RELATING INFORMATION-THEORETIC LIMITS TO THE
LYAPUNOV EXPONENT OF A DYNAMICAL SYSTEM

BY

HANI-JAMES M. EBEID

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010

Urbana, Illinois

Adviser:

Assistant Professor Todd P. Coleman

ABSTRACT

In this thesis we use control theoretic techniques to provide a new perspective for analyzing some problems in information theory. In particular, we explore two related data dissemination problems - channel coding with feedback and source coding with feedforward - and see that the Lyapunov exponent of a related dynamical system emerges as a fundamental quantity. For channel coding with feedback, we show that for a broad class of channels - both with and without memory - the Lyapunov exponent of the transmission function is fundamentally linked to the maximum rate which the scheme can attain. We note that the posterior matching scheme - a provably optimal feedback communication scheme for memoryless channels - has an encoding function with a Lyapunov exponent exactly equal to the communication rate. In the dual problem, source coding with feedforward, the optimal test channel is memoryless. This motivates the idea of dualizing posterior matching for this setting. By exploiting the Lyapunov exponent property, we demonstrate that such a scheme - with low decoder complexity - attains the rate-distortion function. By approaching these problems from a dynamical systems perspective, we hope to provide the intuition to motivate the evaluation and design of new communication schemes.

To my family and my friends

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

AWGN   Additive White Gaussian Noise

i.i.d.   Independent and Identically Distributed

L The Lyapunov Exponent of a System

$\bar{L}$ The Upper Lyapunov Exponent

PHR   Positive Harris Recurrent

PM   Posterior Matching

PMF   Probability Mass Function

$\langle W \rangle_{nR}$ The Quantized Version of the Message Point W

CHAPTER 1

INTRODUCTION

With the increasing complexity of control systems, there is growing interest in relating the areas of control and communication. Towards that end, in this thesis we use control theoretic techniques to provide a new perspective for analyzing some previously studied problems in information theory involving causal side information. In particular we explore two related problems - channel coding with feedback and source coding with feedforward - in which the Lyapunov exponent of a related dynamical system emerges as a fundamental quantity related to the theoretical limits of the problem.

In control systems, feedback provides inherent robustness to system uncertainties, adaption to unknown disturbances, and often a significant reduction in complexity, energy, or other costs. However, even when feedback is dramatically present, the study and use of feedback in information theory has not been explored in as much depth as other aspects of communication. Although the use of feedback information cannot increase the Shannon capacity in point to point communications over general memoryless channels [1], it can reduce the complexity of encoding schemes [2, 3, 4]. Considering the prevalence of recent applications involving many decentralized communication networks with centralized complexity nodes, there is strong motivation to reconsider how feedback should be used in communication systems.

Given that the natural mathematical framework to handle feedback is control theory, we consider the problem of communication over noisy channels with feedback from the dynamical systems perspective, and make use of recent sequential approaches to communication. This viewpoint has been made largely possible by a recent development in the information theory literature - the posterior matching (PM) scheme [5] - which generalizes other known feedback communication schemes where the communication problem was slightly augmented from the standard communication viewpoint [2, 3, 4]. Rather than nR bits, a message point on the interval $[0, 1]$ is considered.

The notion of "decoding nR bits" now becomes equivalent to determining the message point within an interval of length 2^{-nR} at the receiver.

Figure 1.1 on top shows an example of the posterior distribution $f_{W|Y^n}$ after n channel uses, and on bottom shows the discritization of this posterior distribution into the corresponding PMF. The interval $[0, 1]$ is partitioned into 2^{nR} equal-length segments, and the probability that the message point lies in each of these segments is computed. A maximum likelihood decoder selects the interval which has accumulated the most probability at time n. In order to decode nR bits reliably, the posterior distribution must contract fast enough so that arbitrarily large amounts of the a posteriori density accumulate in the same interval of length 2^{-nR} for large enough values of n.

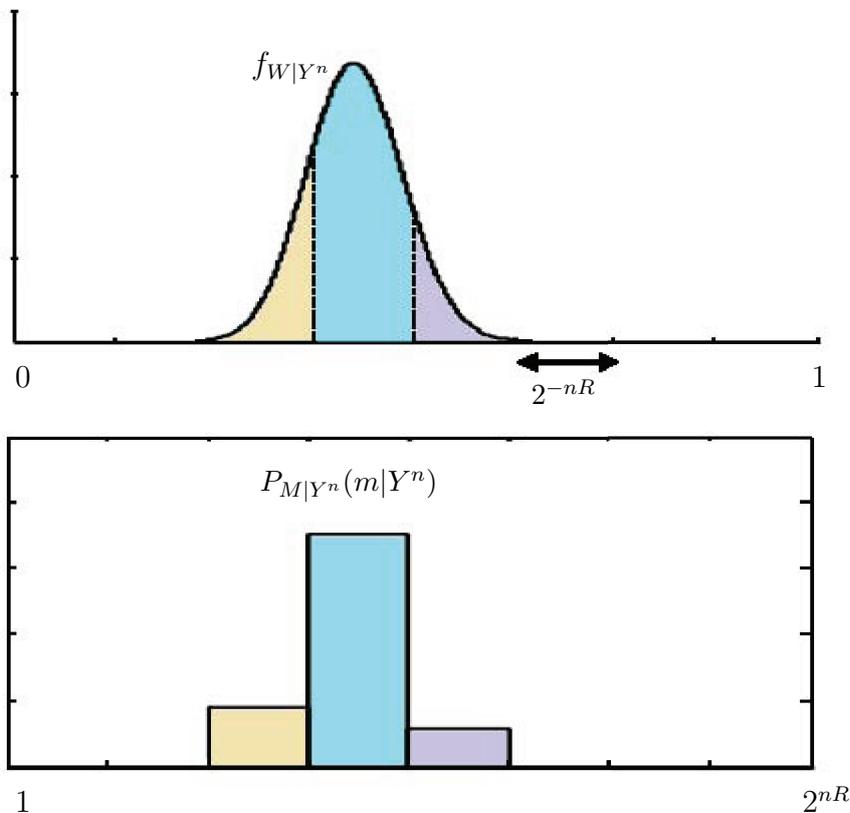Figure 1.1: Quantization of the posterior distribution at time n into 2^{nR} mass points.

The implementational details and fundamental limits are completely in line with traditional communication paradigms, but there are subtle, yet striking differences. Because the message is a point on the $[0, 1]$ line, there is no pre-specified block length; the system operates to sequentially give the user

the information that is "still missing" at the receiver. Moreover, the scheme is remarkably simple.

The high level idea is that the encoder first extracts the information missing at the receiver from the posterior distribution by generating a random variable that is statistically independent of past observations, but when coupled with those observations, deterministically produces the intended message. This information is then matched to the optimal input distribution of the channel to achieve capacity.

Additionally, a generic decoder exploits this recursive representation to provide linear complexity decoding as well. These nice properties for PM style communication schemes led us to analyze the problem of feedback communication under these settings from the viewpoint of control. In [6] it was demonstrated that "achievable rates" with the PM scheme can be interpreted from a stochastic control perspective by defining an appropriate Lyapunov function. Given that these types of schemes hold great promise to design dynamical systems based encoder and decoders for next generation communication systems with feedback, we endeavor to understand a necessary condition for any feedback communication system.

In this thesis we show that for a broad class of channels - including some with memory - the performance of any communication scheme over that channel is related to the Lyapunov exponent of the transmission function, when it exists. More generally, we can upper-bound the set of achievable rates for a given encoding scheme - with or without feedback - by considering a generalization of the Lyapunov exponent of the transmission function. We note that the PM scheme - a provably good feedback communication scheme for memoryless channels - has an encoding function with a Lyapunov exponent equal to the mutual information between channel inputs and outputs. Moreover, in the case of PM, the Lyapunov exponent is exactly equal to the communication rate. In some sense, this explains why the PM scheme is optimal.

The dual of channel coding is source coding [7, 8], and the dual of channel coding with feedback is source coding with feedforward [9]. Source coding with feedforward is a special case of source coding with side information, which was first introduced in [10] for distributed source coding applications. The scenario for reproducing a source with causal side information was introduced by Weissman and Merhav [11] as a competitive prediction problem.

Around the same time, Pradhan and others also began studying source coding with side information, with a particular emphasis on source coding with feedforward information [9, 12, 13]. This model is useful, for example, in sensor networks where some sensors may have information about a random measurement, but they are limited to by power or bandwidth constraints to send only a compressed version of this information.

For source coding with feedforward, an encoder compresses an i.i.d. source into a message, and the decoder takes this message, along with causal noiseless side information about the source, to construct an estimate. Analogous to channel coding with feedback over memoryless channels, with additive distortion measures and i.i.d. sources, feedforward does not change the rate-distortion function. Given that the optimal test channel in source coding with feedforward is a memoryless channel, this motivates the idea of dualizing posterior matching for this setting. By exploiting the Lyapunov exponent property, we demonstrate that such a scheme, with low encoder and decoder complexity, attains the rate-distortion function. By approaching these problems from a dynamical systems perspective, we hope to provide the intuition to motivate the evaluation and design of new communication schemes.

In Chapter 2 we present both problems which are to be considered: channel coding with feedback and source coding with feedforward. We also define the Lyapunov exponent and other important concepts which we will need. In Chapter 3, we focus on channel coding with feedback and develop a necessary condition for an encoding scheme to achieve a given rate. This work also appears in [14]. Chapter 4 concerns the dual problem of source coding with feedforward and develops a provably good scheme based on the posterior matching encoder. This work was presented in [15]. In Chapter 5, we conclude with a discussion.

CHAPTER 2

NOTATIONS AND PRELIMINARIES

Channel coding with feedback and source coding with feedforward are closely related problems, which in some sense are duals of each other [9]. In this chapter, we present the basic model for each problem and define some important concepts which will be needed, including the Lyapunov exponent of a dynamical system.

2.1 Channel Coding with Feedback

A general memoryless channel with noiseless feedback is depicted in Figure 2.1. We assume that the encoder has instantaneous feedback at time n of all channel outputs up to time $n-1$.
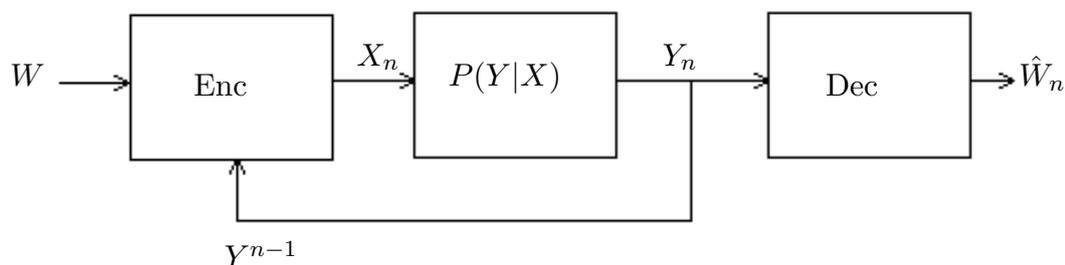

Figure 2.1: Communication system with feedback.

- Let W be a random message point distributed uniformly over the interval $(0, 1)$, representing an infinite sequence of bits to be communicated across the noisy channel:

$$f_W(w) = 1 \ , \ w \in [0, 1] \, . \tag{2.1}$$

- Denote the input alphabet of the channel by $\mathcal{X}$ and the output alphabet by $\mathcal{Y}$.

- We specify the statistical nature of the channel from X to Y with input $x_i \in \mathcal{X}$ and output $y_i \in \mathcal{Y}$ at time i in terms of its transition kernel

$$P_{Y_i|X_i,X^{i-1},Y^{i-1}}(y_i|x_i, x^{i-1}, y^{i-1}). \tag{2.2}$$

The "simplest" class of channels, termed *memoryless* channels, has the property

$$P_{Y_i|X_i,X^{i-1},Y^{i-1}}(y_i|x_i, x^{i-1}, y^{i-1}) = P_{Y|X}(y_i|x_i). \tag{2.3}$$

The PM scheme presented in Section 2.1.3 is optimal for memoryless channels. This fact will be exploited in Chapter 4; however, in Chapter 3 we will work in the more general framework of (2.2).

- Our encoding scheme specifies the sequence of transmission functions which are used to determine the channel inputs

$$X_i = g_i(W, Y^{i-1}). \tag{2.4}$$

Define

$$\partial_w g_i(u, Y^{i-1}) \triangleq \left. \frac{\partial}{\partial w} g_n\left(w, Y^{n-1}\right) \right|_{w=u} \tag{2.5}$$

when the derivative exists.

- Our decoding scheme specifies the sequence of estimates of the message point, based on the received channel outputs:

$$\hat{W}_n = \varrho_n(Y^n). \tag{2.6}$$

- The quantization of a message $W \in [0,1]$ is defined by

$$\langle W \rangle_{nR} \triangleq \frac{\lceil W 2^{nR} \rceil}{2^{nR}} \tag{2.7}$$

and the corresponding index is given by $m \triangleq \langle W \rangle_{nR} 2^{nR}$.

- The *lim sup in probability* of a sequence of random variables $\{X_n\}$ is defined by [16]:

$$\limsup_{\text{in p.}} X_n = \frac{\arg\min}{b} \, s.t. \lim_{n\to\infty} \mathbb{P}\left(X_n \geq b\right) = 0. \qquad (2.8)$$

- Similarly, the *lim inf in probability* of $\{X_n\}$ is [16]:

$$\liminf_{\text{in p.}} X_n = \frac{\arg\max}{a} \, s.t. \lim_{n\to\infty} \mathbb{P}\left(X_n \leq a\right) = 0. \qquad (2.9)$$

2.1.1 Sequential Communication with Feedback

In this section, we specify a non-standard yet illuminating approach to reliable communication with feedback. Traditionally, one assumes that there is a fixed block length of n transmissions, and considers a message lying in one of 2^{nR} possible values, and a coding scheme is designed that maps the possible hypotheses and causal feedback to the next channel input. At time n, the decoder attempts to decode the message.

Here, we consider an alternative approach that has the same fundamental limits. This approach was originated by Horstein to achieve capacity on the binary symmetric channel with feedback [4], applied to achieve capacity on the additive Gaussian channel with feedback in [2], and then generalized to achieve capacity on arbitrary memoryless channels with feedback in [5] - which unifies all previous approaches with a simple recursive interpretation. This approach for feedback communication is particularly attractive practically for the following reasons: there is no pre-specified rate or block length; there is no forward error correction - the encoder can simply adapt on the fly based upon feedback from the decoder; the schemes can admit a simple recursive structure; and the fundamental limit - or capacity - is the same as the more traditional viewpoint. In essence, the question of a rate being achievable is decoder-centric - at time n, it is required that the decoder can resolve the message point to one of 2^{nR} non-overlapping intervals, each of length 2^{-nR}.

The *a posteriori distribution* at time n is the conditional probability distribution on the message point, given the channel outputs y^n, and is denoted by $f_{W|Y^n}(u|y^n)$. For any good communication scheme, as the receiver at-

tains more channel outputs, its posterior belief becomes more concentrated around the message point. In order to communicate nR bits, the decoder at the receiver at time n must be able to guess which interval of length 2^{-nR} contains the message point. The probability of error is the probability that the true message point lies outside of the guessed interval. We will denote the midpoint of the interval chosen by the decoder at time n as $\hat{W}_n$. Then we say that:

Definition 1. *A rate R is achievable if there exists a decoding scheme such that*

$$\lim_{n\to\infty} \mathbb{P}\left(|W - \hat{W}_n| > 2^{-(nR+1)}\right) = 0.$$

That is to say, the probability that the intended message W lies outside of the decoded interval goes to zero.

A simple way to interpret this is that, for any $\delta > 0$, it should be that with high probability,

$$\int_{W-2^{-(nR+1)}}^{W+2^{-(nR+1)}} f_{W|Y^n}(u|Y^n)du \geq 1 - \delta \qquad (2.10)$$

for sufficiently large n.

This implies that the probability mass of the posterior distribution must be concentrated within an interval around the message point W of exponentially decreasing width if we are to communicate reliably at a positive rate. This implies that the height of the posterior must be exponentially increasing as well. In fact, the value of the posterior evaluated at the message point must increase exponentially at a rate R for large n with high probability. This lemma follows closely from [17, Thm 4] under the assumption that the channel law (2.2) is such that the the posterior distribution will be continuous in the message point W.

Lemma 2.1.1. *If a rate R is achievable then*

$$\liminf_{in\ p.} \frac{1}{n} \log f_{W|Y^n}(W|Y^n) \geq R.$$

8

Proof. Define the sets

$$
\begin{aligned}
B_m^n &\triangleq \{y^n \in \mathcal{Y}^n : P_{M|Y^n}(m|y^n) \le 2^{-n\gamma}\} \\
D_m &\triangleq \{y^n \in \mathcal{Y}^n : \left\langle \hat{W}_n \right\rangle_{nR} = \frac{m}{2^{nR}}\}
\end{aligned}
$$

for some $\gamma > 0$. Now consider the conditional probability distribution on the quantized version of the message, given the sequence of received channel outputs

$$
\begin{aligned}
&\mathbb{P}\left(\frac{1}{n} \log \frac{P_{M|Y^n}(\langle W \rangle_{nR} | Y^n)}{P_M(\langle W \rangle_{nR})} \le R - \gamma\right) \\
&= \mathbb{P}\left(\frac{1}{n} \log P_{M|Y^n}(\langle W \rangle_{nR} | Y^n) \le -\gamma\right) \quad\quad (2.11) \\
&= \sum_{m=1}^{2^{nR}} \sum_{y^n \in B_m^n} P_{M,Y^n}(m, y^n) \\
&= \sum_{m=1}^{2^{nR}} \sum_{y^n \in B_m^n \cap D_m^c} P_{M,Y^n}(m, y^n) + \sum_{m=1}^{2^{nR}} \sum_{y^n \in B_m^n \cap D_m} P_{M,Y^n}(m, y^n) \\
&\le \sum_{m=1}^{2^{nR}} \sum_{y^n \in D_m^c} P_{Y^n|M}(y^n|M=m) P_M(m) + \sum_{m=1}^{2^{nR}} \sum_{y^n \in B_m^n \cap D_m} P_{M,Y^n}(m, y^n) \\
&\le P_{error} + \sum_{m=1}^{2^{nR}} \sum_{y^n \in B_m^n \cap D_m} P_{Y^n}(y^n) P_{M|Y^n}(m|y^n) \\
&\le P_{error} + 2^{-n\gamma} \sum_{y^n \in \mathcal{Y}^n} P_{Y^n}(y^n) \quad\quad (2.12) \\
&\le P_{error} + 2^{-n\gamma}.
\end{aligned}
$$

Here, (2.11) follows because the original random variable was uniform, so its quantization is equiprobable over all indices, and (2.12) follows because the sets D_m are disjoint. Because the rate is achievable, the probability of error can be made arbitrarily small and

$$
\mathbb{P}\left(\frac{1}{n} \log \frac{P_{M|Y^n}(\langle W \rangle_{nR} | Y^n)}{P_M(\langle W \rangle_{nR})} \le R\right) \to 0
$$

as n increases.

When the posterior distribution is continuous, the desired result concerning the posterior distribution follows from this analysis of quantized posterior.

Note that since R is an achievable rate, with high probability for any $\delta > 0$, when n is large enough

$$\int_{\langle w \rangle_{nR} - 2^{-nR}}^{\langle w \rangle_{nR}} f_{W|Y^n}(u|y^n)du \;\; = \;\; P_{M|Y^n}(\langle w \rangle_{nR}|y^n) \qquad (2.13)$$

$$\geq \;\; 1 - \delta. \qquad (2.14)$$

By the mean value theorem, almost surely $\exists u \in \left[\langle w \rangle_{nR} - 2^{-nR}, \langle w \rangle_{nR}\right]$ such that $f_{W|Y^n}(u|y^n) \geq (1-\delta)2^{nR}$ and $|u - w| < 2^{-nR}$. By our continuity assumption, as n increases $f_{W|Y^n}(w|y^n) \to f_{W|Y^n}(u|y^n)$, establishing the desired result.

$\square$

2.1.2 A Perspective from the Converse to the Channel Coding Paradigm with Feedback

We note from the converse to the channel coding theorem with feedback that in order to achieve reliable communication at any rate R, it must be that $R \leq \frac{1}{n}I(W;Y^n) + o(n)$, with

$$\frac{1}{n}I(W;Y^n) \;\; = \;\; \frac{1}{n}\sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, W)$$

$$= \;\; \frac{1}{n}\sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, W, X_i)$$

$$= \;\; \frac{1}{n}\sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|X_i) \qquad (2.15)$$

$$\leq \;\; \frac{1}{n}\sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) \qquad (2.16)$$

$$\leq \;\; C \qquad (2.17)$$

where (2.15) follows from the channel being memoryless (2.3); and (2.16) follows if and only if Y_i's are i.i.d and drawn according to P_X^*.

The posterior matching scheme [18, 19] generalizes other feedback communication schemes [4, 2, 3] to provide a recursive scheme for a broad class of memoryless channels with noiseless feedback, and it tightens the converse

inequalities above.

2.1.3 The Posterior Matching Scheme

The posterior matching scheme [5] generalizes other feedback communication schemes [4, 2] to provide a recursive scheme for a broad class of memoryless channels with noiseless feedback. The high level idea is that the encoder first extracts the information missing at the receiver from the a-posteriori probability distribution by generating a random variable that is statistically independent of past observations, but when coupled with those observations, deterministically produces the intended message. This information is then matched to the optimal input distribution of the channel, F_X, to achieve capacity.

This idea is encapsulated by the posterior matching scheme by determining the channel input at time $n + 1$ by the transmission rule

$$X_{i+1} = g_{i+1}(W, Y^i) = F_X^{-1}\left(F_{W|Y^i}\left(W|Y^i\right)\right). \tag{2.18}$$

Note that because $F_{W|Y^n}\left(W|Y^n\right)$ is distributed uniformly on $[0, 1]$, regardless of the sequence Y^n, it follows that [5]

- X_{i+1} is independent of Y^n and so, due to the memoryless nature of the channel, Y_{i+1} is independent of Y^i.

- The marginal distribution on X_{i+1} is P_X, the capacity-achieving distribution. In particular, $\{Y_i\}$ are i.i.d.

The encoder also admits a simple recursive representation which is completely determined by the channel input distribution and the channel transition law, obviating the computationally infeasible task of computing the posterior distribution at each time step:

$$
\begin{aligned}
X_i &= F_X^{-1}(W_i) \\
W_1 &= W \\
W_{i+1} &= S_{Y_i}(W_i).
\end{aligned}
\tag{2.19}
$$

This simple scheme is optimal in the sense that it is able to achieve any rate below capacity for a broad class of memoryless channels. To understand why this scheme performs so well, let us first define the Lyapunov exponent of a dynamical system.

2.2 Lyapunov Exponents

The Lyapunov exponent of a dynamical system indicates how quickly the state of a system diverges from its initial condition. For a channel encoder, we can consider the state of the system at time n to be the channel input that it chooses at that time. Then the Lyapunov exponent of the encoder is given by

$$L = \lim_{n \to \infty} \frac{1}{n} \log \left| \partial_w g_n(W, Y^{n-1}) \right|, \tag{2.20}$$

where the limit is interpreted in the probability sense.

2.2.1 The Upper Lyapunov Exponent

In general, this limit in equation (2.20) may not exist, and we may wish to consider a generalization of the limit. Let us define the upper Lyapunov exponent as follows:

$$\bar{L} = \limsup_{\text{in p.}} \frac{1}{n} \log \left| \partial_w g_n(W, Y^{n-1}) \right|. \tag{2.21}$$

Similarly, the lower Lyapunov exponent is defined by the liminf. When the limit in (2.20) exists, the upper Lyapunov exponent in (2.21) and the lower Lyapunov exponent will be equal to the Lyapunov exponent in (2.20).

2.2.2 Lyapunov Exponent of the PM Scheme

If we consider the channel input at time i to be the state of a dynamical system, then the Lyapunov exponent of this encoding system is a measure of

the divergence between sequences of channel inputs corresponding to different messages. If the Lyapunov exponent is positive, two different messages which are relatively close to each other will still result in increasingly different input sequences to the channel.

To determine the Lyapunov exponent of the PM scheme, we first consider the following result regarding the asymptotic behavior of the posterior distribution, evaluated at the message point [18, lemma 2]:

Lemma 2.2.1. *For the posterior matching scheme, when the Markov chain defined by the channel inputs is positive Harris recurrent (PHR)*

$$\lim_{n \to \infty} \frac{1}{n} \log f_{W|Y^n}(w|y^n) = I(X;Y).$$

Proof. First note that applying Bayes' rule to the posterior distribution yields:

$$
\begin{aligned}
f_{W|Y^n}(w|y^n) &= f_{W|Y^{n-1}}(w|y^{n-1}) \frac{f_{Y_n|W,Y^{n-1}}(y_n|w,y^{n-1})}{f_{Y_n|Y^{n-1}}(y_n|y^{n-1})} \\
&= \prod_{i=1}^{n} \frac{f_{Y_i|W,Y^{i-1}}(y_i|w,y^{i-1})}{f_{Y_i|Y^{i-1}}(y_i|y^{i-1})}.
\end{aligned}
$$

Now taking the logarithm and dividing by n we get

$$
\begin{aligned}
\frac{1}{n} \log f_{W|Y^n}(w|y^n) &= \frac{1}{n} \log \prod_{i=1}^{n} \frac{f_{Y_i|W,Y^{i-1}}(y_i|w,y^{i-1})}{f_{Y_i|Y^{i-1}}(y_i|y^{i-1})} \\
&= \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_{Y_i|W,Y^{i-1}}(y_i|w,y^{i-1})}{f_{Y_i|Y^{i-1}}(y_i|y^{i-1})} \\
&= \frac{1}{n} \sum_{i=1}^{n} \log \frac{f_{Y_i|X_i}(y_i|g_i(w,y^{i-1}))}{f_{Y_i}(y_i)} \qquad (2.22)
\end{aligned}
$$

where (2.22) follows because the channel is memoryless, and the posterior matching scheme results in i.i.d. channel outputs. Now we note that the quantity on the right hand side is the information density. Taking limits we apply the strong law of large numbers to arrive at the desired result with probability one. $\square$

Now we state the following lemma from [18, lemma 3] describing the Lyapunov exponent of the transmission function when the posterior matching

scheme is used:

Lemma 2.2.2. *When the Markov chain (X_i, Y_i) is PHR, the Lyapunov exponent of the encoding system is equal to the mutual information between the channel input and output:*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{\partial g_n(w, Y^{n-1})}{\partial w} \Big|_{w=W} = I(X;Y) \ \ a.s. \tag{2.23}$$

Proof. By taking the derivative of the recursive transmission function (2.19), we have

$$\frac{\partial S(x,y)}{\partial x} = \frac{f_{X|Y}(x|y)}{f_X(S(x,y))}. \tag{2.24}$$

Then the derivative with respect to the message point is

$$\frac{\partial g_n(w, y^{n-1})}{\partial w} = \frac{1}{f_X(g_1(w))} \prod_{i=1}^{n} \frac{\partial S(x_i, y_i)}{\partial x} \tag{2.25}$$

$$= \frac{1}{f_X(g_1(w))} \prod_{i=1}^{n} \frac{f_{X|Y}(x|y)}{f_X(S(x,y))} \tag{2.26}$$

$$= \frac{f_{W|Y^{n-1}}(w|y^{n-1})}{f_X(x_n)}. \tag{2.27}$$

Taking the logarithm and dividing by n, we arrive at

$$\frac{1}{n} \log \frac{\partial g_n(w, y^{n-1})}{\partial w} = \frac{1}{n} \log f_{W|Y^{n-1}}(w|y^{n-1}) - \frac{1}{n} \log f_X(x_n). \tag{2.28}$$

Finally, taking the limit we can apply the previous lemma to arrive at the desired result.

$\square$

2.3   Source Coding with Feedforward

The source coding with feedforward paradigm is a special case of source coding with side information, where the side information made available at the decoder is a noiseless, delayed version of the source as depicted in Figure 2.2.
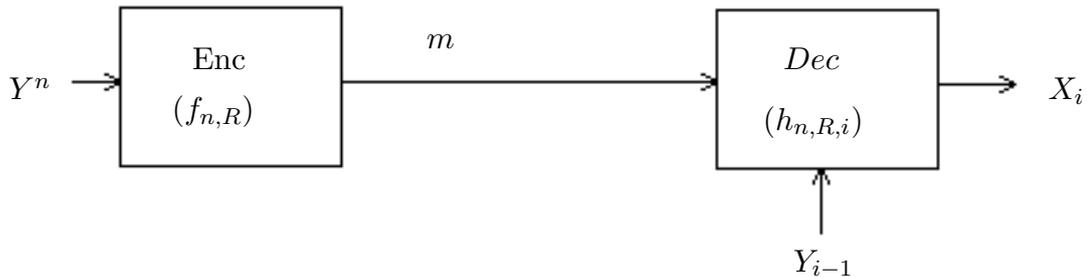
Figure 2.2: General source coding with feedforward.

- The source sequence consists of a sequence of i.i.d. random variables $\{Y_i\}_{i=1}^{\infty}$ with each Y_i distributed according to the distribution P_Y.

- A source encoding scheme of rate R consists of a sequence of mappings $f_{n,R} : \mathcal{Y}^n \to \{1, 2, \cdots, 2^{nR}\}$ from the possible source sequences of length n to an index $M \in \{1, \ldots, 2^{nR}\}$ of nR bits.

- The reconstruction at time i is given by $X_i \in \mathcal{X} \subseteq \mathbb{R}$.

- A decoding scheme consists of a sequence of time-varying decoding functions, denoted $h_{n,R,i} : \{1, 2, \cdots, 2^{nR}\} \times \mathcal{Y}^{i-1} \to \mathcal{X}$, which are parameterized by the rate R, and the time step $i \in \{1, \cdots, n\}$.

- The distortion measure $\rho_n : \mathcal{Y}^n \times \mathcal{X}^n \to \mathbb{R}^+$ is taken to be the average distortion between symbols in the sequence:

$$\rho_n(y^n, x^n) = \frac{1}{n} \sum_{i=1}^{n} d(y_i, x_i) \tag{2.29}$$

for some function $d : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}^+$.

- We say that a rate-distortion pair (R, D) is achievable if there exists a sequence of (n, R) rate-distortion codes such that $\forall \epsilon > 0$, $\exists N_\epsilon$ such that for all $n \geq N_\epsilon$,

$$\mathbb{E}[\rho_n(Y^n, X^n)] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[d(Y_i, X_i)] \leq D + \epsilon.$$

15

2.3.1 The Feedforward Rate-Distortion Function

When the distortion measure is additive and the source is memoryless, feed-forward information cannot improve the rate-distortion region, i.e. $R_{ff}(D) = R(D)$ for all D [20].

For a given source distribution P_Y and additive distortion measure $d(x, y)$, and target rate I, denote the induced joint distribution as $P^I_{Y,X}$ where the marginal P_Y is satisfied, $I = I(X;Y) = I(P^I_{Y,X})$, and the induced distribution $P^I_{Y,X}$ has expected distortion

$$\mathbb{E}_{P^I_{Y,X}}\left[\rho\left(Y, X\right)\right] = D(I).$$

CHAPTER 3

CHANNEL CODING WITH FEEDBACK

Most communication and control systems in practice use feedback periodically to enhance performance. In control systems, feedback provides inherent robustness to system uncertainties, adaption to unknown disturbances, and often a significant reduction in complexity, energy, or other costs. However, even when feedback is dramatically present, the study and use of feedback in information theory has not been explored in as much depth as other aspects of communication. Given the prevalence of recent applications involving many decentralized communication networks with centralized complexity nodes, there is strong motivation to reconsider how feedback should be used.

In this chapter, we demonstrate a converse to the fundamental limit of communication with feedback, that can be stated in terms of the mathematics of dynamical systems in control. Intuitively, it states that if a rate R is achievable, then the upper Lyapunov exponent of the dynamical system acting as the encoder must exceed R. This bound is a property of the encoding scheme rather than the channel and will hold over a broad class of memoryless channels.

Section 3.1 presents the necessary condition for reliable communication that we have developed, explaining the conditions under which it holds, and an example follows in Section 3.2. The main results from this chapter also appear in [14].

3.1   A Necessary Condition for Reliable Communication

Assume that the sequence Y^n is the output of a noisy channel with a sequence of transition laws given by (2.2), where the encoder specifies X_i given feedback Y^n and message W according to some law (2.4). Let $x_i(u)$ denote

17

the sequence of "virtual" channel inputs corresponding to the message point u and the received output sequence y^n, so that $x_i(u) = g(u, y^{i-1})$. Examining the posterior distribution, we see that for a fixed sequence of channel outputs y^n, the posterior distribution evaluated at the message point u depends on that message only through the sequence $x^i(u) \triangleq \{x_i(u), x_{i-1}(u), \ldots, x_1(u)\}$ of channel inputs that would have been generated by that message:

$$
\begin{aligned}
& f_{W|Y^n}(u|y^n) \\
&= \frac{f_{Y^n|W}(y^n|u) f_W(u)}{f_{Y^n}(y^n)} \\
&= \frac{f_{Y^n|W}(y^n|u)}{f_{Y^n}(y^n)} \qquad\qquad\qquad\qquad (3.1\text{a}) \\
&= \prod_{i=1}^{n} \frac{f_{Y_i|u,Y^{i-1}}(y_i|u, y^{i-1})}{f_{Y_i|Y^{i-1}}(y_i|y^{i-1})} \\
&= \prod_{i=1}^{n} \frac{f_{Y_i|u,Y^{i-1},X^i}(y_i|u, y^{i-1}, x^i(u))}{f_{Y_i|Y^{i-1}}(y_i|y^{i-1})} \qquad (3.1\text{b}) \\
&= \prod_{i=1}^{n} \frac{f_{Y_i|Y^{i-1},X^i}(y_i|y^{i-1}, x^i(u))}{f_{Y_i|Y^{i-1}}(y_i|y^{i-1})}, \qquad (3.1\text{c})
\end{aligned}
$$

where (3.1a) follows from (2.1); (3.1b) follows from (2.4).

Now consider the sequence $X^n(W + \epsilon_n)$ generated using the message point $W + \epsilon_n$. If the sequence of transmission functions is continuous, then for ϵ_n small, we expect this sequence of "virtual" channel inputs to be, in some sense, very similar to the channel input sequence resulting from the message W, which actually produced the output sequence Y^n

Lemma 3.1.1. *Define $\epsilon_{n,\delta} \triangleq 2^{-n(\bar{L}+\delta)}$. If the sequence of transmission functions is $g_i(\cdot, y^{i-1})$ almost always continuously differentiable, then for any $\delta > 0$:*

$$
\limsup_{in \ p.} \frac{1}{n} \log |X^n(W + \epsilon_{n,\delta}) - X_n(W)| \leq -\delta.
$$

Proof. Note that since $g_i(\cdot, y^{i-1})$ is almost always continuously differentiable,

18

$\bar{L}$ exists from (2.21). Thus,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\log|X^n(W+\epsilon_{n,\delta}) - X_n(W)| \geq -\delta\right)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\log\left|g_n(W+\epsilon_{n,\delta}, Y^{n-1}) - g_n(W, Y^{n-1})\right| \geq -\delta\right)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\log\left|g'_n(\tilde{W})\epsilon_{n,\delta}\right| \geq -\delta\right) \qquad (3.2)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\log|g'_n(\tilde{W})| + \frac{1}{n}\log|\epsilon_{n,\delta}| \geq -\delta\right)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\log|g'_n(\tilde{W})| \geq \bar{L}\right)$$

$$= 0, \qquad (3.3)$$

where (3.2) holds for some $\tilde{W}$ between W and $W + \epsilon_{n,\delta}$ from the first order Taylor's series representation; (3.3) follows from our assumption that g_i is continuously differentiable and the upper Lyapunov exponent definition (2.21).

$\square$

For a broad class of channels, this exponential decay between the channel inputs corresponding to W and $W + \epsilon_{n,\delta}$ is sufficient to ensure that the posterior distribution evaluated at these two points will be approximately the same. This property will hold for channels which satisfy the following assumption:

Assumption 1. *Define the vector*
$g^i(\delta) = (2^{-\delta}, 2^{-2\delta}, \ldots, 2^{-i\delta})$. *For all $\delta > 0$ and $b \geq 0$, and almost all x^n sequences, the following holds:*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{f_{Y_i|Y^{i-1},X^i}\left(Y_i|Y^{i-1}, x^i + g^i(\delta)\right)}{f_{Y_i|Y^{i-1}}(Y_i|Y^{i-1})} \leq b\right)$$

$$= \lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}\log\frac{f_{Y_i|Y^{i-1},X^i}\left(Y_i|Y^{i-1}, x^i\right)}{f_{Y_i|Y^{i-1}}(Y_i|Y^{i-1})} \leq b\right).$$

We note briefly here that many channels satisfy this assumption. For

example, consider the Gaussian auto-regressive channel:

$$Y_i = \sum_{j=1}^{J} \alpha_j Y_{i-j} + \sum_{k=0}^{K} \beta_k X_{i-k} + N_i, \tag{3.4}$$

where the N_i's are independent, identically distributed, Gaussian random variables with 0 mean and variance σ^2. Note that

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left(N_i + \sum_{k=0}^{K} \beta_k 2^{i-k} \right)^2 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} N_i^2 = \sigma^2 \ \ a.s.$$

and so the assumption holds.

With this, we can now state our main theorem.

Theorem 3.1.2. *If an encoding scheme with continuously differentiable transmission functions $g_i(w, y^{i-1})$ can achieve a rate R with feedback over a noisy channel satisfying Assumption 1, then $\bar{L} \geq R$.*

Proof. We will prove this via contradiction. Assume that $R > \bar{L}$, is achievable by some feedback communication scheme. Let $\delta \in (0, R - \bar{L})$ and consider a sequence of messages $W + \epsilon_{n,\delta}$ so that $2^{-nR} < \epsilon_{n,\delta} < 2^{-nL}$. From Lemma (3.1.1), we see that the difference between the trajectories of $W + \epsilon_{n,\delta}$ and W will decay exponentially. Under assumption 1, this will imply that the posterior distribution evaluated at $W + \epsilon_n$ will be approximately equal to the posterior evaluated at W:

$$\lim_{n \to \infty} \mathbb{P} \left(\frac{1}{n} \log f_{W|Y^n}(W + \epsilon_{n,\delta}|Y^n) \leq R \right)$$

$$= \lim_{n \to \infty} \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^{n} \log \frac{f_{Y_i|Y^{i-1},X^i}\left(Y_i|Y^{i-1}, X^i + g^i(\delta)\right)}{f_{Y_i|Y^{i-1}}(Y_i|Y^{i-1})} \leq R \right) \tag{3.5}$$

$$= \lim_{n \to \infty} \mathbb{P} \left(\frac{1}{n} \sum_{i=1}^{n} \log \frac{f_{Y_i|Y^{i-1},X^i}\left(Y_i|Y^{i-1}, X^i\right)}{f_{Y_i|Y^{i-1}}(Y_i|Y^{i-1})} \leq R \right) \tag{3.6}$$

$$= \lim_{n \to \infty} \mathbb{P} \left(\frac{1}{n} \log f_{W|Y^n}(W|Y^n) \leq R \right) \tag{3.7}$$

$$= 0. \tag{3.8}$$

Here, (3.5) follows from (3.1) and the exponential decay of the trajectories shown in lemma (3.1.1); (3.6) follows from assumption (1); (3.7) follows again from (3.1); and (3.8) follows from lemma (2.1.1).

Thus with high probability the posterior distribution evaluated at $W + \epsilon$ will also be close to 2^{nR} for all $\epsilon \in \left[0, 2^{-n(\bar{L}+\delta)}\right)$. Because $2^{-n\bar{L}} > 2^{-nR}$, this implies that the posterior distribution does not decay rapidly outside an interval of length 2^{-nR}. From above, we see that for large n, with high probability, $f_{W|Y^n}(W + \epsilon_{n,\delta}|y^n) \geq 2^{nR}$ for all $\epsilon_n \in [0, 2^{-n(\bar{L}+\delta)})$. Thus, with high probability, for large n, the conditional probability of error given Y^n, is given by

$$
\begin{aligned}
\lim_{n\to\infty} P(E|Y^n) &\geq \lim_{n\to\infty} \int_{W+2^{-nR}}^{W+\epsilon_{n,\delta}} f_{W|Y^n}(u|y^n)du \\
&\geq \lim_{n\to\infty} 2^{nR}\left(2^{-n(\bar{L}+\delta)} - 2^{-nR}\right) \qquad (3.9) \\
&= \lim_{n\to\infty} 2^{n(R-\bar{L}-\delta)} - 1,
\end{aligned}
$$

where (3.9) holds from (3.8). Since $R - \bar{L} > \delta$, the probability of error does not tend to zero as n increases - thus leading to a contradiction. $\qquad\square$

3.2 Example: AWGN Channel

Consider an additive white Gaussian noise channel with noise power σ^2 and an average power constraint P. In this case, the PM scheme reduces to the Schalkwijk-Kailath scheme, and the transmission scheme is given by

$$
\begin{aligned}
g_{n+1}(W, Y^n) &= S_{Y_n}(X_n) \\
&= \sqrt{1 + \frac{P}{\sigma^2}}\left(X_n - Y_n\frac{\frac{P}{\sigma^2}}{1 + \frac{P}{\sigma^2}}\right).
\end{aligned}
$$

First, note that the derivative of the recursive transmission function with respect to X is

$$
\frac{\partial}{\partial X}S_{Y_i}(X) = \sqrt{1 + \frac{P}{\sigma^2}}.
$$

Now using the chain rule to take the derivative of g_n with respect to the

message point, we get

$$\frac{\partial}{\partial w} g_{n+1}(w, Y^n) = \frac{1}{f_X(w)} \prod_{i=1}^{n} \frac{\partial}{\partial X} S_{Y_i}(X)$$

$$= \frac{1}{f_X(w)} \left(1 + \frac{P}{\sigma^2}\right)^{\frac{n}{2}}.$$

And the Lyapunov exponent is

$$\lim_{n \to \infty} \frac{1}{n} \log \frac{\partial}{\partial w} g_{n+1}(w, Y^n) = \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{f_X(w)} \left(1 + \frac{P}{\sigma^2}\right)^{\frac{n}{2}}$$

$$= \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{f_X(w)} + \frac{1}{n} \log \left(1 + \frac{P}{\sigma^2}\right)^{\frac{n}{2}}$$

$$= \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2}\right),$$

which we recognize as the capacity of an AWGN channel.

Now suppose instead that a particular encoding scheme chooses to use the same recursive functions, but transmit only a fraction $\beta = \frac{1}{\alpha}$ of the time at a power αP, so that

$$g_n(W, Y^n) = \begin{cases} \sqrt{1 + \frac{\alpha P}{\sigma^2}} \left(X_{n_k} - Y_{n_k} \frac{\frac{\alpha P}{\sigma^2}}{1 + \frac{P}{\sigma^2}}\right) & \text{w.p. } \beta \\ 0 & \text{w.p. } 1 - \beta, \end{cases}$$

where n_k was the last time at which the encoder chose to transmit a non-zero message. When $g_{n+1}(W, Y^n) = 0$, the derivative will be equal to zero, since it is constant for all message points. At other times, the derivative will be positive, thus the limit defining the Lyapunov exponent does not exist. The derivative when the encoder transmits will be given by

$$\frac{\partial}{\partial w} g_{n+1}(w, Y^n) = \frac{1}{f_X(w)} \prod_{i=1}^{k} \frac{\partial}{\partial x} S_{Y_{n_i}}(x)$$

$$= \frac{1}{f_X(w)} \left(1 + \frac{\alpha P}{\sigma^2}\right)^{\frac{k}{2}}.$$

Note that for any $w \in (0, 1)$, the following limit holds almost surely:

$$
\begin{aligned}
\lim_{n \to \infty} \frac{1}{n} \log \frac{\partial}{\partial w} g_{n+1}(w, Y^n) &= \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{f_X(w)} + \frac{1}{n} \log \left(1 + \frac{\alpha P}{\sigma^2}\right)^{\frac{k}{2}} \\
&= \frac{k}{2n} \log \left(1 + \frac{\alpha P}{\sigma^2}\right) \\
&= \frac{\alpha}{2} \log \left(1 + \frac{\alpha P}{\sigma^2}\right), \qquad\qquad (3.10)
\end{aligned}
$$

where (3.10) follows from the law of large numbers. It follows that the limit also holds in probability; thus, the upper Lyapunov exponent is given by (3.10). So any achievable rate $R < \frac{\alpha}{2} \log \left(1 + \frac{\alpha P}{\sigma^2}\right) < \frac{1}{2} \log \left(1 + \frac{P}{\sigma^2}\right)$ if this encoding scheme is used with $\alpha < 1$, verifying that this was a suboptimal encoding strategy.

CHAPTER 4

SOURCE CODING WITH FEEDFORWARD

In this chapter, we explore the duality between source coding with feed-forward and channel coding with feedback. By looking at the probabilistic relationship between optimal inputs and outputs in the channel coding problem, we see that a source coding problem exists with the same relationship between the source and reconstruction sequences. This leads us to a provably good scheme for source coding with feedforward. In particular, we show that the posterior matching scheme, which is already known to be an optimal encoder for the channel coding problem, can also be used as an optimal decoder for the source coding problem.

The problem of source coding with side information was first introduced in [10] for distributed source coding applications. The scenario for reproducing a source with causal side information was introduced by Weissman and Merhav [11] as a competitive prediction problem. Around the same time, Pradhan and others also began studying source coding with side information, with a particular emphasis on source coding with the availability of feedforward information [9, 12, 13]. This model is useful, for example, in sensor networks where some sensors may have information about a random measurement before other sensors, but they are limited by power or bandwidth constraints to send only a compressed version of this information.

The dual of source coding is channel coding [7, 8], and the dual of source coding with feedforward is channel coding with feedback [9]. The posterior matching (PM) scheme, introduced by Shayevitz and Feder [5], is a general scheme for communication with feedback over memoryless channels. In this paper, we show that dualizing the role of the PM scheme results in an optimal scheme for source coding with feedforward. Concurrent work by Shayevitz [21] also exploits this duality for lossy compression when the source and reconstruction alphabets are countable. Our scheme differs by making use of the Lyapunov exponent property of the PM scheme to develop a source

coding scheme when the source and reconstruction alphabets are continuous.

4.1 A Perspective from the Converse to the Source Coding Paradigm with Feedback

According to the converse to the source coding with feedforward [20], any source coding scheme that attains an average distortion

$$\limsup_{n\to\infty} \mathbb{E}[\rho_n(Y^n, X^n)] \leq D \tag{4.1}$$

must have a rate $R \geq R_{ff}(D)$. Indeed, for any source Y^n, with message index $M \in \{1, \ldots, 2^{nR}\}$ and reconstruction X^n for which (4.1) holds, we have:

$$
\begin{aligned}
nR &\geq I(M; Y^n) \\
&= \sum_{i=1}^{n} H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1}, M) \\
&= \sum_{i=1}^{n} H(Y_i) - H(Y_i|Y^{i-1}, M, X^i) \tag{4.2} \\
&\geq \sum_{i=1}^{n} H(Y_i) - H(Y_i|X_i) \tag{4.3} \\
&= \sum_{i=1}^{n} I(X_i; Y_i) \\
&\geq nR(D), \tag{4.4}
\end{aligned}
$$

where (4.2) follows because the Y process is i.i.d. and because X^i is a deterministic function of Y^{i-1} and M; (4.3) follows because conditioning reduces entropy; and (4.4) follows from the convexity and monotonicity of the rate-distortion function [20].

Equality in (4.3) holds when the induced channel with message M, input X and output Y is memoryless (2.3). Finally, equality occurs in (4.4) when the minimizing joint distribution on X_i and Y_i for the rate-distortion function is used.

4.2   The Test Channel

To emphasize the relationship between the primal and dual problems, define the normalized index:

$$W_n^* \triangleq \frac{m}{2^{nR}} \in (0, 1].$$

Consider a memoryless channel with transition law $P_{Y|X}(y|x)$ and noiseless feedback. We wish to communicate the message W_n^* over this channel, as suggested by the right side of Figure 4.1 below.
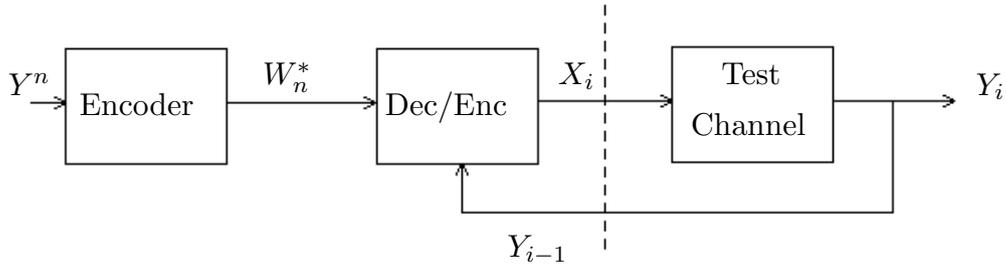

Figure 4.1: Test channel.

For this problem, the posterior matching scheme specifies an optimal encoding strategy in the sense that it can asymptotically achieve capacity [5]. This encoding strategy determines a sequence of channel inputs X_i that are statistically independent of all previous outputs (thus leading to the outputs Y_i being independent). To attain equality in the converse (4.3), the optimal test channel is memoryless, and therefore, the output sequence Y_i will also be i.i.d., determined according to the induced transition law for the test channel $P_{Y|X}^I$.

4.3   Dualizing the PM Encoder

For channel coding with feedback, we start with a memoryless channel and specify the inputs, Z_i, that are randomly mapped to an output sequence of i.i.d. random variables Y_i, which induces a distribution P_Y. Instead, for source coding with feedforward, we start with the source Y that is i.i.d. drawn according to a given distribution P_Y and a sequence of distortion measures ρ_n. For any $R > 0$, we can pick $I > 0$ such that $I < R$. Then for any $I > 0$, the distortion measure induces a joint distribution $P_{Y,Z}^I$ such that the expected distortion $\mathbb{E}_{P_{Y^n,Z^n}^I}[\rho_n(Y^n, X^n)] = D(I)$.

26

Define the test channel to have a transition law given by the induced conditional distribution, $P^I_{Y|Z}$. Since the Y^n sequence is i.i.d., we can interpret them as the channel outputs from a posterior matching scheme algorithm with message point W uniformly distributed on $[0,1]$ and "virtual" channel inputs given by

$$Z_i \triangleq g_i(W, Y^{i-1}). \tag{4.5}$$

Note that in this case, the expected distortion is given by

$$\mathbb{E}_{P^I_{Y^n,Z^n}} \left[\rho_n\left(Y^n, Z^n\right) \right] = D(I). \tag{4.6}$$

Now we consider finding the proposed quantization scheme at the encoder, by taking advantage of how the decoder has causal feedforward side information. After observing Y^n, define the quantizer index $W^*_n \in \{\frac{1}{2^{nR}}, \frac{2}{2^{nR}}, \ldots, 1\}$ and subsequent reproductions X^n as

$$W^*_n \triangleq \begin{array}{c} \arg\min \\ W_n \in \{\frac{1}{2^{nR}}, \frac{2}{2^{nR}}, \cdots, 1\} \end{array} \sum_{i=1}^n d\left(Y_i, g_i\left(W_n, Y^{i-1}\right)\right) \tag{4.7a}$$

$$X_i \triangleq g_i(W^*_n, Y^{i-1}). \tag{4.7b}$$

In order to prove our main theorem, we will need to define a few other pieces of terminology:

$$\tilde{Z}_i \triangleq g_i(\langle W \rangle_{nR}, Y^{i-1}) \tag{4.8}$$

$$\epsilon_n \triangleq 2^{-n(R-I)} \tag{4.9}$$

$$\dot{d}_2(a,b) \triangleq \frac{\partial}{\partial z} d(a,z)\big|_{z=b} \tag{4.10}$$

$$\ddot{d}_2(a,b) \triangleq \frac{\partial^2}{\partial z^2} d(a,z)\big|_{z=b}. \tag{4.11}$$

This will allow us to compare W^*_n to an intermediate pair of quantizer indices and reproductions pertaining to the quantized virtual message point: $\langle W \rangle_{nR} \in \{\frac{1}{2^{nR}}, \frac{2}{2^{nR}}, \ldots, 1\}$, and $\tilde{Z}^n$ - which would be the decoder's reproductions if the PM scheme were used operating on $\langle W \rangle_{nR}$.

Note that the error introduced by the quantizer, $|\langle W \rangle_{nR} - W|$, is bounded in magnitude by the size of the quantization intervals, 2^{-nR}. Since $R > I$,

and since the PM scheme has a Lyapunov exponent of I, we now show that using a PM decoding scheme with $\langle W \rangle_{nR}$ results in $\{Z_i : i = 1, \ldots, n\}$ to be "close to" $\{\tilde{Z}_i : i = 1, \ldots, n\}$ with an approximation error of approximately $\{\epsilon_i = 2^{-i(R-I)}\}$ with very high probability:

Lemma 4.3.1. *With probability one, the following holds:*

$$\limsup_{n \to \infty} \frac{1}{n} \log \left| Z_i - \tilde{Z}_i \right| = -(C - R). \tag{4.12}$$

Proof. Apply Lemma 2.2.2, Assumption 2b, Taylor's theorem, and the fact that $|\langle W \rangle_{nR} - W| < 2^{-nR}$. □

We state another lemma as follows:

Lemma 4.3.2. *If a non-negative sequence $(a_n : n \geq 1)$ satisfies*

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} a_i < \infty, \text{ then}$$

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i a_i = 0.$$

Proof. Suppose $\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i a_i > 0$. Then clearly, $\limsup_{n \to \infty} \epsilon_i a_i = \infty$. Since $\epsilon_i > \frac{1}{i}$ for all i greater than some i_0, $\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} a_i = \infty$, which is a contradiction. □

We state our technical conditions in terms of the following "feasible" set of distortion measures:

Assumption 2. *The following technical conditions hold:*

- ***a:*** *$d(y, \cdot)$ is twice continuously differentiable for almost all $y \in \mathcal{Y}$.*

- ***b:*** $\mathbb{E}_{P_{Y,Z}} \left[\left| \dot{d}_2(Y, Z) \right| \right] < \infty$
  and $\mathbb{E}_{P_{Y,Z}} \left[\left| \ddot{d}_2(Y, Z) \right| \right] < \infty.$

- ***c:*** $\left| \dot{d}_2(y, z) \right| < \infty$ *for almost all $y \in \mathcal{Y}$ and $z \in \mathcal{Z}$.*

- ***d:*** *the Markov chain (Y_i, Z_i) given by (4.5) is Harris-Recurrent.*

Theorem 4.3.3. *If Assumption 2 holds, then the scheme given by (4.7) achieves the rate-distortion function $D(I)$.*

Proof. First note that for any $\omega \in \Omega$,

$$\rho_n(Y^n, X^n) = \frac{1}{n} \sum_{i=1}^{n} d\left(Y_i, g_i(W_i^*, Y^{i-1})\right) \tag{4.13}$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} d\left(Y_i, g_i(\langle W \rangle_{nR}, Y^{i-1})\right) \tag{4.14}$$

$$= \rho_n(Y^n, \tilde{Z}^n), \tag{4.15}$$

where (4.13) follows from (4.7b);(4.14) follows from (4.7a); (4.15) follows from (4.8).

Next, note that for any $\omega \in \Omega$:

$$\left| \rho_n(Y^n, Z^n) - \rho_n(Y^n, \tilde{Z}^n) \right| = \left| \frac{1}{n} \sum_{i=1}^{n} d(Y_i, Z_i) - d(Y_i, \tilde{Z}_i) \right|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left| d(Y_i, Z_i) - d(Y_i, \tilde{Z}_i) \right|$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left| Z_i - \tilde{Z}_i \right| \left| \dot{d}_2(Y_i, \check{Z}_i) \right|, \tag{4.16}$$

where (4.16) follows from Taylor's theorem for some $\check{Z}_i$ between Z_i and $\tilde{Z}_i$.

Note that for almost all $\omega \in \Omega$ and any $\delta > 0$, there exists an $m \equiv m(\delta, \omega)$ such that with

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{i=m}^{n} \left| Z_i - \tilde{Z}_i \right| \left| \dot{d}_2(Y_i, \check{Z}_i) \right|$$

$$\leq \delta + \limsup_{n \to \infty} \frac{1}{n} \sum_{i=m}^{n} \epsilon_i \left| \dot{d}_2(Y_i, Z_i) + \epsilon_i \ddot{d}_2(Y_i, Z_i) \right| \tag{4.17}$$

$$\leq \delta + \limsup_{n \to \infty} \frac{1}{n} \sum_{i=m}^{n} \epsilon_i \left| \dot{d}_2(Y_i, Z_i) \right| + \epsilon_i^2 \left| \ddot{d}_2(Y_i, Z_i) \right|$$

$$= \delta + \limsup_{n \to \infty} \frac{1}{n} \sum_{i=m}^{n} \epsilon_i \left| \dot{d}_2(Y_i, Z_i) \right| \tag{4.18}$$

$$= \delta, \tag{4.19}$$

where (4.17) follows from Lemma 4.3.1 and Assumption 2a; (4.18) follows from the Strong Law of Large Numbers for Markov chains applied to (Y_i, Z_i) for the function $\dot{d}_2(y, z)$ and Assumption 2b; and (4.19) follows from As-

sumption 2d.

Lastly, from Assumption 2c and Lemma 4.3.1, we have that for almost all $w \in \Omega$,

$$\limsup_{n \to \infty} \sum_{i=1}^{m} \left| Z_i - \tilde{Z}_i \right| \left| \dot{d}_2(Y_i, \check{Z}_i) \right| = 0. \tag{4.20}$$

Combining (4.15), (4.16), (4.19),(4.20) with Assumption 2d, we have that

$$\limsup_{n \to \infty} \sum_{i=1}^{n} \rho_n(Y_n, X_n) = \limsup_{n \to \infty} \sum_{i=1}^{n} \rho_n(Y_n, Z_n) = D(I).$$

$\square$

4.4 Example: Gaussian Source with Mean Squared Distortion

Suppose that we have a memoryless zero-mean gaussian source with variance σ^2 so that $Y_i \sim \mathbb{N}(0, \sigma^2)$ for all i. If our distortion measure is the mean-squared difference between corresponding terms of the sequence, and we wish to attain a distortion of no more than D, then our test channel should have independent additive noise with variance D (i.e. $n_i \sim \mathbb{N}(0, D)$). In this case, the PM encoding scheme provides optimal inputs to this channel which are independent Gaussian random variables with a power constraint of $P = \sigma^2 - D$. Now we can compare the input sequence generated by encoding W versus the sequence generated by using a quantized version of the message, $\langle W \rangle_{nR}$.

If we let F_X denote the Gaussian CDF corresponding to an $\mathbb{N}(0, P)$ distribution, then the first terms of the corresponding input sequences are

$$Z_1 = F_X^{-1}(W), \ \tilde{Z}_1 = F_X^{-1}(\langle W \rangle_{nR}).$$

For subsequent terms of the sequence, posterior matching is equivalent to the Schalkwijk-Kailath scheme [2] and channel inputs would be chosen according to the recursive equations:

$$Z_n = \sqrt{\frac{P+D}{D}}\left(Z_{n-1} - \frac{P}{P+D}Y_{n-1}\right)$$

$$\tilde{Z}_n = \sqrt{\frac{P+D}{D}}\left(\tilde{Z}_{n-1} - \frac{P}{P+D}Y_{n-1}\right).$$

We see that the difference between the two terms of these sequences at time n is given by

$$
\begin{aligned}
\left|Z_n - \tilde{Z}_n\right| &= \sqrt{\frac{P+D}{D}}\left|Z_{n-1} - \tilde{Z}_{n-1}\right| \\
&= \left(\frac{P+D}{D}\right)^{\frac{n}{2}}\left|Z_1 - \tilde{Z}_1\right| \\
&= \left(2^I\right)^n\left|Z_1 - \tilde{Z}_1\right|, \quad\quad (4.21)
\end{aligned}
$$

where (4.21) follows because $\log\sqrt{\frac{P+D}{D}}$ is the capacity of a Gaussian channel with input power constrained to P and noise power D. Because we are using the optimal input distribution for our test channel, the mutual information will be equal to the capacity of the channel. If we bound the difference between the initial conditions of these sequences by the Taylor's series approximation, we get

$$
\begin{aligned}
\left|Z_1 - \tilde{Z}_1\right| &= \left|F_X^{-1}(\langle W\rangle_{nR}) - F_X^{-1}(W)\right| \\
&\leq \left(F_X^{-1}\right)'\left(\bar{W}\right)\epsilon_n \quad\quad (4.22) \\
&\leq 2^{\log\left(F_X^{-1}\right)'(\bar{W})}\left(2^{-nR}\right) \\
&\leq 2^{-n(R-\ell)}, \quad\quad (4.23)
\end{aligned}
$$

where (4.22) holds for some $\bar{W} \in [W, \langle W\rangle_{nR}]$ and (4.23) holds for arbitrarily small ℓ when n is large enough. Substituting (4.23) for the difference between Z_1 and $\tilde{Z}_1$ in (4.21), we get

$$
\begin{aligned}
\left|Z_{n+1} - \tilde{Z}_{n+1}\right| &\leq \left(2^I\right)^n\left(2^{-n(R-\ell)}\right) \\
&= 2^{-n(R-I-\ell)}. \quad\quad (4.24)
\end{aligned}
$$

31

So we can bound the mean-squared distortion between $\tilde{Z}_i$ and Y_i as follows:

$$
\begin{aligned}
d\left(\tilde{Z}_i, Y_i\right) &= \left(\tilde{Z}_i - Y_i\right)^2 = \left(\tilde{Z}_i - Z_i + Z_i - Y_i\right)^2 \\
&= \left(\tilde{Z}_i - Z_i\right)^2 + (Z_i - Y_i)^2 \\
&\quad + 2\left(\tilde{Z}_i - Z_i\right)(Z_i - Y_i) \\
&= \left(\tilde{Z}_i - Z_i\right)^2 + N_i^2 + 2\left(\tilde{Z}_i - Z_i\right)(N_i) \\
&\leq \left(2^{-i(R-I-\ell)}\right)^2 + N_i^2 \qquad\qquad (4.25) \\
&\quad + 2\left(2^{-i(R-I-\ell)}\right)(N_i),
\end{aligned}
$$

where (4.25) comes from applying (4.24). Now taking expectations, we get

$$
\begin{aligned}
\mathbb{E}\left[d\left(\tilde{Z}_i, Y_i\right)\right] &\leq \left(2^{-i(R-I-\ell)}\right)^2 + \mathbb{E}\left[N_i^2\right] \\
&\quad + 2\left(2^{-i(R-I-\ell)}\right)\mathbb{E}\left[N_i\right] \\
&= \left(2^{-2i(R-I-\ell)}\right) + D.
\end{aligned}
$$

As long as we are allowed to use a rate $R > I$, we can choose a positive $\ell < R - I$ so that the distortion in our scheme decays to the rate-distortion function at an exponential rate when the source sequence is i.i.d. Gaussian random variables.

CHAPTER 5

CONCLUSION

In this thesis we use control theoretic techniques to provide a new perspective for analyzing two related problems in communication: channel coding with feedback and source coding with feedforward. We see that the Lyapunov exponent of a related dynamical system emerges in both scenarios.

For channel coding with feedback, we show that for a broad class of channels, the Lyapunov exponent of the transmission function - when it exists - is fundamentally linked to the maximum rate which the scheme can attain. More generally, we bound the set of achievable rates for a given encoding scheme by considering a generalization of the Lyapunov exponent of the transmission function. We note that the posterior matching scheme - a provably optimal feedback communication scheme for memoryless channels - has an encoding function with a Lyapunov exponent exactly equal to the communication rate.

In the dual problem, source coding with feedforward, the optimal test channel is memoryless. This motivates the idea of dualizing posterior matching for this setting. By exploiting the Lyapunov exponent property, we demonstrate that such a scheme, with low encoder and decoder complexity, attains the rate-distortion function. By approaching these problems from a dynamical systems perspective, we hope to provide the intuition to motivate the evaluation and design of new communication schemes.

REFERENCES

[1] C. E. Shannon, "Zero rare capacity of something," *IBM Journal of Research and Development*, vol. 2, no. 4, pp. 289–293, 1956.

[2] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 172–182, 1966.

[3] J. Schalkwijk, "A coding scheme for additive noise channels with feedback–II: Band-limited signals," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 183–189, 1966.

[4] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, 1963.

[5] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Transactions on Information Theory*, to be published.

[6] T. P. Coleman, "A stochastic control viewpoint on 'posterior matching'-style feedback communication schemes," in *IEEE International Symposium on Information Theory*, 2009, pp. 1520–1524.

[7] D. MacKay, *Information Theory, Inference, and Learning Algorithms*. New York, NY: Cambridge University Press, 2003.

[8] T. M. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley, 2006.

[9] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side-information case," *IEEE Transactions on Information Theory*, vol. 49, pp. 1181–1203, May 2003.

[10] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, pp. 1–10, 1976.

[11] T. Weissman and N. Merhav, "On competitive prediction and its relation to rate-distortion theory," *IEEE Transactions on Information Theory*, vol. IT-49, pp. 3185–3194, December 2003.

[12] S. S. Pradhan, "Source coding with feedforward: Gaussian sources," in *IEEE International Symposium on Information Theory*, Chicago, IL, 2004, p. 212.

[13] E. Martinian and G. W. Wornell, "Source coding with fixed lag side information," in *42nd Annual Allerton Conference on Communications, Control and Computing*, Monticello, IL, September 2004, pp. 1–10.

[14] H. Ebeid, S. K. Gorantla, and T. P. Coleman, "A necessary condition for reliable communication with feedback based upon Lypunov exponents of dynamical systems," in *MILCOM*, to be published.

[15] H. Ebeid and T. P. Coleman, "Using the posterior matching scheme for source coding with feedforward," in *IEEE International Symposium on Information Theory*, 2010, pp. 96–100.

[16] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, pp. 752–772, 1993.

[17] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[18] O. Shayevitz and M. Feder, "Communication with feedback via posterior matching," in *IEEE International Symposium on Information Theory*, Nice, France, June 2007, pp. 391–395.

[19] O. Shayevitz and M. Feder, "The posterior matching feedback scheme: Capacity achieving and error analysis," in *IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008, pp. 900–904.

[20] R. Venkataramanan and S. S. Pradhan, "Source coding with feedforward: Rate-distortion theorems and error exponents for a general source," *IEEE Transactions on Information Theory*, vol. 53, pp. 2154–2179, June 2007.

[21] O. Shayevitz, "A symbolic dynamical system approach to lossy source coding with feedforward," in *Data Compression Conference*, Snowbird, Utah, March 2010, pp. 317–326.