DIPLOMARBEIT

PERFORMANCE LIMITS OF GAUSSIAN CHANNELS WITH QUANTIZED FEEDBACK

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Diplomingenieurs

unter der Leitung von
Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Gerald Matz
Univ.Ass. Dipl.-Ing. Andreas Winkelbauer
Institute of Telecommunications

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von
Stefan Farthofer
Antonsplatz 16/17
1100 Wien

Wien, im Mai 2014

— To my parents —

Abstract

This thesis studies the performance limits of Gaussian channels with quantized feedback. The channel capacity of a channel with additive white Gaussian noise was already studied by Shannon and is well known as Shannon capacity. The problem of optimal compression of a source was also mathematically formulated by Shannon in his rate distortion theory.

These two aspects are combined in this thesis, as the channel output should be quantized. We show that a rate-distortion optimal compression of the channel output maximizes the mutual information only in the scalar case, but is generally suboptimal in the vector case. We show that the information bottleneck method provides a framework for quantizers which maximize the mutual information. By means of some selected channels we discuss the differences in mutual information and quantify those. We show that the difference is primarily determined by the eigenvalues of the channel.

It is known that perfect feedback does not increase the channel capacity, but the error probability is substantially decreased for finite blocklengths. The performance of such systems with a noisy feedback channel breaks down, i.e., no positive rate is achievable. In this thesis, we study schemes with quantized channel output and quantized feedback, and we derive equations for the achievable rate and error probabilities of such systems. Here different quantization of channel output and feedback corresponds to noisy feedback. Furthermore, we present a scheme where the receiver has the quantized feedback as side-information to achieve positive rates.

Kurzfassung

Diese Diplomarbeit beschäftigt sich mit der Analyse von Kapazitätsgrenzen Gaußscher Kanäle mit quantisiertem Feedback. Die Kanalkapazität eines Kanals mit additivem weißen Gaußschen Rauschen wurde bereits durch Shannon untersucht und ist als Shannon-Kapazität bekannt. Ebenfalls bereits von Shannon mathematisch formuliert wurde das Problem der optimalen Kompression einer Quelle in seiner Rate-Distortion-Theorie.

Diese zwei Aspekte werden in dieser Arbeit kombiniert, indem der Kanalausgang quantisiert werden soll. Es wird gezeigt, dass die im Rate-Distortion Sinne optimale Komprimierung des Kanalausgangs die Transinformation nur im skalaren Fall maximiert, im Vektorfall jedoch im Allgemeinen suboptimal ist. Für den Vektorfall wird gezeigt, dass ein auf Basis der Information-Bottleneck Methode entworfener Quantisierer die Transinformation maximiert. Anhand beispielhafter Kanäle werden die Unterschiede der Transinformationen beider Methoden erörtert und quantifiziert. Es wird gezeigt, dass der Unterschied im Wesentlichen durch die Kanaleigenwerte bestimmt wird.

Es ist ebenfalls bekannt, dass ein perfekter Rückkanal die Kanalkapazität zwar nicht erhöht, die Fehlerwahrscheinlichkeit für endliche Blocklängen jedoch drastisch reduzieren kann. Die Leistung solcher Systeme mit Rückkanal bricht allerdings ein und es kann keine positive Rate erreicht werden, wenn dieser störungsbehaftet ist. In dieser Arbeit werden Systeme mit quantisiertem Kanalausgang und quantisiertem Rückkanal untersucht und Ausdrücke für deren erreichbarer Rate und Fehlerwahrscheinlichkeiten abgeleitet. Unterschiedliche Quantisierung entspricht dabei einem störungsbehaftetem Rückkanal. Es wird jedoch ein Schema gezeigt welches eine positive Rate erreicht, indem der Empfänger die unterschiedlichen quantisierten Signale ausnutzt.

Contents

1

Introduction

1.1 Motivation

In modern digital communication scenarios we want to transmit information from transmitter to receiver over a channel, where the receiver performs a quantization of the received signal, i.e., analog-digital conversion, since the receiver typically processes the data in a digital manner. Even if we want to transmit a message with a finite alphabet, the received signal is analog, since it has to be transmitted over some sort of physical channel. In this thesis the focus is on an important type of channel, namely a channel with additive white Gaussian noise (AWGN). This type of channel was extensively studied with the famous result of the Shannon capacity

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2} \right), \tag{1.1}$$

where σ^2 is the power of the Gaussian noise and P is the power of the transmit signal. Due to the logarithm to base 2 the capacity is measured in bits per channel use. The Shannon capacity is the upper bound for the amount of information we could possibly transmit without errors.

Channel capacity can be interpreted as the maximum amount of mutual information about the message at transmitter and receiver. Or speaking with other words the amount of uncertainty about the message is reduced at most by the value of channel capacity. The amount of uncertainty is measured by the entropy which is defined as

$$H(\mathsf{x}) = -\sum_{x} p(x) \log_2 p(x), \tag{1.2}$$

where $p(x)$ is a probability mass function. Of course this definition makes only sense for discrete distributions. For continuous distributions we define the differential entropy as

$$h(\mathsf{x}) = -\int_S f(x)\log_2 f(x)\mathrm{d}x, \tag{1.3}$$

where S is the support set of the probability density function $f(x)$. The mutual information of two random variables x and y is then defined as

$$\begin{aligned} I(\mathsf{x};\mathsf{y}) &= h(\mathsf{x}) - h(\mathsf{y}|\mathsf{x}) && (1.4)\\ &= h(\mathsf{y}) - h(\mathsf{x}|\mathsf{y}). && (1.5) \end{aligned}$$

One important task of this thesis is to study how feedback affects the performance of a communication system. The most generic case is to assume that the transmitter has knowledge of all previously received symbols as side-information. The question is if such a system has a higher channel capacity as a system without feedback, i.e., $C_{FB} \geq C$? Surprisingly the capacity of the AWGN channel is not increased [10]:

$$C_{FB} = C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right). \tag{1.6}$$

For general Gaussian channels the capacity with feedback may be larger than the capacity without feedback and is limited by [9]

$$C_{FB} \leq \min\left\{2C, C + \frac{1}{2}\right\}, \tag{1.7}$$

measured in bits. Although we may not get an improved channel capacity using feedback, feedback can dramatically improve the performance of communication systems. The main reason is because the channel capacity is an asymptotic measure, i.e., reliable transmission up to the channel capacity requires infinite blocklengths. This is clearly not feasible in real world applications with delay constraints and other processing constraints. Hence, practical communication systems use finite blocklengths and therefore error probabilities are greater than zero. For fixed blocklength, the error probability can be decreased by a factor increasing exponentially in the blocklength [28].

The most common, but also most limiting, assumption is perfect feedback. It is assumed that the transmitter has perfect knowledge of the received values at receiver side. In general the feedback channel is not perfect and may be also modeled as an AWGN channel. Hence, the feedback capacity is limited. Another limitation is due to the fact that the need of quantization of the received signal is system immanent and introduces additional quantization noise. So at least the signal is quantized at the receiver front-end for further digital processing where a high rate approximation may be feasible. To reduce the amount of processing complexity and the rate of the feedback it may be important to reduce the rate of the quantizer. In this case the high rate assumption does not hold and the role of the quantizer, as an element which reduces the total mutual information, has to be studied. Quantizer design is a classical source coding problem and was addressed by many scientific papers, mathematically formulated by Shannon with his rate distortion (RD) theory. Usually the quantizer is designed to introduce the smallest amount of distortion. As a measure of distortion often the mean squared error (MSE) distortion is used. The design goal is then to minimize the MSE of the quantized signal to approximate the source signal as exact as possible, at a given rate.

In the communication scenario, where the received signal should be quantized, this approach however does in general not maximize the information about the transmit signal in the received signal. We show that this desired maximization of the mutual information can be perfectly addressed by the information bottleneck (IB) method [30]. We will show that IB quantizers outperform MSE quantizers and also quantify this performance improvement.

1.2 Goals

- Using the framework of the information bottleneck method, optimal quantizer design parameters should be acquired. Especially for the AWGN channel the closed form expressions for the Gaussian information bottleneck (GIB) [8] is used to obtain expressions for the mutual information depending on the rate of the quantizer. As a formal analogon to the distortion-rate functions and rate-distortion functions in the rate distortion theory, information-rate functions and rate-information functions for the Gaussian vector channel are derived.

- The performance improvement of the GIB quantizers should be proved and quantified. To this end, the information-rate function of the system with RD quantizers is derived and compared with the GIB quantizers.

- The performance limits of communication systems should be studied. Especially the performance of systems with quantized channel output and further quantized feedback signal should be investigated. The performance limits should be quantified by usage of an optimal communication system with linear feedback.

1.3 Outline

This thesis is structured as follows:

- The remainder of this chapter provides the necessary theoretical background.
- In **Chapter 2** we study the capacity of the Gaussian channel with channel output compression and apply the Gaussian information bottleneck method to derive the information-rate function and the rate-information function. Furthermore we compare the performance of information bottleneck optimal quantization and rate distortion optimal quantization.
- In **Chapter 3** we investigate an optimal linear feedback scheme with quantized feedback and use results from Chapter 2 to study the performance in terms of the information-rate tradeoff. Especially we provide asymptotic expressions for capacity and error probabilities of such systems.
- In **Chapter 4** we use the results from Chapter 3 to numerically evaluate the performance of the linear feedback system. We will validate the results by Monte Carlo simulations and discuss the difference to the theoretical findings.
- Finally, **Chapter 5** gives conclusions and an outlook for further possible research.

1.4 Background

1.4.1 Rate Distortion Theory

The rate distortion theory [2] addresses the problem of optimal compression of a source. Often it is necessary to compress a source in order to transmit it over a channel with a given capacity, which is smaller than the rate of the source. Or one just wants to reduce the amount of data to store or transmit, in tradeoff to a distortion, e.g., lossy source coding for speech. The rate distortion theory provides the mathematical formulation for this tradeoff. Let x be the source and $\hat{\mathsf{x}}$ be the compressed source, then we want to find a probabilistic mapping $f(\hat{x}|x)$, which we formulate as

$$R(D) \triangleq \min_{f(\hat{x}|x)} I(\mathsf{x};\hat{\mathsf{x}}) \qquad \text{s.t.} \quad \mathrm{E}\{d(\mathsf{x},\hat{\mathsf{x}})\} \leq D. \tag{1.8}$$

Hence, we want to minimize the rate, which is represented by the mutual information of x and $\hat{\mathsf{x}}$, under the constraint of keeping the average distortion under a specific limit

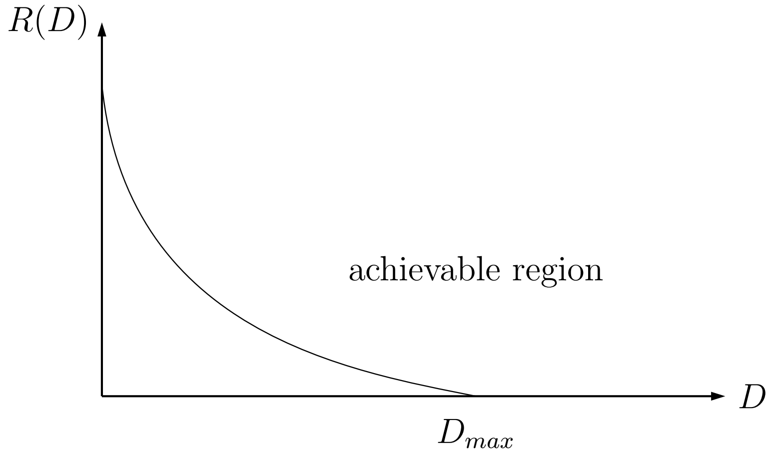

Figure 1.1: Illustration of the rate-distortion function.

D. The average distortion is the expectation of some suitable distortion $d(\cdot,\cdot)$,

$$\mathrm{E}\{d(\mathsf{x},\hat{\mathsf{x}})\} = \iint_{(\mathsf{x},\hat{\mathsf{x}})} f(x)f(\hat{x}|x)d(x,\hat{x})\mathrm{d}x\mathrm{d}\hat{x}. \tag{1.9}$$

The inverse of the rate-distortion function is the distortion-rate function

$$D(R) \triangleq \min_{f(\hat{x}|x)} \mathrm{E}\{d(\mathsf{x},\hat{\mathsf{x}})\} \qquad \text{s.t.} \quad I(\mathsf{x};\hat{\mathsf{x}}) \leq R. \tag{1.10}$$

Thus, the goal is to minimize de distortion at a given rate. Fig. 1.1 shows the general form of the rate-distortion function. The optimal rate-distortion function is bounded by D_{max}, since a source with finite power can be "approximated" by $\hat{x} = 0$ with a distortion that equals the power of the source. Every point on the right hand side of the curve is achievable by simple adding an additional random distortion to the optimal compression.

1.4.2 Information Bottleneck

In many compression problems the question for an appropriate distortion measure arises. However this can not be generally answered. The information-bottleneck method [30] avoids the problem of choosing the "right" distortion measure by a more direct approach. The source x should be compressed in a way that preserves the relevant information. This approach is called *relevance through another variable*. In our context, the relevance variable is denoted by y, and, hence, the fidelity criterion is the mutual information between y and $\hat{\mathsf{x}}$. We want to find an optimal probabilistic map-

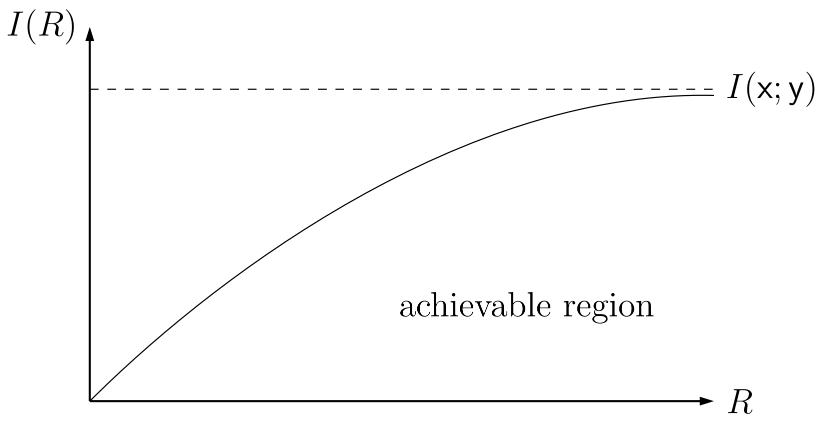

Figure 1.2: Illustration of the information-rate function.

ping $f(\hat{x}|x)$ which maximizes $I(\mathsf{y};\hat{\mathsf{x}})$ subject to a rate constraint. Thus, we define the information-rate function as

$$I(R) \triangleq \max_{f(\hat{x}|x)} I(\mathsf{y};\hat{\mathsf{x}}) \qquad \text{s.t.} \quad I(\mathsf{x};\hat{\mathsf{x}}) \leq R. \tag{1.11}$$

Conversely, the rate-information function is defined as

$$R(I) \triangleq \min_{f(\hat{x}|x)} I(\mathsf{x};\hat{\mathsf{x}}) \qquad \text{s.t.} \quad I(\mathsf{y};\hat{\mathsf{x}}) \geq I. \tag{1.12}$$

The definitions in (1.11) and (1.12) are reminiscent of the distortion-rate function and the rate-distortion function, respectively. Fig. 1.2 shows the general form of the information-rate function. Clearly if we provide no rate, we cannot preserve any information. Thus, the information-rate function always starts at the origin. For increasing rate we expect to preserve more information, hence the information-rate function has to be an increasing function. For infinite rates, the "compressed" $\hat{\mathsf{x}}$ perfectly represents the source x. As a consequence the mutual information $I(\hat{\mathsf{x}};\mathsf{y})$ is bounded by the information of y contained in the original, uncompressed source x.

1.4.3 Channel Capacity

In accordance to the usual definition we define the channel capacity as the maximum mutual information between the source at channel input and the channel output over all source distributions.

Definition 1.1. *The channel capacity with the source* y *and channel output* x *is defined* *by*

$$C \triangleq \max_{f(y)} I(\mathsf{y};\mathsf{x}). \tag{1.13}$$

Channel Capacity of the Gaussian Channel

One of the most important type of channels is the discrete time AWGN channel, which is considered in this thesis. The channel output x is thus given by the sum of the channel input y and the Gaussian distributed noise w, with power σ^2, as

$$\mathsf{x} = \mathsf{y} + \mathsf{w}, \qquad \mathsf{w} \sim \mathcal{N}(0, \sigma^2). \tag{1.14}$$

Obviously the channel capacity would be infinite if we do not constrain the source in any way, since we could just find a source with infinite power. Thus, we modify the definition of the channel capacity with an additional power constraint $\mathrm{E}\{\mathsf{y}^2\} \leq P$.

Definition 1.2. *The channel capacity of the Gaussian channel with the source* y *and channel output* x *is defined by*

$$C \triangleq \max_{f(y):\mathrm{E}\{\mathsf{y}^2\}\leq P} I(\mathsf{y};\mathsf{x}). \tag{1.15}$$

This yields the famous result from Shannon, and is therefore also called Shannon capacity [29],

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right). \tag{1.16}$$

That (1.16) is indeed an upper bound of mutual information can be shown by expanding the mutual information as follows [10]:

$$\begin{aligned} I(\mathsf{y};\mathsf{x}) &= h(\mathsf{x}) - h(\mathsf{x}|\mathsf{y}) && (1.17)\\ &= h(\mathsf{y}) - h(\mathsf{y}+\mathsf{w}|\mathsf{y}) && (1.18)\\ &= h(\mathsf{x}) - h(\mathsf{w}|\mathsf{y}) && (1.19)\\ &= h(\mathsf{x}) - h(\mathsf{w}). && (1.20) \end{aligned}$$

Here we used the fact that the noise is independent of the source. Thus, we have

$$\mathrm{E}\{\mathsf{y}^2\} = \mathrm{E}\{\mathsf{x}^2\} + \mathrm{E}\{\mathsf{w}^2\} = P + \sigma^2. \tag{1.21}$$

Also it is known that the Gaussian distribution maximizes the differential entropy for a source with a power constraint. Hence, we can bound the mutual information by

$$\begin{aligned} I(\mathsf{y};\mathsf{x}) &= h(\mathsf{x}) - h(\mathsf{z}) && (1.22)\\ &\leq \frac{1}{2}\log_2\left(2\pi e(P+\sigma^2)\right) - \frac{1}{2}\log_2\left(2\pi e\sigma^2\right) && (1.23)\\ &= \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\right). && (1.24)\end{aligned}$$

In order to show that this is not only an upper bound, but we a are able to construct codes which achieve this bound we use the common "sphere packing" argument. Consider the N dimensional vectorspace, which contains all possible channel output sequences of length N. Then a channel output sequence, given a specific codeword, is normally distributed around that particular codeword. Following the law of large numbers the received sequence is with probability $P \to 1$ on the hypersphere with radius $\sqrt{N\sigma^2}$ and origin the true codeword as $N \to \infty$. The subspace of all possible received sequences is then also a hypersphere with radius $\sqrt{N(P+\sigma^2)}$. The decoder decides on a codeword, if the received vector is in the hypersphere around the codeword. For a nonambiguous decoding, i.e., error free decoding, these "decoding spheres" must not intersect. Non-intersecting decoding spheres are obtained by an appropriate placement of the codewords in the hypershere with radius $\sqrt{N(P+\sigma^2)}$. The volume of an hypersphere with radius r is given by

$$V = \frac{\pi^{N/2}}{\Gamma(N/2+1)}r^N, \tag{1.25}$$

where $\Gamma(x)$ is the gamma function $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}\mathrm{d}y$. If we want to place M codewords with non-intersecting decoding spheres in the N dimensional hypersphere with radius $\sqrt{N(P+\sigma^2)}$, the ratio of the volumes must fulfill

$$M \leq \frac{V_{P+\sigma^2}}{V_{\sigma^2}} = \frac{\sqrt{N(P+\sigma^2)}^N}{\sqrt{N\sigma^2}^N} = \left(1+\frac{P}{\sigma^2}\right)^{N/2}. \tag{1.26}$$

Hence, we obtain for the rate

$$R = \frac{\log_2 M}{N} \leq \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\right) = C. \tag{1.27}$$

This means we can construct error free asymptotic codes with rate at most equal to the capacity.

Channel Capacity of the Gaussian Channel with Feedback

Since reliable communication close to the channel capacity relies on the sphere hardening effect, this requires large blocklengths. Large blocklengths go along with large coding and decoding delays, which are generally undesired. Thus, we may want to use feedback to obtain comparable performance, but significantly shorter blocklengths and therefore also reduced complexity. The first question which arises is whether feedback increases the channel capacity. As already mentioned, unfortunately this is not the case. As a first motivating example why feedback may be nevertheless beneficial, we present the famous Shalkwijk-Kailath result [28]:

$$P_e^{(N)} \leq 2\exp\left(-e^{2N(C-R)}\right). \tag{1.28}$$

This is true for a simple linear feedback scheme, where the transmitter has perfect strictly causal feedback. Taking a closer look at (1.28) we conclude some important statements:

- If $R < C$ the error probability decreases doubly exponential if we increase the blocklength.
- By increasing the blocklength, the error probability can be made arbitrary small.
- For every $\epsilon > 0$, where $R = C - \epsilon$ the error probability $P_e^{(N)} \to 0$ as $N \to \infty$. Thus, we achieve the channel capacity.

This gives a first insight, why we may want to use feedback in our communication system. In Chapter 3, we will study communication scenarios with feedback in detail.

2

Performance without Feedback

2.1 Introduction

When we want to communicate from A to B, usually the communication is not perfect. The transmitter encodes the message θ and transmits the signal $\boldsymbol{y}$ over the channel, where it is received as the signal $\boldsymbol{x}$ and decoded as the estimate of the message $\hat{\theta}$. A general way to describe a probabilistic and memoryless channel is to describe via a conditional probability density function (pdf) $f(x|y)$ (cf. Fig. 2.1).

The amount of information we can transmit over the channel depends on the statistics of the source and is defined as

$$C \triangleq \max_{f(y)} I(\mathsf{y};\mathsf{x}), \tag{2.1}$$

where $I(\mathsf{y};\mathsf{x})$ is the mutual information shared by transmitter and receiver. The mutual information clearly depends on the channel and the source and is given by

$$I(\mathsf{y};\mathsf{x}) = \iint_{(y,x)} f(y)f(x|y)\log_2\frac{f(x|y)}{f(x)}\mathrm{d}y\mathrm{d}x. \tag{2.2}$$

Because of the use of the logarithm to base 2, the mutual information is measured in bits.

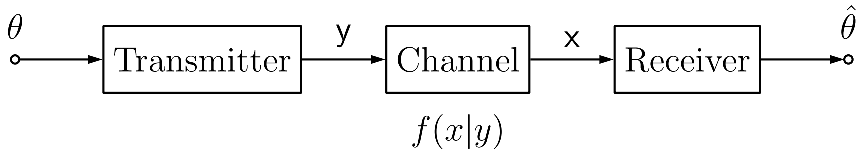

Figure 2.1: Basic communication system.

In this thesis, discrete time channels with additive white Gaussian noise (AWGN) are considered. The received signal x is the sum of the transmitted signal y and an independent noise term w

$$\mathsf{x} = \mathsf{y} + \mathsf{w}, \qquad \mathsf{w} \sim \mathcal{N}(0, \sigma^2). \tag{2.3}$$

The capacity of the channel is again defined as in (2.1), i.e., we have

$$C = \max_{f(y):\mathrm{E}\{\mathsf{y}^2\} \leq P} I(\mathsf{y}; \mathsf{x}), \tag{2.4}$$

with an average power constraint on y. This yields the famous result from Shannon, and is therefore also called Shannon capacity [29],

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2}\right). \tag{2.5}$$

It turns out that the only input distribution which achieves Shannon capacity is also Gaussian and satisfies the average power constraint with equality, hence $\mathsf{y} \sim \mathcal{N}(0, P)$. Thus, the Shannon capacity is an upper bound for all other input distributions.

2.2 System Model

The input of the receiver, the signal x, is continuous-valued, even if the transmit signal y has a finite alphabet. This is because the channel introduces continuous valued noise, in general as $f(x|y)$, and especially in the AWGN case as $\mathsf{w} \sim \mathcal{N}(0, \sigma^2)$. The performance of such systems was intensively studied and is well known. Another important aspect, in order to digitally process the received signal, is the compression (or quantization) of the received signal. This quantization process introduces additional distortion and can be described in the most general way as a probabilistic quantizer with the conditional pdf $f(t|x)$ (cf. Fig. 2.2).

Often the quantizer is assumed to have high quantization rate. Thus, the quantized

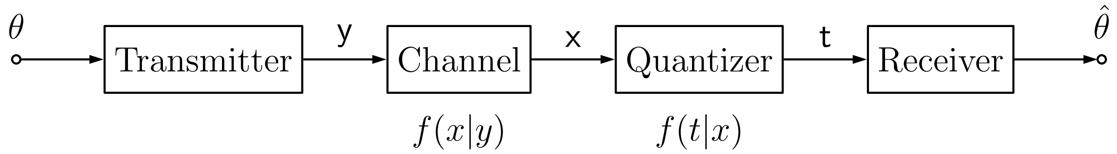

Figure 2.2: Basic communication system with quantizer.

signal $t \approx x$ for further processing. If the high-rate assumption is dropped, $f(t|x)$ has to be optimized in some way. Usually this is done via minimizing the mean-square error (MSE) of the received signal. Rate distortion (RD) theory provides the mathematical framework for such problems [2]. The mutual information of the AWGN channel with RD-optimal compression at the receiver will be studied in the next section. However, our goal is not to minimize the MSE, rather maximizing the mutual information of source and quantized signal. This leads to the information bottleneck (IB) method [30], which is supposed to perform better. Thus, the next sections focus on the analysis of the AWGN channel with IB compression and we will quantify the performance improvement. Therefore, we will formulate, derive and discuss the information-rate function (termed IB-function in [15]) as an analogon to the distortion-rate function in rate-distortion theory.

2.3 The Scalar Case

2.3.1 Rate Distortion Theory

The problem of quantization is mostly addressed by rate distortion theory, as a classical source-coding problem. Rate distortion theory quantifies the trade-off between compression rate R and distortion D. The quantized signal t is therfore an estimation for the received signal x, i.e., $\hat{x} = t$. The choice of the distortion measure is somewhat arbitrary. The most common distortion metric is the squared error

$$d(x, \hat{x}) = (x - \hat{x})^2. \tag{2.6}$$

The rate distortion function is defined as

$$R(D) \triangleq \min_{f(\hat{x}|x)} I(\mathsf{x}; \hat{\mathsf{x}}) \qquad \text{s.t.} \quad \mathrm{E}\{d(\mathsf{x}, \hat{\mathsf{x}})\} \leq D, \tag{2.7}$$

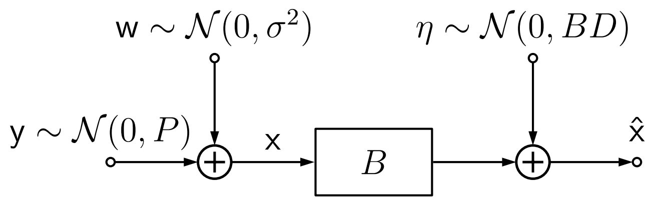

Figure 2.3: Scalar system with RD-optimal compression.

with the average distortion

$$\mathrm{E}\{d(\mathsf{x},\hat{\mathsf{x}})\} = \iint_{(x,\hat{x})} f(x) f(\hat{x}|x) d(x,\hat{x}) \mathrm{d}x \mathrm{d}\hat{x}. \tag{2.8}$$

The inverse of the rate-distortion function is the distortion-rate function

$$D(R) \triangleq \min_{f(\hat{x}|x)} \mathrm{E}\{d(\mathsf{x},\hat{\mathsf{x}})\} \qquad \text{s.t.} \quad I(\mathsf{x};\hat{\mathsf{x}}) \leq R. \tag{2.9}$$

Thus, the goal is to minimize the distortion at a given rate. Recall that $\mathsf{x} = \mathsf{y} + \mathsf{w}$, with $\mathsf{y} \sim \mathcal{N}(0, P)$ and $\mathsf{w} \sim \mathcal{N}(0, \sigma^2)$. Therefore, x is also Gaussian with $\mathsf{x} \sim \mathcal{N}(0, P + \sigma^2)$. It was shown that the optimal rate distortion function can be obtained with a "forward channel" (cf. Fig. 2.3) [2]

$$\hat{\mathsf{x}} = B\mathsf{x} + \eta, \tag{2.10}$$

where $\eta \sim \mathcal{N}(0, BD)$ is independent of x. Hence, $\hat{\mathsf{x}}$ is Gaussian with $\hat{\mathsf{x}} \sim \mathcal{N}(0, B^2 P_{\mathsf{x}} + BD)$ and the distortion

$$D = \mathrm{E}\{d(\mathsf{x},\hat{\mathsf{x}})\} = B^2 P_{\mathsf{x}} + BD + P_{\mathsf{x}} - 2BP_{\mathsf{x}}. \tag{2.11}$$

Solving (2.11) for B yields

$$B = 1 - \frac{D}{P_{\mathsf{x}}} \tag{2.12}$$

and hence $\hat{\mathsf{x}} \sim \mathcal{N}(0, P_{\mathsf{x}} - D)$. The minimum rate R for distortion D is given by

$$R = I(\mathsf{x};\hat{\mathsf{x}}) = h(\hat{\mathsf{x}}) - h(\hat{\mathsf{x}}|\mathsf{x}), \tag{2.13}$$

with $\hat{\mathsf{x}}|\mathsf{x} \sim \mathcal{N}\left(Bx, D - D^2/P_{\mathsf{x}}\right)$. It follows that

$$
\begin{aligned}
R(D) &= \frac{1}{2}\log_2\left(\frac{P_{\mathsf{x}}}{D}\right), && (2.14)\\
D(R) &= 2^{-2R}P_{\mathsf{x}}, && (2.15)
\end{aligned}
$$

which is indeed the famous rate distortion solution for a Gaussian source with variance P_x, which is in our case the output of the AWGN channel. Inserting (2.12) into (2.10) yields

$$
\hat{\mathsf{x}} = \left(1 - \frac{D}{P+\sigma^2}\right)(\mathsf{y}+\mathsf{w}) + \eta. \tag{2.16}
$$

To compare the performance of this solution in terms of information conservation we compute the mutual information $I(\mathsf{y};\hat{\mathsf{x}})$ as

$$
I(R) \triangleq I(\mathsf{y};\hat{\mathsf{x}}) = h(\hat{\mathsf{x}}) - h(\hat{\mathsf{x}}|\mathsf{y}), \tag{2.17}
$$

with $\hat{\mathsf{x}}|\mathsf{y} \sim \mathcal{N}\left((1-D(R)/P_{\mathsf{x}})y, (1-D(R)/P_{\mathsf{x}})^2\sigma^2 + D(R) - D(R)^2/P_{\mathsf{x}}\right)$. $I(R)$ can then be calculated as

$$
\begin{aligned}
I(R) &= \frac{1}{2}\log_2\left(\frac{P_{\mathsf{x}} - D(R)}{\left(1-\frac{D(R)}{P_{\mathsf{x}}}\right)^2\sigma^2 + D(R)\left(1-\frac{D(R)}{P_{\mathsf{x}}}\right)}\right) && (2.18)\\
&= \frac{1}{2}\log_2\left(\frac{P_{\mathsf{x}}\left(1-2^{-2R}\right)}{\left(1-2^{-2R}\right)^2\sigma^2 + 2^{-2R}P_{\mathsf{x}}\left(1-2^{-2R}\right)}\right) && (2.19)\\
&= \frac{1}{2}\log_2\left(\frac{P+\sigma^2}{2^{-2R}P+\sigma^2}\right). && (2.20)
\end{aligned}
$$

Finally we obtain $I(R)$ for the rate distortion solution, as a function of the rate R and the signal-to-noise ratio (SNR) $\frac{P}{\sigma^2}$ and is termed as the information-rate function. Thus, we can formulate following statement.

Corollary 2.1. *The scalar information-rate function with RD-optimal channel output compression is given by*

$$
\boxed{I(R) = \frac{1}{2}\log_2\left(\frac{1+P/\sigma^2}{1+2^{-2R}P/\sigma^2}\right).} \tag{2.21}
$$

The inverse of the information-rate function is called the "rate-information function".

Corollary 2.2. *The scalar rate-information function with RD-optimal channel output compression is given by*

$$\boxed{R(I) = \frac{1}{2}\log_2\left(\frac{P/\sigma^2}{2^{-2I}(1+P/\sigma^2)-1}\right).} \tag{2.22}$$

2.3.2 Information Bottleneck

In communications, we actually we do not want to quantize the received signal with minimum distortion. What we actually want is to preserve the information about y carried by x. This is exactly what the information bottleneck method [30] provides: *relevance through another variable.* The problem of choosing the "right" distortion measure is replaced by relevant information. As in the rate-distortion case the quantization should compress x as much as possible while preserving as much information about y as possible. The rate-distortion function is now replaced by the rate-information function, which is defined as

$$R(I) \triangleq \min_{f(t|x)} I(\mathsf{x};\mathsf{t}) \qquad \text{s.t.} \quad I(\mathsf{y};\mathsf{t}) \geq I, \tag{2.23}$$

and conversely the distortion-rate function is replaced by the information-rate function

$$I(R) \triangleq \max_{f(t|x)} I(\mathsf{y};\mathsf{t}) \qquad \text{s.t.} \quad I(\mathsf{x};\mathsf{t}) \leq R. \tag{2.24}$$

Derivation of the GIB Information-Rate Function for AWGN Channels

For general distributions it is hard to give an analytical expression for the information-rate functions, if even possible. A solution for the case of jointly Gaussian $\boldsymbol{\mathsf{y}}$ and $\boldsymbol{\mathsf{x}}$ has been found in [8], with the problem formulated as the variational problem

$$\min_{p(\boldsymbol{t}|\boldsymbol{x})} I(\boldsymbol{\mathsf{x}};\boldsymbol{\mathsf{t}}) - \beta I(\boldsymbol{\mathsf{t}};\boldsymbol{\mathsf{y}}), \tag{2.25}$$

where the parameter β describes the trade-off between compression and preserved relevant information.

Because of the additive structure of the AWGN channel $\boldsymbol{\mathsf{x}}$ and $\boldsymbol{\mathsf{y}}$ are jointly Gaussian and in the vector case jointly multivariate Gaussian. It was shown in [16] that for jointly Gaussian $\boldsymbol{\mathsf{x}}$ and $\boldsymbol{\mathsf{y}}$ the optimal $\boldsymbol{\mathsf{t}}$ is also jointly Gaussian with $\boldsymbol{\mathsf{x}}$. Thus, $\boldsymbol{\mathsf{t}}$ can be described using the linear transformation

$$\boldsymbol{\mathsf{t}} = \boldsymbol{A}\boldsymbol{\mathsf{x}} + \boldsymbol{\xi}, \qquad \boldsymbol{\xi} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}). \tag{2.26}$$

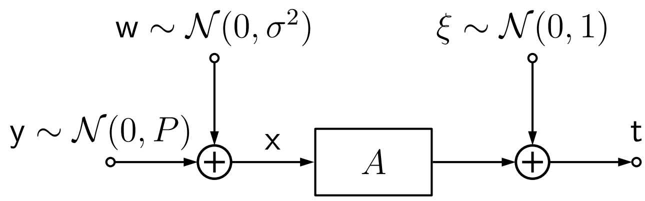

Figure 2.4: GIB equivalent scalar system.

The problem in (2.25) can then be reformulated as

$$\min_{\boldsymbol{A}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}} I(\mathbf{x}; \mathbf{t}) - \beta I(\mathbf{t}; \mathbf{y}). \tag{2.27}$$

The following theorem gives explicit expressions for the optimal $\boldsymbol{A}$ and Σ_ξ.

Theorem 2.3 ([8], Thm. 3.1)**.** *The optimal projection $\mathbf{t} = \boldsymbol{A}\mathbf{x} + \boldsymbol{\xi}$ for a given tradeoff parameter β is given by $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = \boldsymbol{I}_x$ and*

$$\boldsymbol{A} = \begin{cases} \left(\mathbf{0}^T; \quad \dots; \quad \mathbf{0}^T\right), & 0 \le \beta \le \beta_1^c \\ \left(\alpha_1 \boldsymbol{v}_1^T; \quad \mathbf{0}^T; \quad \dots; \quad \mathbf{0}^T\right), & \beta_1^c \le \beta \le \beta_2^c \\ \left(\alpha_1 \boldsymbol{v}_1^T; \quad \alpha_2 \boldsymbol{v}_2^T; \quad \mathbf{0}^T; \quad \dots; \quad \mathbf{0}^T\right), & \beta_2^c \le \beta \le \beta_3^c \\ \vdots \end{cases} \tag{2.28}$$

where $\{\boldsymbol{v}_1^T, \boldsymbol{v}_2^T, \dots, \boldsymbol{v}_{n_x}^T\}$ are left eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}} \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}$ sorted by their corresponding ascending eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_{n_x}$, $\beta_i^c = \frac{1}{1-\lambda_i}$ are critical β values, α_i are coefficients defined by $\alpha_i = \sqrt{\frac{\beta(1-\lambda_i)-1}{\lambda_i r_i}}$, $r_i = \boldsymbol{v}_i{}^T \boldsymbol{\Sigma}_{\mathbf{x}} \boldsymbol{v}_i$, $\mathbf{0}^T$ is an n_x dimensional row vector of zeros, and semicolons separate rows in the matrix $\boldsymbol{A}$.

Scalar Channel

As mentioned before y and x are jointly Gaussian and the optimum t is also jointly Gaussian with y and therefore form a Markov chain

$$\mathsf{y} - \mathsf{x} - \mathsf{t}, \tag{2.29}$$

where $\mathsf{x} = \mathsf{y} + \mathsf{w}$, $\mathsf{y} \sim \mathcal{N}(0, P)$ and $\mathsf{w} \sim \mathcal{N}(0, \sigma^2)$ (cf. Fig. 2.4).

In the scalar case the covariance matrices simplify to

$$
\begin{aligned}
\Sigma_{\mathsf{y}} &= P, && (2.30)\\
\Sigma_{\mathsf{w}} &= \sigma^2, && (2.31)\\
\Sigma_{\mathsf{x}} &= \Sigma_{\mathsf{y}} + \Sigma_{\mathsf{w}} = P + \sigma^2, && (2.32)\\
\Sigma_{\mathsf{x}|\mathsf{y}} &= \Sigma_{\mathsf{w}} = \sigma^2. && (2.33)
\end{aligned}
$$

The desired eigenvalues from Theorem 2.3 therefore collapse to one single eigenvalue,

$$\Sigma_{\mathsf{x}|\mathsf{y}}\Sigma_{\mathsf{x}}^{-1} = \frac{\sigma^2}{\sigma^2 + P} = \frac{1}{1 + P/\sigma^2} = \lambda. \tag{2.34}$$

Using the explicit tradeoff parameter β, $I_\beta(\mathsf{t};\mathsf{x})$ and $I_\beta(\mathsf{t};\mathsf{y})$ are derived in [8]

$$I_\beta(\mathsf{t};\mathsf{x}) = \frac{1}{2}\sum_{i=1}^{n(\beta)} \log_2\left((\beta - 1)\frac{1-\lambda_i}{\lambda_i}\right), \tag{2.35}$$

where

$$n(\beta) = \max\{n : \beta \geq \beta_n^c\}. \tag{2.36}$$

In the scalar case there is only one critical β^c, because there is only one eigenvalue λ. The only interesting case is $\beta \geq \beta^c$, otherwise $I_\beta(\mathsf{t};\mathsf{x}) = 0$. Therfore $n(\beta) = 1$.

$$I_\beta(\mathsf{t};\mathsf{x}) = \frac{1}{2}\log_2\left((\beta - 1)\frac{1-\lambda}{\lambda}\right) \triangleq R. \tag{2.37}$$

Since the mutual information of t and x is a measure for the compression, it can also be seen as the rate of the quantization. Making β explicit yields

$$\beta = 2^{2R}\frac{\lambda}{1-\lambda} + 1. \tag{2.38}$$

The mutual information between t and y is

$$I_\beta(\mathsf{t};\mathsf{y}) = I(\mathsf{t};\mathsf{x}) - \frac{1}{2}\sum_{i=1}^{n(\beta)} \log_2\left(\beta(1-\lambda_i)\right), \tag{2.39}$$

where again in the scalar case $n(\beta) = 1$. Substituting $I(\mathsf{t};\mathsf{x})$ with (2.37), β with (2.38) and λ with (2.34) yields

$$\begin{aligned} I_\beta(\mathsf{t};\mathsf{y}) &= R - \frac{1}{2}\log_2(\beta(1-\lambda)) && (2.40)\\ &= R - \frac{1}{2}\log_2\left(2^{2R}\lambda + 1 - \lambda\right) && (2.41)\\ &= \frac{1}{2}\log_2\left(\frac{2^{2R}}{2^{2R}\lambda + 1 - \lambda}\right) && (2.42)\\ &= \frac{1}{2}\log_2\left(\frac{2^{2R}}{2^{2R}\frac{1}{1+P/\sigma^2} + 1 - \frac{1}{1+P/\sigma^2}}\right) && (2.43)\\ &= \frac{1}{2}\log_2\left(\frac{2^{2R}(1+P/\sigma^2)}{2^{2R}+P/\sigma^2}\right) \triangleq I(R). && (2.44)\end{aligned}$$

This can also be directly obtained, without using $I_\beta(\mathsf{t};\mathsf{x})$ and $I_\beta(\mathsf{t};\mathsf{y})$ from [8]. We just need the identity $\mathsf{t} = A\mathsf{x} + \xi = A(\mathsf{y}+\mathsf{w}) + \xi$. The rate and the mutual information are again defined as

$$\begin{aligned} R &\triangleq I(\mathsf{t};\mathsf{x}) = h(\mathsf{t}) - h(\mathsf{t}|\mathsf{x}), && (2.45)\\ I &\triangleq I(\mathsf{t};\mathsf{y}) = h(\mathsf{t}) - h(\mathsf{t}|\mathsf{y}). && (2.46)\end{aligned}$$

These random variables are all Gaussian distributed as follows:

$$\begin{aligned} \mathsf{t} &\sim \mathcal{N}(0, A^2(P+\sigma^2)+1), && (2.47)\\ \mathsf{t}|\mathsf{x} &\sim \mathcal{N}(0,1), && (2.48)\\ \mathsf{t}|\mathsf{y} &\sim \mathcal{N}(0, A^2\sigma^2+1). && (2.49)\end{aligned}$$

Using (2.47)–(2.49) to express R and I yields

$$\begin{aligned} R &= \frac{1}{2}\log_2\left(A^2(P+\sigma^2)+1\right), && (2.50)\\ I &= \frac{1}{2}\log_2\left(\frac{A^2(P+\sigma^2)+1}{A^2\sigma^2+1}\right). && (2.51)\end{aligned}$$

Making A^2 explicit in (2.50) yields

$$A^2 = \frac{2^{2R}-1}{P+\sigma^2}. \tag{2.52}$$

Substituting A^2 in (2.51) with (2.52) yields the information-rate function in terms of the rate R and the SNR $\frac{P}{\sigma^2}$.

Corollary 2.4. *The scalar information-rate function with GIB-optimal channel output compression is given by*

$$\boxed{I(R) = \frac{1}{2}\log_2\left(\frac{1+P/\sigma^2}{1+2^{-2R}P/\sigma^2}\right).} \tag{2.53}$$

The inverse of the information-rate function is called the rate-information function and is given by the following corollary.

Corollary 2.5. *The scalar rate-information function with GIB-optimal channel output compression is given by*

$$\boxed{R(I) = \frac{1}{2}\log_2\left(\frac{P/\sigma^2}{2^{-2I}(1+P/\sigma^2)-1}\right).} \tag{2.54}$$

2.4 The Vector Case

2.4.1 System Model

Now we consider the vector case with some restrictions for the sake of simplicity. These restrictions are dropped in the next subsection and can be removed using algebraic transformations on source and channel. Let the random vector source be Gaussian distributed as $\mathbf{y} = \begin{pmatrix} \mathsf{y}_1 & \mathsf{y}_2 & \dots & \mathsf{y}_n \end{pmatrix}^T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D})$, where $\boldsymbol{D}$ is a diagonal matrix. In the simplest case the available transmit power is evenly distributed on the independent $y_1 \dots y_n$. Therefore, the elements of $\mathbf{y}$ are i.i.d. and $\mathbf{y}$ has the covariance matrix

$$\boldsymbol{\Sigma}_{\mathbf{y}} = \boldsymbol{D} = \frac{P}{n}\boldsymbol{I}. \tag{2.55}$$

The noise is again additive and independent of $\mathbf{y}$, thus the input/output relation reads

$$\mathbf{x} = \boldsymbol{H}\mathbf{y} + \mathbf{w}, \tag{2.56}$$

where $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is deterministic (cf. Fig. 2.5).
In this restricted model the noise is also modeled to be i.i.d., i.e., $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$. The covariance matrices are then

$$\boldsymbol{\Sigma}_{\mathbf{w}} = \sigma^2\boldsymbol{I}, \tag{2.57}$$

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \boldsymbol{H}\boldsymbol{\Sigma}_{\mathbf{y}}\boldsymbol{H}^T + \boldsymbol{\Sigma}_{\mathbf{w}} = \frac{P}{n}\boldsymbol{H}\boldsymbol{H}^T + \sigma^2\boldsymbol{I}, \tag{2.58}$$

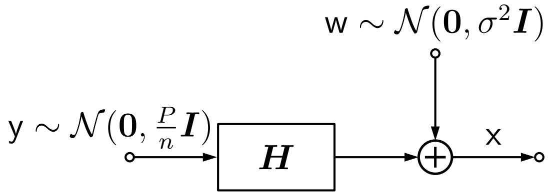

Figure 2.5: Vector system model.

since $\mathbf{y}$ and $\mathbf{w}$ are assumed to be independent.

2.4.2 Generalization of the System

Previously the system was restricted to an i.i.d. source $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \frac{P}{n}\boldsymbol{I})$ and noise $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$. Then $\mathbf{x} = \boldsymbol{H}\mathbf{y} + \mathbf{w} \sim \mathcal{N}(0, \frac{P}{n}\boldsymbol{H}\boldsymbol{H}^T + \sigma^2\boldsymbol{I})$. Now we drop this restriction and let $\mathbf{y}$ and $\mathbf{w}$ be independent Gaussian random vectors with full-rank covariance matrices. We therefore have

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{y}}), \tag{2.59}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{w}}). \tag{2.60}$$

P and σ^2 can then be defined as

$$P = \mathrm{E}\{\mathbf{y}^T\mathbf{y}\}, \tag{2.61}$$

$$\sigma^2 = \frac{1}{n}\mathrm{E}\{\mathbf{w}^T\mathbf{w}\}. \tag{2.62}$$

Then the covariance matrix of $\mathbf{x}$ is

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \boldsymbol{H}\boldsymbol{\Sigma}_{\mathbf{y}}\boldsymbol{H}^T + \boldsymbol{\Sigma}_{\mathbf{w}}. \tag{2.63}$$

Whitening the noise in $\mathbf{x}$ and decorrelating the signal yields

$$\tilde{\mathbf{x}} = \sqrt{\sigma^2}\boldsymbol{U}^T\boldsymbol{\Sigma}_{\mathbf{w}}^{-1/2}\mathbf{x}, \tag{2.64}$$

where $\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ is the eigen decomposition of $\sqrt{\frac{n\sigma^2}{P}}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1/2}\boldsymbol{H}\boldsymbol{\Sigma}_{\mathbf{y}}\boldsymbol{H}^T\boldsymbol{\Sigma}_{\mathbf{w}}^{-1/2}\sqrt{\frac{n\sigma^2}{P}}$. Hence,

$$\tilde{\mathbf{x}} \sim \mathcal{N}\left(\mathbf{0}, \frac{P}{n}\boldsymbol{\Lambda} + \sigma^2\boldsymbol{I}\right). \tag{2.65}$$

The previous transformations are all invertible and, hence, do not change the mutual information. Therefore, we can write an equivalent system as

$$\tilde{\mathbf{x}} = \tilde{\boldsymbol{H}}\tilde{\mathbf{y}} + \tilde{\mathbf{w}}. \tag{2.66}$$

This equivalent system has a diagonal channel $\tilde{\boldsymbol{H}}$ and i.i.d. signal $\tilde{\mathbf{y}}$ and noise $\tilde{\mathbf{w}}$, which are given as

$$\tilde{\boldsymbol{H}} = \boldsymbol{\Lambda}^{1/2}, \tag{2.67}$$

$$\tilde{\mathbf{w}} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}), \tag{2.68}$$

$$\tilde{\mathbf{y}} \sim \mathcal{N}\left(\mathbf{0}, \frac{P}{n}\boldsymbol{I}\right). \tag{2.69}$$

Since the channel is diagonal, the eigenvalues of $\tilde{\boldsymbol{H}}\tilde{\boldsymbol{H}}^T$ are the diagonal elements of $\boldsymbol{\Lambda}$,

$$\lambda_{\tilde{\boldsymbol{H}}_i} = \boldsymbol{\Lambda}_{ii}. \tag{2.70}$$

To simplify the calculations in the next sections we will work with the system model defined in the previous section which is equivalent to (2.66) with (2.67)–(2.69).

2.4.3 Rate Distortion Theory

As already described for the scalar case, rate distortion theory aims at compressing a signal with minimum distortion at a given rate. As a common distortion metric we again use the squared-error

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \mathrm{E}\{(\mathbf{x} - \hat{\mathbf{x}})^T(\mathbf{x} - \hat{\mathbf{x}})\} = \sum_i d(\mathrm{x}_i, \hat{\mathrm{x}}_i). \tag{2.71}$$

We note that each scalar distortion $d(\mathrm{x}_i, \hat{\mathrm{x}}_i)$ contributes to the overall distortion, but in general scalar compression of the x_i's is suboptimal. The optimal solution, which jointly compresses $\mathbf{x}$ can again be obtained using a "forward channel" [2]

$$\hat{\mathbf{x}} = \boldsymbol{B}\mathbf{x} + \boldsymbol{\eta}. \tag{2.72}$$

The main challenge is to optimally allocate the rate to the individual modes. The optimal rate allocation is obtained by applying the reverse waterfilling algorithm. For general multivariate Gaussian distributions the algorithm is applied to the Karhunen-Loève eigenvalues λ_k. The λ_k are the eigenvalues of the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}} = \frac{P}{n}\boldsymbol{H}\boldsymbol{H}^T + \sigma^2\boldsymbol{I}$. Hence, the rate is allocated to the scalar modes according to the values of λ_k.

The eigenvalue problem for λ_k ($\lambda_k \neq \sigma^2$) reads

$$\begin{aligned}
\det(\mathbf{\Sigma_x} - \lambda \boldsymbol{I}) &= 0 && (2.73)\\
\det\left(\frac{P}{n}\boldsymbol{H}\boldsymbol{H}^T - (\lambda - \sigma^2)\boldsymbol{I}\right) &= 0 && (2.74)\\
\det\left(\frac{P}{n(\sigma^2 - \lambda)}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T + \boldsymbol{I}\right) &= 0 && (2.75)\\
\prod_{i=1}^{r}\left(\frac{P}{n(\sigma^2 - \lambda)}\lambda_{H_i} + 1\right) &= 0, && (2.76)
\end{aligned}$$

where r is the rank of the channel and λ_{H_i} are the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^T$.

Since (2.76) must be fulfilled for all λ_k one can simplify (2.76) to

$$\lambda_{H_k} = \frac{n(\lambda_k - \sigma^2)}{P} \qquad \Leftrightarrow \qquad \lambda_k = \frac{P}{n}\lambda_{H_k} + \sigma^2. \tag{2.77}$$

In the case $\lambda_k = \sigma^2$, (2.74) reduces to $\det(\boldsymbol{H}\boldsymbol{H}^T) = 0$. As a consequence $\lambda_{H_k} = 0$ and therefore (2.77) is still fulfilled.

The Rate-Distortion Function

The reverse waterfilling algorithm provides the rate allocation and reads [2]

$$\begin{aligned}
D(\theta) &= \sum_{k=1}^{n} \min\{\theta, \lambda_k\}, && (2.78)\\
R(\theta) &= \sum_{k=1}^{n} \max\left\{0, \frac{1}{2}\log_2\left(\frac{\lambda_k}{\theta}\right)\right\}, && (2.79)
\end{aligned}$$

where θ is the waterlevel. This is a parametric form of the rate-distortion function in the vector case.

Then, if the eigenvalues λ_k are sorted in descending order, the number of active "modes" is

$$n(\theta) = \arg\max_i \lambda_i, \quad \lambda_i \in (\lambda_k > \theta). \tag{2.80}$$

The critical rates R_c where new modes are added are then

$$\begin{aligned}
R_c(n) &= \frac{1}{2}\sum_{k=1}^{n} \log_2\left(\frac{\lambda_k}{\lambda_n}\right) && (2.81)\\
&= \frac{1}{2} n \log_2\left(\frac{\overline{\lambda}_i}{\lambda_n}\right). && (2.82)
\end{aligned}$$

Note the structural analogies of (2.144) and (2.82). This means $n(\theta)$ is incremented at each R_c and is therefore now a function of R ($n(\theta) \Rightarrow n(R)$):

$$n(R) = \max\{n : R \geq R_c(n)\}. \tag{2.83}$$

At rate R there are $n(R)$ active scalar modes.

Rewriting the distortion (2.78) and the rate (2.79) as the sum of the per-mode distortions and rates, as

$$D(\theta) = \sum_{k=1}^{n} D_k(\theta), \tag{2.84}$$

$$R(\theta) = \sum_{k=1}^{n} R_k(\theta), \tag{2.85}$$

yields the "mode" distortions and rates

$$D_k(\theta) = \min\{\theta, \lambda_k\}, \tag{2.86}$$

$$R_k(\theta) = \max\left\{0, \frac{1}{2}\log_2\left(\frac{\lambda_k}{\theta}\right)\right\}. \tag{2.87}$$

Since, in the scalar case, $X \sim \mathcal{N}(0, P+\sigma^2)$, the distortion-rate function is given by $D(R) = 2^{-2R}(P+\sigma^2)$. In the vector case we have $n(\theta)$ independent parallel Gaussian channels, where $P \Rightarrow \frac{P}{n}\lambda_{H_k}$ and $R \Rightarrow R_k(\theta)$.
Proof:

$$D(R_k(\theta)) = 2^{-2R_k(\theta)}\left(\frac{P}{n}\lambda_{H_k} + \sigma^2\right) \tag{2.88}$$

$$= \begin{cases} 2^{-2\left(\frac{1}{2}\log_2\left(\frac{\lambda_k}{\theta}\right)\right)}\left(\frac{P}{n}\lambda_{H_k} + \sigma^2\right) & \text{if} \quad \lambda_k \geq \theta \\ \frac{P}{n}\lambda_{H_k} + \sigma^2 & \text{else} \end{cases} \tag{2.89}$$

$$= \begin{cases} \frac{\theta}{\lambda_k}\lambda_k & \text{if} \quad \lambda_k \geq \theta \\ \lambda_k & \text{else} \end{cases} \tag{2.90}$$

$$= \min(\theta, \lambda_k) = D_k(\theta). \tag{2.91}$$

The Information-Rate Function

From the previous section we know that, in the scalar case, the information-rate function of the rate distortion solution is the same as the Gaussian information bottleneck solution (which has also been shown in [32]),

$$I(R) = I(\mathsf{y};\mathsf{t}) = h(\mathsf{y}) - h(\mathsf{y}|\mathsf{t}) = \frac{1}{2}\log_2\left(\frac{1+P/\sigma^2}{1+2^{-2R}P/\sigma^2}\right). \tag{2.92}$$

An intuitive explanation for this is the fact that the RD optimal t is also jointly Gaussian with x. Hence, the scalar $I(\mathsf{t};\mathsf{x})$ has the same structure in RD and GIB. Therefore, the information curve of the k^{th} "mode" is

$$I_k(R_k(\theta)) = \frac{1}{2}\log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_k}}{1+2^{-2R_k(\theta)}\frac{P}{n\sigma^2}\lambda_{H_k}}\right). \tag{2.93}$$

In the case of n independent Gaussian channels

$$\begin{aligned} I(R) = I(\mathbf{y};\mathbf{t}) &= h(\mathsf{y}_1;\mathsf{y}_2;\ldots;\mathsf{y}_n) - \\ &\quad h(\mathsf{y}_1;\mathsf{y}_2;\ldots;\mathsf{y}_n|\mathsf{t}_1;\mathsf{t}_2;\ldots;\mathsf{t}_n) &(2.94)\\ &= \sum_{k=1}^{n}(h(\mathsf{y}_k) - h(\mathsf{y}_k|\mathsf{t}_k)) &(2.95)\\ &\hat{=} \sum_{k=1}^{n} I_k(R_k(\theta)). &(2.96)\end{aligned}$$

The equivalence of information-rate function $I(R)$ of the RD solution with (2.96) is formally proved in [31]. The information-rate function of the rate distortion solution is then given by (2.79) and (2.96):

$$\begin{aligned} R(\theta) &= \sum_{k=1}^{n}\max\left\{0, \frac{1}{2}\log_2\left(\frac{\lambda_k}{\theta}\right)\right\}, &(2.97)\\ I(\theta) &= \frac{1}{2}\sum_{k=1}^{n}\log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_k}}{1+2^{-2R_k(\theta)}\frac{P}{n\sigma^2}\lambda_{H_k}}\right). &(2.98)\end{aligned}$$

This is already the information-rate function in parametric form, since it is obtained by the reverse waterfilling solution of the rate-distortion solution. To obtain an explicit expression for the information-rate function, in a first step we sort the eigenvalues in descending order, instead of the max operator in (2.97), and then rewrite (2.97) as

$$\begin{aligned} R(\theta) &= \sum_{k=1}^{n(R)}\frac{1}{2}\log_2\left(\frac{\lambda_k}{\theta}\right) &(2.99)\\ &= \frac{1}{2}n(R)\log_2\left(\frac{\overline{\lambda}_n}{\theta}\right), &(2.100)\end{aligned}$$

where it is ensured by $n(R)$ that $\log_2\left(\frac{\lambda_k}{\theta}\right) \geq 0$ and therefore the max operator in (2.97) can be dropped. $\overline{\lambda}_n = \overline{\lambda}_{n(R)}$ is the geometric mean $\prod_{i=1}^{n(\beta)} \lambda_i^{1/n(R)}$. Making θ explicit yields

$$\theta \;=\; 2^{\frac{-2R}{n(R)}}\overline{\lambda}_n. \tag{2.101}$$

Now we can replace θ in (2.87), which yields the explicit rate allocation

$$R_k(R) \;=\; \max\left\{0, \frac{1}{2}\left(\log_2\left(\frac{\lambda_k}{\overline{\lambda}_n}\right) + \frac{2R}{n(R)}\right)\right\}. \tag{2.102}$$

Basically the rate R is evenly distributed among all $n(R)$ active modes up to a correction term. Depending on whether the eigenvalue of the mode is greater or smaller than $\overline{\lambda}_n$ the rate $R/n(R)$ is increased or decreased by $\log_2(\lambda_k/\overline{\lambda}_n)$. Interestingly, all rates of the scalar modes have the same slope

$$\frac{\partial R_k(R)}{\partial R} = \frac{1}{n(R)}, \tag{2.103}$$

were $n(R)$ is constant. $R_k(R)$ is not differentiable at the critical rates . This means that a differential increase $\mathrm{d}R$ is evenly distributed among all $n(R)$ active modes. Inserting (2.102) in (2.98) yields an explicit expression for the information-rate function.

Corollary 2.6. *The information-rate function with RD-optimal channel output compression is given by*

$$\boxed{I(R) = \frac{1}{2}\sum_{k=1}^{n(R)} \log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_k}}{1+2^{-\frac{2R}{n(R)}}\frac{P}{n\sigma^2}\lambda_{H_k}\frac{\overline{\lambda}_n}{\lambda_k}}\right)}, \tag{2.104}$$

where the number of active modes is given by

$$n(R) = \max\{n : R \geq R_c(n)\}, \tag{2.105}$$

and the critical rates $R_c(n)$ are given by

$$R_c(n) = \frac{1}{2} n \log_2\left(\frac{\overline{\lambda}_i}{\lambda_n}\right). \tag{2.106}$$

Equivalently we can write (2.104) as the sum of $n(R)$ scalar modes with the rate

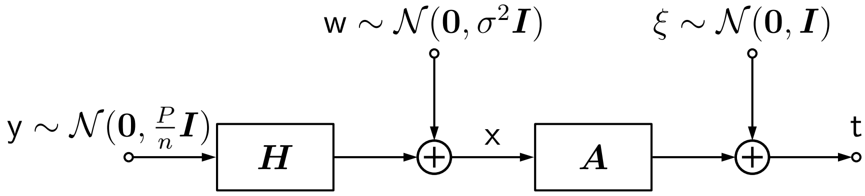

Figure 2.6: GIB equivalent vector system.

allocation (2.102) as follows:

$$I(R) = \sum_{k=1}^{n(R)} I_k(R) \tag{2.107}$$

$$= \frac{1}{2}\sum_{k=1}^{n(R)} \log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_k}}{1+2^{-2R_k(R)}\frac{P}{n\sigma^2}\lambda_{H_k}}\right). \tag{2.108}$$

We identify (2.108) as the sum of scalar information-rate functions with the mode rates $R_k(R)$. Of course the mode rates have to sum up to the total rate, i.e., we have

$$R = \sum_{k=1}^{n(R)} R_k(R). \tag{2.109}$$

2.4.4 Information Bottleneck

Now we consider jointly Gaussian random vectors $\mathsf{y}, \mathsf{x}, \mathsf{t}$, which form a Markov chain

$$\mathsf{y} - \mathsf{x} - \mathsf{t}, \tag{2.110}$$

where $\mathsf{x} = \boldsymbol{H}\mathsf{y} + \mathsf{w}$, $\mathsf{y} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D})$ and $\mathsf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$, $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ (cf. Fig. 2.6). Recall the covariance matrices, in the simplest case where it is assumed that the transmit power is evenly distributed among the uncorrelated $y_1 \dots y_n$,

$$\boldsymbol{\Sigma}_{\mathsf{y}} = \boldsymbol{D} = \frac{P}{n}\boldsymbol{I}, \tag{2.111}$$

$$\boldsymbol{\Sigma}_{\mathsf{w}} = \sigma^2\boldsymbol{I}, \tag{2.112}$$

$$\boldsymbol{\Sigma}_{\mathsf{x}} = \boldsymbol{H}\boldsymbol{\Sigma}_{\mathsf{y}}\boldsymbol{H}^T + \boldsymbol{\Sigma}_{\mathsf{w}} = \frac{P}{n}\boldsymbol{H}\boldsymbol{H}^T + \sigma^2\boldsymbol{I}. \tag{2.113}$$

The conditional covariance matrix $\mathbf{\Sigma}_{\mathsf{x}|\mathsf{y}}$, which is needed to calculate the necessary eigenvalues from Theorem 2.3, can be calculated by the Schur complement as

$$\mathbf{\Sigma}_{\mathsf{x}|\mathsf{y}} = \mathbf{\Sigma}_{\mathsf{x}} - \mathbf{\Sigma}_{\mathsf{x},\mathsf{y}}\mathbf{\Sigma}_{\mathsf{y}}^{-1}\mathbf{\Sigma}_{\mathsf{x},\mathsf{y}}^{T} \quad = \quad \mathbf{\Sigma}_{\mathsf{w}} = \sigma^2\boldsymbol{I}. \tag{2.114}$$

Using these covariance matrices to calculate the eigenvalues from Theorem 2.3 then yields

$$\begin{aligned}
\mathbf{\Sigma}_{\mathsf{x}|\mathsf{y}}\mathbf{\Sigma}_{\mathsf{x}}^{-1} &= \sigma^2\boldsymbol{I}\left(\frac{P}{n}\boldsymbol{H}\boldsymbol{H}^T + \sigma^2\boldsymbol{I}\right)^{-1} && (2.115)\\
&= \left(\frac{P}{n\sigma^2}\boldsymbol{H}\boldsymbol{H}^T + \boldsymbol{I}\right)^{-1} && (2.116)\\
&= \left(\frac{P}{n\sigma^2}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T + \boldsymbol{I}\right)^{-1} =: \boldsymbol{B}^{-1}, && (2.117)
\end{aligned}$$

where the channel matrix is diagonalized as $\boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$. The resulting eigenvalue problem is

$$\begin{aligned}
\boldsymbol{v}\boldsymbol{B}^{-1} &= \lambda\boldsymbol{v} && (2.118)\\
\Rightarrow \boldsymbol{v} &= \lambda\boldsymbol{v}\boldsymbol{B} && (2.119)\\
\Rightarrow \frac{1}{\lambda}\boldsymbol{v} &= \boldsymbol{v}\boldsymbol{B} && (2.120)\\
\lambda_B\boldsymbol{v} &= \boldsymbol{v}\boldsymbol{B} && (2.121)\\
\lambda &= \frac{1}{\lambda_B}. && (2.122)
\end{aligned}$$

The characteristic equation for λ_B ($\lambda_B \neq 1$) is

$$\begin{aligned}
\det(\boldsymbol{B} - \lambda_B\boldsymbol{I}) &= 0 && (2.123)\\
\det\left(\frac{P}{n\sigma^2}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T - (\lambda_B - 1)\boldsymbol{I}\right) &= 0 && (2.124)\\
\det\left(\frac{P}{n\sigma^2(1-\lambda_B)}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T + \boldsymbol{I}\right) &= 0 && (2.125)\\
\prod_{i=1}^{r}\left(\frac{P}{n\sigma^2(1-\lambda_B)}\lambda_{H_i} + 1\right) &= 0, && (2.126)
\end{aligned}$$

where r is the rank of the channel and λ_{H_i} are the eigenvalues of $\boldsymbol{H}\boldsymbol{H}^T$. Since (2.126) must be fulfilled for all λ_{B_i} one can simplify this to

$$\Rightarrow n\sigma^2(1-\lambda_{B_i}) + P\lambda_{H_i} = 0 \tag{2.127}$$

$$\lambda_{B_i} = 1 + \frac{P}{n\sigma^2}\lambda_{H_i} \tag{2.128}$$

$$\lambda_i = \frac{1}{\lambda_{B_i}} = \frac{1}{\frac{P}{n\sigma^2}\lambda_{H_i} + 1}. \tag{2.129}$$

In the case $\lambda_B = 1$, (2.124) reduces to $\det(\boldsymbol{H}\boldsymbol{H}^T) = 0$. As a consequence $\lambda_{H_i} = 0$ and therfore (2.128) and (2.129) are still fulfilled.

The Information-Rate Function

The mutual information between $\mathbf{t}$ and $\mathbf{x}$ is again a measure for the compression and can be interpreted as a compression rate R. This yields

$$I_\beta(\mathbf{t};\mathbf{x}) = \frac{1}{2}\sum_{i=1}^{n(\beta)} \log_2\left((\beta-1)\frac{1-\lambda_i}{\lambda_i})\right) \tag{2.130}$$

$$= \frac{1}{2}\log_2\left((\beta-1)^{n(\beta)}\prod_{i=1}^{n(\beta)}\frac{1-\lambda_i}{\lambda_i}\right) \triangleq R. \tag{2.131}$$

Then making β explicit yields following expression

$$\Rightarrow \beta = 2^{\frac{2R}{n(\beta)}}\prod_{i=1}^{n(\beta)}\left(\frac{\lambda_i}{1-\lambda_i}\right)^{\frac{1}{n(\beta)}} + 1. \tag{2.132}$$

Using the expression for the mutual information between $\mathbf{t}$ and $\mathbf{y}$ as in the scalar case and substituting β with (2.132) reads

$$\begin{aligned}
I_\beta(\mathbf{t};\mathbf{y}) &= I(\mathbf{t};\mathbf{x}) - \frac{1}{2}\sum_{i=1}^{n(\beta)} \log_2\left(\beta(1-\lambda_i)\right) && (2.133)\\
&= R - \frac{1}{2}\log_2\left(\beta^{n(\beta)}\prod_{i=1}^{n(\beta)}(1-\lambda_i)\right) && (2.134)\\
&= R - \frac{1}{2}\times \\
&\quad \log_2\left(\left(2^{\frac{2R}{n(\beta)}}\prod_{i=1}^{n(\beta)}\left(\frac{\lambda_i}{1-\lambda_i}\right)^{\frac{1}{n(\beta)}}+1\right)^{n(\beta)}\prod_{i=1}^{n(\beta)}(1-\lambda_i))\right) && (2.135)\\
&= R - \frac{1}{2}n(\beta)\log_2\left(2^{\frac{2R}{n(\beta)}}\prod_{i=1}^{n(\beta)}\lambda_i^{\frac{1}{n(\beta)}} + \prod_{i=1}^{n(\beta)}(1-\lambda_i)^{\frac{1}{n(\beta)}}\right) && (2.136)\\
&= \frac{1}{2}n(\beta)\log_2\left(\frac{2^{\frac{2R}{n(\beta)}}}{2^{\frac{2R}{n(\beta)}}\prod_{i=1}^{n(\beta)}\lambda_i^{\frac{1}{n(\beta)}} + \prod_{i=1}^{n(\beta)}(1-\lambda_i)^{\frac{1}{n(\beta)}}}\right). && (2.137)
\end{aligned}$$

Now substituting λ_i with (2.129) yields

$$\begin{aligned}
I_\beta(\mathbf{t};\mathbf{y}) &= \frac{1}{2}n(\beta)\log_2\left(\frac{2^{\frac{2R}{n(\beta)}}}{\frac{2^{\frac{2R}{n(\beta)}}+\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{\prod_{i=1}^{n(\beta)}\left(1+\frac{P}{n\sigma^2}\lambda_{H_i}\right)^{1/n(\beta)}}}\right) && (2.138)\\
&= \frac{1}{2}n(\beta)\log_2\left(\frac{2^{\frac{2R}{n(\beta)}}\prod_{i=1}^{n(\beta)}\left(1+\frac{P}{n\sigma^2}\lambda_{H_i}\right)^{1/n(\beta)}}{2^{\frac{2R}{n(\beta)}}+\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) && (2.139)\\
&= \frac{1}{2}\log_2\left(\prod_{i=1}^{n(\beta)}\frac{2^{\frac{2R}{n(\beta)}}(1+\frac{P}{n\sigma^2}\lambda_{H_i})}{2^{\frac{2R}{n(\beta)}}+\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) && (2.140)\\
&= \frac{1}{2}\sum_{i=1}^{n(\beta)}\log_2\left(\frac{2^{\frac{2R}{n(\beta)}}(1+\frac{P}{n\sigma^2}\lambda_{H_i})}{2^{\frac{2R}{n(\beta)}}+\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) \triangleq I(R), && (2.141)
\end{aligned}$$

where $\overline{\lambda}_{H_n} = \overline{\lambda}_{H_{n(R)}}$ is the geometric mean $\prod_{i=1}^{n(\beta)}\lambda_{H_i}^{1/n(\beta)}$. Again substituting λ_i with (2.129) and inserting the critical β values $\beta_c = 1/(1-\lambda_n)$ yields critical rates R_c when using (2.130):

$$R_c(n) = \frac{1}{2}\sum_{i=1}^{n}\log_2\left(\frac{\lambda_n}{1-\lambda_n}\frac{1-\lambda_i}{\lambda_i}\right) \quad (2.142)$$

$$= \frac{1}{2}\sum_{i=1}^{n}\log_2\left(\frac{\lambda_{H_i}}{\lambda_{H n}}\right) \quad (2.143)$$

$$= \frac{1}{2}n\log_2\left(\frac{\overline{\lambda}_{H_n}}{\lambda_{H_n}}\right). \quad (2.144)$$

Hence, the eigenvalues λ_i are sorted in ascending order and $\lambda_{H_i} \propto 1/\lambda_i$, the λ_{H_i} are in descending order.

At the critical rates new modes are added to $I(R)$ (2.141). This means $n(\beta)$ is incremented at each R_c and is therefore a function explicitly in R ($n(\beta) \Rightarrow n(R)$):

$$n(R) = \max\{n : R \geq R_c(n)\}. \quad (2.145)$$

Or in other words, at rate R there are $n(R)$ active modes. These are the modes with the $n(R)$ largest eigenvalues λ_{H_i}. The following corollary gives an explicit expression for the information-rate function.

Corollary 2.7. *The information-rate function with GIB-optimal channel output compression is given by*

$$\boxed{I(R) = \frac{1}{2}\sum_{i=1}^{n(R)}\log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-\frac{2R}{n(R)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right)}, \quad (2.146)$$

where the number of active modes $n(R)$ is given by

$$n(R) = \max\{n : R \geq R_c(n)\}, \quad (2.147)$$

and the critical rates $R_c(n)$ are given by

$$R_c(n) = \frac{1}{2}n\log_2\left(\frac{\overline{\lambda}_{H_n}}{\lambda_{H_n}}\right). \quad (2.148)$$

We identify (2.146) as the sum of $n(R)$ scalar information-rate functions, i.e., we have

$$I(R) = \sum_{i=1}^{n(R)} I_i(R) \tag{2.149}$$

$$= \frac{1}{2}\sum_{i=1}^{n(R)} \log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-2R_i(R)}\frac{P}{n\sigma^2}\lambda_{H_i}}\right). \tag{2.150}$$

Comparing (2.146) with (2.150), yields the rate allocation

$$R_i(R) = \max\left\{0, \frac{R}{n(R)} + \frac{1}{2}\log_2\frac{\lambda_{H_i}}{\overline{\lambda}_{H_n}}\right\}. \tag{2.151}$$

This is essentially the same way of rate allocation as in the RD case, where the rate R is evenly distributed among all $n(R)$ active modes up to a correction term. Depending on whether the eigenvalue of the mode is greater or smaller than $\overline{\lambda}_{H_n}$ the rate $R/n(R)$ is increased or decreased by $\log_2(\lambda_{H_i}/\overline{\lambda}_{H_n})$. All rates of the scalar modes have the same slope

$$\frac{\partial R_i(R)}{\partial R} = \frac{1}{n(R)}, \tag{2.152}$$

where $n(R)$ is constant. As before, $R_i(R)$ is not differentiable on the critical rates. This means that a differential increase $\mathrm{d}R$ is evenly distributed among all $n(R)$ active modes. Of course the scalar rates again have to sum up to the total rate

$$R = \sum_{i=1}^{n(R)} R_i(R). \tag{2.153}$$

Rewriting (2.146) in the form $I(R) = \frac{1}{2}\sum_{i=1}^{n(R)}\log_2(1+SNR_i)$ yields the "mode" SNRs

$$SNR_i = \frac{P}{n\sigma^2}\frac{2^{\frac{2R}{n(R)}}\lambda_{H_i} - \overline{\lambda}_{H_n}}{2^{\frac{2R}{n(R)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}. \tag{2.154}$$

As in the scalar case, the information-rate function can also be directly obtained, without using $I_\beta(\mathsf{t};\mathsf{x})$ and $I_\beta(\mathsf{t};\mathsf{y})$ from [8]. Again we just need the identity $\mathsf{t} = \boldsymbol{A}\mathsf{x}+\boldsymbol{\xi} = \boldsymbol{A}(\mathsf{y}+\mathsf{w})+\boldsymbol{\xi}$. As presented in Subsection 2.4.2 an equivalent diagonalized system can be given with the diagonal channel $\boldsymbol{H} = \mathrm{diag}\left\{\sqrt{\lambda_{H_i}}\right\}_{i=1}^{n}$. The problem reduces then to n scalar information-rate functions with $\mathsf{t}_i = A_i\mathsf{x}_i + \xi_i$, where we still have to find the optimal rate allocation. The resulting information-rate function is the sum of the

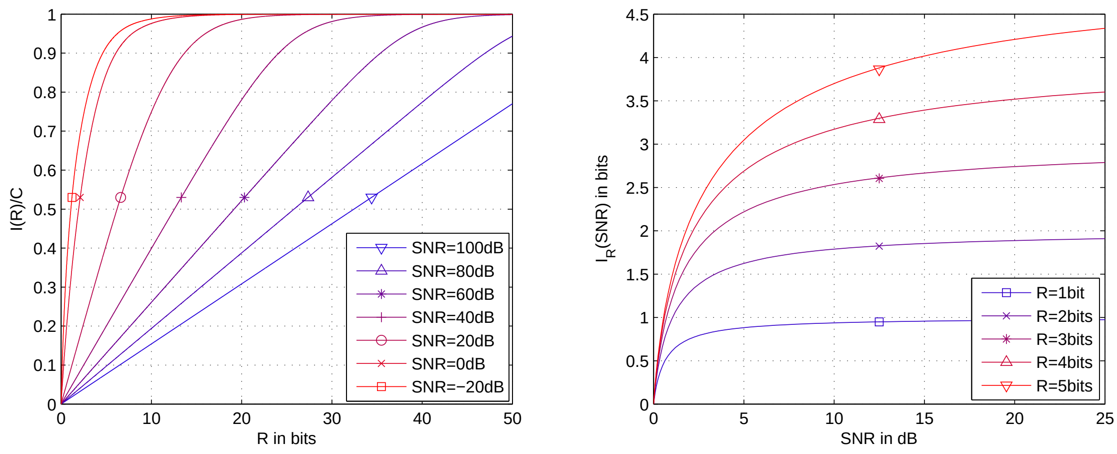

Figure 2.7: $I(R)/C$ (left) and $I_R(SNR)$ (right).

individual scalar information-rate functions (2.150)

$$I(R) = \frac{1}{2}\sum_{i=1}^{n} \log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-2R_i}\frac{P}{n\sigma^2}\lambda_{H_i}}\right), \tag{2.155}$$

with the rate constraint $R = \sum_{i=1}^{n} R_i$. The optimal rate allocation can then be obtained using Lagrange multipliers. That this solution is indeed equal to (2.146) will be shown in Subsection 2.6.

The Rate-Information Function

Next we calculate the inverse to the information-rate function, the rate-information function. Rewriting (2.133) as

$$R(I) = I - \frac{1}{2}\sum_{i=1}^{n(\beta)} \log_2\left(\beta(1-\lambda_i)\right), \tag{2.156}$$

then using (2.130) to substitute $R(I)$ and making I explicit yields

$$\begin{aligned} I(\beta) &= \frac{1}{2}\sum_{i=1}^{n(\beta)} \log_2\left(\frac{\beta-1}{\beta}\frac{1}{\lambda_i}\right) && (2.157)\\ &= \frac{1}{2}n(\beta)\log_2\left(\frac{\beta-1}{\beta}\right) - \frac{1}{2}\log_2\left(\lambda_i\right). && (2.158)\end{aligned}$$

The above equation can be rewritten as

$$\log_2\left(\frac{\beta-1}{\beta}\right) = \frac{2I}{n(\beta)} + \frac{1}{n(\beta)}\sum_{i=1}^{n(\beta)}\log_2(\lambda_i) \quad (2.159)$$

$$= \frac{2I}{n(\beta)} + \log_2\left(\bar{\lambda}_n\right). \quad (2.160)$$

Making β explicit yields

$$\beta = \frac{1}{1-2^{\frac{2I}{n(\beta)}}\bar{\lambda}_n}. \quad (2.161)$$

Inserting the expression for β in (2.161) into (2.130) yields

$$R(I) = \frac{1}{2}\sum_{i=1}^{n(\beta)}\log_2\left(\frac{2^{\frac{2I(R)}{n(\beta)}}\bar{\lambda}_n}{1-2^{\frac{2I}{n(\beta)}}\bar{\lambda}_n}\frac{1-\lambda_i}{\lambda_i}\right). \quad (2.162)$$

The last step is to substitute λ_i with (2.129) and rearrange the equation to

$$R(I) = \frac{1}{2}\sum_{i=1}^{n(\beta)}\log_2\left(\frac{\frac{P}{n\sigma^2}\lambda_{H_i}}{1/\left(2^{\frac{2I}{n(\beta)}}\bar{\lambda}_n\right)-1}\right). \quad (2.163)$$

However, $R(I)$ still depends on β through $n(\beta)$. Analog to the critical rates we can calculate critical mutual information values from (2.157) if we use the critical beta values $\beta_c = 1/(1-\lambda_n)$:

$$I_c(n) = \frac{1}{2}\sum_{i=1}^{n}\log_2\left(\frac{\lambda_n}{\lambda_i}\right) \quad (2.164)$$

$$= \frac{1}{2}n\log_2\left(\frac{\lambda_n}{\bar{\lambda}_n}\right). \quad (2.165)$$

At these critical information values I_c new modes are added to $R(I)$ (2.163). This means $n(\beta)$ is incremented at each I_c and is therefore a function of I ($n(\beta) \Rightarrow n(I)$):

$$n(I) = \max\{n : I \geq I_c(n)\}. \quad (2.166)$$

Or in other words, at mutual information I there are $n(I)$ active modes. These are the modes with the $n(I)$ smallest eigenvalues λ_i. The following corollary gives an explicit expression for the rate-information function.

Corollary 2.8. *The rate-information function with GIB optimal channel output com-*

pression is given by

$$\boxed{R(I) = \frac{1}{2}\sum_{i=1}^{n(I)} \log_2 \left(\frac{\frac{P}{n\sigma^2}\lambda_{H_i}}{1/\left(2^{\frac{2I}{n(I)}}\bar{\lambda}_n\right) - 1} \right)}, \tag{2.167}$$

where the number of active modes $n(R)$ *is given by*

$$n(I) = \max\{n : I \geq I_c(n)\}, \tag{2.168}$$

and the critical mutual information values $I_c(n)$ *are given by*

$$I_c(n) = \frac{1}{2} n \log_2 \left(\frac{\lambda_n}{\bar{\lambda}_n} \right). \tag{2.169}$$

We identify (2.167) as the sum of $n(I)$ scalar rate-information functions, i.e., we have

$$\begin{aligned} R(I) &= \sum_{i=1}^{n(I)} R_i(I) && (2.170) \\ &= \frac{1}{2}\sum_{i=1}^{n(I)} \log_2 \left(\frac{\frac{P}{n\sigma^2}\lambda_{H_i}}{2^{-2I_i(I)}\left(1 + \frac{P}{n\sigma^2}\lambda_{H_i}\right) - 1} \right). && (2.171) \end{aligned}$$

Comparing (2.167) with (2.171) yields

$$I_i(I) = \frac{I}{n(I)} + \frac{1}{2}\log_2 \left(\frac{\bar{\lambda}_n}{\lambda_i} \right) \tag{2.172}$$

and

$$I = \sum_{i=1}^{n(I)} I_i(I). \tag{2.173}$$

2.4.5 MSE-Optimal Quantization

Although we showed that in the scalar case RD and GIB yield the same information-rate function, this is not true in the vector case. While the information bottleneck directly maximizes the mutual information it is clearly the upper bound for all information-rate functions. RD minimizes the MSE $\mathrm{E}\{d(t,x)\} = \mathrm{E}\{d(\hat{x},x)\}$, i.e., it compresses the received signal x with smallest distortion. Actually a more reasonable approach would be to compress x in a manner that minimizes the MSE of the desired signal y, so $\mathrm{E}\{d(t,y)\} = \mathrm{E}\{d(\hat{y},y)\}$. Mathematically formulating this problem analogous to RD reads as

$$R(D) \triangleq \min_{f(\hat{y}|x)} I(\mathsf{x};\hat{\mathsf{y}}) \qquad \text{s.t.} \quad \mathrm{E}\{d(\mathsf{y},\hat{\mathsf{y}})\} \leq D, \tag{2.174}$$

In [2] it was shown that this problem can be optimally solved by a Wiener filter, which estimates y with smallest MSE, followed by reverse waterfilling on the estimate. [24] studied MSE optimal quantization, not only for the case with a Wiener filter, but for arbitrary linear filters, followed by reverse waterfilling. In particular [24] derives the information-rate function, where the derivation is analogous to the regular RD quantization. The resulting information-rate function is given in the following theorem [24].

Theorem 2.9. *The information-rate function with MSE optimal channel output compression is given by*

$$\boxed{I(R) = \frac{1}{2}\sum_{i=1}^{n(R)} \log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-\frac{2R}{n(R)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}\frac{1+\lambda_{H_i}}{\lambda_{H_i}}\frac{\overline{\lambda_{H_n}}}{\overline{\lambda_{H_n}+1}}}\right)}. \tag{2.175}$$

where the number of active modes $n(R)$ is given by

$$n(R) = \max\{n : R \geq R_c(n)\}, \tag{2.176}$$

and the critical rates $R_c(n)$ are given by

$$R_c(n) \;=\; \frac{1}{2}\sum_{i=1}^{n} \log_2\left(\frac{\lambda_{H_i}^2}{1+\lambda_{H_i}}\frac{1+\lambda_{H_n}}{\lambda_{H_n}^2}\right). \tag{2.177}$$

This can again be identified as the sum of $n(R)$ active scalar modes

$$I(R) \;=\; \sum_{i=1}^{n(R)} I_i(R) \tag{2.178}$$

$$\;=\; \frac{1}{2}\sum_{i=1}^{n(R)} \log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-2R_i(R)}\frac{P}{n\sigma^2}\lambda_{H_i}}\right), \tag{2.179}$$

because by comparison of coefficients in (2.175) and (2.179), we have

$$R_i(R) = \frac{R}{n(R)} + \frac{1}{2}\log_2\frac{\lambda_{H_i}^2}{1+\lambda_{H_i}}\frac{\overline{\lambda_{H_n}+1}}{\overline{\lambda^2}_{H_n}}. \tag{2.180}$$

2.5 Discussion of the Optimal Information-Rate Function

Now we discuss and prove the general properties of the optimal information-rate function, i.e., the information-rate function of the Gaussian information bottleneck. These properties are true for all Gaussian sources and noise and arbitrary channel realizations. The information-rate function $I(R)$ has the following properties:

1. $I(R)$ is concave ($\cap$) in R on $[0, \infty)$.
2. $I(R)$ is strictly increasing in R, i.e., $I(R) > I(R')$ iff $R > R'$.
3. $I(R) \leq R$.
4. $I(0) = 0$.
5. $\lim\limits_{R\to\infty} I(R) = I(\mathbf{x};\mathbf{y})$.
6. $I'(R) = \frac{1}{n(R)} \sum_{i=1}^{n(R)} \frac{\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{2^{\frac{2R}{n(R)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}} \leq 1$ and $I'(R)$ is continuous.

Consequently, the rate-information function $R(I)$ has the following properties:

1. $R(I)$ is convex ($\cup$) in I on $[0, I(X;Y))$.
2. $R(I)$ is strictly increasing in I, i.e., $R(I) > R(I')$ iff $I > I'$.
3. $R(I) \geq I$.
4. $R(0) = 0$.
5. $\lim\limits_{I\to I(X;Y)} R(I) = \infty$.
6. $R'(I) = \frac{1}{n(I)} \sum_{i=1}^{n(I)} \frac{1/\overline{\lambda}_n}{2^{\frac{2R}{n(I)}} - 1/\overline{\lambda}_n} \geq 1$ and $R'(I)$ is continuous.

Fig. 2.8 shows an illustration of the general form of the information-rate function. The area below the curve is the *feasible region*, where every point is reachable. For rates $R < C$ the feasible region is limited by the fact that $I(R) \leq R$ and therefore called *rate limited region*. For $R > C$ the feasible region is limited by the channel capacity, i.e., $I(R) \leq C$ and thus called *capacity limited region*.

Next we prove the general properties of the information-rate function. The properties of the rate-information function follow from the fact that it is the inverse of the information-rate function.

Proof ($I(R)$):

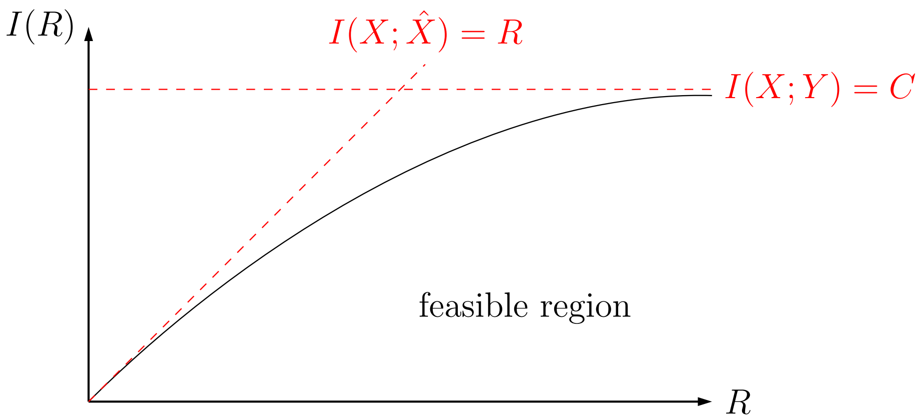

Figure 2.8: Illustration of the information-rate function.

1. From the definition of concavity we have

$$I(\alpha R_1 + (1-\alpha)R_2) \geq \alpha I(R_1) + (1-\alpha)I(R_2), \qquad 0 \leq \alpha \leq 1. \tag{2.181}$$

Without loss of generality we assume that $R_1 \leq R_2$ and define $R_\alpha = \alpha R_1 + (1-\alpha)R_2)$. Because a sum of concave functions is again a concave function we have to show that each summand in (2.146) is concave:

$$\begin{aligned}
\log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-\frac{2R_\alpha}{n(R_\alpha)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) &\geq \alpha\log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-\frac{2R_1}{n(R_1)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) + \\
&\quad (1-\alpha)\log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_i}}{1+2^{-\frac{2R_2}{n(R_2)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) &(2.182)\\
\log_2\left(\frac{1+2^{-\frac{2R_2}{n(R_2)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{1+2^{-\frac{2R_\alpha}{n(R_\alpha)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) &\geq \alpha\log_2\left(\frac{1+2^{-\frac{2R_2}{n(R_2)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{1+2^{-\frac{2R_1}{n(R_1)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right) &(2.183)\\
\frac{1+2^{-\frac{2R_2}{n(R_2)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{1+2^{-\frac{2R_\alpha}{n(R_\alpha)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}} &\geq \left(\frac{1+2^{-\frac{2R_2}{n(R_2)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{1+2^{-\frac{2R_1}{n(R_1)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}\right)^\alpha &(2.184)\\
\left(1+2^{-\frac{2R_2}{n(R_2)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}\right)^{1-\alpha} &\geq \frac{1+2^{-\frac{2R_\alpha}{n(R_\alpha)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{\left(1+2^{-\frac{2R_1}{n(R_1)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}\right)^\alpha} &(2.185)\\
1+2^{-\frac{2R_\alpha}{n(R_\alpha)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n} &\leq \left(1+2^{-\frac{2R_2}{n(R_2)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}\right)^{1-\alpha} \times \\
&\quad \left(1+2^{-\frac{2R_1}{n(R_1)}}\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}\right)^{\alpha}. &(2.186)
\end{aligned}$$

If we weaken the inequality to prove partly concavity on intervals where $n(R_\alpha) = n(R_1) = n(R_2)$, we define the equivalent inequality

$$(1+a)^\alpha(1+b)^{1-\alpha} \geq 1 + a^\alpha b^{1-\alpha}, \tag{2.187}$$

where $a = 2^{-\frac{2R_1}{n(R_1)}} \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}$ and $b = 2^{-\frac{2R_2}{n(R_2)}} \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}$, hence $a \geq b$. This can be seeen as a weighted geometric mean. We reformulate the equivalent inequality as

$$g(x) = (1+x)^\alpha(1+b)^{1-\alpha} - 1 - x^\alpha b^{1-\alpha} \geq 0. \tag{2.188}$$

Then $g(b) = 0$ and the first derivate is

$$g'(x) = \alpha\left((1+x)^{\alpha-1}(1+b)^{1-\alpha} - x^{\alpha-1}b^{1-\alpha}\right) = \alpha\left(\frac{1+b}{1+x}\right)^{1-\alpha} - \alpha\left(\frac{b}{x}\right)^{1-\alpha}. \tag{2.189}$$

It can be easily seen that $g'(x) < 0$ for $0 < x < b$ and $g'(x) > 0$ for $x > b$. It follows that $g(x) \geq 0$, and hence each summand is partly concave. This statement is not true in general, i.e., if $R_\alpha = R_c^+$ and $R_1 = R_c^-$, then $n(R_\alpha) = n(R_1) + 1$. Because we already showed partly concavity on intervals where $n(R_\alpha) = n(R_1)$, it is sufficient to show that $I'(R)$ is continuous. This is done in 6).

An equivalent way to prove the concavity of $I(R)$ is to show that $I(R)$ is continuously differentiable and $I'(R)$ is strictly decreasing, that is

$$I'(R_1) \geq I'(R_2), \qquad R_1 \leq R_2. \tag{2.190}$$

Since $I'(R)$ is continuous (will be shown in 6)), it is sufficient to show (2.190) for rates R_1, R_2 where $n(R_1) = n(R_2)$. The proof then simplifies and it remains to show that each summand in $I'(R)$ is strictly decreasing:

$$\begin{aligned}
\frac{\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{2^{\frac{2R_1}{n(R_1)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}} &\geq \frac{\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{2^{\frac{2R_2}{n(R_2)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}} && (2.191)\\
2^{\frac{2R_1}{n(R_1)}} &\leq 2^{\frac{2R_2}{n(R_2)}} && (2.192)\\
R_1 &\leq R_2. && (2.193)
\end{aligned}$$

Hence, $I(R)$ is concave.

2. Since all eigenvalues λ_{H_i} are nonnegative, the arguments of each logarithm in (2.146) are strictly increasing in R. The logarithm is a strictly increasing func-

tion and hence $I(R)$ is strictly increasing in R.

3. This property can be justified by the data processing inequality. Recall that we have the Markov chain $\mathbf{y}-\mathbf{x}-\mathbf{t}$ and $R = I(\mathbf{t};\mathbf{x})$ and $I = I(\mathbf{t};\mathbf{y})$. Loosely speaking the data processing inequality states that data processing can never increase the mutual information. Therefore,

$$I(\mathbf{t};\mathbf{x}) \geq I(\mathbf{t};\mathbf{y}) \quad \Leftrightarrow \quad R \geq I(R). \tag{2.194}$$

This can also be shown by directly evaluating (2.146) as

$$\begin{aligned} I(R) &= \frac{1}{2}\sum_{i=1}^{n(R)} \log_2 \left(\frac{2^{\frac{2R}{n(R)}}(1+\frac{P}{n\sigma^2}\lambda_{H_i})}{2^{\frac{2R}{n(R)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}} \right) & (2.195) \\ &= R - \frac{1}{2}\sum_{i=1}^{n(R)} \log_2 \left(\frac{2^{\frac{2R}{n(R)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{1+\frac{P}{n\sigma^2}\lambda_{H_i}} \right). & (2.196) \end{aligned}$$

To fulfill $I(R) \leq R$ it is sufficient that $I(0) = 0$ (property 4)) and $\max_R I'(R) \leq 1$ (from property 6)).

4.
$$I(0) = \frac{1}{2}\log_2 \left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_1}}{1+\frac{P}{n\sigma^2}\lambda_{H_1}} \right) = 0. \tag{2.197}$$

5.
$$\lim_{R\to\infty} I(R) = \frac{1}{2}\sum_{i=1}^{n} \log_2 \left(1+\frac{P}{n\sigma^2}\lambda_{H_i} \right) = I(\mathbf{x};\mathbf{y}). \tag{2.198}$$

6. Taking the derivate of (2.196) with respect to R yields

$$I'(R) = \frac{1}{n(R)}\sum_{i=1}^{n(R)} \frac{\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{2^{\frac{2R}{n(R)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}. \tag{2.199}$$

From (2.199) and the property of concavity it follows that $\max_R I'(R) = I'(0) = 1-(1+\frac{P}{n\sigma^2}\lambda_{H_1})^{-1} \leq 1$ and $\min_R I'(R) = \lim_{R\to\infty} I'(R) = 0$. $I'(R)$ is obviously continuous for constant $n(R)$ because all summands are continuous. Hence, we have to prove that $I'(R)$ is continuous at the critical rates $R_c(j)$. Note that

$$I'(R) = \frac{\partial I(R)}{\partial R} = \sum_{i=1}^{n(R)} \frac{\partial I_i(R)}{\partial R} = \sum_{i=1}^{n(R)} \frac{\partial I_i(R_i(R))}{\partial R} = \sum_{i=1}^{n(R)} \frac{\partial I_i(R_i)}{\partial R_i}\frac{\partial R_i}{\partial R}. \quad (2.200)$$

$\partial I_i(R_i)/\partial R_i$ can easily be calculated from the scalar information-rate function (2.53) as

$$\frac{\partial I_i(R_i)}{\partial R_i} = \frac{\frac{P}{n\sigma^2}\lambda_{H_i}}{2^{2R_i} + \frac{P}{n\sigma^2}\lambda_{H_i}}. \quad (2.201)$$

$\partial R_i/\partial R$ can be obtained from the equation for the mode rate allocation (2.151) as

$$\frac{\partial R_i}{\partial R} = \frac{1}{n(R)} \qquad R \neq R_c(j). \quad (2.202)$$

By evaluating (2.201) at the specific allocated rates $R_i(R)$ (2.151), we get

$$\left.\frac{\partial I_i(R_i)}{\partial R_i}\right|_{R_i=R_i(R)} = \frac{\frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}{2^{\frac{2R}{n(R)}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_n}}, \quad (2.203)$$

which we identify as the summands in (2.199). Hence, all modes have the same slope at the rates $R_i(R)$. To prove that $I'(R)$ is continuous, it is therefore sufficient to prove that $\left.\frac{\partial I_i(R_i)}{\partial R_i}\right|_{R_i=R_i(R)}$ is continuous. We show that

$$\lim_{R\to R_c(j)^-} \left.\frac{\partial I_i(R_i)}{\partial R_i}\right|_{R_i=R_i(R_c(j)^-)} = \lim_{R\to R_c(j)^+} \left.\frac{\partial I_i(R_i)}{\partial R_i}\right|_{R_i=R_i(R_c(j)^+)} \quad (2.204)$$

$$\frac{\frac{P}{n\sigma^2}\overline{\lambda}_{H_{j-1}}}{2^{\frac{2R_c(j)}{j-1}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_{j-1}}} = \frac{\frac{P}{n\sigma^2}\overline{\lambda}_{H_j}}{2^{\frac{2R_c(j)}{j}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_j}} \quad (2.205)$$

$$\frac{\overline{\lambda}_{H_{j-1}}}{\left(\frac{\overline{\lambda}_{H_j}}{\lambda_{H_j}}\right)^{j/(j-1)} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_{j-1}}} = \frac{\overline{\lambda}_{H_j}}{\frac{\overline{\lambda}_{H_j}}{\lambda_{H_j}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_j}} \quad (2.206)$$

$$\frac{\overline{\lambda}_{H_{j-1}}}{\frac{\overline{\lambda}_{H_{j-1}}}{\lambda_{H_j}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_{j-1}}} = \frac{\overline{\lambda}_{H_j}}{\frac{\overline{\lambda}_{H_j}}{\lambda_{H_j}} + \frac{P}{n\sigma^2}\overline{\lambda}_{H_j}} \quad (2.207)$$

$$\frac{1}{\frac{1}{\lambda_{H_j}} + \frac{P}{n\sigma^2}} = \frac{1}{\frac{1}{\lambda_{H_j}} + \frac{P}{n\sigma^2}}, \quad (2.208)$$

where we used the identity $\overline{\lambda}_{H_j}^{j} = \overline{\lambda}_{H_{j-1}}^{j-1}\lambda_{H_j}$. Thus, $I'(R)$ is also continuous.

Proof ($R(I)$):
1) follows directly from property 1) and 5) of $I(R)$.
2) follows directly from property 2) of $I(R)$.
3) follows directly from property 3) of $I(R)$.
4) follows directly from property 4) of $I(R)$.
5) follows from property 1) and 3).

2.6 Comparison of GIB and RD Solution

To distinguish between the solution for the Gaussian information bottleneck and the rate distortion, the information-rate functions are from now on called $I_{GIB}(R)$, respectivly $I_{RD}(R)$. Obviously there can only be a difference in the solutions, if a channel is present, so $\sigma^2 > 0$. Also at least one eigenvalue has to be different than the others, because the eigenvalues determine the rates where additional modes get active. If all eigenvalues are the same, the rate is just evenly distributed among all modes, at all rates.

Theorem 2.10. *Iff* $\max \lambda_{H_i} \neq \min \lambda_{H_i}$ *and* $\sigma^2 > 0$*, then* $\lambda_k > \frac{P}{n}\lambda_{H_k}$, $\overline{\lambda}_n > \frac{P}{n}\overline{\lambda}_{H_n}$ *and*

$$I_{GIB}(R) > I_{RD}(R), \tag{2.209}$$

otherwise

$$I_{GIB}(R) = I_{RD}(R). \tag{2.210}$$

If all eigenvalues are equal, then

$$I_{RD}(R) = I_{GIB}(R). \tag{2.211}$$

In the high SNR region ($\sigma^2 \ll P$) $\lambda_k \approx \frac{P}{n}\lambda_{H_k}$, hence

$$I_{RD}(R) \approx I_{GIB}(R). \tag{2.212}$$

Proof: In the previous section it was shown that the RD solution allocates the rate on $n(R)$ scalar modes according to the reverse waterfilling algorithm on the eigenvalues λ_k. It turns out that the GIB solution is also mathematically equivalent to the reverse waterfilling algorithm, but on the eigenvalues λ_{H_i}. So we will show that the GIB solution is indeed also the solution of the constrained maximization problem formulated with

Lagrange multipliers

$$\Lambda(R_1, \ldots, R_n, \mu) = \sum_{i=1}^{n} I_i(R_i) - \mu \left(\sum_{i=1}^{n} R_i - R \right), \tag{2.213}$$

where μ is the Lagrange multiplier. Setting the derivative to zero yields

$$\frac{\partial \Lambda(R_1, \ldots, R_n, \mu)}{\partial R_i} = 2 \frac{\frac{P}{n\sigma^2} \lambda_{H_i}}{2^{2R_i} + \frac{P}{n\sigma^2} \lambda_{H_i}} - \mu = 0. \tag{2.214}$$

Solving this equation for R_i yields

$$R_i = \max \left\{ 0, \frac{1}{2} \log_2 \left(\frac{\frac{P}{n\sigma^2} \lambda_{H_i}}{\theta} \right) \right\}, \tag{2.215}$$

which is the reverse waterfilling formulation with the "waterlevel" θ, where $\mu = 2\theta \frac{P}{n\sigma^2} / (1 + \theta \frac{P}{n\sigma^2})$. Using (2.215) in $\sum_{i=1}^{n} R_i = R$ to express θ in R yields $\theta = 2^{2R/n(R)} / (\frac{P}{n\sigma^2} \bar{\lambda}_{H_n})$. Substituting this back in (2.215) formulates as

$$R_i = \max \left\{ 0, \frac{R}{n(R)} + \frac{1}{2} \log_2 \frac{\lambda_{H_i}}{\bar{\lambda}_{H_n}} \right\}. \tag{2.216}$$

This equation is identical to (2.151). Thus, this rate allocation is optimal, i.e., it maximizes $I(R)$ at a given rate and every other rate allocation is suboptimal. ■

2.6.1 Number of Active Modes

Although we showed that in the scalar case $I(R) = I_{RD}(R)$, this is in general not true in the vector case. This is caused by the fact that the information-rate function is the sum of the information-rate functions of the individual modes and the total rate R has to be distributed among these modes. How the rate is distributed is different in GIB and RD, also the number of active modes $n(R)$ may be different. This fact is summarized in the following theorem.

Theorem 2.11. *The number of active modes using GIB-optimal channel output compression is always less or equal to the number of active modes using RD-optimal channel output compression*

$$n_{RD}(R) \geq n(R), \tag{2.217}$$

or equivalently the critical rates using GIB-optimal channel output compression are al-

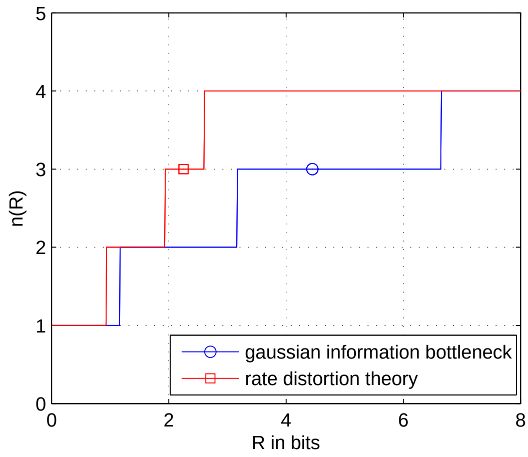

Figure 2.9: Example for number of active "modes".

ways greater or equal to the critical rates using RD-optimal channel output compression

$$R_{c_{RD}}(k) \leq R_c(k). \tag{2.218}$$

Proof: Comparing (2.82) and (2.144) yields

$$\frac{1}{2}k\log_2\left(\frac{\bar{\lambda}_k}{\lambda_k}\right) \leq \frac{1}{2}k\log_2\left(\frac{\bar{\lambda}_{H_k}}{\lambda_{H_k}}\right) \tag{2.219}$$

$$\frac{\bar{\lambda}_k}{\bar{\lambda}_{H_k}} \leq \frac{\lambda_k}{\lambda_{H_k}} \tag{2.220}$$

$$\prod_{k=1}^{k}\left(\frac{P}{n}+\frac{\sigma^2}{\lambda_{H_k}}\right)^{1/k} \leq \frac{P}{n}+\frac{\sigma^2}{\lambda_{H_k}}. \tag{2.221}$$

The last inequality in (2.221) is obviously true, since the eigenvalues are sorted in descending order and therefore λ_{H_k} is the smallest of the first k eigenvalues. ■

2.6.2 Quantitative Comparison of Some Selected Channels

Beamforming Scenario

In a typical beamforming scenario, where only one eigenvalue of $\boldsymbol{HH}^T$ is present (i.e., $\lambda_{H_1} = 1$, $\lambda_{H_i} = 0$), the largest differences between GIB and RD are expected. Defining the SNR as

$$\rho = \frac{P}{n\sigma^2} \tag{2.222}$$

yields

$$\lambda_{H_1} = 1, \qquad \lambda_1 = (1+\rho)\sigma^2, \tag{2.223}$$
$$\lambda_{H_i} = 0, \qquad \lambda_i = \sigma^2. \tag{2.224}$$

For low rates R only one mode is active in GIB and RD. Therefore, GIB and RD are equivalent for small rates. The critical rate where a second mode becomes active $R_c(2)$ is in RD always lower than in GIB. In the beamforming scenario only one mode is active at all rates in the GIB case. In the following we denote the critical rates in RD as $R_{c_{\mathrm{RD}}}(n)$.

$R \leq R_{c_{\mathrm{RD}}}(2)$: $I_{\mathrm{RD}}(R) = I_{\mathrm{GIB}}(R)$.
Calculating the critical rate $R_{c_{\mathrm{RD}}}(2)$ from (2.82) yields

$$R_{c_{\mathrm{RD}}}(2) = \frac{1}{2}\log_2 \frac{\frac{P}{n}+\sigma^2}{\sigma^2} = \frac{1}{2}\log_2\left(1+\frac{P}{n\sigma^2}\right) = \frac{1}{2}\log_2(1+\rho) = C. \tag{2.225}$$

So for rates lower than channel capacity both information-rate functions are equal,

$$R \leq C \qquad \Rightarrow \qquad I_{\mathrm{RD}}(R) = I_{\mathrm{GIB}}(R) \tag{2.226}$$

and thus the difference $\Delta I = I_{\mathrm{GIB}}(R) - I_{\mathrm{RD}}(R)$ is generally bounded by

$$\Delta I \leq C - I(C) = \frac{1}{2}\log_2\left(1+\rho\frac{1}{1+\rho}\right), \tag{2.227}$$

since $I(R)$ is a strictly increasing function, $I_{GIB}(R) \leq C$ and $I_{RD}(R) \geq I(C)$ if $R \geq C$. The bound becomes tight for $R \gg n \gg 1$.

$R > R_{c_{\mathrm{RD}}}(2) = C$: $I_{\mathrm{RD}}(R) < I_{\mathrm{GIB}}(R)$.
In the GIB case $n(R) = 1$ for all rates and therefore one can calculate from (2.146)

$$I_{\mathrm{GIB}}(R) = \frac{1}{2}\log_2\left(\frac{1+\rho}{1+\rho\, 2^{-2R}}\right). \tag{2.228}$$

Otherwise in the RD case $n(R) = n$, but only the mode with $\lambda_{H_1} = 1$ carries mutual information. So one can calculate from (2.104)

$$I_{\mathrm{RD}}(R) = \frac{1}{2}\log_2\left(\frac{1+\rho}{1+\rho\, 2^{-2R/n}(1+\rho)^{1/n-1}}\right), \tag{2.229}$$

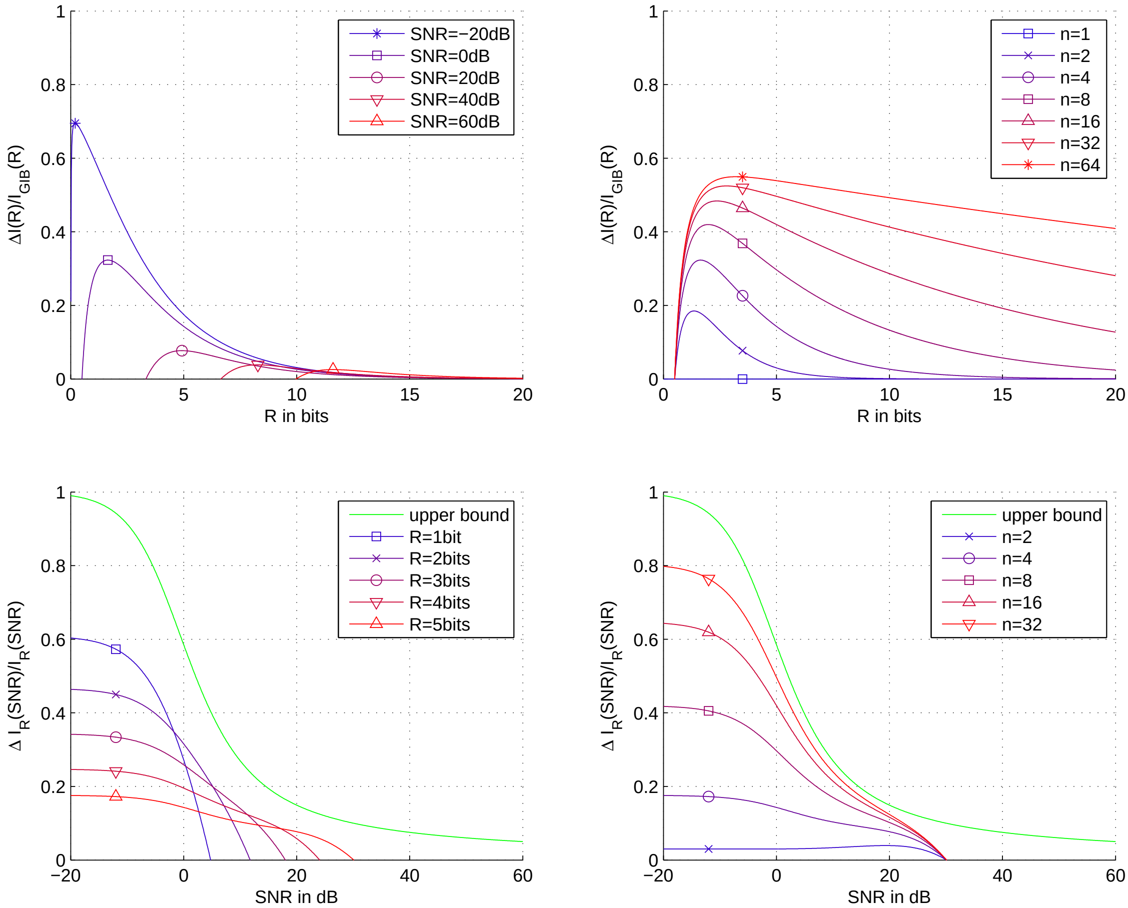

Figure 2.10: $\Delta I(R)/I(R)$ (top) and $\Delta I_R(SNR)/I_R(SNR)$ (bottom); $\lambda_{H_1} = 1$, $\lambda_{H_2} = 0$; $n = 4$ (left); $SNR = 0$dB (top right); $R = 5$bits (bottom right).

and

$$\Delta I(R) = \frac{1}{2} \log_2 \left(\frac{1 + \rho\, 2^{-2R/n}(1+\rho)^{1/n-1}}{1 + \rho\, 2^{-2R}} \right). \tag{2.230}$$

Fig. 2.10 shows ΔI for various SNRs. It can be observed that the difference increases in the high dimensional case, but decreases for increasing ρ. The points where $\Delta I = 0$ are determined by (2.225), where $R = R_{c_{\text{RD}}}(2) = C$. In Fig. 2.10 (bottom right) this corresponds to the ρ value where $R = \frac{1}{2}\log_2(1+\rho)$, since R is fixed.

General 2×2 Case

Because in the general $n \times n$ case many critical rates for both RD and GIB exist, it is not possible to give a handy expression for $\Delta I(R)$. In the 2×2 case we only have two critical rates, which we therefore denote by $R_{c_{GIB}}$ (from (2.144)) and $R_{c_{RD}}$ (from

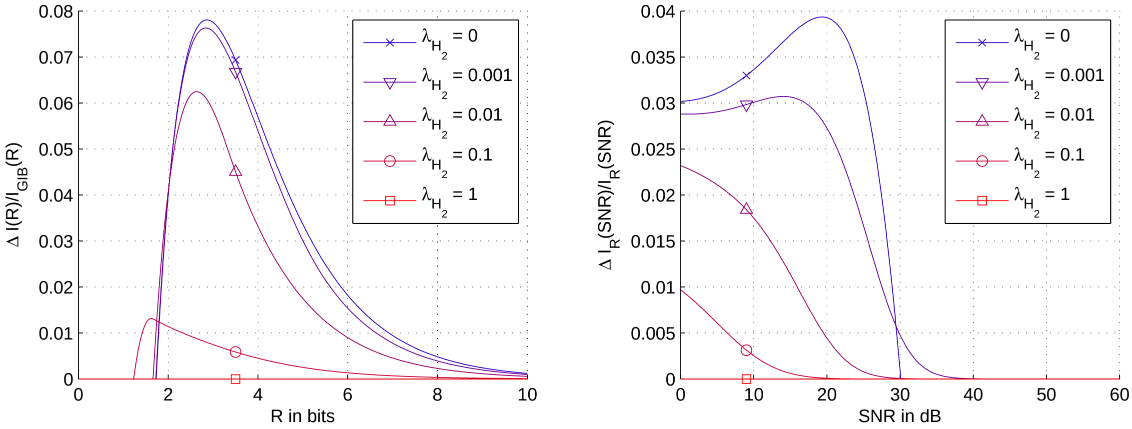

Figure 2.11: $\Delta I(R)/I(R)$ with $SNR = 0\text{dB}$ (left) and $\Delta I_R(SNR)/I_R(SNR)$ with $R = 5\,\text{bits}$ (right).

(2.82)):

$$R_{c_{GIB}} = \frac{1}{2}\log_2 \frac{\lambda_{H_1}}{\lambda_{H_2}}, \tag{2.231}$$

$$R_{c_{RD}} = \frac{1}{2}\log_2 \frac{\lambda_1}{\lambda_2} = \frac{1}{2}\log_2 \frac{\frac{P}{n\sigma^2}\lambda_{H_1}+1}{\frac{P}{n\sigma^2}\lambda_{H_2}+1}. \tag{2.232}$$

Then the information-rate functions are

$$I_{GIB}(R) = \frac{1}{2}\begin{cases} \frac{1+\frac{P}{n\sigma^2}\lambda_{H_1}}{1+2^{-2R}\frac{P}{n\sigma^2}\lambda_{H_1}}, & R < R_{c_{GIB}} \\ \frac{1+\frac{P}{n\sigma^2}\lambda_{H_1}}{1+2^{-R}\frac{P}{n\sigma^2}\sqrt{\lambda_{H_1}\lambda_{H_2}}} + \frac{1+\frac{P}{n\sigma^2}\lambda_{H_2}}{1+2^{-R}\frac{P}{n\sigma^2}\sqrt{\lambda_{H_1}\lambda_{H_2}}}, & R \geq R_{c_{GIB}} \end{cases}, \tag{2.233}$$

$$I_{RD}(R) = \frac{1}{2}\begin{cases} \frac{1+\frac{P}{n\sigma^2}\lambda_{H_1}}{1+2^{-2R}\frac{P}{n\sigma^2}\lambda_{H_1}}, & R < R_{c_{RD}} \\ \frac{1+\frac{P}{n\sigma^2}\lambda_{H_1}}{1+2^{-R}\frac{P}{n\sigma^2}\lambda_{H_1}\sqrt{\lambda_1\lambda_2}/\lambda_1} + \frac{1+\frac{P}{n\sigma^2}\lambda_{H_2}}{1+2^{-R}\frac{P}{n\sigma^2}\lambda_{H_2}\sqrt{\lambda_1\lambda_2}/\lambda_2}. & R \geq R_{c_{RD}} \end{cases}. \tag{2.234}$$

We have $R_{c_{RD}} \leq R_{c_{GIB}}$ and therefore we can write the difference of the information-rate functions $\Delta I(R)$ (cf. Fig. 2.11) as

$$\Delta I(R) = \frac{1}{2}\begin{cases} 0, & R < R_{c_{RD}} \\ \log_2 \frac{1+2^{-R}\frac{P}{n\sigma^2}\lambda_{H_1}\sqrt{\lambda_1\lambda_2}/\lambda_1}{1+2^{-2R}\frac{P}{n\sigma^2}\lambda_{H_1}} + \log_2 \frac{1+2^{-R}\frac{P}{n\sigma^2}\lambda_{H_2}\sqrt{\lambda_1\lambda_2}/\lambda_2}{1+\frac{P}{n\sigma^2}\lambda_{H_2}}, & R_{c_{RD}} \leq R < R_{c_{GIB}} \\ \log_2 \frac{1+2^{-R}\frac{P}{n\sigma^2}\lambda_{H_1}\sqrt{\lambda_1\lambda_2}/\lambda_1}{1+2^{-R}\frac{P}{n\sigma^2}\sqrt{\lambda_{H_1}\lambda_{H_2}}} + \log_2 \frac{1+2^{-R}\frac{P}{n\sigma^2}\lambda_{H_2}\sqrt{\lambda_1\lambda_2}/\lambda_2}{1+2^{-R}\frac{P}{n\sigma^2}\sqrt{\lambda_{H_1}\lambda_{H_2}}}, & R \geq R_{c_{GIB}} \end{cases}. \tag{2.235}$$

3

Performance with Feedback

3.1 Introduction

Until now we studied scenarios without feedback, and it is well known that feedback does not increase the channel capacity [10] of AWGN channels,

$$C_{FB} = C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right). \tag{3.1}$$

So why study feedback? Schalkwijk and Kailath showed that a scheme with feedback ([28], [27]), although it does not increase the capacity, can be very beneficial. They demonstrated that a relatively simple iterative scheme can drastically reduce the error probability and can even achieve channel capacity.

In our context we understand feedback as the additional knowledge of the transmitter about the received signal at the receiver side. In many practical systems feedback is used to give the transmitter knowledge about the channel (channel state information). We instead, in the most generic system, directly feed back the received signal and assume the transmitter has perfect knowledge about it, i.e., the feedback channel is perfect. Such systems would have very high requirements regarding the feedback channel and therefore are not feasible. Here it is important to reduce the rate of the feedback to a reasonable amount and quantizers become an issue. We study the performance of systems with GIB quantizers and therefore the basic requirement is that all signals are Gaussian.

This requirement restricts the system design to a communication system where the initial transmit symbol is Gaussian and also the later transmitted iterations are

Gaussian. The later transmitted iterations are a function of the transmit message and the previously received feedback signals. The Gaussianity of all signals can be guaranteed by restricting to systems with linear feedback.

3.2 Source Signal Model

The optimization in the previous chapters was done with a Gaussian source signal, yielding the Gaussian information bottleneck. If we want to measure the performance of systems, we usually want to give error probabilites, i.e., symbol error probability or bit error probability. Therefore, we have to discretize the source. In order to combine the contradictory requirements of the Gaussian source and the discrete source we approximate the Gaussian source by a discrete source. To derive the discrete signal we use definitions and derivations of the Riemann integral.
Given an interval $[a, b]$ and $f : [a, b] \to \mathbb{R}$, a bounded function. We partition the interval $[a, b]$ in intervals of size Δ, where $n\Delta = b - a$ and $x_k = a + k\Delta$. Then we define the upper sum and lower sum as

$$O(\Delta) \triangleq \sum_{k=1}^{n} \Delta \sup_{x_{k-1}<x<x_k} f(x), \tag{3.2}$$

$$U(\Delta) \triangleq \sum_{k=1}^{n} \Delta \inf_{x_{k-1}<x<x_k} f(x). \tag{3.3}$$

By the definition of integrability we have

$$\int_a^b f(x)\mathrm{d}x \triangleq \lim_{\Delta\to 0} U(\Delta) = \lim_{\Delta\to 0} O(\Delta), \tag{3.4}$$

if the last equality holds. Clearly if we do not take $x_{k-1} < x < x_k$, which achieves the supremum in $O(\Delta)$ and the infimum in $U(\Delta)$, but set $x = x_k$ and define

$$S(\Delta) \triangleq \sum_{k=1}^{n} \Delta f(x_k), \tag{3.5}$$

we get a sum which is bounded by $O(\Delta)$ and $U(\Delta)$

$$O(\Delta) \geq S(\Delta) \geq U(\Delta). \tag{3.6}$$

If (3.4) is fulfilled, we get

$$\int_a^b f(x)\mathrm{d}x = \lim_{\Delta \to 0} S(\Delta). \tag{3.7}$$

To come back to the desired discrete approximation $p(x)$ of our Gaussian distribution $f(x)$, the approximation should ideally fulfill

$$\int_a^b f(x)\mathrm{d}x = \sum_{k=1}^{n} p(x_k) \qquad a, b \in \mathbb{R}. \tag{3.8}$$

Comparing this equation with (3.7) yields

$$p(x_k) = \lim_{\Delta \to 0} \Delta f(x_k). \tag{3.9}$$

In this case, $\Delta \to 0$, we say both are equivalent. If $\Delta > \epsilon > 0$ (3.8) is no longer exactly fulfilled, but again we define the approximation as (Fig. 3.1)

$$p(x_k) \approx \Delta f(x_k). \tag{3.10}$$

The relation of differential entropy and the entropy of the discretized version is derived in [10]. The differential entropy of the original Gaussian source is denoted as $h(\mathsf{x})$. The source should have the rate R, which is the discrete entropy $H(\mathsf{x})$

$$R = H(\mathsf{x}) = -\sum_{x} p(x_k) \log_2 p(x_k). \tag{3.11}$$

Now substituting $p(x_k)$ with (3.10) yields

$$\begin{aligned} R &= -\sum_{-\infty}^{\infty} \Delta f(x_k) \log_2(\Delta f(x_k)) && (3.12)\\ &= -\sum_{-\infty}^{\infty} \Delta f(x_k) \log_2 f(x_k) - \sum_{-\infty}^{\infty} \Delta f(x_k) \log_2 \Delta && (3.13)\\ &\approx \underbrace{-\int_{-\infty}^{\infty} f(x_k) \log_2 f(x_k)\mathrm{d}x}_{=h(\mathsf{x})} - \underbrace{\int_{-\infty}^{\infty} f(x_k)\mathrm{d}x}_{=1} \log_2 \Delta && (3.14)\\ &= h(\mathsf{x}) - \log_2 \Delta. && (3.15) \end{aligned}$$

This approximation gets tight for high rates and is exactly fulfilled as $R \to \infty$. The

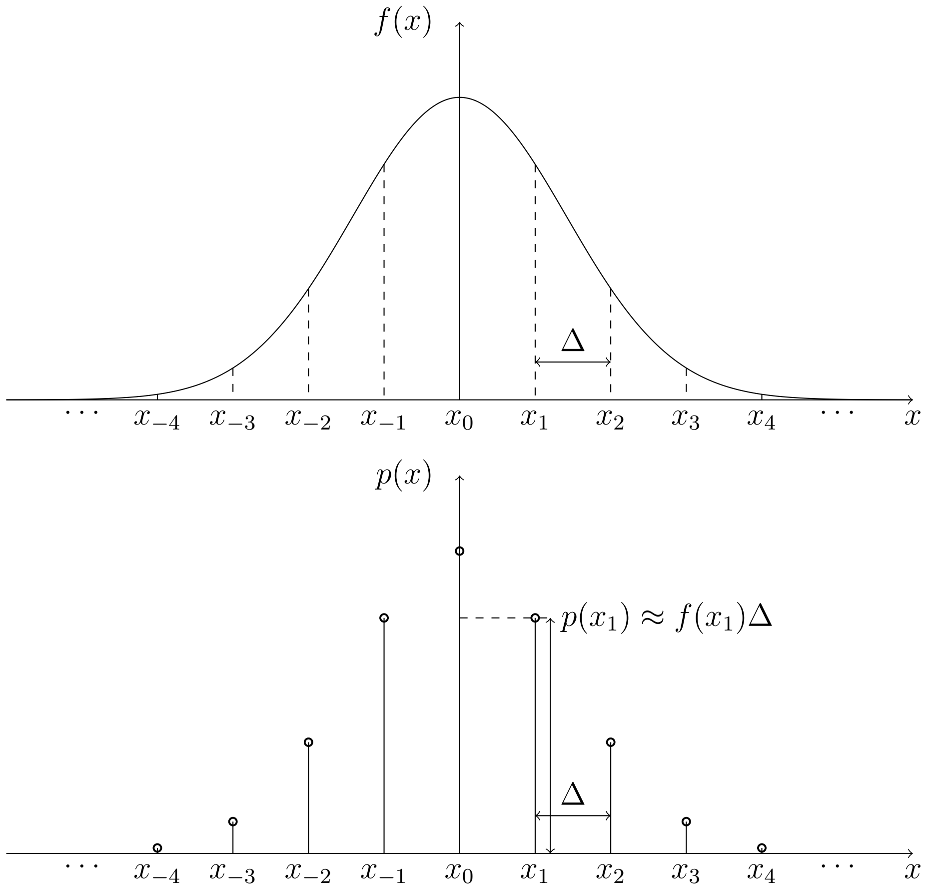

Figure 3.1: Gaussian distribution (top) and resulting discrete approximation (bottom).

differential entropy of a Gaussian source is

$$h(\mathsf{x}) = \frac{1}{2}\log_2(2\pi e\sigma^2). \tag{3.16}$$

Making Δ explicit then yields

$$\Rightarrow \Delta^2 = \frac{\sigma^2}{2^{2R}} 2\pi e. \tag{3.17}$$

An uniform distribution would result in

$$\Delta_U^2 = \frac{\sigma^2}{2^{2R}}\frac{1}{12}. \tag{3.18}$$

Then the ratio of (3.17) and (3.18) is called the *ultimate shaping gain* [13] and formulates

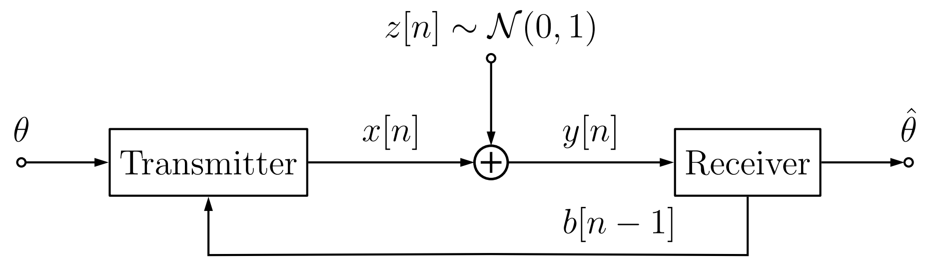

Figure 3.2: Generic system model.

as

$$c_0^2 = \frac{\Delta^2}{\Delta_U^2} = \frac{\pi e}{6} \triangleq 1.53\text{dB}. \tag{3.19}$$

So the discrete approximation of the Gaussian source with the differential entropy $h(\mathsf{x})$ goes to an entropy $H(\mathsf{x}) \to \infty$ if $\Delta \to 0$. From a mathematical point of view, every continous source has to have infinite entropy, since the realization is chosen from a set of real numbers, which are innumerable, even if the set is bounded.

3.3 Generic System Model

The most generic system model with feedback is shown in Fig. 3.2. We want to transmit the message θ over the normalized channel with additive white Gaussian noise. Without loss of generality the variance of the noise is $\sigma^2 = 1$, hence the power of the transmit signal $x[n]$ is both a measure for the transmit power and received SNR. θ is chosen from any sort of alphabet and is transmitted as a part of the signal $x[n]$. $x[n]$ is generally a function of θ and the feedback signal $b[n]$. Therefore,

$$x[n] = f(\theta, b[1], b[2], \ldots, b[n-1]), \tag{3.20}$$

in order to be a strictly causal function. This signal is received as

$$y[n] = x[n] + z[n]. \tag{3.21}$$

The feedback signal $b[n]$ is generally a causal function of the previously received $y[n]$, i.e., we have

$$b[n] = f(y[1], y[2], \ldots, y[n]). \tag{3.22}$$

To simplify the problem we often set $b[n] = y[n]$. This means the transmitter has the side-information of all previously received $y[n]$, and therefore also the previous noise realizations $z[n]$, since it has of course knowledge of the transmit signal $x[n]$ and the noise is just $z[n] = y[n] - x[n]$. In what follows, N is considered to be arbitrary, but fixed. After all N transmissions, the receiver estimates the transmit message using all previously received values $y[1], y[2], \ldots, y[N]$. N is therfore called *blocklength*, or *number of iterations*. We formulate the estimate as

$$\hat{\theta} = f(y[1], y[2], \ldots, y[N]). \tag{3.23}$$

These functions do not necessarily have to be deterministic. For example in the case of noisy feedback, $b[n]$ is a probabilistic function of the previously received values.

If this iterative system only involves linear operations and only Gaussian noise is introduced, we can describe $f(\hat{\theta}|\theta)$ as a Gaussian channel with capacity

$$C_\theta = \frac{1}{2}\log_2(1 + SNR). \tag{3.24}$$

Since a new message is transmitted every N^{th} iteration, we define the "superchannel", normalized by the number of channel usages, as

$$C_S = \frac{1}{N}\frac{1}{2}\log_2(1 + SNR). \tag{3.25}$$

3.4 Perfect Feedback

3.4.1 Scalar Channel

A communication system with quantized feedback can be seen as a system with perfect feedback, if the feedback signal is quantized with the same quantizer as the received signal[1] (Fig. 3.3), e.g., $t_{FB} = t$.

[1]With an abuse of notation the transmit signal is termed y, the noise w, the received signal before information bottleneck quantization is termed x and after the quantization t, if the actual architecture of feedback and quantization should be pointed out, since this nomenclature is common in information bottleneck literature. If the focus is on the whole communication system including the transmit message generation the usual notation is used, where $x[n]$ is the transmit signal, $z[n]$ is the noise and $y[n]$ is

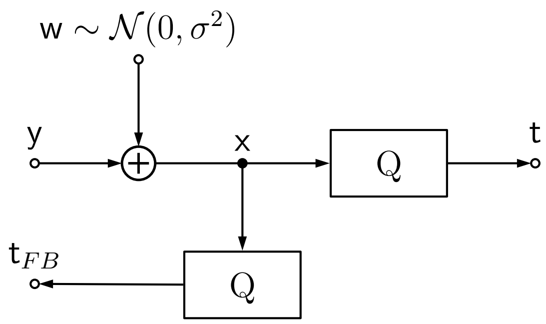

Figure 3.3: Communication system with perfect feedback (same quantizers).

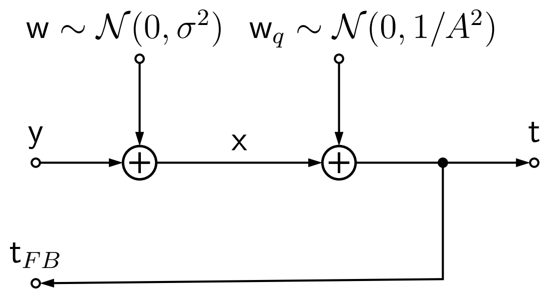

Figure 3.4: Equivalent system.

If we include the quantizer in this analysis, the quantizer is considered to be a part of the channel. Clearly every quantization process introduces additional noise. In some special cases the quantization noise can be approximated as an independent Gaussian signal. However in our analysis, where the quantizer is a Gaussian information bottleneck, this is exactly true. Using the GIB solution for the quantizer leads to an equivalent system (Fig. 3.4), where we can define a new channel including the quantizer. The resulting channel from y to t is then again an AWGN channel with the noise $\mathsf{z} = \mathsf{w} + \mathsf{w_q}$, where $\mathsf{z} \sim \mathcal{N}(0, \sigma^2 + 1/A^2)$.

From [8] we get

$$A(\beta) = \sqrt{\frac{\beta(1-\lambda)-1}{r\lambda}}v, \tag{3.26}$$

with $\beta = 2^{2R}\frac{\lambda}{1-\lambda} + 1$ (2.38), $\lambda = \frac{\sigma^2}{P+\sigma^2}$ (2.34), the resulting eigenvector in the scalar case $v = 1$ and $r = v^T\Sigma_x v = P + \sigma^2$. Writing now A as a function of the rate of the

the received signal.

quantizer R_Q yields[2]

$$A(R_Q) = \sqrt{\frac{2^{2R_Q} - 1}{P + \sigma^2}}. \tag{3.27}$$

The capacity $C(R_Q)$ of the resulting channel[3], including the quantizer, is then equivalent to the information-rate function, since the information-rate function is the mutual information $I(\mathsf{t};\mathsf{y})$:

$$\begin{aligned} C(R_Q) &= \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2 + 1/A(R_Q)^2}\right) & (3.28)\\ &= \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2 + \frac{P+\sigma^2}{2^{2R_Q}-1}}\right) & (3.29)\\ &= \frac{1}{2}\log_2\left(\frac{P/\sigma^2 + 1}{2^{-2R_Q}P/\sigma^2 + 1}\right) = I(R_Q). & (3.30) \end{aligned}$$

Using a Schalkwijk-Kailath (SK) coding scheme ([28], [27]) it was schown in [28] that the probability of an error P_e, with a source rate R, decreases as a second-order exponential in N and achieves channel capacity.

Definition 3.1. *A rate R is said to be achievable, if*

$$\lim_{N\to\infty} P_e^{(N)}(R) = 0. \tag{3.31}$$

Figure 3.5 shows the feedback system where θ is the message to transmit. The SK scheme is iteratively formulated as

$$\begin{aligned} x[0] &= \theta, & (3.32)\\ x[1] &= \alpha_1(y[0] - x[0]) = \alpha_1 z[0], & (3.33)\\ x[i] &= \alpha_i(z[0] - \hat{z}_i[0]), \qquad i = 2, 3, \ldots, N, & (3.34) \end{aligned}$$

where $\hat{z}_i[0]$ is the minimum mean square error (MMSE) estimate of $z[0]$ given all previously received values $y[0], \ldots, y[i-1]$. Basically at the initial transmission the actual meassage is transmitted and the iterations are used to cancel the noise introduced at the initial transmission. The scaling factors α_i are chosen to meet the power constraint

[2]From now on the rate of the quantizer is termed as R_Q to avoid confusion with the previously introduced rate of the source R.

[3]Because of the fact that the resulting channel is a Gaussian channel including the quantizer (with rate R_Q) we term the capacity of the resulting channel as $C(R_Q)$. Especially the term $C(R_Q)$ is used if the equation where it is used is derived from an equation referring to the channel capacity C. To avoid confusion the channel capacity without quantization is termed as C_0 ($C_0 = C(\infty)$).

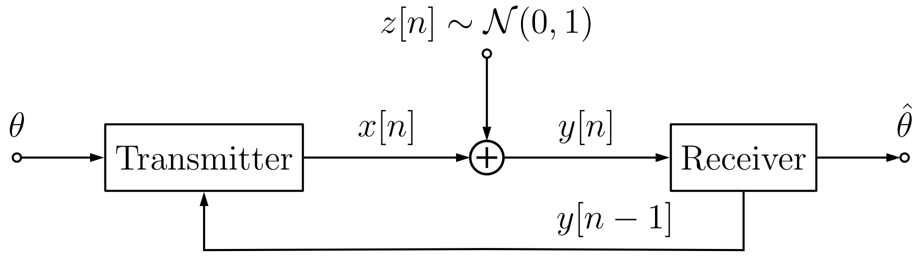

Figure 3.5: Equivalent communication system with perfect feedback and quantizer in forward path.

$\mathrm{E}\{\mathsf{x}[i]^2\}$ for all transmit values. This is rather an estimation approach, than an information theoretical approach. However the MMSE estimate ensures that $x[n]$ is Gaussian for all n. Thus, the information-rate function is still the correct measure for the capacity of the channel including the quantizer. A bound for the error probability was given in [14] as

$$P_e \leq 2\mathcal{Q}(2^{N(C-R)}), \tag{3.35}$$

where

$$\mathcal{Q}(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \tag{3.36}$$

is the Q-function. As showed, the capacity of the resulting channel is equivalent to the information-rate function and therefore give the following Corollaries.

Corollary 3.1. *The error probability of the linear feedback scheme with perfect feedback decreases doubly exponential as*

$$P_e \leq 2\mathcal{Q}(2^{N(C(R_Q)-R)}), \tag{3.37}$$

if $R < I(R_Q) = C(R_Q)$.

Corollary 3.2. *If the rate is limited by* $R < I(R_Q) = C(R_Q)$*, the error probability tends to 0, i.e.,*

$$\lim_{N\to\infty} P_e^{(N)}(R) = 0. \tag{3.38}$$

Thus, the achievable rate is $I(R_Q) = C(R_Q)$.

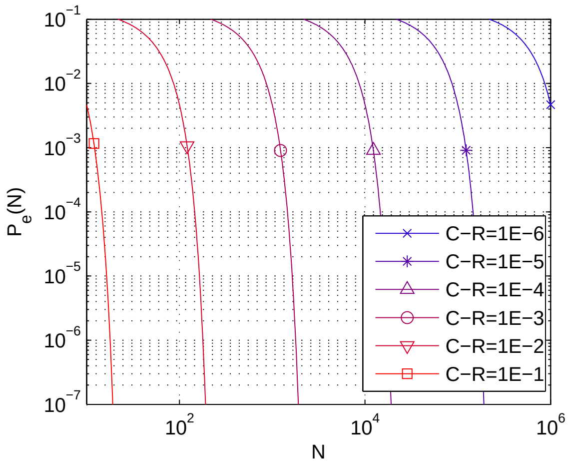

Figure 3.6: Error probability $P_e(N)$.

The expression in (3.37) can be bounded by an approximation of the Gaussian cumulative density function, as in [18], by

$$P_e \leq 2 \exp\left(-\frac{c_0^2 2^{2N(C(R_Q)-R)}}{2}\right), \tag{3.39}$$

where c_0 is a fixed constant which depends only on the message constellation. The proof in [18] is based on θ being a real number in the interval $[-1, 1]$ with 2^{2NR} steps. We instead set θ to be discrete Gaussian distributed with rate NR. So c_0 in (3.39) is equal to (3.19), which is the shaping gain of the Gaussian distribution over the uniform distribution (ultimate shaping gain). Fig. 3.6 and Fig. 3.7 show the tighter bound (3.37) for the error probabilities.

More sophisticated systems were studied, which show a error probability decreasing faster than exponentially in blocklength. However these systems are nonlinear, e.g., [1] uses partial sequential feedback. These systems were extended to error probabilities even decreasing in any exponential order in [21], [11] and [26]. [25] studied a scheme where the error probability is of the form $P_e = \exp(-O(n))$. So the error probability decreases exponentially with order of the blocklength. The drawback of all these systems is the nonlinearity. Unfortunately linearity is necessary in our treatment, since Gaussian signals are required. Thus, we will introduce an optimal communication system with linear feedback in the next section.

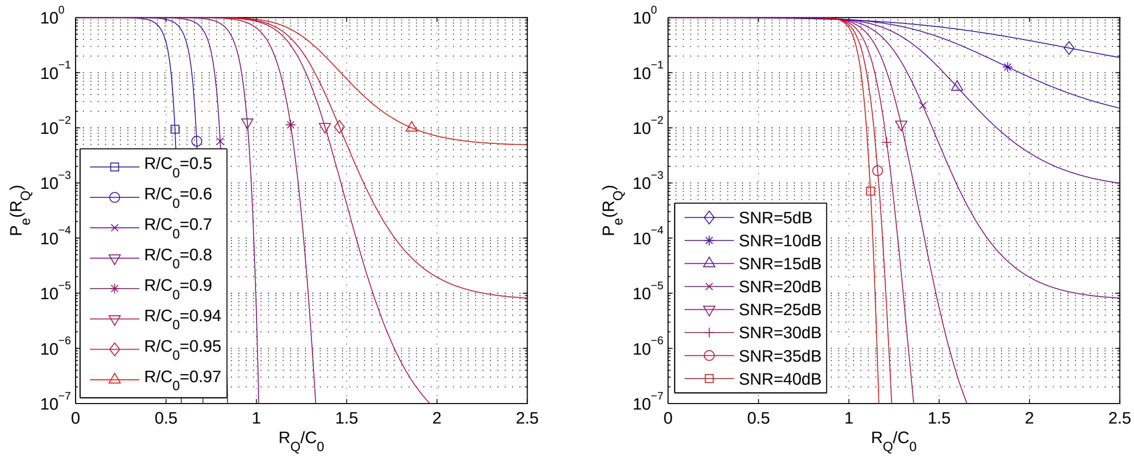

Figure 3.7: $P_e(R_Q)$ for $SNR = 20\text{dB}$, $N = 10$ (left); $P_e(R_Q)$ for $R = 0.95C_0$, $N = 10$ (right).

3.4.2 Vector Channel

We study the equivalent diagonalized system, where $n(R_Q)$ modes are active. We have a total information-rate, which is the sum of the capacities of the active modes,

$$I(R_Q) = \sum_{k=1}^{n(R_Q)} C_k(R_Q). \tag{3.40}$$

The sum of all scalar capacities does not provide the capacity of the vector channel, since the power allocation on the modes is fixed. In order to obtain the capacity of the vector channel we would have to jointly optimize power allocation and quantization of the scalar modes, which is a much harder problem. As shown previously the capacity has to be equal to the information-rate functions, (2.146) and (2.104),

$$I(R_Q) = C(R_Q). \tag{3.41}$$

Therfore if a GIB optimized quantizer is used, the resulting capacity is given by

$$C_k(R_Q) = \frac{1}{2}\log_2\left(\frac{\frac{P}{n\sigma^2}\lambda_{H_k} + 1}{\frac{P}{n\sigma^2}\overline{\lambda}_{H_n} 2^{-\frac{2R_Q}{n(R_Q)}} + 1}\right), \tag{3.42}$$

or in the case of an RD optimized quantizer the capacity is given by

$$C_k(R_Q) = \frac{1}{2}\log_2\left(\frac{1+\frac{P}{n\sigma^2}\lambda_{H_k}}{1+2^{-\frac{2R_Q}{n(R_Q)}}\frac{P}{n\sigma^2}\lambda_{H_k}\frac{\overline{\lambda}_n}{\lambda_k}}\right). \tag{3.43}$$

The difference of the capacities of the channel including the quantizer is due to the fact that the critical rates are different in GIB and RD. The critical rates determine the number of active modes $n(R_Q)$ and the actual allocation of the total quantization rate for each mode. In general the critical rates are different and so are the mode capacities.

Optimal source rate allocation

Due to different mode capacities we allocate the source rate in order to minimize the average probability of a symbol error. The probability is given as the average of a symbol error over all active $n(R_Q)$ modes,

$$\overline{P_e} \leq \frac{1}{n(R_Q)}\sum_{k=1}^{n(R_Q)} P_e(R_k|R_k>0) \tag{3.44}$$

$$= \frac{1}{n(R_Q)}\sum_{k=1}^{n(R_Q)} 2\exp\left(-\frac{c_0^2 2^{2N(C_k(R_Q)-R_k)}}{2}\right) I(R_k>0), \tag{3.45}$$

under the constraint

$$\sum_{k=1}^{n(R_Q)} R_k = R. \tag{3.46}$$

Here, $I(R_k > 0)$ is the indicator function

$$I(R_k>0) = \begin{cases} 1 & \text{if} \quad R_k > 0 \\ 0 & \text{else} \end{cases}, \tag{3.47}$$

since not all active modes necessarily carry information. In other words R_k may be 0 for an index $k \leq n(R_Q)$. Clearly $R_k = 0$ for all indices $k > n(R_Q)$, because these modes are inactive and therefore do not provide any capacity. If the individual error probabilities $P_e(R_k)$ are low, a valid approximation is to assume that the only occurring errors are errors to the neighbor symbol. As a consequence, if a Gray code is assumed, a symbol error is equivalent to the error of one bit. A bound for the average probability of an bit error can then be bounded by

$$\overline{P_b} \leq \frac{1}{R} \sum_{k=1}^{n(R_Q)} P_e(R_k | R_k > 0) \tag{3.48}$$

$$= \frac{1}{R} \sum_{k=1}^{n(R_Q)} 2 \exp\left(-\frac{c_0^2 2^{2N(C_k(R_Q)-R_k)}}{2}\right) I(R_k > 0), \tag{3.49}$$

again under the constraint

$$\sum_{k=1}^{n(R_Q)} R_k = R. \tag{3.50}$$

Minimizing the average probability of an error is a constrained optimization problem, which can be solved by Lagrange multipliers. The problem formulates as

$$\Lambda(R_1, \ldots, R_n, \lambda) = \overline{P_e}(R_1, \ldots, R_n) + \lambda \left(\sum_{k=1}^{n(R_Q)} R_k - R\right) \tag{3.51}$$

$$= \sum_{k=1}^{n(R_Q)} \left(2\frac{1}{R} \exp\left(-\frac{c_0^2 2^{2N(C_k(R_Q)-R_k)}}{2}\right) + \lambda R_k\right) - \lambda R, \tag{3.52}$$

$$\frac{\partial \Lambda(R_1, \ldots, R_n, \lambda)}{\partial R_k} = \underbrace{\frac{\partial}{\partial R_k}\left(2\frac{1}{R} \exp\left(-\frac{c_0^2 2^{2N(C_k(R_Q)-R_k)}}{2}\right)\right)}_{\text{const.}} + \lambda = 0. \tag{3.53}$$

Since the constant term defines a transcendent equation one can give only an implicit solution as

$$R_k(\theta) = \left\{R_k : \frac{\partial}{\partial R_k} \exp\left(-\frac{c_0^2 2^{2N(C_k(R_Q)-R_k)}}{2}\right) = \theta\right\} \tag{3.54}$$

$$= \left\{R_k : e^{2N(C_k(R_Q)-R_k)} \exp\left(-\frac{c_0^2 e^{2N(C_k(R_Q)-R_k)}}{2}\right) = \theta\right\}, \tag{3.55}$$

where $C_k(R_Q)$ and R_k in (3.55) are in nats to avoid additional factors and in the other equations in bits.

The first exponential is strictly increasing in $C_k(R_Q) - R_k$, while the second exponential is striclty decreasing. Hence, the resulting equation may be nonmonotonic and therefore multiple solutions may exist. Since the probability of an error decreases doubly exponential with $C_k(R_Q) - R_k$ one optimal way to allocate the rates of the source R_k

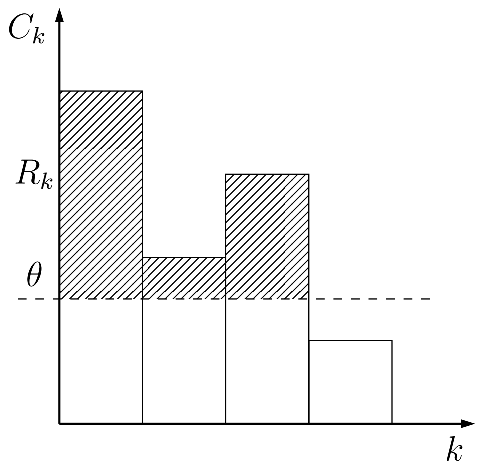

Figure 3.8: Rate allocation.

would be to keep $C_k(R_Q) - R_k$ constant for all k (Fig. 3.8):

$$R_k(\theta) = \max(0, C_k(R_Q) - \theta). \tag{3.56}$$

3.5 Noisy Feedback

In the previous chapter the assumption was that the feedback is perfect, i.e., the transmitter perfectly receives the feedback signal sent by the receiver. Since this is a very convenient assumption, which simplifies analysis, the perfect feedback case got most attention in literature. In reality the feedback would be hardly perfect, at least the feedback is quantized, what is discussed in the following chapter. It turns out that doubly exponential decreasing error probability is not always achieved with noisy feedback. Actually it is only achieved in special cases. Generally the noise accumulates in the iterative process and the performance of the system breaks down with further increasing blocklength.

3.5.1 Scalar Channel

In [7] the scalar case of noisy feedback was investigated and an optimal linear feedback scheme was derived, which is similar to the SK-scheme. Since the SK-scheme is designed for perfect feedback, it suffers from noise accumulation. The central idea of an optimal scheme with noisy feedback is that the iterations do not only aim in canceling the noise initially introduced while transmitting the message at the initial transmission, but finding an optimal tradeoff in cancelling the noise and retransmitting the message. In [6] the noisy feedback was specialized to quantization noise in the feedback path.

Fig. 3.9 shows the communication system with feedback and quantizers in forward an backward path and the equivalent system as in [6]. In the original system the forward path introduces the noise of the channel $z[n]$ and the noise of the quantizer. Without loss of generality the total introduced noise is normalized to have unit power and termed as $z[n]$. The basic assumption is that, since the transmitter knows all previously transmitted values, this is equivalent to having the side-information $z_q[n'] = z[n'] + n_q[n'] \quad n' < n$ at iteration n. This allows to describe the system as a scheme with linear feedback. So

$$x[n] \quad = \quad \sum_{n'<n} \underbrace{F[n,n'](z[n'] + n_q[n'])}_{\text{noise cancellation part}} + \underbrace{g[n]\theta}_{\text{message part}} \quad , \tag{3.57}$$

where $F[n, n']$ and $g[n]$ define the tradeoff between noise cancellation and message (re)transmission and are specified later. This can also be compactly written in matrix-vector notation as

$$\mathbf{x} = \boldsymbol{F}\mathbf{z_q} + \boldsymbol{g}\theta, \tag{3.58}$$

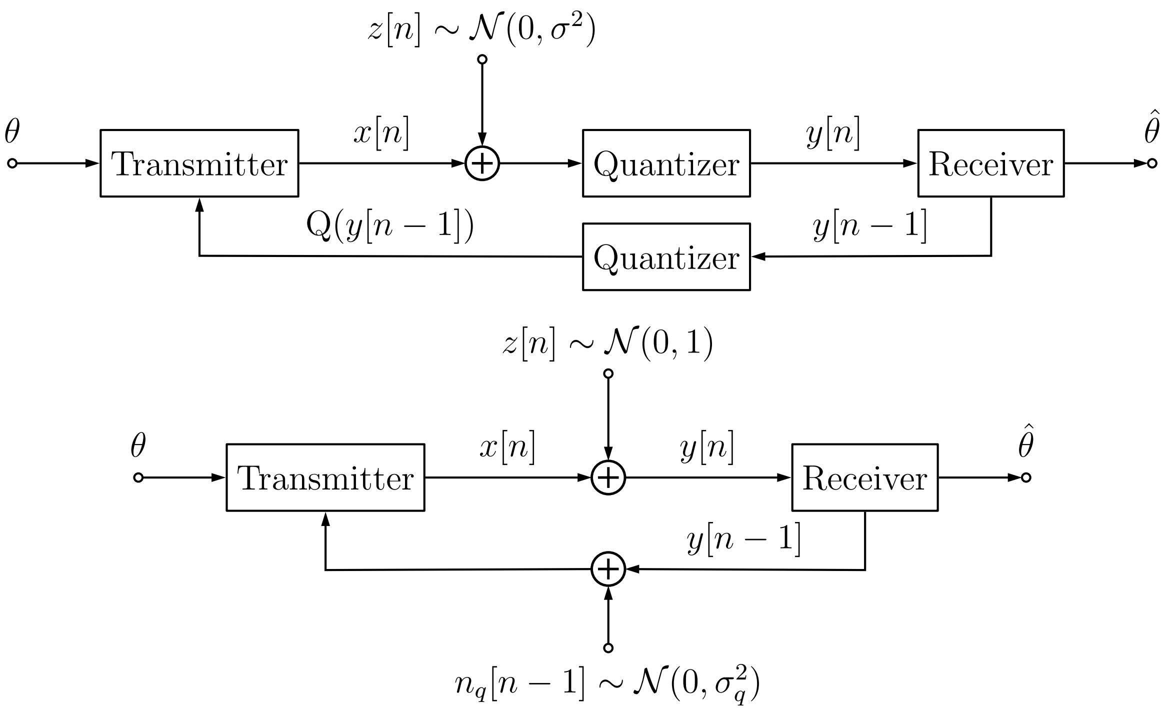

Figure 3.9: Communication system with feedback and quantizers in forward and feedback path (top); equivalent system as in [6] (bottom).

---

where the encoding matrix $\boldsymbol{F} \in \mathbb{R}^{N \times N}$ is a strictly lower triangular matrix in order to describe a strictly causal system, $\mathbf{n_q} \in \mathbb{R}^N$ is the quantization error of the received signal $\mathbf{y}$, $\boldsymbol{g} \in \mathbb{R}^N$ is a unit norm vector, and θ is the message to be sent. The signal $\mathbf{x}$ fulfills the average power constraint ρ

$$\mathrm{E}\{\mathbf{x}^T \mathbf{x}\} \leq N\rho. \tag{3.59}$$

Additionaly we know from the Elias result [12] that the optimal solution for a linear feedback scheme fulfills the power constraint not only on average, but

$$\mathrm{E}\{\mathrm{x}_i^2\} = \rho. \tag{3.60}$$

Note, since $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$, ρ is also the SNR of the forward path. An estimate of the message θ is obtained as a linear combination of the received signal $y[n]$:

$$\hat{\theta} \;=\; \sum_n q[n] y[n]. \tag{3.61}$$

We compactly write (3.61) in vector notation as

$$\hat{\theta} = \boldsymbol{q}^T \mathbf{y}, \tag{3.62}$$

where $\boldsymbol{q} \in \mathbb{R}^N$ is the combining vector. Inserting (3.58) in (3.62) yields

$$\begin{aligned} \hat{\theta} &= \boldsymbol{q}^T(\mathbf{x} + \mathbf{z}) & (3.63)\\ &= \boldsymbol{q}^T(\boldsymbol{F}\mathbf{z}_{\mathbf{q}} + \boldsymbol{g}\theta + \mathbf{z}) & (3.64)\\ &= \underbrace{\boldsymbol{q}^T\boldsymbol{g}\theta}_{\text{message part}} + \underbrace{\boldsymbol{q}^T(\boldsymbol{F} + \boldsymbol{I})\mathbf{z} + \boldsymbol{q}^T\boldsymbol{F}\mathbf{n}_q}_{\text{noise part}}. & (3.65) \end{aligned}$$

Since the message is assumed to be Gaussian with $\theta \sim \mathcal{N}(0, P_\theta)$ and the system consists only of linear operations and all involved noise processes are also Gaussian the estimate $\hat{\theta}$ and $\hat{\theta}|\theta$ are also Gaussian with

$$\begin{aligned} \hat{\theta} &\sim \mathcal{N}(0, P_\theta|\boldsymbol{q}^T\boldsymbol{g}|^2 + \|\boldsymbol{q}(\boldsymbol{F} + \boldsymbol{I})\|^2 + \sigma_q^2\|\boldsymbol{q}^T\boldsymbol{F}\|^2), & (3.66)\\ \hat{\theta}|\theta &\sim \mathcal{N}(\theta, \|\boldsymbol{q}(\boldsymbol{F} + \boldsymbol{I})\|^2 + \sigma_q^2\|\boldsymbol{q}^T\boldsymbol{F}\|^2). & (3.67) \end{aligned}$$

Thus, the mutual information $I(\theta; \hat{\theta})$ equals the capacity of the Gaussian channel

$$\begin{aligned} I(\theta; \hat{\theta}) &= h(\hat{\theta}) - h(\hat{\theta}|\theta) & (3.68)\\ &= \frac{1}{2}\log_2(1 + SNR) & (3.69)\\ &= \frac{1}{2}\log_2\left(1 + \frac{P_\theta|\boldsymbol{q}^T\boldsymbol{g}|^2}{\|\boldsymbol{q}^T(\boldsymbol{I} + \boldsymbol{F})\|^2 + \sigma_q^2\,\|\boldsymbol{q}^T\boldsymbol{F}\|^2}\right). & (3.70) \end{aligned}$$

The optimal values for $\boldsymbol{F}$, $\boldsymbol{g}$ and $\boldsymbol{q}$ which maximize the mutual information, or equivalently the SNR, are derived in [7] as

$$\begin{aligned} \boldsymbol{F} &= \begin{pmatrix} 0 & & \cdots & & 0\\ -\frac{1-\beta_0^2}{(1+\sigma_q^2)\beta_0} & 0 & & & \\ -\frac{1-\beta_0^2}{(1+\sigma_q^2)} & \ddots & \ddots & & \vdots\\ \vdots & \ddots & & & \\ -\frac{1-\beta_0^2}{(1+\sigma_q^2)}\beta_0^{N-3} & \cdots & -\frac{1-\beta_0^2}{(1+\sigma_q^2)} & -\frac{1-\beta_0^2}{(1+\sigma_q^2)\beta_0} & 0 \end{pmatrix}, & (3.71)\\ \boldsymbol{g} &= \sqrt{\frac{1-\beta_0^2}{1-\beta_0^{2N}}}\begin{pmatrix} 1 & \beta_0 & \beta_0^2 & \cdots & \beta_0^{N-1} \end{pmatrix}^T, & (3.72)\\ \boldsymbol{q} &= \boldsymbol{g}, & (3.73) \end{aligned}$$

where β_0 is the smallest positive root of

$$\beta^{2N} - \left[N + (1+\sigma_q^2)N\gamma\rho\right]\beta^2 + (N-1) \tag{3.74}$$

and $\gamma \in [0,1)$ is a power trade-off factor between transmitting side-information and transmitting the message. A power trade-off factor $\gamma = 0$ would mean the transmit power is used to transmit the message θ only. This would result in a repetition code, which is clearly suboptimal (except in some trivial cases, e.g., zero feedback rate). A power trade-off factor $\gamma = 1$ would mean the transmit power is used to transmit the side-information. So focusing on the transmit power constraint yields

$$\begin{aligned}
\mathrm{E}\{\mathbf{x}^T\mathbf{x}\} &= \operatorname{tr}\left(\boldsymbol{F}\mathrm{E}\{(\mathbf{z}+\mathbf{n}_q)(\mathbf{z}+\mathbf{n}_q)^T\}\boldsymbol{F}^T\right) + \|\boldsymbol{g}\|^2\mathrm{E}\{\theta\}^2 && (3.75)\\
&= \underbrace{(1+\sigma_q^2)\|\boldsymbol{F}\|_F^2}_{\text{noise-cancelation power}} + \underbrace{\mathrm{E}\{\theta\}^2}_{\text{signal power}} && (3.76)\\
&\leq N\rho. && (3.77)
\end{aligned}$$

Using the tradeoff factor γ we can formulate the noise-cancellation and signal power as

$$\begin{aligned}
(1+\sigma_q^2)\|\boldsymbol{F}\|_F^2 &= \gamma N\rho, && (3.78)\\
\mathrm{E}\{\theta\}^2 &= (1-\gamma)N\rho. && (3.79)
\end{aligned}$$

For finite blocklenghts the optimal γ is in between, but can be only calculated numerically [7]. The choice of the value of the trade-off factor plays a crucial role in performance of the system. We will give optimal values for some asymptotic cases.

No knowledge of the quantized feedback at the receiver

If the receiver has no knowledge of the quantized feedback, but only of course knowledge of the unquantized feedback, this is equivalent to the transmission over an AWGN feedback channel, since the GIB quantizer introduces additional independent noise. The signal-to-noise ratio of the superchannel is then according to (3.70) and [7]

$$\begin{aligned}
SNR &= \frac{\mathrm{E}\{\theta^2\}|\boldsymbol{q}^T\boldsymbol{g}|^2}{\left\|\boldsymbol{q}^T(\boldsymbol{I}+\boldsymbol{F})\right\|^2 + \sigma_q^2\left\|\boldsymbol{q}^T\boldsymbol{F}\right\|^2} && (3.80)\\
&= \frac{(1+\sigma_q^2)N(1-\gamma)\rho}{\sigma_q^2 + \beta_0^{2(N-1)}}. && (3.81)
\end{aligned}$$

Since there is no closed-form solution for β_0 an approximation is given in [7] as

$$\beta_0 \approx \sqrt{\frac{N-1}{N+(1+\sigma_q^2)N\gamma\rho}} \approx \sqrt{\frac{1}{1+(1+\sigma_q^2)\gamma\rho}}, \tag{3.82}$$

where the second approximation is tight for $N \gg 1$. An upper bound for the SNR can be obtained if γ is chosen appropriately. If $N \to \infty$, $N\gamma \to \infty$, and $\gamma \to 0$, we have

$$SNR \overset{N\to\infty}{\to} \frac{1+\sigma_q^2}{\sigma_q^2} N\rho. \tag{3.83}$$

If we quantize the forward path with a GIB optimal quantizer the capacity is

$$C_{FW} = \frac{1}{2}\log_2(1+\rho) \triangleq I(R_q) = \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\frac{1-2^{-2R_q}}{1+P/\sigma^2 2^{-2R_q}}\right). \tag{3.84}$$

The first term is just the Shannon capacity for the Gaussian channel and the second term is the information-rate function in the scalar case. We showed that these terms have to be equal in (3.29)-(3.30). Consequently,

$$\rho = \frac{P}{\sigma^2}\frac{1-2^{-2R_q}}{1+P/\sigma^2 2^{-2R_q}}. \tag{3.85}$$

Also quantizing the feedback path with a GIB optimal quantizer yields

$$C_{FB} = \frac{1}{2}\log_2(1+SNR_{FB}) = \frac{1}{2}\log_2\left(1+\frac{\rho+1}{\sigma_q^2}\right), \tag{3.86}$$

because the power of the received signal, which is fed back is $\rho+1$, since the noise of the forward channel is normalized to have unit variance and the assumption that no additional noise, except the quantization noise, is added in the backward channel. Therefore, the quantization noise is

$$\sigma_q^2 = \frac{\rho+1}{2^{2R_{q_{FB}}}-1}. \tag{3.87}$$

Substituting ρ with (3.85) and σ_q^2 with (3.87) in the bound for the signal-to-noise ratio (3.83) yields

$$SNR \leq \underbrace{\frac{\rho+2^{2R_{q_{FB}}}}{\rho+1}}_{\geq 1} \underbrace{\underbrace{\frac{1-2^{-2R_q}}{1+P/\sigma^2 2^{-2R_q}}}_{\leq 1} \frac{P}{\sigma^2}}_{\rho=SNR_{FW}} N. \tag{3.88}$$

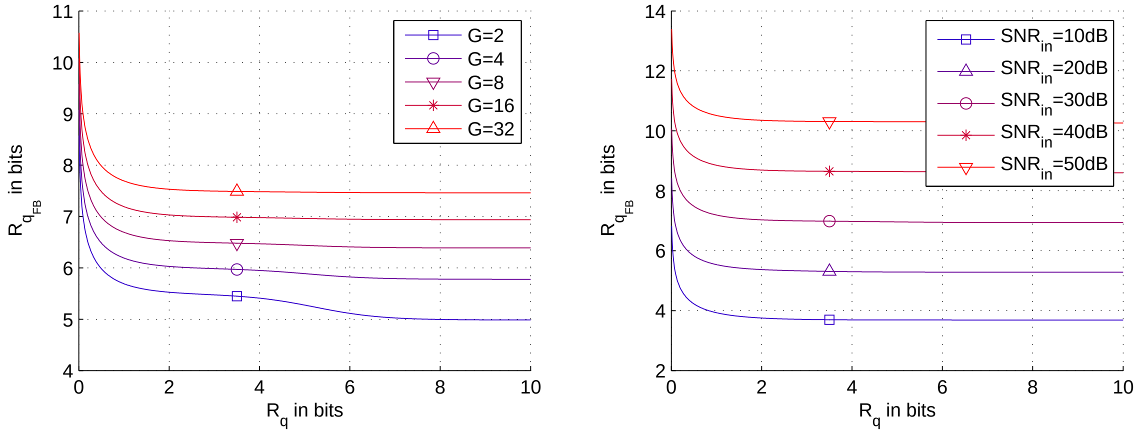

Figure 3.10: Trade-off between R_q and $R_{q_{FB}}$ for fixed $\frac{P}{\sigma^2} = SNR_{in} = 30$dB (left) and fixed $G = G_{rep} = 16$ (right).

Obviously the contribution of the quantized feedback path to the SNR is always positive and the quantziation of the forward path decreases the SNR. $\frac{P}{\sigma^2}N$ is just the SNR of the repetition code. So we define a SNR gain over the repetition code as

$$G_{rep} = \frac{SNR}{\frac{P}{\sigma^2}N}. \tag{3.89}$$

If we keep the gain constant for a given $\frac{P}{\sigma^2}$ there is a trade-off between the necessary quantization rates of the forward and feedback path (Fig. 3.10):

$$R_{q_{FB}} \geq \frac{1}{2}\log_2\left(G_{rep}\frac{P}{\sigma^2}\frac{\rho+1}{\rho} - \rho\right). \tag{3.90}$$

Achievable rate

In [3] it was shown that the SNR would have to have the form $(1+\rho)^N$ in order to achieve channel capacity. A rate R is defined to be achievable, if the error probability goes to zero if the blocklength goes to infinity. Hence, a bound for the achievable rate can be given as

$$\begin{aligned} R &\leq \lim_{N\to\infty}\frac{1}{N}\frac{1}{2}\log_2\left(1+(1+\rho)^N\right) && (3.91)\\ &= \frac{1}{2}\log_2(1+\rho). && (3.92)\end{aligned}$$

The factor $1/N$ is due to the fact that we need N channel uses to transmit one message. Comparing the capacity achieving form $(1+\rho)^N$ with (3.88) makes clear that for finite quantizer rates the achievable rate is zero.

In the case of a binary pulse amplitude modulated (PAM) signal the bit error probability is given as

$$P_e = \mathcal{Q}(\sqrt{2SNR}). \tag{3.93}$$

For general M-PAM signals, the probability of an error is [17]

$$P_e = 2\frac{M-1}{M}\mathcal{Q}\left(\sqrt{\frac{SNR}{M^2-1}}c_0\right), \tag{3.94}$$

for $M \gg 1$ we have

$$P_e \approx 2\mathcal{Q}\left(\frac{\sqrt{SNR}}{M}c_0\right), \tag{3.95}$$

where M is the size of the source alphabet and is related to the rate as

$$M = 2^{NR}. \tag{3.96}$$

Using the capacity achieving SNR yields

$$P_e \approx 2\mathcal{Q}\left(\left(\frac{1+\rho}{2^{2R}}\right)^{N/2} c_0\right) = 2\mathcal{Q}\left(2^{N(C-R)}c_0\right). \tag{3.97}$$

The error probability is again doubly exponential decreasing if the SNR has the capacity achieving form and if $R < \frac{1}{2}\log_2(1+\rho) = C$. In this case

$$\lim_{N\to\infty} P_e = 0. \tag{3.98}$$

On the other side, using the resulting SNR of the super channel (cf. (3.88)) yields

$$P_e \approx 2\mathcal{Q}\left(\left(2^{-2NR}\frac{\rho + 2^{2R_{q_{FB}}}}{\rho+1} \underbrace{\frac{1-2^{-2R_q}}{1+P/\sigma^2 2^{-2R_q}}\frac{P}{\sigma^2}}_{\rho} N\right)^{1/2} c_0\right). \tag{3.99}$$

Since the term 2^{-NR} decreases faster than the linear term in N, the error probability is not decreasing doubly exponential anymore. In fact the error probability is even

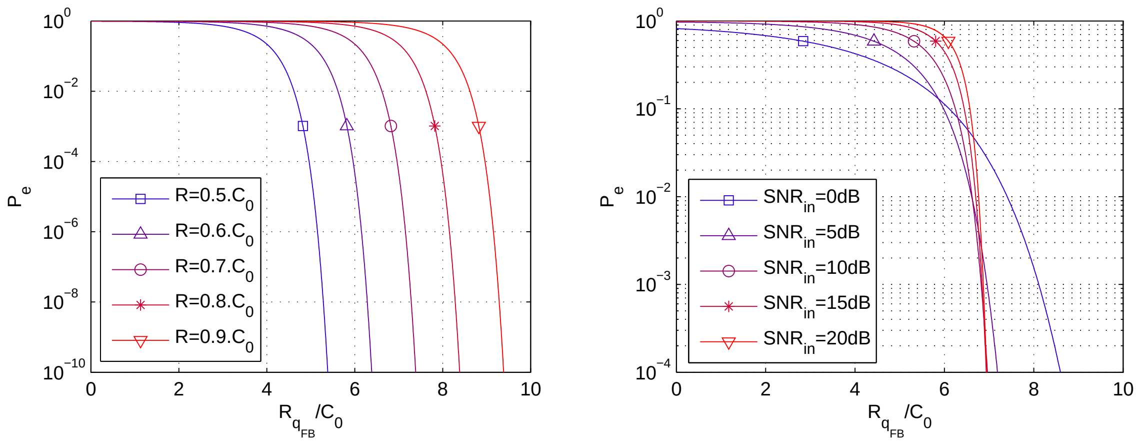

Figure 3.11: Error probability over $R_{q_{FB}}$ for fixed $N = 10$ and $SNR_{in} = 10$dB (left), $R = 0.7C_0$ (right).

---

increasing for increasing blocklengths. Basically this setup shows similar performance as a repetition code up to a constant factor (the previously introduced gain G_{rep}). As a consequence for finite rate feedback $R_{q_{FB}}$ we have

$$P_e > 0 \qquad \forall N. \tag{3.100}$$

This means no rate is achievable.

Corollary 3.3. *The achievable rate with feedback rate $R_{q_{FB}} < \infty$ is zero.*

Fig. 3.11 shows the error probability over the rate of the quantized feedback. The error decreases nearly doubly exponential in the feedback rate, which is due to the factor $2^{2R_{q_{FB}}}$ in (3.99). For small ρ, the doubly exponential slope is close. The impact of ρ and therefore of SNR_{in} on the slope is shown in Fig. 3.11 (right). Note that the x-axis is normalized to different capacities C_0 depending on SNR_{in}.

Letting $R_{q_{FB}}$ be possibly infinite, in order to achieve channel capacity, yields

$$\frac{\rho + 2^{2R_{q_{FB}}}}{\rho + 1}\rho N \;\hat{=}\; (1+\rho)^N \tag{3.101}$$

$$\Rightarrow R_{q_{FB}} = \frac{1}{2}\log_2\left(\frac{(1+\rho)^{N+1}}{\rho N} - \rho\right). \tag{3.102}$$

So $R_{q_{FB}}$ would have to grow linearly in the blocklength, which is clearly not feasible for large blocklengths.

Corollary 3.4. *The achievable rate with feedback rate $R_{q_{FB}} = O(N)$ equals capacity.*

3.6 Quantized Feedback

In the previous chapters the feedback was either perfect or noisy and the receiver had no knowledge about the feedback noise. If the noise in the feedback path is introduced by a quantizer the assumption that the noise is known at receiver side is valid, since the quantization process takes place at receiver side.

Perfect knowledge of the quantized feedback at the receiver

Since the system is assumed to be linear, the contribution of the feedback quantization noise can be canceled at the receiver, if known. This is because the fact that the received signal

$$\mathbf{y} = (\boldsymbol{F} + \boldsymbol{I})\mathbf{z} + \underbrace{\boldsymbol{F}\mathbf{n_q}}_{\text{feedback noise part}} + \boldsymbol{g}\theta, \tag{3.103}$$

consists of the feedback quantization noise part. Thus, we can cancel the feedback quantization noise at receiver side and therefore define the noise-canceled received signal $\mathbf{y}'$ as

$$\mathbf{y}' = \mathbf{y} - \boldsymbol{F}\mathbf{n_q} = (\boldsymbol{F} + \boldsymbol{I})\mathbf{z} + \boldsymbol{g}\theta. \tag{3.104}$$

Then the estimated message $\hat{\theta}$ can be written as

$$\hat{\theta} = \boldsymbol{q}^T\mathbf{y}' = \boldsymbol{q}^T((\boldsymbol{F} + \boldsymbol{I})\mathbf{z} + \boldsymbol{g}\theta), \tag{3.105}$$

which is equivalent to the noiseless feedback case. The SNR is then given by (3.81) with $\sigma_q^2 = 0$, i.e.,

$$SNR = \frac{N(1-\gamma)\rho}{\beta_0^{2(N-1)}}. \tag{3.106}$$

Inserting the approximation (3.82) for β_0 yields

$$SNR = N(1-\gamma)\rho(1+\gamma\rho)^{N-1}. \tag{3.107}$$

Although the system is equivalent to the system with perfect feedback, the average transmit power ρ increases, since the noise contribution is only subtracted at the receiver, but is physically present as in the previous section. This done by maximizing the SNR by designing $\boldsymbol{F}$, $\boldsymbol{g}$ and $\boldsymbol{q}^T$ as there would be no noise in the backward channel.

This would result in a power constraint (3.59) given by [7]

$$\begin{aligned} \mathrm{E}\{\mathbf{x}^T\mathbf{x}\} &= \mathrm{tr}\left(\boldsymbol{F}\mathrm{E}\{\mathbf{z}^T\mathbf{z}\}\boldsymbol{F}^T\right) + ||\boldsymbol{g}||^2\mathrm{E}\{\theta\}^2 && (3.108) \\ &= \underbrace{||\boldsymbol{F}||_F^2}_{\text{noise-cancelation power}} + \underbrace{\mathrm{E}\{\theta\}^2}_{\text{signal power}} && (3.109) \\ &\leq N\rho. && (3.110) \end{aligned}$$

Then by using the power tradeoff parameter γ, the powers are given by

$$\begin{aligned} ||\boldsymbol{F}||_F^2 &= N\gamma\rho, && (3.111) \\ \mathrm{E}\{\theta\}^2 &= N(1-\gamma)\rho. && (3.112) \end{aligned}$$

However, the quantization noise of the backward channel is physically present, so the actual transmit power is given by

$$\begin{aligned} \mathrm{E}\{\mathbf{x}^T\mathbf{x}\} &= \mathrm{tr}\left(\boldsymbol{F}\mathrm{E}\{(\mathbf{z}+\mathbf{n}_q)^T(\mathbf{z}+\mathbf{n}_q\}\boldsymbol{F}^T\right) + ||\boldsymbol{g}||^2\mathrm{E}\{\theta\}^2 && (3.113) \\ &= \underbrace{(1+\sigma_q^2)||\boldsymbol{F}||_F^2}_{\text{new noise-cancelation power}} + \underbrace{\mathrm{E}\{\theta\}^2}_{\text{signal power}} && (3.114) \\ &= \underbrace{(1+\sigma_q^2)N\gamma\rho}_{\text{new noise-cancelation power}} + \underbrace{(1-\gamma)N\rho}_{\text{signal power}} && (3.115) \\ &= (1+\gamma\sigma_q^2)N\rho. && (3.116) \end{aligned}$$

Inserting the new average power, the average power scaled by the factor $(1+\gamma\sigma_q^2)^{-1}$, in (3.107) yields

$$SNR = N\rho\frac{1-\gamma}{1+\gamma\sigma_q^2}\left(1+\frac{\gamma}{1+\gamma\sigma_q^2}\rho\right)^{N-1}. \tag{3.117}$$

As shown before, the SNR would have to have the form $(1+\rho)^N$ in order to achieve channel capacity. For finite quantization rates ($\sigma_q^2 > 0$), the SNR does not have this form. Hence, it can not achieve channel capacity. But fortunately if γ is chosen appropriately as $\gamma \to 1$, then it can be shown that the asymptotic SNR is

$$
\begin{aligned}
SNR &\overset{N\to\infty}{\to} \left(1 + \frac{1}{1+\sigma_q^2}\rho\right)^N && (3.118)\\
&= \left(1 + \frac{2^{2R_{q_{FB}}}-1}{2^{2R_{q_{FB}}}+\rho}\rho\right)^N && (3.119)\\
&= \left(1 + \underbrace{\frac{2^{2R_{q_{FB}}}-1}{2^{2R_{q_{FB}}}+\rho}}_{\leq 1} \underbrace{\underbrace{\frac{1-2^{-2R_q}}{1+P/\sigma^2 2^{-2R_q}}}_{\leq 1} \frac{P}{\sigma^2}}_{\rho = SNR_{FW}}\right)^N. && (3.120)
\end{aligned}
$$

The capacity of the (Gaussian) superchannel is given as

$$C_S = \frac{1}{N}\frac{1}{2}\log_2\left(1 + SNR\right), \tag{3.121}$$

since a message is transmitted every N^{th} channel use. The achievable rate is limited by $R < C_S$, so we can give the bound as

$$
\begin{aligned}
R &\leq \lim_{N\to\infty} \frac{1}{N}\frac{1}{2}\log_2\left(1 + SNR\right) && (3.122)\\
&= \frac{1}{2}\log_2\left(1 + \frac{1}{1+\sigma_q^2}\rho\right), && (3.123)
\end{aligned}
$$

which is strictly positive. We identify (3.123) as the channel capacity of the closed loop, so we can state following

Corollary 3.5. *The achievable rate is*

$$R \leq \lim_{N\to\infty} \frac{1}{N}\frac{1}{2}\log_2\left(1 + SNR\right) = C_{CL}, \tag{3.124}$$

where C_{CL} is the channel capacity of the closed loop $\mathbf{TX} \overset{\mathbf{FW}}{\to} \mathbf{RX} \overset{\mathbf{FB}}{\to} \mathbf{TX}$*, if the realization of the feedback is known at the receiver. Otherwise the achievable rate is* 0*.*

We define a gain over the SNR with perfect quantization ($R_q \to \infty$, $R_{q_{FB}} \to \infty$) as

$$G = \frac{1 + \frac{1}{1+\sigma_q^2}\rho}{1 + \frac{P}{\sigma^2}} = \frac{2^{2C_{CL}}}{2^{2C_0}} = 2^{-2(C_0 - C_{CL})}. \tag{3.125}$$

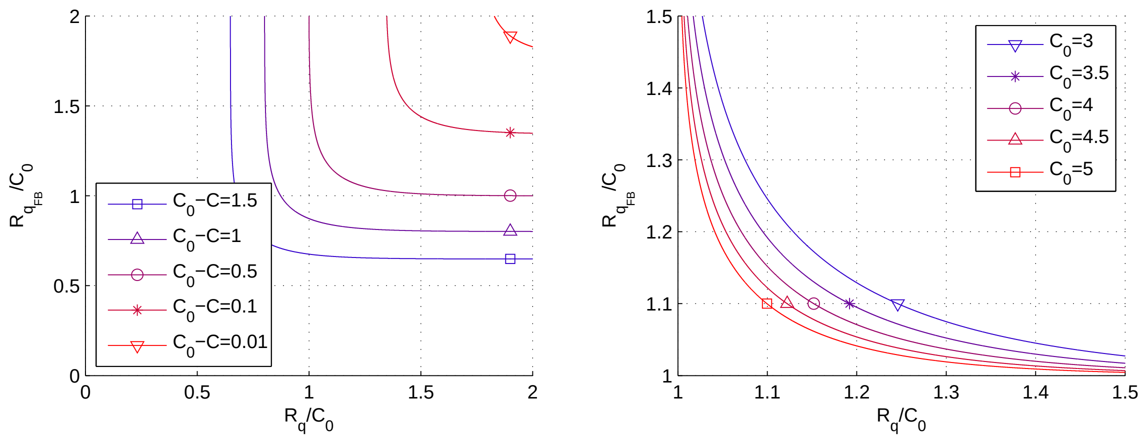

Figure 3.12: Trade-off between R_q and $R_{q_{FB}}$ for fixed $C_0 = 4$ (left) and fixed $C_0 - C = 0.5$ (right).

Thus, a constant gain corresponds to a constant $C_0 - C_{CL}$ and if we keep the gain constant for a given $\frac{P}{\sigma^2}$ there is a trade-off between the necessary quantization rates of the forward and feedback path (Fig. 3.12):

$$R_{q_{FB}} \geq \frac{1}{2}\log_2\left(\frac{G(1+P/\sigma^2)}{1-(G(1+P/\sigma^2)-1)/\rho}\right). \tag{3.126}$$

The probability of an error is given by (cf. (3.94))

$$P_e \approx 2\mathcal{Q}\left(\frac{\sqrt{SNR}}{2^{NR}}c_0\right), \tag{3.127}$$

if M-PAM with alphabet size $M = 2^{NR}$ is used. Inserting (3.118) in (3.127) yields

$$\begin{aligned}
P_e &\approx 2\mathcal{Q}\left(\left(\frac{1+\frac{1}{1+\sigma_q^2}\rho}{2^{2R}}\right)^{N/2} c_0\right) &(3.128)\\
&= 2\mathcal{Q}\left(\left(\frac{1+\frac{2^{2R_{qFB}}-1}{2^{2R_{qFB}}+\rho}\frac{1-2^{-2R_q}}{1+P/\sigma^2 2^{-2R_q}}\frac{P}{\sigma^2}}{2^{2R}}\right)^{N/2} c_0\right) &(3.129)\\
&= 2\mathcal{Q}\left(2^{N(C_{CL}-R)}\right). &(3.130)
\end{aligned}$$

Again the error probability decreases doubly exponential in blocklength. This is true if $R < C_{CL}$, while in the perfect feedback case the constraint was $R < C$. For finite

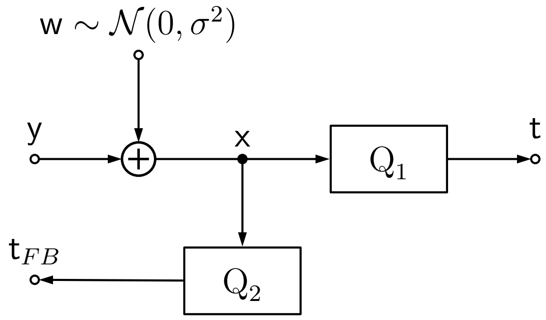

Figure 3.13: Communication system with feedback (different quantizers).

quantization rates (corresponding to $\sigma_q^2 > 0$) we have $C_{CL} < C$ because

$$C_{CL} = \frac{1}{2}\log_2\left(1 + \frac{1}{1+\sigma_q^2}\rho\right) < \frac{1}{2}\log_2(1+\rho) = C. \tag{3.131}$$

Corollary 3.6. *If the rate is limited by $R < C_{CL} < C$, the error probability tends to 0 as*

$$\lim_{N\to\infty} P_e^{(N)}(R) = 0. \tag{3.132}$$

Thus, the maximum achievable rate is C_{CL}.

Corollary 3.7. *For large blocklengths N, the error probability decreases doubly exponential as*

$$P_e \leq 2\mathcal{Q}(2^{N(C_{CL}-R)}), \tag{3.133}$$

if $R < C_{CL} < C$.

Alternative representation of the quantizers

Using GIB optimized quantizers, with different rates, in the forward path and in the feedback path, where the introduced noise is assumed to be correlated, yields an equivalent system (Fig. 3.14). Basically we jointly quantize the forward and feedback signal and it is assumed that $R_q \geq R_{q_{FB}}$.

We showed that the information bottleneck quantization can equivalently be model by an additive zero-mean Gaussian noise term, now called $\mathsf{w}_{q_{FW}}$ and $\mathsf{w}_{q_{FB}}$, for different quantization rates. The joint quantization is then performed in a way that

$$\mathsf{w}_{q_{FB}} = \mathsf{w}_{q_{FW}} + \Delta\mathsf{w}_q, \tag{3.134}$$

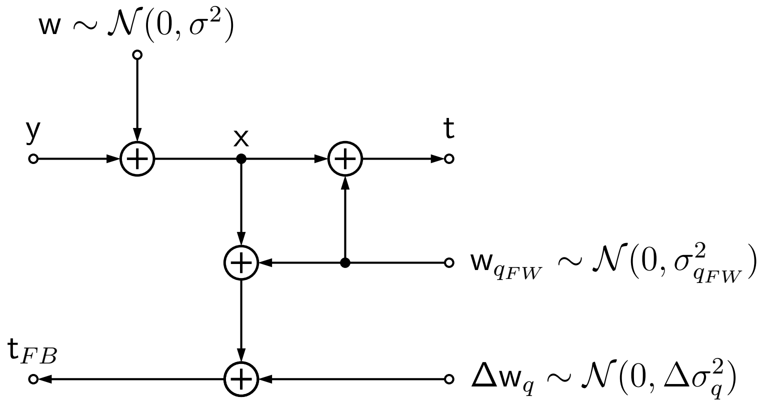

Figure 3.14: Equivalent system with feedback (different quantizers).

with $\Delta\mathsf{w}_q \sim \mathcal{N}(0, \sigma^2_{q_{FB}} - \sigma^2_{q_{FW}})$. Thus, the mutual information is still given by

$$I(\mathsf{t};\mathsf{y}) = I(R_q), \tag{3.135}$$

$$I(\mathsf{t}_{FB};\mathsf{y}) = I(R_{q_{FB}}), \tag{3.136}$$

and $\Delta\mathsf{w}_q$ can be interpreted as a further quantization, that is $\mathsf{t}_{FB} = \mathsf{t} + \Delta\mathsf{w}_q$. This means that the variance of the additional introduced noise $\Delta\mathsf{w}_q$ is the difference of the both quantizer noise variances. Thus, normalizing the total noise in the forward path $\mathsf{z} \sim \mathcal{N}(0,1)$, as in Fig. 3.9, yields

$$\Delta\sigma^2_q = \sigma^2_{q_{FB}} - \sigma^2_{q_{FW}} \tag{3.137}$$

$$= \left(\frac{P}{\sigma^2} + 1\right) \frac{2^{2R_q} - 2^{2R_{q_{FB}}}}{\left(2^{2R_q} + P/\sigma^2\right)\left(2^{2R_{q_{FB}}} - 1\right)}. \tag{3.138}$$

The resulting SNR is then again given by (3.117), with σ^2_q replaced by $\Delta\sigma^2_q$:

$$SNR = N\rho \frac{1-\gamma}{1+\gamma\Delta\sigma^2_q}\left(1 + \frac{\gamma}{1+\gamma\Delta\sigma^2_q}\rho\right)^{N-1}. \tag{3.139}$$

If the term growing exponentially in N is in the same magnitude as the term growing linearly in N, there is a trade-off between R_q and $R_{q_{FB}}$ (Fig. 3.15). For growing N the SNR gets less sensitive in R_q and therefore the achievable rate and SNR for $N \to \infty$

are independent of the actual rate R_q, according to Corollary 3.5:

$$\begin{aligned} SNR &\stackrel{N\to\infty}{\to} \left(1+\frac{2^{2R_{q_{FB}}}-1}{2^{2R_{q_{FB}}}+P/\sigma^2}\frac{P}{\sigma^2}\right)^N, && (3.140)\\ R &\leq C_{CL} = I(R_{q_{FB}}), && (3.141)\\ P_e &\approx 2\mathcal{Q}\left(2^{N(C_{CL}-R)}\right) = 2\mathcal{Q}\left(2^{N(I(R_{q_{FB}})-R)}\right). && (3.142) \end{aligned}$$

These equations can also directly be obtained by substituting σ_q^2 with (3.138) in (3.139) and optimally letting $\gamma \to 1$ as $N \to \infty$ and we follow

Corollary 3.8. *The achievable rate is*

$$R \leq \lim_{N\to\infty} \frac{1}{N}\frac{1}{2}\log_2(1+SNR) = I(R_{q_{FB}}), \tag{3.143}$$

if the feedback quantization noise is known at the receiver and $R_q \geq R_{q_{FB}}$. Otherwise the achievable rate is 0*.*

Since in this representation of the quantizers the capacity of the loop $C_{CL} = I(R_{q_{FB}})$, if N goes to infinity, we can concretize corollary 3.6 and 3.7 as follows:

Corollary 3.9. *If the rate is limited by $R < I(R_{q_{FB}}) < C_0$ and $R_q \geq R_{q_{FB}}$, the error probability tends to 0 as*

$$\lim_{N\to\infty} P_e^{(N)}(R) = 0. \tag{3.144}$$

Thus, the achievable rate is $I(R_{q_{FB}})$.

Corollary 3.10. *For large blocklengths N, the error probability decreases doubly exponential as*

$$P_e \leq 2\mathcal{Q}(2^{N(I(R_{q_{FB}})-R)}), \tag{3.145}$$

if $R < I(R_{q_{FB}})$.

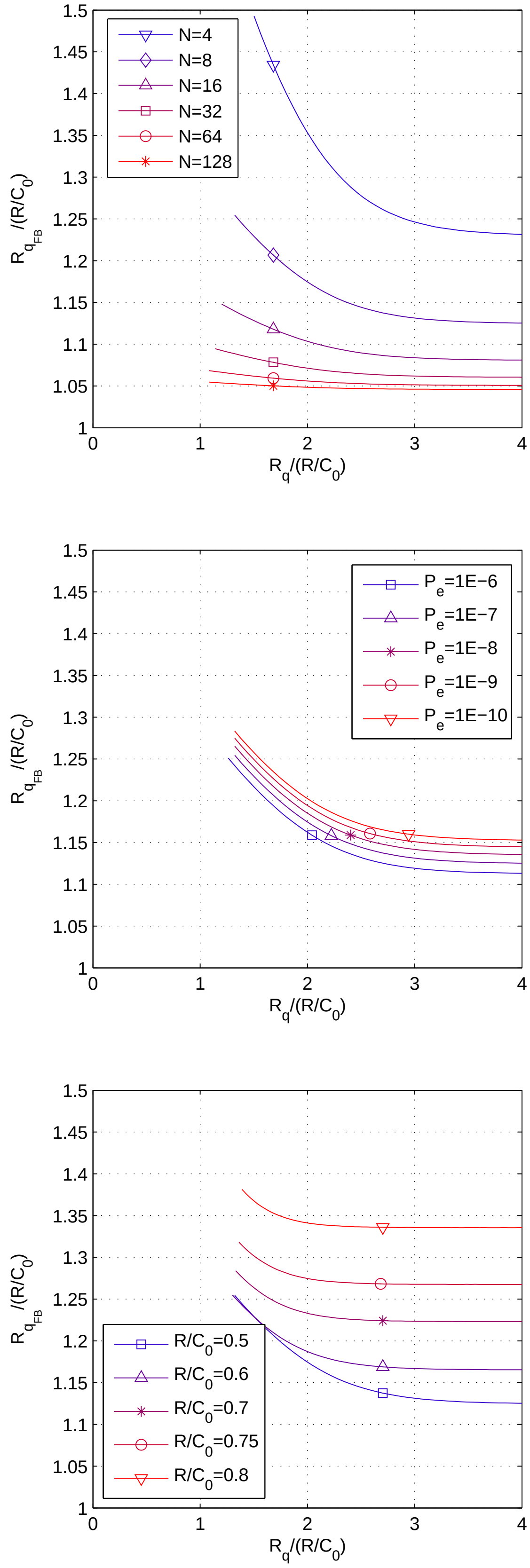

Figure 3.15: Trade-Off between R_q and $R_{q_{FB}}$ for fixed $SNR = 20$dB: $R/C_0 = 0.5$ and $P_e = 10^{-7}$ (top); $R/C_0 = 0.5$ and $N = 8$ (middle); $P_e = 10^{-7}$ and $N = 8$ (bottom).

3.6.1 Vector Channel

In this section we study the optimal feedback quantization rate allocation, for given quantization in the forward path. The quantization rate allocation of the forward path is given by the optimal rate allocation, which maximizes the information-rate function, as showed in the previous chapter. The SNRs of the individual modes of the forward path are denoted as SNR_{FW_i}. We know that the i.i.d. (Gaussian) message vector $\boldsymbol{\theta}$ is transmitted over the Gaussian vector channel, which can be decomposed as a sum of parallel scalar channels (modes). The information-rate I_S of the superchannel is then

$$I_S = \frac{1}{N} I(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}) \triangleq \frac{1}{N} \sum_{i}^{n} C_i = \frac{1}{N} \sum_{i}^{n} I(\hat{\theta}_i; \theta_i), \tag{3.146}$$

since we need N channel uses to transmit one message $\boldsymbol{\theta}$. Note that in what follows we consider the power allocation on the individual modes or equivalently the SNR_{FW_i} to be fixed. Thus we do not obtain the global optimum $I_S = C_S$. In order to obtain the capacity of the vector channel we would have to jointly optimize power allocation and quantization of the scalar modes, which is a much harder problem.

Optimal feedback quantization rate allocation (no knowledge of the quantized feedback at the receiver)

The information-rate of a Gaussian vector channel is given as the sum of capacities of the parallel scalar channels with SNR_i as in (3.88):

$$\begin{aligned} I = \sum_{i=1}^{n} C_i &= \frac{1}{2} \sum_{i=1}^{n} \log_2 (1 + SNR_i) && (3.147) \\ &= \frac{1}{2} \sum_{i=1}^{n} \log_2 \left(1 + \frac{SNR_{FW_i} + 2^{2R_{q_{FB}}}}{SNR_{FW_i} + 1} SNR_{FW_i} N \right), && (3.148) \end{aligned}$$

where for simpler notation, R_i are the feedback quantization rates of the individual modes and fulfill

$$R_{q_{FB}} = \sum_{i=1}^{n} R_i. \tag{3.149}$$

The maximization of the overall information-rate is a constrained maximization problem, can be solved by Lagrange multipliers and reads

$$\begin{aligned}\Lambda(R_1,R_2,\ldots,R_n,\lambda) &= \frac{1}{2}\sum_{i=1}^{n}\log_2\left(1+\frac{SNR_{FW_i}+2^{2R_{q_{FB}}}}{SNR_{FW_i}+1}SNR_{FW_i}N\right)\\ &\quad +\lambda\left(\sum_{i=1}^{n}R_i-R\right), \qquad (3.150)\\ \frac{\partial\Lambda(R_1,R_2,\ldots,R_n,\lambda)}{\partial R_i} &= \frac{2^{2R_i}SNR_{FW_i}}{2^{2R_i}SNR_{FW_i}+SNR_{FW_i}^2+(SNR_{FW_i}+1)/N} \qquad (3.151)\\ &\quad +\lambda=0.\end{aligned}$$

A parameterized solution for the rates is

$$\begin{aligned}R_i(\theta) &= \max\left(0,\frac{1}{2}\log_2\frac{SNR_{FW_i}^2+(SNR_{FW_i}+1)/N)}{\theta SNR_{FW_i}}\right) \qquad (3.152)\\ &\approx \max\left(0,\frac{1}{2}\log_2\frac{SNR_{FW_i}}{\theta}\right). \qquad (3.153)\end{aligned}$$

Optimal feedback quantization rate allocation (perfect knowledge of the quantized feedback at the receiver)

The information-rate of a Gaussian vector channel is again, with SNR_i as in (3.120):

$$\begin{aligned}I=\sum_{i=1}^{n}C_i &= \frac{1}{2}\sum_{i=1}^{n}\log_2\left(1+SNR_i\right) \qquad (3.154)\\ &= \frac{1}{2}\sum_{i=1}^{n}\log_2\left(1+\left(1+\frac{2^{2R_i}-1}{2^{2R_i}+SNR_{FW_i}}SNR_{FW_i}\right)^N\right), \qquad (3.155)\end{aligned}$$

where for simpler notation, R_i are the feedback quantization rates of the individual modes and fulfill

$$R_{q_{FB}}=\sum_{i=1}^{n}R_i. \qquad (3.156)$$

The maximization of the overall information-rate is a constrained maximization problem, can be solved by Lagrange multipliers and formulates reads

$$\begin{aligned}\Lambda(R_1, R_2, \ldots, R_n, \lambda) &= \frac{1}{2}\sum_{i=1}^{n}\log_2\left(1+\left(1+\frac{2^{2R_i}-1}{2^{2R_i}+SNR_{FW_i}}SNR_{FW_i}\right)^N\right) \\ &\quad +\lambda\left(\sum_{i=1}^{n}R_i - R\right), &(3.157)\\ \frac{\partial\Lambda(R_1, R_2, \ldots, R_n, \lambda)}{\partial R_i} &= N\frac{SNR_{FW_i}}{2^{2R_i}+SNR_{FW_i}}+\lambda = 0, &(3.158)\end{aligned}$$

for $N \to \infty$. A parameterized solution for the rates is

$$R_i(\theta) = \max\left(0, \frac{1}{2}\log_2\frac{SNR_{FW_i}}{\theta}\right). \tag{3.159}$$

Optimal feedback quantization rate allocation (perfect knowledge of the quantized feedback at the receiver; alternative representation of the quantizers)

We showed that in the case of large blocklength the SNR and therefore also the capacity is independent of the actual quantization rates of the forward path as long as $R_q > R_{q_{FW}}$. Actually we showed that the information-rate is equal to the information rate function $I(R_{q_{FB}})$. Thus, the problem reduces to the maximization of $I(R_{q_{FB}})$, which we already studied in the previous chapter.

4

Numerical Results

4.1 Numerical Evaluation

In the previous chapter we gave equations for mutual information, SNR and error probabilities including the power tradeoff factor γ. Generally it is not possible to give analytic values of the optimal γ for finite blocklengths. So the last chapter focused on the asymptotic performance for blocklengths $N \to \infty$, where we were able to give explicit values for optimal γ. Now we study the performance for finite blocklength and give numerical solutions and performance comparisons. Since our iterative scheme transmits a message every N^{th} iteration one performance measure is the normalized mutual information per channel use, which is equivalent to the channel capacity of the super-channel

$$C_S = \frac{1}{N} I(\hat{\theta}; \theta) = \frac{1}{N} \frac{1}{2} \log_2 (1 + SNR) \,. \tag{4.1}$$

The second performance measure is the error probability for the normalized source rate NR

$$P_e \approx 2Q \left(\frac{\sqrt{SNR}}{2^{NR}} \right) . \tag{4.2}$$

We will study the performance of perfect feedback, noisy feedback and quantized feedback (noise known at the receiver). Recall the most general expressions for the SNRs. The SNR of just the AWGN channel is given by

$$SNR_0 = \frac{P}{\sigma^2}. \tag{4.3}$$

The SNR for perfect feedback ($\sigma_q^2 = 0$) and noisy feedback is given by (3.81)

$$SNR_{nf} = \frac{(1+\sigma_q^2)N(1-\gamma)\rho}{\sigma_q^2 + \beta_0^{2(N-1)}}, \qquad \beta_0 \approx \sqrt{\frac{N-1}{N+(1+\sigma_q^2)N\gamma\rho}}, \tag{4.4}$$

$$SNR_{pf} = \frac{N(1-\gamma)\rho}{\beta_0^{2(N-1)}}, \qquad \beta_0 \approx \sqrt{\frac{N-1}{N+N\gamma\rho}}. \tag{4.5}$$

In the case of known quantized feedback noise the SNR is given by (3.106) and (3.116)

$$SNR_{qf} = \frac{N(1-\gamma)\rho}{(1+\gamma\sigma_q^2)\beta_0^{2(N-1)}}, \qquad \beta_0 \approx \sqrt{\frac{N-1}{N+N\gamma\rho/(1+\gamma\sigma_q^2)}}. \tag{4.6}$$

Note that generally β_0 is given by (3.82), but is different in all three cases, since it depends on the optimal γ. The perfect feedback case, is equivalent to the noisy feedback case with $\sigma_q^2 = 0$ and if the quantized feedback noise is known this case is equivalent to the perfect feedback case, but ρ scaled by $(1+\gamma\sigma_q^2)^{-1}$ as argued in (3.116). ρ is the SNR of the forward channel and therefore given by (3.85)

$$\rho = \frac{P}{\sigma^2}\frac{1-2^{-2R_q}}{1+P/\sigma^2 2^{-2R_q}}. \tag{4.7}$$

We consider σ_q^2 as the additional introduced noise of the quantizer in the feedback path, what we studied in the previous chapter. σ_q^2 is then given by (3.138)

$$\sigma_q^2 = \left(\frac{P}{\sigma^2}+1\right)\frac{2^{2R_q}-2^{2R_{q_{FB}}}}{\left(2^{2R_q}+P/\sigma^2\right)\left(2^{2R_{q_{FB}}}-1\right)}. \tag{4.8}$$

Fig. 4.1 shows the numerical optimization of the power tradeoff factor γ for an exemplary operation at $R = 0.8C_0 = 1$bit ($C_0 = 1.25$bits). Clearly the case of perfect feedback with $R_Q = 10$bits (plotted in green with star type markers) is an upper performance bound, where the mutual information $\frac{1}{N}I(\hat{\theta};\theta) = I(R_q) \approx C_0$, since the channel output quantizer can be considered to have high rate $R_Q \gg C_0$ in this example. The mutual information for known feedback quantization noise (plotted in red with triangle type markers) asymptotically approaches $I(R_{q_{FB}})$. If the feedback quantization noise is unkown (plotted in blue with cross type markers) the mutual information can be kept constant up to a specific blocklength, if the feedback noise is low. For longer blocklengths the noise accumulation is too high and the performance quickly breaks down. However there may be an optimal blocklength, where the error probability can be decreased, compared to the error probability of just the AWGN channel. This is the case if the feedback quantizer rate is high, or equivalently the feedback quantization noise is low. If the feedback quantization noise is known and it is ensured that $I(R_{q_{FB}}) > R$, the error probability tends to zero, since $\lim_{n\to\infty} \frac{1}{N} I(\hat{\theta};\theta) = I(R_{q_{FB}})$. Although in the asymptotic regime the performance is independent of the quantizer in the forward path (as long as $R_Q \geq R_{q_{FB}}$), this is not true for finite blocklength. If we compare the error probabilities of perfect feedback ($R_Q = 2$bits, plotted in green with star type marker) and known quantized feedback noise ($R_Q = 10$bits and $R_{q_{FB}} = 2$bits, plotted in red with triangle type marker), the higher rate quantization in the forward path provides a substantial performance improvement, even though they have the same asymptotic performance.

4.2 Monte Carlo Simulation

We next validate our error probability expressions using Monte Carlo simulations. We use the presented source signal model to generate Gaussian distributed symbols. Recall that theses Gaussian distributed symbols are quantized with a distance of the quantization points given in (3.17) as

$$\Delta = \frac{\sqrt{2\pi e \sigma^2}}{2^{NR}}, \tag{4.9}$$

to match a source rate R. These iterative schemes require a symbol alphabet growing exponentially in the blocklength, which is numerically problematic: In the simulation this quantization process involves a rounding operation with an accuracy of 2^{NR}bits. Generally this numerical accuracy is not feasible for higher rates or longer blocklengths. The Monte Carlo simulations were performed using Matlab, which provides a relative accuracy of 2^{-52} [23]. Thus, the remaining accuracy of the distance of adjacent symbols is approximately $(NR - 52)$bits.

Fig. 4.2 shows the error probabilities of the Monte Carlo simulation compared to the numerically evaluated error probabilities of the previous section. Again the scheme exemplarily operates at $R = 0.8C_0 = 1$bit ($C_0 = 1.25$bits) up to a blocklength of 15, to avoid numerical errors. It can be seen that in the case of perfect feedback ($R_q = R_{q_{FB}} = 2$ and $R_q = R_{q_{FB}} = 10$) the simulations perfectly meet the theoretical results. In the case of known feedback quantization noise the performance in the simulation has the same properties as the theoretical results, but performs better in the low feedback rate regime ($I(R_{q_{FB}}) \approx R$). For higher feedback rates the simulation results are close to the theoretical findings. This is due to the fact that the discrete approximation of the Gaussian source becomes more exact for higher rates.

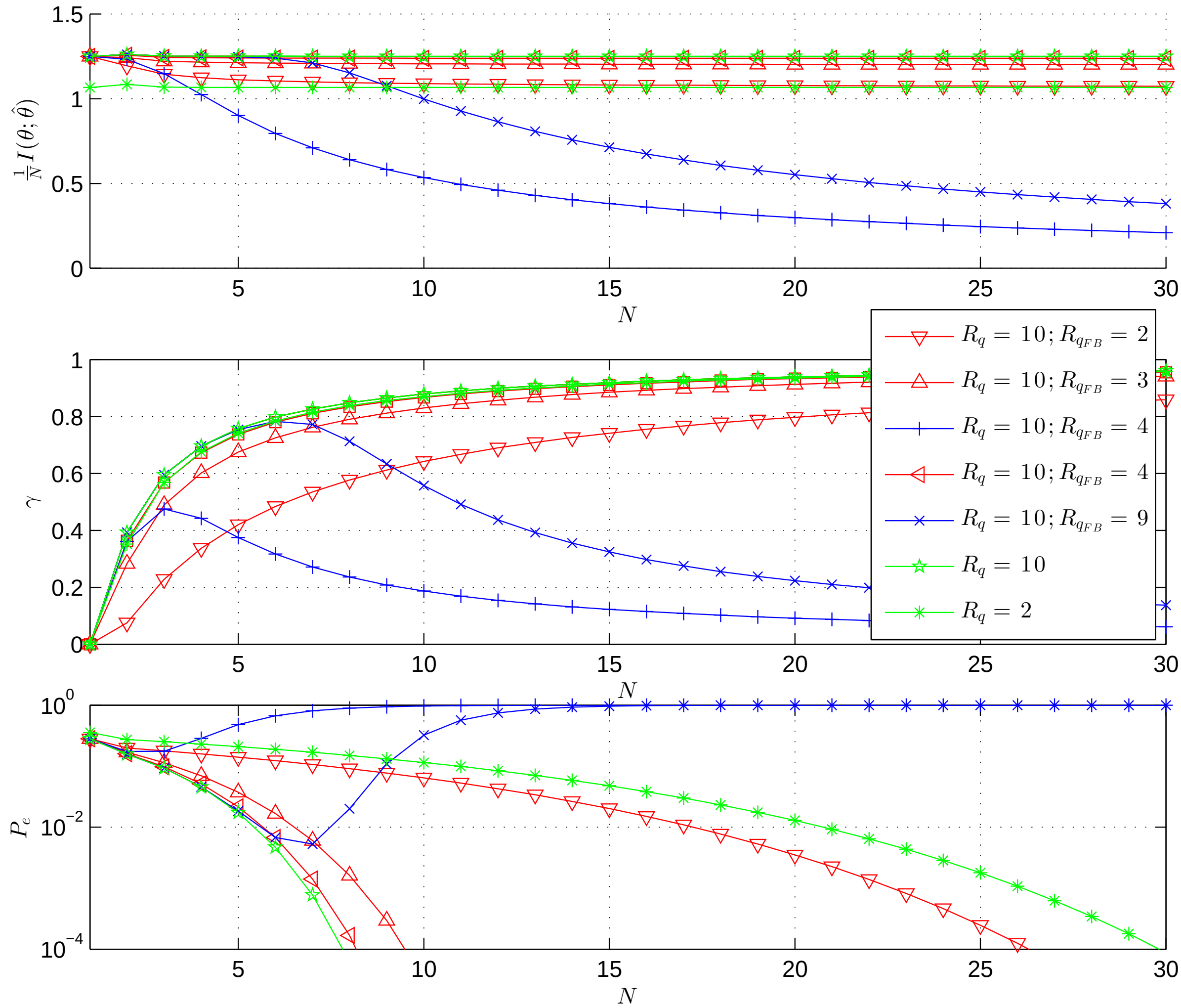

Figure 4.1: Example of numerical optimization of the tradeoff γ over the blocklength: mutual information $I(\hat{\theta};\theta)$ (top); optimal γ (middle); error probabilities P_e (bottom). $R = 0.8C_0 = 1$bit, $C_0 = 1.25$bits; perfect feedback (green, star type markers), noisy feedback (blue, cross type markers), modified scheme with known quantized noisy feedback (red, triangle type markers).

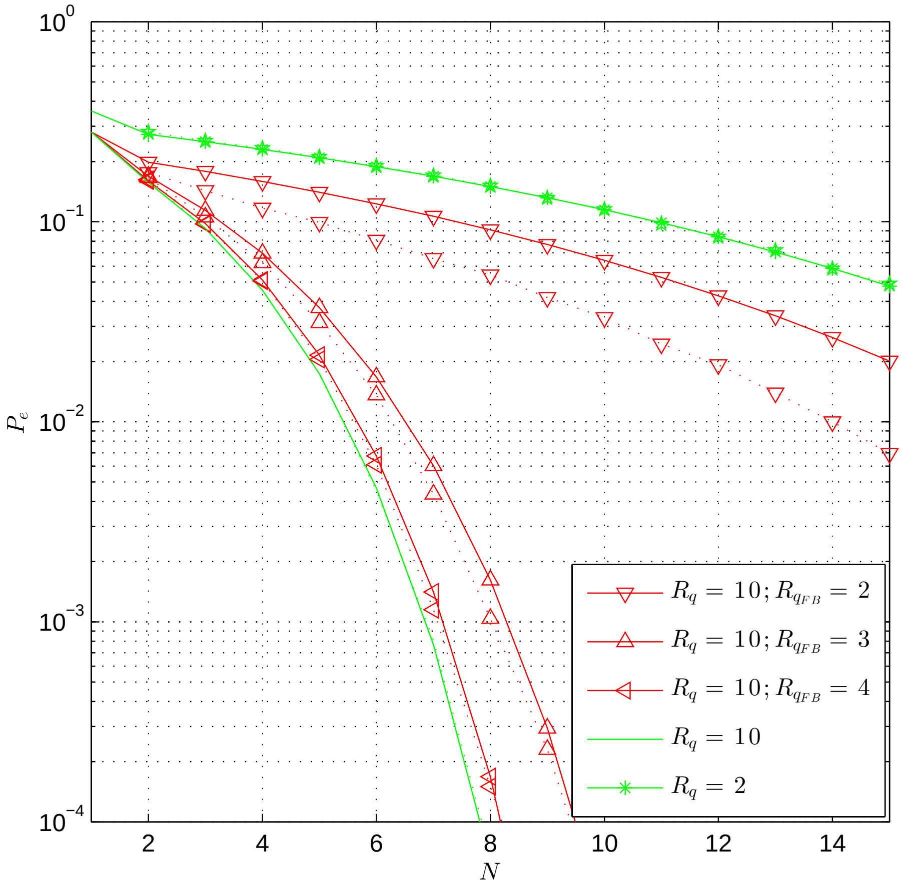

Figure 4.2: Comparison of theoretical (solid lines) and simulated (dashed lines) error probabilities P_e. $R = 0.8C_0 = 1$bit, $C_0 = 1.25$bits; perfect feedback (green, star type markers), modified scheme with known quantized noisy feedback (red, triangle type markers).

5

Conclusion and Outlook

It is known that communication systems with feedback do not increase the channel capacity of the AWGN channel. The channel capacity is the maximum rate at which error free communication is possible. However the channel capacity is an asymptotic measure which requires infinite blocklengths. So for communication systems with finite blocklengths the error probability is greater than zero. Altough feedback does not increase the channel capacity of AWGN channels, it dramatically improves the performance in terms of SNR and error probability at finite blocklengths. In [28] a linear system was suggested which has an error probability doubly exponential decreasing in blocklength. In [14] it was even shown that a nonlinear system with an error probability decreasing not only exponential in blocklength, but with an exponential order in blocklength, is possible.

As a drawback these systems require perfect feedback as their performance quickly collapses if the feedback is noisy. Also they do not consider quantization as a source of feedback noise. [22] studied quantized feedback, where the quantization noise was modelled as bounded noise. Basically the assumption is that the quantizer is deterministic and uniform and in order to achieve bounded quantization noise, the source signal also has to be bounded. They showed if the noise is appropriately bounded, this system also provides an error probability doubly exponential decreasing in blocklength.

In [20], [19], and [18] it was shown that in fact no *linear* system with noisy feedback can provide error probabilities decreasing doubly exponential in blocklength and also no linear system with noisy feedback can achieve any positive rate. This does not contradict the findings in [22], since very special requirements at the feedback noise were assumed.

In [4] and [5], an optimal linear system with possibly noisy feedback was developed, which serves as a basis in this thesis. They showed that the Schalkwijk-Kailath scheme presented in [28] is a suboptimal special case of their scheme. The noisy feedback was specialised in [6] to quantization noise where it was shown that the quantizer rate would have to grow linearly in blocklength in order to achieve a positive rate.

What all these systems have in common is the prerequisite of a PAM source signal equally spaced in some interval. However, we studied a schemes with quantizers modelled by the Gaussian information bottleneck which requires Gaussian signals. So we specialized the scheme to Gaussian messages via the approximation of the Gaussian distribution as a discrete distribution. As a consequence the source is of course not bounded anymore and so is the quantization noise. Therefore, systems which rely on bounded noise, as in [22] are not possible.

Thus, we formulated the general idea of designing a quantizer, which compresses the channel output in order to maximize the mutual information of the desired transmit signal. We formulated the information-rate function as a measure for the mutual information and showed that the AWGN channel including the Gaussian information bottleneck can also be interpreted as an AWGN channel for a Gaussian source.

We reassured that systems with GIB quantized feedback also cannot provide any positive rate, if the feedback quantization noise is assumed to be unknown as usual. As an important difference to the usual analysis the feedback quantization noise was then assumed to be known at receiver side. This assumption may be valid if for example the quantization of the feedback is considered as a re-quantization with a lower rate in order to fulfill rate requirements on the feedback channel. The contribution of the feedback noise can then be cancelled at the receiver, so the performance is basically the same as for systems with perfect feedback if power requirements on the source are neglected. Unfortunately the feedback noise is only mathematically canceled at receiver side, but of course physically present in the system. So the feedback noise increases the transmit power or conversely the feedback power reduces the available power of the desired signal. We also showed that in contrast to the system where the feedback noise is unknown, a positive rate is achievable and shows the same asymptotic performance as a system with perfect feedback with reduced channel capacity. It turned out that the resulting channel capacity of the system is equal to the channel capacity of the closed loop of forward channel and backward channel.

The channel capacity of the closed loop is only an asymptotic performance measure, since it is only achieved for infinite blocklengths. For finite blocklengths the system with noise cancelation performs better in terms of SNR and error probability than the system with perfect feedback but channel capacity of the closed loop, although both have the same achievable rate. The system with unknown feedback noise performs the same up to a specific number of iterations because the introduced error of the noise can be reduced via the trade off factor γ by choosing it slightly lower and so decrease the influence of the noise. This is only possible up to a specific number of iterations from where the accumulated error dominates. Hence, for further increasing number of iterations less and less power is allocate for retransmission of the feedback until the performance goes to the performance of the repetition code, which does not achieve any positive rate.

5.1 Outlook

- In this thesis we studied memoryless Gaussian channel with Gaussian inputs and thus applied the Gaussian information bottleneck to derive the information-rate function. The interpretation of the information-rate function as the resulting channel capacity including the quantizer is only valid in the sense that the information-rate function is indeed the limit of the mutual information for a Gaussian source. However the general meaning of the channel capacity is the maximum of mutual information over *all* possible source distributions. Thus, it is still an open problem how to jointly find the optimal information bottleneck and source distribution to give the information-rate function the general meaning of a channel capacity under output compression.

- For vector channels and jointly Gaussian source and channel output an open question is how to jointly design the source, i.e., optimal power allocation, and the quantizers, i.e., optimal rate allocation. It is known that a power allocation according to the waterfilling algorithm on the noise levels of the channel maximizes the mutual information for a given channel. Also we showed that the GIB quantizers maximize the mutual information for a given source, which can be interpreted as an reverse waterfilling rate allocation. That separate optimization generally results in suboptimal solutions becomes evident if we consider a low rate case where at least the weakest mode is inactive (provided with zero quantization rate), although transmit power was allocated on this mode.

- Another open question is if feedback schemes exist, which achieve a strictly positive rate in the presence of noisy feedback. It was shown by [18] that in fact no scheme with linear feedback achieves any positive rate.

Bibliography

[1] M. Agarwal, D. Guo, and M.L. Honig. Error Exponent for Gaussian Channels With Partial Sequential Feedback. *IEEE Trans. Inf. Theory*, 59(8):4757–4766, Aug. 2013.

[2] T. Berger. *Rate Distortion Theory.* Prentice Hall, Englewood Cliffs (NJ), 1971.

[3] S. Butman. A general formulation of linear feedback communication systems with solutions. *IEEE Trans. Inf. Theory*, 15(3):392–400, 1969.

[4] Z. Chance and D.J. Love. On Linear Processing for an AWGN Channel with Noisy Feedback. *CoRR*, abs/0909.0105, 2009.

[5] Z. Chance and D.J. Love. A noisy feedback encoding scheme for the Gaussian channel. In *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2010)*, pages 3482–3485, March 2010.

[6] Z. Chance and D.J. Love. On quantization of channel output feedback for the Gaussian channel. In *International Waveform Diversity and Design Conference (WDD 2010)*, pages 76–80, 2010.

[7] Z. Chance and D.J. Love. Concatenated Coding for the AWGN Channel With Noisy Feedback. *IEEE Trans. Inf. Theory*, 57(10):6633–6649, 2011.

[8] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss. Information Bottleneck for Gaussian Variables. *Journal of Machine Learning Research*, 6:165–188, 2005.

[9] T.M. Cover and S. Pombra. Gaussian feedback capacity. *IEEE Trans. Inf. Theory*, 35(1):37–43, Jan. 1989.

[10] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York, NY, USA, 1991.

[11] A.G. D'yachkov. Upper bounds on the error probability for discrete memoryless channels with feedback. *Problemy Peredachi Informatsii*, 11(4):13–28, 1975.

[12] P. Elias. Channel capacity without coding,. *MIT Research Laboratory of Electronics, Quarterly Progress Report*, pages 90–93, 1956.

[13] R.F.H. Fischer. *Precoding and Signal Shaping for Digital Transmission.* Wiley, 2005.

[14] R.G. Gallager and B. Nakiboglu. Variations on a Theme by Schalkwijk and Kailath. *IEEE Trans. Inf. Theory*, 56(1):6–17, Jan. 2010.

[15] R. Gilad-Bachrach, A. Navot, and N. Tishby. An Information Theoretic Tradeoff between Complexity and Accuracy. In *Proceedings of the COLT*, pages 595–609. Springer, 2003.

[16] A. Globerson and N. Tishby. On the optimality of the Gaussian information bottleneck curve. Technical report, The Hebrew University of Jerusalem, Feb. 2004.

[17] A. Goldsmith and A. Nin. *Wireless Communications*. Cambridge University Press, 2005.

[18] Young-Han Kim. On reliability of Gaussian channels with noisy feedback. In *Proceedings of the 44th Allerton Conference on Communication, Control, and Computing*, 2006.

[19] Young-Han Kim, A. Lapidoth, and T. Weissman. Bounds on the Error Exponent of the AWGN Channel with AWGN-Corrupted Feedback. In *24th Convention of Electrical and Electronics Engineers in Israel, 2006*, pages 184–188, Nov. 2006.

[20] Young-Han Kim, A. Lapidoth, and T. Weissman. The Gaussian Channel with Noisy Feedback. In *IEEE International Symposium on Information Theory (ISIT 2007)*, pages 1416–1420, June 2007.

[21] A. Kramer. Improving communication reliability by use of an intermittent feedback channel. *IEEE Trans. Inf. Theory*, 15(1):52–60, Jan. 1969.

[22] N.C. Martins and T. Weissman. Coding for Additive White Noise Channels With Feedback Corrupted by Quantization or Bounded Noise. *IEEE Trans. Inf. Theory*, 54(9):4274–4282, Sept. 2008.

[23] MathWorks. eps. http://www.mathworks.de/de/help/matlab/ref/eps.html.

[24] M. Meidlinger, A. Winkelbauer, and G. Matz. On the relation between the Gaussian information bottleneck and MSE-optimal rate-distortion quantization. In *Proc. IEEE Workshop on Statistical Signal Processing (SSP 2014)*, June 2014.

[25] R. Mirghaderi, A. Goldsmith, and T. Weissman. Achievable Error Exponents in the Gaussian Channel With Rate-Limited Feedback. *IEEE Trans. Inf. Theory*, 59(12):8144–8156, Dec. 2013.

[26] M.S. Pinsker. The probability of error in block transmission in a memoryless Gaussian channel with feedback. *Problemy Peredachi Informatsii*, 4(4):3–19, 1968.

[27] J. Schalkwijk. A coding scheme for additive noise channels with feedback–ii: Band-limited signals. *IEEE Trans. Inf. Theory*, 12(2):183–189, 1966.

[28] J. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback–i: No bandwidth constraint. *IEEE Trans. Inf. Theory*, 12(2):172–182, 1966.

[29] C. E. Shannon and W. W. Weaver. The Mathematical Theory of Communication,. *University of Illinois Press, Urbana, IL*, 1949.

[30] N. Tishby, F. Pereira, and W. Bialek. The Information Bottleneck Method. pages 368–377, 1999.

[31] A. Winkelbauer, S. Farthofer, and G. Matz. The rate-information trade-off for Gaussian vector channels. In *Proc. IEEE Int. Symp. Information Theory (ISIT 2014)*, 2014.

[32] A. Winkelbauer and G. Matz. Rate-information-optimal Gaussian channel output compression. In *Proc. 48th Annual Conference on Information Sciences and Systems (CISS 2014)*, March 2014.