e_M\,\mathbb{E}\left[e^{t|Z_M-e_M|}\right] \geq e_M^2 e^{2te_M} f_{Z_M}(-2e_M) - e_M^2\frac{M}{\sqrt{2\pi}}e^{te_M} - e_M^2\frac{M}{\sqrt{2\pi}}e^{te_M} \tag{5.113}$$

$$= e_M^2\left(f_{Z_M}(-2e_M)\; e^{te_M} - M\sqrt{\frac{2}{\pi}}\right) e^{te_M}. \tag{5.114}$$

For $t = t^*_M$, the right side of (5.114) is equal to 0, and therefore $u_{\delta,M}(t^*_M) > 0$. □

$(L,\, M,\, \delta)$	Chernoff Bound (5.90)	Bound in [26, (15)] (5.115)
(10, 100, 0.25)	$1.95 \cdot 10^{-2}$	> 1
(10, 1000, 0.1)	$7.96 \cdot 10^{-1}$	> 1
(10, 10^4, 0.25)	$4.15 \cdot 10^{-11}$	$4.7 \cdot 10^{-2}$

Table 5.3: Comparison between upper bounds on p_2

Table 5.3 compares the bound in (5.90) with the corresponding bound from [26, (15)],

$$p_2 < \left(\frac{M^{2\delta}}{8 \log M}\right)^{-L}. \tag{5.115}$$

It should be mentioned that our probability p_2 is not identical to the one in [26], since the deviations $\{\epsilon_i\}_{i=1}^{L}$ are defined differently; however, the two probabilities should be similar for large values of L.

5.5 Summary

The following theorem summarizes the results we have proved in Sections 5.1–5.4.

Theorem 2. *Let* $\mathbf{S}$ *be an i.i.d. source sequence of length* n*, generated according to the Gaussian distribution* $\mathcal{N}(0, \sigma^2)$*. Let* $\delta_0, \delta_1, \delta_2$ *be positive constants such that*

$$\Delta \triangleq \delta_0 + 5R\,(\delta_1 + \delta_2) < \frac{1}{2}. \tag{5.116}$$

Let $\mathbf{A}$ *be as defined in Section 3.2. For the SPARC with the matrix* $\mathbf{A}$ *and for* $L \geq 10R$*, the encoding algorithm in Section 3.2 produces a codeword* $\mathbf{A}\hat{\beta}$*, for which*

$$\begin{aligned}\mathbb{P}\left[|\mathbf{S} - \mathbf{A}\hat{\beta}|^2 > \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2\right] &< 1 - \bar{\gamma}\left(\frac{n}{2}, \frac{n(1+\delta_0)^2}{2}\right) + \bar{\gamma}\left(\frac{n}{2}, \frac{n(1-\delta_0)^2}{2}\right) \\ &+ \left(\inf_{0<t<\frac{1}{2}} \left\{\left(\frac{e^{t(1-\delta_1)}}{\sqrt{1+2t}}\right)^n \bar{\gamma}\left(\tfrac{n}{2}, \tfrac{n}{2} + nt\right) + \left(\frac{e^{-t(1+\delta_1)}}{\sqrt{1-2t}}\right)^n \left[1 - \bar{\gamma}\left(\tfrac{n}{2}, \tfrac{n}{2} - nt\right)\right]\right\}\right)^L \\ &+ \left(\inf_{t>0} \left\{e^{-te_M\delta_2} \int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right\}\right)^L,\end{aligned} \tag{5.117}$$

with e_M *as defined in* (2.8) *and where* $f_{Z_M}(\cdot)$ *the probability density function in* (2.13).

Proof. Apply Lemma 8, Lemma 9 and Lemma 10 on p_0, p_1 and p_2 respectively in Theorem 1. □

Table 5.4 compares the bound in (5.117) with the bound in [26],

$$\mathbb{P}\left[|\mathbf{S} - \mathbf{A}\hat{\beta}|^2 > \sigma^2 e^{-2R}\left(1 + e^R \Delta\right)^2\right] < 2e^{-\frac{3n\delta_0^2}{4}} + 2MLe^{-\frac{n\delta_1^2}{8}} + \left(\frac{M^{2\delta_2}}{8 \log M}\right)^{-L}, \tag{5.118}$$

$(n,\, L,\, M,\, \delta_0,\, \delta_1,\, \delta_2)$	Our Bound (5.117)	Bound in [26] (5.118)
$(10^3,\, 10,\, 10^3,\, 0.05,\, 0.1,\, 0.1)$	0.82	> 1
$(10^3,\, 50,\, 2.5 \cdot 10^3,\, 0.05,\, 0.1,\, 0.1)$	0.044	> 1
$(5 \cdot 10^3,\, 50,\, 1.25 \cdot 10^5,\, 0.03,\, 0.2,\, 0.2)$	0.0027	0.069

Table 5.4: Comparison between upper bounds on the probability of excess distortion

which is proven for sufficiently large M and L. The probabilities in (5.117) and (5.118) describe different events, as is evident from their expressions. Furthermore, the encoding algorithm in Section 3.2 is an altered version of the one in [26]. However, for large values of n and for small values of Δ, they should be similar.

Chapter 6

Gap to the Distortion-Rate vs. Complexity

Throughout Chapter 5, we show that Theorem 1 gives an upper bound on the probability of excess distortion for the SPARC from Section 3.2; however, Chapter 5 does not give explicit indication as to how rapidly the excess distortion decays with n, i.e. how fast the distortion of the SPARC tends to the distortion-rate function of a memoryless Gaussian source, as function of the source block size n. In this section, an upper bound on this convergence rate is proved for the case of a memoryless Gaussian source, and it is shown that this bound cannot be significantly improved via Theorem 1.

Let σ, L, M and n be as defined in the encoding algorithm in Section 3.2. For some positive $\Delta < \frac{1}{2}$, define $\alpha_{L,\Delta}$ as

$$\alpha_{L,\Delta} \triangleq \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2, \tag{6.1}$$

with e_M as defined in (2.8) and w as in (5.3). $\alpha_{L,\Delta}$ is the "reference point" to which the excess distortion is compared in Theorem 1; as previously shown in Chapter 5, from Proposition 1 and (5.35),

$$\lim_{\Delta \to 0} \lim_{L \to \infty} \alpha_{L,\Delta} = \sigma^2 e^{-2R}, \tag{6.2}$$

the distortion-rate function of a memoryless Gaussian source. Let $D_{L,\Delta}$ be

$$D_{L,\Delta} \triangleq \left|\alpha_{L,\Delta} - \sigma^2 e^{-2R}\right|, \tag{6.3}$$

the difference between the Gaussian distortion-rate function and the reference point. From (6.2), it is clear that

$$\lim_{\Delta \to 0} \lim_{L \to \infty} D_{L,\Delta} = 0, \tag{6.4}$$

hence the reference point tends to the Gaussian distortion-rate function when $L \to \infty$ and

$\Delta \to 0$. Two points remain unclear, however:

1. What is a sufficient condition on Δ as function of L, so that the upper bound to the probability of excess distortion in Theorem 1 tends to 0 with n.

2. Using Δ that holds the condition of item 1, what is the convergence rate of $D_{L,\Delta}$ to 0.

In order to resolve item 1, we show that if

$$\Delta = \frac{A \log\log L}{b \log L} \tag{6.5}$$

for some $A > \frac{1}{4}$, then the upper bound in Theorem 1 decreases to 0 with n for the case of a memoryless Gaussian source.

Lemma 11. *Let p_0 and p_1 be the probabilities defined in* (5.4)–(5.5). *If the source sequence* $\mathbf{S}$ *is generated by a memoryless Gaussian distribution, then for $\delta_0 = \delta_1 = \frac{1}{\log L}$,*

$$\lim_{L\to\infty} p_0 + p_1 = 0. \tag{6.6}$$

Proof. From (5.71) and (5.87),

$$p_0 + p_1 \le 2\exp\left(-\frac{3n\delta_0^2}{4}\right) + 2ML\exp\left(-\frac{n\delta_1^2}{8}\right). \tag{6.7}$$

Therefore, if $\delta_0 = \delta_1 = \frac{1}{\log L}$,

$$\begin{aligned}
\limsup_{L\to\infty}\ p_0 + p_1 &\le \limsup_{L\to\infty}\left(2\exp\left(-\frac{3n}{4\log^2 L}\right) + 2ML\exp\left(-\frac{n}{8\log^2 L}\right)\right) && (6.8)\\
&= \limsup_{L\to\infty}\left(2\exp\left(-\frac{3bL}{4R\log L}\right) + 2ML\exp\left(-\frac{bL}{8R\log L}\right)\right) && (6.9)\\
&= 0, && (6.10)
\end{aligned}$$

where (6.9) is due to (3.1). Since $p_0 + p_1 \ge 0$, the limit in (6.6) exists and is equal to 0. □

Lemma 12. *Let Z_M and e_M be defined as in* (2.7) *and* (2.8) *respectively. Then,*

$$\limsup_{M\to\infty} \frac{1}{e_M}\int_{-\infty}^{\infty} e^{e_M|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z < 13.6 + \frac{1}{2\sqrt{2}}. \tag{6.11}$$

Proof: See Appendix 3.

Corollary 2. *Let p_2 be the probability defined in* (5.6). *If $\delta_2 = \frac{A\log\log L}{b\log L}$ for some constant $A > \frac{1}{4}$, then*

$$\lim_{L\to\infty} p_2 = 0. \tag{6.12}$$

Proof. From (5.90),

$$p_2 \leq \exp\left[-L \sup_{t>0}\left\{\left(te_M\delta_2 - \log\left(\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right)\right)\right\}\right] \tag{6.13}$$

$$\leq \exp\left[-L\left(\frac{Ae_M^2 \log\log M}{\log M} - \log\left(\int_{-\infty}^{\infty} e^{e_M|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right)\right)\right], \tag{6.14}$$

where (6.14) follows from (3.1) and by choosing $t = e_M$. According to Lemma 12, for a large enough M,

$$\log\left(\int_{-\infty}^{\infty} e^{e_M|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right) \leq \log(e_M \cdot C), \tag{6.15}$$

where $C = 13.6 + \frac{1}{2\sqrt{2}}$. Thus, from (2.113), for a large enough M,

$$\log\left(\int_{-\infty}^{\infty} e^{e_M|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right) \leq \frac{Ae_M^2 \log\log M}{\log M} \tag{6.16}$$

when $A > \frac{1}{4}$. From (6.14) and (6.16), we have

$$\lim_{L\to\infty} p_2 = 0. \tag{6.17}$$

$\square$

Lemma 11 and Corollary 2 imply that for a memoryless Gaussian source, choosing $\Delta = \frac{A \log\log L}{b \log L}$ for some $A > \frac{1}{4}$ guarantees that there exists a partition of Δ into δ_0, δ_1 and δ_2 (see (5.1)), such that

$$\lim_{L\to\infty} p_0 + p_1 + p_2 = 0, \tag{6.18}$$

where p_0–p_2 are defined as in (5.4)–(5.6). Hence, for such a Δ, the probability of excess distortion in Theorem 1 tends to 0 with n. The next step is finding the convergence rate of $D_{L,\Delta}$ to 0 when $\Delta = \frac{A \log\log L}{b \log L}$. To this end, we first prove the following lemma:

Lemma 13. *Let M, L, n, R and b as defined in the encoding algorithm in Section 3.2. Then,*

$$\lim_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1 - \frac{e_M^2}{n}\right)^L - e^{-2R}\right) = \frac{R}{b} e^{-2R}, \tag{6.19}$$

with e_M as given in (2.8).

Proof: See Appendix 4.

Lemma 14. *Let M, L, n, R and b as defined in the encoding algorithm in Section 3.2. Then,*

$$\limsup_{L\to\infty}\; L\left(w^L - e^R\right) = R^2 e^R, \tag{6.20}$$

with w as given in (5.3).

Proof. From (5.34),

$$\limsup_{L\to\infty} L\left(w^L - R\right) \leq \limsup_{L\to\infty} L\left(\exp\left(\frac{e_M^2 L}{2(n-e_M^2)}\right) - e^R\right) \tag{6.21}$$

$$\leq \limsup_{L\to\infty} L\left(\exp\left(\frac{bL\log L}{2(\frac{b}{R}L\log L - b\log L)}\right) - e^R\right) \tag{6.22}$$

$$= \limsup_{L\to\infty} Le^R\left(\exp\left(\frac{R^2}{L-R}\right) - 1\right), \tag{6.23}$$

where (6.22) follows from (3.1). Using L'Hôpital's rule,

$$\limsup_{L\to\infty} L\left(w^L - R\right) \leq \limsup_{L\to\infty} \frac{R^2L^2e^R}{(L-R)^2}\exp\left(\frac{R^2}{L-R}\right) \tag{6.24}$$

$$= R^2e^R. \tag{6.25}$$

□

Corollary 3. *Let L, R, b and σ be as defined in the encoding algorithm in Section 3.2. Then, for a large enough L,*

$$\frac{\log\log L}{\log L}\cdot\frac{R\sigma^2e^{-2R}}{b} + O\left(\frac{\epsilon(L)\log\log L}{\log L}\right) \leq D_{L,\Delta}, \tag{6.26}$$

where $D_{L,\Delta}$ is as defined in (6.3)*, and where $\epsilon(\cdot)$ is some non-negative function such that*

$$\lim_{L\to\infty}\epsilon(L) = 0. \tag{6.27}$$

Furthermore, if

$$\Delta = \frac{A\log\log L}{b\log L} \tag{6.28}$$

for some constant $A > \frac{1}{4}$, then for a large enough L,

$$D_{L,\Delta} \leq \frac{\log\log L}{\log L}\left(Re^{-2R} + 2Ae^{-R}\right)\frac{\sigma^2}{b} + O\left(\frac{\epsilon(L)\log\log L}{\log L}\right). \tag{6.29}$$

Proof. From Lemma 13, there exists a non-negative function $\epsilon(\cdot)$ such that

$$\lim_{L\to\infty}\epsilon(L) = 0, \tag{6.30}$$

and

$$\frac{R}{b}e^{-2R} - \epsilon(L) \leq \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right) \leq \frac{R}{b}e^{-2R} + \epsilon(L); \tag{6.31}$$

therefore,

$$e^{-2R} + \frac{\log \log L}{\log L} \left(\frac{R}{b} e^{-2R} - \epsilon(L) \right) \leq \left(1 - \frac{e_M^2}{n} \right)^L \leq e^{-2R} + \frac{\log \log L}{\log L} \left(\frac{R}{b} e^{-2R} + \epsilon(L) \right). \tag{6.32}$$

Similarly, from Lemma 14,

$$w^L \leq e^R + O\left(\frac{1}{L}\right). \tag{6.33}$$

By definition of $\alpha_{L,\Delta}$ in (6.1),

$$\alpha_{L,\Delta} = \sigma^2 \left(1 - \frac{e_M^2}{n} \right)^L \left(1 + w^L \Delta \right)^2, \tag{6.34}$$

and so, according to (6.32) and (6.33),

$$\alpha_{L,\Delta} \leq \sigma^2 \left(e^{-2R} + \frac{\log \log L}{\log L} \cdot \frac{R}{b} e^{-2R} + \frac{\epsilon(L) \log \log L}{\log L} \right) \left(1 + e^R \Delta + \Delta O\left(\frac{1}{L}\right) \right)^2 \tag{6.35}$$

$$\leq \sigma^2 \left(e^{-2R} + \frac{\log \log L}{\log L} \cdot \frac{R}{b} e^{-2R} + \frac{\epsilon(L) \log \log L}{\log L} \right) \left(1 + e^{2R} \Delta^2 + 2e^R \Delta + \Delta O\left(\frac{1}{L}\right) \right) \tag{6.36}$$

$$\leq \sigma^2 \left[e^{-2R} + \frac{\log \log L}{\log L} \left(\frac{R}{b} e^{-2R} + \frac{2A}{b} e^{-R} \right) \right] + O\left(\frac{\epsilon(L) \log \log L}{\log L} \right), \tag{6.37}$$

where (6.37) is due to (5.35) and (6.28). Therefore,

$$D_{L,\Delta} \leq \frac{\log \log L}{\log L} \left(\frac{R\sigma^2}{b} e^{-2R} + \frac{2A\sigma^2}{b} e^{-R} \right) + O\left(\frac{\epsilon(L) \log \log L}{\log L} \right). \tag{6.38}$$

On the other hand, from (6.32),

$$\alpha_{L,\Delta} \geq \sigma^2 \left(e^{-2R} + \frac{\log \log L}{\log L} \cdot \frac{R}{b} e^{-2R} - \frac{\epsilon(L) \log \log L}{\log L} \right) \left(1 + w^L \Delta \right)^2 \tag{6.39}$$

$$\geq \sigma^2 \left(e^{-2R} + \frac{\log \log L}{\log L} \cdot \frac{R}{b} e^{-2R} \right) + O\left(\frac{\epsilon(L) \log \log L}{\log L} \right), \tag{6.40}$$

where (6.40) is due to $w, \Delta \geq 0$, as per their definitions in (5.1) and (5.3). Subsequently, we have

$$D_{L,\Delta} \geq \frac{\log \log L}{\log L} \cdot \frac{R\sigma^2 e^{-2R}}{b} + O\left(\frac{\epsilon(L) \log \log L}{\log L} \right). \tag{6.41}$$

□

Two conclusions can be drawn from Corollary 3. Firstly, the inequality in (6.26) gives an asymptotic lower bound on $D_{L,\Delta}$ regardless of the distribution of the source sequence **S** and of

Δ; thus, it shows that using Theorem 1, the best possible upper bound to the gap between the distortion of the SPARC and the distortion-rate function of a memoryless Gaussian source is

$$\frac{\log\log L}{\log L}\cdot\frac{R\sigma^2 e^{-2R}}{b}+O\left(\frac{\epsilon(L)\log\log L}{\log L}\right), \tag{6.42}$$

where $\epsilon(\cdot)$ is some non-negative function such that $\epsilon(L)\xrightarrow[L\to\infty]{}0$. Secondly, for the case of a memoryless Gaussian source sequence $\mathbf{S}$, (6.29) asymptotically upper bounds $D_{L,\Delta}$ for a specific sequence of Δ, such that the probability of excess distortion in Theorem 1 decays to 0. Hence, it proves that asymptotically,

$$\frac{\log\log L}{\log L}\left(\frac{R\sigma^2}{b}e^{-2R}+\frac{2A\sigma^2}{b}e^{-R}\right)+O\left(\frac{\epsilon(L)\log\log L}{\log L}\right) \tag{6.43}$$

is an upper bound to the gap with probability that tends to 1. The asymptotic lower bound in (6.42) suggests that this upper bound cannot be significantly improved using Theorem 1, since the difference between the lower and upper bounds is only in the coefficient, rather than in their asymptotic behavior.

In [26], a similar result can be derived, wherein the upper bound on the gap between the distortion of the SPARC and the distortion-rate function of a memoryless Gaussian source is

$$\frac{\log\log L}{\log L}\cdot\frac{\sigma^2 e^{-R}}{b}+O\left(\left(\frac{\log\log L}{\log L}\right)^2\right); \tag{6.44}$$

in comparison with the upper bound in (6.43), our coefficient is smaller, especially for low values of R.

Remark. The demand for the source sequence to be memoryless Gaussian is necessary solely to be able to apply the upper bound on p_0 in Lemma 11. For other memoryless source distributions, the same results as in Corollary 3 can be achieved, as long as $p_0\xrightarrow[L\to\infty]{}0$ for $\delta_0=\frac{1}{\log L}$.

Chapter 7

Computer Simulations

7.1 Comparison Between the Performance of the Two Encoding Algorithms

As previously stated, the algorithm presented in Section 3.2 is slightly different than the algorithm in [26], as the set of coefficients $\{c_i\}_{i=1}^{L}$ was changed to minimize the distortion. The computer simulation in this section compares the performance of the two SPARCs, with regard to the distortion as function of the rate. The simulation was conducted by generating an i.i.d. Gaussian source sequence $\mathbf{S}$ with zero mean and unit variance, with $b = 2, 3$ and over a range of values for the rate R and L (the number of sections in the matrix $\mathbf{A}$). Each data point in Figures 7.1a and 7.1b is an average over a total of 100 iterations; every iteration was simulated by randomly generating a source sequence $\mathbf{S}$ and a matrix $\mathbf{A}$, and calculating the distortion between the codeword and the source sequence for both algorithms.

Figures 7.1a and 7.1b show that for all R's and L's examined, and for both $b = 2$ and $b = 3$, the algorithm presented in Section 3.2 performs better than the one in [26]. The improvement is especially significant for higher values of R, because then the less accurate approximation used in [26] for the maximum of M i.i.d. Gaussian random variables is more prominent than for lower R's.

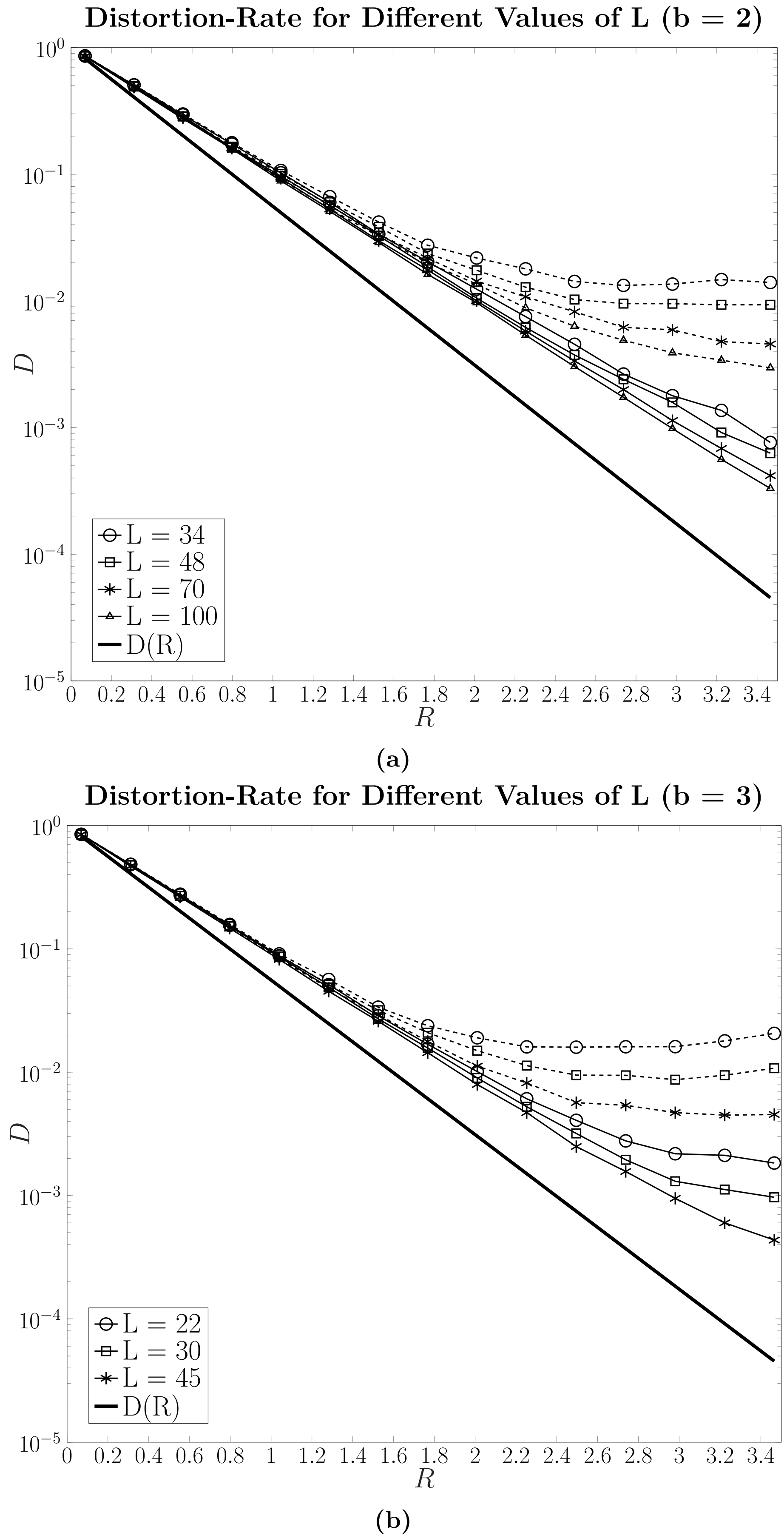

Figure 7.1: The distortion as function of the rate for several values of L with $b = 2, 3$. The solid lines were calculated using the algorithm in Section 3.2; the dashed lines were calculated according to the algorithm in [26]; the bold line is the Distortion-Rate function.

7.2 Comparison Between the Upper Bound on the Probability of Excess Distortion and the Simulated Results

In Chapter 5, an upper bound on the probability of excess distortion is achieved for a finite source blocklength n, which significantly improves the existing bound from [26]. In this section, we compare the improvement of the bound while taking into account the difference between the two encoding algorithms; furthermore, we analyze the tradeoff between the performance of the SPARC and its computational complexity derived by Theorem 2, and compare it to a computer simulation.

For the analysis of the tradeoff between complexity and performance guaranteed by Theorem 2, we fix the parameters R, b, σ^2 and ϵ, and calculate the minimal blocklength n such that

$$\mathbb{P}\left[|\mathbf{S}-\mathbf{A}\hat{\beta}|^2 > \sigma^2 e^{-2R}(1+\eta)\right] < \epsilon, \tag{7.1}$$

for a range of values of $\eta > 0$. From (5.2), the condition in (7.1) boils down to the following conditions,

- $$p_0 + p_1 + p_2 \leq \epsilon, \tag{7.2}$$
- $$\left(1 - \frac{e_M^2}{n}\right)^L \left(1 + w^L \Delta\right)^2 \leq e^{-2R}(1+\eta). \tag{7.3}$$

Therefore, finding the minimal blocklength n derived from Theorem 2 was performed numerically as follows:

1. Set the parameters R, b, σ^2, ϵ and η. Pick an initial value $n = n_0$.
2. For $i \geq 0$,
   - 2.1. Compute L and M given $n = n_i$ (see (3.1) and recall that $M = L^b$).
   - 2.2. Find a maximal value for Δ from (7.3).
   - 2.3. Compute the minimal value of $P \triangleq p_0 + p_1 + p_2$ (see (5.64), (5.74), (5.90)) under condition (5.1). The minimization is taken with respect to the two free parameters δ_0 and δ_1 which then restrict δ_2 in (5.1).
   - 2.4. If $P > \epsilon$, set $n_{i+1} = 2n_i$ and return to Step 2.1. If $P < \epsilon$, set $n_{\max} = n_i$, $n_{\min} = \frac{n_i}{2}$ and exit the loop.
3. Perform the bisection method in the interval $[n_{\min}, n_{\max}]$:
   - 3.1. Set $n_i = \left\lceil \frac{1}{2}(n_{\max} + n_{\min}) \right\rceil$.
   - 3.2. Perform Steps 2.1-2.3 with $n = n_i$.

3.3. If $P > \epsilon$, set $n_{\min} = n$ and return to Step 3.1. If $P < \epsilon$, set $n_{\max} = n$ and return to Step 3.1. If $n_{\max} = n_{\min} + 1$, then exit the loop.

In addition, a simulation was conducted to find the actual necessary minimal blocklength n for the SPARC from Section 3.2, such that it has a distortion greater than D with probability of at most ϵ. The simulation was carried out by generating random source sequences $\mathbf{S}$ and matrices $\mathbf{A}$ for different values of n, and calculating the distortion D that ϵ of them exceed.

Figure 7.2 shows a comparison between the minimal blocklength n derived from Theorem 2, and the minimal blocklength n acquired via the simulation of the SPARC. It is evident from Figure 7.2 that a considerable difference exists between the two values; this gap implies that the bound in Theorem 2 is not tight. The reason for this difference is mainly due to taking the "worst case scenario" for $|\Delta_i|$ in each of the L steps of Lemma 6, and then summing over them (See (B.11)–(B.15)).

Figure 7.3 shows a comparison between the minimal blocklength n derived from Theorem 2 and the one derived from [26, Theorem 1]. Since the upper bound on the probability of excess distortion was improved, the minimal blocklength n from Theorem 2 is significantly lower, as expected.

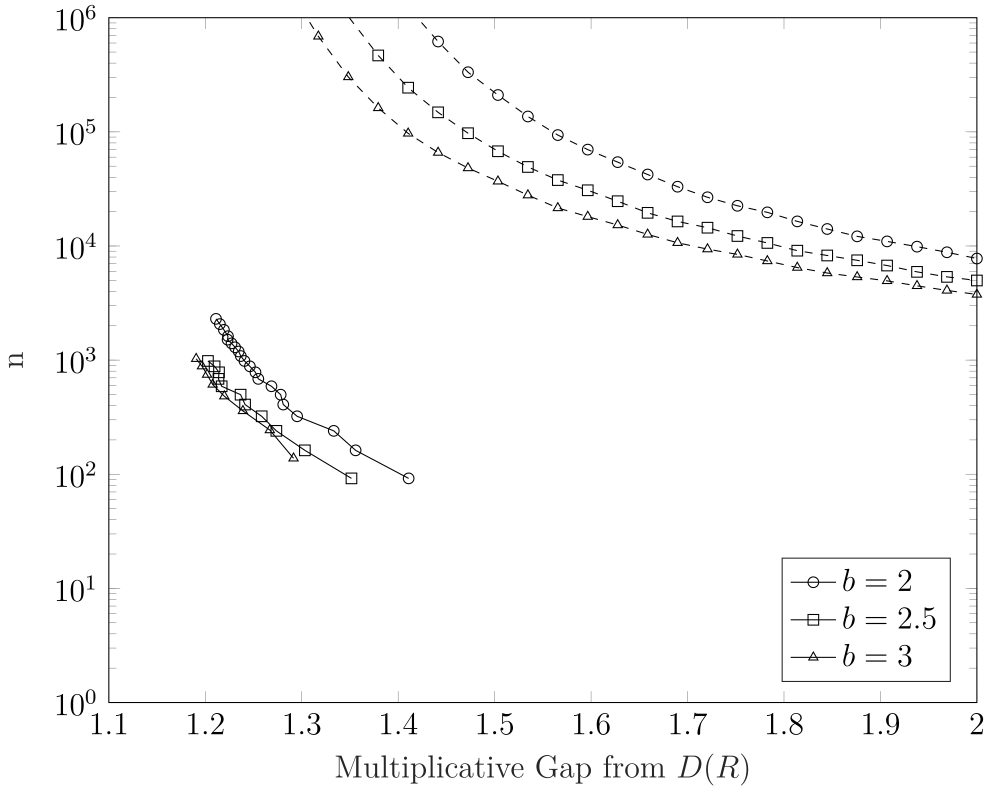

Figure 7.2: Comparison of the minimum blocklength n as function of the multiplicative gap from the Rate-Distortion, with $R = 0.5$, $\sigma^2 = 1$ and $\epsilon = 0.01$. The solid lines are from the simulation, the dotted lines are derived from our bound.

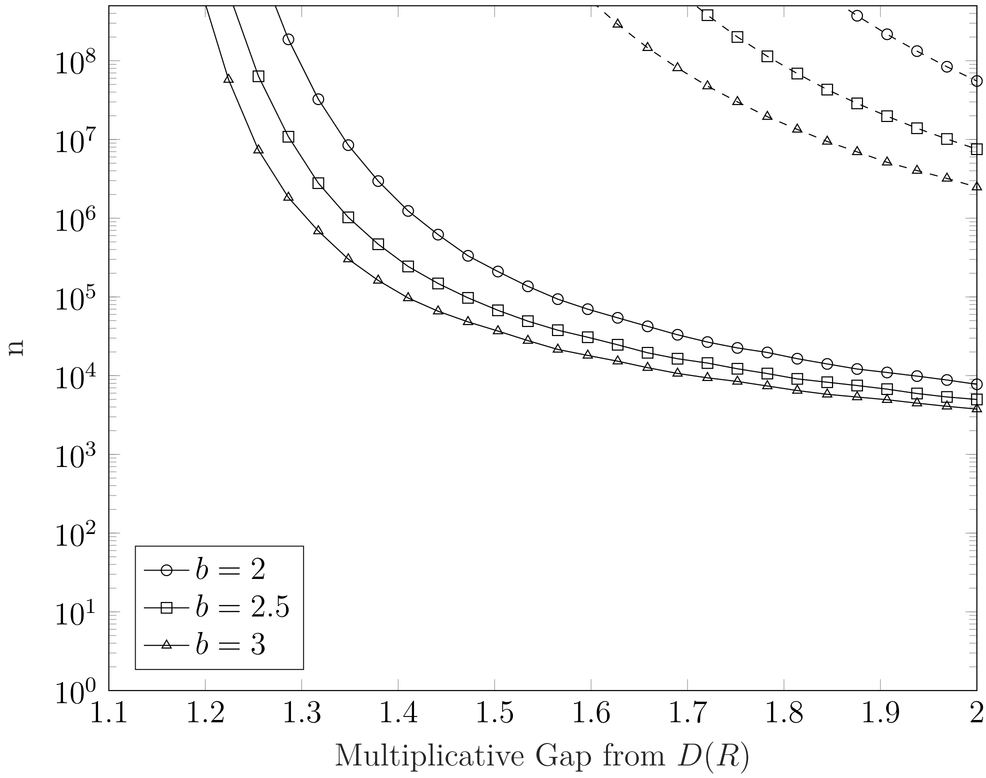

Figure 7.3: Comparison of the minimum blocklength n as function of the multiplicative gap from the Rate-Distortion, with $R = 0.5$, $\sigma^2 = 1$ and $\epsilon = 0.01$. The solid lines are derived from our bound, the dotted lines are derived from the bound in [26].

Chapter 8

Research Summary and Open Questions

8.1 Research Results

The main goal of this thesis was to develop bounds to the probability of excess distortion of SPARCs for finite-length lossy compression with sparse regressing codes, and thereby to find the tradeoff between the complexity of the codes and their performance. During the process toward this goal, several contributions were made.

In Chapter 2, new tight upper and lower bounds on the Lambert W function are derived, which constitute a major improvement to the existing lower and upper bounds. The Lambert W function was then used in the estimate of expected value of the maximum among standard i.i.d. Gaussian random variables (denoted by e_M).

In Chapter 3, a modified version of the SPARC encoding algorithm is presented, on the basis of e_M rather than its approximated value (which is an upper bound on e_M). The modified algorithm improves the performance of the SPARCs, as it is supported by our computer simulations, especially for the higher rates. Subsequently, it enables us to derive improved bounds on the probability of excess distortion.

After providing a preliminary analysis in Chapter 4, our main result is introduced and proved in Chapter 5. This gives an upper bound on the probability of excess distortion for lossy compression of an ergodic source using the modified SPARC encoding algorithm from Chapter 3. In contrast to previous works, Theorem 1 is proved for finite blocklength n. The bound in Theorem 1 is expressed as the sum of three separate probabilities, one of which depends on the distribution of the source; the three probabilities are bounded individually in Chapter 5, and the complete bound for the Gaussian i.i.d. case is provided in Theorem 2.

Relying on Theorem 1, Chapter 6 tackles the rate in which the distortion of our SPARC lossy compression scheme approaches the distortion-rate function, as the blocklength n grows. An asymptotic upper bound on this rate is proved, which scales in a similar fashion to previously known results, albeit with smaller coefficients; however, we also show that no better asymptotic bound can be developed, as long as we rely on Theorem 1.

8.2 Open Questions

Several subjects which were discussed in this research are left open for further research. The principal issue is the tightness of the bound in Theorem 1: while it is a significant improvement to the previously known bound in [26], the computer simulations show that there still is a non-negligible gap between the bound and the empirical performance of the SPARCs. The reason for this disparity revolves around the method in which the probability of excess distortion is bounded. As explained thoroughly in Chapter 3, the lossy compression encoding is performed in L steps, where at each step $i \in \{1, \ldots, L\}$, a column is picked in section i of the design matrix $\mathbf{A}$. In order to prove Theorem 1, the absolute value of the deviation Δ_i (see definition in (5.9)) is bounded, and the bound allows for the "worst case scenario": the absolute value of the deviation from the previous step, $|\Delta_{i-1}|$, is added to the absolute value of the deviation created in the current step. In practice, it is very unlikely that all L deviations have the same sign, which justifies the gap between the bound and the empirical results. To reach a tight bound to the probability of excess distortion, we must take into account the sign of the individual deviation at each step.

A related open question is the asymptotic scaling of the gap between the distortion-rate function and the performance of the SPARCs. In Chapter 6, for the case of an i.i.d. Gaussian source, it is shown that the gap tends to 0 in n at least as fast as $O\left(\frac{\log\log n}{\log n}\right)$; for a general memoryless source, the rate of the decay depends on the distribution of the source. However, it is still not clear whether the gap indeed decays to 0 at this rate or in a (much) faster rate. We have only shown that no bound can be developed which is asymptotically better while using Theorem 1.

Another possible future research direction is to develop a generalization to the lossy compression scheme of the SPARCs. The main theorem of the thesis, Theorem 1, is applicable to all ergodic and memoryless sources, not only a memoryless Gaussian source. However, it proves that all memoryless and ergodic sources approach the distortion-rate function of a memoryless Gaussian source when compressed by SPARCs, rather than their corresponding distortion-rate functions; as per [23], this is the best possible result that can be reached using a Gaussian codebook. An interesting open question is whether it is possible to apply the sparse regression method to other sources, even finite alphabet sources, and reach their respective asymptotic informational-theoretic limits.

From a more practical perspective, the aim of any code for lossy compression is not only to have good asymptotic performance, but also to work well for finite, reasonable blocklengths. Using the proposed algorithm in Chapter 3, the compression complexity increases polynomially in n. Although this growth rate in complexity is reasonable, in practice, it is not feasible to use the encoding algorithm with blocklengths exceeding several hundred source symbols. One improvement is to use a Hadamard-based design matrix $\mathbf{A}$ instead of the structureless i.i.d. standard Gaussian design matrix from Chapter 3, as suggested in [18]; this significantly decreases the complexity of the encoder, since it is possible to use the fast Hadamard transform. However, no theoretical guarantees for SPARCs with Hadamard-based design matrices have been developed so far (to the best of our knowledge), which could be an interesting direction for further research.

A different path to achieve better performance by SPARCs could be obtained by spatially coupled SPARCs (SC-SPARCs), where the design matrix $\mathbf{A}$ contains coupling between blocks. Recent results in [20] and [21] indicate promising empirical performance of the SC-SPARCs for channel coding, although they have not yet been adapted for lossy compression, and no finite blocklength analysis has been performed for these modern coding techniques.

Appendix A

Proof of Lemma 5

We first show that if (5.20) holds for some $i \in \{1, \ldots, L\}$, then $\Delta_i > -\frac{1}{2}$. Since $L > 2R$, from (2.15),

$$n = \frac{L}{R} \log M > 2 \log M \geq e_M^2, \tag{A.1}$$

$$\Rightarrow \frac{2e_M^2}{n - e_M^2} > 0. \tag{A.2}$$

Furthermore, from (3.1),

$$\frac{2e_M^2 L}{n - e_M^2} \leq \frac{4L \log M}{n - 2\log M} = \frac{4bL \log L}{\frac{b}{R} L \log L - 2b \log L} = \left(\frac{4\frac{L}{R}}{\frac{L}{R} - 2} \right) R. \tag{A.3}$$

Let R be fixed. Then the right side of (A.3) is a function of the form $f(x) = \frac{dx}{x-2}$ of the free parameter $x = \frac{L}{R}$. Since f is a monotonically decreasing function on (a, ∞) when $d > 0$, under the assumption that $\frac{L}{R} \geq 10$, the maximum of the right side of (A.3) is attained at $\frac{L}{R} = 10$. Substituting this value in the right side of (A.3) implies that

$$\frac{2e_M^2 L}{n - e_M^2} \leq 5R. \tag{A.4}$$

We conclude from (5.20), (A.2) and (A.4) that conditioning on $\mathcal{A}$, for $i \in \{1 \ldots, L\}$,

$$\Delta_i \geq \Delta_0 - \frac{2e_M^2 L}{n - e_M^2} \sum_{j=1}^{i} \frac{|\gamma_j| + |\epsilon_j|}{L} \tag{A.5}$$

$$\geq -\delta_0 - 5R(\delta_1 + \delta_2) \tag{A.6}$$

$$> -\frac{1}{2}, \tag{A.7}$$

where (A.7) holds due to the assumption in (5.1).

We next prove (5.20) by induction. For $i = 0$, it is trivial. Assume that (5.20) holds for some $i - 1$, and prove for $i \leq L$. The induction hypothesis together with (A.5)–(A.7) imply that

$$1+\Delta_{i-1}>\frac{1}{2}, \tag{A.8}$$

and therefore, using (5.15) and (A.2) yields

$$(1+\Delta_i)^2 \geq (1+\Delta_{i-1})^2 - \frac{e_M^2}{n-e_M^2}\Big(|\gamma_i| + 2|\epsilon_i|(1+\Delta_{i-1})\Big) \tag{A.9}$$

$$= (1+\Delta_{i-1})^2\left[1-\frac{e_M^2}{n-e_M^2}\left(\frac{|\gamma_i|}{(1+\Delta_{i-1})^2}+\frac{2|\epsilon_i|}{1+\Delta_{i-1}}\right)\right]. \tag{A.10}$$

To prove that the expression in the second term in the right side of (A.10) is positive, notice that from (A.8),

$$\frac{e_M^2}{n-e_M^2}\left(\frac{|\gamma_i|}{(1+\Delta_{i-1})^2}+\frac{2|\epsilon_i|}{1+\Delta_{i-1}}\right) \leq \frac{4e_M^2}{n-e_M^2}\left(|\gamma_i|+|\epsilon_i|\right), \tag{A.11}$$

and from (5.17)–(5.18), (A.4) and (A.11),

$$\frac{e_M^2}{n-e_M^2}\left(\frac{|\gamma_i|}{(1+\Delta_{i-1})^2}+\frac{2|\epsilon_i|}{1+\Delta_{i-1}}\right) \leq 10R(\delta_1+\delta_2) < 1, \tag{A.12}$$

where the last inequality holds due to (5.1). This proves that the right side of (A.10) is positive, which allows taking the square root of both sides of the inequality,

$$1+\Delta_i \geq (1+\Delta_{i-1})\left[1-\frac{e_M^2}{n-e_M^2}\left(\frac{|\gamma_i|}{(1+\Delta_{i-1})^2}+\frac{2|\epsilon_i|}{1+\Delta_{i-1}}\right)\right]^{\frac{1}{2}} \tag{A.13}$$

$$\geq (1+\Delta_{i-1})\left[1-\frac{e_M^2}{n-e_M^2}\left(\frac{|\gamma_i|}{(1+\Delta_{i-1})^2}+\frac{2|\epsilon_i|}{1+\Delta_{i-1}}\right)\right] \tag{A.14}$$

$$= 1+\Delta_{i-1} - \frac{e_M^2}{n-e_M^2}\left(\frac{|\gamma_i|}{1+\Delta_{i-1}}+2|\epsilon_i|\right), \tag{A.15}$$

where (A.14) holds since $\sqrt{x} \geq x$ for all $x \in [0,1]$. Consequently,

$$\Delta_i \geq \Delta_{i-1} - \frac{e_M^2}{n-e_M^2}\left(\frac{|\gamma_i|}{1+\Delta_{i-1}}+2|\epsilon_i|\right) \tag{A.16}$$

$$\geq \Delta_{i-1} - \frac{2e_M^2\left(|\gamma_i|+|\epsilon_i|\right)}{n-e_M^2} \tag{A.17}$$

$$\geq \Delta_0 - \frac{2e_M^2 L}{n-e_M^2}\sum_{j=1}^{i}\frac{|\gamma_j|+|\epsilon_j|}{L}, \tag{A.18}$$

where (A.17) is due to (A.8), and (A.18) from the induction hypothesis, i.e. that (5.20) holds for some $i-1$.

Appendix B

Proof of Lemma 6

We prove the lemma by induction. For $i = 1$, by (5.15) and (A.2),

$$(1+\Delta_1)^2 = (1+\Delta_0)^2 + \frac{e_M^2}{n-e_M^2}\left(\Delta_0^2 + \gamma_1 - 2\epsilon_1(1+\Delta_0)\right) \tag{B.1}$$

$$\leq (1+|\Delta_0|)^2 + \frac{e_M^2}{n-e_M^2}\left(\Delta_0^2 + |\gamma_1| + 2|\epsilon_1|\left(1+|\Delta_0|\right)\right) \tag{B.2}$$

and therefore,

$$1+\Delta_1 \leq (1+|\Delta_0|)\left[1 + \frac{e_M^2}{n-e_M^2}\left(\frac{\Delta_0^2}{(1+|\Delta_0|)^2} + \frac{|\gamma_1|}{(1+|\Delta_0|)^2} + \frac{2|\epsilon_1|}{1+|\Delta_0|}\right)\right]^{\frac{1}{2}} \tag{B.3}$$

$$\leq (1+|\Delta_0|)\left[1 + \frac{e_M^2}{2(n-e_M^2)}\left(\frac{\Delta_0^2}{(1+|\Delta_0|)^2} + \frac{|\gamma_1|}{(1+|\Delta_0|)^2} + \frac{2|\epsilon_1|}{1+|\Delta_0|}\right)\right], \tag{B.4}$$

where (B.4) relies of the inequality $\sqrt{1+x} \leq 1 + \frac{x}{2}$ which holds for $x \geq 0$. Consequently,

$$\Delta_1 \leq |\Delta_0| + \frac{e_M^2}{2(n-e_M^2)}\left(\frac{\Delta_0^2}{1+|\Delta_0|} + \frac{|\gamma_1|}{1+|\Delta_0|} + 2|\epsilon_1|\right) \tag{B.5}$$

$$\leq |\Delta_0| + \frac{e_M^2}{2(n-e_M^2)}\left(|\Delta_0| + 2|\gamma_1| + 2|\epsilon_1|\right) \tag{B.6}$$

$$\leq |\Delta_0|\left(1 + \frac{e_M^2}{2(n-e_M^2)}\right) + \frac{e_M^2}{n-e_M^2}\left(|\gamma_1| + |\epsilon_1|\right). \tag{B.7}$$

From Lemma 5,

$$\Delta_1 \geq \Delta_0 - \frac{2e_M^2}{n-e_M^2}\left(|\gamma_1| + |\epsilon_1|\right) \geq -|\Delta_0| - \frac{2e_M^2}{n-e_M^2}\left(|\gamma_1| + |\epsilon_1|\right). \tag{B.8}$$

Combining (B.7) and (B.8) yields

$$|\Delta_1| \leq |\Delta_0|\left(1 + \frac{e_M^2}{2(n-e_M^2)}\right) + \frac{2e_M^2}{n-e_M^2}\left(|\gamma_1| + |\epsilon_1|\right) \tag{B.9}$$

$$= |\Delta_0| w + \frac{2e_M^2}{n-e_M^2}\left(|\gamma_1| + |\epsilon_1|\right), \tag{B.10}$$

where the inequality in (B.10) holds by the definition in (5.3). We next assume that (5.21) holds for some $i-1$, and prove for $i \le L$. From (5.15), and using identical arguments to (B.1)–(B.7), it follows that

$$\Delta_i \le |\Delta_{i-1}| \left(1 + \frac{e_M^2}{2(n-e_M^2)}\right) + \frac{e_M^2}{n-e_M^2} \left(|\gamma_i| + |\epsilon_i|\right). \tag{B.11}$$

From (A.17),

$$\Delta_i \ge \Delta_{i-1} - \frac{2e_M^2}{n-e_M^2} \left(|\gamma_i| + |\epsilon_i|\right) \ge -|\Delta_{i-1}| - \frac{2e_M^2}{n-e_M^2} \left(|\gamma_i| + |\epsilon_i|\right), \tag{B.12}$$

and combining (B.11) and (B.12) yields

$$|\Delta_i| \le |\Delta_{i-1}| w + \frac{2e_M^2}{n-e_M^2} \left(|\gamma_i| + |\epsilon_i|\right). \tag{B.13}$$

with w as defined in (5.3). Using the induction hypothesis,

$$|\Delta_i| \le \left[|\Delta_0| w^{i-1} + \frac{2e_M^2}{n-e_M^2} \sum_{j=1}^{i-1} w^{i-j-1} (|\gamma_j| + |\epsilon_j|) \right] w + \frac{2e_M^2}{n-e_M^2} \left(|\gamma_i| + |\epsilon_i|\right) \tag{B.14}$$

$$= |\Delta_0| w^i + \frac{2e_M^2}{n-e_M^2} \sum_{j=1}^{i} w^{i-j} (|\gamma_j| + |\epsilon_j|). \tag{B.15}$$

Appendix C

Proof of Lemma 12

$$\int_{-\infty}^{\infty} e^{e_M|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z = \int_{e_M}^{\infty} e^{e_M(z-e_M)} f_{Z_M}(z)\,\mathrm{d}z + \int_{-\infty}^{e_M} e^{-e_M(z-e_M)} f_{Z_M}(z)\,\mathrm{d}z. \tag{C.1}$$

For the first integral in the right side of (C.1), from (2.13),

$$\int_{e_M}^{\infty} e^{e_M(z-e_M)} f_{Z_M}(z)\,\mathrm{d}z = \frac{M\,e^{-e_M^2}}{\sqrt{2\pi}} \int_{e_M}^{\infty} e^{e_M z - \frac{1}{2}z^2 + (M-1)\log(1-Q(z))}\,\mathrm{d}z. \tag{C.2}$$

Since $\log(1-Q(z)) \leq 0$ for all $z \in \mathbb{R}$,

$$\begin{aligned}
\int_{e_M}^{\infty} e^{e_M(z-e_M)} f_{Z_M}(z)\,\mathrm{d}z &\leq \frac{M\,e^{-e_M^2}}{\sqrt{2\pi}} \int_{e_M}^{\infty} e^{e_M z - \frac{1}{2}z^2}\,\mathrm{d}z && \text{(C.3)}\\
&= \frac{M\,e^{-e_M^2}}{\sqrt{2\pi}} \int_{e_M}^{\infty} e^{-\frac{1}{2}(z-e_M)^2 + \frac{1}{2}e_M^2}\,\mathrm{d}z && \text{(C.4)}\\
&= \frac{M\,e^{-\frac{1}{2}e_M^2}}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{1}{2}u^2}\,\mathrm{d}u && \text{(C.5)}\\
&= \frac{1}{2} M\,e^{-\frac{1}{2}e_M^2}. && \text{(C.6)}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\limsup_{M\to\infty} \frac{1}{e_M} \int_{e_M}^{\infty} e^{e_M(z-e_M)} f_{Z_M}(z)\,\mathrm{d}z &\leq \limsup_{M\to\infty} \frac{e^{\log M - \frac{1}{2}e_M^2}}{2e_M} && \text{(C.7)}\\
&= \frac{1}{2\sqrt{2}}, && \text{(C.8)}
\end{aligned}$$

where in (C.8) we used the following limit,

$$\lim_{M\to\infty} \frac{\log(M) - \frac{1}{2}e_M^2}{\log\log(M)} = \frac{1}{2}, \tag{C.9}$$

which follows from (2.73). For the second integral in the right side of (C.1), from (2.13),

$$\int_{-\infty}^{e_M} e^{-e_M(z-e_M)} f_{Z_M}(z)\,\mathrm{d}z = \int_{-\infty}^{e_M} \frac{M\,e^{e_M^2}}{\sqrt{2\pi}}\, e^{-e_M z-\frac{1}{2}z^2+(M-1)\log(1-Q(z))}\,\mathrm{d}z \tag{C.10}$$

$$\leq \int_{-\infty}^{0} g_M(z)\,\mathrm{d}z + \int_{0}^{e_M-\frac{1}{2}} g_M(z)\,\mathrm{d}z + \int_{e_M-\frac{1}{2}}^{e_M} g_M(z)\,\mathrm{d}z, \tag{C.11}$$

where

$$g_M(z) \triangleq \frac{M}{\sqrt{2\pi}} \exp\left(e_M^2 - e_M z - \frac{1}{2}z^2 + (M-1)\log\left(1-Q(z)\right)\right). \tag{C.12}$$

We now bound each of the three integrals in (C.11) individually. For the first,

$$\frac{M\,e^{e_M^2}}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-e_M z-\frac{1}{2}z^2+(M-1)\log(1-Q(z))}\,\mathrm{d}z \leq \frac{M^3}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-e_M z-\frac{1}{2}z^2-(M-1)\log(2)}\,\mathrm{d}z \tag{C.13}$$

$$= \frac{M^3}{\sqrt{2\pi}} e^{-(M-1)\log(2)} \int_{0}^{\infty} e^{e_M z-\frac{1}{2}z^2}\,\mathrm{d}z \tag{C.14}$$

$$= \frac{M^3}{\sqrt{2\pi}} e^{-(M-1)\log(2)} \int_{0}^{\infty} e^{-\frac{1}{2}(z-e_M)^2+\frac{1}{2}e_M^2}\,\mathrm{d}z \tag{C.15}$$

$$= \frac{M^3}{\sqrt{2\pi}} e^{-(M-1)\log(2)+\frac{1}{2}e_M^2} \int_{-e_M}^{\infty} e^{-\frac{1}{2}u^2}\,\mathrm{d}u \tag{C.16}$$

$$\leq M^3\, e^{-(M-1)\log(2)+\frac{1}{2}e_M^2}, \tag{C.17}$$

where (C.13) is due to $Q(x) \geq \frac{1}{2}$ for $x \leq 0$ and due to (2.15), and (C.17) follows from $Q(x) \leq 1$ for $x \in \mathbb{R}$. From (C.17),

$$\limsup_{M\to\infty} \int_{-\infty}^{0} g_M(z)\,\mathrm{d}z \leq \limsup_{M\to\infty} e^{-(M-1)\log(2)+\frac{1}{2}e_M^2+3\log M} = 0. \tag{C.18}$$

The second integral in (C.11) is similarly bounded as follows,

$$\frac{M\,e^{e_M^2}}{\sqrt{2\pi}} \int_{0}^{e_M-\frac{1}{2}} e^{-e_M z-\frac{1}{2}z^2+(M-1)\log(1-Q(z))}\,\mathrm{d}z$$

$$\leq \frac{M^3}{\sqrt{2\pi}} e^{(M-1)\log\left(1-Q\left(e_M-\frac{1}{2}\right)\right)} \int_{0}^{e_M-\frac{1}{2}} e^{-e_M z-\frac{1}{2}z^2}\,\mathrm{d}z \tag{C.19}$$

$$= \frac{M^3}{\sqrt{2\pi}} e^{(M-1)\log\left(1-Q\left(e_M-\frac{1}{2}\right)\right)+\frac{1}{2}e_M^2} \int_{0}^{e_M-\frac{1}{2}} e^{-\frac{1}{2}(z+e_M)^2}\,\mathrm{d}z \tag{C.20}$$

$$= M^3\, e^{(M-1)\log\left(1-Q\left(e_M-\frac{1}{2}\right)\right)+\frac{1}{2}e_M^2} \left(Q(e_M) - Q\left(2e_M-\tfrac{1}{2}\right)\right) \tag{C.21}$$

$$\leq M^3\, e^{-(M-1)Q\left(e_M-\frac{1}{2}\right)+\frac{1}{2}e_M^2} Q(e_M) \tag{C.22}$$

$$\leq M^3\, e^{-(M-1)Q\left(e_M-\frac{1}{2}\right)}, \tag{C.23}$$

where (C.19) is due to (2.15) and since $Q(x)$ is monotonically decreasing for $x \in \mathbb{R}$, (C.22) follows from the inequality $\log(1-x) \leq -x$ for $x \in [0,1)$, and (C.23) follows from $Q(x) \leq e^{-\frac{1}{2}x^2}$.

From the known inequality $Q(x) \geq \frac{x}{x^2+1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$,

$$3\log(M) - (M-1)Q\left(e_M - \tfrac{1}{2}\right) \leq 3\log(M) - \frac{e_M - \frac{1}{2}}{\left(e_M - \frac{1}{2}\right)^2 + 1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(e_M - \frac{1}{2})^2 + \log(M-1)} \tag{C.24}$$

$$= 3\log(M) - \frac{e_M - \frac{1}{2}}{\left(e_M - \frac{1}{2}\right)^2 + 1} \cdot \frac{1}{\sqrt{2\pi}} e^{\log(M-1) - \frac{1}{2}e_M^2 + \frac{1}{2}e_M - \frac{1}{8}}. \tag{C.25}$$

From (2.15), (C.23) and (C.25),

$$\limsup_{M\to\infty} \int_0^{e_M - \frac{1}{2}} g_M(z)\, \mathrm{d}z \leq \limsup_{M\to\infty} \exp\left(3\log(M) - \frac{e_M - \frac{1}{2}}{\left(e_M - \frac{1}{2}\right)^2 + 1} \cdot \frac{1}{\sqrt{2\pi}} e^{\log(M-1) - \frac{1}{2}e_M^2 + \frac{1}{2}e_M - \frac{1}{8}}\right) \tag{C.26}$$

$$\leq \limsup_{M\to\infty} \exp\left(3e^{\log\log(M)} - \frac{e_M - \frac{1}{2}}{\left(e_M - \frac{1}{2}\right)^2 + 1} \cdot \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}e_M - \frac{1}{8}}\right) = 0, \tag{C.27}$$

since $e_M \approx \sqrt{2\log M}$ (see (2.113)). The third integral in (C.11) can be bounded as follows,

$$\frac{M\, e^{e_M^2}}{\sqrt{2\pi}} \int_{e_M - \frac{1}{2}}^{e_M} e^{-e_M z - \frac{1}{2}z^2 + (M-1)\log(1-Q(z))}\, \mathrm{d}z \leq \frac{M\, e^{e_M^2}}{2\sqrt{2\pi}} \max_{e_M - \frac{1}{2} \leq z \leq e_M} \left\{ e^{-e_M z - \frac{1}{2}z^2 + (M-1)\log(1-Q(z))} \right\} \tag{C.28}$$

$$\leq \frac{M\, e^{e_M^2}}{2\sqrt{2\pi}} \max_{e_M - \frac{1}{2} \leq z \leq e_M} \left\{ e^{-e_M z - \frac{1}{2}z^2 - (M-1)Q(z)} \right\} \tag{C.29}$$

where (C.29) follows from the inequality $\log(1-x) \leq -x$ for $x \in [0, 1)$. Define

$$f(x) \triangleq e_M x + \frac{x^2}{2} + (M-1)Q(x) \tag{C.30}$$

for $x \in [e_M - \frac{1}{2},\, e_M]$. The derivative of $f(x)$ is

$$f'(x) = e_M + x - \frac{(M-1)}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{C.31}$$

To find the minimum of $f(x)$ in the interval, we set the derivative to zero and define $\epsilon \triangleq e_M - x$, which yields the following implicit equation,

$$2e_M - \epsilon - \frac{M-1}{\sqrt{2\pi}} \exp\left(-\frac{(e_M - \epsilon)^2}{2}\right) = 0 \tag{C.32}$$

$$\Rightarrow \epsilon = e_M - \sqrt{2\log\left(\frac{M-1}{(2e_M - \epsilon)\sqrt{2\pi}}\right)}. \tag{C.33}$$

To make sure that there exists $\epsilon \in [0, \frac{1}{2}]$ such that the implicit equation in (C.33) has a solution, the following step proves that the expression on the right side of (C.33) is positive and arbitrarily small for a large enough M. On the one hand,

$$e_M - \sqrt{2\log\left(\frac{M-1}{(2e_M-\epsilon)\sqrt{2\pi}}\right)} \leq \sqrt{2\log M}\left(1 - \sqrt{\frac{\log(M-1)}{\log M} - \frac{\log\left((2e_M-\epsilon)\sqrt{2\pi}\right)}{\log M}}\right) \tag{C.34}$$

$$\leq \sqrt{2\log M}\left(1 - \frac{\log(M-1)}{\log M} + \frac{\log\left((2e_M-\epsilon)\sqrt{2\pi}\right)}{\log M}\right) \tag{C.35}$$

$$= \sqrt{\frac{2}{\log M}} \cdot \log\left(\frac{M}{M-1}\right) + \frac{\sqrt{2}\log\left((2e_M-\epsilon)\sqrt{2\pi}\right)}{\sqrt{\log M}}, \tag{C.36}$$

where (C.34) follows from (2.15) and (C.35) due to $\sqrt{x} \geq x$ for $x \in [0, 1]$. Therefore,

$$\limsup_{M\to\infty}\left(e_M - \sqrt{2\log\left(\frac{M-1}{(2e_M-\epsilon)\sqrt{2\pi}}\right)}\right) \leq 0. \tag{C.37}$$

On the other hand,

$$e_M - \sqrt{2\log\left(\frac{M-1}{(2e_M-\epsilon)\sqrt{2\pi}}\right)} \geq \sqrt{2\log M} - \frac{\log\log M + \log(4\pi)}{2\sqrt{2\log M}} - \sqrt{2\log\left(\frac{M-1}{(2e_M-\epsilon)\sqrt{2\pi}}\right)} \tag{C.38}$$

$$\geq \sqrt{2\log M}\left(1 - \frac{\log\log M + \log(4\pi)}{4\log M} - \sqrt{1 - \frac{\log\left((2e_M-\epsilon)\sqrt{2\pi}\right)}{\log M}}\right) \tag{C.39}$$

$$\geq \sqrt{2\log M}\left(\frac{\log\left((2e_M-\epsilon)\sqrt{2\pi}\right)}{2\log M} - \frac{\log\log M + \log(4\pi)}{4\log M}\right) \tag{C.40}$$

$$= \frac{1}{\sqrt{2\log M}}\left(\log\left((2e_M-\epsilon)\sqrt{2\pi}\right) - \log\left(\sqrt{4\pi\log M}\right)\right) \tag{C.41}$$

$$= \frac{1}{\sqrt{2\log M}}\log\left(\frac{2e_M-\epsilon}{\sqrt{2\log M}}\right), \tag{C.42}$$

where (C.38) follows from (2.114) for a large enough M, and (C.40) is due to $\sqrt{1-x} \leq 1 - \frac{x}{2}$ for $x \in [0, 1]$. From (2.113), (C.42) implies that for a large enough M,

$$e_M - \sqrt{2\log\left(\frac{M-1}{(2e_M-\epsilon)\sqrt{2\pi}}\right)} \geq 0. \tag{C.43}$$

Since for a large enough M the right side of (C.33) is almost constant for $\epsilon \in [0, \frac{1}{2}]$, the implicit equation in (C.33) has a solution for $\epsilon \in [0, \frac{1}{2}]$ according to the intermediate value theorem. Therefore, the minimum of $f(x)$ when $x \in [e_M - \frac{1}{2}, e_M]$ is at

$$x^* = \sqrt{2 \log \left(\frac{M-1}{(2e_M - \epsilon)\sqrt{2\pi}} \right)}, \tag{C.44}$$

for some $\epsilon \in [0, \frac{1}{2}]$. Substituting (C.44) into (C.29) we have

$$\begin{aligned}
\int_{e_M - \frac{1}{2}}^{e_M} g_M(z)\,\mathrm{d}z &\leq \frac{1}{\sqrt{8\pi}} \exp\left(\log M + e_M^2 - e_M \sqrt{2 \log \left(\frac{M-1}{(2e_M - \epsilon)\sqrt{2\pi}} \right)} - \log \left(\frac{M-1}{(2e_M - \epsilon)\sqrt{2\pi}} \right) \right) && \text{(C.45)} \\
&= \frac{M}{M-1} \left(e_M - \frac{\epsilon}{2} \right) \exp\left(e_M^2 - e_M \sqrt{2 \log \left(\frac{M-1}{(2e_M - \epsilon)\sqrt{2\pi}} \right)} \right) && \text{(C.46)} \\
&\leq \frac{M}{M-1} \left(e_M - \frac{\epsilon}{2} \right) \exp\left(\sqrt{2 \log M} \left(e_M - \sqrt{2 \log \left(\frac{M-1}{\sqrt{16\pi \log M}} \right)} \right) \right), && \text{(C.47)}
\end{aligned}$$

where (C.45) follows from $Q(x) \geq 0$ for $x \in \mathbb{R}$ and (C.47) from (2.15) and since $\epsilon \geq 0$. From (2.114), for large enough M,

$$\int_{e_M - \frac{1}{2}}^{e_M} g_M(z)\,\mathrm{d}z \leq \frac{M}{M-1} \left(e_M - \frac{\epsilon}{2} \right) \exp\left(2 \log M - \frac{\log \log M - 2\gamma}{2} - 2\sqrt{\log M \log \left(\frac{M-1}{\sqrt{16\pi \log M}} \right)} \right). \tag{C.48}$$

Define

$$g(x) \triangleq 2x - \frac{\log x}{2} - 2\sqrt{x^2 - x \log(\sqrt{16\pi x})}. \tag{C.49}$$

Then,

$$\begin{aligned}
g'(x) &= 2 - \frac{1}{2x} - \left(x^2 - x \log(\sqrt{16\pi x}) \right)^{-0.5} \left(2x - \log(\sqrt{16\pi x}) - x \cdot \frac{1}{\sqrt{16\pi x}} \cdot \sqrt{\frac{4\pi}{x}} \right) && \text{(C.50)} \\
&= 2 - \frac{1}{2x} - \frac{2x - \log(\sqrt{16\pi x}) - \frac{1}{2}}{\sqrt{x^2 - x \log(\sqrt{16\pi x})}} && \text{(C.51)} \\
&= 2 - \frac{1}{2x} - \frac{2x - \log(\sqrt{16\pi x})}{\sqrt{x^2 - x \log(\sqrt{16\pi x})}} + \frac{1}{2\sqrt{x^2 - x \log(\sqrt{16\pi x})}} && \text{(C.52)} \\
&= 2 \left(1 - \sqrt{1 + \frac{\frac{1}{4} \log^2(\sqrt{16\pi x})}{x^2 - x \log(\sqrt{16\pi x})}} \right) + \frac{1}{2x} \left(\frac{1}{\sqrt{1 - \frac{\log(\sqrt{16\pi x})}{x}}} - 1 \right). && \text{(C.53)}
\end{aligned}$$

Since

$$\frac{1}{\sqrt{1-t}} - 1 \leq t, \qquad \text{for } t \in [0, \tfrac{\sqrt{5}-1}{2}], \tag{C.54}$$

$$1 - \sqrt{1+t} \leq -\tfrac{t}{4} \quad \text{for } t \in [0, 8], \tag{C.55}$$

then for a large enough x,

$$g'(x) \leq -\frac{\frac{1}{8}\log^2(\sqrt{16\pi x})}{x^2 - x\log(\sqrt{16\pi x})} + \frac{\log(\sqrt{16\pi x})}{2x^2} \tag{C.56}$$

$$\leq 0. \tag{C.57}$$

When $x \geq \frac{1}{16\pi}e^8 \approx 59.3$, the conditions in (C.54), (C.55) and (C.57) are held, and therefore

$$g'(x) \leq 0 \tag{C.58}$$

for $x > 60$. Thus, when $M \geq \exp(60)$ the exponent in (C.48) is monotonically decreasing with M, and subsequently for a large enough M,

$$\exp\left(2\log M - \frac{\log\log M - 2\gamma}{2} - 2\sqrt{\log M \log\left(\frac{M-1}{\sqrt{16\pi \log M}}\right)}\right) < 13.6, \tag{C.59}$$

and therefore,

$$\limsup_{M\to\infty} \frac{1}{e_M}\int_{e_M - \frac{1}{2}}^{e_M} g_M(z)\,\mathrm{d}z < 13.6. \tag{C.60}$$

Combining (C.1), (C.8), (C.11), (C.18), (C.27) and (C.60), we have (6.11).

Appendix D

Proof of Lemma 13

From (2.114),

$$\limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right)$$

$$\leq \limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{2\log M - \log\log M - \log 4\pi}{n}\right)^L - e^{-2R}\right) \tag{D.1}$$

$$= \limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{2R}{L}+\frac{\log\log L + \log 4\pi b}{\frac{b}{R}L\log L}\right)^L - e^{-2R}\right) \tag{D.2}$$

$$= \limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{2R}{L}\left(1-f(L)\right)\right)^L - e^{-2R}\right), \tag{D.3}$$

where (D.2) follows from (3.1) and with

$$f(L) \triangleq \frac{\log\log L + \log 4\pi b}{2b\log L}. \tag{D.4}$$

Continuing from (D.3),

$$\limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right)$$

$$\leq \limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\exp\left(L\log\left(1-\frac{2R}{L}\left(1-f(L)\right)\right)\right) - e^{-2R}\right) \tag{D.5}$$

$$\leq \limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(e^{2Rf(L)}-1\right)e^{-2R}, \tag{D.6}$$

where (D.6) follows from the inequality $\log(1-x) \leq -x$ for $x \in [0,1)$. Define $\xi \triangleq \frac{\log\log L}{\log L}$. Then,

$$\limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right) \leq \limsup_{\xi\to 0} \frac{1}{\xi}\left(e^{\frac{R}{b}(\xi+\epsilon)}-1\right)e^{-2R}, \tag{D.7}$$

for an arbitrarily small $\epsilon > 0$. Hence, from L'Hôpital's rule,

$$\limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right) \leq \limsup_{\xi\to 0} \frac{R}{b} e^{\frac{R}{b}(\xi+\epsilon)-2R} \tag{D.8}$$

$$= \frac{R}{b} e^{-2R+R\epsilon/b}. \tag{D.9}$$

Since the result in (D.9) is true for an arbitrarily small $\epsilon > 0$,

$$\limsup_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right) \leq \frac{R}{b} e^{-2R}. \tag{D.10}$$

On the other hand, from (2.114),

$$\liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right) \geq \liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{2R}{L}(1-g(L))\right)^L - e^{-2R}\right), \tag{D.11}$$

where

$$g(L) \triangleq \frac{\log\log L + \log b - 2\gamma}{2b\log L} - \frac{1}{b\log L}\left(\frac{\log\log L + \log b - 2\gamma}{2\sqrt{2b\log L}}\right)^2. \tag{D.12}$$

Continuing from (D.11),

$$\liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right)$$

$$\geq \liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\exp\left(L\log\left(1-\frac{2R}{L}(1-g(L))\right)\right) - e^{-2R}\right) \tag{D.13}$$

$$\geq \liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\exp\left(-2R(1-g(L)) - \frac{4R^2}{L}(1-g(L))^2\right) - e^{-2R}\right) \tag{D.14}$$

$$= \liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\exp\left(-2R + 2Rg(L) - \frac{4R^2}{L}(1-g(L))^2\right) - e^{-2R}\right), \tag{D.15}$$

where in (D.14) we used the inequality $\log(1-x) \geq -x - x^2$ for $x \in [0, \frac{1}{2}]$. Define $\xi \triangleq \frac{\log\log L}{\log L}$. Then, for an arbitrarily small $\epsilon > 0$,

$$\liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right) \geq \liminf_{\xi\to 0} \frac{1}{\xi}\left(\exp\left(-2R + \frac{R}{b}\xi - \epsilon\right) - e^{-2R}\right) \tag{D.16}$$

$$= \frac{R}{b} e^{-2R-\epsilon}, \tag{D.17}$$

where (D.17) is due to L'Hôpital's rule; since ϵ is an arbitrarily small positive number,

$$\liminf_{L\to\infty} \frac{\log L}{\log\log L}\left(\left(1-\frac{e_M^2}{n}\right)^L - e^{-2R}\right) \geq \frac{R}{b}e^{-2R}, \tag{D.18}$$

Combining (D.10) and (D.18) yields (6.19).

Bibliography

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July-October 1948.

[2] C. Berrou and A. Glavieux, “Near optimum error correcting coding and decoding: turbo-codes,” *IEEE Trans. on Communications*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.

[3] T. Richardson and R. Urbanke, *Modern Coding Theory*, Cambridge University Press, 2008.

[4] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. on Information Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[5] S. Korada and R. Urbanke, “Polar codes are optimal for lossy source coding,” *IEEE Trans. on Information Theory*, vol. 56, no. 4, pp. 1751–1768, Apr. 2010.

[6] S. Kudekar, T. Richardson and R. Urbanke, “Spatially coupled ensembles universally achieve capacity under belief propagation,” *IEEE Trans. on Information Theory*, vol. 59, no. 12, pp. 7761–7813, Sep. 2013.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, second edition, John Wiley & Sons, 2006.

[8] J. Hamkins and K. Zeger, “Gaussian source coding with spherical codes,” *IEEE Trans. on Information Theory*, vol. 48, no. 11, pp. 2980–2989, Nov. 2002.

[9] R. Venkataramanan, S. Tatikonda and A. Barron, “Sparse Regression Codes,” *IEEE Information Theory Society Newsletter*, pp. 7–15, Dec. 2016.

[10] G. Böcherer, F. Steiner and P. Schulte, “Bandwidth efficient and rate-matched low-density parity-check coded modulation,” *IEEE Trans. on Communications*, vol. 63, no. 12, pp. 4651–4665, Oct. 2015.

[11] U. Erez and R. Zamir, “Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding,” *IEEE Trans. on Information Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[12] R. M. Corless, G. H. Gonnet, D.E.G. Hare, D. J. Jeffrey and D. E. Knuth, “On the Lambert W function,” *Advances in Computational Mathematics*, vol. 5, pp. 329–359, 1996.

[13] A. Joseph and A. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. on Information Theory*, vol. 58, no. 5, pp. 1891–1897, Feb. 2012.

[14] A. Joseph and A. Barron, "Fast sparse superposition codes have near exponential error probability for $R < \mathcal{C}$," *IEEE Trans. on Information Theory*, vol. 60, pp. 919–942, Feb. 2014.

[15] A. Barron and S. Cho, "High-rate sparse superposition codes with iteratively optimal estimates," *IEEE International Symposium on Information Theory Proceedings (ISIT) 2012*, pp. 120–124, July 2012.

[16] S. Cho and A. Barron, "Approximate iterative bayes optimal estimates for high-rate sparse superposition codes," in *Sixth Workshop on Information-Theoretic Methods in Science and Engineering*, 2013.

[17] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb. 2011.

[18] C. Rush, A. Grieg and R. Venkataramanan, "Capacity-achieving sparse regression codes via approximate message passing decoding," *IEEE Trans. on Information Theory*, vol. 63, no. 3, pp. 1476–1500, June 2017.

[19] C. Rush and R. Venkataramanan, "The error exponent of sparse regression codes with AMP decoding," *IEEE International Symposium on Information Theory (ISIT) 2017*, pp. 2478–2482, 2017.

[20] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity-achieving sparse superposition codes," *IEEE Trans. on Information Theory*, vol. 63, no. 8, pp. 4894–4927, 2017.

[21] K. Hsieh, C. Rush and R. Venkataramanan, "Spatially coupled sparse regression codes: design and state evolution analysis," ArXiv e-prints arXiv:1801.01796v2, Apr. 2018.

[22] A. Greig and R. Venkataramanan, "Techniques for improving the finite length performance of sparse regression codes," *IEEE Trans. on Communications*, vol. 66, no. 3, pp. 905–917, Mar. 2018.

[23] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. on Information Theory*, vol. 43, no. 1, pp. 38–47, Jan. 1997.

[24] R. Venkataramanan, A. Joseph and S. Tatikonda, "Lossy compression via sparse linear regression: performance under minimum-distance encoding," *IEEE Trans. on Information Theory*, vol. 60, no. 6, pp. 3254–3264, June 2014.

[25] R. Venkataramanan and S. Tatikonda, “The rate-distortion function and excess-distortion exponent of sparse regression codes with optimal encoding,” *IEEE Trans. on Information Theory*, vol. 63, no. 8, pp. 5228–5243, Aug. 2017.

[26] R. Venkataramanan, T. Sarkar and S. Tatikonda, “Lossy compression via sparse linear regression: computationally efficient encoding and decoding,” *IEEE Trans. on Information Theory*, vol. 63, no. 6, pp. 3265–3278, June 2014.

[27] S. Chatterjee, *Superconcentration and Related Topics,* Springer, 2014.

[28] P. Massart, “Exponential and information inequalities,” in *Concentration Inequalities and Model Selection, Lecture Notes in Math* vol. 1896, Berlin, Springer, 2007.

[29] H. Cramér, “Asymptotic properties of sampling distribution,” in *Mathematical Methods of Statistics,* Princeton University Press, 1946, pp. 370-376.

[30] A. Hoorfar and M. Hassani, “Inequalities on the Lambert W function and hyperpower function,” *J. Inequalities in Pure and Applied Mathematics*, vol. 9, no. 2, pp. 1–5, 2008.

בפרק 3 מפורט אלגוריתם הקידוד של ה-SPARCs אשר פותח בעבודה קודמת, ולאחריו מוצע שינוי באלגוריתם המשפר את ביצועי הקוד, בעיקר בקצבי קוד גבוהים. מודיפיקציה זאת לסכימת הקידוד מבוססת על אומדנים יותר מדויקים מבעבר.

בפרק 4 מובאת אנליזה אסימפטוטית לביצועי אלגוריתם הקידוד החדש של ה-SPARCs, המציגה ניתוח פשוט (אם כי לא ריגורוזי) לכך שקודים אלה מסוגלים, עבור אורך בלוק גדול מספיק, לדחוס מקורות חסרי זיכרון כלשהם לפונקציית הקצב-עיוות של מקור גאוסי חסר-זיכרון.

בפרק 5, התוצאה המרכזית בעבודת המחקר מוצגת ומוכחת. המשפט שאנו מוכיחים בפרק זה נותן חסם עליון על ההסתברות שיתקבל עיוות יתר בקידוד עם SPARCs. המשפט מחדש ביחס לתוצאות קודמות בכמה מישורים: ראשית, הוא מתבסס על אלגוריתם הקידוד החדש שפורט בפרק 3; שנית, ההוכחה מתייחסת לכניסות בעלות אורך סופי, ולא רק אסימפטוטי כפי שהיה קודם לכן; ולבסוף, החסמים על ההסתברות לעיוות-יתר הודקו באופן משמעותי. בנוסף, אנו מראים כי המשפט החדש חל גם על מקורות סטציונריים וחסרי זיכרון שאינם גאוסיים, ושה-SPARCs מסוגלים לדחוס בהצלחה כל מקור סטציונרי וחסר זיכרון עם שונות סופית, לקצב דחיסה המתאים לפונקציית הקצב-עיוות של מקור גאוסי חסר זיכרון עם אותה שונות.

בהסתמך על המשפט מפרק 5, בפרק 6 מוצג חסם עליון אסימפטוטי לפער בין העיוות כתוצאה מקידוד בעזרת SPARCs של מקור גאוסי חסר זיכרון, לבין פונקציית העיוות-קצב, כתלות באורך הבלוק. בנוסף, מוכח כי לא ניתן להגיע לחסם עליון טוב יותר (עד כדי קבוע) כל עוד משתמשים במשפט זה.

פרק 7 מכיל סימולציות מחשב של ביצועי אלגוריתם הקידוד החדש לדחיסה עם עיוות בעזרת SPARCs, בהשוואה לאלגוריתם הקיים. בנוסף, איכות החסם שלנו על ההסתברות לעיוות יתר נבחנת אל מול החסם הקיים הקודם ואל מול ביצועי קודי ה-SPARCs עצמם.

סיכום תוצאות המחקר ניתנות בקצרה בפרק 8, וכן מוצגים נושאים עתידיים להמשך מחקר.

תקציר

אחת מהמטרות העיקריות של תורת האינפורמציה ושל תורת הקידוד היא תכנון של קודים מעשיים לתקשורת אמינה מעל ערוצים רועשים ולדחיסת נתונים עם עיוות, המגיעים לגבולות האינפורמציוניים התאורטיים היסודיים עם סיבוכיות ברת ביצוע. החל משנות ה-90, מספר קודים מעשיים כאלה פותחו, כגון קודי טורבו, קודי LDPC וקודי קיטוב; יחד עם זאת, קודים אלה נחקרו בעיקר לערוצים ולמקורות עם אלפבית דיסקרטי.

לאחרונה, סוג חדש של קודים הוצע עבור תקשורת אמינה מעל ערוצים חסרי זיכרון, ועבור דחיסה עם עיוות של מקורות סטציונריים וחסרי זיכרון, עם אלפבית רציף. קודים אלה, הנקראים קודי רגרסיה דלילה (באנגלית: sparse regression codes ובקיצור SPARCs), מבוססים על טכניקת קידוד שבה מילות הקוד הן צירופים לינאריים של עמודות מטריצת התכן של הקוד. ה-SPARCs פותחו במקור לתקשורת מעל ערוץ רעש לבן גאוסי אדיטיבי (AWGN), והוכח שהם מגיעים אסימפטוטית לקיבול הערוץ. לאחר מכן, ה-SPARCs הותאמו לדחיסה עם עיוות, והראו כי הם מצליחים להשיג אסימפטוטית את פונקציית הקצב-עיוות של מקור גאוסי חסר-זיכרון עם סיבוכיות זמן שגדלה פולינומיאלית עם אורך הבלוק.

עבודת מחקר זאת מתמקדת בבחינת ביצועי ה-SPARCs עבור דחיסה עם עיוות של מקורות חסרי זיכרון, על ידי הידוק חסמים אסימפטוטיים קיימים על ההסתברות לעיוות יתר, ועל ידי התאמת חסמים אלה לאורכי בלוק סופיים. זאת ועוד, בעבודת המחקר מוצע שיפור לסכימת הקידוד של דחיסה עם עיוות בעזרת SPARCs, ושיפור זה נבחן הן תאורטית והן באמצעות סימולציות מחשב. נושא נוסף אשר נידון בעבודה הוא ה-tradeoff שבין ביצועי קודי SPARCs ובין הסיבוכיות שלהם בהקשר של דחיסה עם עיוות.

העבודה מאורגנת כך: בפרק 2, מוצגים סימונים ומשפטים בסיסיים בהם השתמשנו לאורך התזה. בנוסף, מוכחים חסם עליון וחסם תחתון חדשים לפונקציית W של Lambert, אשר הדוקים באופן משמעותי מהחסמים הקיימים, ואשר רלוונטיים לאנליזה בעבודת מחקר זאת.

המחקר נעשה בהנחיית פרופ׳ יגאל ששון בפקולטה
להנדסת חשמל

הכרת תודה

ברצוני להודות למנחה שלי פרופ׳ יגאל ששון, עבור
ההדרכה והתמיכה שהעניק לי במשך עבודת מחקר זאת.
לא היה באפשרותי להשלים את לימודיי ללא המסירות
והסבלנות שלו, ואני באמת ובתמים אסיר תודה על הזמן
הרב שהוא הקדיש לשם כך.

אני מודה לטכניון על התמיכה הכספית הנדיבה
בהשתלמות

אנליזת אורך-בלוק סופי לדחיסה עם עיוות בעזרת קודי רגרסיה דלילה

חיבור על מחקר לשם מילוי חלקי של הדרישות לקבלת התואר מגיסטר למדעים בהנדסת חשמל

גל לבני

הוגש לסנאט הטכניון – מכון טכנולוגי לישראל

אייר תשע״ט חיפה מאי 2019