Semi-Analytical Solution for Rate Distortion Function and OPTA for Sources with Arbitrary Distribution

Matthias Rüngeler, Birgit Schotsch, Peter Vary
Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{ruengeler | schotsch | vary}@ind.rwth-aachen.de

This documents includes the paper published at the SCC 2010 and this additional page with one erratum and one additional figure.

I. Erratum

In equation (7) the sign is incorrect. The correct equation states:

$$\mathrm{MMSE}(\mathrm{cSNR}) = \frac{1}{1+\mathrm{cSNR}} - 2\frac{d}{d\mathrm{cSNR}} D(p_{\mathcal{Y}} || p_{\mathcal{Y}'}, \mathrm{cSNR}).$$

II. Addendum

The poster published at the conference included an additional figure to the published paper which helps to clarify the first paragraphs in Section III:

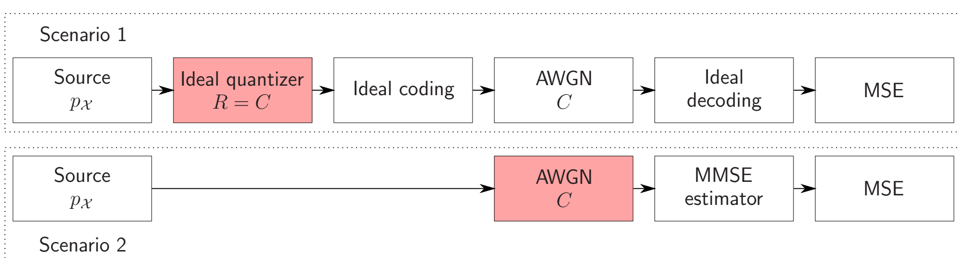

Fig. 1. Two equivalent scenarios for interpretation

Semi-Analytical Solution for Rate Distortion Function and OPTA for Sources with Arbitrary Distribution

Matthias Rüngeler, Birgit Schotsch, Peter Vary
Institute of Communication Systems and Data Processing (ind)
RWTH Aachen University, Germany
{ruengeler | schotsch | vary}@ind.rwth-aachen.de

***Abstract*—The rate distortion function is a widely used theoretical bound which describes the minimum mean square error (MMSE) distortion for a given number of quantization bits when quantizing a scalar random variable. An analytical solution for this function is only available for a small number of probability density functions (pdf), such as the Gaussian pdf. For arbitrary pdfs, the Blahut-Arimoto algorithm [1], [2] needs to be applied to iteratively estimate the rate distortion function. We propose a novel (semi-)analytical and non-iterative method to calculate the rate distortion function for sources with arbitrary pdfs. This method is based on the Guo–Shamai–Verdú (GSV) theorem [3]. Furthermore, it is possible to apply the proposed method for calculating the Optimum Performance Theoretically Attainable (OPTA) for arbitrarily distributed input symbols observed through an AWGN channel.**

***Index Terms*—Blahut-Arimoto Algorithm, Optimum Performance Theoretically Attainable (OPTA) for AWGN Channels, Rate Distortion Function, Guo–Shamai–Verdú (GSV) theorem**

I. Introduction

When encoding realizations of a continuous-amplitude discrete-time scalar random variable, the well known rate distortion function expresses how many (information) bits are at least required to obtain a certain minimum mean square error (MMSE) distortion. The rate distortion function depends on the input symbol probability density function (pdf) and analytical solutions are available for some pdfs like the Gaussian or the bipolar pdf [1], [2]. For arbitrarily distributed input symbols, e.g. with uniform or Laplacian pdfs, the rate distortion function can be calculated numerically using the iterative Blahut-Arimoto algorithm [4], [5]. We propose a novel (semi-)analytical and non-iterative approach for obtaining rate distortion functions for sources with arbitrary pdfs.

A theoretical bound which describes the maximum input-output symbol signal-to-noise ratio (sSNR) of symbols observed through an AWGN channel with a certain channel SNR (cSNR) is the Optimum Performance Theoretically Attainable (OPTA). OPTA is closely related to the rate distortion function, yet it is not possible to use the Blahut-Arimoto algorithm to directly calculate OPTA. The proposed approach also provides (semi-)analytical solutions for OPTA of random variables with arbitrary pdfs observed through additive white Gaussian noise. Section II provides the derivation of the proposed approach which is interpreted and illustrated with some examples in Section III. Section IV summarizes the results.

II. (Semi-)Analytical Approach to Rate Distortion Functions and OPTA

The derivation of the rate distortion function and OPTA consists mainly of three steps. First, the input-output mutual information for arbitrarily distributed input symbols is derived, and second, a recently found relationship between information theory and estimation theory, the Guo–Shamai–Verdú (GSV) theorem [3] is applied. The third step consists of an interpretation (Sec. III) to show the equivalence of the results of the proposed approach and the Blahut-Arimoto algorithm.

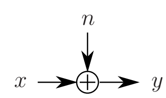

Fig. 1. A generic AWGN channel

We consider a generic scalar AWGN channel (Fig. 1) with input symbols x with an arbitrary pdf $p_{\mathcal{X}}$ and a variance of σ_x^2, additive white Gaussian noise symbols n with a variance σ_n^2 and output symbols $y = x + n$. The channel quality is cSNR $= \frac{\sigma_x^2}{\sigma_n^2}$.

The input-output mutual information $I(\mathcal{X};\mathcal{Y})$ of the AWGN channel is

$$I(\mathcal{X};\mathcal{Y}) = h(\mathcal{Y}') - D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}) - h(\mathcal{N}) \qquad (1)$$

which will be derived and further reformulated in the following.

The noise and the input symbols are statistically independet and $\mathcal{Y}'$ is a random variable with a zero-mean Gaussian pdf $p_{\mathcal{Y}'}$ which has the same variance as $\mathcal{Y}$ ($\sigma_{y'}^2 = \sigma_y^2$). This pdf $p_{\mathcal{Y}'}$ is called *equivalent Gaussian pdf* in the following. The function $h(\cdot)$ stands for the differential entropy[1] $h(\mathcal{A}) = -\int p_{\mathcal{A}}(a) \cdot \ln(p_{\mathcal{A}}(a))\, da$ and $D(f(a)||g(a)) = \int f(a) \ln\left(\frac{f(a)}{g(a)}\right) da$ is the Kullback-Leibler distance [1, (8.46)].

To derive (1), the statistical independence of the input symbols and the noise is used while expanding the mutual information:

[1]To improve readability, limits of integrals are omitted when integrating from $-\infty$ to ∞.

ISBN 978-3-8007-3211-1

$$
\begin{aligned}
I(\mathcal{X};\mathcal{Y}) &= h(\mathcal{Y}) - h(\mathcal{Y}|\mathcal{X}) = h(\mathcal{Y}) - h(\mathcal{X}+\mathcal{N}|\mathcal{X}) \\
&= h(\mathcal{Y}) - h(\mathcal{N}|\mathcal{X}) = h(\mathcal{Y}) - h(\mathcal{N}) \\
&= -\int p_{\mathcal{Y}}(y)\cdot \ln\left(p_{\mathcal{Y}}(y)\right) dy - h(\mathcal{N}) \\
&= -\int p_{\mathcal{Y}}(y)\cdot \ln\left(p_{\mathcal{Y}}(y)\frac{p_{\mathcal{Y}'}(y)}{p_{\mathcal{Y}'}(y)}\right) dy - h(\mathcal{N}) \\
&= -\int p_{\mathcal{Y}}(y)\cdot \ln\left(p_{\mathcal{Y}'}(y)\right) dy \\
&\quad -\int p_{\mathcal{Y}}(y)\cdot \ln\left(\frac{p_{\mathcal{Y}}(y)}{p_{\mathcal{Y}'}(y)}\right) dy - h(\mathcal{N}) \qquad (2) \\
&= -\int p_{\mathcal{Y}'}(y)\cdot \ln\left(p_{\mathcal{Y}'}(y)\right) dy \\
&\quad -\int p_{\mathcal{Y}}(y)\cdot \ln\left(\frac{p_{\mathcal{Y}}(y)}{p_{\mathcal{Y}'}(y)}\right) dy - h(\mathcal{N}) \qquad (3) \\
&= h(\mathcal{Y}') - D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}) - h(\mathcal{N}).
\end{aligned}
$$

The conversion from (2) to (3) can also be found in [1, (8.72)-(8.76)] but is elaborated with more detail for convenience in the following using $\sigma_y^2 = \sigma_{y'}^2$:

$$
\begin{aligned}
&-\int p_{\mathcal{Y}}(y)\cdot \ln\left(p_{\mathcal{Y}'}(y)\right) dy \\
&= -\int p_{\mathcal{Y}}(y)\cdot \ln\left(\frac{1}{\sqrt{2\pi\sigma_{y'}^2}}\cdot \mathrm{e}^{\frac{-y^2}{2\sigma_{y'}^2}}\right) dy \\
&= \ln\left(\sqrt{2\pi\sigma_{y'}^2}\right)\underbrace{\int p_{\mathcal{Y}}(y)dy}_{=1} + \frac{1}{2\sigma_{y'}^2}\underbrace{\int y^2 p_{\mathcal{Y}}(y)dy}_{=\sigma_y^2} \\
&= \ln\left(\sqrt{2\pi\sigma_{y'}^2}\right)\underbrace{\int p_{\mathcal{Y}'}(y)dy}_{=1} + \frac{1}{2\sigma_{y'}^2}\underbrace{\int y^2 p_{\mathcal{Y}'}(y)dy}_{=\sigma_{y'}^2} \\
&= -\int p_{\mathcal{Y}'}(y)\cdot \ln\left(p_{\mathcal{Y}'}(y)\right) dy. \qquad (4)
\end{aligned}
$$

Equation (1) can be futher reformulated using the differential entropy of a continuous-amplitude Gaussian random variable [1]:

$$
\begin{aligned}
I(\mathcal{X};\mathcal{Y}) &= h(\mathcal{Y}') - h(\mathcal{N}) - D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}) \\
&= \frac{1}{2}\ln\left(2\pi\mathrm{e}\sigma_{y'}^2\right) - \frac{1}{2}\ln\left(2\pi\mathrm{e}\sigma_n^2\right) - D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}) \\
&= \frac{1}{2}\ln\left(\frac{\sigma_{y'}^2}{\sigma_n^2}\right) - D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}) \\
&= \frac{1}{2}\ln\left(\frac{\sigma_n^2+\sigma_{x'}^2}{\sigma_n^2}\right) - D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}) \\
&= \frac{1}{2}\ln\left(1+\text{cSNR}\right) - D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}). \qquad (5)
\end{aligned}
$$

The first term of (5) is the well known channel capacity for AWGN channels. The output pdf $p_{\mathcal{Y}}$ which is needed to calculate the Kullback-Leibler distance is the convolution of the input pdf $p_{\mathcal{X}}$ and the Gaussian pdf $p_{\mathcal{N}}$ of the noise ($p_{\mathcal{Y}} = p_{\mathcal{X}} * p_{\mathcal{N}}$). Therefore, the Kullback-Leibler distance and hence the mutual information can be obtained analytically for pdfs which allow closed form convolution and integration to calculate the Kullback-Leibler distance. For other pdfs, convolution and integration can be carried out semi-analytically.

In [3] an analytical relationship between the mutual information in nats and the minimum MSE (MMSE) of arbitrarily distributed input symbols x and their observations y through an AWGN channel (GSV theorem) is derived:

$$\text{MMSE}(\text{cSNR}) = 2\cdot\frac{d}{d\text{cSNR}}I(\mathcal{X};\mathcal{Y},\text{cSNR}). \qquad (6)$$

By inserting (5) into (6) we obtain our key equation:

$$
\begin{aligned}
\text{MMSE}(\text{cSNR}) = &\frac{1}{1+\text{cSNR}} \\
&+2\frac{d}{d\text{cSNR}}D(p_{\mathcal{Y}}||p_{\mathcal{Y}'},\text{cSNR}). \qquad (7)
\end{aligned}
$$

In the context of rate distortion functions, the rate equals the mutual information $I(\mathcal{X};\mathcal{Y})$ and the distortion equals the MMSE. The different operating points on the rate distortion function can be obtained by varying cSNR.

Using our key equation (7), the MMSE and hence the maximum sSNR for a given cSNR can be calculated to directly obtain OPTA:

$$\text{sSNR}(\text{cSNR}) = \frac{\sigma_x^2}{\text{MMSE}(\text{cSNR})}. \qquad (8)$$

III. Interpretation and Examples

The rate distortion function in the context of quantization describes the minimum distortion which can be achieved by an optimal quantizer with a given rate, i.e. a given number of quantization bits per source symbol. The rate distortion function can e.g. be obtained iteratively using the Blahut-Arimoto algorithm. We consider two scenarios to obtain the rate distortion function in a different way:

Scenario 1: The function does not change when it is calculated for a transmission system in which the quantization bits are ideally channel coded and transmitted over an AWGN channel whose capacity equals the rate of the quantizer. The quantization bits are not altered by the appended transmission and thus the distortion is only introduced by the quantizer.

Scenario 2: The source symbols are directly transmitted without quantization (therefore no quantization noise is introduced) and without channel coding over the AWGN channel which has the same capacity as in scenario 1. Now, the distortion is only introduced by the noise of the AWGN channel.

In both scenarios, the overall information rate is the same and since only ideal components are assumed, the minimum distortion which can be achieved is also the same. In scenario 2, the mutual information or the information rate, respectively, is limited by the channel noise and not by the number of quantization bits as in scenario 1. Therefore the formulas which hold for scenario 2 to calculate the information rate i.e. mutual information (5) of the transmission system and the corresponding MMSE distortion (6) can be used to calculate the rate distortion function of quantizing arbitrarily distributed input symbols.

In contrast to the Blahut-Arimoto algorithm [4] where an AWGN channel and hence the cSNR is not part of the algorithm, the algorithm proposed in this paper explicitly uses the AWGN channel for calculating the rate distortion function. Therefore, the corresponding cSNR for each rate-distortion-pair is also available and OPTA can easily be obtained.

An advantage of the Blahut-Arimoto algorithm is the possibility to use arbitrary distortion measures whereas the proposed algorithm is limited to the MSE distortion measure. But the proposed algorithm calculates the rate distortion function and OPTA (semi-)analytically which yields more accurate results with lower computational complexity than the Blahut-Arimoto algorithm.

A. Rate Distortion Functions

To calculate the rate distortion functions, (5) and (7) can be solved for arbitrary input pdfs. Due to the statistical independence of the signal and the noise, the output pdf $p_{\mathcal{Y}}$ is the convolution of the pdfs of the source symbols and the noise symbols:

$$p_{\mathcal{Y}} = p_{\mathcal{X}} * p_{\mathcal{N}}. \tag{9}$$

The equivalent Gaussian pdf $p_{\mathcal{Y}'}$ has the same variance as $p_{\mathcal{Y}}$: $\sigma_{y'}^2 = \sigma_y^2 = \sigma_x^2 + \sigma_n^2$.

Without loss of generality, we set $\sigma_n^2 = 1$ and $\alpha = \frac{\sigma_x^2}{\sigma_n^2} =$ cSNR. For some exemplary arbitrary input pdfs, the resulting output pdfs are the following:

Uniform input pdf $p_{\mathcal{X}}$:

$$p_{\mathcal{Y}}(y,\alpha) = \frac{1}{4\sqrt{3\alpha}}\left[\mathrm{erf}\left(\frac{y+\sqrt{3\alpha}}{\sqrt{2}}\right) - \mathrm{erf}\left(\frac{y-\sqrt{3\alpha}}{\sqrt{2}}\right)\right].$$

Laplacian input pdf $p_{\mathcal{X}}$:

$$p_{\mathcal{Y}}(y,\alpha) = \frac{1}{2\sqrt{2\alpha}}\Bigg\{\mathrm{e}^{\frac{1}{\alpha}-y\sqrt{\frac{2}{\alpha}}}\left[1-\mathrm{erf}\left(\frac{1}{\sqrt{\alpha}}-\frac{y}{\sqrt{2}}\right)\right] + \mathrm{e}^{\frac{1}{\alpha}+y\sqrt{\frac{2}{\alpha}}}\left[1-\mathrm{erf}\left(\frac{1}{\sqrt{\alpha}}+\frac{y}{\sqrt{2}}\right)\right]\Bigg\}.$$

Bipolar input pdf $p_{\mathcal{X}}$:

$$p_{\mathcal{Y}}(y,\alpha) = \frac{1}{2\sqrt{2\pi}}\left(\mathrm{e}^{-\frac{(y+\sqrt{\alpha})^2}{2}} + \mathrm{e}^{-\frac{(y-\sqrt{\alpha})^2}{2}}\right).$$

Figure 2 shows the mutual information $I(\mathcal{X};\mathcal{Y})$ in bits[2] for different input symbol pdfs. For continuous-amplitude input symbol pdfs, the mutual information rises approximately linearly with cSNR, but in the bipolar case, the mutual information saturates at 1 bit. For a fixed cSNR, the highest rates are achieved by the Gaussian input pdf and the rates for the uniform and Laplacian pdfs are just below.

[2]All calculations have been carried out in nats to avoid frequent normalizations with $\ln(2)$. The graphs are plotted in bits, though.

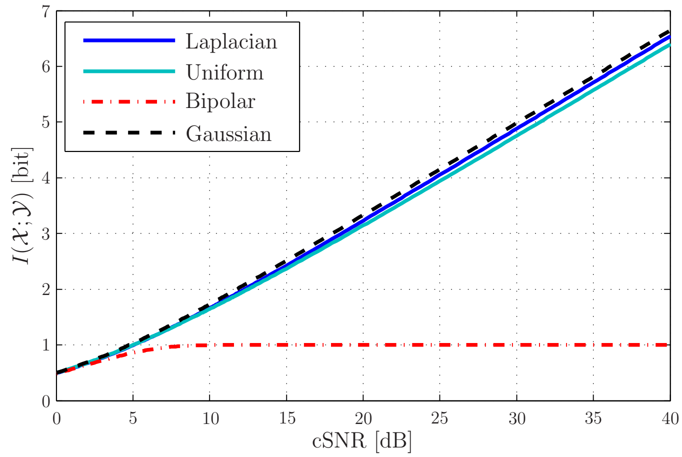

Fig. 2. Mutual information of a transmission system using an AWGN channel for different input symbol pdfs

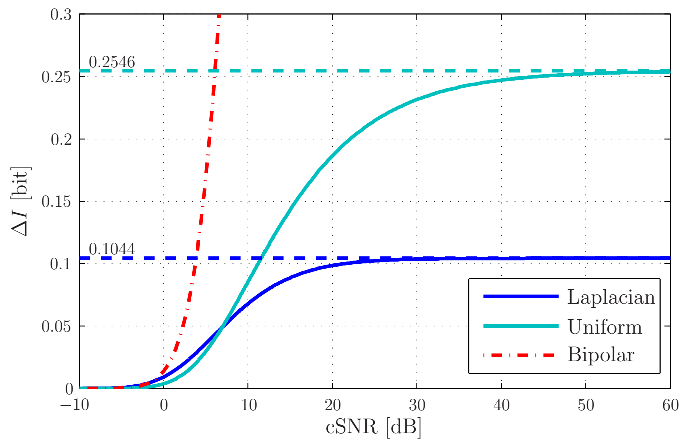

Fig. 3. Difference between the mutual information of Gaussian input symbols and arbitrarily distributed input symbols and their observation through additive white Gaussian noise

Figure 3 shows the difference $\Delta I = I^{[\mathrm{Gauss}]} - I^{[\mathrm{arbit.}]}$ between the mutual information of Gaussian input symbols and arbitrarily distributed input symbols. For vanishing noise (cSNR $\to \infty$) the output symbol pdf converges to the input symbol pdf $p_{\mathcal{Y}} \to p_{\mathcal{X}}$ and ΔI saturates at the shaping gain i.e. the Kullback-Leibler distance between the input symbol pdf and the equivalent Gaussian pdf: $D(p_{\mathcal{Y}}||p_{\mathcal{Y}'}) \to D(p_{\mathcal{X}}||p_{\mathcal{Y}'}) = h(\mathcal{Y}') - h(\mathcal{X})$. Using the differential entropy of uniform, Laplacian and Gaussian random variables, ΔI saturates at $\Delta I_{\text{max, uniform}} = 0.2546$ bit and $\Delta I_{\text{max, Laplacian}} = 0.1044$ bit, indicated by the dashed lines in Fig. 3. The Kullback-Leibler distance between a discrete and a continuous-amplitude random variable is infinite [1] and therefore ΔI for the bipolar input pdf goes to infinity for rising cSNR.

Figure 4 shows the rate distortion functions calculated with (7) for arbitrary input symbol pdfs. This rate distortion functions are equal to the curves published in [6, Fig. 2] using the Blahut-Arimoto algorithm. Especially the characteristic crossovers of the curves of the Laplacian and uniform pdf in the high distortion region can be observed in both graphs.

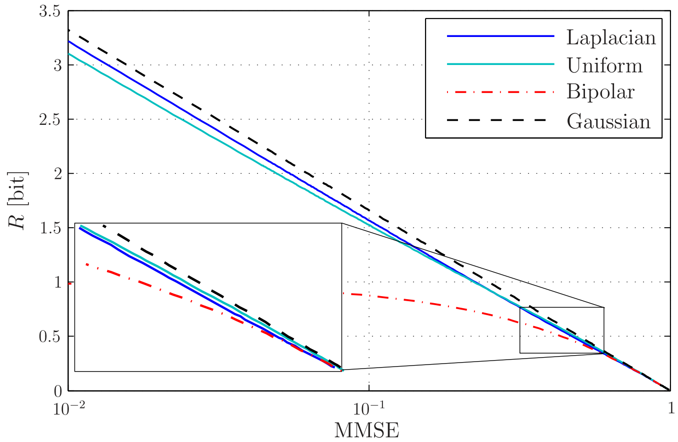

Fig. 4. Rate distortion functions for sources with different pdfs

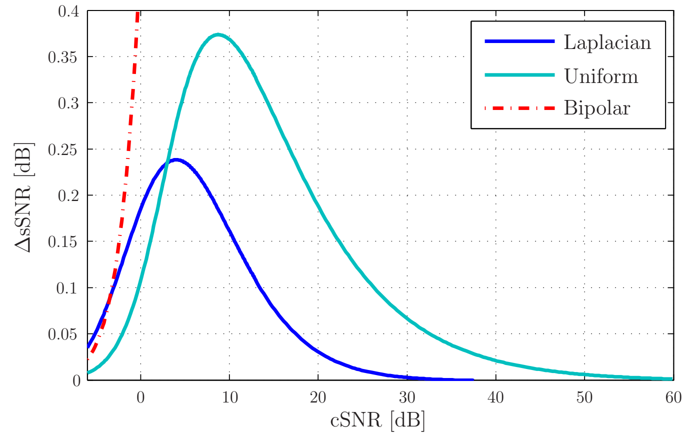

Fig. 6. Difference between OPTA of a transmission system using an AWGN channel for arbitrary input pdfs and Gaussian input pdfs

B. Optimum Performance Theoretically Attainable (OPTA)

OPTA for an AWGN channel describes the theoretical bound for the sSNR for a given channel quality cSNR and a fixed input symbol pdf. The Blahut-Arimoto algorithm iteratively computes rate-distortion-pairs, but it has no connection to the channel quality of an AWGN channel. Equation (5) can be used to calculate the cSNR for a given rate calculated with the Blahut-Arimoto algorithm to obtain OPTA, however numerical inaccuracies can easily produce misleading results. The proposed algorithm uses the cSNR as an inherent parameter and therefore, OPTA can be obtained directly from (5) and (8).

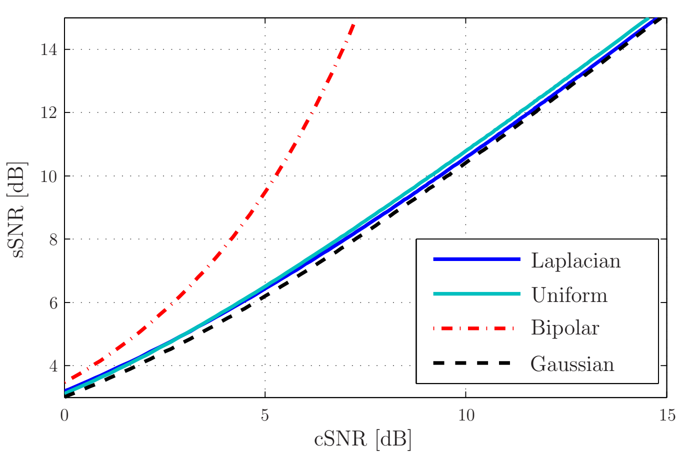

Fig. 5. OPTA of a transmission system using an AWGN channel for different input symbol pdfs

Figure 5 shows OPTA of a transmission system using an AWGN channel with different channel qualities and input symbol pdfs. The lowest sSNR for a fixed cSNR is obtained by the Gaussian input symbol pdf. One can observe that the input pdf with the highest mutual information for a given cSNR does not automatically generate the highest sSNR, but in contrast, yields the lowest sSNR.

The difference in sSNR between an arbitrary pdf and the Gaussian pdf is shown in Fig. 6. Interestingly, for the two evaluated continuous-amplitude pdfs, the difference approaches zero for small and large cSNR, and a maximum can be observed at around cSNR $= 4$ and 9 dB, respectively. For very good or very bad channels, a decoder which is optimized to decode input symbols with a uniform or Laplacian pdf cannot yield a gain over a linear estimator, which is optimal for a Gaussian input symbol pdf. Only for channels between approximately 0 dB and 25 dB a ΔsSNR of at least 0.1 dB can be expected for e.g. a uniform input symbol pdf. This behavior can be explained with the saturation of the difference ΔI between the mutual information of Gaussian and arbitrarily distributed input symbols in Fig. 3. The MMSE depends on the derivative of the mutual information, and therefore, when the difference in mutual information saturates, the difference in MMSE goes to zero. This relationship can also be shown analytically: Without loss of generality, we set $\sigma_x^2 = 1$ and define $I^{[\text{arbit.}]} = I(\mathcal{X};\mathcal{Y},\text{cSNR})$, $I^{[\text{Gauss}]} = \frac{1}{2}\ln(1+\text{cSNR})$, $\Delta\text{sSNR} = 10\log_{10}\left(\text{sSNR}^{[\text{arbit.}]}\right) - 10\log_{10}\left(\text{sSNR}^{[\text{Gauss}]}\right)$ and $\Delta I = I^{[\text{Gauss}]} - I^{[\text{arbit.}]} = D(p_{\mathcal{Y}}||p_{\mathcal{Y}'},\text{cSNR})$

Using (6) it follows:

$$
\begin{aligned}
2\frac{d}{d\text{cSNR}}\Delta I &= 2\frac{d}{d\text{cSNR}}I^{[\text{Gauss}]} - 2\frac{d}{d\text{cSNR}}I^{[\text{arbit.}]} \\
&= \text{MMSE}^{[\text{Gauss}]} - \text{MMSE}^{[\text{arbit.}]} \\
&= \frac{1}{\text{sSNR}^{[\text{Gauss}]}} - \frac{1}{\text{sSNR}^{[\text{arbit.}]}} \\
&= \frac{1}{\text{sSNR}^{[\text{Gauss}]}}\left(1 - \frac{\text{sSNR}^{[\text{Gauss}]}}{\text{sSNR}^{[\text{arbit.}]}}\right)
\end{aligned}
$$

$$
\Leftrightarrow \quad \frac{\text{sSNR}^{[\text{Gauss}]}}{\text{sSNR}^{[\text{arbit.}]}} = 1 - 2\text{sSNR}^{[\text{Gauss}]}\frac{d}{d\text{cSNR}}\Delta I. \tag{10}
$$

The $\Delta\text{sSNR}(\text{cSNR})$ can be calculated using (10) and $\text{sSNR}^{[\text{Gauss}]} = \frac{1}{2\frac{d}{d\text{cSNR}}\frac{1}{2}\ln(1+\text{cSNR})} = 1 + \text{cSNR}$:

$$\begin{aligned}\Delta\text{sSNR}(\text{cSNR}) &= 10\log_{10}\left(\text{sSNR}^{[\text{arbit.}]}\right) - 10\log_{10}\left(\text{sSNR}^{[\text{Gauss}]}\right)\\ &= 10\log_{10}\left(\frac{\text{sSNR}^{[\text{arbit.}]}}{\text{sSNR}^{[\text{Gauss}]}}\right)\\ &= -10\log_{10}\left(1-2(1+\text{cSNR})\frac{d}{d\text{cSNR}}\Delta I\right)\\ &= -10\log_{10}\left(1-2(1+\text{cSNR})\frac{d}{d\text{cSNR}}D(p_{\mathcal{Y}}||p_{\mathcal{Y}'})\right).\end{aligned} \tag{11}$$

Equation (11) clearly shows that a saturation in D yields a vanishing ΔsSNR.

IV. Summary

In this paper a (semi-)analytical solution is proposed to calculate the rate distortion function for arbitrarily distributed source symbols. The approach is not iterative and therefore has great advantages over the iterative Blahut-Arimoto algorithm. Furthermore, OPTA can also be calculated for AWGN channels and arbitrarily distributed input symbols which could not be achieved directly with the Blahut-Arimoto algorithm. The approach delivers deeper insights into estimating random variables observed through additive white Gaussian noise and gives a mathematical explanation for the convergence of the sSNR for high and low cSNR for arbitrarily distributed continuous-amplitude discrete-time input symbols.

References

[1] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley-Interscience New York, 2006.

[2] R.E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1987.

[3] D. Guo, S. Shamai, and S. Verdu, "Mutual Information and MMSE in Gaussian Channels," *IEEE International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.*, pp. 349–349, 2004.

[4] R.E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Transactions on Information Theory*, vol. 18, no. 4, pp. 460–473, Jul 1972.

[5] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 18, no. 1, pp. 14–20, Jan 1972.

[6] P. Noll and R. Zelinski, "Bounds on quantizer performance in the low bit-rate region," *IEEE Transactions on Communications*, vol. 26, no. 2, pp. 300–304, Feb 1978.