Graz University of Technology

Master Project

Reconstruction of Nonlinearly Distorted Signals

conducted at the
Signal Processing and Speech Communications Laboratory
Graz University of Technology, Austria

by
Stefan Wakolbinger

Supervisors:
Bernhard Geiger

Assessors:
Bernhard Geiger

Graz, August 2, 2011

Abstract

If a random variable is passed through a nonlinear system, where multiple input values are mapped onto the same output value, information is lost. This means that by observing the output of the system, the input values cannot be reconstructed correctly with certainty. Analytic results for calculating the information loss in terms of the conditional entropy exist [3] [4], but are not always computable since logarithms of sums are involved. Similar to Fano's Inequality [1], the information loss can be bounded in terms of the probability of making a reconstruction error when using a Maximum a Posteriori (MAP) estimator. If the analytic result for the information loss exists, these bounds can be used conversely to bound the error probability of the MAP estimator.

In this work, two different scenarios are explored: In the first scenario, independent, identically distributed (i.i.d.) samples of a continuous random variable are passed through the system. The information loss is bounded by using the error probability of a sample-by-sample estimator, which attempts to reconstruct every input value x_n by just observing its corresponding output value y_n, therefore $\hat{x}_n = f(y_n)$, where $\hat{x}_n$ is the estimated input value.

In a second scenario, the input samples are not independent from each other, but form a Markov process. Because of this dependency it makes sense to observe more than just one output value, since also past output values contain information about the current input value. Therefore, the information loss can be bounded using the error probability of an estimator with memory. In this work, only estimators using the past two output values are examined ($\hat{x}_n = f(y_n, y_{n-1})$). Three different examples are presented which illustrate the application of the analytic results.

Contents

1 Introduction

The entropy H of a random variable X over the alphabet $1, ..., M$ with probability distribution $\{p(x)\}_{x=1}^{M}$ is measured in bits and defined as

$$H(X) = -\sum_{x=1}^{M} p(x)\, \log_2(p(x)) \tag{1.1}$$

The logarithm is taken to the base of two throughout this work, therefore the notation $log(x)$ is used instead of $log_2(x)$ in the remainder of this paper. The entropy is a measure for the complexity of a random variable. The higher the entropy, the harder it is to predict its value. In this work, the conditional entropy $H(X|Y)$ between the input random variable X and the random variable Y at the output of the system is examined. This measure is called equivocation and describes the uncertainty about the random variable X if the random variable Y is given. In other words, it describes how much additional information is required to determine the value at the input of the system if the corresponding value at the output is known, or equivalently, how much information is lost by passing through the nonlinear system. Similar to the entropy of a single random variable, the equivocation is defined as [2]

$$H(X|Y) = -\int_{y} p(x|y)\, \log(p(x|y))\, dP(y). \tag{1.2}$$

In general, the systems considered in this work are described by nonlinear, deterministic functions $Y = g(X)$. It is possible that not every input value is mapped bijectively, which means that there can be several different input values which map onto the same output value. Therefore, information is lost by passing a random variable through the system, since one cannot determine with certainty which input value was mapped onto the value observed at the output.

The remainder of this paper is structured as follows: In Section 2, the problem is described in more detail, specifying the properties of both the input process and the system. Furthermore, an analytic expression derived in [3] for computing the information loss in a system with an i.i.d. input process is presented. In Section 3, different estimators are presented in detail, which were derived in previous work [3] [4]. Results exist both for a sample-by-sample estimation with an i.i.d. input process and a scenario where the input of the system is a Markov process and the input values are estimated based on more than just the current output value. Section 4 gives a compact overview over computing bounds of the information loss, which is also carried out in detail in [3] and [4]. The analytic results are illustrated by using a set of examples in Section 5. The results obtained in these examples are finally discussed in Section 6.

2 Information Loss in Static Nonlinearities

2.1 Problem Statement

A continuous random variable is passed through a nonlinear system, which is described by a surjective function $g(x)$ of the following type:

Definition 1. Let $g : \mathcal{X} \to \mathcal{Y},\ \mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}$, be a bounded, measurable function which is piecewise strictly monotone on L disjoint, proper intervals $\mathcal{X}_l,\ l = 1, ..., L$:

$$g(x) = \begin{cases} g_1(x) & x \in \mathcal{X}_1 \\ g_2(x) & x \in \mathcal{X}_2 \\ \vdots & \\ g_L(x) & x \in \mathcal{X}_L \end{cases} \tag{2.1}$$

This means, that the surjective function $g(x)$ is divided into L bijective functions $g_l(x)$. The corresponding intervals $\mathcal{X}_l$ unite to $\mathcal{X}$:

$$\bigcup_{l=1}^{L} \mathcal{X}_l = \mathcal{X} \tag{2.2}$$

The intervals $\mathcal{Y}_l$ also unite to $\mathcal{Y}$, but overlap if the system is non-injective, i.e. if more than one input value is mapped onto the same output value.

In this work two different scenarios are considered. In the first scenario, the input of the system is described by a continuous random variable X, where $f_X(x)$ is positive on $\mathcal{X}$ and zero elsewhere. The input samples are independent and identically distributed and have a continuous cumulative distribution function. According to [3], the marginal probability density function (PDF) of Y is given as

$$f_Y(y) = \sum_{i \in \mathbb{I}(y)} \frac{f_X(x_i)}{|g'(x_i)|} \tag{2.3}$$

where x_i are the roots of y and $\mathbb{I}(y)$ is the index set pointing to these roots.

In the second scenario, the input samples are not independent from each other, but taken from an ergodic, discrete-time, continuous process $\mathbf{X}$ with Markov property:

$$f_{X_n|X_0^{n-1}}(x_0^n) = f_{X_n|X_{n-1}}(x_{n-1}, x_n) \tag{2.4}$$

Furthermore, this process is time-homogeneous, i.e.:

$$f_{X_n|X_{n-1}}(x_{n-1}, x_n) = f_{X_1|X_0}(x_0, x_1) = f_M(x_1|x_0) \tag{2.5}$$

Again, the support of the marginal distribution is identical to $\mathcal{X}$. The joint PDF of the sequence

x_0^n can then be obtained by

$$f_{X_0^n}(x_0^n) = f_X(x_0) \prod_{k=1}^{n} f_M(x_n|x_{n-1}). \tag{2.6}$$

The n-th order joint density $f_{Y_0^n}(y_0^n)$ can be computed by [4]

$$f_{Y_0^n}(y_0^n) = \sum_{i_0^n \in \mathbb{I}(y_0^n)} \frac{f_X(x_0^{(i_0)})}{|g'(x_0^{(i_0)})|} \prod_{k=1}^{n} \frac{f_M(x_k^{(i_k)}|x_{k-1}^{(i_{k-1})})}{|g'(x_k^{(i_k)})|}. \tag{2.7}$$

2.2 Analytical Computation of Information Loss

If a surjective function $g(\cdot)$ is applied to the input samples x, information is lost, since the mapping is not unique and a single output value y can have more than one root. This loss of information can be measured in terms of the conditional entropy $H(X|Y)$, which is the entropy of the input random variable X given the output Y.

An analytical result for this conditional entropy for the first scenario (input samples are i.i.d.) was derived in [3]:

$$H(X|Y) = \int_{\mathcal{X}} f_X(x) log \left(\frac{\sum_{i \in \mathbb{I}(g(x))} \frac{f_X(x_i)}{|g'(x_i)|}}{\frac{f_X(x)}{|g'(x)|}} \right) dx \tag{2.8}$$

For the second scenario, where the input to the system is taken from a Markov process, only n-th order bounds on the information loss can be computed, which are presented in Section 4. If certain conditions are fulfilled, these bounds are tight and therefore equal to the analytic result.

3 Reconstruction of Input Samples

If information is lost by passing through a nonlinear system, the question arises how the input samples can be reconstructed with a minimum error probability by just observing the output of the system.

A maximum a posteriori probability estimator (MAP estimator) minimizes the probability of a reconstruction error by maximizing the conditional probability of the input given the output:

$$\hat{x}(y) = \underset{x}{\operatorname{argmax}} \left\{ f_{X|Y}(x, y) \right\} \tag{3.1}$$

As carried out in [3] for the estimation of an input sample x_n based on one output sample y_n, the MAP estimator chooses the most likely root of the output sample y_n, i.e.,

$$\hat{x}(y) = g_k^{-1}(y) \tag{3.2}$$

where

$$k = \underset{i \,\in\, \mathbb{I}(y)}{\operatorname{argmax}} \left\{ \frac{f_X(x_i)}{|g'(x_i)|} \right\}. \tag{3.3}$$

The probability of a reconstruction error using this MAP estimator can be computed as follows [3]:

$$P_e = 1 - \int_{\mathcal{Y}} \max_{i \,\in\, \mathbb{I}(y)} \left\{ \frac{f_X(x_i)}{|g'(x_i)|} \right\} \, dy \tag{3.4}$$

Since the MAP estimator can be unable or very hard to compute analytically, a less difficult, but suboptimal estimator was derived in [3], which chooses the root which lies in the interval with the highest probability mass:

$$\hat{x}(y) = \begin{cases} g^{-1}(y) & \text{if } y \in \mathcal{Y}_b \\ g_k^{-1}(y) & \text{if } k \in \mathbb{I}(y),\ y \notin \mathcal{Y}_b \\ x:\ x \in \mathcal{X}_k & \text{if } k \notin \mathbb{I}(y),\ y \notin \mathcal{Y}_b \end{cases} \tag{3.5}$$

where

$$k = \underset{i}{\operatorname{argmax}} \int_{\mathcal{X}_i \cup \mathcal{X}_b} f_X(x) \, dx \tag{3.6}$$

The error probability for this estimator is given by:

$$\hat{P}_e = 1 - P_b - \int_{\mathcal{X}_k \setminus \mathcal{X}_b} f_X(x) \, dx \tag{3.7}$$

where

$$P_b = \int\limits_{\mathcal{X}_b} f_X(x)\ dx = \int\limits_{\mathcal{Y}_b} f_Y(y)\ dy \tag{3.8}$$

The value of P_b is the probability mass which is mapped bijectively. If an input value is mapped bijectively, it can be reconstructed by computing the unique inverse of $g(\cdot)$ for the corresponding output value. For these input values the probability of a reconstruction error is zero, therefore the probability mass containing these values has to be subtracted.

The probability of an estimation error of this suboptimal estimator can be bounded as follows [3]:

$$\hat{P}_e \leq \frac{L-1}{L} \tag{3.9}$$

If the input samples are not independent from each other, it makes sense to consider also past output values in order to estimate the current input value. If the input is taken from a Markov process (Section 2.1), the MAP estimator computes the most probable input value as [4]

$$\hat{x}_n(y_0^n) = \arg \max_{x_n^{(i_n)} : i_n \in \mathbb{I}(y_n)} \left\{ \sum_{i_0^{n-1} \in \mathbb{I}(y_0^{n-1})} \frac{f_X(x_0^{(i_0)})}{|g'(x_0^{(i_0)})|} \prod_{k=1}^{n} \frac{f_M(x_k^{(i_k)} | x_{k-1}^{(i_{k-1})})}{|g'(x_k^{(i_k)})|} \right\}. \tag{3.10}$$

The resulting input value $\hat{x}_n$ is estimated based on multiple output values y_0^n. The computational difficulty of this estimator increases with the number of output values that are taken into account.

4 Bounding Information Loss

4.1 First Scenario: $\hat{\mathbf{x}}_\mathbf{n} = \mathbf{f}(\mathbf{y}_\mathbf{n})$

If the input values of the nonlinear system are i.i.d., the information loss can be expressed using Equation 2.8. Unfortunately, the analytic result is often impossible or very difficult to compute, since the logarithm of a sum is involved.
Thus, upper and lower bounds for the information loss have been derived in [3]. These bounds will be shortly presented in this Section.

If the error probability of the MAP estimator can be determined according to Equation 3.4, the following upper bounds of the information loss can be computed:

$$\begin{aligned} H(X|Y) &\leq \min\{1 - P_b, H_2(P_e)\} + P_e \log\left(\max_{y\in\mathcal{Y}}\{|\mathbb{I}(y)|\} - 1\right) && (4.1)\\ &\leq \min\{1 - P_b, H_2(P_e)\} + P_e \log(L-1) && (4.2)\end{aligned}$$

Furthermore, a lower bound based on P_e can be computed as follows:

$$H(X|Y) \geq \Phi(P_e) \tag{4.3}$$

where

$$\Phi(x) = \left(x - \frac{i-1}{i}\right)(i+1)\ i \log\left(1 + \frac{1}{i}\right) + \log(i) \tag{4.4}$$
$$\text{for } \frac{i-1}{i} \leq x \leq \frac{i}{i+1}$$

The Φ-function is a continuous function with a piecewise continuous derivative, composed of linear segments. More information about this function can be found in [2].

If the suboptimal estimator (Equation 3.5) is used instead of the MAP estimator, the following upper bounds can be computed:

$$H(X|Y) \leq 1 - P_b + \hat{P}_e \log(L-1) \tag{4.5}$$

$$H(X|Y) \leq \hat{P}_e \log(L-1) + \frac{\hat{P}_e}{L} \log\left(\frac{L}{L-1}\right) + (1 + \hat{P}_e)H_2(\hat{P}_e) \tag{4.6}$$

$$H(X|Y) \leq \hat{P}_e \log(L) + H_2(\hat{P}_e) \tag{4.7}$$

In addition, the following upper bounds, which only depend on the probability distributions and the intervals of $g(\cdot)$, have been derived in [3]:

$$H(X|Y) \leq \int_{\mathcal{Y}} f_Y(y)\ \log(|\mathbb{I}(y)|)\ dy \tag{4.8}$$

$$\leq log\left(\sum_{l=1}^{L}\int_{\mathcal{Y}_l} f_Y(y)\ dy\right) \tag{4.9}$$

$$\leq \log(L) \tag{4.10}$$

4.2 Second Scenario: $\hat{\mathbf{x}}_\mathbf{n} = \mathbf{f}(\mathbf{y}_\mathbf{0}^\mathbf{n})$

If the input values are taken from a Markov process as described in Section 2, the n-th order bounds of the information loss can be computed as follows [4]:

$$\overline{H}(\mathbf{X}|\mathbf{Y}) \geq \iiint_{\mathcal{X}^{n+1}} f_{X_0^n}(x_0^n)\ \log\left(\frac{|g'(x_n)| f_{Y_n|Y_0^{n-1}}(g(x_0^n))}{f_M(x_n|x_{n-1})}\right)\ dx_0^n \tag{4.11}$$

$$\overline{H}(\mathbf{X}|\mathbf{Y}) \leq \iiint_{\mathcal{X}^{n+1}} f_{X_0^n}(x_0^n)\ \log\left(\frac{|g'(x_n)| f_{Y_n|Y_1^{n-1}X_0}(x_0, g(x_1^n))}{f_M(x_n|x_{n-1})}\right)\ dx_0^n \tag{4.12}$$

For the first order bounds, these expressions simplify to [4]:

$$\overline{H}(\mathbf{X}|\mathbf{Y}) \geq \iint_{\mathcal{X}^2} f_M(x_1|x_0) f_X(x_0)\ \log\left(\frac{|g'(x_1)| \sum\limits_{i_0^1 \in \mathbb{I}(g(x_0^1))} \frac{f_M(x_1^{(i_1)}|x_0^{(i_0)}) f_X(x_0^{(i_0)})}{|g'(x_1^{i_1}) g'(x_0^{(i_0)})|}}{f_M(x_1|x_0) \sum\limits_{i_0 \in \mathbb{I}(g(x_0))} \frac{f_X(x_0^{(i_0)})}{|g'(x_0^{(i_0)})|}}\right)\ dx_1 dx_0 \tag{4.13}$$

$$\overline{H}(\mathbf{X}|\mathbf{Y}) \leq \iint_{\mathcal{X}^2} f_M(x_1|x_0) f_X(x_0)\ \log\left(\frac{|g'(x_1)| \sum\limits_{i_1 \in \mathbb{I}(g(x_1))} \frac{f_M(x_1^{(i_1)}|x_0)}{|g'(x_1^{i_1})|}}{f_M(x_1|x_0)}\right)\ dx_1 dx_0 \tag{4.14}$$

If both the input and the output processes are first-order Markov processes, these bounds are equal and can therefore be used to compute the analytic result of the information loss. The following conditions have to be fulfilled for a first-order Markov output process [4]:

$$\frac{f_X(x_0^{(i_0)})}{|g'(x_0^{(i_0)})|} = \frac{f_X(x_0)}{|g'(x_0)|} \tag{4.15}$$

$$\sum_{i_0^1 \in \mathbb{I}(g(x_0^1))} \frac{f_M(x_1^{(i_1)}|x_0^{(i_0)})}{|g'(x_1^{(i_1)})|} = |\mathbb{I}(g(x_0))| \sum_{i_1 \in \mathbb{I}(g(x_1))} \frac{f_M(x_1^{(i_1)}|x_0)}{|g'(x_1^{(i_1)})|} \tag{4.16}$$

5 Examples

In this Section three examples will be presented, which illustrate the results of the previous Sections. In the first two examples, the input samples are independent to each other, and therefore the reconstruction of a single output sample is based only on the according input sample. In the third example, estimation based on the last two output samples is shown for an autoregressive ($AR(1)$) process as input to the system.

All simulations were carried out in MATLAB. In order to calculate the mutual information between two vectors, a library from Rudy Moddemeijer [6] was used.

5.1 Example 1 - Third Order Polynomial, Normal Distribution

First, consider a Gaussian random variable $X \sim \mathcal{N}(0, \sigma^2)$. Let this random variable be the input to a nonlinear system described by a surjective function

$$g(x) = x^3 - 100x \tag{5.1}$$

which is depicted in Figure 5.1.

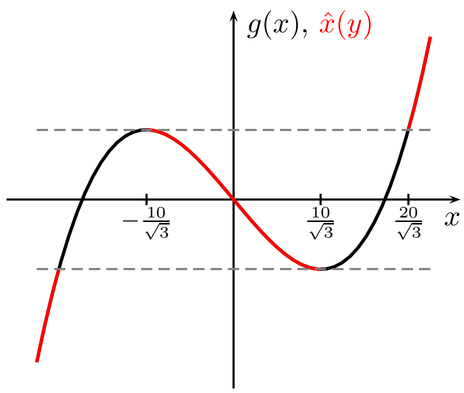

Figure 5.1: Third-order Polynomial of Example 1 and Reconstruction Domains of the Sub-Optimal Estimator

The calculation of the analytic result for the information loss $H(X|Y)$ using Equation 2.8 is prevented by the logarithm of sums. Therefore, one can only bound the information loss using the results of Section 4.

By computing the extrema of $g(x)$, three piecewise monotone intervals can be defined, which are $\mathcal{X}_1 = (-\infty, -\frac{10}{\sqrt{3}}]$, $\mathcal{X}_2 = (-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}})$, and $\mathcal{X}_3 = [\frac{10}{\sqrt{3}}, \infty)$. Therefore, the parameter L is 3.

Further, the domain which is mapped bijectively is

$$\mathcal{X}_b = \left(-\infty, -\frac{20}{\sqrt{3}}\right] \cup \left[\frac{20}{\sqrt{3}}, \infty\right). \tag{5.2}$$

Thus, the probability mass P_b which is mapped bijectively can be computed as

$$P_b = 2F_X\left(-\frac{20}{\sqrt{3}}\right) = 2Q\left(\frac{20}{\sqrt{3}\sigma}\right). \tag{5.3}$$

Using Equation 4.8 and Equation 3.8 leads to the following upper bound for $H(X|Y)$:

$$H(X|Y) \leq \int_{\mathcal{Y}} f_Y(y)\, \log(|\mathbb{I}(y)|)\ dy \tag{5.4}$$

$$= \int_{\mathcal{Y}\setminus\mathcal{Y}_b} f_Y(y)\, \log(3)\ dy + \int_{\mathcal{Y}_b} f_Y(y)\, \log(1)\ dy \tag{5.5}$$

$$= (1 - P_b)\, \log(3) \tag{5.6}$$

Since the output values y which lie in the bijectively mapped domain $\mathcal{Y}_b$ only have one possible root, $|\mathbb{I}(y)|$ is one for these values and the second integral vanishes. For all other values of y, $|\mathbb{I}(y)|$ is 3, since there are three possible roots which map to the output value.

In order to compute some other bounds, first the error probability of the suboptimal estimator presented in Section 3 is determined. As depicted in Equation 3.5, the suboptimal estimator decides for the root which lies in the interval containing the highest probability mass.
If $y \in Y_b$, only one root exists and therefore the choice is unambiguous. If this is not the case, each value y has three possible roots, one in each of the intervals $(-\frac{20}{\sqrt{3}}, -\frac{10}{\sqrt{3}}]$, $(-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}})$ and $[\frac{10}{\sqrt{3}}, \frac{20}{\sqrt{3}})$.

Since the maximum of the zero-mean Gaussian PDF lies in the second interval, and the width of this interval is double the width of the other intervals, most probability mass lies in the second interval, independent of the parameter σ. Thus, the suboptimal estimator always decides for the root which lies in the interval $(-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}})$. The estimated input values $\hat{x}$ are depicted in Figure 5.1.

The error probability of this estimator can be computed according to Equation 3.7, where $\mathcal{X}_k = (-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}})$:

$$\hat{P}_e = 1 - P_b - \int_{\mathcal{X}_k\setminus\mathcal{X}_b} f_X(x)\ dx = 1 - P_b - \int_{-\frac{10}{\sqrt{3}}}^{\frac{10}{\sqrt{3}}} f_X(x)\ dx = 2Q\left(\frac{10}{\sqrt{3}\sigma}\right) - 2Q\left(\frac{20}{\sqrt{3}\sigma}\right) \tag{5.7}$$

Using the values of P_b, $\hat{P}_e$ and L, the bounds of Equations 4.5, 4.6 and 4.7 can be computed as follows:

$$H(X|Y) \leq 1 - P_b + \hat{P}_e\, \log(L-1) = 1 - P_b + \hat{P}_e \tag{5.8}$$

$$H(X|Y) \leq \hat{P}_e\, \log(L-1) + \frac{\hat{P}_e}{L}\, \log\left(\frac{L}{L-1}\right) + (1+\hat{P}_e)H_2(\hat{P}_e) \tag{5.9}$$

$$= \hat{P}_e + \frac{\hat{P}_e}{3}\, \log\left(\frac{3}{2}\right) + (1+\hat{P}_e)H_2(\hat{P}_e) \tag{5.10}$$

$$H(X|Y) \leq \hat{P}_e \log(L) + H_2(\hat{P}_e) = \hat{P}_e \log(3) + H_2(\hat{P}_e) \tag{5.11}$$

In order to compute the bounds based on MAP estimation, the error probability of this estimator has to be determined. According to Equation 5.41, the root with the highest value of $\frac{f_x(x_i)}{|g'(x_i)|}$ is chosen.

For this specific example, the denominator of this term is

$$|g'(x)| = \begin{cases} 100 - 3x^2 & \text{if } -\frac{10}{\sqrt{3}} \leq x \leq \frac{10}{\sqrt{3}} \\ 3x^2 - 100 & \text{elsewhere} \end{cases} . \tag{5.12}$$

Since the input PDF is Gaussian with zero mean, $f_X(x_i)$ is always the highest for the root in the interval $(-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}})$. Since also $|g'(x)|$ grows very fast outside of this interval, one can suspect that the MAP estimator also always chooses the root which lies in the interval including the point of origin.
This assumption turns out to be true, as the following proof shows:

Proof. Since the input PDF is Gaussian with zero-mean and arbitrary variance σ^2, the assumption has to hold for every value of σ. If the variance of the input PDF is small, more probability mass lies in the interval around the origin, and the MAP estimator is more likely to choose the root which lies in this interval. Therefore, a Gaussian PDF with infinitely large variance is assumed for this proof, which leads to a flat distribution. If the proof holds for this distribution, it also holds for smaller values of σ^2.
If the variance is infinitely large, the criterion for choosing the most probable root becomes

$$k = \underset{i \,\in\, \mathbb{I}(y)}{\text{argmax}} \left\{ \frac{f_X(x_i)}{|g'(x_i)|} \right\} = \underset{i \,\in\, \mathbb{I}(y)}{\text{argmax}} \left\{ \frac{1}{|g'(x_i)|} \right\} \tag{5.13}$$

since $f_X(x_i)$ is constant.

This criterion is equivalent to

$$k = \underset{i \,\in\, \mathbb{I}(y)}{\text{argmin}} \left\{ |g'(x_i)| \right\} . \tag{5.14}$$

Thus, it has to be shown that $g(x_i)$ always has the lowest absolute derivative for the root in the interval $(-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}})$.

The roots of a given output value y of the third order polynomial can be determined by using Cardano's method [5]. The three resulting roots $x_i, i = 1, ..., 3$ of any output value $y \in \mathcal{Y} \backslash \mathcal{Y}_b$ result to be

$$x_i = 2\sqrt{\frac{100}{3}} \cos\left(\frac{\varphi + 2\pi i}{3}\right) \tag{5.15}$$

with

$$\varphi = \arccos\left(\frac{y}{2\sqrt{\frac{10^6}{27}}}\right) . \tag{5.16}$$

There is one root in every interval of $\mathcal{X} \backslash \mathcal{X}_b$, i.e. $x_i \in \mathcal{X}_i$.

Due to symmetry, it is sufficient to show that

$$|g'(x_3)| \geq |g'(x_2)| \; \forall \; y \in \mathcal{Y} \backslash \mathcal{Y}_b. \tag{5.17}$$

By inspecting the function $g(x)$, it is obvious that then also $|g'(x_1)| \geq |g'(x_2)| \; \forall \; y \in \mathcal{Y} \backslash \mathcal{Y}_b$.

Using 5.12 in 5.17 yields

$$3x_3^2 - 100 \geq 100 - 3x_2^2 \tag{5.18}$$

which simplifies to

$$x_3^2 + x_2^2 \geq \frac{200}{3} \; . \tag{5.19}$$

Inserting the roots (5.15) into this inequality yields

$$\left(2\sqrt{\frac{100}{3}}\cos\left(\frac{\varphi + 6\pi}{3}\right)\right)^2 + \left(2\sqrt{\frac{100}{3}}\cos\left(\frac{\varphi + 4\pi}{3}\right)\right)^2 \geq \frac{200}{3} \tag{5.20}$$

and simplifying leads to

$$\cos^2\left(\frac{\varphi}{3}\right) + \cos^2\left(\frac{\varphi}{3} + \frac{4\pi}{3}\right) \geq \frac{1}{2} \; . \tag{5.21}$$

Using the formula

$$\cos^2(x) = \frac{1}{2}\cos(2x) + \frac{1}{2} \tag{5.22}$$

yields

$$\cos\left(\frac{2\varphi}{3}\right) + \frac{1}{2}\cos\left(\frac{2\varphi}{3} + \frac{8\pi}{3}\right) \geq -1. \tag{5.23}$$

This can be further simplified by using the formula

$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \tag{5.24}$$

leading to the inequality

$$\cos\left(\frac{\pi}{3}\right)\cos\left(\frac{2\varphi}{3}\right) - \sin\left(\frac{\pi}{3}\right)\sin\left(\frac{2\varphi}{3}\right) \geq -1. \tag{5.25}$$

Using

$$\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta) \tag{5.26}$$

yields

$$\cos\left(\frac{2\varphi}{3} + \frac{\pi}{3}\right) \geq -1 \tag{5.27}$$

which completes the proof. □

This proves that the absolute value of the derivative of the root in the interval $(-\frac{10}{\sqrt{3}}, \frac{10}{\sqrt{3}})$ is always lower that the one at the roots in the other intervals. Thus, the MAP estimater chooses the same roots as the suboptimal estimator.

As a consequence, the error probability of the MAP estimator is the same as the one of the suboptimal estimator:

$$P_e = \hat{P}_e = 2Q\left(\frac{10}{\sqrt{3}\sigma}\right) - 2Q\left(\frac{20}{\sqrt{3}\sigma}\right) \tag{5.28}$$

Using Equation 4.1 yields the following upper bound for the information loss:

$$H(X|Y) \leq \min\{1-P_b, H_2(P_e)\}+P_e \log\left(\max_{y\in\mathcal{Y}}\{|\mathbb{I}(y)|\} - 1\right) = \min\{1-P_b, H_2(P_e)\}+P_e \tag{5.29}$$

A lower bound based on P_e can be computed using Equation 4.3.

Figure 5.2 shows all the computed bounds of $H(X|Y)$ together with the simulated result in MATLAB.

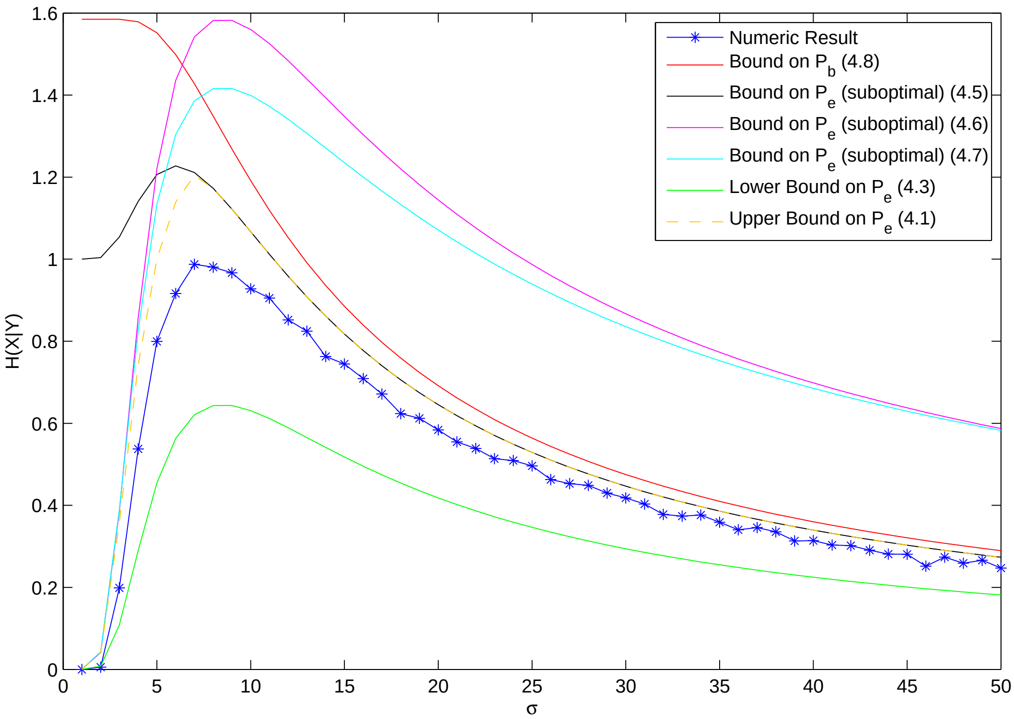

Figure 5.2: Bounds of $H(X|Y)$ and Simulated Result

Inversely, if the information loss is known, the error probability of the MAP estimator can be bounded using Equations 4.1 and 4.3. Because these inequalities are difficult to solve for P_e, a MATLAB script was used, which computes the result iteratively.

Since the analytic result for $H(X|Y)$ cannot be computed in this example, the simulated value was used. Based on these values, an upper and lower bound for the estimation error of the MAP estimator was computed in order to double-check the results. The MAP estimator was also simulated in MATLAB and its error probability computed by comparing the real input values with the estimated ones. The results are shown in Figure 5.3.

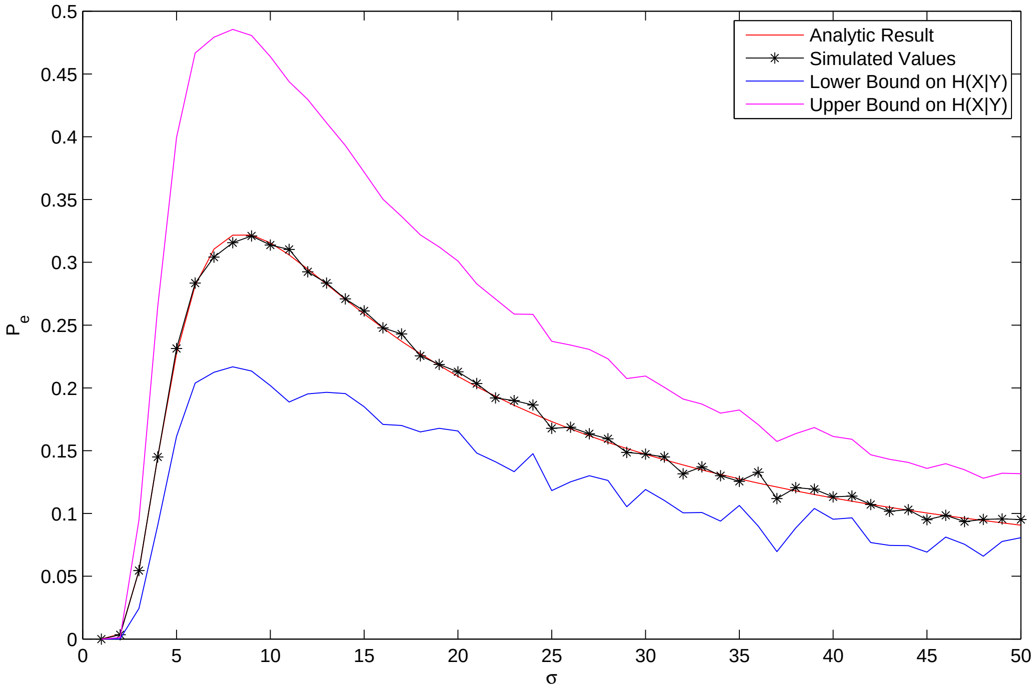

Figure 5.3: Bounds of P_e

5.2 Example 2 - Sinusoidal Function, Shifted Triangular Distribution

Next, consider a sinusoidal function

$$g(x) = \sin(\pi x) \tag{5.30}$$

and a triangular PDF with mean m:

$$f_X(x) = \begin{cases} x - m + 1 & \text{if } -1 + m \leq x \leq m \\ 1 - x + m & \text{if } m \leq x \leq m + 1 \\ 0 & \text{elsewhere} \end{cases} \tag{5.31}$$

Figure 5.4 illustrates both $g(x)$ and $f_X(x)$ for $m = -\frac{1}{4}$.

Unlike in the first example, the analytical computation of the information loss using Equation 2.8 is possible here.

In the following, the calculation of the information loss for a shift of the triangular PDF in the interval $m \in [-0.5, 0]$ is depicted. The integration range has to be split into four intervals,

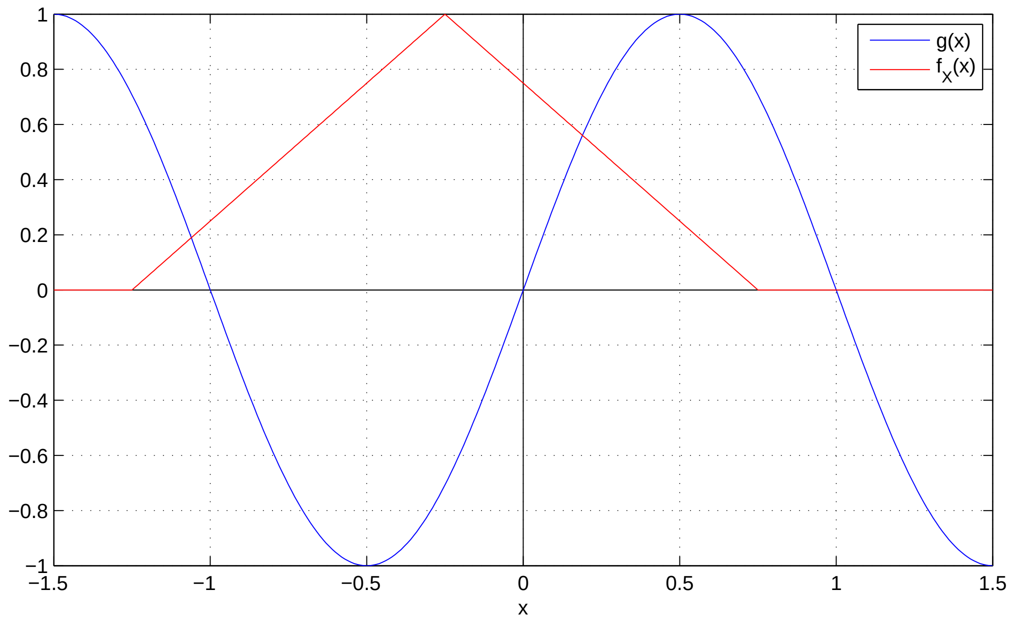

Figure 5.4: Sinusoidal Function and Triangular PDF

since the roots can either both lie on the increasing, resp. non-increasing side of the triangular distribution, or on different sides. The derivative of $g(x)$ disappears, since it is the same for every root of an output value y.

$$H(X|Y) = \int_{\mathcal{X}} f_X(x) log \left(\frac{\sum_{i \in \mathbb{I}(g(x))} \frac{f_X(x_i)}{|g'(x_i)|}}{\frac{f_X(x)}{|g'(x)|}} \right) dx \tag{5.32}$$

$$\begin{aligned} &= \int_{m-1}^{-m-1} (x-m+1) \log\left(\frac{2x+3}{x-m+1}\right) dx + \int_{-m-1}^{m} (x-m+1) \log\left(\frac{-2m+1}{x-m+1}\right) dx \\ &+ \int_{m}^{-m} (1-x+m) \log\left(\frac{-2x+1}{1-x+m}\right) dx + \int_{-m}^{m+1} (1-x+m) \log\left(\frac{2m+1}{1-x+m}\right) dx \end{aligned} \tag{5.33}$$

A closed solution for these integrals exists. The terms can be further simplified using basic algebra, which leads to the final result for $m \in [-0.5, 0]$:

$$H(X|Y) = \frac{2(2m+1)\log(e) + (3-4m-4m^2)\log(1-2m) + (1+4m+4m^2)\log(1+2m)}{4} \tag{5.34}$$

For a triangular PDF which is shifted by $m \in [0, 0.5]$ the calculations are very similar and yield

$$H(X|Y) = \frac{2(1-2m)\log(e) + (3+4m-4m^2)\log(1+2m) + (1-4m+4m^2)\log(1-2m)}{4}. \tag{5.35}$$

These results can be combined leading to a single solution for $m \in [-0.5, 0.5]$, which is

$$H(X|Y) = \frac{2(1-2|m|)\log(e) + (3+4|m|-4m^2)\log(1+2|m|)) + (1-4|m|+4m^2)\log(1-2|m|)}{4}. \tag{5.36}$$

The interval $m \in [-0.5, 0.5]$ is the basic interval of the periodic function $H(X|Y)$ depending on m. This means, that for higher, resp. lower, values of m the shift can be mapped into this basic interval in order to compute the information loss. Figure 5.5 shows the information loss in dependence of the shift m.

As a next step, the suboptimal estimator (Equation 3.5) for this example is derived.
In general, every output value y has two roots. An exception are the extrema at $y = 1$ and $y = -1$, which have only one root, and the value $y = 0$, which can have three roots if $m = 0.5 + k,\ k \in \mathbb{Z}$. Since the input random variable is continuous, these values can be neglected, because they are met exactly with probability zero.
The domain $\mathcal{X}$ can be split into subdomains $\mathcal{X}_l$, in which each input value is mapped bijectively. These subdomains always lie between a minimum and a maximum of $g(x)$. By inspecting $f_X(x)$ and $g(x)$ it is obvious that for every output value one root always lies in the subdomain $\mathcal{X}_l$ that contains the peak of the input PDF. The following proof will show that this interval always contains most of the probability mass, and therefore this root is always the one that is chosen by the suboptimal estimator.

Proof. Because of periodicity, it is sufficient to prove this statement only for any $m \in [-0.5, 0.5]$. For this range of values for m, the interval containing the peak of the PDF is $\mathcal{X}_z = [-\frac{1}{2}, \frac{1}{2}]$. The probability mass lying in this interval can be computed as follows:

$$\int\limits_{\mathcal{X}_z} f_X(x)\ dx = \left(\frac{x^2}{2} - mx + x\right)\Bigg|_{-\frac{1}{2}}^{m} + \left(-\frac{x^2}{2} + mx + x\right)\Bigg|_{m}^{\frac{1}{2}} \tag{5.37}$$

$$= \frac{3}{4} - m^2 \geq \frac{1}{2} \quad \forall\ m \in [-0.5, 0.5] \tag{5.38}$$

Thus, the interval containing the maximum of the triangular input PDF always contains at least half of the probability mass, which completes the proof. □

The error probability of the suboptimal estimator is computed as (3.7)

$$\hat{P}_e = 1 - P_b - \int\limits_{\mathcal{X}_k \setminus \mathcal{X}_b} f_X(x)\ dx = m^2 + \frac{1}{4} \quad \forall\ m \in [-0.5, 0.5] \tag{5.39}$$

since $P_b = 0$, because no probability mass is mapped bijectively.

Using the result for $\hat{P}_e$, the bounds for the information loss can be computed, as done in Example 1. The non-zero values of the input PDF are lying in three different intervals $\mathcal{X}_l$, if m is not exactly at a maximum or minimum of $g(x)$. If that is the case, only two intervals are covered by the input PDF. Thus,

$$L = \begin{cases} 2 & \text{if } m = 0.5 + k,\ k \in \mathbb{Z} \\ 3 & \text{else} \end{cases}. \tag{5.40}$$

The three upper bounds based on the values of $\hat{P}_e$, P_b and L are depicted in Figure 5.5.

The question arises, if the MAP estimator performs better than the suboptimal estimator in this example. Recall that the MAP estimator chooses the root

$$\hat{x}(y) = g_k^{-1}(y) \tag{5.41}$$

where

$$k = \underset{i \,\in\, \mathbb{I}(y)}{\operatorname{argmax}} \left\{ \frac{f_X(x_i)}{|g'(x_i)|} \right\}. \tag{5.42}$$

Because of the symmetry and periodicity of the sinusoidal function,

$$|g'(x_i)| = const. \quad \forall\, i \in \mathbb{I}(y). \tag{5.43}$$

Thus, the MAP estimator decides for the root which maximizes $f_X(x)$. The probability is the highest near the peak of the input PDF and decreases linearly with the distance to it. Therefore, because there is always exactly one root in the interval containing this peak, the MAP estimator also always decides for this root. As in Example 1, the MAP estimator is equivalent to the suboptimal estimator, which means that the error probabilities are the same:

$$P_e = \hat{P}_e = m^2 + \frac{1}{4} \quad \forall\, m \in [-0.5, 0.5] \tag{5.44}$$

The bounds of the information loss based on P_e (Equations 4.1 and 4.3) are also included in Figure 5.5.

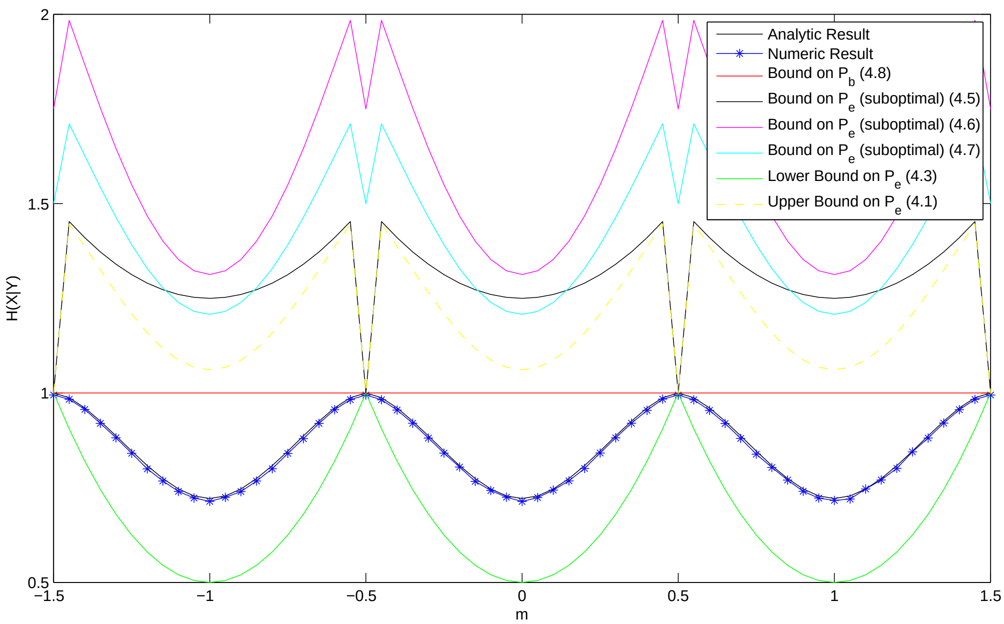

Figure 5.5: Bounds of Information Loss an Analytic Result

An analytic result for the information loss could be derived in this example, which can be used to bound the error probability of the MAP estimator as done in Example 1. The resulting bounds together with the computed and simulated values for the error probability are depicted in Figure 5.6.

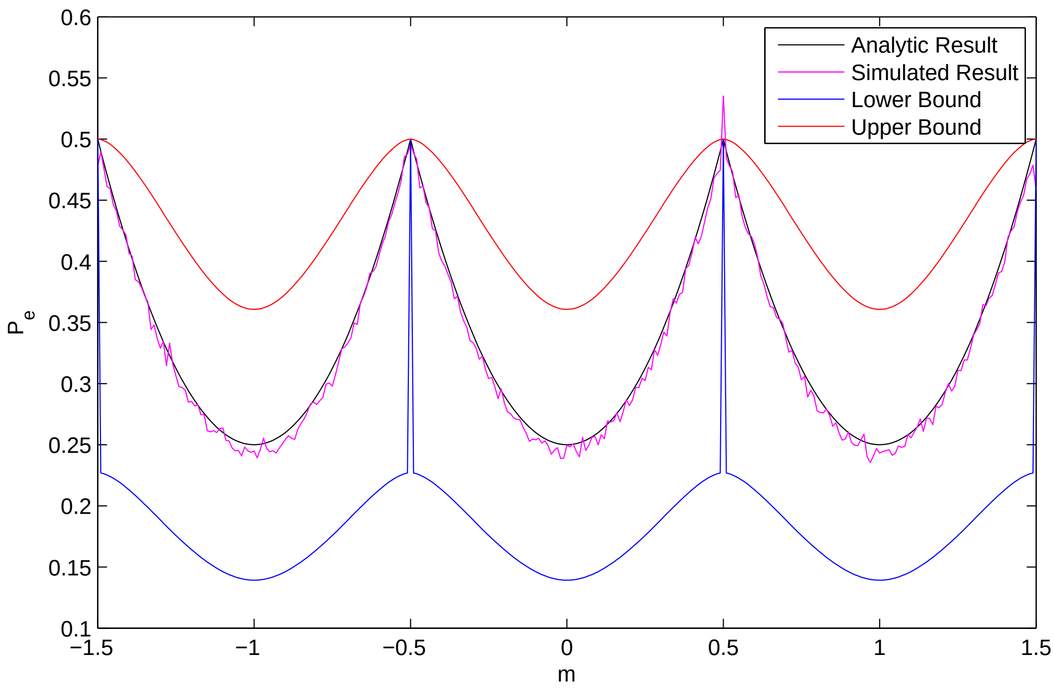

Figure 5.6: Bounds of P_e

5.3 Example 3 - AR(1) Model

Finally, consider the system depicted in Figure 5.7. The input Z of this system is uniformely distributed between $[-a, a]$. Therefore, $\mathbf{X}$ is an autoregressive model of first order (AR(1) model) and satisfies the Markov property. This means that each value of x only depends on the previous value. The Markov kernel in dependence of the filter coefficient z_0 is given as

$$f_M(x_1|x_0) = \begin{cases} \frac{1}{2a} & -a < x_1 - z_0 x_0 < a \\ 0 & \text{else} \end{cases} . \tag{5.45}$$

As long as z_0 is sufficiently close to 1, $\mathbf{X}$ can be approximated by a zero-mean Gaussian random process with variance

$$\sigma^2 = \frac{a^2}{3(1 - z_0^2)} . \tag{5.46}$$

The function $g(x)$ is a magnitude function, which is shifted along the x-axis by the parameter m:

$$g(x) = |x - m| = \begin{cases} -x + m & x \leq m \\ x - m & x > m \end{cases} \tag{5.47}$$

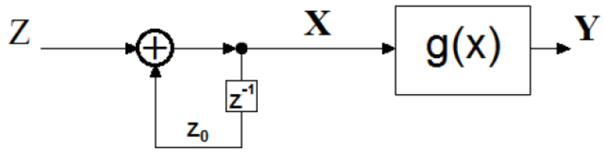

Figure 5.7: AR(1) Model

The first order upper bound can be computed using Equation 4.14:

$$\overline{H}(\mathbf{X}|\mathbf{Y}) \leq \iint\limits_{\mathcal{X}^2} f_M(x_1|x_0) f_X(x_0) \log \left(\frac{|g'(x_1)| \sum\limits_{i_1 \in \mathbb{I}(g(x_1))} \frac{f_M(x_1^{(i_1)}|x_0)}{|g'(x_1^{i_1})|}}{f_M(x_1|x_0)} \right) dx_1 dx_0 \tag{5.48}$$

$$= \iint\limits_{\mathcal{X}^2} f_M(x_1|x_0) f_X(x_0) \log \left(\frac{\sum\limits_{i_1 \in \mathbb{I}(g(x_1))} f_M(x_1^{(i_1)}|x_0)}{f_M(x_1|x_0)} \right) dx_1 dx_0 \tag{5.49}$$

since $|g'(x)| = 1 \; \forall \; x$. Next, the limits for x_0 and x_1 are inserted. While x_0 can have any value, x_1 has to be in a neighborhood $[z_0 x_0 - a, z_0 x_0 + a]$ around x_0. Using Equation 5.45 yields

$$= \int\limits_{x_0=-\infty}^{\infty} \int\limits_{x_1=z_0x_0-a}^{z_0x_0+a} \frac{1}{2a} f_X(x_0) \log \left(2a \sum_{i_1 \in \mathbb{I}(g(x_1))} f_M(x_1^{(i_1)}|x_0) \right) dx_1 dx_0 \tag{5.50}$$

The roots of $g(x_1)$ are x_1 and $2m - x_1$, thus the sum can be evaluated as

$$= \frac{1}{2a} \int\limits_{x_0=-\infty}^{\infty} f_X(x_0) \int\limits_{x_1=z_0x_0-a}^{z_0x_0+a} log\left(2a(f_M(x_1|x_0) + f_M(2m - x_1|x_0))\right) dx_1 dx_0. \tag{5.51}$$

Using again Equation 5.45 and the fact that due to symmetry

$$\int\limits_{-\infty}^{\infty} f_X(x_0) \; dx_0 = 2 \int\limits_{0}^{\infty} f_X(x_0) \; dx_0 \tag{5.52}$$

leads to

$$= \frac{1}{a} \int\limits_{x_0=0}^{\infty} f_X(x_0) \int\limits_{x_1=z_0x_0-a}^{z_0x_0+a} log\left(1 + 2a \; f_M(2m - x_1|x_0)\right) dx_1 dx_0. \tag{5.53}$$

The Markov kernel $f_M(2m - x_1|x_0)$ is non-zero only for $x_1 \in [x_0 z_0 - a, 2m - x_0 z_0 + a]$. Therefore, the equation simplifies to

$$= \frac{1}{a} \int\limits_{x_0=0}^{\frac{m+a}{z_0}} f_X(x_0) \int\limits_{x_1=x_0z_0-a}^{2m-x_0z_0+a} dx_1 dx_0 = \frac{2}{a} \int\limits_{x_0=0}^{\frac{m+a}{z_0}} f_X(x_0)(m - x_0 z_0 + a) \; dx_0. \tag{5.54}$$

This integral can be solved using integration by parts:

$$= \frac{2}{a}(m+a)\left(F_X\left(\frac{m+a}{z_0}\right) - F_X(0)\right) - \frac{2z_0}{a}\int_0^{\frac{m+a}{z_0}} f_X(x_0)x_0\, dx_0 \tag{5.55}$$

$$= \frac{2}{a}(m+a)\left(Q\left(\frac{m+a}{z_0\sigma}\right) - \frac{1}{2}\right) - \frac{2z_0\sigma}{a\sqrt{2\pi}}\left(\mathrm{e}^{-\frac{(m+a)^2}{2\sigma^2 z_0^2}}\right) \tag{5.56}$$

Using Equation 5.46, the final result for the upper bound of the information loss is the following:

$$\overline{H}(\mathbf{X}|\mathbf{Y}) \leq \frac{2}{a}(m+a)Q\left(\frac{(m+a)\sqrt{3-3z_0^2}}{z_0 a}\right) - \frac{m+a}{a} - \sqrt{\frac{2z_0^2}{3\pi(1-z_0^2)}}\left(1-\mathrm{e}^{-\frac{3(m+a)^2(1-z_0^2)}{2z_0^2a^2}}\right) \tag{5.57}$$

This bound is tight only for the case $m = 0$. In any other case, the conditions for a first-order Markov output process (Equations 4.15 and 4.16) are not fulfilled.

Since the computation of the first order lower bound of $\overline{H}(\mathbf{X}|\mathbf{Y})$, i.e. Equation 4.13, is much more difficult, it was not carried out in this work.

Figure 5.8 shows the first order upper bound of $\overline{H}(\mathbf{X}|\mathbf{Y})$ for different values of m.

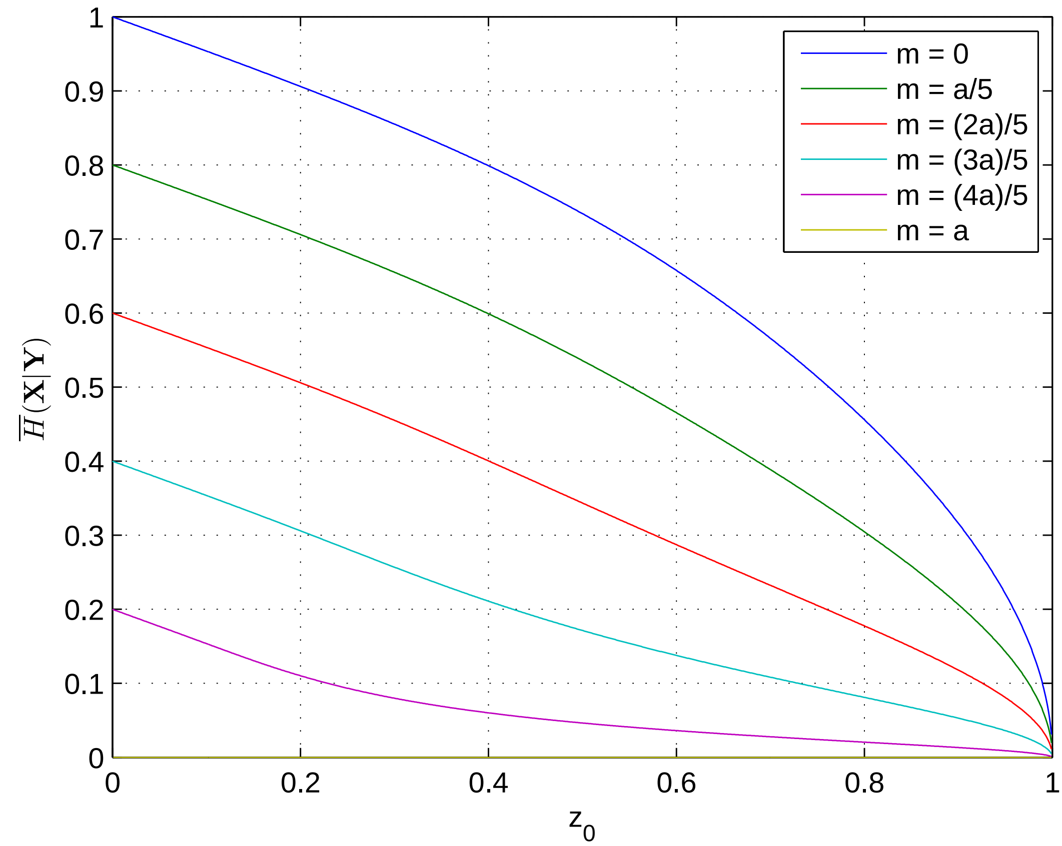

Figure 5.8: First Order Upper Bound of $\overline{H}(\mathbf{X}|\mathbf{Y})$

The MAP estimator for this example was simulated in MATLAB. Recall that the MAP estimator for the reconstruction of a single output value based on all input values is defined as

$$\hat{x}_n(y_0^n) = \arg \max_{x_n^{(i_n)}: i_n \in \mathbb{I}(y_n)} \left\{ \sum_{i_0^{n-1} \in \mathbb{I}(y_0^{n-1})} \frac{f_X(x_0^{(i_0)})}{|g'(x_0^{(i_0)})|} \prod_{k=1}^{n} \frac{f_M(x_k^{(i_k)}|x_{k-1}^{(i_{k-1})})}{|g'(x_k^{(i_k)})|} \right\}. \tag{5.58}$$

Since only the reconstruction based on the last two output values is considered in this work, the estimator simplifies to

$$\hat{x}_n(y_{n-1}^n) = \arg \max_{x_n^{(i_n)}: i_n \in \mathbb{I}(y_n)} \left\{ \sum_{i_{n-1} \in \mathbb{I}(y_{n-1})} f_X\left(x_{n-1}^{(i_{n-1})}\right) f_M\left(x_n^{(i_n)}|x_{n-1}^{(i_{n-1})}\right) \right\}. \tag{5.59}$$

Every output value y has two roots, namely

$$x^{(0)} = m - y \tag{5.60}$$

$$x^{(1)} = m + y \tag{5.61}$$

Inserting these roots into Equation 5.59 yields

$$P_{x^{(0)}} = f_X(m - y_{n-1}) f_M(m - y_n|m - y_{n-1}) + f_X(m + y_{n-1}) f_M(m - y_n|m + y_{n-1}) \tag{5.62}$$

$$P_{x^{(1)}} = f_X(m - y_{n-1}) f_M(m + y_n|m - y_{n-1}) + f_X(m + y_{n-1}) f_M(m + y_n|m + y_{n-1}) \tag{5.63}$$

$$\hat{x}_n(y_{n-1}^n) = \begin{cases} x^{(0)} & \text{if } P_{x^{(0)}} > if\ P_{x^{(1)}} \\ x^{(1)} & \text{else} \end{cases} \tag{5.64}$$

The error probability P_e of this estimator can be lower bounded by using the upper bound for $H(X|Y)$ in Equation 4.3.
As a lower bound for P_e, the error probability of a sample-by-sample MAP estimator as in the first two examples is used. Since $|g(x)'| = 1 \ \forall x$, this estimator chooses the root with the highest value of the input PDF. If m is positive, this is always the root on the left, i.e. $x^{(0)}$. The root $x^{(1)}$ is always chosen if m is negative. The error probability of the sample-by-sample estimator is

$$\tilde{P}_e = 1 - \int_{\mathcal{Y}} \max_{i \,\in\, \mathbb{I}(y)} \left\{ \frac{f_X(x_i)}{|g'(x_i)|} \right\} \, dy. \tag{5.65}$$

Thus, for this example the error probability is

$$\tilde{P}_e = Q\left(\frac{|m|}{\sigma}\right). \tag{5.66}$$

Figure 5.9 shows the simulated error probability of the MAP estimator together with the upper and lower bound for different values of the filter parameter z_0 and a mean value of $m = \frac{a}{2}$. The results for $m = \frac{a}{4}$ are depicted in Figure 5.10.

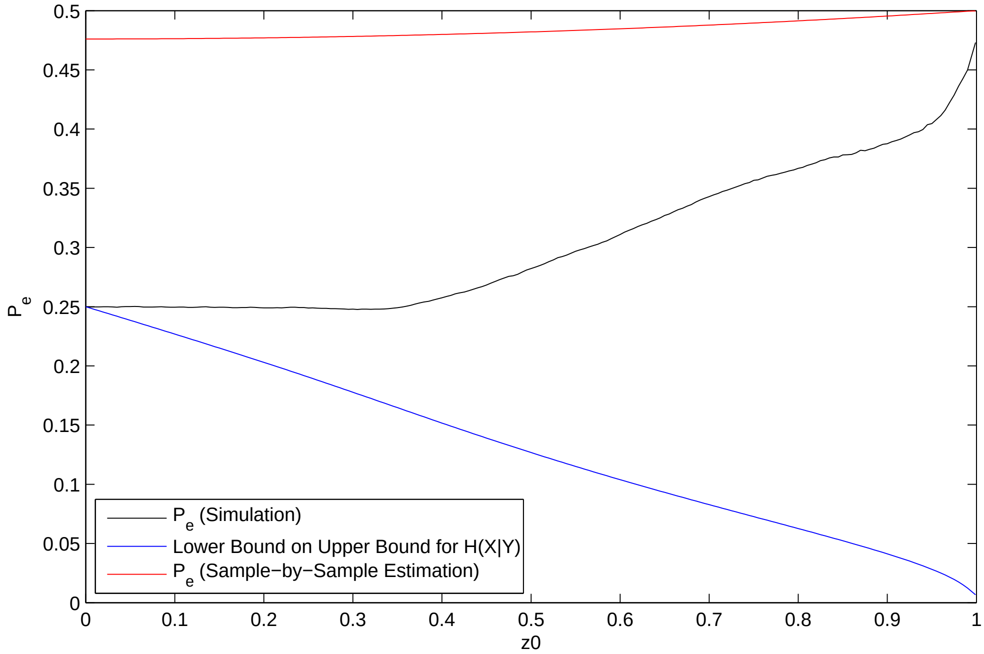

Figure 5.9: Bounds of P_e, $m = a/2$

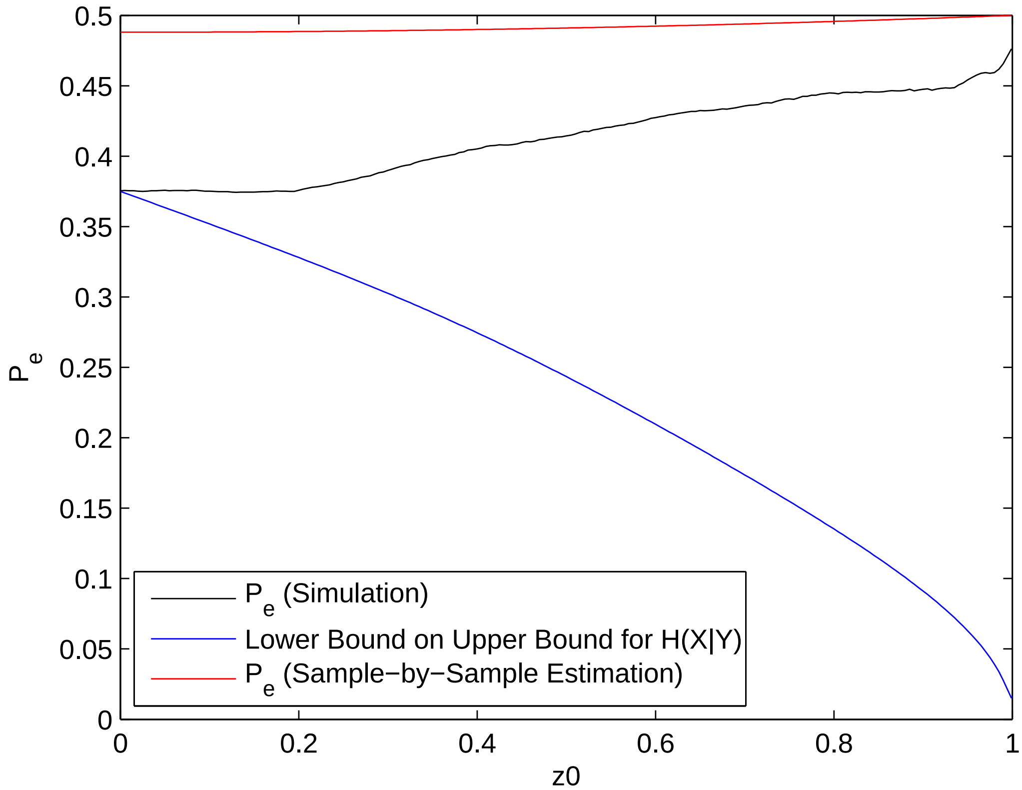

Figure 5.10: Bounds of P_e, $m = a/4$

6 Discussion and Outlook

The three examples elaborated in this work have shown that it can be difficult to compute the information loss of a RV passing through a nonlinear system analytically. Fortunately, several bounds exist, which can be computed in a much easier way. These bounds are sometimes not very tight, but can often limit the possible range quite well. Also the MAP estimator can be difficult to compute in some cases, but the examples have shown that similar, easier but suboptimal estimators can achieve the same error probabilities if certain conditions on the symmetry of the function are met. In the remainder of this Section the results of each of the examples are discussed.

In Example 1, it can be seen in Figure 5.2 that the information loss first increases with the variance of the input PDF, but after reaching its peak it decreases with higher values of σ. If the variance is very low, the vast majority of the probability mass is located in the interval around the origin, which means that only very few output values originate from the outer intervals and little information is lost. If the variance rises, more and more probability mass is located in the other intervals, and the error probability of the estimator grows. After reaching a certain value of σ, the information loss decreases again, because for very large variances more and more probability mass is located in the intervals which are mapped bijectively, which means that many input values can be reconstructed without loss of information.
It is also notable that the bounds based on the reconstruction error are quite good, especially for large values of σ. Therefore, the information loss can be estimated quite accurately without knowing its exact value.

The information loss in Example 2, depicted in Figure 5.5, is exactly 1 Bit, if the peak of the triangular input PDF is located at an extremum of the sinusoidal function. In that case, one can only guess from which of the two roots the output values origin, due to the symmetry of both the PDF and the sinusoidal function. The least information is lost if the PDF is shifted to a zero of the function. In that case, one root is always quite a lot more probable than the other, since the other one is much further away from the peak of the PDF.
Unfortunately, the bounds of $H(X|Y)$ are not very tight in this example. An exception is the case where the input PDF is located at an extremum of $g(x)$. Here, both a tight upper and lower bound exist, and therefore give an exact result for the information loss. In any other case, the upper bounds are useless, since all of them only indicate that the information loss is below 1 Bit. Since there are only two roots associated with any output value, this is an obvious result, since guessing blindly between two values is equivalent to an information loss of 1 Bit.
Since an analytic result for $H(X|Y)$ is present, it can be used to bound the error probability of the MAP estimator. The estimator was also simulated in MATLAB in order to evaluate the bounds. Again, both a tight upper and lower bound exist if the input PDF is located at an extremum.
Both in Example 1 and Example 2 the suboptimal estimator performed equivalently to the MAP estimator, but was easier to derive.

Considering Example 3, it is shown that the information loss decreases with the shift parameter m. This is quite intuitive, since the input PDF is symmetric around the origin. If $g(x)$ is shifted away from the origin, the probabilities of the roots differ more and more, and estimation is

easier. The most notable fact in the third example is that the MAP estimator $\hat{x}_n = f(y_n, y_{n-1})$ performes worse with increasing filter parameter z_0, even though the information loss decreases with increasing z_0. As a suggestion for future work, it would be interesting if taking into account more than two output values affects this behavior. Also the derivation of an analytic result for the error probability of the MAP estimator would be desirable.
Figure 5.9 and 5.10 show that by taking into account more than just one output value, the error probability of the estimator can be lowered. It could be subject of future work to investigate the effects of not only estimating one input value by several output values, but also several input values by several output values, i.e. $(\hat{x}_n ... \hat{x}_{n-L}) = f(y_n, ... y_{n-L})$, which would probably lead to improved results.

Bibliography

[1] R. Fano. Transmission of information. Cambridge, Mass., M.I.T. Press, 1961.
[2] M. Feder and N. Merhav. Relations between entropy and error probability, 1992.
[3] B. C. Geiger, C. Feldbauer, and G. Kubin. Information loss in static nonlinearities. arXiv:1102.4794 [cs.IT], 2011.
[4] B. C. Geiger, C. Feldbauer, and G. Kubin. Loss rate of information in static nonlinearities, In Progress.
[5] N. Jacobsen. Basic algebra. page 210, 2009.
[6] R. Moddemeijer. Matlab library. http://www.cs.rug.nl/ rudy/matlab/index.html, 2010.