Nonasymptotic noisy lossy source coding

Victoria Kostina, Sergio Verdú
Dept. of Electrical Engineering, Princeton University, NJ 08544, USA

***Abstract*—This paper shows new general nonasymptotic achievability and converse bounds and performs their dispersion analysis for the lossy compression problem in which the compressor observes the source through a noisy channel. While this problem is asymptotically equivalent to a noiseless lossy source coding problem with a modified distortion function, nonasymptotically there is a difference in how fast their minimum achievable coding rates approach the rate-distortion function, providing yet another example where at finite blocklengths one must put aside traditional asymptotic thinking.**

***Index Terms*—Achievability, converse, finite blocklength regime, lossy data compression, noisy source coding, strong converse, dispersion, memoryless sources, Shannon theory.**

I. Introduction

Consider a lossy compression setup in which the encoder has access only to a noise-corrupted version X of a source S, and we are interested in minimizing (in some stochastic sense) the distortion $\mathsf{d}(S, Y)$ between the true source S and its rate-constrained representation Y (see Fig. 1). This problem arises if the object to be compressed is the result of an uncoded transmission over a noisy channel, or if it is observed data subject to errors inherent to the measurement system. Some examples include speech in a noisy environment, or photographic images corrupted by noise introduced by the image sensor and circuitry. Since we are concerned with preserving the original information in the source rather than preserving the noise, the distortion measure is defined with respect to the source.

The noisy source coding setting was first discussed by Dobrushin and Tsybakov [1], who showed that when the goal is to minimize the average distortion, the noisy source coding problem is asymptotically equivalent to a particular noiseless source coding problem. More precisely, for stationary memoryless sources observed through a stationary memoryless channel under a separable distortion measure, the noisy rate-distortion function is given by

$$R(d) = \min_{\substack{P_{\mathsf{Y}|\mathsf{X}}:\\ \mathbb{E}[\mathsf{d}(\mathsf{S},\mathsf{Y})]\leq d\\ \mathsf{S}-\mathsf{X}-\mathsf{Y}}} I(\mathsf{X};\mathsf{Y}) \tag{1}$$

$$= \min_{\substack{P_{\mathsf{Y}|\mathsf{X}}:\\ \mathbb{E}\left[\bar{\mathsf{d}}(\mathsf{X},\mathsf{Y})\right]\leq d}} I(\mathsf{X};\mathsf{Y}) \tag{2}$$

where

$$\bar{\mathsf{d}}(a, b) = \mathbb{E}\left[\mathsf{d}(\mathsf{S}, b) | \mathsf{X} = a\right] \tag{3}$$

i.e. in the limit of infinite blocklengths, the problem is equivalent to the classical lossy source coding problem where the distortion measure is the conditional average of the original distortion measure given the noisy observation of the source. Berger [2, p.79] used the modified distortion measure (3) to streamline the proof that the right side of (2) is the fundamental limit. Witsenhausen [3] explored the strength of distortion measures defined through conditional expectations such as in (3) to treat various so-called indirect rate distortion problems.

Sakrison [4] showed that if both the source and its noise-corrupted version take values in a separable Hilbert space and the fidelity criterion is mean squared error, then asymptotically, an optimal code can be constructed by first creating a minimum mean-square estimate of the source outcome based on its noisy observation, and then vector-quantizing this estimate as if it were noise-free. Wolf and Ziv [5] showed that Sakrison's result holds even nonasymptotically, namely, that the minimum average distortion achievable in one-shot noisy compression of the object S can be written as

$$D^\star(M) = \mathbb{E}\left[|S - \mathbb{E}\left[S|X\right]|^2\right] + \inf_{\mathsf{f},\mathsf{c}} \mathbb{E}\left[|\mathsf{c}(\mathsf{f}(X)) - \mathbb{E}\left[S|X\right]|^2\right] \tag{4}$$

where the infimum is over all encoders $\mathsf{f}\colon \mathcal{X} \mapsto \{1, \ldots, M\}$ and all decoders $\mathsf{c}\colon \{1, \ldots, M\} \mapsto \mathcal{Y}$, and $\mathcal{X}$ and $\mathcal{Y}$ are the alphabets of the channel output and the decoder output, respectively. Note that (4) is a direct consequence of the choice of the mean squared error distortion and does not hold in general. For vector quantization of a Gaussian signal corrupted by an additive independent Gaussian noise under weighted squared error distortion measure, Ayanoglu [6] found explicit expressions for the optimum quantizer values and the optimum quantization rule. Wolf and Ziv's result was extended to waveform vector quantization under weighted quadratic distortion measures and to autoregressive vector quantization under the Itakura-Saito distortion measure by Ephraim and Gray [7], as well as to a model in which the encoder and decoder have access to the history of their past inputs and outputs, allowing exploitation of inter-block dependence, by Fisher, Gibson and Koo [8]. Thus, the cascade of the optimal estimator followed by the optimal quantizer achieves the minimum average distortion in those settings as well.

All efforts [1]–[8] focused on the average (over the source and noise statistics) distortion between the source and its representation by the decoder. This paper considers the excess distortion constraint, i.e. our fidelity criterion is the probability of exceeding a certain distortion threshold. The excess distortion criterion is relevant for example in applications where, if more than a given fraction of bits is erroneous, the entire

This work was supported in part by the National Science Foundation (NSF) under Grant CCF-1016625 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under Grant CCF-0939370.

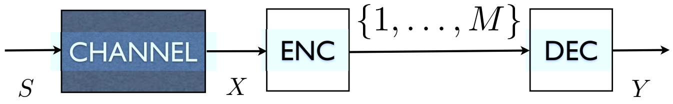

Fig. 1. Noisy source coding.

packet must be discarded. Moreover, as we argued in [9], excess distortion is a natural way to look at lossy compression problems at finite blocklengths where the entire distribution of the distortion observed at the output of the decoder (and not just its average) matters. Evaluating the excess distortion probability at all distortion thresholds gives the full distribution of the distortion at the output of the decoder, providing a more complete description of system performance than that afforded by knowing just the average value of that distortion.

In this paper, we give new nonasymptotic achievability and converse bounds for the noisy source coding problem, which generalize the noiseless source coding bounds in [9]. We observe that at finite blocklenghs, the noisy coding problem is not equivalent to the noiseless coding problem with the modified distortion measure in (3). Essentially, the reason is that the averaging in (3) dismisses the randomness introduced by the noisy channel in Fig. 1, which nonasymptotically cannot be neglected. That additional randomness slows down the rate of approach to the asymptotic fundamental limit in the noisy source coding problem compared to the asymptotically equivalent noiseless problem. Specifically, for noiseless source coding of stationary memoryless sources with separable distortion measure, we showed previously [9] (see also [10] for an alternative proof in the finite alphabet case) that the minimum number M of representation points compatible with a given probability ϵ of exceeding distortion threshold d can be written as

$$\log M^\star(k, d, \epsilon) = kR(d) + \sqrt{k\mathcal{V}(d)}Q^{-1}(\epsilon) + o\left(\sqrt{k}\right) \tag{5}$$

where $\mathcal{V}(d)$ is the rate-dispersion function, explicitly identified in [9], and $Q^{-1}(\cdot)$ denotes the inverse of the complementary standard Gaussian cdf. In this paper, we show that for noisy source coding of a discrete stationary memoryless source over a discrete stationary memoryless channel under a separable distortion measure, $\mathcal{V}(d)$ in (5) is replaced by the noisy rate-dispersion function $\tilde{\mathcal{V}}(d)$, which can be expressed as

$$\tilde{\mathcal{V}}(d) = \mathcal{V}(d) + \lambda^{\star 2}\mathrm{Var}\left(\mathbb{E}\left[\mathsf{d}(\mathsf{S}, \mathsf{Y}^\star) - \bar{\mathsf{d}}(\mathsf{X}, \mathsf{Y}^\star)|\mathsf{S}, \mathsf{X}\right]\right) \tag{6}$$

where $\lambda^\star = -R'(d)$, $\mathsf{S} - \mathsf{X} - \mathsf{Y}^\star$ and $P_{\mathsf{Y}^\star|\mathsf{X}}$ is the minimizer in (2).

Section II shows new nonasymptotic bounds in the general single-shot setting, Section III presents their asymptotic analysis, and Section IV considers the example of a binary source observed through an erasure channel.

II. Nonasymptotic bounds

Consider the setup in Fig. 1 where we are given the distribution P_S on the alphabet $\mathcal{M}$ and the transition probability kernel $P_{X|S}\colon \mathcal{M} \to \mathcal{X}$. We are also given the distortion measure $\mathsf{d}\colon \mathcal{M} \times \mathcal{Y} \mapsto [0, +\infty]$, where $\mathcal{Y}$ is the representation alphabet. An (M, d, ϵ) code is a pair of mappings $P_{Z|X}\colon \mathcal{X} \mapsto \{1, \ldots, M\}$ and $P_{Y|Z}\colon \{1, \ldots, M\} \mapsto \mathcal{Y}$ such that $\mathbb{P}\left[\mathsf{d}(S, Y) > d\right] \le \epsilon$.

Define

$$\mathbb{R}_X(d) = \inf_{\substack{P_{Y|X}\colon \\ \mathbb{E}\left[\bar{\mathsf{d}}(X,Y)\right] \le d}} I(X; Y) \tag{7}$$

where $\bar{\mathsf{d}}\colon \mathcal{X} \times \mathcal{Y} \mapsto [0, +\infty]$ is given by

$$\bar{\mathsf{d}}(x, y) = \mathbb{E}\left[\mathsf{d}(S, y)|X = x\right] \tag{8}$$

and assume that the infimum is achieved by a unique $P_{Y^\star|X}$ such that the constraint is satisfied with equality. Noting that this assumption guarantees differentiability of $\mathbb{R}_X(d)$, denote

$$\tilde{\jmath}_{S,X}(s, x, y, d) = \imath_{X;Y^\star}(x; y) + \lambda^\star \mathsf{d}(s, y) - \lambda^\star d \tag{9}$$

$$\lambda^\star = -\mathbb{R}'_X(d) \tag{10}$$

$$\imath_{X;Y^\star}(x; y) = \frac{dP_{X|Y^\star=y}}{dP_X}(x) \tag{11}$$

Theorem 1 (Converse)**.** *Any* (M, d, ϵ) *code must satisfy*

$$\epsilon \ge \inf_{P_{Y|X}} \sup_{\gamma \ge 0} \left\{\mathbb{P}\left[\tilde{\jmath}_{S,X}(S, X, Y, d) \ge \log M + \gamma\right] - \exp(-\gamma)\right\} \tag{12}$$

Proof: Let the encoder and decoder be the random transformations $P_{Z|X}$ and $P_{Y|Z}$, where Z takes values in $\{1, \ldots, M\}$. Denote

$$B_d(s) = \{y \in \mathcal{Y}\colon \mathsf{d}(s, y) \le d\} \tag{13}$$

We have, for any $\gamma \ge 0$

$$\begin{aligned}
&\mathbb{P}\left[\tilde{\jmath}_{S,X}(S, X, Y, d) \ge \log M + \gamma\right] \\
&= \mathbb{P}\left[\tilde{\jmath}_{S,X}(S, X, Y, d) \ge \log M + \gamma, d(S, Y) > d\right] \\
&+ \mathbb{P}\left[\tilde{\jmath}_{S,X}(S, X, Y, d) \ge \log M + \gamma, d(S, Y) \le d\right] && (14) \\
&\le \epsilon + \sum_{x \in \mathcal{X}} P_X(x) \sum_{s \in \mathcal{M}} P_{S|X}(s|x) \sum_{z=1}^{M} P_{Z|X}(z|x) \\
&\quad \cdot \sum_{y \in B_d(s)} P_{Y|Z}(y|z) 1\left\{M \le \exp\left(\tilde{\jmath}_{S,X}(s, x, y, d) - \gamma\right)\right\} && (15) \\
&\le \epsilon + \exp(-\gamma) \sum_{x \in \mathcal{X}} P_X(x) \sum_{s \in \mathcal{M}} P_{S|X}(s|x) \\
&\quad \cdot \sum_{z=1}^{M} \frac{1}{M} \sum_{y \in B_d(s)} P_{Y|Z}(y|z) \exp\left(\tilde{\jmath}_{S,X}(s, x, y, d)\right) && (16) \\
&\le \epsilon + \exp(-\gamma) \sum_{z=1}^{M} \frac{1}{M} \sum_{y \in \mathcal{Y}} P_{Y|Z}(y|z) \\
&\quad \cdot \sum_{x \in \mathcal{X}} P_X(x) \exp\left(\imath_{X;Y^\star}(x; y)\right) \sum_{s \in \mathcal{M}} P_{S|X}(s|x) && (17) \\
&= \epsilon + \exp(-\gamma) \sum_{z=1}^{M} \frac{1}{M} \sum_{y \in \mathcal{Y}} P_{Y|Z}(y|z) \sum_{x \in \mathcal{X}} P_{X|Y^\star}(x|y) && (18) \\
&= \epsilon + \exp(-\gamma) && (19)
\end{aligned}$$

where

- (16) follows by upper-bounding

$$P_{Z|X}(z|x)1\left\{M \le \exp\left(\tilde{\jmath}_{S,X}(s,x,y,d)-\gamma\right)\right\} \le \frac{\exp(-\gamma)}{M}\exp\left(\tilde{\jmath}_{S,X}(s,x,y,d)\right) \tag{20}$$

  for every $(x,z) \in \mathcal{M} \times \{1,\ldots,M\}$,
- (17) uses

$$1\{\mathsf{d}(s,y) \le d\} \le \exp\left(\lambda^\star d - \lambda^\star \mathsf{d}(s,y)\right) \tag{21}$$

Finally, (12) follows by choosing γ that gives the tightest bound and $P_{Y|X}$ that gives the weakest in order to obtain a code-independent converse. ■

Corollary 1. *Any* (M,d,ϵ) *code must satisfy*

$$\epsilon \ge \sup_{\gamma \ge 0}\left\{\mathbb{E}\left[\min_{y\in\mathcal{Y}}\mathbb{P}\left[\tilde{\jmath}_{S,X}(S,X,y,d) \ge \log M + \gamma | X\right]\right] - \exp(-\gamma)\right\} \tag{22}$$

Proof: We weaken (12) using

$$\inf_{P_{Y|X}}\mathbb{P}\left[\tilde{\jmath}_{S,X}(S,X,Y,d) \ge \log M + \gamma\right] = \mathbb{E}\left[\inf_{P_{Y|X}}\mathbb{P}\left[\tilde{\jmath}_{S,X}(S,X,Y,d) \ge \log M + \gamma | X\right]\right] \tag{23}$$

$$= \mathbb{E}\left[\inf_{y\in\mathcal{Y}}\mathbb{P}\left[\tilde{\jmath}_{S,X}(S,X,y,d) \ge \log M + \gamma | X\right]\right] \tag{24}$$

where we used $S - X - Y$. ■

Remark 1. For the asymptotically equivalent noiseless source coding problem, we know that almost surely [9], [11] ,

$$\jmath_X(x,d) = \imath_{X;Y^\star}(x;Y^\star) + \lambda^\star \bar{\mathsf{d}}(x,Y^\star) - \lambda^\star d \tag{25}$$

where $\jmath_X(x,d)$ is the d-tilted information in $x \in \mathcal{X}$, the intuitive meaning of which is the number of bits required to represent x within distortion d in the conventional noiseless observations setting. Using (25), we may express (9) for all $y \in \text{supp}(P_{Y^\star})$ as

$$\tilde{\jmath}_{S,X}(s,x,y,d) = \jmath_X(x,d) + \lambda^\star \mathsf{d}(s,y) - \lambda^\star \bar{\mathsf{d}}(x,y) \tag{26}$$

Remark 2. If $\text{supp}(P_{Y^\star}) = \mathcal{Y}$ and $P_{X|S}$ is the identity mapping so that $\mathsf{d}(S,y) = \bar{\mathsf{d}}(X,y)$ almost surely, for every y, and Theorem 1 reduces to the noiseless converse in [9, Theorem 7].

Remark 3. Careful examination of the proof of Theorem 1 reveals that it is unnecessary to force the choice $P_{X|Y^\star}$ in (11); the result holds for an arbitrary $P_{X|Y}$. Moreover, $\lambda^\star$ can be replaced by any nonnegative scalar; that scalar can be chosen differently for each (s,x,y,d). Tuning those choices, it is possible to obtain a tighter finite blocklength bound, at the expense of a more cumbersome expression.

Theorem 2 (Achievability)**.** *There exists an* (M,d,ϵ) *code with*

$$\epsilon \le \inf_{P_Y}\int_0^1 \mathbb{E}\left[\left(1 - P_Y(\tilde{B}_t(X))\right)^M\right]dt \tag{27}$$

where

$$\tilde{B}_t(x) = \{y \in \mathcal{Y}\colon \pi(x,y) \le t\} \tag{28}$$

$$\pi(x,y) = \mathbb{P}\left[\mathsf{d}(S,y) > d | X = x\right] \tag{29}$$

Proof: The proof appeals to a random coding argument. Given M codewords $(c_1,\ldots,c_M)$, the encoder f and decoder c achieving minimum excess distortion probability attainable with the given codebook operate as follows. Having observed $x \in \mathcal{X}$, the optimum encoder chooses

$$i^\star \in \arg\min_i \pi(x,c_i) \tag{30}$$

with ties broken arbitrarily, so $\mathsf{f}(x) = i^\star$ and the decoder simply outputs $c_{i^\star}$, so $\mathsf{c}(\mathsf{f}(x)) = c_{i^\star}$.

The excess distortion probability achieved by the scheme is given by

$$\mathbb{P}\left[d(S,\mathsf{c}(\mathsf{f}(X))) > d\right] = \mathbb{E}\left[\pi(X,\mathsf{c}(\mathsf{f}(X)))\right] \tag{31}$$

$$= \int_0^1 \mathbb{P}\left[\pi(X,\mathsf{c}(\mathsf{f}(X))) > t\right]dt \tag{32}$$

$$= \int_0^1 \mathbb{E}\left[\mathbb{P}\left[\pi(X,\mathsf{c}(\mathsf{f}(X))) > t | X\right]\right]dt \tag{33}$$

Now, we notice that

$$1\left\{\pi(x,\mathsf{c}(\mathsf{f}(x))) > t\right\} = 1\left\{\min_{i\in 1,\ldots,M}\pi(x,c_i) > t\right\} \tag{34}$$

$$= \prod_{i=1}^M 1\left\{\pi(x,c_i) > t\right\} \tag{35}$$

and we average (33) with respect to the codewords $Y_1,\ldots,Y_M$ drawn i.i.d. from P_Y, independently of any other random variable, so that $P_{XY_1\ldots Y_M} = P_X \times P_Y \times \ldots \times P_Y$, to obtain

$$\int_0^1 \mathbb{E}\left[\prod_{i=1}^M \mathbb{P}\left[\pi(X,Y_M) > t | X\right]\right]dt \tag{36}$$

$$= \int_0^1 \mathbb{E}\left[\mathbb{P}^M\left[\pi(X,\bar{Y}) > t | X\right]\right]dt \tag{37}$$

where $P_{X\bar{Y}} = P_X P_Y$. Since there must exist a codebook achieving the excess distortion probability below or equal to the average over codebooks, (27) follows. ■

Remark 4. Notice that we have actually shown that the right-hand side of (27) gives the exact minimum excess distortion probability of random coding, averaged over codebooks drawn i.i.d. from P_Y.

Remark 5. In the noiseless case, $S = X$, for all $t \in [0,1)$,

$$\pi(x,y) = 1\left\{\mathsf{d}(x,y) > d\right\} \tag{38}$$

Therefore

$$\tilde{B}_t(x) = B_d(x) = \{y \in \mathcal{Y}\colon \mathsf{d}(x,y) \le d\} \tag{39}$$

and the bound in Theorem 2 reduces to the noiseless random coding bound in [9, Theorem 10].

The bound in (27) can be weakened to obtain the following result, which generalizes Shannon's bound for noiseless lossy compression (see e.g. [9, Theorem 1]).

Corollary 2. *There exists an* (M, d, ϵ) *code with*

$$\epsilon \leq \inf_{\gamma \geq 0,\, P_{Y|X}} \Big\{ \mathbb{P}\left[d(S,Y) > d\right] + \mathbb{P}\left[\imath_{X;Y}(X;Y) > \log M - \gamma\right] + e^{-\exp(\gamma)} \Big\} \quad (40)$$

where $P_{SXY} = P_S P_{X|S} P_{Y|X}$.

Proof: Fix $\gamma \geq 0$ and transition probability kernel $P_{Y|X}$. Let $P_X \to P_{Y|X} \to P_Y$ (i.e. P_Y is the marginal of $P_X P_{Y|X}$), and let $P_{X\bar{Y}} = P_X P_Y$. We use the nonasymptotic covering lemma [12, Lemma 5] to establish

$$\mathbb{E}\left[\mathbb{P}^M\left[\pi(X,\bar{Y}) > t | X\right]\right] \leq \mathbb{P}\left[\pi(X,Y) > t\right] + \mathbb{P}\left[\imath_{X;Y}(X;Y) > \log M - \gamma\right] + e^{-\exp(\gamma)} \quad (41)$$

Applying (41) to (37) and noticing that

$$\int_0^t \mathbb{P}\left[\pi(X,Y) > t\right] = \mathbb{E}\left[\pi(X,Y)\right] dt \quad (42)$$

$$= \mathbb{P}\left[\mathsf{d}(S,Y) > d\right] \quad (43)$$

we obtain (40). ∎

III. Asymptotic analysis

In this section, we pass from the single shot setup of Section II to a block setting by letting the alphabets be Cartesian products $\mathcal{M} = \mathcal{S}^k$, $\mathcal{X} = \mathcal{A}^k$, $\mathcal{Y} = \mathcal{B}^k$, and we study the second order asymptotics in k of $M^\star(k,d,\epsilon)$, the minimum achievable number of representation points compatible with the excess distortion constraint $\mathbb{P}\left[\mathsf{d}(S^k, Y^k) > d\right] \leq \epsilon$. We make the following assumptions.

(i) $P_{S^k X^k} = P_{\mathsf{S}} P_{\mathsf{X}|\mathsf{S}} \times \ldots \times P_{\mathsf{S}} P_{\mathsf{X}|\mathsf{S}}$ and

$$\mathsf{d}(s^k, y^k) = \frac{1}{k}\sum_{i=1}^k \mathsf{d}(s_i, y_i) \quad (44)$$

(ii) The alphabets $\mathcal{S}$, $\mathcal{A}$, $\mathcal{B}$ are finite sets.

(iii) The distortion level satisfies $d_{\min} < d < d_{\max}$, where

$$d_{\min} = \inf\{d \colon\ R(d) < \infty\} \quad (45)$$

and $d_{\max} = \inf_{\mathsf{y} \in \mathcal{B}} \mathbb{E}\left[d(\mathsf{X}, \mathsf{y})\right]$, where the expectation is with respect to the unconditional distribution of X.

The following result is obtained via a technical second order analysis (omitted due to space constraints) of Theorem 2 and the generalized Corollary 1 (according to Remark 3).

Theorem 3 (Gaussian approximation)**.** *For* $0 < \epsilon < 1$,

$$\log M^\star(k,d,\epsilon) = kR(d) + \sqrt{k\tilde{\mathcal{V}}(d)}Q^{-1}(\epsilon) + o\left(\sqrt{k}\right) \quad (46)$$

$$\tilde{\mathcal{V}}(d) = \mathrm{Var}\left(\mathbb{E}\left[\jmath_{\mathsf{S};\mathsf{X}}(\mathsf{S},\mathsf{X},\mathsf{Y}^\star,d)|\mathsf{S},\mathsf{X}\right]\right) \quad (47)$$

where $P_{\mathsf{SXY}^\star} = P_{\mathsf{S}} P_{\mathsf{X}|\mathsf{S}} P_{\mathsf{Y}^\star|\mathsf{X}}$*, and* $P_{\mathsf{Y}^\star|\mathsf{X}}$ *achieves the infimum in* (2).

Remark 6. The rate-dispersion function of the asymptotically equivalent noiseless problem is given by [9]

$$\mathcal{V}(d) = \mathrm{Var}\left(\jmath_{\mathsf{X}}(\mathsf{X},d)\right) \quad (48)$$

where $\jmath_{\mathsf{X}}(\mathsf{X},d)$ is defined in (25). To verify that the decomposition (6) indeed holds, which implies that $\tilde{\mathcal{V}}(d) > \mathcal{V}(d)$ unless there is no noise, write

$$\tilde{\mathcal{V}}(d) = \mathrm{Var}\left(\jmath_{\mathsf{X}}(\mathsf{X},d) + \lambda^\star \mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Y}^\star) - \bar{\mathsf{d}}(\mathsf{X},\mathsf{Y}^\star)|\mathsf{S},\mathsf{X}\right]\right) \quad (49)$$

$$= \mathrm{Var}\left(\jmath_{\mathsf{X}}(\mathsf{X},d)\right) + \lambda^{\star 2}\mathrm{Var}\left(\mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Y}^\star) - \bar{\mathsf{d}}(\mathsf{X},\mathsf{Y}^\star)|\mathsf{S},\mathsf{X}\right]\right) + 2\lambda^\star \mathrm{Cov}\left(\jmath_{\mathsf{X}}(\mathsf{X},d), \mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Y}^\star) - \bar{\mathsf{d}}(\mathsf{X},\mathsf{Y}^\star)|\mathsf{S},\mathsf{X}\right]\right) \quad (50)$$

where the covariance is zero:

$$\mathbb{E}\left[\left(\jmath_{\mathsf{X}}(\mathsf{X},d) - R(d)\right)\mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Y}^\star) - \bar{\mathsf{d}}(\mathsf{X},\mathsf{Y}^\star)|\mathsf{S},\mathsf{X}\right]\right]$$

$$= \mathbb{E}\left[\jmath_{\mathsf{X}}(\mathsf{X},d)\mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Y}^\star) - \bar{\mathsf{d}}(\mathsf{X},\mathsf{Y}^\star)|\mathsf{S},\mathsf{X}\right]\right] \quad (51)$$

$$= \mathbb{E}\left[\jmath_{\mathsf{X}}(\mathsf{X},d)\mathbb{E}\left[\mathsf{d}(\mathsf{S},\mathsf{Y}^\star) - \bar{\mathsf{d}}(\mathsf{X},\mathsf{Y}^\star)|\mathsf{X}\right]\right] \quad (52)$$

$$= 0 \quad (53)$$

IV. Example

A. Erased fair coin flips

Let a binary equiprobable source be observed by the encoder through a binary erasure channel with erasure rate δ. The goal is to minimize the bit error rate with respect to the source. For $\frac{\delta}{2} \leq d \leq \frac{1}{2}$, the rate-distortion function is given by

$$R(d) = (1-\delta)\left(\log 2 - h\left(\frac{d - \frac{\delta}{2}}{1-\delta}\right)\right) \quad (54)$$

where $h(\cdot)$ is the binary entropy function, and (54) is obtained by solving the optimization in (2) which is achieved by

$$P^\star_{\mathsf{X}|\mathsf{Y}}(a|b) = \begin{cases} 1 - d - \frac{\delta}{2} & b = a \\ d - \frac{\delta}{2} & b \neq a \neq ? \\ \delta & a = ? \end{cases} \quad (55)$$

where $a \in \{0,1,?\}$ and $b \in \{0,1\}$, and $P^\star_{\mathsf{Y}}(0) = P^\star_{\mathsf{Y}}(1) = \frac{1}{2}$.

$$\jmath_{\mathsf{S},\mathsf{X}}(\mathsf{S},\mathsf{X},b,d) = \imath_{\mathsf{X};\mathsf{Y}^\star}(\mathsf{X},b) + \lambda^\star d(\mathsf{S},b) - \lambda^\star d \quad (56)$$

$$= -\lambda^\star d + \begin{cases} \log\frac{2}{1+\exp(-\lambda^\star)} & \text{w.p. } 1-\delta \\ \lambda^\star & \text{w.p. } \frac{\delta}{2} \\ 0 & \text{w.p. } \frac{\delta}{2} \end{cases} \quad (57)$$

The rate-dispersion function is given by the variance of (57):

$$\tilde{\mathcal{V}}(d) = \delta(1-\delta)\log^2\cosh\left(\frac{\lambda^\star}{2\log e}\right) + \frac{\delta}{4}\lambda^{\star 2} \quad (58)$$

$$\lambda^\star = -R'(d) = \log\frac{1 - \frac{\delta}{2} - d}{d - \frac{\delta}{2}} \quad (59)$$

Note that as d approaches $\frac{\delta}{2}$, the rate-dispersion function grows without limit. This is to be expected, because for $d = \frac{\delta}{2}$, even in the hypothetical case in which the decoder knew X^k, the number of erroneously represented bits is binomially distributed with mean $k\frac{\delta}{2}$, so no code can achieve probability of distortion exceeding $d = \frac{\delta}{2}$ lower than $\frac{1}{2}$. Therefore, the validity of (46) for $\epsilon < \frac{1}{2}$ requires $\tilde{V}(\delta/2) = \infty$. For $d = \frac{\delta}{2}$, codes with rate approaching the rate-distortion function were constructed in [13].

Bounds to the minimum achievable rate exploiting the symmetry of the erased coin flips setting were shown in [9].

B. Erased fair coin flips: asymptotically equivalent problem

According to (3), the distortion measure of the asymptotically equivalent noiseless problem is given by $\bar{\mathsf{d}}(1,1) = \bar{\mathsf{d}}(0,0) = 0$, $\bar{\mathsf{d}}(1,0) = \bar{\mathsf{d}}(0,1) = 1$, $\bar{\mathsf{d}}(?,1) = \bar{\mathsf{d}}(?,0) = \frac{1}{2}$. According to (26), the d-tilted information can be obtained by taking the expectation of (57) with respect to S:

$$\jmath_{\mathsf{X}}(\mathsf{X}, d) = -\lambda^\star d + \begin{cases} \log \frac{2}{1+\exp(-\lambda^\star)} & \text{w.p. } 1-\delta \\ \frac{\lambda^\star}{2} & \text{w.p. } \delta \end{cases} \tag{60}$$

Its variance is equal to

$$\mathcal{V}(d) = \delta(1-\delta)\log^2 \cosh\left(\frac{\lambda^\star}{2\log e}\right) \tag{61}$$

For convenience, denote

$$\left\langle {k \atop j} \right\rangle = \sum_{i=0}^{j} \binom{k}{i} \tag{62}$$

with the convention $\left\langle {k \atop j} \right\rangle = 0$ if $j < 0$ and $\left\langle {k \atop j} \right\rangle = \left\langle {k \atop k} \right\rangle$ if $j > k$.

The following tight achievability and converse bounds are obtained similarly to [9, Theorem 32] and [9, Theorem 33].

Theorem 4 (asymptotically equivalent erased coin flips)**.**

$$\sum_{j=0}^{\lfloor 2kd \rfloor} \binom{k}{j} \delta^j (1-\delta)^{k-j} \left[1 - \frac{M^\star(k,d,\epsilon)}{2^{k-j}} \left\langle {k-j \atop \lfloor kd - \frac{1}{2}j \rfloor} \right\rangle\right]^+$$
$$\leq \epsilon \tag{63}$$
$$\leq \sum_{j=0}^{k} \binom{k}{j} \delta^j (1-\delta)^{k-j} \left(1 - \frac{1}{2^{k-j}} \left\langle {k-j \atop \lfloor kd - \frac{1}{2}j \rfloor} \right\rangle\right)^{M^\star(k,d,\epsilon)} \tag{64}$$

The bounds in [9, Theorem 32], [9, Theorem 33] and the approximation in Theorem 3 (with the remainder term taken to be equal to 0), as well as the bounds in (63) and (64) for the asymptotically equivalent problem together with their Gaussian approximation, are plotted in Fig. 2. We note the following.

- The achievability and converse bounds are extremely tight.
- Despite the fact that the asymptotically achievable rate in the two problems is the same, we would be making a gross approximation error if we were to use the finite-blocklength bounds and dispersion approximation for the asymptotically equivalent problem in lieu of the results of this paper. For example, at blocklength 1000, the penalty over the rate-distortion function is 9% for erased coin flips and only 4% for the asymptotically equivalent problem.

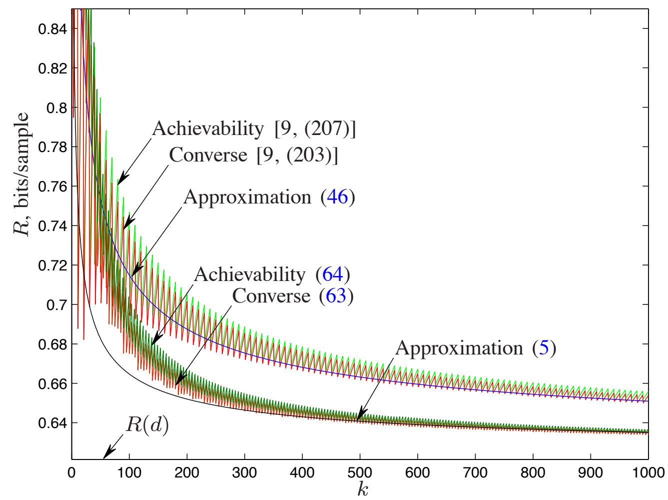

Fig. 2. The minimum achievable rate for the fair binary source observed through an erasure channel, as well as that for the asymptotically equivalent problem, with $\delta = 0.1$, $d = 0.1$, $\epsilon = 0.1$.

References

[1] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293 –304, Sep. 1962.

[2] T. Berger, *Rate distortion theory*. Prentice-Hall Englewood Cliffs, NJ, 1971.

[3] H. S. Witsenhausen, "Indirect rate distortion problems," *IEEE Transactions on Information Theory*, vol. 26, no. 5, pp. 518–521, 1980.

[4] D. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 165–167, 1968.

[5] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Transactions on Information Theory*, vol. 16, no. 4, pp. 406–411, 1970.

[6] E. Ayanoglu, "On optimal quantization of noisy sources," *IEEE Transactions on Information Theory*, vol. 36, no. 6, pp. 1450–1452, 1990.

[7] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 826–834, 1988.

[8] T. Fischer, J. Gibson, and B. Koo, "Estimation and noisy source coding," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 1, pp. 23–34, 1990.

[9] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.

[10] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *Data Compression Conference (DCC)*, Snowbird, UT, Mar. 2011, pp. 53–62.

[11] I. Csiszár, "On an extremum problem of information theory," *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, no. 1, pp. 57–71, Jan. 1974.

[12] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Allerton, IL: IEEE, Oct. 2012, pp. 1–8.

[13] E. Martinian and J. Yedidia, "Iterative quantization using codes on graphs," *Proceedings of the 41st Annual Allerton Conference on Communication, Control, and Computing; Monticello, IL*, Sep. 2004.