Remote Source Coding and AWGN CEO Problems

Krishnan Eswaran

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2006-2
http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-2.html

January 20, 2006

Remote Source Coding and AWGN CEO Problems

by

Krishnan Eswaran

B.S. Cornell University 2003

A thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science, Plan II
in

Engineering - Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Michael Gastpar
Professor Kannan Ramchandran

Fall 2005

Remote Source Coding and AWGN CEO Problems

by

Krishnan Eswaran

Acknowledgements

I thank my advisors Michael Gastpar and Kannan Ramchandran for all their advice and feedback while I was working on this. I would also like to thank Aaron Wagner and Vinod Prabhakaran for discussions about converses to the CEO problem. In particular, Aaron Wagner's outer bound was instrumental in giving a non-trivial lower bound to the sum-rate-distortion function for the AWGN CEO problem. Anand Sarwate's comments inspired the "no binning" results presented in this thesis. Vinod Prabhakaran conveyed the results given in Section 4.4.2. Stimulating discussions with Paolo Minero, Bobak Nazer, and Stark Draper are also gratefully acknowledged.

There are others who helped make writing this thesis a lot easier. Ruth Gjerde went the extra mile to help with paperwork. Niels Hoven's thesis template saved me what would have been hours of formatting. Neelam Ihsanullah helped proofread various incarnations of the writing. Finally, to everyone else that helped make my life easier while I was working on this thesis, thanks for everything.

Contents

Contents **ii**

List of Figures **iv**

List of Tables **v**

1 Introduction **1**

1.1 Mean Squared Error and Estimation 4

1.2 Background, Definitions, and Notation 8

2 Remote Source Coding **13**

2.1 Remote Rate-Distortion Function with a Single Observation 14

2.2 Remote Rate-Distortion Function with Multiple Observations 17

2.3 Squared Error Distortion and Encoder Estimation 19

2.4 Examples 22

2.5 Discussion 23

3 AWGN CEO Problems **24**

3.1 Background and Notation 25

3.2 Lower Bound 27

3.3 Upper Bounds 29

3.3.1 Maximum Entropy Upper Bound 30

3.3.2 Rate Loss Upper Bound 32

3.3.3 Rate Loss Upper Bound vs. Maximum Entropy Upper Bound 34

3.4 Discussion 37

4 Scaling Laws **38**

4.1 Definitions and Notation 39

4.2 Remote Source Coding Problem 40

4.3 CEO Problem 42

4.4 No Binning 44

4.4.1 Quadratic AWGN CEO Problem 46

4.4.2 Binary Erasure 48

4.5 Discussion 50

5 Conclusion and Future Work **51**

Bibliography **53**

References 53

A Lower Bounds on MMSE **55**

B Unified Entropy Power Inequality **59**

B.1 Main Result 59

B.2 Definitions and Notation 60

B.3 Fisher Information Inequalities 62

B.4 Gaussian Smoothing 64

C Sufficient Statistics in Remote Source Coding **67**

C.1 Sufficient Statistics in Additive White Gaussian Noise 67

C.2 Remote Rate-Distortion Functions 67

D Rate Loss **69**

D.1 Definitions and Notation 70

D.2 General Rate Loss Expression 71

E Linear Algebra **74**

List of Figures

1.1 Classical Source Coding 1

1.2 Remote Source Coding 2

1.3 CEO Problem 3

1.4 Two Pulse Source. 7

2.1 Remote Source Coding with a Single Observation 13

2.2 Remote Source Coding with M Observations 14

2.3 Plot of bounds for Gaussian, Laplacian, and uniform sources 22

3.1 AWGN CEO problem 25

3.2 Plots for the Laplacian and uniform source in AWGN CEO problem. 31

3.3 When the curve crosses log(2), the rate loss approach provides a better bound than the maximum entropy bound for the BPSK source and large enough M. 34

3.4 The rate loss upper bound outperforms the maximum entropy upper bound for the BPSK source for certain distortions. 36

4.1 Scaling behavior for Laplacian source in AWGN CEO Problem . . . 44

4.2 Scaling behavior for Logistic source in AWGN CEO Problem 45

A.1 Example of a source that decays as s^{-1}. 56

B.1 Desired behavior of $s(t)$ 64

D.1 Remote Source Coding Problem 69

D.2 AWGN CEO Problem 70

List of Tables

4.1 Performance loss for a "no binning" coding strategy in the quadratic AWGN CEO problem. 48

A.1 Distributions and their differential entropies. 55

Chapter 1

Introduction

Claude Shannon introduced the problem of source coding with a fidelity criterion in his 1959 paper [1]. In this problem, one is interested in specifying an encoding rate for which one can represent a source with some fidelity. Fidelity is measured by a fidelity criterion or distortion function. The minimal encoding rate for which one can reconstruct a source with respect to a target distortion is called the rate-distortion function. Because the source does not need to be reconstructed perfectly, this has also been called the lossy source coding or data compression problem. For the setup shown in Figure 1.1, Shannon reduced the solution to an optimization problem. Since his work, rate-distortion theory has experienced significant development, much of which has been chronicled in the survey paper of Berger and Gibson [2].

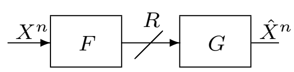

Figure 1.1. Classical Source Coding

Recognizing that computing the rate-distortion function can be a hard problem for a particular source and distortion function[1], Shannon computed upper and lower bounds to the rate-distortion function for difference distortions. Since then, Blahut [3], Arimoto [4], and Rose [5] have devised algorithms to help compute the rate-distortion function.

When one moves beyond the classical source coding problem, new issues arise. To motivate one such case, consider a monitoring or sensing system. In such a system, the encoder may not have direct access to the source of interest. Instead, only a corrupted or noisy version of the source is available to the encoder. This problem has been termed the remote source coding problem and was first studied by Dobrushin and Tsyabakov [6], who expressed the remote rate-distortion function in terms of an optimization problem. Wolf and Ziv considered a version of this problem shown in Figure 1.2, in which the source is corrupted by additive noise [7]. While the expression for the remote rate-distortion function can be reduced to the classical rate-distortion function with a modified distortion (see e.g. [8]), these expressions can be difficult to simplify into a closed form, and the bounds provided by Shannon for the classical problem [1] are not always applicable.

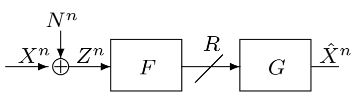

Figure 1.2. Remote Source Coding

In other cases, a general expression for the rate-distortion function or rate region is unknown. This has primarily been the case for distributed or multiterminal source coding problems. In these problems, there can be multiple encoders and/or decoders

[1] When Toby Berger asked Shizuo Kakutani for help with a homework problem as an undergraduate, Kakutani purportedly responded, “I know the general solution, but it doesn’t work in any particular case.”

that only have partial access to the sources of interest. General inner and outer bounds to such problems were first given by Berger [9], Tung [10], and Housewright [22]; an improved outer bound was later developed by Wagner and Anantharam [11], [12]. Unfortunately, these bounds are not computable.

One extension of the remote source coding problem to the distributed setting has been called the CEO problem, introduced in [13]. In the CEO problem, a chief executive officer (CEO) is interested in an underlying source. M agents observe independently corrupted observations of the source. Each has a noiseless, rate-constrained channel to the CEO. Without collaborating, the agents must send the CEO messages across these channels so that the CEO can reconstruct an estimate of the source to within some fidelity. The special case of Gaussian source and noise statistics with squared error distortion is called the quadratic Gaussian CEO problem and was introduced [14]. For this case, the rate region is known [15], [16]. When these assumptions no longer hold, not even a general expression for the rate region is not known.

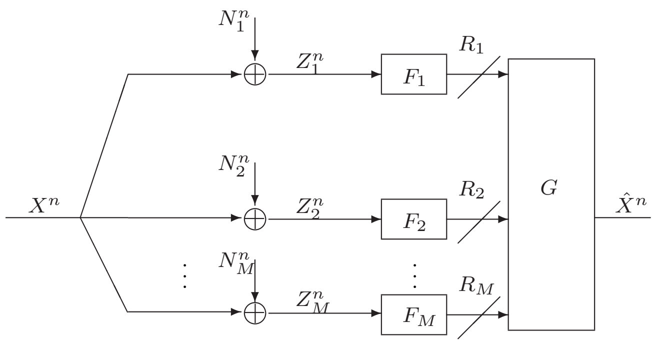

Figure 1.3. CEO Problem

Rather than taking an algorithmic approach to address these problems, we take an approach similar to the one considered by Shannon and derive closed form upper and lower bounds to the remote source coding problems previously described above. In

Chapter 2, we derive bounds to the remote rate-distortion function under an additive noise model. We apply these bounds to analyze the case of a mean squared error distortion and compare them with previously known results in this area. In Chapter 3, we derive bounds to the sum-rate-distortion function for CEO problem. We focus on the case of additive white Gaussian noise and mean squared error distortion, and we analyze an upper bound approach that relies on a connection between the gap of remote joint compression and remote distributed compression. In Chapter 4, we consider what happens as the number of observations gets large and present the scaling behavior of the rate-distortion functions for both the remote source coding problem as well as the CEO problem. We draw conclusions about these results in Chapter 5 and consider future research directions.

The remainder of this chapter establishes preliminaries that will be useful in interpreting the results found in subsequent chapters. In the next section, we consider the problem of minimum mean squared error (MMSE) estimation and its relationship to remote source coding problems with a mean squared error distortion. The remainder of the chapter establishes definitions and notation that are used in the rest of the thesis.

1.1 Mean Squared Error and Estimation

A distortion that we will pay particular attention to in this work is mean squared error. This also arises in problems in which one is trying to minimize the mean squared error between an underlying source and an estimator given noisy source observations. Note that such a problem is like the remote source coding problem with a mean squared error distortion, except that we no longer require that the estimate is a compressed version of the observations. Thus, the distortion obtained by any code in the remote source coding problem must be at least as large as the MMSE given the noisy observations. Indeed, this relationship exhibits itself in the bounds we derive, so we study the behavior of MMSE under additive noise models.

Example 1.1. Consider a Gaussian random variable $X \sim \mathcal{N}(0, \sigma_X^2)$ viewed through additive Gaussian noise $N \sim \mathcal{N}(0, \sigma_N^2)$ as $Z = X + N$. We assume X and N are independent. For this problem, the minimum mean squared estimate is

$$\hat{X} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2} Z,$$

and the corresponding minimum mean squared error is

$$\begin{aligned} E(X - \hat{X})^2 &= E\left(\frac{\sigma_N^2}{\sigma_X^2 + \sigma_N^2} X + \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2} N \right)^2 && (1.1) \\ &= \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2 + \sigma_N^2} && (1.2) \\ &= \sigma_X^2 \frac{1}{s+1}, && (1.3) \end{aligned}$$

where s is the signal-to-noise ratio, $s = \sigma_X^2 / \sigma_N^2$.

Since we have only used the second order statistics of X and N to calculate the mean squared error in this problem, we have the following well-known fact.

Corollary 1.2. *Let (X', N') be random variables with the same second order statistics as $(X, N) \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_N^2 \end{bmatrix}\right)$. Let $Z = X + N$ and define Z' similarly. Then*

$$E(X' - E[X'|Z'])^2 \leq E(X - E[X|Z])^2. \tag{1.4}$$

We have shown the above fact is true simply by using the linear estimator given in Example 1.1. This raises the possibility that a non-linear estimator can allow the MMSE to decay faster when the source and/or noise statistics are non-Gaussian. It turns out that for a large class of source distributions and noise distributions, the MMSE decays inversely with the signal-to-noise ratio (see Appendix A). However, the following counterexample shows that it is not always the case.

Example 1.3. Consider the same setup in Example 1.1, except now our source $X = \pm\sigma_X$, each with probability $\frac{1}{2}$. We call this the BPSK source. An exact expression for the minimum mean squared error (see e.g. [17]) is

$$\sigma_X{}^2 \left(1 - \int f_Z(z)(\tanh(s \cdot z/\sigma_X))^2 dz\right), \tag{1.5}$$

where $f_Z(z)$ is the probability density function of Z and $s = \sigma_X^2/\sigma_N^2$. The following bound shows that it scales exponentially with s. We will use the maximum likelihood estimator, which decides $+\sigma_X$ when Z is positive, and $-\sigma_X$ otherwise. Notice that the squared error will only be nonzero in the event that noise N is large enough to counteract the sign of X. Using this fact, along with the symmetry of the two errors, then for $s > 1$ we get that

$$E(X - E[X|Z])^2 \leq 4\sigma_X{}^2 P(N > \sigma_X) \tag{1.6}$$

$$< 4\sigma_X^2 \exp\{-s/2\}. \tag{1.7}$$

In fact, for $s \leq 1$, (1.7) continues to hold since in this range, the error of the linear estimator given in (1.3) is less than the right-hand side of (1.7).

While a considering the mean squared error for a discrete source might seem strange, we can show that we can get a quasi-exponential decay with a continuous source that sufficiently concentrates around its standard deviation.

Example 1.4. Now consider the case in which X has pdf

$$f_X(x) = \begin{cases} \frac{1}{4c\epsilon} & (1-\epsilon)c < |x| < (1+\epsilon)c \\ 0 & \text{otherwise} \end{cases}, \tag{1.8}$$

where $c = \frac{\sigma_X}{\sqrt{1+\epsilon^2/3}}$. We call this the two pulse source because its pdf has the shape shown in Figure 1.4. Using a similar ML approach to Example 1.3, we can show that

$$E(X - E[X|Z])^2 \leq \frac{\sigma_X^2}{1+\epsilon^2/3}\left(\epsilon^2 + e^{-\frac{(1-\epsilon)^2 s}{2(1+\epsilon^2/3)}}(2+\epsilon)^2\right) \tag{1.9}$$

Note that while this source does show a quasi-exponential behavior, it decays inversely with the signal-to-noise ratio asymptotically in s. Referring to Table A.1 and Theorem A.3 in Appendix A, we find that for the two pulse source,

$$E(X - E[X|Z])^2 \geq \frac{\frac{8\epsilon^2}{\pi e(2+\epsilon^2/3)}\sigma_X^2}{s+1}. \tag{1.10}$$

For a fixed s, however, we can make this bound arbitrarily small by choosing ϵ small enough while extending the quasi-exponential behavior shown in (1.9).

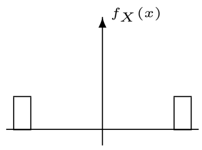

Figure 1.4. Two Pulse Source.

We can combine the upper bound as a consequence of Corollary 1.2 and the lower bound from Lemma A.2 to get the following result.

Corollary 1.5. *Let (X, N) be independent random variables with covariance matrix* $\begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_N^2 \end{bmatrix}$. *Then*

$$\frac{Q_X \cdot Q_N}{Q_{X+N}} \leq E(X - E[X|Z])^2 \leq \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2 + \sigma_N^2}, \tag{1.11}$$

where Q_X, Q_N, Q_{X+N} are the entropy powers of $X, N, X+N$ respectively. Here, we use entropy power of a random variable to refer the variance of a Gaussian random variable with the same differential entropy.

It will turn out that the gap between our upper and lower bounds for the remote rate-distortion function and the sum-rate-distortion function in the CEO problem will be related intimately to the gap between the upper and lower bounds to the MMSE in (1.11).

1.2 Background, Definitions, and Notation

To provide adequate background for the remainder of this work, this section introduces previous results, definitions, and notation that we will use in subsequent chapters. We use capital letters X, Y, Z to represent random variables, and the calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ to represent their corresponding sets. Vectors of length n are written as X^n, Y^n, Z^n. We denote a set of random variables as $\mathbf{H}_A = \{H_i, i \in A\}$ for some subset $A \subseteq \{1, \ldots, M\}$. Likewise, $\bar{H}_A = \frac{1}{|A|} \sum_{i \in A} H_i$. For convenience, we define $\mathbf{H} = \{H_i\}_{i=1}^M$ and $\bar{H}$ correspondingly.

To avoid getting mired in measure-theoretic notation, we present the following definitions for entropy and mutual information. While these definitions are not applicable to all cases presented in this work, the interested reader can find generally applicable definitions in Pinsker [18] and Wyner [19]. In our notation, nats are considered the standard unit of information, so all our logarithms are natural.

Definition 1.6. Given a random variable X with density $f(x)$, its differential entropy is defined as

$$H(X) = -\int f(x) \log f(x) dx. \tag{1.12}$$

Further, its entropy power is

$$Q_X = \frac{e^{2H(X)}}{2\pi e}. \tag{1.13}$$

Definition 1.7. The mutual information between two random variables X and Y with joint pdf $f(x, y)$ is

$$I(X; Y) = \int f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dxdy \tag{1.14}$$

$$= H(X) - H(X|Y), \tag{1.15}$$

where

$$H(X|Y) = -\int f(x, y) \log f(x|y) dx.$$

We now consider the classical source coding problem.

Definition 1.8. A distortion function d is a measurable map $d : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbf{R}_+$, where $\mathbf{R}_+$ is the set of positive reals.

Definition 1.9. A difference distortion is a distortion function $d : \mathbf{R} \times \mathbf{R} \to \mathbf{R}_+$ with the property $d(x, y) = \tilde{d}(x - y)$.

Definition 1.10. A direct code (n, N, Δ) is specified by an encoder function F and decoding function G such that

$$F : \mathcal{X}^n \to \mathcal{I}_N, \tag{1.16}$$

$$G : \mathcal{I}_N \to \mathcal{X}^n, \tag{1.17}$$

$$E\frac{1}{n}\sum_{k=1}^{n} d\left(X(k), \hat{X}(k)\right) = \Delta, \tag{1.18}$$

where $\mathcal{I}_N = \{1, \ldots, N\}$ and $\hat{X}^n = G(F(X^n))$.

Definition 1.11. A pair (R, D) is directly achievable if, for all $\epsilon > 0$ and sufficiently large n, there exists a direct code (n, N, Δ) such that

$$N \leq \exp\{n(R + \epsilon)\}, \tag{1.19}$$

$$\Delta \leq D + \epsilon. \tag{1.20}$$

Definition 1.12. The minimal rate R for a distortion D such that (R, D) is directly achievable is called the direct rate-distortion function, denoted $R_X(D)$.

The direct rate-distortion function is well known and, for an i.i.d. source, is characterized by the following single letter mutual information expression [20], [8].

$$R_X(D) = \min_{\substack{\hat{X} \text{ s.t.} \\ Ed(X,\hat{X}) \leq D}} I(X; \hat{X}), \tag{1.21}$$

Upper and lower bounds to the direct rate-distortion function are given by [8, p. 101]

$$\frac{1}{2}\log\left(\frac{Q_X}{D}\right) \leq R_X(D) \leq \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right), \tag{1.22}$$

From the direct rate-distortion problem, we move to the remote rate-distortion problem.

Definition 1.13. A remote code (n, N, Δ) is specified by an encoder function F_R and decoding function G_R such that

$$F_R : \mathcal{Z}_1^n \times \cdots \times \mathcal{Z}_M^n \to \mathcal{I}_N, \tag{1.23}$$

$$G_R : \mathcal{I}_N \to \mathcal{X}^n, \tag{1.24}$$

$$E\frac{1}{n}\sum_{k=1}^{n} d\left(X(k), \hat{X}_R(k)\right) = \Delta, \tag{1.25}$$

where $\hat{X}_R^n = G_R(F_R(Z_1^n, \ldots, Z_M^n))$.

Definition 1.14. A pair (R, D) is remotely achievable if, for all $\epsilon > 0$ and sufficiently large n, there exists a remote code (n, N, Δ) such that

$$N \leq \exp\{n(R + \epsilon)\}, \tag{1.26}$$

$$\Delta \leq D + \epsilon. \tag{1.27}$$

Definition 1.15. The minimal rate R for a distortion D such that (R, D) is remotely achievable is called the remote rate-distortion function, denoted $R_X^R(D)$.

The remote rate-distortion function is known and, for an i.i.d. source with i.i.d. observations, is characterized by the following single letter mutual information expression [8, p. 79].

$$R_X^R(D) = \min_{\hat{X}_R \in \hat{\mathcal{X}}_X^R(D)} I(Z_1, \ldots, Z_M; \hat{X}_R), \tag{1.28}$$

$$\hat{\mathcal{X}}_X^R(D) = \left\{\hat{X} : X \to Z_1, \ldots, Z_M \to \hat{X},\ E(X - f(\hat{X}))^2 \leq D,\ \text{for some } f.\right\}$$

Since a direct code could always corrupt its source according to the same statistics as the observations and then use a remote code, it should be clear that the remote rate distortion function is always at least as large as the direct rate distortion function, or $R_X^R(D) \geq R_X(D)$.

Definition 1.16. A CEO code $(n, N_1, \ldots, N_M, \Delta)$ is specified by M encoder functions $F_1, \ldots, F_M$ corresponding to the M agents and a decoder function G corresponding to the CEO

$$F_i : \mathcal{Z}_i^n \to \mathcal{I}_{N_i}, \tag{1.29}$$

$$G : \mathcal{I}_{N_1} \times \cdots \times \mathcal{I}_{N_M} \to \mathcal{X}^n, \tag{1.30}$$

where $\mathcal{I}_j = \{1, \ldots, j\}$. Such a code satisfies the condition

$$E\frac{1}{n}\sum_{k=1}^{n} d\left(X(k), \hat{X}(k)\right) = \Delta, \tag{1.31}$$

where $\hat{X}^n = G(F_1(Z_1^n), \ldots, F_M(Z_M^n))$.

Definition 1.17. A sum-rate distortion pair (R, D) is achievable if, for all $\epsilon > 0$ and sufficiently large n, there exists a CEO code $(n, N_1, \ldots, N_M, \Delta)$ such that

$$\begin{aligned} N_1 &\leq \exp\{n(R_1 + \epsilon)\} \\ N_2 &\leq \exp\{n(R_2 + \epsilon)\} \\ &\vdots \\ N_M &\leq \exp\{n(R_M + \epsilon)\}, \end{aligned} \tag{1.32}$$

$$\sum_{i=1}^{M} R_i = R, \tag{1.33}$$

$$\Delta \leq D + \epsilon. \tag{1.34}$$

Definition 1.18. We call the minimal sum-rate R for a distortion D such that (R, D) is achievable the sum-rate-distortion function, which we denote $R_X^{CEO}(D)$.

No single-letter characterization of the sum-rate-distortion function is known for the CEO problem. In Chapter 3, we will examine inner and outer bounds to this function.

The following definition will be useful when we consider the squared error distortion.

Definition 1.19. Let X be a random variable with variance σ_X^2. $\mathcal{T}_X$ is the set of functions $\mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ such that for all $t \in \mathcal{T}_X$,

$$E(X - E[X|X+V])^2 \leq t(s, \sigma_X^2), \tag{1.35}$$

where $V \sim \mathcal{N}(0, \sigma_X^2/s)$.

To show this set is not empty, we have the following lemma.

Lemma 1.20. *Define the function $t_l : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ as*

$$t_l(s, \sigma_X^2) = \frac{\sigma_X^2}{s+1}. \tag{1.36}$$

Then $t_l \in \mathcal{T}_X$.

Proof. The result follows immediately from Corollary 1.2 and Example 1.1. □

Chapter 2

Remote Source Coding

Although an expression for the remote rate-distortion function is given in (1.28), this function is difficult to evaluate in general. In this chapter, we derive upper and lower bounds for the remote rate-distortion function viewed in additive noise for the model in Figure 2.1, and for the case of additive Gaussian noise for the model in Figure 2.2.

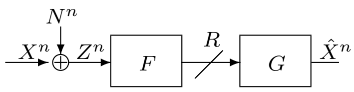

Figure 2.1. Remote Source Coding with a Single Observation

We start with the case of the single observation. For our analysis, we assume an additive noise model for an i.i.d. source process $\{X(k)\}_{k=1}^{\infty}$, an i.i.d. noise process $\{N(k)\}_{k=1}^{\infty}$, the observation process is described as

$$Z(k) = X(k) + N(k), \quad k \geq 1. \tag{2.1}$$

We then consider the case of multiple observations in which the noise is additive

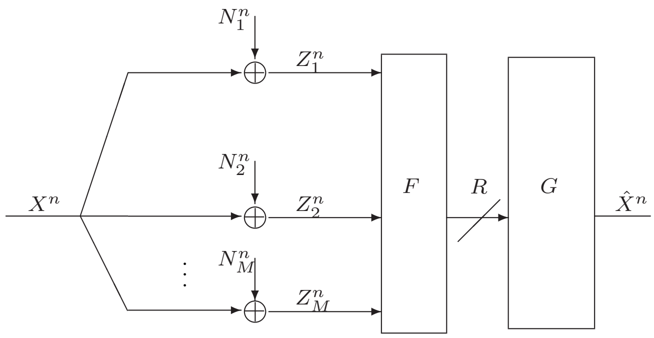

Figure 2.2. Remote Source Coding with M Observations

white Gaussian at each sensor and independent among sensors. For the M-observation model, we have, for $1 \leq i \leq M$,

$$Z_i(k) = X(k) + N_i(k), \quad k \geq 1, \tag{2.2}$$

where $N_i(k) \sim \mathcal{N}(0, \sigma_{N_i}^2)$.

We specialize these bounds for the case of mean-squared error and compare our results to previous work related to this case [7], [21].

2.1 Remote Rate-Distortion Function with a Single Observation

Recall the remote-rate distortion expression in (1.28). For the case of a single observation, this specializes to

$$R_X^R(D) = \min_{\hat{X} \in \hat{\mathcal{X}}^R(D)} I(Z; \hat{X}), \tag{2.3}$$

$$\hat{\mathcal{X}}^R(D) = \left\{ \hat{X} : X \to Z \to \hat{X}, Ed\Big(X, f(\hat{X})\Big) \leq D, \text{ for some } f. \right\}$$

For the case of a Gaussian source and noise statistics and a squared error distortion, the remote rate-distortion function $R^R_{X,\mathcal{N}}(D)$ is known to be [8]

$$R^R_{X,\mathcal{N}}(D) = \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right) + \frac{1}{2}\log\left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2 - \sigma_N^2\frac{\sigma_X^2}{D}}\right). \tag{2.4}$$

$$= R_{X,\mathcal{N}}(D) + \frac{1}{2}\log\left(\frac{\sigma_X^2}{\sigma_Z^2 - \sigma_N^2 e^{2R_{X,\mathcal{N}}(D)}}\right), \tag{2.5}$$

where $R_{X,\mathcal{N}}(D)$ is the direct rate-distortion function for a Gaussian source and mean-squared error distortion. The upper and lower bounds that we derive in this section will have a similar form to the remote rate-distortion function for Gaussian statistics and squared error distortion. In fact, the bounds will be tight for that case. We start by stating and proving the lower bound.

Theorem 2.1. *Consider the remote source coding problem with a single observation. Then, a lower bound for the remote rate-distortion function is*

$$R^R_X(D) \geq R_X(D) + \frac{1}{2}\log\left(\frac{Q_X}{Q_Z - Q_N e^{2R_X(D)}}\right). \tag{2.6}$$

Proof. The key tool involved is a new entropy power inequality (see Appendix B), which considers remote processing of data corrupted by additive noise. By Theorem B.1, we know that

$$e^{-2I(Z;\hat{X})} \leq \frac{e^{2(H(Z)-I(X;\hat{X}))} - e^{2H(N)}}{e^{2H(X)}}. \tag{2.7}$$

Simplifying this equation, we get that

$$I(Z;\hat{X}) \geq \frac{1}{2}\log\left(\frac{e^{2H(X)}}{e^{2(H(Z)-I(X;\hat{X}))} - e^{2H(N)}}\right) \tag{2.8}$$

$$\geq \frac{1}{2}\log\left(\frac{e^{2H(X)}}{e^{2(H(Z)-R_X(D))} - e^{2H(N)}}\right) \tag{2.9}$$

$$= R_X(D) + \frac{1}{2}\log\left(\frac{e^{2H(X)}}{e^{2H(Z)} - e^{2(H(N)+R_X(D))}}\right) \tag{2.10}$$

Since the above inequality is true for all choices of $\hat{X}$, we conclude that

$$R^R_X(D) \geq R_X(D) + \frac{1}{2}\log\left(\frac{e^{2H(X)}}{e^{2H(Z)} - e^{2(H(N)+R_X(D))}}\right). \tag{2.11}$$

Normalizing the numerator and denominator in the second term of (2.11) by $2\pi e$ gives the result. $\square$

Note that when the source and noise statistics are Gaussian and the distortion is squared error, the lower bound in Theorem 2.1 is tight. We will find that the upper bound described below also satisfies this property.

Theorem 2.2. *Let the source and observation process have fixed second order statistics. If a function* $t : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ *satisfies*

$$\min_f Ed\left(X, f(X + N + V)\right) \leq t(s, \sigma_X^2) \tag{2.12}$$

for $V \sim \mathcal{N}(0, \sigma_X^2/s - \sigma_N^2)$ *and all* $s < \frac{\sigma_X^2}{\sigma_N^2}$*, then*

$$R_X^R(D) \leq r + \frac{1}{2}\log\left(\frac{\sigma_X^2}{\sigma_Z^2 - \sigma_N^2 e^{2r}}\right) \tag{2.13}$$

where r *is the solution to* $D = t\left(e^{2r} - 1, \sigma_X^2\right)$.

Proof. Let $\hat{X} = Z + V$. Then, for $D = t(s, \sigma_X^2)$, $\hat{X} \in \hat{\mathcal{X}}^R(D)$, so

$$\begin{aligned}
R_X^R(D) &\leq I(Z; \hat{X}) & (2.14)\\
&= H(\hat{X}) - H(V) & (2.15)\\
&\leq \frac{1}{2}\log\left(2\pi e \sigma_X^2 \frac{1+s}{s}\right) - H(V) & (2.16)\\
&= \frac{1}{2}\log\left(\sigma_X^2 \frac{1+s}{\sigma_X^2 - s\sigma_N^2}\right) & (2.17)\\
&= \frac{1}{2}\log(1+s) + \frac{1}{2}\log\left(\frac{\sigma_X^2}{\sigma_X^2 - s\sigma_N^2}\right). & (2.18)
\end{aligned}$$

Letting $s = e^{2r} - 1$ completes the result. $\square$

For Gaussian source and noise statistics and a squared error distortion, the function f in the theorem is just the MMSE estimator. Then, for the set $\mathcal{T}_X$ given in Definition 1.19, any $t \in \mathcal{T}_X$ satisfies the condition (2.12), so we use the function given in Lemma 1.20 to get a tight result.

2.2 Remote Rate-Distortion Function with Multiple Observations

To handle upper and lower bounds, we restrict ourselves to cases in which the noise statistics are Gaussian. For this case, the minimal sufficient statistic is a scalar and can be represented as the source corrupted by independent additive noise. In fact, this requirement is all that is necessary to provide a lower bound in this problem. Of course, we can always give an upper bound, regardless of whether such a condition is satisfied.

Using Lemma C.2, we can return to the framework of the single observation problem as long as we can find an appropriate scalar sufficient statistic for X given $Z_1, \ldots, Z_M$. Lemma C.1 tells us that

$$\tilde{Z}(k) = \frac{1}{M} \sum_{i=1}^{M} \frac{\sigma_{\tilde{N}}^2}{\sigma_{N_i}^2} Z_i(k) \tag{2.19}$$

$$= X(k) + \frac{1}{M} \sum_{i=1}^{M} \frac{\sigma_{\tilde{N}}^2}{\sigma_{N_i}^2} N_i(k), \tag{2.20}$$

is a sufficient statistic for $X(k)$ given $Z_1(k), \ldots, Z_M(k)$ where $\sigma_{\tilde{N}}^2 = \frac{1}{\frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sigma_{N_i}^2}}$. From this, we can now use our single observation results to get upper and lower bounds for the M-observation case.

Theorem 2.3. *Consider the M-observation remote source coding problem with additive white Gaussian noise. Then, a lower bound for the remote rate-distortion function is*

$$R_X^R(D) \geq R_X(D) + \frac{1}{2} \log \left(\frac{M Q_X}{M Q_{\tilde{Z}} - \sigma_{\tilde{N}}^2 e^{2R_X(D)}} \right). \tag{2.21}$$

Proof. Lemma C.2 states that using a sufficient statistic for a source given its observations does not change the remote rate-distortion function. Thus, by Lemma C.1 and Theorem 2.1, we have (2.21). □

Theorem 2.4. *Consider the M-observation remote source coding problem with additive white Gaussian noise and a source with fixed second order statistics. If a function* $t : \mathbf{R}_+ \times \mathbf{R}_+ \to \mathbf{R}_+$ *satisfies*

$$\min_f Ed\left(X, f(X+V)\right) \le t(s, \sigma_X^2). \tag{2.22}$$

for $V \sim \mathcal{N}(0, \sigma_X^2/s)$ *and all* $s < \frac{M\sigma_X^2}{\sigma_{\tilde{N}}^2}$, *then*

$$R_X^R(D) \le r + \frac{1}{2}\log\left(\frac{M\sigma_X^2}{M\sigma_{\tilde{Z}}^2 - \sigma_{\tilde{N}}^2 e^{2r}}\right), \tag{2.23}$$

where r *is the solution to* $D = t\left(e^{2r} - 1, \sigma_X^2\right)$.

Proof. Simply form the sufficient statistic given in Lemma C.1 and apply Theorem 2.2. □

Again, Theorems 2.3 and 2.4 are tight for Gaussian statistics and squared error distortion. To see why, we rewrite the results in terms of entropy powers and variances in the following corollary. Tightness follows for the Gaussian case since the entropy power of a Gaussian is that same as its variance.

Corollary 2.5. *Consider the M-observation remote source coding problem with additive white Gaussian noise and a source with fixed second order statistics and squared error distortion* $d(x, \hat{x}) = (x - \hat{x})^2$. *Let* t *be a function in the set* $\mathcal{T}_X$ *given in Definition 1.19. Then a lower bound to the rate-distortion function is*

$$R_X^R(D) \ge \frac{1}{2}\log\left(\frac{Q_X}{D}\right) + \frac{1}{2}\log\left(\frac{MQ_X}{MQ_{\tilde{Z}} - \frac{Q_X}{D}\sigma_{\tilde{N}}^2}\right), \tag{2.24}$$

where $\sigma_{\tilde{N}}^2 = \frac{1}{\frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sigma_{N_i}^2}}$. *Further, an upper bound to the rate-distortion function is*

$$R_X^R(D) \le \frac{1}{2}\log\left(\frac{\sigma_X^2}{D_l}\right) + \frac{1}{2}\log\left(\frac{M\sigma_X^2}{M\sigma_{\tilde{Z}}^2 - \frac{\sigma_X^2}{D_l}\sigma_{\tilde{N}}^2}\right), \tag{2.25}$$

where D_l *is the solution to the equation* $D = t\left(\frac{\sigma_X^2}{D_l} - 1, \sigma_X^2\right)$.

Proof. Applying the direct rate-distortion lower bound in (1.22) to Theorem 2.3, we get (2.24) by noting that $Q_{\tilde{N}} = \sigma^2_{\tilde{N}}$. From Definition 1.19, we know that t satisfies (2.22), so applying Theorem 2.4 with $r = \frac{1}{2}\log\left(\frac{\sigma_X^2}{D_l}\right)$ gives (2.25). Note that Lemma 1.20 implies $D \leq D_l$. □

2.3 Squared Error Distortion and Encoder Estimation

Having established upper and lower bounds for the remote rate-distortion function, we now want to compare our results to previously known bounds. In this section, we present such a set of upper and lower bounds for the case of squared error distortion and additive white Gaussian noise. These bounds are based upon the arguments provided by Wolf and Ziv [7] for the single observation case and by Gastpar [21] for the multiple observation case . They amount to performing an MMSE estimate at the encoder and then using the optimal squared error codebook for the case in which the MMSE estimate is the source.

Theorem 2.6. *Define*

$$D_0 = E\left(X - E\left[X \,\middle|\, X + \frac{1}{M}\sum_{i=1}^{M}\frac{\sigma^2_{\tilde{N}}}{\sigma^2_{N_i}}N_i\right]\right)^2, \tag{2.26}$$

where $\sigma^2_{\tilde{N}} = \frac{1}{\frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sigma^2_{N_i}}}$. *Then for the M-observation remote source coding problem, upper and lower bounds to the rate-distortion function are [7], [21]*

$$R_X^R(D) \leq \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right) + \frac{1}{2}\log\left(\frac{1-\frac{D_0}{\sigma_X^2}}{1-\frac{D_0}{D}}\right), \tag{2.27}$$

$$R_X^R(D) \geq \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right) + \frac{1}{2}\log\left(\frac{1-\frac{D_0}{\sigma_X^2}}{1-\frac{D_0}{D}}\right) - \log\left(\frac{\sigma_V^2}{Q_V}\right). \tag{2.28}$$

Proof. The reasoning comes from the fact that we can modify the distortion (see [8] for a detailed discussion) for $d(x, \hat{x}) = (x - \hat{x})^2$ in terms of the observations to get

$$\begin{aligned}\tilde{d}(\tilde{z}, \hat{x}) =& E[(X - \hat{x})^2 | \tilde{Z} = \tilde{z}] \\ =& E[(X - E[X|\tilde{Z}] + E[X|\tilde{Z}] - \bar{x})^2 | \tilde{Z} = \tilde{z}] \\ =& E[(X - E[X|\tilde{Z}])^2 | \tilde{Z} = \tilde{z}] + (E[X|\tilde{Z} = \tilde{z}] - \hat{x})^2 \\ &+ 2E[(X - E[X|\tilde{Z}])(E[X|\tilde{Z}] - \hat{x}) | \tilde{Z} = \tilde{z}].\end{aligned}$$

This simplifies further since $X \to \tilde{Z} \to \hat{X}$ and thus $E(X - E[X|\tilde{Z}])(E[X|\tilde{Z}] - \bar{X}) = 0$. This gives us an equivalent distortion of

$$\tilde{d}(\tilde{z}, \hat{x}) = E[(X - E[X|\tilde{Z}])^2 | \tilde{Z} = \tilde{z}] + (E[X|\tilde{Z} = \tilde{z}] - \hat{x})^2.$$

When Wolf and Ziv [7] considered this problem, they found that the direct rate-distortion function for $E[X|\tilde{Z}]$ shifted by $E(X - E[X|\tilde{Z}])^2$ is the same as the remote rate-distortion function for $\tilde{Z}$. To simplify notation, we let

$$V = E[X|\tilde{Z}].$$

Since we are now just considering a shifted direct rate-distortion for source V and squared error distortion, we can apply our bounds in (1.22) to get

$$\frac{1}{2} \log \left(\frac{Q_V}{D - D_0} \right) \le R_X^R(D) \le \frac{1}{2} \log \left(\frac{\sigma_V^2}{D - D_0} \right), \tag{2.29}$$

where $D_0 = E(X - E[X|\tilde{Z}])^2$.

Since V is the MMSE estimate for X given $\mathbf{Z}$, the orthogonality principle implies that $\sigma_V^2 = \sigma_X^2 - D_0$. This allows us to rewrite the upper bound as

$$R_X^R(D) \le \frac{1}{2} \log \left(\frac{\sigma_X^2}{D} \right) + \frac{1}{2} \log \left(\frac{1 - \frac{D_0}{\sigma_X^2}}{1 - \frac{D_0}{D}} \right). \tag{2.30}$$

Note that Q_V does not depend on D. This fact implies that the lower bound is simply a vertically displaced version of the upper bound. That is,

$$R_X^R(D) \ge \frac{1}{2} \log \left(\frac{\sigma_X^2}{D} \right) + \frac{1}{2} \log \left(\frac{1 - \frac{D_0}{\sigma_X^2}}{1 - \frac{D_0}{D}} \right) - \log \left(\frac{\sigma_V^2}{Q_V} \right). \tag{2.31}$$

□

We now examine a few properties of these bounds. The first is the monotonicity of D_0.

Lemma 2.7. *Fix $\sigma_X^2 > 0$ and $0 < D \leq \sigma_X^2$. Then*

$$\frac{1}{2}\log\left(\frac{1-\frac{D_0}{\sigma_X^2}}{1-\frac{D_0}{D}}\right)$$

is nonnegative and monotonically decreasing in $D_0 > D$.

Proof. Nonnegativity follows immediately from how σ_X^2 and D are defined. To prove the monotonicity, consider any $0 < \delta \leq D_0$. Then,

$$\frac{1}{2}\log\left(\frac{1-\frac{D_0-\delta}{\sigma_X^2}}{1-\frac{D_0-\delta}{D}}\right) = \frac{1}{2}\log\left(\frac{1-\frac{D_0}{\sigma_X^2}}{1-\frac{D_0-\delta}{D}} + \frac{\frac{\delta}{\sigma_X^2}}{1-\frac{D_0-\delta}{D}}\right) \tag{2.32}$$

$$= \frac{1}{2}\log\left(\frac{1-\frac{D_0}{\sigma_X^2}}{1-\frac{D_0}{D}}\left(\frac{1-\frac{D_0}{D}+\frac{\delta}{\sigma_X^2}}{1-\frac{D_0}{D}+\frac{\delta}{D}}\right)\right) \tag{2.33}$$

$$\leq \frac{1}{2}\log\left(\frac{1-\frac{D_0}{\sigma_X^2}}{1-\frac{D_0}{D}}\right) \tag{2.34}$$

□

Thus, we can use Corollary 1.5 to give upper and lower bounds on D_0, which in turn will allow us to compare the bounds (2.28) and (2.27) to the ones in (2.21) and (2.23). While the bounds in equations (2.27) and (2.28) are correct to within a constant shift, in practice, computing the differential entropy of the conditional mean of X given an observation $\tilde{Z}$ can be impractical. However, unlike our previous lower bound (2.24), Q_V can be positive even when the Q_X is zero, which makes it useful for studying discrete sources.

2.4 Examples

We now plot these bounds for different source distributions observed in additive white Gaussian noise and with a squared error distortion. The source variance is $\sigma_X^2 = 1$ and the noise variances are $\sigma_{N_i}^2 = 1$ for each observation, and there are a total of $M = 20$ observations. We consider three source distributions: Gaussian, Laplacian, and uniform. For the Gaussian source, as expected, the upper and lower bounds are both tight. In computing the lower bounds for the Laplacian and uniform sources, we use the maximum entropy bound for $Q_{\tilde{Z}}$ rather than compute it. While this gives worse lower bounds, it makes these bounds easy to compute. This gives the bounds shown in Figure 2.3.

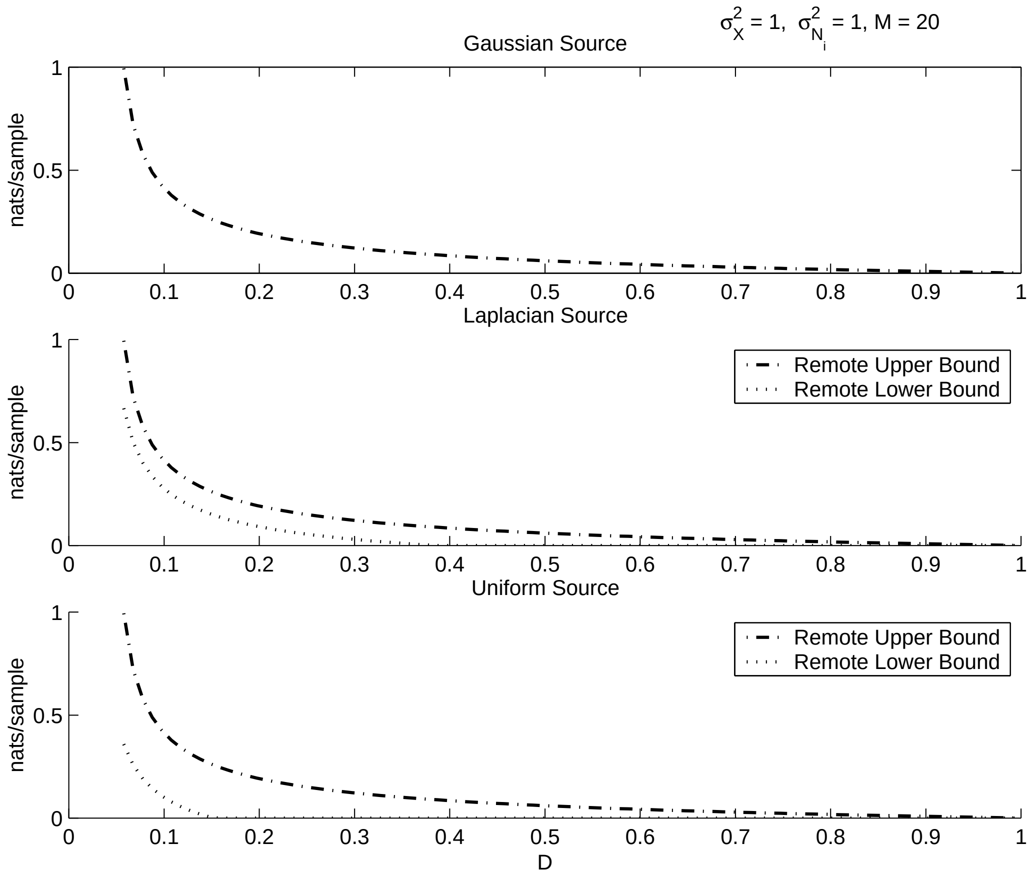

Figure 2.3. Plot of bounds for Gaussian, Laplacian, and uniform sources

2.5 Discussion

In this chapter, we have established general upper and lower bounds for the remote rate-distortion function when the observations are viewed in additive noise. For the case of Gaussian source and noise statistics and a squared error distortion, our bounds are tight. Further, the gap between the two bounds for non-Gaussian sources can be expressed in terms of the gap between the entropy power and variance of non-Gaussian sources. We then presented previously known results for the case of additive white Gaussian noise and squared error distortions. Unlike those bounds, ours are easy to evaluate for non-Gaussian source distributions.

Chapter 3

AWGN CEO Problems

In the previous chapter, we studied the remote source coding problem, for which the rate-distortion function can be characterized by a single-letter mutual information expression. In this chapter, we consider the CEO problem, a distributed version of the remote source coding problem, as seen in Figure 3.1. Except for special cases, the rate region for this problem is unknown.

Our focus is on the case in which the observations are corrupted by additive white Gaussian noise, which we call the AWGN CEO problem. Thus, we have, for $1 \leq i \leq M$,

$$Z_i(k) = X(k) + N_i(k), \quad k \geq 1, \tag{3.1}$$

where $N_i(k) \sim \mathcal{N}(0, \sigma_{N_i}^2)$.

In the proceeding sections, we derive lower and upper bounds for the sum-rate distortion function of AWGN CEO problems. While our arguments hold in greater generality, we restrict our attention to the case of squared error distortion to give better interpretations of our results. We begin by presenting well known inner and outer bounds to the rate region of the CEO problem. Using the outer bound and the entropy power inequality given in Appendix B, we derive an outer bound to

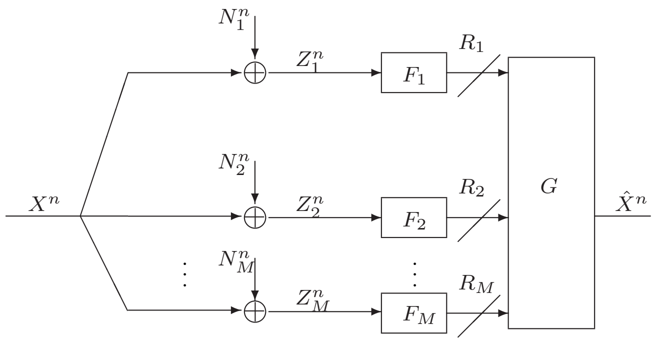

Figure 3.1. AWGN CEO problem

the sum-rate-distortion function for AWGN CEO problems. We then consider two upper bounds. The first upper bound follows from the maximum entropy bound. The second follows from a novel technique that bounds the difference between the sum-rate-distortion function for the CEO problem and the remote rate-distortion function.

3.1 Background and Notation

The Berger-Tung region [9] [10] provides inner and outer bounds in terms of single-letter mutual information expressions for discrete memoryless sources and bounded distortion measures, which apply to the CEO problem. Results by Housewright show that these bounds continue to hold for abstract alphabets and suitably smooth distortion functions [22]. The squared error distortion is one such case [19]. We rely on the Berger-Tung inner bound to get an upper bound to the sum-rate distortion function, which for $\mathbf{U} \in \mathcal{U}_X^{CEO}(D)$ is [23]

$$R_X^{CEO}(D) \leq I(\mathbf{Z}; \mathbf{U}), \tag{3.2}$$

$$\mathcal{U}_X^{CEO}(D) = \left\{ \mathbf{U} : U_i \to Z_i \to (X, Z_{\{i\}^c}, U_{\{i\}^c}), \exists f, Ed\left(X, f(\mathbf{U})\right) \leq D \right\}.$$

For our lower bound, we use an outer bound introduced by Wagner and Anantharam [11], [12] that subsumes the Berger-Tung outer bound.

Definition 3.1. If the set of random variables $(X, \mathbf{Z}, \mathbf{U}, W, T, \hat{X})$ is in the collection $\mathcal{W}_X^{CEO}(D)$, then $Ed(X, \hat{X}) \leq D$ and

(i) (W, T) is independent of $(X, \mathbf{Z})$,

(ii) $\mathbf{U}_A \leftrightarrow (\mathbf{Z}_A, W, T) \leftrightarrow (X, \mathbf{Z}_{A^c}, \mathbf{U}_{A^c})$ for all $A \subseteq \{1, \ldots M\}$,

(iii) $(X, W, T) \leftrightarrow (\mathbf{U}, T) \leftrightarrow \hat{X}$, and

(iv) the conditional distribution of U_i given W and T is discrete for each i.

For $(X, \mathbf{Z}, \mathbf{U}, W, T, \hat{X}) \in \mathcal{W}_X^{CEO}(D)$, the outer bound is given by the following collection of inequalities on subsets $A \subseteq \{1, \ldots M\}$ [12, p. 109]:

$$\sum_{i \in A} R_i \geq I(X; \mathbf{U}, T) - I(X; \mathbf{U}_{A^c} | T) + \sum_{i \in A} I(Z_i; U_i | X, W, T) \tag{3.3}$$

$$\geq I(X; \mathbf{U}, T) - I(X; \mathbf{U}_{A^c} | W, T) + \sum_{i \in A} I(Z_i; U_i | X, W, T), \tag{3.4}$$

where R_i is the rate at which agent i. Condition (iii) implies that $I(X; \mathbf{U}, T) \geq I(X; \hat{X}) \geq R_X(D)$, so we have the following lower bound on the sum-rate:

$$R_X^{CEO}(D) \geq \min_{\mathbf{U}, W, T \in \mathcal{W}_X^{CEO}(D)} R_X(D) - I(X; \mathbf{U}_{A^c} | W, T) + \sum_{i=1}^{M} I(Z_i; U_i | X, W, T). \tag{3.5}$$

For convenience, we will use the following shorthand to refer to scalar sufficient statistics for X given $\mathbf{Z}_A$.

$$\tilde{Z}_A = \frac{1}{|A|} \sum_{i \in A} \frac{\sigma^2_{\hat{N}_A}}{\sigma^2_{N_i}} Z_i \tag{3.6}$$

$$= X + \hat{N}_A, \tag{3.7}$$

where $\sigma^2_{\tilde{N}_A} = \frac{1}{\frac{1}{|A|}\sum_{i\in A}\frac{1}{\sigma^2_{N_i}}}$ and $\hat{N}_A = \frac{1}{|A|}\sum_{i\in A}\frac{\sigma^2_{\tilde{N}_A}}{\sigma^2_{N_i}}N_i$. Note that the variance of $\hat{N}_A \sim \mathcal{N}\left(0, \sigma^2_{\tilde{N}_A}/|A|\right)$. We remove the subscript A when $A = \{1, \ldots, M\}$.

3.2 Lower Bound

By applying the outer bound given in the previous section and a new entropy power inequality given in Theorem B.1, we can give the following lower bound to the sum-rate distortion functions of AWGN CEO problems with a single source. Note the similarities between it and the remote rate distortion lower bound in equation (2.21).

Theorem 3.2. *For the M-agent AWGN CEO problem for a source X, the sum-rate distortion function is lower bounded by*

$$R_X^{CEO}(D) \geq R_X(D) + \frac{M}{2}\log\left(\frac{MQ_X}{MQ_{\tilde{Z}} - \sigma^2_{\tilde{N}}e^{2R_X(D)}}\right) \tag{3.8}$$

To prove the theorem, we start by considering the following lemma. This is a generalization of the lemma proved by Oohama [16] that holds for non-Gaussian sources, as well.

Lemma 3.3. *Let $r_i = I(Z_i; U_i|X, W, T)$ and $A \subseteq \{1, \ldots, M\}$. Then*

$$e^{2I(X;\mathbf{U}_A|W,T)} \leq \frac{e^{2H(\tilde{Z}_A)}}{e^{2H(\tilde{N}_A)}} - \frac{e^{2H(X)}}{|A|}\sum_{i\in A}\frac{e^{-2r_i}}{\sigma^2_{N_i}/\sigma^2_{\tilde{N}_A}}. \tag{3.9}$$

Proof. Since (W, T) is independent of $(X, \mathbf{Z})$ when condition (i) holds, we know that we preserve the Markov chain $X \to \tilde{Z}_A \to \mathbf{Z}_A \to \mathbf{U}_A$ when we condition on any realization of (W, T), so Theorem B.1 implies

$$e^{2H(\tilde{Z}_A)}e^{-2I(X;\mathbf{U}_A|W=w,T=t)} \geq e^{2H(X)}e^{-2I(\tilde{Z};\mathbf{U}_A|W=w,T=t)} + e^{2H(\hat{N}_A)} \tag{3.10}$$

$$= \frac{e^{2H(X)}}{e^{2H(\tilde{Z}_A)}}e^{2H(\tilde{Z}_A|\mathbf{U}_A,W=w,T=t)} + e^{2H(\hat{N}_A)}. \tag{3.11}$$

Now,

$$
\begin{aligned}
&H(\tilde{Z}_A|\mathbf{U}_A, W = w, T = t)\\
&\quad = H(\tilde{Z}_A|\mathbf{U}_A, X, W = w, T = t) + I(X; \tilde{Z}_A|\mathbf{U}_A, W = w, T = t) && (3.12)\\
&\quad = H(\tilde{Z}_A|\mathbf{U}_A, X, W = w, T = t) + I(X; \tilde{Z}_A, \mathbf{U}_A|W = w, T = t)\\
&\qquad - I(X; \mathbf{U}_A|W = w, T = t) && (3.13)\\
&\quad = H(\tilde{Z}_A|\mathbf{U}_A, X, W = w, T = t) + I(X; \tilde{Z}_A|W = w, T = t)\\
&\qquad - I(X; \mathbf{U}_A|W = w, T = t) && (3.14)\\
&\quad = H(\tilde{Z}_A|\mathbf{U}_A, X, W = w, T = t) + I(X; \tilde{Z}_A)\\
&\qquad - I(X; \mathbf{U}_A|W = w, T = t), && (3.15)
\end{aligned}
$$

where (3.14) follows from the Markov chain $X \to \tilde{Z}_A \to \mathbf{Z}_A \to \mathbf{U}_A$ and (3.15) from (i). Now it's simply a matter of bounding $H(\tilde{Z}_A|\mathbf{U}_A, X, W = w, T = t)$. However, we note that we can write

$$
\begin{aligned}
e^{2H(\tilde{Z}_A|\mathbf{U}_A,X,W=w,T=t)} &= e^{2H\left(\frac{1}{|A|}\sum_{i\in A}\frac{\sigma^2_{\tilde{N}_A}}{\sigma^2_{N_i}}Z_i\middle|\mathbf{U}_A,X,W=w,T=t\right)} && (3.16)\\
&\geq \sum_{i\in A}\left(\frac{\sigma^2_{\tilde{N}_A}}{\sigma^2_{N_i}}\right)^2 \frac{e^{2H(Z_i|U_i,X,W=w,T=t)}}{|A|^2} && (3.17)
\end{aligned}
$$

where (3.17) follows by (ii) and the entropy power inequality. But

$$
\begin{aligned}
H(Z_i|U_i, X, W = w, T = t) &= H(Z_i|X, W = w, T = t) - I(Z_i; U_i|X, W = w, T = t) && (3.18)\\
&= H(N_i) - I(Z_i; U_i|X, W = w, T = t). && (3.19)
\end{aligned}
$$

Combining (3.11), (3.15), (3.17), and (3.19) gives

$$
\begin{aligned}
e^{2H(\tilde{Z}_A)}e^{-2I(X;\mathbf{U}_A|W=w,T=t)} \geq & \frac{e^{2H(X)}}{e^{2H(\tilde{Z}_A)}} \cdot \frac{e^{2I(X;\tilde{Z}_A)}}{e^{2I(X;\mathbf{U}_A|W=w,T=t)}}\\
& \cdot \sum_{i\in A}\left(\frac{\sigma^2_{\tilde{N}_A}}{\sigma^2_{N_i}}\right)^2 \frac{e^{2H(N_i)-2I(Z_i;U_i|X,W=w,T=t)}}{|A|^2} + e^{2H(\hat{N}_A)}. && (3.20)
\end{aligned}
$$

Solving for $e^{2I(X;\mathbf{U}_A|W=w,T=t)}$, we get that

$$e^{2I(X;\mathbf{U}_A|W=w,T=t)} \leq e^{-2H(\tilde{N}_A)} \left[e^{2H(\tilde{Z}_A)} - \frac{e^{2H(X)} e^{2I(X;\tilde{Z}_A)}}{e^{2H(\tilde{Z}_A)}} \sum_{i \in A} \frac{e^{2H(\hat{N}_A) - 2I(Z_i;U_i|X,W=w,T=t)}}{|A| \sigma^2_{N_i} / \sigma^2_{\tilde{N}_A}} \right] \tag{3.21}$$

$$= \frac{e^{2H(\tilde{Z}_A)}}{e^{2H(\tilde{N}_A)}} - \frac{e^{2H(X)}}{|A|} \sum_{i \in A} \frac{e^{-2I(Z_i;U_i|X,W=w,T=t)}}{\sigma^2_{N_i} / \sigma^2_{\tilde{N}_A}}. \tag{3.22}$$

We complete the proof by taking the expectation over W, T and applying Jensen's inequality twice, once to the left-hand side and once to the right-hand side. □

At this point, we can substitute this inequality into (3.4) to get

$$\sum_{i \in A} R_i \geq R_X(D) - \frac{1}{2} \log \left(\frac{e^{2H(\tilde{Z}_{A^c})}}{e^{2H(\tilde{N}_{A^c})}} - \frac{e^{2H(X)}}{|A^c|} \sum_{i \in A^c} \frac{e^{-2r_i}}{\sigma^2_{N_i} / \sigma^2_{\tilde{N}_{A^c}}} \right) + \sum_{i \in A} r_i. \tag{3.23}$$

Optimizing over this set of inequalities gives us lower bounds on r_i and substituting these choices into (3.5) gives (3.8), thus proving the theorem.

Corollary 3.4. *For the M-agent quadratic AWGN CEO problem for a source X, the sum-rate distortion function is lower bounded by*

$$R_X^{CEO}(D) \geq \frac{1}{2} \log \left(\frac{Q_X}{D} \right) + \frac{M}{2} \log \left(\frac{M Q_X}{M Q_{\tilde{Z}} - \frac{Q_X}{D} \sigma^2_{\tilde{N}}} \right) \tag{3.24}$$

Proof. The result follows from (3.8) and the lower bound for the direct rate-distortion function given in (1.22). □

3.3 Upper Bounds

In this section, we consider upper bounds to the sum-rate distortion function. We focus exclusively on the mean squared error case. The first bound follows from a maximum entropy bound. The second bound takes advantage of the difference between the AWGN CEO sum-rate-distortion function and the remote rate-distortion function.

3.3.1 Maximum Entropy Upper Bound

Our first upper bound requires one to know the MMSE performance of the source through additive Gaussian noise.

Theorem 3.5. *Let t be a function in the set $\mathcal{T}_X$ given in Definition 1.19. Then an upper bound for the sum-rate distortion function for the AWGN CEO problem is*

$$R_X^{CEO}(D) \leq \frac{1}{2}\log\left(\frac{\sigma_X^2}{D_l}\right) + \frac{M}{2}\log\left(\frac{M\sigma_X^2}{M\sigma_{\tilde{Z}}^2 - \frac{\sigma_X^2}{D_l}\sigma_{\tilde{N}}^2}\right), \tag{3.25}$$

where $\frac{\sigma_X^2\sigma_{\tilde{N}}^2}{M\sigma_X^2+\sigma_{\tilde{N}}^2} < D_l \leq \sigma_X^2$ is the solution to

$$D = t\left(\frac{\sigma_X^2}{D_l} - 1, \sigma_X^2\right). \tag{3.26}$$

Proof. We simply let the $U_i = Z_i + W_i$, where the W_i are independent, jointly Gaussian, independent of X and the N_i. Since our previous bounds have involved the sufficient statistic $\tilde{Z}$ for X given $\mathbf{Z}$, we select the W_i such that the sufficient statistic for X given $\mathbf{U}$ is a noisy version of $\tilde{Z}$. Thus, we choose W_i such that $W_i + N_i$ has variance $\beta\sigma_{N_i}^2$. That is, we know that

$$D \leq t\left(\frac{M\sigma_X^2}{\beta\sigma_{\tilde{N}}^2}, \sigma_X^2\right), \tag{3.27}$$

so set $\frac{\sigma_X^2}{D_l} - 1 = \frac{M\sigma_X^2}{\beta\sigma_{\tilde{N}}^2}$ and substitute it into future expressions in place of β. Finally, taking advantage of the fact that the source has a second moment constraint, we apply the maximum entropy theorem and Lemma E.1 to get the bound in (3.25). □

By Lemma 1.20, we know that for any X, there exists a function $t \in \mathcal{T}_X$ that when applied to Theorem 3.5 gives $D = D_l$. Thus,

$$R_X^{CEO}(D) \leq \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right) + \frac{M}{2}\log\left(\frac{M\sigma_X^2}{M\sigma_X^2 + \sigma_{\tilde{N}}^2 - \frac{\sigma_X^2}{D}\sigma_{\tilde{N}}^2}\right). \tag{3.28}$$

However, this is just the sum-rate distortion function for the quadratic Gaussian CEO problem [24]. One can also verify that for this case, the lower bound in Corollary 3.4 is tight. Thus, the quadratic Gaussian CEO sum-rate distortion function is an upper bound for all sources. Notice that (3.28) is a valid bound for any source X with variance σ_X^2. Thus, Gaussian statistics are a worst case for the sum-rate-distortion function, just as they are for estimation.

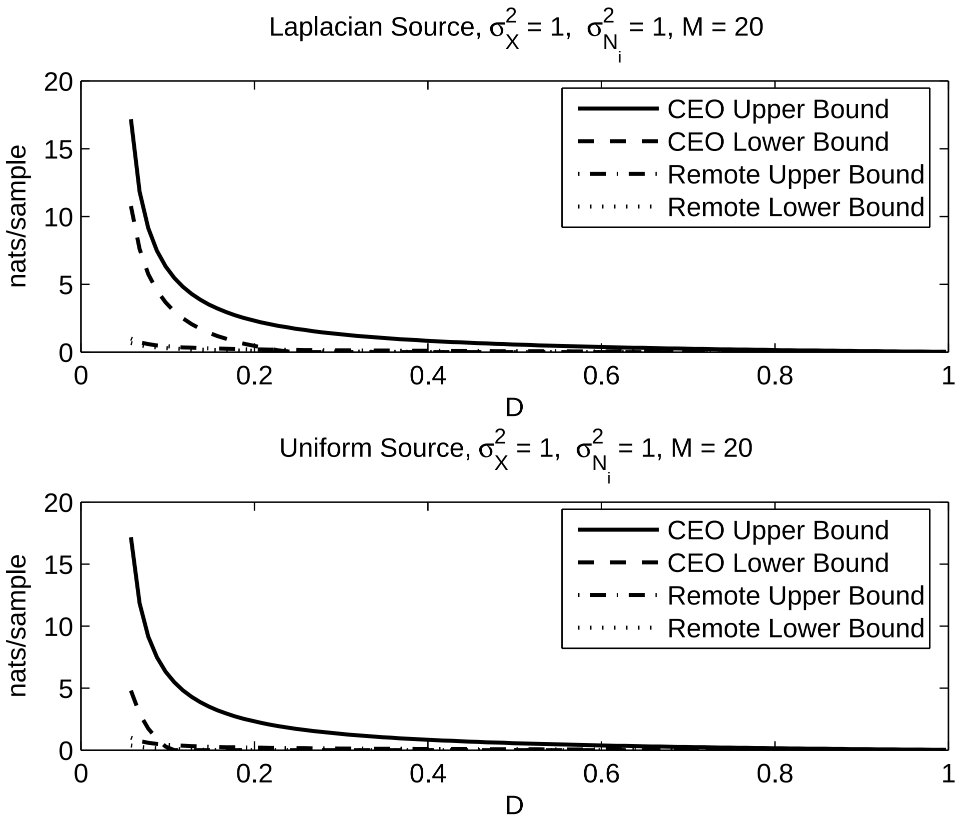

Figure 3.2. Plots for the Laplacian and uniform source in AWGN CEO problem.

In Figure 3.2, we compare how closely the upper bound in (3.28) compares to the lower bound given in Corollary 3.4. In particular, we consider the case of the Laplacian and uniform source. Note that the lower bounds improve on the remote rate-distortion lower bound for M-observations, which is also a valid lower bound for the AWGN CEO problem. However, there is still a gap between the bounds since neither the Laplacian nor uniform source meet the maximum entropy bound. The approach considered in the next section is one way to improve upon this weakness.

3.3.2 Rate Loss Upper Bound

We now restrict our attention to cases in which the noise variances are equal. That is, $\sigma_{N_i}^2 = \sigma_N^2$. For such cases, that harmonic mean $\sigma_{\tilde{N}}^2 = \sigma_N^2$. For convenience, we will use the following shorthand for the right-hand side of (3.25).

$$R_1(D) = \frac{1}{2} \log \left(\frac{\sigma_X^2}{D_l} \right) + \frac{M}{2} \log \left(\frac{M\sigma_X^2}{M\sigma_X^2 + \sigma_N^2 - \frac{\sigma_X^2}{D_l}\sigma_{\tilde{N}}^2} \right). \tag{3.29}$$

While the $R_1(D)$ is straightforward to calculate, maximizing the mutual information under the assumption that the source is Gaussian makes it somewhat coarse. We consider an alternate approach and rely on the fact that we can express the sum-rate distortion function in the CEO problem as

$$R_X^{CEO}(D) = R_X^R(D) + \left(R_X^{CEO}(D) - R_X^R(D)\right) \tag{3.30}$$

$$= R_X^R(D) + L(D), \tag{3.31}$$

where we define $L(D) = R_X^{CEO}(D) - R_X^R(D)$.

Unlike $R_X^{CEO}(D)$, for which an information theoretic expression is unknown, $R_X^R(D)$ is given exactly by (1.28) and can be computed in principle [5]. Otherwise, we can use (2.30) as an upper bound. Thus, our second upper bound approach is to find an upper bound to $L(D)$ and either find an upper bound to $R_X^R(D)$ or compute it exactly. We can bound $L(D)$ as follows.

Theorem 3.6. *Let t be a function in the set $\mathcal{T}_X$ given in Definition 1.19. Then an upper bound for the rate loss between the CEO sum-rate-distortion function and the* *remote rate-distortion function is*

$$L(D) \leq \frac{M-1}{2}\log\left(\frac{M\sigma_X^2}{M{\sigma_X}^2 - s{\sigma_N}^2}\right) + \frac{1}{2}\log\left(\frac{M\cdot(D+2\sqrt{D\cdot\sigma_N^2})s + M\sigma_X^2}{M{\sigma_X}^2 - s{\sigma_N}^2}\right) \tag{3.32}$$

where $0 \leq s < M\sigma_X^2/\sigma_N^2$ *is the solution to*

$$D = t\left(s, \sigma_X^2\right). \tag{3.33}$$

Proof. By Lemma D.2 and Theorem D.4, we have that

$$R_X^{CEO}(D) - R_X^R(D) \leq \max_{\underline{H}\in\mathcal{H}(D)} I(\underline{H}; \underline{H} + \underline{W}), \tag{3.34}$$

$$\mathcal{H}(D) = \Big\{\underline{H} : \exists 0 \leq a \leq 1, \exists -1 \leq b \leq 1, \; E{H_i}^2 = aD + 2b\sqrt{aD\sigma_N^2} + \sigma_N^2, \; EH_iH_j = aD + 2b\sqrt{aD\sigma_N^2}\Big\},$$

where $\underline{W}$ can be any random vector independent of $\underline{H}, X, \underline{Z}$ that satisfies

$$E(X - E[X|Z_1 + W_1, \ldots, Z_M + W_M])^2 \leq D. \tag{3.35}$$

To make our expressions easy to evaluate, we will assume $\underline{W}$ is i.i.d. jointly Gaussian. By the maximum entropy theorem [20, Thm. 9.6.5, p. 234], $H_1, \ldots, H_M$ will be jointly Gaussian, and the optimal choices for a and b are $a = b = 1$. This simplifies the right-hand side of (3.34) to

$$\frac{M-1}{2}\log\left(1 + \frac{\sigma_N^2}{\sigma_W^2}\right) + \frac{1}{2}\log\left(1 + \frac{M\cdot(D+2\sqrt{D\cdot\sigma_N^2}) + \sigma_N^2}{\sigma_W^2}\right). \tag{3.36}$$

All that is left is to select σ_W^2 to satisfy (3.35). Noting that our sufficient statistic is just the sum $\sum Z_i + W_i$, all we have to do is set

$$D = t\left(\frac{M\sigma_X^2}{\sigma_N^2 + \sigma_W^2}, \sigma_X^2\right). \tag{3.37}$$

Defining $s = \frac{M\sigma_X^2}{\sigma_N^2+\sigma_W^2} < \frac{M\sigma_X^2}{\sigma_N^2}$, solving for σ_W^2, and substituting it into (3.36) gives us our bound in (3.32). □

3.3.3 Rate Loss Upper Bound vs. Maximum Entropy Upper Bound

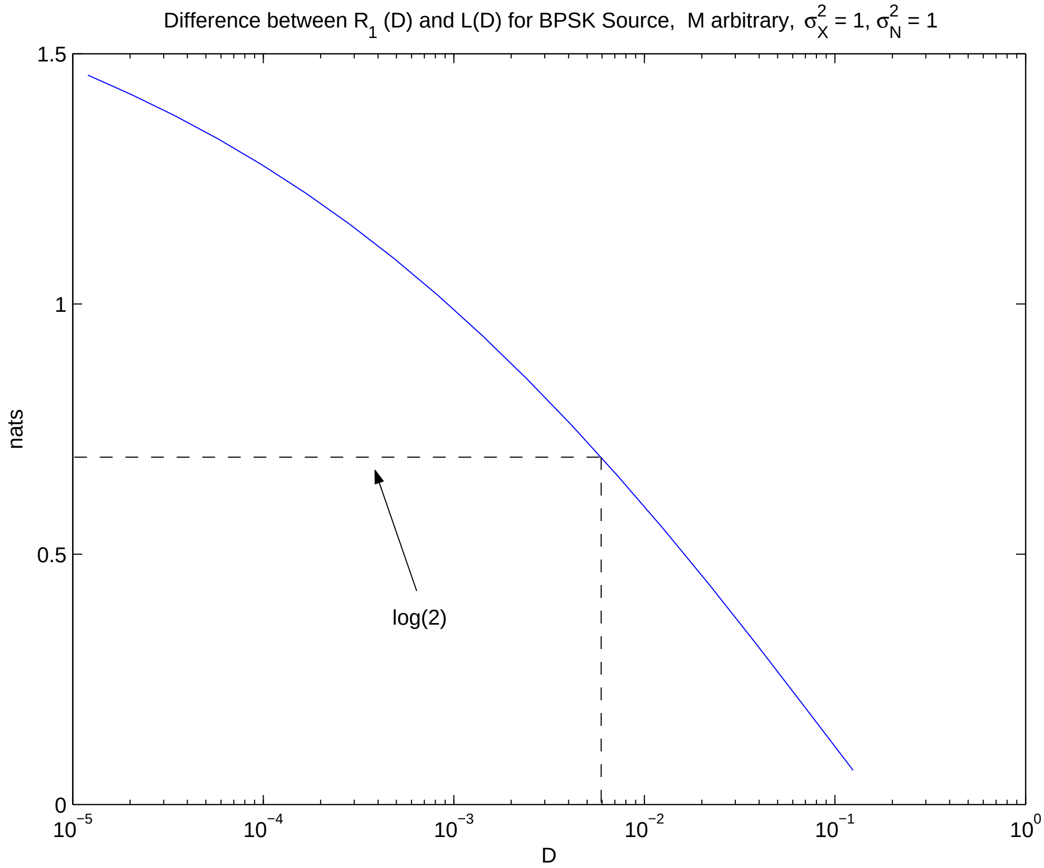

Figure 3.3. When the curve crosses log(2), the rate loss approach provides a better bound than the maximum entropy bound for the BPSK source and large enough M.

Instead of computing $R_X^R(D)$ exactly, we consider approaches to find an upper bound for it. A maximum entropy bound on $R_X^R(D)$ results in a bound that is worse than $R_1(D)$ (see Appendix). For this reason, the rate loss approach gives a worse bound than $R_1(D)$ for the Gaussian case, for which the latter is tight. Thus, we move away from bounds $R_X^R(D)$ that hold for all possible sources, and consider specializing it to specific sources. To see how good such bounds on $R_X^R(D)$ need to be, consider what happens when we subtract (3.32) from (3.29).

$$R_1(D) - L(D) \geq \frac{1}{2}\log\left(\frac{\sigma_X^2}{D + 2\sqrt{D\cdot\sigma_N^2} + D_l}\right). \tag{3.38}$$

Thus, when $R_X^R(D)$ is smaller than (3.38) for some choice of D, then our second bound will be strictly better than $R_1(D)$. Indeed, this turns out to be true for the BPSK source (Example 1.3), as we now show.

Consider the following coding strategy for the BPSK source in the remote source coding setting. First, the encoder averages its observations $Z_i(k)$, giving

$$\hat{Z}(k) = \frac{1}{M}\sum_{i=1}^{M} Z_i(k) = X(k) + \frac{1}{M}\sum_{i=1}^{M} N_i(k). \tag{3.39}$$

Next, we quantize these observations to $+\sigma_X$ when $\hat{Z}(k)$ is positive and to $-\sigma_X$ when it is negative. We call the quantized version $\tilde{Z}(k)$. By the same arguments in Example 1.3, this gives

$$E(X(k) - \tilde{Z}(k))^2 \leq 4{\sigma_X}^2 \exp\left\{-\frac{M\sigma_X^2}{2\sigma_N^2}\right\}, \tag{3.40}$$

which we get simply by replacing snr with $\frac{M\sigma_X^2}{\sigma_N^2}$ in the right hand side of (1.7). If we now apply noiseless source coding to the sequence $\tilde{Z}(k)$, it is clear that $R = \log 2$ is sufficient to reconstruct $\tilde{Z}(k)$ with arbitrarily small error probability as the block length gets large, and that with small enough error probability, we can approach the distortion on the right-hand side of (3.40) arbitrarily closely.

For large enough M, the right-hand side of (3.40) can be made arbitrarily small. Thus, for all $\delta > 0$ and for large enough M, $(R, D) = (\log 2, \delta)$ is achievable in the remote source coding problem for the BPSK source. Observing that the difference curve in (3.38) does not depend on M, it is sufficient to show that for the BPSK source and for some $D > 0$, this curve is larger than $\log 2$. This is evident in Figure 3.3. We also plot an example for the rate loss approach in Figure 3.4

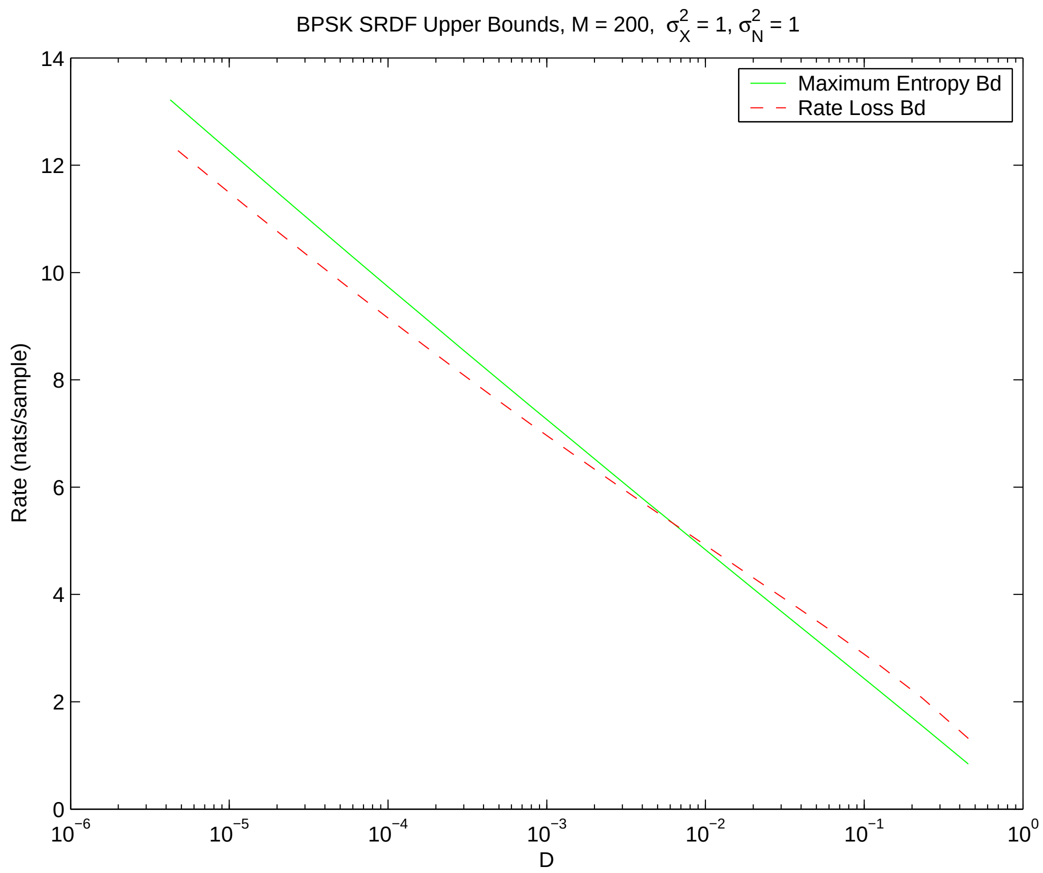

Figure 3.4. The rate loss upper bound outperforms the maximum entropy upper bound for the BPSK source for certain distortions.

3.4 Discussion

We presented a lower bound and two upper bounds on the sum-rate-distortion function for the AWGN CEO problem. The lower bound and the maximum entropy upper bound are as tight as the gap between the entropy powers and variances found in their expressions, respectively. To reduce this gap, our rate loss upper bound provides an improvement on the maximum entropy bound for certain non-Gaussian sources and certain target distortion values. One disadvantage of this approach is that there is no simple closed form expression for such a bound, unlike the maximum entropy bound. One might notice a relationship between the lower bound and maximum entropy upper bound presented in this chapter with the bounds presented in Chapter 2. In Chapter 4, we explore this relationship in greater detail by considering what happens as the number of observations gets large in both problems.

Chapter 4

Scaling Laws

In this chapter, we examine what happens as the number of observations gets large in the remote source coding and CEO problems. Recall that in Chapters 2 and 3, we assumed a finite number of observation M and for $1 \leq i \leq M$,

$$Z_i(k) = X(k) + N_i(k), \quad k \geq 1, \tag{4.1}$$

where $N_i(k) \sim \mathcal{N}(0, \sigma_{N_i}^2)$. In this chapter, we examine what happens as we let M get large. We will focus exclusively on the case of squared error distortion for both models. That is, $d(x, \hat{x}) = (x - \hat{x})^2$.

In the next section, we provide definitions and notation that will be useful in proving our scaling laws. We then present scaling laws for the remote source coding problem and show that as the number of observations increases, the remote rate-distortion function converges to the direct rate-distortion function. For the CEO problem, we find that the sum-rate-distortion function does not converge to the classical rate-distortion function and that there is, in fact, a penalty. It turns out that this penalty results in a different scaling behavior for the CEO sum-rate-distortion function. As a cautionary tale on scaling laws, we consider a coding strategy for the CEO problem that does not exploit the redundancy among the distributed observations. This "no binning" approach ends up exhibiting the same scaling behavior as the sum-rate-distortion function in the CEO problem.

4.1 Definitions and Notation

The following definition will allows us to state our scaling law results precisely.

Definition 4.1. Two functions $f(D)$ and $g(D)$ are asymptotically equivalent, denoted $f(D) \sim g(D)$, if there exist positive real numbers K_1 and K_2 such that

$$K_1 \leq \liminf_{D \to 0} \frac{f(D)}{g(D)} \leq \limsup_{D \to 0} \frac{f(D)}{g(D)} \leq K_2. \tag{4.2}$$

For convenience, we will use the following shorthand to refer to scalar sufficient statistics for X given $\mathbf{Z}$.

$$\begin{aligned} \tilde{Z} &= \frac{1}{M} \sum_{i=1}^{M} \frac{\sigma_{\tilde{N}}^2}{\sigma_{N_i}^2} Z_i & (4.3) \\ &= X + \hat{N}, & (4.4) \end{aligned}$$

where the harmonic mean $\sigma_{\tilde{N}}^2 = \frac{1}{\frac{1}{M}\sum_{i=1}^{M} \frac{1}{\sigma_{N_i}^2}}$ and $\hat{N} = \frac{1}{M} \sum_{i=1}^{M} \frac{\sigma_{\tilde{N}}^2}{\sigma_{N_i}^2} N_i$. Note that the variance of $\hat{N} \sim \mathcal{N}\left(0, \sigma_{\tilde{N}}^2/M\right)$. We assume that the harmonic mean $\sigma_{\tilde{N}}^2$ stays fixed as the number of observations M increases. One important case in which this holds is the equi-variance case in which $\sigma_{N_i}^2 = \sigma_N^2$.

Since the number of observations M is no longer a fixed parameter in our analysis, we now denote the M-observation remote rate-distortion function and CEO sum-rate-distortion function as $R_X^{R,M}(D)$ and $R_X^{CEO,M}(D)$, respectively. Further, the notation $R_X^{R,\infty}(D) = \lim_{M\to\infty} R_X^{R,M}(D)$ and $R_X^{CEO,\infty}(D) = \lim_{M\to\infty} R_X^{CEO,M}(D)$ will be useful when stating our scaling laws.

4.2 Remote Source Coding Problem

Upper and lower bounds for the remote rate-distortion function with squared error distortion were given in Corollary 2.5. While one can take the limit of the upper and lower bounds to derive our scaling law, we will show a slightly stronger result. Before doing so, we establish the following result for the direct rate-distortion function.

Lemma 4.2. *When $Q_X > 0$, the direct rate-distortion function behaves as*

$$R_X(D) \sim \log \frac{1}{D}. \tag{4.5}$$

Proof. Recalling the upper and lower bounds to the direct-rate-distortion function in (1.22), we know that

$$\frac{1}{2} \log \left(\frac{Q_X}{D} \right) \leq R_X(D) \leq \frac{1}{2} \log \left(\frac{\sigma_X^2}{D} \right). \tag{4.6}$$

From this, we can conclude that

$$\limsup_{D \to 0} \frac{R_X(D)}{\log \frac{1}{D}} \leq 1, \tag{4.7}$$

$$\liminf_{D \to 0} \frac{R_X(D)}{\log \frac{1}{D}} \geq 1. \tag{4.8}$$

This satisfies the conditions in the definition, so we have established the result. □

Theorem 4.3. *For the AWGN remote source coding problem with M-observations and a squared error distortion, the remote rate-distortion function converges to the direct rate-distortion function as $M \to \infty$. That is,*

$$R_X^{R,M}(D) \to R_X(D) \tag{4.9}$$

as $M \to \infty$.

Proof. It is clear that the direct rate-distortion function for X is less than the remote rate-distortion for X given $Z_1, \ldots, Z_M$. That is, $R_X(D) \leq R_X^{R,M}(D)$ for all M.

Thus, all we have to establish is that the remote rate-distortion function for X given $Z_1, \ldots, Z_M$ converges to a function that is at most the direct rate-distortion function for X.

By Lemma C.1 and Lemma C.2, we know that it is sufficient to consider the remote rate-distortion function for X given the sufficient statistic $\tilde{Z}$ defined in (C.2). By the Cauchy-Schwartz inequality, we know that if $E(\tilde{Z} - U)^2 = \delta$, then

$$\delta - \sqrt{\delta \frac{\sigma_{\tilde{N}}^2}{M}} \leq E(X - U)^2 \leq \delta + \sqrt{\delta \frac{\sigma_{\tilde{N}}^2}{M}}. \tag{4.10}$$

Similarly, if $E(X - U)^2 = \delta$,

$$\delta - \sqrt{\delta \frac{\sigma_{\tilde{N}}^2}{M}} \leq E(\tilde{Z} - U)^2 \leq \delta + \sqrt{\delta \frac{\sigma_{\tilde{N}}^2}{M}}. \tag{4.11}$$

Thus, by the single-letter characterizations for the direct and remote rate-distortion functions given in (1.21) and (1.28), respectively, we can conclude that the remote rate-distortion function for X given $\tilde{Z}$ converges to the direct rate-distortion function for $\tilde{Z}$ (denoted $R_{\tilde{Z}}^{M}(D)$). That is, as $M \to \infty$,

$$\left| R_X^{R,M}(D) - R_{\tilde{Z}}^{M}(D) \right| \to 0. \tag{4.12}$$

By the same argument, we know that the remote-rate distortion function for $\tilde{Z}$ given X (denoted $R_{\tilde{Z}}^{R}(D)$) converges to the direct rate-distortion function for X. That is, as $M \to \infty$,

$$R_{\tilde{Z}}^{R,M}(D) \to R_X(D). \tag{4.13}$$

However, we also know that $R_{\tilde{Z}}^{R,M}(D) \geq R_{\tilde{Z}}^{M}(D)$. Thus, we can establish that $R_X^{R,M}(D)$ converges to a function that is at most $R_X(D)$, which completes our proof. □

Theorem 4.3 implies that the scaling behavior of the remote source coding problem is that same as in Lemma 4.2. We summarize this in the following corollary.

Corollary 4.4. *For the AWGN remote source coding problem with M-observations and a squared error distortion, the remote rate-distortion function scales as $\frac{1}{D}$ in the limit as $M \to \infty$. That is,*

$$R_X^{R,\infty}(D) \sim \log \frac{1}{D}. \tag{4.14}$$

4.3 CEO Problem

We now establish a scaling law for the sum-rate-distortion function for the AWGN CEO problem.

Theorem 4.5. *When $Q_X > 0$ and the limit of the right-hand side of (3.24) exists as $M \to \infty$,*

$$R_X^{CEO,\infty}(D) \sim \frac{1}{D}. \tag{4.15}$$

In fact, the following upper and lower bounds hold.

$$\begin{aligned} \frac{1}{2}\log\left(\frac{Q_X}{D}\right) + \frac{\sigma_{\tilde{N}}^2}{2}\left(\frac{1}{D} - J(X)\right) \\ \leq \quad R_X^{CEO,\infty}(D) \quad \leq \\ \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right) + \frac{\sigma_{\tilde{N}}^2}{2}\left(\frac{1}{D} - \frac{1}{\sigma_X^2}\right). \end{aligned} \tag{4.16}$$

Proof. If we can establish (4.16), the scaling law result follows immediately since

$$\liminf_{D\to 0} DR_X^{CEO,\infty}(D) = \limsup_{D\to 0} DR_X^{CEO,\infty}(D) = \frac{\sigma_N^2}{2}. \tag{4.17}$$

Taking the limit in (3.24) as $M \to \infty$ gives

$$R_X^{CEO,\infty}(D) \geq \log\left(\frac{Q_X}{D}\right) + \frac{\sigma_{\tilde{N}}^2}{2}\left(\frac{1}{D} - J(X)\right), \tag{4.18}$$

Likewise, we can take the as $M \to \infty$ for the upper bound given in (3.28) to get

$$R_X^{CEO,\infty}(D) \leq \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right) + \frac{\sigma_{\tilde{N}}^2}{2}\left(\frac{1}{D} - \frac{1}{\sigma_X^2}\right). \tag{4.19}$$

Thus, we have shown the desired results. □

Clearly, the above result holds for a Gaussian source. However, we are interested in finding non-Gaussian sources for which we know this scaling behavior holds. The following examples show it holds for a Laplacian source as well as a logistic source.

Example 4.6. Consider a data source with a Laplacian distribution. That is,

$$f(x) = \frac{1}{\sqrt{2}\sigma_X} e^{-\frac{\sqrt{2}|x|}{\sigma_X}}.$$

For this data source, the Fisher information is

$$J(X) = \frac{2}{\sigma_X^2}$$

and the differential entropy is

$$H(X) = \frac{1}{2}\log 2e^2\sigma_X^2.$$

Thus, for this case, the limit exists for the lower bound in (3.24) and is

$$R_X^{CEO,\infty}(D) \geq \frac{1}{2}\log\left(\frac{e\sigma_X^2}{\pi D}\right) + \frac{\sigma_{\tilde{N}}^2}{2}\left(\frac{1}{D} - \frac{2}{\sigma_X^2}\right), \tag{4.20}$$

where the inequality follows from (1.22). Thus, we can conclude that for the Laplacian source, $R_X^{CEO,\infty}(D) \sim \frac{1}{D}$. The gap between the direct and CEO sum-rate-distortion function is shown in Figure 4.1.

Example 4.7. Consider a data source with a Logistic distribution. That is,

$$f(x) = \frac{e^{-\frac{x}{\beta}}}{\beta\left(1 + e^{-\frac{x}{\beta}}\right)^2}.$$

For this data source, the Fisher information is

$$J(X) = \frac{1}{3\beta^2},$$

the entropy power is [20, p. 487]

$$Q_X = \frac{e^3\beta^2}{2\pi},$$

and the variance is

$$\sigma_X^2 = \frac{\pi^2\beta^2}{3}.$$

Thus, for this case, the limit exists for the lower bound in (3.24), so we can conclude that for the Logistic source, $R_X^{CEO,\infty}(D) \sim \frac{1}{D}$. The gap between the direct and CEO sum-rate-distortion function is shown in Figure 4.2. Notice that the gap is even smaller than in Example 4.6.

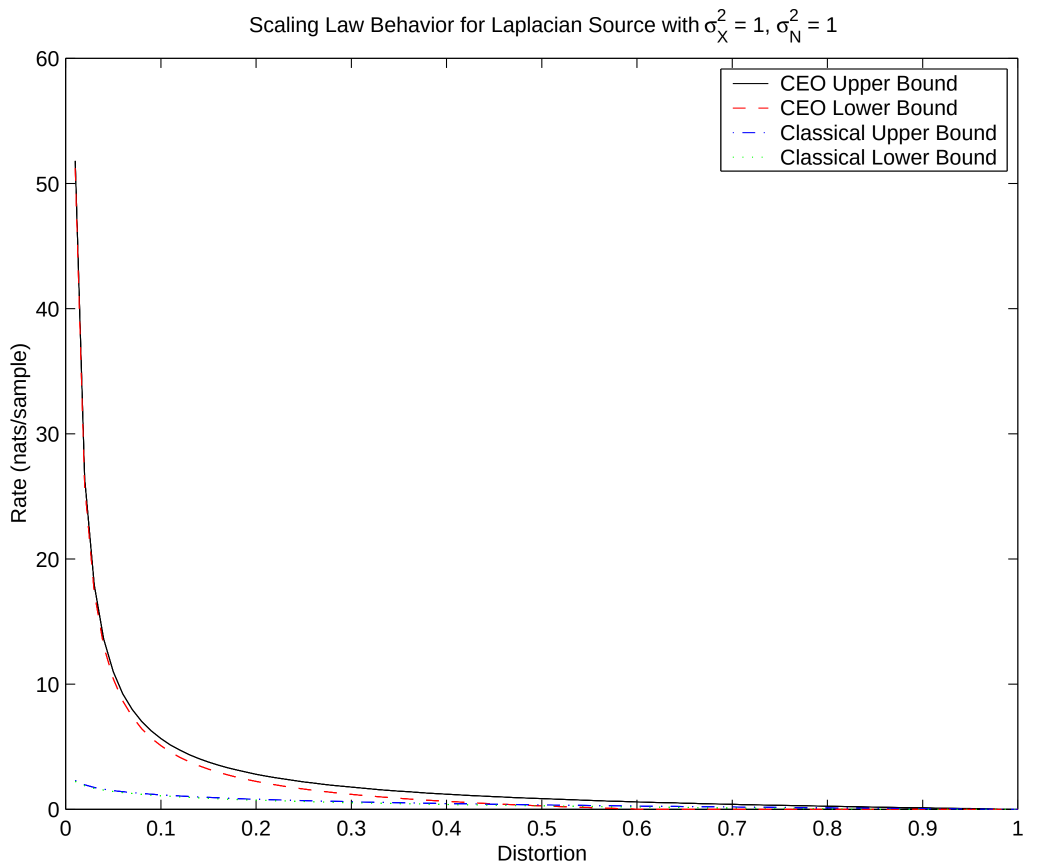

Figure 4.1. Scaling behavior for Laplacian source in AWGN CEO Problem

4.4 No Binning

As a cautionary tale about the utility of scaling laws, we consider a coding strategy in which encoders do not exploit the correlation with observations at other encoders.

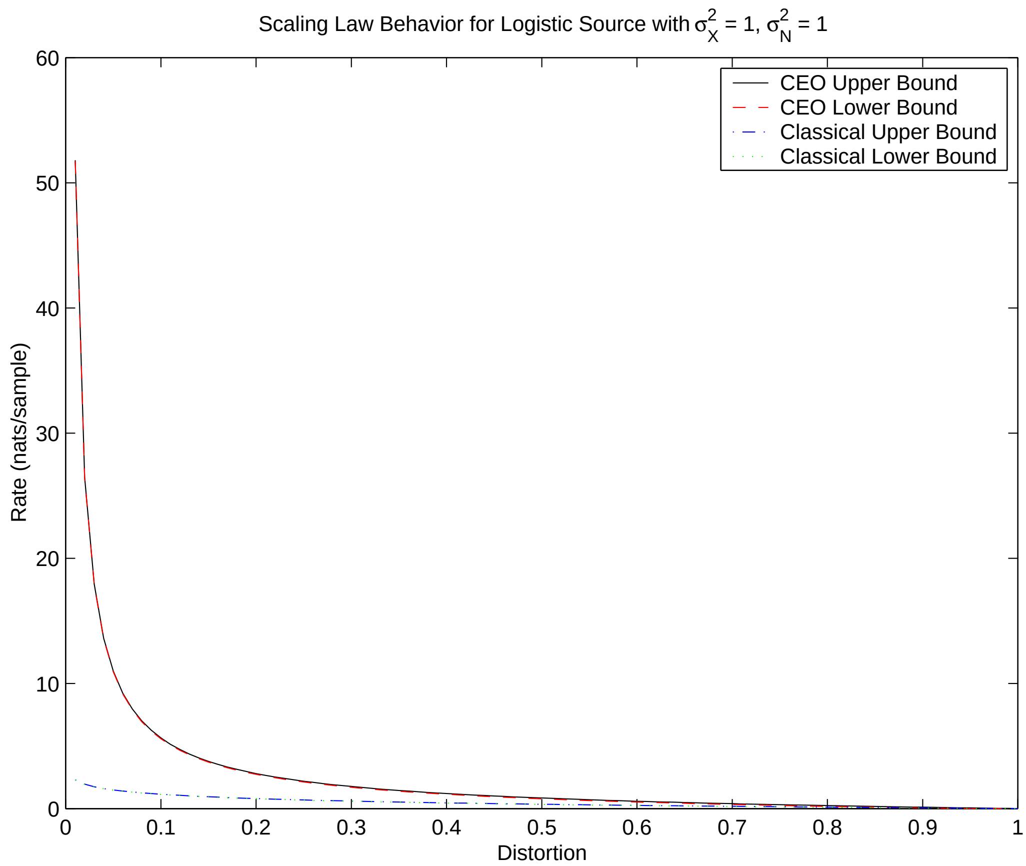

Figure 4.2. Scaling behavior for Logistic source in AWGN CEO Problem

We call this approach the "no binning" strategy and denote the minimal achievable sum-rate-distortion pairs by the function $R_X^{NB,M}(D)$ for M-observations. It turns out that for certain cases, the scaling behavior remains the same as the sum-rate-distortion function. We consider two such cases. The first is for the case of the quadratic AWGN CEO problem, which we have already considered. The second is based on a different CEO problem introduced by Wagner and Anantharam [11], [25]. The strategies that we consider are closely related to special cases of robust coding strategies considered by Chen et. al. [26].

4.4.1 Quadratic AWGN CEO Problem

Our first case involves the quadratic AWGN CEO problem. That is, the AWGN CEO problem with a squared error distortion. The coding strategy simply involves vector quantizing the observations and then performing an estimate at the decoder. This is similar to the coding strategy we used in our upper bound for the AWGN CEO sum-rate-distortion function, except now we have removed the binning stage.

Theorem 4.8. *When $Q_X > 0$ and the limit of the right-hand side of (3.24) exists as $M \to \infty$, then the minimal sum-rate distortion pairs achievable by this coding strategy is*

$$R_X^{NB,\infty}(D) \sim \frac{1}{D}. \tag{4.21}$$

Proof. The lower bound follows immediately from the previous lower bound in (4.18). Thus, it is simply a matter of providing an upper bound on the performance of codes with our structure. For such codes, we can show that random quantization arguments give

$$R = \sum_{i=1}^{M} I(Z_i; U_i), \tag{4.22}$$

$$D = E(X - E[X|U_1, \ldots, U_M])^2 \tag{4.23}$$

as an achievable sum-rate distortion pair for auxiliary random variables U_i satisfying $U_i \leftrightarrow Z_i \leftrightarrow X, U_{\{i\}^c}, \mathbf{Z}_{\{i\}^c}$. Defining $U_i = Z_i + W_i$ where the W_i are independent Gaussian random variables and applying the maximum entropy bound for $H(Z_1, \ldots, Z_M)$ [20, Thm. 9.6.5, p. 234], we get that

$$\begin{aligned} R_X^{NB,M}(D) \leq & \frac{M}{2} \log \left(1 + \frac{\frac{\sigma_X^2}{D} - 1}{M} \right) \\ & + \frac{M}{2} \log \left(\frac{M\sigma_X^2}{M\sigma_X^2 - \left(\frac{\sigma_X^2}{D} - 1 \right) \sigma_{\tilde{N}}^2} \right) \end{aligned} \tag{4.24}$$

Taking the limit as $M \to \infty$ gives

$$R_X^{NB,\infty}(D) \leq \frac{\sigma_X^2 + \sigma_{\tilde{N}}^2}{2} \left(\frac{1}{D} - \frac{1}{\sigma_X^2} \right). \tag{4.25}$$

Thus, we have that

$$\begin{aligned} \frac{\sigma_N^2}{2} & \leq \liminf_{D \to 0} D R_X^{NB,\infty}(D) \\ & \leq \limsup_{D \to 0} D R_X^{NB,\infty}(D) \leq \frac{\sigma_X^2 + \sigma_N^2}{2}, \end{aligned} \tag{4.26}$$

and we have proved the desired result. □

While the above shows that we can achieve the same scaling behavior, the performance loss can still be large in some instances. The following result bounds the performance loss.

Theorem 4.9. *Let $D_X^{NB,\infty}(R)$ denote the inverse function of $R_X^{NB,\infty}(D)$ and likewise for $D_X^{CEO,\infty}(R)$. Then, as $R \to \infty$,*

$$\begin{aligned} & 10 \log_{10} D_X^{NB,\infty} - 10 \log_{10} D_X^{CEO,\infty} \\ & \leq 10 \log_{10} \left(1 + \frac{\sigma_X^2}{\sigma_N^2} \right) \; dB. \end{aligned} \tag{4.27}$$

Proof. We can bound the performance loss for this robustness by rearranging (4.25) and (4.16) to get, for large enough R,

$$\begin{aligned} &10\log_{10} D_X^{NB,\infty}(R) - 10\log_{10} D_X^{CEO,\infty}(R) \\ &\leq 10\log_{10}\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right) + 10\log_{10}\left(\frac{R+J(X)\left(\frac{\sigma_N^2}{2}\right)}{R+\frac{1}{\sigma_X^2}\left(\frac{\sigma_X^2+\sigma_N^2}{2}\right)}\right) \text{ dB.} \end{aligned} \quad (4.28)$$

Thus, at high rates ($R \to \infty$), the performance loss is upper bounded by

$$10\log_{10}\left(1+\frac{\sigma_X^2}{\sigma_N^2}\right) \text{ dB.} \quad (4.29)$$

This completes the result □

For Gaussian sources, the above bound is valid for any choice of R. A summary for the performance loss for different SNR inputs at each sensor is given in Table 4.1.

SNR (dB)	Loss (dB)
10	10.41
5	6.19
1	3.54
0	3.01
-1	2.53
-5	1.19
-10	0.41

Table 4.1. Performance loss for a "no binning" coding strategy in the quadratic AWGN CEO problem.

4.4.2 Binary Erasure

For our second example, we consider a different CEO problem introduced by Wagner and Anantharam [11], [25]. In this problem, the source $X = \pm 1$, each with probability 1/2. Each of the encoders views the output of X through an independent binary erasure channel with crossover probability ϵ. Thus, $Z_i \in \{-1, 0, 1\}$. The distortion of interest to the CEO is

$$d(x,\hat{x}) = \begin{cases} K \gg 1, & \hat{x} \neq x, \hat{x} \neq 0 \\ 1 & \hat{x} = 0 \\ 0 & \hat{x} = x \end{cases}.$$

It turns out that as K gets large, the asymptotic sum-rate-distortion function for this binary erasure CEO problem is [25]

$$R_{BE}^{CEO,\infty}(D) = (1-D)\log 2 + \log\left(\frac{1}{D}\right)\log\left(\frac{1}{1-\epsilon}\right). \tag{4.30}$$

Theorem 4.10. *For the binary erasure the sum-rate distortion pairs achievable by a "no binning" strategy have the property*

$$R_{BE}^{NB,\infty}(D) \sim R_{BE}^{CEO,\infty}(D). \tag{4.31}$$

Proof. Since $R_{BE}^{NB,\infty}(D) \geq R_{BE}^{CEO,\infty}(D)$, the lower bound is clear. By random quantization arguments, we can show that for an appropriately chosen f,

$$R = \sum_{i=1}^{M} I(Z_i; U_i), \tag{4.32}$$

$$D = Pr\left(f(U_1, \ldots, U_M) = 0\right) \tag{4.33}$$

is an achievable sum-rate distortion pair for auxiliary random variables U_i satisfying $U_i \leftrightarrow Z_i \leftrightarrow X, U_{\{i\}^c}, \mathbf{Z}_{\{i\}^c}$. Defining $U_i = Z_i \cdot Q_i$ where $Q_i \in \{0,1\}$ are Bernoulli-q random variables, then D is just simply the probability that all the U_i are 0, which is just

$$D = \left(1 - (1-\epsilon)(1-q)\right)^M. \tag{4.34}$$

Taking the limit as $M \to \infty$ gives

$$R_X^{NB,\infty}(D) \leq \log\left(\frac{1}{D}\right) + \log\left(\frac{1}{D}\right)\log\left(\frac{1}{1-\epsilon}\right). \tag{4.35}$$

Thus, we have that

$$\limsup_{D\to 0} \frac{R_X^{NB,\infty}(D)}{R_{BE}^{CEO,\infty}(D)} \leq 1 + \left(\log\left(\frac{1}{1-\epsilon}\right)\right)^{-1}, \tag{4.36}$$

and we have proved the desired result. □

4.5 Discussion

In this chapter, we have presented bounds for the remote rate-distortion function and CEO sum-rate-distortion function as the number of observations increases. While the remote rate-distortion function converges to the direct rate-distortion function, there is still a rate loss asymptotically in the AWGN CEO problem. It turns out that even significantly suboptimal coding strategies can yield the same scaling behavior, leading one to question the sufficiency of scaling laws to characterize tradeoffs in such problems.

Chapter 5

Conclusion and Future Work

In 1959, Claude Shannon characterized the direct rate-distortion function and gave closed-form upper and lower bounds to it [1]. In this thesis, we presented extensions of these bounds to remote source coding problems. In particular, we considered the case in which the observations were corrupted by additive noise and the distortion was squared error.

We first gave bounds for the case of centralized encoding and decoding. Like Shannon's bounds for squared error distortion, the upper and lower bounds had similar forms with the upper bound matching the Gaussian remote rate-distortion function. The lower bound met the upper bound for the Gaussian source and entropy powers took the place of variances for non-Gaussian sources. Unlike previously known lower bounds for this problem, the lower bound in this problem was easier to compute for non-Gaussian sources.

We then gave bounds for the case of distributed encoding and centralized decoding, the so-called CEO problem. The lower bound appears to be the first non-trivial lower bound to the sum-rate-distortion function for AWGN CEO problems. We also considered two upper bounds for the sum-rate-distortion function. The second, while not as elegant as the first, proved to be more useful for certain non-Gaussian sources. Again, for the sum-rate-distortion function, the upper bounds and lower bounds matched for the case of the Gaussian source.

Using these bounds, we derived scaling laws for these problems. We found that while the case of centralized encoding and decoding could overcome the noise and converge to the direct rate-distortion function in the limit as the number of observations increases, the CEO sum-rate-distortion function converges to a larger function that has a different scaling behavior at low distortions. We also noted that one can still maintain this scaling behavior even none of the encoders take advantage of correlation among the different observations in the distributed coding scheme.

The results presented here pave the way for new research directions. One would be to consider the case in which the source or noise processes have memory. One can also consider how to handle different types of noise distributions as well as different distortions. Further, the bounds presented about the CEO problem appear to are related to the μ-sums problem considered by Wagner et. al. in [25]. Using the results given here, one might be able to provide similar results for the case of Gaussian mixtures.

References

[1] C. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Convention Rec.*, vol. 7, p. 142, 1959.

[2] T. Berger and J. Gibson, "Lossy source coding," *IEEE Transactions on Information Theory*, vol. 44, no. 6, p. 2693, 1998.

[3] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, p. 460, 1972.

[4] S. Arimoto, "An algorithm for calculating the capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, p. 14, 1972.

[5] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Transactions on Information Theory*, vol. 40, p. 1939, 1994.

[6] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IEEE Transactions on Information Theory*, vol. 8, p. 293, 1962.

[7] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Transactions on Information Theory*, vol. 16, pp. 406–411, 1970.

[8] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, ser. Information and System Sciences Series. Englewood Cliffs, NJ, USA: Prentice-Hall, 1971.

[9] ——, "Multiterminal source coding," in *Lecture Notes presented at CISM Summer School on the Information Theory Approach to Communications*, 1977.

[10] S. Tung, "Multiterminal source coding," Ph.D. dissertation, Cornell University, 1977.

[11] A. Wagner and V. Anantharam, "An improved outer bound for the multiterminal source coding problem," in *International Symposium on Information Theory*, 2005.

[12] A. B. Wagner, "Methods of offline distributed detection: Interacting particle models and information-theoretic limits," Ph.D. dissertation, University of California, Berkeley, 2005.

[13] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Transactions on Information Theory*, vol. 42, pp. 887–902, May 1996.

[14] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. 43, pp. 1549–1559, 1997.

[15] V. Prabhakaran, D. Tse, and K. Ramchandran, "Rate region of the quadratic Gaussian CEO problem," in *Proceedings of ISIT*, 2004.

[16] Y. Oohama, "Rate-distortion theory for Gaussian multiterminal source coding systems with several side informations at the decoder," *IEEE Transactions on Information Theory*, vol. 51, pp. 2577–2593, 2005.

[17] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, pp. 1261–1282, 2005.

[18] M. Pinsker, *Information and Information Stability of Random Variables and Processes.* San Francisco: Holden-Day, 1964, translated by A. Feinstein.

[19] A. Wyner, "The rate-distortion function for source coding with side information at the decoder II: General sources," *Information and Control*, vol. 19, pp. 60–80, 1978.

[20] T. Cover and J. Thomas, *Elements of Information Theory.* John Wiley and Sons, 1991.

[21] M. Gastpar, "A lower bound to the AWGN remote rate-distortion function," in *2005 Statistical Signal Processing Workshop, Bordeaux, France*, 2005.

[22] K. B. Housewright, "Source coding studies for multiterminal systems," Ph.D. dissertation, University of California, Los Angeles, 1977.

[23] J. Chen, X. Zhang, T. Berger, and S. B. Wicker, "An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem," *IEEE Journal on Selected Areas in Communications: Special Issue on Sensor Networks*, pp. 1–10, 2003.

[24] Y. Oohama, "Multiterminal source coding for correlated memoryless Gaussian sources with several side informations at the decoder," in *ITW*, 1999.

[25] A. B. Wagner, S. Tavildar, and P. Viswanath, "The rate region of the quadratic Gaussian two-terminal source-coding problem," 2005. [Online]. Available: http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:cs/0510095

[26] J. Chen and T. Berger, "Robust distributed source coding," *IEEE Transactions on Information Theory*, submitted.

[27] N. Blachman, "The convolution inequality for entropy powers," *IEEE Transactions on Information Theory*, vol. 11, p. 267, 1965.

[28] H. V. Poor, *Introduction to Signal Detection and Estimation*, ser. Texts in Electrical Engineering. New York: Springer-Verlag, 1994.

[29] M. Costa, "A new entropy power inequality," *IEEE Transactions on Information Theory*, vol. 31, p. 751, 1985.

[30] R. Durrett, *Probability: Theory and Examples, 3rd ed*, ser. Information and System Sciences Series. Duxbury, 2004.

[31] H. Witsenhausen, "On the structure of real-time source coders," *Bell System Technical Journal*, vol. 58, no. 6, pp. 1437–1451, 1979.

[32] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE Transactions on Information Theory*, vol. 42, pp. 2073–2084, Nov. 1996.

Appendix A

Lower Bounds on MMSE

In this appendix, we show that for a large class of sources corrupted by additive Gaussian noise have mean squared error decaying as $\Theta\left(\frac{1}{\text{snr}}\right)$ where snr is the signal-to-noise ratio of the source viewed in additive noise. The main conditions on the sources are that they have finite variance and are continuous.

Although we specialize our proofs to cases in which the noise is additive Gaussian, the same arguments work for any additive noise N with variance σ^2 for which $\frac{1}{2}\log\sigma^2 - h(N) = K$, where K is a constant that does not depend σ^2. The only thing that changes is the scaling constant. Table A.1 lists examples of sources for which this is true.

Distribution	Density	Variance	Entropy
Gaussian$(0,\sigma^2)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-x^2/2\sigma^2}$	σ^2	$\frac{1}{2}\log(2\pi e\sigma^2)$
Laplacian(λ)	$f(x) = \frac{\lambda}{2}e^{-\lambda\|x\|}$	$\sigma^2 = \frac{2}{\lambda^2}$	$\frac{1}{2}\log(2e^2\sigma^2)$
Uniform$(-a,a)$	$f(x) = \frac{1}{2a}1_{\{x\in[-a,a]\}}$	$\sigma^2 = \frac{1}{3}a^2$	$\frac{1}{2}\log(12\sigma^2)$
Two Pulse(c,ϵ)	$f(x) = \frac{1}{4c\epsilon}1_{\{x\in[-c(1+\epsilon),-c(1-\epsilon)] \cup [c(1-\epsilon),c(1+\epsilon)]\}}$	$\sigma^2 = c^2(1+\epsilon^2/3)$	$\frac{1}{2}\log\left(\frac{16\epsilon^2\sigma^2}{1+\epsilon^2/3}\right)$

Table A.1. Distributions and their differential entropies.

In the sequel, we denote our source as the random variable X with variance σ_X^2. X is corrupted by independent additive noise N with variance σ_X^2/s. This gives the signal-to-noise ratio s. We are interested in the performance of the estimator $E[X|X+N]$ as a function of s for fixed σ_X^2 .

Lemma A.1. *Define* $m(s) = E(X - E[X|X+N])^2$. *Then*

$$m(s) \leq \frac{\sigma_X^2}{s+1}. \tag{A.1}$$

Proof. Using the linear estimator $\hat{X} = \frac{\sigma_X^2}{\sigma_X^2+\sigma_X^2/s}(X+N)$, we get a mean squared error upper bound of

$$E(X - E[X|X+N])^2 \leq E(X - \hat{X})^2 = \frac{\sigma_X^2}{s+1}. \quad \text{(A.2)}$$

□

We have now established that the decay in MMSE is always at least as fast as s^{-1}. In fact, the bound in Lemma A.1 is tight for the Gaussian case. However, we are still left with the following question. Are there sources for which this decay is faster than s^{-1}? The answer to this question is yes. For discrete sources, the MMSE decays exponentially in s.

We now modify the question. Are there any continuous sources for which the decay is faster than s^{-1}? Consider the one in Figure A.1. While it is not discrete, its probability is concentrated around two regions. Despite this fact, it will turn out that even the source in Figure A.1 decays as s^{-1}. The key observation is to note that $H(X)$, the differential entropy of X, is finite.

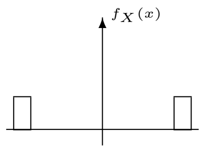

Figure A.1. Example of a source that decays as s^{-1}.

The following lemma will help us obtain this result.

Lemma A.2. *Let Q_X, Q_N, and Q_{X+N} be the normalized (by $2\pi e$) entropy powers of X, N, and $X+N$, respectively. Then*

$$E(X - E[X|X+N])^2 \geq \frac{Q_X Q_N}{Q_{X+N}}. \quad \text{(A.3)}$$

Proof. The result is related to a derivation of the Shannon lower bound for squared error distortion (see e.g. [8]). That is,

$$\begin{aligned} I(X; X+N) &= H(X) - H(X|X+N) && \text{(A.4)} \\ &= H(X) - H(X - E[X|X+N]|X+N) && \text{(A.5)} \\ &\geq H(X) - H(X - E[X|X+N]) && \text{(A.6)} \\ &\geq H(X) - \frac{1}{2}\log\left(2\pi e E(X - E[X|X+N])^2\right). && \text{(A.7)} \end{aligned}$$

However, since $I(X; X+N) = H(X+N) - H(N)$, we can rearrange and recollect terms in (A.7) to conclude the result. □

Theorem A.3. *If $H(X) > -\infty$, then for $N \sim \mathcal{N}(0, \sigma_X^2/s)$ independent of X, then*

$$m(s) = E(X - E[X|X+N])^2 = \Theta\left(s^{-1}\right). \tag{A.8}$$

Further,

$$m(s) \geq \frac{Q_X}{s+1}. \tag{A.9}$$

Proof. We already established the upper bound in Lemma A.1. Further, by Lemma A.2, we know that

$$\geq \frac{Q_X Q_N}{Q_{X+N}}. \tag{A.10}$$

Since N is Gaussian, its normalized entropy power $Q_N = \sigma_X^2/s$. Further, by the maximum entropy theorem under second moment constraints, $Q_{X+N} \leq \sigma_X^2 + \sigma_X^2/s$. With these facts, we can conclude the lower bound in (A.9). □

One can generalize the result easily to the following case.

Corollary A.4. *If $H(X|U) > -\infty$, then for $N \sim \mathcal{N}(0, \sigma_x^2/s)$ independent of X, U,*

$$m(s) = E(X - E[X|U, X+N])^2 = \Theta\left(s^{-1}\right). \tag{A.11}$$

Further,

$$m(s) \geq \frac{\frac{1}{2\pi e}\exp\{2H(X|U)\}}{s+1}. \tag{A.12}$$

Proof. The upper bound in Lemma A.1 continues to hold. We now have the bounds

$$H(X|U) - \frac{1}{2}\log(2\pi e m(s)) \leq I(X; X+N|U) \leq I(X, U; X+N). \tag{A.13}$$

where the lower bound follows as before. We get the upper bound by writing

$$\begin{aligned} I(X, U; X+N) &\leq H(X+N) - H(X+N|X, U) && \text{(A.14)} \\ &= H(X+N) - H(N|X, U) && \text{(A.15)} \\ &= H(X+N) - H(N) && \text{(A.16)} \\ &\leq \frac{1}{2}\log(1+s). && \text{(A.17)} \end{aligned}$$

Now, by the same arguments given in Theorem A.3, we can conclude (A.12) and thereby (A.11). □

Suppose now we consider a source that is either discrete or continuous conditioned on a random variable T. In particular, suppose a source X is 0 when $T = d$ and X has a pdf like the one in Figure A.1 when $T = c$. Can this scale faster than s^{-1}? The answer again turns out to be no.

Corollary A.5. *If $H(X|T=c) > -\infty$ and $Var(X|T=c) = \alpha\sigma_X^2$ for some $\alpha < \infty$, then for $N \sim \mathcal{N}(0, \sigma_X^2/s)$ independent of X, T,*

$$m(s) = E(X - E[X|X+N])^2 = \Theta\left(s^{-1}\right). \tag{A.18}$$

Further,

$$m(s) \geq P(T=c)\frac{\frac{1}{2\pi e\alpha}\exp\{2H(X|T=c)\}}{s+1}. \tag{A.19}$$

Proof. This follows immediately from Theorem A.3 by conditioning on the event $\{T=c\}$. □

Appendix B

Unified Entropy Power Inequality

We prove an inequality that specializes to the entropy power inequality in one case, Costa's entropy power inequality [29] in a second, and an entropy power inequality used by Oohama to prove a converse to the sum-rate-distortion function for Gaussian CEO problem in a third [16]. The generalization allows us to generalize his converse approach to give lower bounds to the sum-rate-distortion function for non-Gaussian sources in the AWGN CEO problem (see Chapter 3). It is also useful for new lower bounds for the remote rate-distortion function (see Chapter 2).

In the next section, we state the main result and give some interpretations of it. The remainder of this appendix is devoted to the proof. Our goal is to define a function on the positive reals that is the ratio between the right- and left-hand side of our inequality at 0 and show that it increases monotonically to 1, as in Figure B.1. To do this, we find derivatives of the differential entropies in our expressions and then establish inequalities about these derivatives. We then applies these inequalities on the function we defined and show that it has a positive derivative at all points.

B.1 Main Result

In this section, we state the main result. We also briefly describe how one can get other inequalities in this case.

Theorem B.1. $\mathbf{X}$ *and* $\mathbf{N}$ *are independent random vectors in* $\mathbf{R}^n$*. Let* $\mathbf{Z} = \mathbf{X} + \mathbf{N}$ *and require that* $\mathbf{X} \leftrightarrow \mathbf{Z} \leftrightarrow W$*, where* W *is some auxiliary random variable. Then*

$$\frac{e^{\frac{2}{n}H(\mathbf{Z})}}{e^{\frac{2}{n}I(\mathbf{X};W)}} \geq \frac{e^{\frac{2}{n}H(\mathbf{X})}}{e^{\frac{2}{n}I(\mathbf{Z};W)}} + e^{\frac{2}{n}H(\mathbf{N})}. \tag{B.1}$$

The original entropy power inequality follows when $\mathbf{Z}$ is independent of W. When $\mathbf{N}$ is Gaussian and $W = \mathbf{Z} + \mathbf{N}_2$, where $\mathbf{N}_2$ is additional Gaussian noise, we get

Costa's entropy power inequality [29]. Note that we can rewrite (B.1) as

$$e^{\frac{2}{n}H(\mathbf{X}|W)} \geq \frac{e^{\frac{2}{n}H(\mathbf{Z}|W)}}{\left(e^{\frac{2}{n}I(\mathbf{N};\mathbf{Z})}\right)^2} + \frac{e^{\frac{2}{n}H(\mathbf{N})}}{e^{\frac{2}{n}I(\mathbf{N};\mathbf{Z})}}. \tag{B.2}$$

For the special case when $\mathbf{X}$ and $\mathbf{N}$ are Gaussian, this becomes Oohama's entropy power inequality [16, eq. (48)]. It can be thought of as a "reverse" entropy power inequality in the sense that one of the addends in the traditional entropy power inequality is on the left-hand side of this inequality. Note that unlike his case, which follows from the relationship between independence and uncorrelatedness for Gaussian random variables, our result shows that the inequality holds simply by Bayes' rule. Another way of writing the inequality is

$$\frac{1}{n}I(\mathbf{X};W) \leq \frac{1}{n}I(\mathbf{X};\mathbf{Z}) - \frac{1}{2}\log\left(\frac{e^{-\frac{2}{n}I(\mathbf{N};\mathbf{Z})} + e^{-\frac{2}{n}I(\mathbf{X};Z)}e^{2I(\mathbf{Z};W)}}{e^{-\frac{2}{n}I(\mathbf{X};\mathbf{Z})}e^{\frac{2}{n}I(\mathbf{Z};W)}}\right). \tag{B.3}$$

This gives a tighter bound than the data processing inequality under additional structure assumptions.

B.2 Definitions and Notation

The proof follows along the lines of Blachman's proof [27] of the original entropy power inequality, except we treat the vector case directly. Before moving on to the proof, we introduce Fisher information and provide a few basic properties of it.

Definition B.2. Let $\mathbf{X}$ be a random vector in $\mathbf{R}^n$. If $\mathbf{X}$ has a pdf $f(\mathbf{x})$, the trace of the Fisher information matrix of $\mathbf{X}$ is

$$J(\mathbf{X}) = \int_{\mathbf{R}^n} \frac{\|\nabla f(\mathbf{x})\|^2}{f(\mathbf{x})} d\mathbf{x}. \tag{B.4}$$

This definition is based on the case in which one is interested in estimating means, so it differs slightly from [28]. The relationship is described in [20, p. 194]

Definition B.3. Let $\mathbf{X}$ be a random vector in $\mathbf{R}^n$ and W a random variable. If $\mathbf{X}$ has a conditional pdf $f(\mathbf{x}|w)$, the conditional Fisher information of $\mathbf{X}$ given W is

$$J(\mathbf{X}|W) = \int \mu(dw) \int_{\mathbf{R}^n} \frac{\|\nabla_{\mathbf{x}} f(\mathbf{x}|w)\|^2}{f(\mathbf{x}|w)} d\mathbf{x}. \tag{B.5}$$

One might recall de Bruijn's identity [20, Theorem 16.6.2]. The following is a statement of the vector version.

Lemma B.4. *Let $\mathbf{X}$ be a random vector in $\mathbf{R}^n$. We denote $\mathbf{X}_t = \mathbf{X} + \sqrt{t}\mathbf{V}$, where $\mathbf{V}$ is a standard normal Gaussian random vector. We denote the pdf of $\mathbf{X}_t$ with $f_t(\mathbf{x})$. Then*

$$\frac{d}{dt}H(\mathbf{X}_t) = \frac{1}{2}J(\mathbf{X}_t). \tag{B.6}$$

Proof. See [29]. □

The following is a conditional version of the identity.

Lemma B.5. *Let* $\mathbf{X}$ *be a random vector in* $\mathbf{R}^n$*. We denote* $\mathbf{X}_t = \mathbf{X} + \sqrt{t}\mathbf{V}$*, where* $\mathbf{V}$ *is a standard normal Gaussian random vector independent of* $(\mathbf{X}, W)$*. We denote the pdf of* $\mathbf{X}_t$ *with* $f_t(\mathbf{x})$*. Then,*

$$\frac{d}{dt}H(\mathbf{X}_t|W) = \frac{1}{2}J(\mathbf{X}_t|W). \tag{B.7}$$

Proof. Smoothing the distribution of $\mathbf{X}$ by a Gaussian $\sqrt{t}\mathbf{V}$ is equivalent to smoothing the conditional distribution of X on any realization $W = w$ by $\sqrt{t}\mathbf{V}$. A formal argument requires applying Fubini's theorem and the uniqueness of the Radon-Nikodym derivative [30], but the observation should be intuitively obvious. From this fact, we find that, for $\mathbf{U}$ independent of $(\mathbf{X}, \mathbf{V}, W)$,

$$0 \leq H(\mathbf{X}_t + \sqrt{h}\mathbf{U}|W = w) - H(\mathbf{X}_t|W = w) \leq \frac{1}{2}\log\left(\frac{t+h}{t}\right) \leq \frac{h}{2t}, \tag{B.8}$$

where the first inequality follows from the non-negativity of mutual information, the second from the data processing inequality, and the third from the inequality $\log(1 + x) \leq x$. Thus,

$$0 \leq \frac{H(\mathbf{X}_t + \sqrt{h}\mathbf{U}|W = w) - H(\mathbf{X}_t|W = w)}{h} \leq \frac{1}{2t}, \tag{B.9}$$

so by bounded convergence, we can swap the derivative and expectation over W to conclude the result. □

Lemma B.6. *Let* $\mathbf{X}$ *be a random vector in* $\mathbf{R}^n$*. We denote* $\mathbf{X}_t = \mathbf{X} + \sqrt{t}\mathbf{V}$*, where* $\mathbf{V}$ *is a standard Gaussian random vector independent of* $(\mathbf{X}, W)$*. Then,*

$$J(\mathbf{X}_t|W) \geq J(\mathbf{X}_t). \tag{B.10}$$

Proof. Let $\mathbf{U}$ be a standard Gaussian random vector independent of $(\mathbf{X}, \mathbf{V}, W)$. Then, by the data processing inequality,

$$I(\mathbf{X}_t; W) \geq I(\mathbf{X}_t + \sqrt{h}\mathbf{U}; W). \tag{B.11}$$

Thus, for all $h > 0$,

$$0 \geq \frac{I(\mathbf{X}_t + \sqrt{h}\mathbf{U}; W) - I(\mathbf{X}_t; W)}{h} \tag{B.12}$$

$$= \frac{H(\mathbf{X}_t + \sqrt{h}\mathbf{U}) - H(\mathbf{X}_t)}{h} - \frac{H(\mathbf{X}_t + \sqrt{h}\mathbf{U}|W) - H(\mathbf{X}_t|W)}{h}. \tag{B.13}$$

Letting $h \to 0$ and applying Lemmas B.4 and B.5 gives the desired result. □

B.3 Fisher Information Inequalities

We now consider Fisher information inequalities that will allow us to show that a function we define momentarily has a positive derivative at all points. The following inequality is sufficient to prove the classical entropy power inequality shown by Blachman [27].

Lemma B.7. *Let* $\mathbf{X}, \mathbf{N}$ *be independent random vectors and* $\mathbf{Z} = \mathbf{X} + \mathbf{N}$ *with differentiable, nowhere vanishing densities* $f_{\mathbf{X}}(\cdot)$*,* $f_{\mathbf{N}}(\cdot)$*, and* $f_{\mathbf{Z}}(\cdot)$*. Then*

$$\frac{1}{J(\mathbf{Z})} \geq \frac{1}{J(\mathbf{X})} + \frac{1}{J(\mathbf{N})} \tag{B.14}$$

Proof. See Blachman [27] for the scalar version. The vector proof is the same except derivatives are replaced by gradients. □

For our generalization of the entropy power inequality, we require an additional inequality that we now state.

Lemma B.8. *Let* $\mathbf{X}, \mathbf{N}$ *be independent random vectors and* $\mathbf{Z} = \mathbf{X}+\mathbf{N}$ *with differentiable, nowhere vanishing densities* $f_{\mathbf{X}}(\cdot)$*,* $f_{\mathbf{N}}(\cdot)$*, and* $f_{\mathbf{Z}}(\cdot)$*. If* $\mathbf{X} \leftrightarrow \mathbf{Z} \leftrightarrow W$ *for some random variable* W *and* $f_{\mathbf{X}|W}(\cdot|w)$ $f_{\mathbf{Z}|W}(\cdot|w)$ *are differentiable, nowhere vanishing conditional densities, then*

$$J(\mathbf{X}|W) - J(\mathbf{X}) \leq \frac{J(\mathbf{N})\left(J(\mathbf{Z}|W) - J(\mathbf{Z})\right)}{J(\mathbf{Z}|W) - J(\mathbf{Z}) + J(\mathbf{N})}. \tag{B.15}$$

Note that when $J(\mathbf{X}|W) > J(\mathbf{X})$ *and* $J(\mathbf{Z}|W) > J(\mathbf{Z})$*, we can rewrite (B.15) as*

$$\frac{1}{J(\mathbf{X}|W) - J(\mathbf{X})} \geq \frac{1}{J(\mathbf{Z}|W) - J(\mathbf{Z})} + \frac{1}{J(\mathbf{N})} \tag{B.16}$$

Proof. We follow a similar argument in Blachman [27] for Lemma B.7. By a straightforward application of Bayes' rule, we can write

$$f_{\mathbf{X}|W}(\mathbf{x}|w) = \int_{\mathbf{R}^n} \frac{f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} f_{\mathbf{Z}|W}(\mathbf{z}|w) d\mathbf{z}. \tag{B.17}$$

$$= f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbf{R}^n} \frac{f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} f_{\mathbf{Z}|W}(\mathbf{z}|w) d\mathbf{z}. \tag{B.18}$$

Since we have assumed differentiability, we can write the gradient as

$$\begin{aligned}\nabla f_{\mathbf{X}|W}(\mathbf{x}|w) = (\nabla f_{\mathbf{X}}(\mathbf{x})) &\int_{\mathbf{R}^n} \frac{f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} f_{\mathbf{Z}|W}(\mathbf{z}|w) d\mathbf{z} \\ + f_{\mathbf{X}}(\mathbf{x}) &\int_{\mathbf{R}^n} \frac{f_{\mathbf{Z}|W}(\mathbf{z}|w) f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})} \frac{\nabla f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})}{f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})} d\mathbf{z}.\end{aligned} \tag{B.19}$$

Dividing (B.19) by (B.18) gives

$$\frac{\nabla f_{\mathbf{X}|W}(\mathbf{x}|w)}{f_{\mathbf{X}|W}(\mathbf{x}|w)} = \frac{\nabla f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X}}(\mathbf{x})} + \int_{\mathbf{R}^n} \frac{f_{\mathbf{Z}|W}(\mathbf{z}|w) f_{\mathbf{N}}(\mathbf{x}-\mathbf{z}) f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Z}}(\mathbf{z}) f_{\mathbf{X}|W}(\mathbf{x}|w)} \frac{\nabla f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})}{f_{\mathbf{N}}(\mathbf{x}-\mathbf{z})} d\mathbf{z}. \tag{B.20}$$

$$= E\left[\frac{\nabla f_{\mathbf{X}}(\mathbf{X})}{f_{\mathbf{X}}(\mathbf{X})} + \frac{\nabla f_{\mathbf{N}}(\mathbf{N})}{f_{\mathbf{N}}(\mathbf{N})} \middle| \mathbf{X} = x, W = w \right]. \tag{B.21}$$

Observing that the integral in (B.18) is simply a convolution, the same arguments give us

$$\frac{\nabla f_{\mathbf{X}|W}(\mathbf{x}|w)}{f_{\mathbf{X}|W}(\mathbf{x}|w)} = E\left[\frac{\nabla f(\mathbf{X})}{f_{\mathbf{X}}(\mathbf{X})} + \frac{\nabla f_{\mathbf{Z}|W}(\mathbf{Z}|W)}{f_{\mathbf{Z}|W}(\mathbf{Z}|W)} - \frac{\nabla f_{\mathbf{Z}}(\mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{Z})} \middle| \mathbf{X} = x, W = w \right]. \tag{B.22}$$

We can combine (B.21) and (B.22) by the linearity of expectations to give

$$\begin{aligned} (a+b)\frac{\nabla f_{\mathbf{X}|W}(\mathbf{x}|w)}{f_{\mathbf{X}|W}(\mathbf{x}|w)} =& E\Bigg[(a+b)\frac{\nabla f(\mathbf{X})}{f_{\mathbf{X}}(\mathbf{X})} + a\frac{\nabla f_{\mathbf{N}}(\mathbf{N})}{f_{\mathbf{N}}(\mathbf{N})} \\ & + b\left(\frac{\nabla f_{\mathbf{Z}|W}(\mathbf{Z}|W)}{f_{\mathbf{Z}|W}(\mathbf{Z}|W)} - \frac{\nabla f_{\mathbf{Z}}(\mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{Z})} \right) \Bigg| \mathbf{X} = x, W = w \Bigg]. \end{aligned} \tag{B.23}$$

Conditional Jensen's inequality implies that

$$\begin{aligned} &(a+b)^2 \left\| \frac{\nabla f_{\mathbf{X}|W}(\mathbf{x}|w)}{f_{\mathbf{X}|W}(\mathbf{x}|w)} \right\|^2 \leq \\ &E\left[\left\| (a+b)\frac{\nabla f(\mathbf{X})}{f_{\mathbf{X}}(\mathbf{X})} + a\frac{\nabla f_{\mathbf{N}}(\mathbf{N})}{f_{\mathbf{N}}(\mathbf{N})} + b\left(\frac{\nabla f_{\mathbf{Z}|W}(\mathbf{Z}|W)}{f_{\mathbf{Z}|W}(\mathbf{Z}|W)} - \frac{\nabla f_{\mathbf{Z}}(\mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{Z})} \right) \right\|^2 \middle| \begin{array}{l} \mathbf{X} = \mathbf{x}, \\ W = w \end{array} \right]. \end{aligned} \tag{B.24}$$

It is straightforward to check that

$$E\left[\left\langle \frac{\nabla f_{\mathbf{Z}|W}(\mathbf{Z}|W)}{f_{\mathbf{Z}|W}(\mathbf{Z}|W)}, \frac{\nabla f_{\mathbf{Z}}(\mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{Z})} \right\rangle \right] = J(\mathbf{Z}),$$

$$E\left[\left\langle \frac{\nabla f_{\mathbf{Z}|W}(\mathbf{Z}|W)}{f_{\mathbf{Z}|W}(\mathbf{Z}|W)}, \frac{\nabla f_{\mathbf{N}}(\mathbf{N})}{f_{\mathbf{N}}(\mathbf{N})} \right\rangle \right] = E\left[\left\langle \frac{\nabla f_{\mathbf{Z}}(\mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{Z})}, \frac{\nabla f_{\mathbf{N}}(\mathbf{N})}{f_{\mathbf{N}}(\mathbf{N})} \right\rangle \right],$$

$$E\left[\left\langle \frac{\nabla f_{\mathbf{Z}|W}(\mathbf{Z}|W)}{f_{\mathbf{Z}|W}(\mathbf{Z}|W)}, \frac{\nabla f(\mathbf{X})}{f_{\mathbf{X}}(\mathbf{X})} \right\rangle \right] = E\left[\left\langle \frac{\nabla f_{\mathbf{Z}}(\mathbf{Z})}{f_{\mathbf{Z}}(\mathbf{Z})}, \frac{\nabla f(\mathbf{X})}{f_{\mathbf{X}}(\mathbf{X})} \right\rangle \right].$$

Taking the expectation of (B.23) and knowing these facts, we get

$$(a+b)^2 J(\mathbf{X}|W) \leq (a+b)^2 J(\mathbf{X}) + a^2 J(\mathbf{N}) + b^2 \left(J(\mathbf{Z}|W) - J(\mathbf{Z}) \right). \tag{B.25}$$

We can rewrite this as

$$J(\mathbf{X}|W) - J(\mathbf{X}) \leq \frac{a^2 J(\mathbf{N}) + b^2 \left(J(\mathbf{Z}|W) - J(\mathbf{Z}) \right)}{(a+b)^2}. \tag{B.26}$$

When $J(\mathbf{Z}|W) - J(\mathbf{Z}) = 0$, we show that (B.15) holds by making b arbitrarily large. When $J(\mathbf{Z}|W) - J(\mathbf{Z}) > 0$, we simply set $a = \frac{1}{J(\mathbf{N})}$ and $b = \frac{1}{J(\mathbf{Z}|W) - J(\mathbf{Z})}$ to give (B.15). □

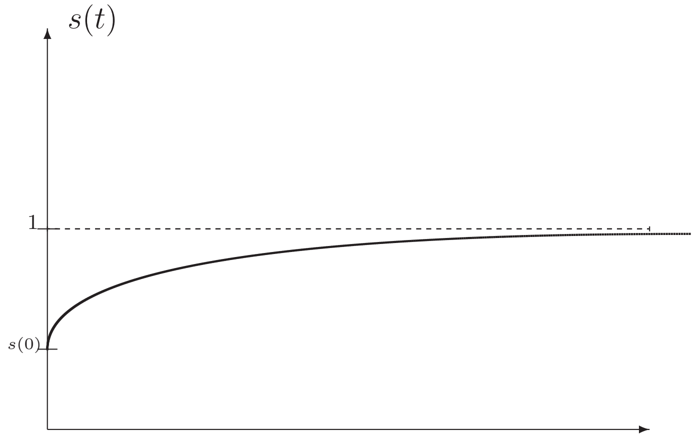

Figure B.1. Desired behavior of $s(t)$

B.4 Gaussian Smoothing

To finish the proof, the idea is that since we know the result is Gaussian, we smooth all the distributions until they approach a Gaussian, and then show that this smoothing can only increase the ratio. We smooth $\mathbf{X}$ by a an iid Gaussian vector $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, f(t)\mathbb{I})$ independent of $(\mathbf{X}, \mathbf{N}, W)$ and smooth $\mathbf{N}$ by an iid Gaussian vector $\mathbf{V} \sim \mathcal{N}(\mathbf{0}, g(t)\mathbb{I})$ independent of $(\mathbf{X}, \mathbf{N}, W, \mathbf{U})$, where $\mathbb{I}$ is the identity matrix. This induces smoothing on $\mathbf{Z}$ by an iid Gaussian vector of variance $h(t) = f(t) + g(t)$. We denote the smoothed random variables as $\mathbf{X}_f = \mathbf{X} + \mathbf{U}$, $\mathbf{N}_g = \mathbf{N} + \mathbf{V}$, and $\mathbf{Z}_h = \mathbf{Z} + \mathbf{U} + \mathbf{V}$, respectively. This will also result in smoothed the conditional distributions. We now define

$$s(t) = \frac{e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h; W)\right)} + e^{2/nH(\mathbf{N}_g)}}{e^{2/n\left(H(\mathbf{Z}_h) - I(\mathbf{X}_f; W)\right)}} \tag{B.27}$$

We let $f(0) = g(0) = h(0) = 0$. Thus, our goal is to show that $s(0) \leq 1$, which we do by showing that for $f'(t) = e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h; W)\right)}$ and $g'(t) = e^{2/nH(\mathbf{N}_g)}$, $s'(t) \geq 0$ and then showing that either $s(+\infty) = 1$ or $s(0) = 1$ trivially.

Lemma B.9. *Let $f(0) = g(0) = 0$, $f'(t) = e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h; W)\right)}$, and $g'(t) = e^{2/nH(\mathbf{N}_g)}$. Then, for all t,*

$$s'(t) \geq 0 \tag{B.28}$$

Proof. Taking the derivative,

$$e^{2/n\left(H(\mathbf{Z}_h) - I(\mathbf{X}_f; W)\right)} s'(t) = e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h; W)\right)} \left(f'(t)J(\mathbf{X}_f) + h'(t)(J(\mathbf{Z}_h|W) - J(\mathbf{Z}_h))\right)$$

$$+ e^{2/nH(\mathbf{N}_g)} g'(t) J(\mathbf{N}_g)$$
$$- \left[e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)} + e^{2/nH(\mathbf{N}_g)} \right]$$
$$\cdot (h'(t)J(\mathbf{Z}_h) + f'(t)(J(\mathbf{X}_f|W) - J(\mathbf{X}_f))). \tag{B.29}$$

Applying Lemmas B.7 and B.8 to the final term and factoring yields

$$e^{2/n\left(H(\mathbf{Z}_h) - I(\mathbf{X}_f;W)\right)} s'(t) \geq$$
$$\frac{(f'(t)J(\mathbf{X}_f) - g'(t)J(\mathbf{N}_g)) \left(e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)} J(X_f) - e^{2/nH(\mathbf{N}_g)} J(\mathbf{N}_g) \right)}{J(\mathbf{X}_f) + J(\mathbf{N}_g)}$$
$$+ (f'(t) + g'(t)) \frac{(J(\mathbf{Z}_h|W) - J(\mathbf{Z}_h))^2 e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)}}{J(\mathbf{Z}_h|W) - J(\mathbf{Z}_h) + J(\mathbf{N}_g)}$$
$$+ \left(e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)} g'(t) - e^{2/nH(\mathbf{N}_g)} f'(t) \right) \frac{J(\mathbf{N}_g)}{J(\mathbf{Z}_h|W) - J(\mathbf{Z}_h) + J(\mathbf{N}_g)} \tag{B.30}$$

Setting $f'(t) = e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)}$ and $g'(t) = e^{2/nH(\mathbf{N}_g)}$, this simplifies to

$$e^{2/n\left(H(\mathbf{Z}_h) - I(\mathbf{X}_f;W)\right)} s'(t) \geq \frac{\left(e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)} J(\mathbf{X}_f) - e^{2/nH(\mathbf{N}_g)} J(\mathbf{N}_g) \right)^2}{J(\mathbf{X}_f) + J(\mathbf{N}_g)} +$$
$$\left(e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)} + e^{2/nH(\mathbf{N}_g)} \right) \frac{(J(\mathbf{Z}_h|W) - J(\mathbf{Z}_h))^2 e^{2/n\left(H(\mathbf{X}_f) - I(\mathbf{Z}_h;W)\right)}}{J(\mathbf{Z}_h|W) - J(\mathbf{Z}_h) + J(\mathbf{N}_g)} \geq 0. \tag{B.31}$$

This establishes the result. □

Proof of Theorem B.1. With Lemma B.9, all that is left is to establish that $s(+\infty) = 1$ and when $s'(t) = 0$ for all t, then $s(0) = 1$ trivially. We establish the latter case first. The cases in which $s'(t) = 0$ for all t happen when any differential entropy or mutual information is infinite. Note that when any of $H(N)$, $H(X)$, and $H(Z)$ are infinite, the inequality is satisfied automatically if one appeals to the non-negativity of mutual information ($H(Z) \geq H(N)$ $H(Z) \geq H(X)$) and the data processing inequality ($I(\mathbf{X};W) \leq I(\mathbf{X};\mathbf{Z})$ and $I(\mathbf{X};W) \leq I(\mathbf{Z};W)$). Further, our inequality is an immediate consequence of the data processing inequality when $I(Z;W) = \infty$. Thus, we only need to consider the case in which these terms are finite.

Since $f'(t) > 0$ and increasing when $I(\mathbf{Z};W) < \infty$, the Gaussian smoothing eventually dominates. If the convergence is uniform, then $\frac{1}{n}H(\mathbf{X}_f) - \frac{1}{2}\log 2\pi e f \to 0$, $\frac{1}{n}H(\mathbf{N}_g) - \frac{1}{2}\log 2\pi e g \to 0$, and $\frac{1}{n}H(\mathbf{Z}_h) - \frac{1}{2}\log 2\pi e(f+g) \to 0$. Further, $\frac{1}{n}H(\mathbf{X}_f|W) - \frac{1}{2}\log 2\pi e f \to 0$, so $I(\mathbf{X}_f;W) \to 0$, and by the data processing inequality, $I(\mathbf{Z}_h;W) \to 0$.

To show that we can get the result with absolute convergence, we rely on a truncation argument. Consider a channel with independent input $\mathbf{X}(k)$ with density $f(\mathbf{x})$ and noise $\mathbf{N}(k)$ with density $g(\mathbf{x})$ at each time k. The output of the channel is

$\mathbf{Z}(k) = \mathbf{X}(k) + \mathbf{N}(k)$, then this output passes through another channel generating $W(k)$ at the output. The Markov chain $\mathbf{X}(k) \leftrightarrow \mathbf{Z}(k) \leftrightarrow W(k)$ holds. The mutual information rate for the first channel is $I(\mathbf{X};\mathbf{Z})$ and the mutual information rate of the second channel is $I(\mathbf{Z};W)$.

We now consider a second pair of channels that behave like the first, except now, if at any time k, the values of any $|X_i(k)|$, $|N_i(k)|$, $f(\mathbf{X}(k)|W(k))$, $g(\mathbf{N}(k))$ exceeds a value L, then the channels skip over time k without delay to time $k+1$. We define the probability that this event occurs as $1 - P(L)$. For this pair of channels, we know that the mutual information rates cannot exceed $I(\mathbf{X};\mathbf{Z})/P(L)$ and $I(\mathbf{Z};W)/P(L)$, respectively, since it can convey no more information than the first channel, and in $P(L)$ times the original number of channel uses.

Note that the inputs and noise for the second pair of channels is bounded with bounded densities. Here, we truncate smoothed versions of our random vectors. We use the notation $\hat{\mathbf{X}}_f$ to mean that $\mathbf{X}_f$ is truncated. Our truncation enables uniform convergence, so we get

$$\frac{e^{2/nH(\hat{\mathbf{Z}}_h)}}{e^{2/nI(\hat{\mathbf{X}}_f;W)}} \geq \frac{e^{2/nH(\hat{\mathbf{X}}_f)}}{e^{2I(\hat{\mathbf{Z}}_h;W)}} + e^{2/nH(\hat{\mathbf{N}}_g)} \tag{B.32}$$

$$\geq \frac{e^{2/nH(\hat{\mathbf{X}}_f)}}{e^{2/nI(\mathbf{Z}_h;W)/P(L)}} + e^{2/nH(\hat{\mathbf{N}}_g)} \tag{B.33}$$

for any $t \geq 0$. Since $H(\hat{\mathbf{Z}}_h) = I(\hat{\mathbf{X}}_f;\hat{\mathbf{Z}}_h) + H(\hat{\mathbf{N}}_g)$, we know that

$$H(\hat{\mathbf{Z}}_h) \leq \frac{H(\mathbf{Z}_h) - H(\mathbf{N}_g)}{P(L)} + H(\hat{\mathbf{N}}_g). \tag{B.34}$$

and find that

$$\frac{e^{2/n\left[\frac{H(\mathbf{Z}_h)-H(\mathbf{N}_g)}{P(L)}+H(\hat{\mathbf{N}}_g)\right]}}{e^{2/nI(\hat{\mathbf{X}}_f;W)}} \geq \frac{e^{2/nH(\hat{\mathbf{X}}_f)}}{e^{2/nI(\mathbf{Z}_h;W)/P(L)}} + e^{2/nH(\hat{\mathbf{N}}_g)} \tag{B.35}$$

For any $t > 0$, we can let $L \to \infty$, causing $P(L) \to 1$, $H(\hat{\mathbf{X}}_f) \to H(\mathbf{X}_f)$, $H(\hat{\mathbf{N}}_g) \to H(\mathbf{N}_g)$, and $H(\hat{\mathbf{X}}_f|W) \to H(\mathbf{X}_f|W)$, giving

$$\frac{e^{2/nH(\mathbf{Z}_h)}}{e^{2/nI(\mathbf{X}_f;W)}} \geq \frac{e^{2/nH(\mathbf{X}_f)}}{e^{2/nI(\mathbf{Z}_h;W)}} + e^{2/nH(\mathbf{N}_g)} \tag{B.36}$$

Finally, letting $t \to 0$, we get our result for the case in which the convergence is absolute. □

Appendix C

Sufficient Statistics in Remote Source Coding

In this appendix, we establish some basic results about sufficient statistics that allow us to simplify remote source coding problems with multiple observations.

C.1 Sufficient Statistics in Additive White Gaussian Noise

Lemma C.1. *Consider the M observations of a random variable X. For $1 \leq i \leq M$,*

$$Z_i = X + N_i, \tag{C.1}$$

where $N_i(k) \sim \mathcal{N}(0, \sigma_{N_i}^2)$. Now define

$$\tilde{Z} = \frac{1}{M} \sum_{i=1}^{M} \frac{\sigma_{\tilde{N}}^2}{\sigma_{N_i}^2} Z_i, \tag{C.2}$$

where $\sigma_{\tilde{N}}^2 = \frac{1}{\frac{1}{M}\sum_{i=1}^{M} \frac{1}{\sigma_{N_i}^2}}$ is the harmonic mean of the noise variances $\sigma_{N_i}^2$. Then, $\tilde{Z}$ is a sufficient statistic for X given $Z_1, Z_2, \ldots, Z_M$.

Proof. Since this a well known result, we only sketch the proof. One can whiten the noise, which makes the noise isotropic since the noise statistics are Gaussian. Projecting in the direction of the signal then gives the sufficient statistic. □

C.2 Remote Rate-Distortion Functions

We want to show that considering the sufficient statistic is equivalent to considering the original source give the same remote rate distortion function. That is, we want to show that we lose nothing in terms of the rate distortion sense by considering a sufficient statistic in the remote source coding problem.

Lemma C.2. *Given a sufficient statistic $T_i(Z_i)$ for a memoryless source X_i given a memoryless observation process Z_i (i.e., $X_i \to T_i(Z_i) \to Z_i$), then we have two equivalent single letter characterizations of the remote rate distortion function for a distortion measure $d(\cdot,\cdot)$.*

$$R^R(D) = \min_{U \in \mathcal{U}^R(D)} I(Z;U) = \min_{U \in \mathcal{U}'^R(D)} I(T;U), \tag{C.3}$$

$$\mathcal{U}'^R(D) = \Big\{ U : X \to T \to U, Ed(X,U) \leq D. \Big\}$$

Proof. This proof can be split into two parts. First, we show that if our encoder is a function on the $f(T(Z)^N)$ instead of $f(Z^N)$, the distortion can only be smaller. Based on an argument by Witsenhausen in [31], this is equivalent to showing that the conditional expectation of the distortion measure on Z^N is equal to the conditional expectation of the distortion measure on $T(Z)^N$. To avoid hairy notation in the below expression, we express everything in terms of the single letter terms. Since we are making the assumption that the source is memoryless and the observations are viewed by a memoryless channel, it is clear that this argument will continue to hold over blocks of any length.

$$\begin{aligned}
E[d(X,u)|Z=z] &= E[d(X,u)|Z=z, T(Z)=t(z)] && \text{(C.4)}\\
&= E[d(X,u)|T(Z)=t(z)] && \text{(C.5)}\\
&= E[d(X,u)|T=t(z)]. && \text{(C.6)}
\end{aligned}$$

Second, we show that considering the sufficient statistic will not change the rate. This is equivalent to showing that $I(Z;U) = I(T;U)$ since we know both rate distortion functions.

$$\begin{aligned}
I(Z;U) &= I(T(Z), Z; U) && \text{(C.7)}\\
&= I(T(Z);U) + I(Z;U|T(Z)) && \text{(C.8)}\\
&= I(T(Z);U) = I(T;U). && \text{(C.9)}
\end{aligned}$$

Thus, we can only do better with the sufficient statistic since it does not affect the rate and can only improve the distortion. Since the sufficient statistic is just a function of Z, this means that the rate distortion functions of the two problems are the same. ☐

Appendix D

Rate Loss

Recall the model considered for the remote source coding problem with multiple observations shown in Figure D.1 and the AWGN CEO problem in Figure D.2. In particular, consider the case in which the noise variances were equal. We could then describe the observation as

$$Z_i(k) = X(k) + N_i(k), \quad k \geq 1, \tag{D.1}$$

where $N_i(k) \sim \mathcal{N}(0, \sigma_N^2)$ and for $1 \leq i \leq M$. We derived upper and lower bounds for the (sum-)rate-distortion functions for these problems in Chapters 2 and 3. Clearly, the sum-rate-distortion function in the CEO problem is always at least as large as the remote rate-distortion function.

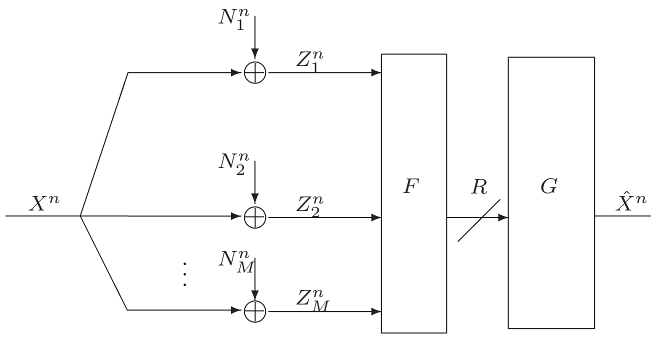

Figure D.1. Remote Source Coding Problem

In this appendix, we want to characterize the "gap" between these two functions. A simple way to find this gap or rate loss this is to take differences between the bounds derived in previous chapters. The disadvantage to such an approach is that the bounds are vacuous when the entropy power of the source is zero. We consider a novel approach introduced by Zamir [32] to determine an upper bound to the rate loss between the sum-rate-distortion function for the CEO problem and the remote rate-distortion function.

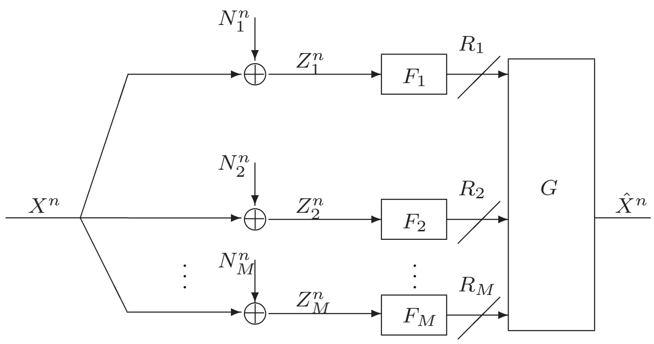

Figure D.2. AWGN CEO Problem

D.1 Definitions and Notation

Our results will apply to all difference distortions, which are distortions of the form $d(x, u) = f(x-u)$. For our purposes, it will be sufficient to consider the squared error distortion case.

Definition D.1. A **weak constraint** $E\mathbf{g}(\mathbf{Z} - \hat{X}) \leq \mathbf{K}(\rho)$ for some functions $\mathbf{g} : \mathcal{Z}^M \to \mathbf{R}^p$ and $\mathbf{K} : \mathbf{R}^q \to \mathbf{R}^p$ and for some $\mathbf{a}, \mathbf{b} \in \mathbf{R}^q$, is such that for all $\rho \in \mathbf{R}^q$ satisfying $a_i \leq \rho_i \leq b_i$, $Ed(X, \hat{X}) \leq D$ implies $E\mathbf{g}(\mathbf{Z} - \hat{X}) \leq \mathbf{K}(\rho)$.

Lemma D.2. *For* $d(x, \hat{x}) = (x - \hat{x})^2$,

$$E(Z_i - \hat{X})^2 = aD + 2b\sqrt{aD \cdot \sigma_N^2} + \sigma_N^2. \tag{D.2}$$

$$E(Z_i - \hat{X})(Z_j - \hat{X}) = aD + 2b\sqrt{aD \cdot \sigma_N^2}. \tag{D.3}$$

are weak constraints for all $i \neq j$.

Proof. We can write $E(X - \hat{X})^2 \leq D$ as $E(X - \hat{X})^2 = aD$ for some $0 \leq a \leq 1$. The Cauchy-Schwartz inequality implies

$$E(Z_i - \hat{X})^2 = E(X - \hat{X})^2 + 2b\sqrt{E(X - \hat{X})^2 \cdot \sigma_N^2} + \sigma_N^2,$$

$$E(Z_i - \hat{X})(Z_j - \hat{X}) = E(X - \hat{X})^2 + 2b\sqrt{E(X - \hat{X})^2 \cdot \sigma_N^2},$$

where $-1 \leq b \leq 1$. Note that we use the same choice of b for all values since $\sum_i Z_i$ is a sufficient statistic for X, and by symmetry of the Z_i distributions all the b will be the same. Lemma C.2 implies that $\sum_i Z_i$ will satisfy the same remote rate distortion function as $\mathbf{Z}$. □

Definition D.3. The **minimax capacity** with respect to a weak constraint $E\mathbf{g}(\mathbf{Z}-U) \leq \mathbf{K}(\rho)$,$a_i \leq \rho_i \leq b_i$ is

$$C(D) = \min_{\mathbf{W}\in\mathcal{L}(D)} C(D,\mathbf{W}), \tag{D.4}$$

where

$$\mathcal{L}(D) = \left\{\mathbf{W} : \mathbf{W} \perp\!\!\!\perp (X,U,\mathbf{Z}), \exists f, Ed(X, f(\mathbf{Z}+\mathbf{W})) \leq D\right\},$$

$$C(D,\mathbf{W}) = \max_{\substack{p(\mathbf{h}),\rho: E\mathbf{g}(\mathbf{H})\leq\mathbf{K}(\rho),\\ a_i\leq\rho_i\leq b_i}} I(\mathbf{H};\mathbf{H}+\mathbf{W}), \tag{D.5}$$

where $\perp\!\!\!\perp$ denotes statistical independence.

D.2 General Rate Loss Expression

We first find a bound on the rate loss in the CEO problem.

Theorem D.4. *We are given an i.i.d. source X, and M observations $Z_1,\ldots,Z_M$ viewed through a channel that is memoryless over time, and a difference distortion $d(\cdot,\cdot)$. The remote rate distortion function for this problem is described by*

$$R_X^R(D) = \min_{U\in\mathcal{X}_X^R(D)} I(Z_1,\ldots,Z_M;U). \tag{D.6}$$

Suppose there is a weak constraint $E\mathbf{g}(\mathbf{Z}-U) \leq \mathbf{K}(\rho)$ for $Ed(X,U) \leq D$. Then,

$$R_X^{CEO}(D) - R_X^R(D) \leq C(D). \tag{D.7}$$

The rest of this section is concerned with the proof of this theorem.

We begin with a lemma that generalizes Zamir's information inequality in [32]. This will lead to a general expression for the rate loss in the CEO problem. We then equate the terms in this expression to rate distortion quantities to get our rate loss bound.

Lemma D.5. *For any joint distribution on $(\mathbf{W},U,\mathbf{Z})$ such that $\mathbf{W}$ is independent of $(U,\mathbf{Z})$,*

$$I(\mathbf{Z};\mathbf{Z}+\mathbf{W}) - I(\mathbf{Z};U) \leq I(\mathbf{Z}-U;\mathbf{Z}-U+\mathbf{W}). \tag{D.8}$$

Proof.

$$
\begin{aligned}
& I(\mathbf{Z};\mathbf{Z}+\mathbf{W}) - I(\mathbf{Z};U) \\
& = -I(\mathbf{Z};U|\mathbf{Z}+\mathbf{W}) + I(\mathbf{Z};\mathbf{Z}+\mathbf{W}|U) && \text{(D.9)} \\
& \leq I(\mathbf{Z};\mathbf{Z}+\mathbf{W}|U) && \text{(D.10)} \\
& = I(\mathbf{Z}-U;\mathbf{Z}-U+\mathbf{W}|U) && \text{(D.11)} \\
& = I(\mathbf{Z}-U,U;\mathbf{Z}-U+\mathbf{W}) - I(U;\mathbf{Z}-\mathbf{U}+\mathbf{W}) && \text{(D.12)} \\
& \leq I(\mathbf{Z}-U,U;\mathbf{Z}-U+\mathbf{W}) && \text{(D.13)} \\
& = I(\mathbf{Z}-\mathbf{U};\mathbf{Z}-U+\mathbf{W}) + I(U;\mathbf{Z}-U+\mathbf{W}|\mathbf{Z}-U) && \text{(D.14)} \\
& = I(\mathbf{Z}-U;\mathbf{Z}-U+\mathbf{W}) + I(U;\mathbf{W}|\mathbf{Z}-U) && \text{(D.15)} \\
& = I(\mathbf{Z}-U;\mathbf{Z}-U+\mathbf{W}) + I(U,\mathbf{Z}-U;\mathbf{W}) - I(\mathbf{Z}-U;\mathbf{W}) && \text{(D.16)} \\
& = I(\mathbf{Z}-U;\mathbf{Z}-U+\mathbf{W}), && \text{(D.17)}
\end{aligned}
$$

where we justify the steps with
(D.9) the chain rule for mutual information;
(D.10) the non-negativity of mutual information;
(D.11) one-to-one transformations preserve mutual information;
(D.12) the chain rule for mutual information;
(D.13) the non-negativity of mutual information;
(D.14) the chain rule for mutual information;
(D.15) one-to-one transformations preserve mutual information;
(D.16) the chain rule for mutual information;
(D.17) our assumption about the joint distribution of $(\mathbf{W}, U, \mathbf{Z})$.

□

Our next step is to use this inequality to get a worst case upper bound on the rate loss in the CEO problem. To do this, we will make the second term on the left-hand side of (D.8) satisfy conditions to be the remote rate distortion function evaluated at distortion D and then the first term satisfy conditions for an inner (achievable) bound on the sum rate for distortion D.

We first choose $U \in \mathcal{X}_X^R(D)$ to minimize $I(U;\mathbf{Z})$ in (D.8). This gives us the remote-rate-distortion function in (1.28), and so

$$R_X^R(D) = I(Z_1, \ldots, Z_M; U). \tag{D.18}$$

With these constraints on U, we find from (D.8) that

$$I(\mathbf{Z};\mathbf{Z}+\mathbf{W}) - R_X^R(D) \leq I(\mathbf{Z}-U;\mathbf{Z}-U+\mathbf{W}). \tag{D.19}$$

Notice that we have the constraint $Ed(X, U) \leq D$. To simplify expressions, we want a series of constraints on $\mathbf{Z} - U$. To do this, we find a weak constraint by specifying functions $\mathbf{g}(\cdot)$, $\mathbf{K}(\cdot)$, $\mathbf{a}$, and $\mathbf{b}$.

We now allow ourselves to choose any distribution for $\mathbf{Z} - \mathbf{U}$ that satisfies the weak constraints. This can be larger than the right side of (D.19), and so

$$I(\mathbf{Z};\mathbf{Z}+\mathbf{W}) - R_X^R(D) \leq C(D, \mathbf{W}). \tag{D.20}$$

We now turn our attention to $I(\mathbf{Z}; \mathbf{Z}+\mathbf{W})$. To satisfy the conditions for our inner bound inequality in (3.2), we set $U_i = Z_i + W_i$ and require that $(W_1, \ldots, W_M)$ are mutually independent and independent of $(U, Z_1, \ldots, Z_M, X)$ to satisfy the Markov chain condition and $\mathbf{W}$ "small enough" to satisfy the distortion condition. A fortiori, this satisfies the assumption in Lemma D.5 for the information inequality to hold. Observe further that for difference distortions, $\mathbf{0} \in \mathcal{L}(D)$ whenever $\mathcal{L}(D)$ is nonempty, so we can always find a $\mathbf{W}$ such that

$$R_X^{CEO}(D) \leq I(Z_1, \ldots, Z_M; Z_1 + W_1, \ldots, Z_M + W_M). \tag{D.21}$$

From this, we conclude

$$R_{CEO,X}(D) - R_X^R(D) \leq C(D, \mathbf{W}). \tag{D.22}$$

If we minimize this mutual information over choices of $\mathbf{W} \in \mathcal{L}(D)$, we have proven the theorem.

Appendix E

Linear Algebra

This appendix gives the determinant of a specific type of matrix that is of particular interest in computing the determinant of a covariance matrix. This is most likely a well known result, so the derivation is provided here for completeness.

Lemma E.1. *For a matrix of the following form:*

$$A_n = \begin{bmatrix} a_0 + a_1 & a_0 \dots & & a_0 \\ a_0 & a_0 + a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_0 \\ a_0 & \dots & a_0 & a_0 + a_n \end{bmatrix}.$$

Then, for $n \geq 2$*,*

$$\det A_n = \sum_{k=0}^{n} \prod_{i \neq k} a_i. \tag{E.1}$$

Proof. Base case: $n = 2$,

$$\det \begin{bmatrix} a_0 + a_1 & a_0 \\ a_0 & a_0 + a_2 \end{bmatrix} = (a_0 + a_1)(a_0 + a_2) - a_0^2 \tag{E.2}$$

$$= a_0 a_1 + a_0 a_2 + a_1 a_2. \tag{E.3}$$

Induction hypothesis: Suppose this is true for some n and any choice of $a_0, \dots, a_n$. We will show it must hold for $n + 1$. Thus,

$$A_{n+1} = \begin{bmatrix} A_n & a_0 \\ a_0 \dots a_0 & a_0 + a_{n+1} \end{bmatrix}. \tag{E.4}$$

The determinant is defined recursively, so we can write

$$\begin{aligned}\det A_{n+1} =&(a_0+a_1)\det\begin{bmatrix} a_0+a_2 & a_0\dots & a_0 \\ a_0\dots a_0 & \ddots & a_0\dots a_0 \\ a_0\dots & a_0 & a_0+a_{n+1}\end{bmatrix} \\ &-a_0\sum_{i=2}^{n+1}(-1)^{2i}\det\begin{bmatrix} a_0 & a_0\dots a_0 \\ a_0 & B_{n,i}\end{bmatrix},\end{aligned} \tag{E.5}$$

where the second $(-1)^i$ comes from row swapping $B_{n,i}$ is the same as A_n except we replace all a_k with a_{k+1} for $k \geq i$. Note that all of these matrices are special cases of the matrices covered by our induction hypothesis. Thus, we get that

$$\det A_{n+1} = (a_0+a_1)\sum_{k=0,k\neq 1}^{n+1}\prod_{i\neq k,i\neq 1} a_i - a_0\sum_{j=2}^{n+1}\prod_{i\neq j} a_i \tag{E.6}$$

$$= a_0\prod_{i\neq 0,i\neq 1} a_i + a_1\sum_{k=0,k\neq 1}^{n+1}\prod_{i\neq k,i\neq 1} a_i \tag{E.7}$$

$$= \prod_{i\neq 1} a_i + \sum_{k=0,k\neq 1}^{n+1}\prod_{i\neq k} a_i \tag{E.8}$$

$$= \sum_{k=0}^{n+1}\prod_{i\neq k} a_i. \tag{E.9}$$

Thus, the formula is satisfied for $n+1$, and we conclude that it holds for all $n \geq 2$. □