Upper-Bounding Rate Distortion Functions based on the Minimum Mean-Square Error

Meik Dörpinghaus*, Matthias Rüngeler**, and Rudolf Mathar*

*Institute for Theoretical Information Technology, RWTH Aachen University,
52056 Aachen, Germany, {doerpinghaus, mathar}@ti.rwth-aachen.de

**Institute for Communication Systems and Data Processing, RWTH Aachen University,
52056 Aachen, Germany, ruengeler@ind.rwth-aachen.de

Abstract—**We derive a new upper bound on the rate distortion function for arbitrary memoryless sources, which is based on the relation between mutual information and minimum mean-square error discovered by Guo et al. This upper bound is in general tighter than the well known upper bound given by the rate distortion function of a Gaussian source with an equal variance found by Shannon and becomes tight for Gaussian sources. We evaluate the new upper bound for various source distributions and compare it to the Shannon lower and upper bound and to the rate distortion function calculated with the Blahut-Arimoto algorithm. This shows that the new upper bound is quite tight.**

I. INTRODUCTION AND SETUP

It is well known that the rate distortion function states which rate R is at least required to describe a source X, such that it can be reconstructed up to a given distortion D for a specific distortion measure.

Let X be a random source that generates independent identically distributed (i.i.d.) symbols with respect to the probability distribution function (PDF) $p(x)$, i.e., $X \sim p(x)$. When encoding, source sequences X^n consisting of n source symbols are mapped onto indices

$$f_n : X^n \to \{1, 2, \dots, 2^{nR}\}. \tag{1}$$

Here, R is the rate of the encoded sequence. By this mapping the source sequences are represented by nR bits. Based on this index representation the decoder is able to generate an estimate $\hat{X}^n$ of the source sequence X^n, i.e., the decoding function is

$$g_n : \{1, 2, \dots, 2^{nR}\} \to \hat{X}^n. \tag{2}$$

The encoding and decoding given by the functions f_n and g_n is often referred to as a $(2^{nR}, n)$-rate distortion code.

The question is: How *close* can the source sequences X^n be reconstructed when they are encoded with rate R? I.e., how large is the expected distortion D between the source sequences X^n and their reconstruction $\hat{X}^n$ with

$$D = \mathrm{E}\left[d(X^n, g_n(f_n(X^n)))\right] \qquad \text{where}$$

$$d(x^n, \hat{x}^n) = d(x^n, g_n(f_n(x^n))) = \frac{1}{n}\sum_{i=1}^{n} d(x_i, \hat{x}_i) \tag{3}$$

is the distortion between the source sequence x^n and its reconstruction $\hat{x}^n$. In addition, $d(x_i, \hat{x}_i)$ is the distortion between the individual source symbols x_i and their estimates $\hat{x}_i$.

The distortion D and the rate R form a rate distortion pair (R, D). A rate distortion pair is achievable if a sequence of $(2^{nR}, n)$-rate distortion codes (f_n, g_n) exists such that

$$\lim_{n \to \infty} \mathrm{E}\left[d(X^n, g_n(f_n(X^n)))\right] \le D. \tag{4}$$

The rate distortion function $R(D)$ is defined as the minimum rate R that is required to encode the source X for a given distortion D. In [1], [2] Shannon has shown that for an i.i.d. source X with distribution $p(x)$ and bounded distortion function $d(x^n, \hat{x}^n)$ the rate distortion function is given by

$$R(D) = \min_{p(\hat{x}|x) \in \mathcal{S}} \mathcal{I}(X; \hat{X}) \tag{5}$$

where $\mathcal{I}(X; \hat{X})$ is the mutual information of X and $\hat{X}$ and

$$\mathcal{S} = \left\{ p(\hat{x}|x) \,\middle|\, \iint p(\hat{x}|x)p(x)d(x, \hat{x})\, \mathrm{d}x\, \mathrm{d}\hat{x} \le D \right\} \tag{6}$$

see also [3, Th. 10.2.1]. Minimization in (5) is hence over all $p(\hat{x}|x)$ for which $p(x, \hat{x})$ fulfills the distortion constraint D.

For certain distortion measures $d(x, \hat{x})$ and source distributions $p(\cdot)$, $R(D)$ is known. E.g., it is a well known result that for a Gaussian source $X \sim \mathcal{N}(0, \sigma_X^2)$ and the mean-square error as distortion measure, i.e., $\mathrm{E}[(X - \hat{X})^2] \le D$, the rate distortion function is given by[1], see, e.g., [3, Theorem 10.3.2]

$$R(D) = \begin{cases} \frac{1}{2}\log\left(\frac{\sigma_X^2}{D}\right) & \text{for } 0 \le D \le \sigma_X^2, \\ 0 & \text{for } D > \sigma_X^2. \end{cases} \tag{7}$$

While for a few combinations of distortion measures and source distributions the rate distortion functions are known, they are in general unknown. Existing bounds like the Shannon lower bound [2] and the upper bound given by the fact that the rate distortion function for a Gaussian source in (7) is an upper bound to the rate distortion function of arbitrarily distributed sources X with the same variance σ_X^2, see [1], are in general not tight. This is the motivation for the present work. We derive a new upper bound on the rate distortion function for the mean-square error as distortion measure and arbitrary source distributions. This upper bound is based on the relation between the minimum mean-square error (MMSE) and the mutual information given in [4]. Recently, in [5] this relation has already been applied in the context of rate distortion theory. However, in contrast to the present work [5] does not consider minimization over the reproduction distribution for the study of $R(D)$.

II. NEW UPPER BOUND ON $R(D)$

Theorem 1. *Let X be a random source that generates i.i.d. symbols according to $X \sim p(x)$ with $\mathrm{var}(X) = \sigma_X^2$. The minimal required rate R such that $\mathrm{E}[(X - \hat{X})^2] \leq D$ is upper-bounded by*

$$R(D) \leq \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) - \frac{1}{2}\int_0^{\frac{1}{\sigma_N^2}}\left(\frac{\sigma_X^2}{1+\sigma_X^2\gamma} - mmse(\gamma)\right)\mathrm{d}\gamma \tag{8}$$

$$= \frac{1}{2}\log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) - \mathcal{D}(p_{X+N} \| p_{X'+N}) \tag{9}$$

where $mmse(\gamma)$ is the minimum mean-squared error when estimating X disturbed by additive Gaussian noise N, i.e.

$$mmse(\gamma) = \mathrm{E}\left[(X - \mathrm{E}[X|X+N])^2\right] \tag{10}$$

and where γ is the normalized SNR, i.e., $\gamma = \frac{1}{\mathrm{E}[N^2]}$. Furthermore, σ_N^2 is related to the distortion D by

$$D = mmse\left(1/\sigma_N^2\right) \tag{11}$$

$$= \frac{\sigma_X^2\sigma_N^2}{\sigma_N^2 + \sigma_X^2} - 2\frac{\mathrm{d}}{\mathrm{d}\left(\frac{1}{\sigma_N^2}\right)}\mathcal{D}(p_{X+N} \| p_{X'+N}). \tag{12}$$

Here $\mathcal{D}(p_{X+N} \| p_{X'+N})$ is the Kullback-Leibler divergence between the PDF of $X+N$ and of $X'+N$ where X' is a Gaussian random variable with the same mean and variance as X.

Theorem 1 has also the following interpretation. The rate distortion function $R(D)$ for an arbitrarily distributed source X with variance σ_X^2 is upper-bounded by the mutual information of an AWGN channel with the input $X - \mathrm{E}[X]$ and a noise variance σ_N^2, which is chosen such that the distortion is equal to $D = \mathrm{mmse}(1/\sigma_N^2)$. As (9), i.e., the upper bound on $R(D)$, corresponds to the mutual information of an AWGN channel, $\mathrm{mmse}(1/\sigma_N^2)$ and hence D is given by two times the derivative of (9) with respect to $1/\sigma_N^2$ yielding (12), see [4] for the relation between mutual information and MMSE.

The representations in (9) and (12) using the Kullback-Leibler divergence allow an easy numerical evaluation of the upper bound on the rate distortion function. To calculate $\mathcal{D}(p_{X+N} \| p_{X'+N})$ the PDF p_{X+N} is required. As X and the additive noise N are independent, p_{X+N} is given by the convolution of p_X and $p_N = \mathcal{N}(0, \sigma_N^2)$. In case p_{X+N} cannot be obtained in closed-form, a numerical convolution is required to calculate the upper bound on $R(D)$. Thus, evaluation of (9) gives an upper bound on the rate R for a given σ_N^2. The corresponding distortion D can be calculated from (12). Hence, to evaluate the upper bound on the rate distortion function pairs of the upper bound on R given by (9) and D given by (12) have to be calculated by varying the parameter σ_N^2.

III. PROOF OF THEOREM 1

As the rate distortion function is independent of the mean $\mathrm{E}[X]$ of the source, i.e., the rate distortion functions $R(D)$ for the sources X and $X - \mathrm{E}[X]$ are equal, in the following proof we assume that $\mathrm{E}[X] = 0$ without loss of generality.

With (5) the rate distortion function is given by

$$R(D) = \min_{p(\hat{x}|x)\in\mathcal{S}} \mathcal{I}(X;\hat{X}) \tag{13}$$

where the minimization is over all $p(\hat{x}|x)$ such that

$$\mathrm{E}\left[(X - \hat{X})^2\right] \leq D \tag{14}$$

i.e., all $p(\hat{x}|x)$ in the set $\mathcal{S}$ in (6) with $d(x, \hat{x}) = (x - \hat{x})^2$.

As $R(D)$ is monotonically decreasing in D, all $p(\hat{x}|x)$ being solutions of (13) fulfill (14) with equality. Obviously, any choice $p(\check{x}|x) \in \mathcal{S}$ different from the optimal $p(\hat{x}|x)$ which minimizes $\mathcal{I}(X;\hat{X})$ yields an upper bound on $R(D)$:

$$R(D) \leq \mathcal{I}(X;\check{X}). \tag{15}$$

Thus, we construct an upper bound on the rate distortion function $R(D)$ by choosing some $p(\check{x}|x)$ satisfying

$$\mathrm{E}\left[(X - \check{X})^2\right] = D \tag{16}$$

for which $\mathcal{I}(X;\check{X})$ can be easily evaluated and at the same time is a sufficiently tight upper bound. Therefor, we define

$$Y = X + N \tag{17}$$

where N is zero-mean additive white Gaussian noise with variance σ_N^2. Furthermore, let

$$\check{X} = \mathrm{E}[X|Y] \tag{18}$$

be the MMSE estimate of X from Y. Now, the variance σ_N^2 has to be chosen such that (16) is fulfilled. As the MMSE estimate $\check{X}$ is a sufficient statistic of Y, see [6], it does not change the mutual information [3, Sect. 2.9] and it holds that

$$\mathcal{I}(X;\check{X}) = \mathcal{I}(X;Y). \tag{19}$$

Thus, an upper bound on $R(D)$ is given by the mutual information of an AWGN channel with the source X at its input and the variance of the additive Gaussian noise σ_N^2 depending on the distortion D and implicitly given by (10) and (11).

Hence, to prove Theorem 1 it remains to show that (9) is equal to $\mathcal{I}(X;Y)$, i.e., the mutual information of the AWGN channel in (17) with the source X at its input and noise variance σ_N^2. Therefor, we rewrite $\mathcal{I}(X;Y)$ as follows, see [7][2]:

$$\mathcal{I}(Y;X) = h(Y) - h(Y|X) = h(Y) - h(N)$$

$$= -\int p_Y(y)\log\left(p_Y(y)\right)\mathrm{d}y - h(N)$$

$$= -\int p_Y(y)\log\left(p_{Y'}(y)\right)\mathrm{d}y - \int p_Y(y)\log\left(\frac{p_Y(y)}{p_{Y'}(y)}\right)\mathrm{d}y - h(N)$$

$$= -\int p_{Y'}(y)\log\left(p_{Y'}(y)\right)\mathrm{d}y - \int p_Y(y)\log\left(\frac{p_Y(y)}{p_{Y'}(y)}\right)\mathrm{d}y - h(N) \tag{20}$$

$$= h(Y') - \mathcal{D}\left(p_Y \| p_{Y'}\right) - h(N) \tag{21}$$

---

[2]Note that [7] states that inequality (8) in Theorem 1 holds with equality. However, this statement is wrong in general. The erroneous assumption in [7] is that the information rate of a system with optimal source and channel coding is equal to the information rate of an AWGN channel where the source symbols are transmitted without coding. It can be shown that this statement does not hold in general and, therefore, (8) does not hold with equality in general.

where $h(\cdot)$ denotes the differential entropy, $p_Y(y)$ is the PDF of Y, and $p_{Y'}(y)$ is the PDF of a corresponding zero-mean Gaussian variable with the same variance

$$\sigma_{Y'}^2 = \sigma_Y^2 = \sigma_X^2 + \sigma_N^2. \tag{22}$$

Moreover, $\mathcal{D}(p_Y \| p_{Y'})$ is the Kullback-Leibler divergence between $p_Y(y)$ and $p_{Y'}(y)$. Furthermore, (20) holds because

$$-\int p_Y(y) \log(p_{Y'}(y)) \, \mathrm{d}y = -\int p_Y(y) \log\left(\frac{\exp\left(\frac{-y^2}{2\sigma_Y^2}\right)}{\sqrt{2\pi\sigma_Y^2}}\right) \mathrm{d}y$$

$$= \frac{1}{2} \log\left(2\pi e \sigma_Y^2\right) = -\int p_{Y'}(y) \log\left(p_{Y'}(y)\right) \mathrm{d}y. \tag{23}$$

As Y' and N are Gaussian, (21) is given by

$$\mathcal{I}(Y;X) = \frac{1}{2} \log\left(2\pi e \sigma_Y^2\right) - \frac{1}{2} \log\left(2\pi e \sigma_N^2\right) - \mathcal{D}(p_Y \| p_{Y'})$$

$$= \frac{1}{2} \log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) - \mathcal{D}(p_{X+N} \| p_{X'+N}) \tag{24}$$

where for (24) we have used (22). This shows that (9) is the mutual information of an AWGN channel with noise variance σ_N^2 and an arbitrarily distributed source X with zero-mean and variance σ_X^2 at its input.

The alternative representation of the upper bound on $R(D)$ in (8) holds, as $\mathcal{D}(p_Y \| p_{Y'})$ in (24) is given by

$$\mathcal{D}(p_Y \| p_{Y'}) = h(Y') - h(Y) \tag{25}$$

$$= \mathcal{I}(Y';X') + h(Y'|X') - \mathcal{I}(Y;X) - h(Y|X)$$

$$= \mathcal{I}(Y';X') - \mathcal{I}(Y;X) + h(X'+N|X') - h(X+N|X) \tag{26}$$

$$= \frac{1}{2} \int_0^{\frac{1}{\sigma_N^2}} \left(\frac{\sigma_X^2}{1+\sigma_X^2\gamma} - \mathrm{mmse}(\gamma)\right) \mathrm{d}\gamma \tag{27}$$

where X' is a zero-mean Gaussian random variable with variance σ_X^2. If X' is the input to the AWGN channel in (17), this results in the zero-mean Gaussian output Y', yielding (26). Moreover, (25) and (27) follow from the results given in [8].

IV. PROPERTIES OF THE UPPER BOUND

A. Gaussian Sources

For a Gaussian source X, $\mathrm{mmse}(\gamma)$ in the upper bound (8) becomes $\sigma_X^2/(1+\sigma_X^2\gamma)$ and, thus, the integral in (8) is zero. Hence, the upper bound is equal to $\frac{1}{2} \log\left(1 + \sigma_X^2/\sigma_N^2\right)$, which is the capacity of the AWGN channel. However, for a Gaussian source the capacity of the AWGN channel and the rate distortion function $R(D)$ in (7) are equal, as (11) becomes equal to

$$D = \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2 + \sigma_N^2}. \tag{28}$$

which follows from (12) as the Kullback-Leibler divergence in (12) is zero for Gaussian inputs. Thus, for a Gaussian source the upper bound on $R(D)$ in (8) and (9) is tight.

This discussion also shows another well known fact. In case we estimate the Gaussian source X after transmission over an AWGN channel based on Y using an MMSE estimator, the MMSE is given by D in (28), see (10) to (12). To achieve

this distortion no code at all is required. On the other hand, we can source encode the source X using a rate-distortion code, then use a channel code to transmit it reliably over the AWGN channel. Following Shannon's source channel coding separation theorem [1], this separation of source and channel coding is optimal. At the receiver the signal can be channel and source decoded. To achieve a maximum distortion D, the source encoder has to encode the source with rate $R(D)$ given in (7). For reliable transmission over the AWGN channel its capacity has to correspond to $R(D)$, i.e., for $0 \le D \le \sigma_X^2$

$$\frac{1}{2} \log\left(1 + \frac{P}{\sigma_N^2}\right) \doteq R(D) \stackrel{(7),(28)}{=} \frac{1}{2} \log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \tag{29}$$

where P is the required transmit power. Obviously $P = \sigma_X^2$. However, this means that for a given transmit power, there is no advantage of the use of source and channel coding when a Gaussian source has to be transmitted over an AWGN channel. Directly transmitting uncoded source symbols over the AWGN channel yields the same distortion D as perfect source and channel coding. In addition, the latter one implies an infinite delay. This is a well known result, see e.g. [9], [10].

B. Gap to Shannon's Lower Bound

Shannon gave a lower bound on $R(D)$, which for the mean-square error distortion measure becomes [2]

$$R(D) \ge h(X) - \frac{1}{2} \log\left(2\pi e D\right) = \mathrm{LB_{Sha}}. \tag{30}$$

The gap Δ between the new upper bound on $R(D)$ given in Theorem 1 and Shannon's lower bound in (30) is given by

$$\Delta = \mathrm{UB_{new}} - \mathrm{LB_{Sha}} = -h(X|Y) + \frac{1}{2} \log\left(2\pi e\, \mathrm{mmse}\left(\frac{1}{\sigma_N^2}\right)\right). \tag{31}$$

where we have used (9), (30), (11), and (24). The variance of the MMSE estimate $\mathrm{E}[X|Y]$ is equal to $\mathrm{mmse}(1/\sigma_N^2)$. Furthermore, for a given variance Gaussian random variables are entropy maximizers and the second term on the RHS of (31) corresponds to the differential entropy of a Gaussian random variable with variance $\mathrm{mmse}(1/\sigma_N^2)$. In general, the estimation error of the MMSE estimate $\mathrm{E}[X|Y]$ is non-Gaussian. Thus, $h(X|Y)$ in (31) is upper-bounded by the second term on the RHS of (31). Δ corresponds to the negative difference of the entropy of X conditioned on Y and the corresponding entropy in case X is Gaussian with the same variance. Hence, Δ becomes zero for X being Gaussian.

C. Comparison of the New Upper Bound on $R(D)$ with Shannon's Upper Bound on the Rate Distortion Function

Shannon has shown that for an arbitrary distributed source X with variance σ_X^2, the rate distortion function is upper-bounded by the rate distortion function of a Gaussian source with the same variance, which is given in (7). I.e., Gaussian sources are hardest to encode, cf. [11].

The difference between this upper bound in (7) given by Shannon and the new upper bound on $R(D)$ stated in

Theorem 1 is for $D \leq \sigma_X^2$ given by

$$\text{UB}_{\text{Sha}} - \text{UB}_{\text{new}} = \frac{1}{2} \log \left(\frac{\sigma_X^2}{\text{mmse}\left(\frac{1}{\sigma_N^2}\right)} \right) - \frac{1}{2} \int_0^{\frac{1}{\sigma_N^2}} \text{mmse}(\gamma) d\gamma \tag{32}$$

where we have used (7), (11), and (8). The difference on the RHS of (32) depends on $\text{mmse}(\gamma)$, which is bounded by $0 \leq \text{mmse}(\gamma) \leq \frac{\sigma_X^2}{1+\sigma_X^2 \gamma}$, where the RHS is the MMSE in case X is Gaussian. The upper bound on $\text{mmse}(\gamma)$ holds as Gaussian random variables are hardest to estimate [12, Prop. 15].

UB_{new} is in general tighter than UB_{Sha}, as, by using $\text{mmse}(\gamma) \leq \frac{\sigma_X^2}{1+\sigma_X^2 \gamma}$ and the relation between the MMSE and the mutual information given in [4], (32) is lower-bounded by

$$(32) \geq \frac{1}{2} \log \left(1 + \frac{\sigma_X^2}{\sigma_N^2} \right) - \mathcal{I}(X+N;X) \geq 0 \tag{33}$$

where the RHS of (33) follows from the fact that $\mathcal{I}(X+N;X)$ is maximized if X is Gaussian. For $\text{mmse}(\gamma) = \sigma_X^2/(1+\sigma_X^2\gamma)$ (32) is equal to zero.

Finally, for an arbitrary distribution of the source X and $D \to \sigma_X^2$, which by (11) corresponds to $\sigma_N^2 \to \infty$, the MMSE converges to the MMSE in case the source is Gaussian:

$$\lim_{1/\sigma_N^2 = \gamma \to 0} \left\{ \text{mmse}(\gamma) - \frac{\sigma_X^2}{1+\sigma_X^2\gamma} \right\} = 0. \tag{34}$$

Thus, $\text{UB}_{\text{Sha}} - \text{UB}_{\text{new}}$ converges to zero for $D \to \sigma_X^2$.

D. Evaluation of the Upper Bound in Theorem 1

In Fig. 1 we have evaluated[3] the new upper bound on the rate distortion function given by Theorem 1 for a uniform source distribution (Fig. 1(a)), a source with Laplace distribution and a bipolar input distribution (both Fig. 1(b)) all with zero-mean and variance $\sigma_X^2 = 1$, i.e.,

$$p_{X,\text{uniform}}(x) = \begin{cases} \frac{1}{\sqrt{12}} & \text{for } |x| \leq \sqrt{3}, \\ 0 & \text{otherwise}, \end{cases} \tag{35}$$

$$p_{X,\text{Laplace}}(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}|x|\right), \tag{36}$$

$$p_{X,\text{bipolar}}(x) = \begin{cases} \frac{1}{2} & \text{for } x = \pm 1, \\ 0 & \text{otherwise}. \end{cases} \tag{37}$$

For comparison the Shannon upper bound in (7) corresponding to the rate distortion function for a Gaussian source, the Shannon lower bound in (30), and the actual rate distortion function calculated numerically based on the Blahut-Arimoto algorithm [13], [14] are shown. The Shannon lower bound (30) is not shown for the bipolar source as $h(X)$ does not exist in this case.

Fig. 1 shows that the new upper bound and the Shannon lower bound on the rate distortion function become tight for $D \to 0$. On the other hand, for $D \to \sigma_X^2$ the new upper bound converges to the Shannon upper bound. The comparison of the new upper bound with the numerical evaluation of $R(D)$ based on the Blahut-Arimoto shows that the new upper bound is quite tight for the displayed source distributions.

---

[3]The MATLAB source code is available on our websites.

(a) Uniform source distribution

(b) Laplace and bipolar source distribution
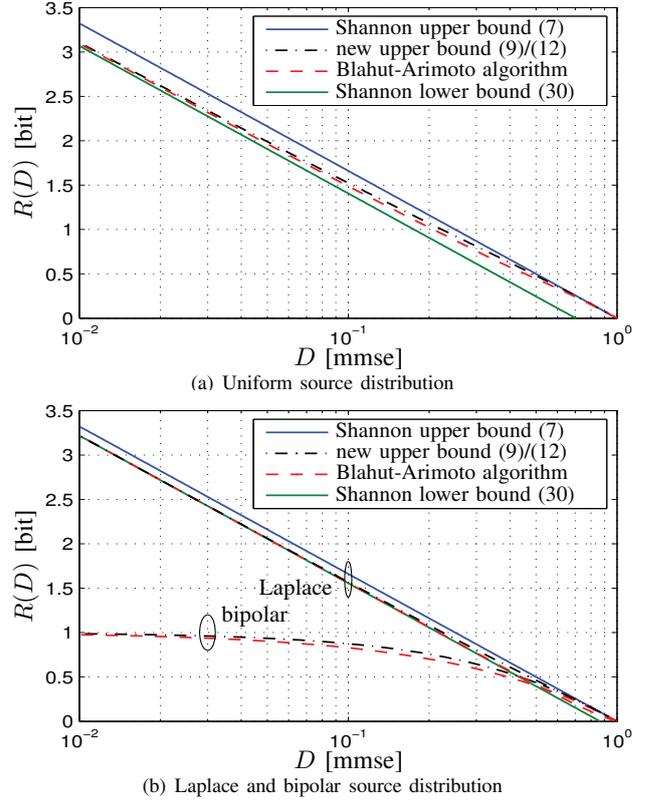
Fig. 1. Comparison of the new upper bound on $R(D)$ given by Theorem 1 with the lower bound (30) and the upper bound (7) given by Shannon and the rate distortion function calculated based on the Blahut-Arimoto algorithm.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 623–656, 1948.

[2] ——, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec., Pt. 4*, pp. 142–163, 1959.

[3] T. Cover and J. Thomas, *Elements of Information Theory, 2nd edition*. New York: Wiley & Sons, 2006.

[4] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[5] N. Merhav, "Rate-distortion function via minimum mean square error estimation," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3196–3206, Jun. 2011.

[6] G. D. Forney Jr., "Shannon meets Wiener II: On MMSE estimation in successive decoding schemes," in *Proc. 42nd Annu. Allerton Conf. Commun., Control, and Computing*, Monticello, IL, Sep./Oct. 2004.

[7] M. Rüngeler, B. Schotsch, and P. Vary, "Semi-analytical solution for rate distortion function and OPTA for sources with arbitrary distribution," in *Proc. 2010 International ITG Conference on Source and Channel Coding (SCC)*, Siegen, Germany, Jan. 2010.

[8] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3712–3720, Aug. 2010.

[9] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[10] M. Gastpar, "To code or not to code," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Dec. 2002.

[11] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 38–47, Jan. 1997.

[12] D. Guo, Y. Wu, S. Shamai (Shitz), and S. Verdú, "Estimation in Gaussian noise: Properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.

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

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