A Generalization of the Rate-Distortion Function for Wyner-Ziv Coding of Noisy Sources in the Quadratic-Gaussian Case*

David Rebollo-Monedero and Bernd Girod
Information Systems Lab., Electrical Eng. Dept.
Stanford University, Stanford, CA 94305
{drebollo,bgirod}@stanford.edu

Abstract

We extend the rate-distortion function for Wyner-Ziv coding of noisy sources with quadratic distortion, in the jointly Gaussian case, to more general statistics. It suffices that the noisy observation Z be the sum of a function of the side information Y and independent Gaussian noise, while the source data X must be the sum of a function of Y, a linear function of Z, and a random variable N such that the conditional expectation of N given Y and Z is zero, almost surely. Furthermore, the side information Y may be arbitrarily distributed in any alphabet, discrete or continuous. Under these general conditions, we prove that no rate loss is incurred due to the unavailability of the side information at the encoder. In the noiseless Wyner-Ziv case, i. e., when the source data is directly observed, the assumptions are still less restrictive than those recently established in the literature. We confirm, theoretically and experimentally, the consistency of this analysis with some of the main results on high-rate Wyner-Ziv quantization of noisy sources.

1 Introduction

In numerous data compression applications, an indirect observation of some source data, for instance, an image corrupted by noise, is to be encoded and transmitted to a decoder. The decoder has access to some local side information, for example, a previously decoded image. The side information is not available at the encoder but the statistical dependence among the source data, the noisy observation and the side information is known, and in principle it may be exploited in the design of both the encoder and the decoder, in order to improve the rate-distortion performance. Intuitively, if the side information were available to the encoder, it could be jointly used with the noisy observation to produce an estimate of the original data. In most cases, reducing the noise would decrease the number of bits required for transmission, for a given distortion. In addition, the dependence with the side information, known

*This work was supported by NSF under Grant No. CCR-0310376.

also at the decoder, could be exploited to further reduce the bit rate. However, supposing that the side information is *not* available at the encoder, we wish to know whether it is possible to achieve the same rate-distortion performance.

The information-theoretic rate-distortion bounds for memoryless coding of directly observed data with side information at the decoder, also called Wyner-Ziv (WZ) coding, were established in [1,2]. Furthermore, the same work showed that if the source data and the side information are jointly Gaussian sequences, and mean-squared error is used as a distortion measure, then no rate loss is incurred by not having access to the side information at the encoder, and a closed-form expression for the rate-distortion function was provided. Using an argument based on the duality between source and channel coding, [3] recently generalized the absence of rate loss and the rate-distortion formula in the case of quadratic distortion, requiring only that the source data be the sum of arbitrarily distributed side information and independent, zero-mean Gaussian noise.

The earliest one-letter characterization of the rate-distortion function for *noisy* WZ coding, i. e., lossy coding of noisy observations with side information at the decoder, appeared in [4]. It was also proved that if the source data, the noisy observation and the side information are jointly Gaussian, and the distortion is quadratic, then no rate loss is incurred, and a closed formula was given for the rate-distortion function. Independent, similar work was carried out shortly afterwards [5], and more recently [6]. A method to reduce rate-distortion problems with noisy sources (and also noisy reconstructions) to noiseless cases by using modified distortion functions was presented in [7]. However, the reduction of the general noisy WZ problem requires side-information-dependent distortion functions, which were studied later [8,9].

In this paper, we extend the validity of the rate-distortion function for noisy WZ coding with quadratic distortion in the jointly Gaussian case to much more general statistics, proving that no rate loss is incurred. In particular, no restriction on the alphabet or distribution of the side information is made. In the noiseless, direct case, the requirements are still more general than those established in [3].

Sec. 2 contains the definitions and fundamental results used in the theoretic analysis, presented in Sec. 3. A noisy WZ coding problem is investigated experimentally in Sec. 4, confirming our main theoretic result.

2 Definitions and preliminaries

Throughout the paper, the measurable space in which a random variable (r. v.) takes values will be called alphabet. We shall follow the convention of using uppercase letters for r. v., lowercase letters for particular values they take on, and uppercase script letters for their alphabets. As usual, a. e. abbreviates 'almost everywhere' or 'almost every' with respect to an underlying probability measure.

Let X, Y and Z be r. v. defined on a common probability space, taking values in arbitrary alphabets $\mathscr{X}$, $\mathscr{Y}$ and $\mathscr{Z}$, respectively. Let $((X_i, Y_i, Z_i))_{i\in\mathbb{Z}^+}$ be a sequence of independent, identically distributed drawings of (X, Y, Z). A distortion function is a measurable function $d : \mathscr{X} \times \hat{\mathscr{X}} \to [0, \infty) \subset \mathbb{R}$, where $\hat{\mathscr{X}}$ is a measurable space.

For all positive integers n, M and non-negative real Δ, a (n, M, Δ)-code is defined by two measurable mappings, namely an encoding or quantization function $q : \mathscr{Z}^n \to \{1, \ldots, M\}$, and a decoding or reconstruction function $\hat{x}^n : \{1, \ldots, M\} \times \mathscr{Y}^n \to \hat{\mathscr{X}}^n$, with associated distortion per sample $\Delta = \mathrm{E}\, \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i)$. $Q = q(Z^n)$ represents a quantization index in $\{1, \ldots, M\}$, and $\hat{X}^n = \hat{x}^n(Q, Y^n)$ a block of reconstructed values with distortion Δ with respect to the encoded block X^n. Fig. 1 depicts a noisy WZ coder. A pair of non-negative real numbers $(\mathcal{R}, \mathcal{D})$, representing a rate and a distortion respectively, is said to be achievable if for every positive real ϵ there exists a code with arbitrarily large n, such that $M \leqslant 2^{n(\mathcal{R}+\epsilon)}$ and $\Delta \leqslant \mathcal{D} + \epsilon$. For $\mathcal{D} \geqslant 0$, the WZ rate-distortion function for noisy sources is defined as the extended real-valued function $\mathcal{R}^{\mathrm{NWZ}}_{XZ|Y}(\mathcal{D}) = \inf\{\mathcal{R} \,|\, (\mathcal{R}, \mathcal{D}) \text{ achievable}\}$, with the convention $\inf \varnothing = \infty$ [(a)]. Under mild technical conditions, [4] provides the following one-letter characterization:

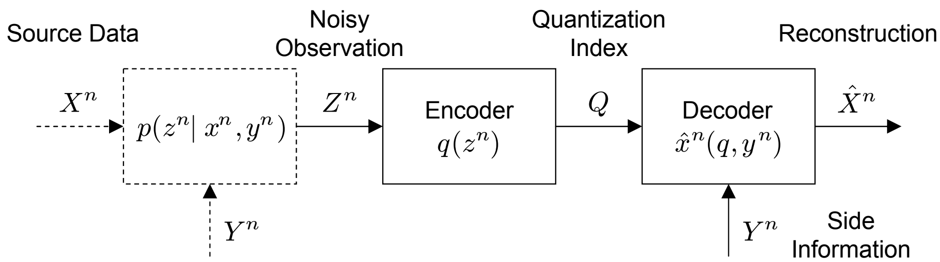

Figure 1: Noisy WZ coder.

$$\mathcal{R}^{\mathrm{NWZ}}_{XZ|Y}(\mathcal{D}) = \inf_{\substack{Q,\, \hat{x}(q,y) \\ (X,Y) \leftrightarrow Z \leftrightarrow Q \\ \mathrm{E}\, d(X, \hat{X}) \leqslant \mathcal{D}}} \mathrm{I}(Z; Q | Y), \tag{1}$$

where the infimum is taken over all r. v. Q, representing a quantization index, in an arbitrary alphabet $\mathscr{Q}$, such that Q and (X, Y) are conditionally independent given Z, and over all measurable reconstruction functions $\hat{x} : \mathscr{Q} \times \mathscr{Y} \to \hat{\mathscr{X}}$, subject to the constraint $\mathrm{E}\, d(X, \hat{X}) \leq \mathcal{D}$.

The case in which the side information is available at the encoder will be referred to as noisy conditional coding, and the corresponding rate-distortion function denoted by $\mathcal{R}^{\mathrm{N}}_{XZ|Y}(\mathcal{D})$. The one-letter characterization is the same as (1), with $X \leftrightarrow (Y, Z) \leftrightarrow Q$ in place of $(X, Y) \leftrightarrow Z \leftrightarrow Q$. The rate-distortion functions for the WZ and the conditional cases when the source data is directly observed, i. e., $Z = X$, will be denoted by $\mathcal{R}^{\mathrm{WZ}}_{X|Y}(\mathcal{D})$ and $\mathcal{R}_{X|Y}(\mathcal{D})$, respectively. Define $\log^+ x = \log x$ for all $x \geqslant 1$, and $\log^+ x = 0$ for all $x \in [0, 1)$.

3 Theoretic Results

We start by proving in Proposition 1 that conditional Gaussian statistics with arbitrarily distributed side information maximize the conditional differential entropy

[(a)] It can be shown that for a given $\mathcal{D}$, the set of achievable pairs is closed, thus when not empty, inf can be replaced by min.

with an *unconditional* quadratic distortion constraint. This result will be used in the noiseless WZ rate-distortion functions of Theorem 2.

Proposition 1 *Let $\mathcal{D} > 0$, and let X be a real-valued r. v. If $X|\,y \sim \mathcal{N}(0, \mathcal{D})$ for a. e. $y \in \mathscr{Y}$, then X maximizes $\mathrm{h}(X|\,Y)$ subject to the constraint $\mathrm{E}\, X^2 \leqslant \mathcal{D}$.*

Proof: Since we are interested in maximizing $\mathrm{h}(X|Y)$, without loss of generality assume that for all $y \in \mathscr{Y}$, $\mathrm{h}(X|\,y) > -\infty$. Then,

$$\mathrm{h}(X|\,Y) + \mathrm{D}(X|\,Y \| X'|\,Y) = \mathrm{E}_{XY} \log \frac{1}{p_{X'|\,Y}(X|\,Y)}.$$

Set $X'|\,y \sim \mathcal{N}(0, \mathcal{D})$ for all $y \in \mathscr{Y}$. Since $\mathrm{E}\, X^2 \leqslant \mathcal{D}$,

$$\mathrm{h}(X|\,Y) + \mathrm{D}(X|\,Y \| X'|\,Y) = \frac{1}{2} \log(2\pi\mathcal{D}) + \log \mathrm{e}\, \frac{1}{2} \frac{\mathrm{E}\, X^2}{\mathcal{D}} \leqslant \frac{1}{2} \log(2\pi\mathrm{e}\, \mathcal{D}).$$

Finally, observe that $\mathrm{D}(X|\,Y \| X'|\,Y) \geqslant 0$, with equality if (and only if) for a. e. $y \in \mathscr{Y}$, $X|\,y$ and $X'|\,y$ are equal in distribution.

Alternatively, this can be proven from the analogous result for non-distributed coding, using the inequalities $\mathrm{h}(X|\,Y) \leq \mathrm{h}(X) \leq \frac{1}{2} \log(2\pi\mathrm{e}\, \mathcal{D})$. ■

The following theorem presents an extension of the WZ rate-distortion function in the quadratic-Gaussian case, which slightly relaxes the hypotheses required in [3], together with a direct proof that does not use duality arguments. Observe that the hypothesis $X|\,y \sim \mathcal{N}(\mu(y), \sigma^2)$ in the theorem is equivalent to $X \overset{\text{a.e.}}{=} \mu(Y) + N$, with $N \sim \mathcal{N}(0, \sigma^2)$, independent from Y. That is, the source data X is a noisy version of any real-valued measurable function μ of the side information $Y \in \mathscr{Y}$, where $\mathscr{Y}$ is an arbitrary measurable space, and the noise is additive, Gaussian and independent from Y.

Theorem 2 *Let $\mathscr{X} = \hat{\mathscr{X}} = \mathbb{R}$, $d(x, \hat{x}) = (x - \hat{x})^2$, $\mu : \mathscr{Y} \to \mathbb{R}$ measurable, and $\sigma^2 > 0$. Suppose that for a. e. $y \in \mathscr{Y}$, $X|\,y \sim \mathcal{N}(\mu(y), \sigma^2)$. Then, for all $\mathcal{D} > 0$,*

$$\mathcal{R}^{\mathrm{WZ}}_{X|\,Y}(\mathcal{D}) = \mathcal{R}_{X|\,Y}(\mathcal{D}) = \frac{1}{2} \log^+ \frac{\sigma^2}{\mathcal{D}}.$$

Proof: We prove $\mathcal{R}^{\mathrm{WZ}}_{X|\,Y}(\mathcal{D}) = \frac{1}{2} \log^+ \frac{\sigma^2}{\mathcal{D}}$. Eliminating the WZ constraint $Y \leftrightarrow X \leftrightarrow Q$ makes the proof valid for $\mathcal{R}_{X|\,Y}(\mathcal{D})$. Since $\hat{X} = \hat{x}(Q, Y)$ and $\mathrm{h}(X|Y)$ is finite,

$$\begin{aligned} \mathrm{I}(X; Q|\,Y) = \mathrm{h}(X|\,Y) - \mathrm{h}(X|\,Q, Y) = \\ = \mathrm{h}(X|\,Y) - \mathrm{h}(X - \hat{X}|\,Q, Y) \geqslant \mathrm{h}(X|\,Y) - \mathrm{h}(X - \hat{X}|\,Y), \end{aligned}$$

with equality if and only if $X - \hat{X} \leftrightarrow Y \leftrightarrow Q$. Proposition 1 implies

$$\begin{aligned} \mathcal{R}^{\mathrm{WZ}}_{X|\,Y}(\mathcal{D}) \geqslant \mathrm{h}(X|\,Y) - \sup_{\substack{Q,\, \hat{x}(q,y) \\ Y \leftrightarrow X \leftrightarrow Q \\ \mathrm{E}\,(X - \hat{X})^2 \leqslant \mathcal{D}}} \mathrm{h}(X - \hat{X}|\,Y) \geqslant h(X|\,Y) - \frac{1}{2} \log(2\pi\mathrm{e}\, \mathcal{D}) = \\ = \frac{1}{2} \log(2\pi\mathrm{e}\, \sigma^2) - \frac{1}{2} \log(2\pi\mathrm{e}\, \mathcal{D}) = \frac{1}{2} \log \frac{\sigma^2}{\mathcal{D}}. \end{aligned}$$

To complete the proof, it suffices to find Q and $\hat{X}$ such that the inequalities in the previous derivation hold with equality, and the constraints in the supremum are met. Precisely, we need $\hat{X} = \hat{x}(Q, Y)$, $Y \leftrightarrow X \leftrightarrow Q$, $X - \hat{X} \leftrightarrow Y \leftrightarrow Q$, and $X - \hat{X} | \, y \sim \mathcal{N}(0, \mathcal{D})$ for a. e. y. Define $d = \mathcal{D}/\sigma^2$. The case $d \geq 1$ is trivial. For all $d \in (0, 1)$ set

$$\forall x, y \quad Q | \, x, y = Q | \, x \sim \mathcal{N}((1 - d)x, \, d(1 - d)\sigma^2),$$

and $\hat{X} = Q + d\,\mu(Y)$ (well defined since μ is measurable), which by definition satisfy $Y \leftrightarrow X \leftrightarrow Q$ and $\hat{X} = \hat{x}(Q, Y)$. Clearly,

$$\hat{X} | \, x, y \sim \mathcal{N}((1 - d)x + d\,\mu(y), \, d(1 - d)\sigma^2).$$

It is easy to check that $\hat{X} | \, y \sim \mathcal{N}(\mu(y), \, (1 - d)\sigma^2)$ and $X - \hat{X} | \, y \sim \mathcal{N}(0, d\,\sigma^2)$ for a. e. y (the associated test channel is represented in Fig. 2). Observe that for a. e. y, $\hat{X} | \, y$ and $X - \hat{X} | \, y$ are Gaussian and their variances add up to the variance of their sum, $X | \, y$. This implies that $\hat{X}$ and $X - \hat{X}$ are conditionally independent given Y,

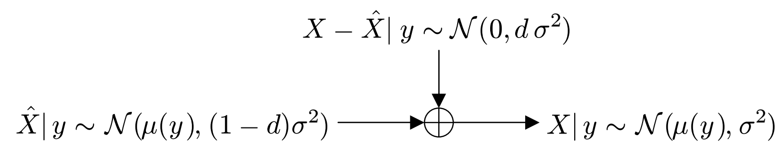

Figure 2: Test channel for the proof of Theorem 2.

thus by the definition of Q, $X - \hat{X} \leftrightarrow Y \leftrightarrow Q$. ■

The extension of the noisy WZ rate-distortion function for the quadratic-Gaussian case to more general conditions, presented in Theorem 5, will make use of Propositions 3 and 4, on side-information-dependent and modified distortion functions.

Proposition 3 *Let $\mathscr{X} = \hat{\mathscr{X}} = \mathbb{R}$. Consider the side-information-dependent*[(b)] *distortion function $d(x, \hat{x}, y) = (\alpha\, x - \hat{x} + f(y))^2$, with $\alpha \in \mathbb{R}$ and $f : \mathscr{Y} \to \mathbb{R}$ measurable, and let $\mathcal{R}^{\mathrm{WZ}}_{X|Y}$ denote the corresponding WZ rate-distortion function. If $\tilde{\mathcal{R}}^{\mathrm{WZ}}_{X|Y}$ is the WZ rate-distortion function corresponding to the quadratic distortion function $\tilde{d}(x, \hat{x}) = (x - \hat{x})^2$, then*

$$\mathcal{R}^{\mathrm{WZ}}_{X|Y}(\mathcal{D}) = \begin{cases} 0 & \text{if} \quad \alpha = 0 \\ \tilde{\mathcal{R}}^{\mathrm{WZ}}_{X|Y}(\frac{\mathcal{D}}{\alpha^2}) & \text{if} \quad \alpha \neq 0 \end{cases}.$$

Furthermore, the analogous result also holds for the conditional rate-distortion function $\mathcal{R}_{X|Y}$ and the quadratic version $\tilde{\mathcal{R}}_{X|Y}$.

Proof: The distortion per sample of a (n, M, Δ)-code is

$$\Delta = \mathrm{E}\, \frac{1}{n} \sum_{i=1}^{n} d(X_i, \hat{X}_i, Y_i) = \mathrm{E}\, \frac{1}{n} \sum_{i=1}^{n} \Big(\alpha\, X_i - \hat{X}_i + f(Y_i)\Big)^2.$$

[(b)] In the (operational) definition of the WZ rate-distortion function, replace $d(x, \hat{x})$ by $d(x, \hat{x}, y)$. It can be shown that the one-letter characterization of the WZ rate-distortion function for arbitrary alphabets is still valid in the case in which the distortion function d is allowed to depend on the side information [8, 9]. In fact, the proof is a straightforward modification of the classical one [1, 2].

Define the measurable vector extension of f as $f(y^n) = (f(y_i))_{i=1}^n$. If $\alpha = 0$, set $M = 1$, $q(z^n) = 1$, and $\hat{x}(q, y^n) = f(y^n)$. Such code satisfies $\Delta = 0$, thus for all $\mathcal{D} \geqslant 0$, $\mathcal{R}^{\mathrm{WZ}}_{X|Y}(\mathcal{D}) = 0$. On the other hand, if $\alpha \neq 0$, define $\Delta' = \Delta/\alpha^2$ and $\hat{x}'(q, y^n) = (\hat{x}(q, y^n) - f(y^n))/\alpha$. We have

$$\Delta' = \frac{\Delta}{\alpha^2} = \mathrm{E}\,\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \frac{\hat{X}_i}{\alpha} + \frac{f(Y_i)}{\alpha}\right)^2 = \mathrm{E}\,\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{X}'_i\right)^2.$$

This shows that given $\mathcal{D} \geqslant 0$, for any $\epsilon > 0$, any code for the original, side-information-dependent WZ problem satisfying $\Delta \leqslant \mathcal{D} + \epsilon$ can be converted into a code for the quadratic WZ problem satisfying $\Delta' \leqslant \mathcal{D}/\alpha^2 + \epsilon/\alpha^2$, and conversely. Finally, observe that the above reasoning is true regardless of whether the encoder mapping depends on the side information, thus this also proves the proposition in the conditional coding case. ■

The following proposition, based on [7], can be used to reduce a noisy WZ problem with a very general form of distortion function to a WZ problem for a clean source with a side-information-dependent distortion function, and similarly in the conditional case. In particular, the one-letter characterization of the noisy WZ rate-distortion function (1) can be extended to distortion functions of the form $d(x, \hat{x}, y, z)$, simply by using the WZ theorem on clean sources and side-information-dependent distortion functions, and the modified distortion function of the proposition. The r. v. Z, which plays the role of the noisy observation of the clean source data in the original problem, represents the clean source data itself in the equivalent problem. The alternative Markov conditions in the proposition correspond to the WZ and conditional cases, respectively.

Proposition 4 *Let $d : \mathscr{X} \times \hat{\mathscr{X}} \times \mathscr{Y} \times \mathscr{Z} \to [0, \infty)$ be measurable. Define $\tilde{d}(z, \hat{x}, y) = \mathrm{E}[d(X, \hat{x}, y, z)|\, y, z]$. Let Q be a r. v. in some alphabet $\mathscr{Q}$. Assume that $(X, Y) \leftrightarrow Z \leftrightarrow Q$ or $X \leftrightarrow (Y, Z) \leftrightarrow Q$, and that there exists a measurable function $\hat{x} : \mathscr{Q} \times \mathscr{Y} \to \hat{\mathscr{X}}$ such that $\hat{X} = \hat{x}(Q, Y)$. Then, $\mathrm{E}\, d(X, \hat{X}, Y, Z) = \mathrm{E}\, \tilde{d}(Z, \hat{X}, Y)$.*

Proof: First, observe that for any independent r. v. $U \in \mathscr{U}$ and $V \in \mathscr{V}$, and any measurable function $g : \mathscr{U} \times \mathscr{V} \to \mathbb{R}$,

$$\mathrm{E}_{UV}\, g(U, V) = \mathrm{E}_U\, \mathrm{E}_{V|U}[g(U, V)|U] = \mathrm{E}_U\, [\mathrm{E}_V\, g(u, V)]_{u=U}. \qquad (2)$$

By assumption, $(X, Y) \leftrightarrow Z \leftrightarrow Q$ or $X \leftrightarrow (Y, Z) \leftrightarrow Q$, which implies $X \leftrightarrow (Y, Z) \leftrightarrow Q$ in either case. Therefore, $X \leftrightarrow (Y, Z) \leftrightarrow (Y, Q)$, and since $\hat{X}$ is a function of (Q, Y), then $X \leftrightarrow (Y, Z) \leftrightarrow \hat{X}$. Use the preliminary observation (2) and the conditional independence of X and $\hat{X}$ given (Y, Z) to obtain

$$\mathrm{E}[d(X, \hat{X}, Y, Z)|Y, Z] = \mathrm{E}\left[[\mathrm{E}[d(X, \hat{x}, Y, Z)|Y, Z]]_{\hat{x}=\hat{X}}\, |Y, Z\right] = \mathrm{E}[\tilde{d}(Z, \hat{X}, Y)|Y, Z].$$

Apply iterated expectation on (Y, Z) to complete the proof. ■

The next theorem extends Theorem 2 to the noisy case. Observe that condition (i) in Theorem 5 is equivalent to $Z \overset{\text{a.e.}}{=} \mu(Y) + N$, with $N \sim \mathcal{N}(0, \sigma^2)$, independent from Y, and any measurable $\mu : \mathscr{Y} \to \mathbb{R}$, similarly to Theorem 2. On the other hand, condition (ii) is equivalent to $X \stackrel{\text{a.e.}}{=} f(Y) + \alpha\, Z + N$, where $\alpha \in \mathbb{R}$, $f : \mathscr{Y} \to \mathbb{R}$ measurable, and N is a r.v. satisfying $\mathrm{E}[N|\, y, z] = 0$ for a.e. $y \in \mathscr{Y}$, $z \in \mathbb{R}$. The jointly Gaussian case is clearly a particular one.

Theorem 5 *Let $\mathscr{X} = \hat{\mathscr{X}} = \mathscr{Z} = \mathbb{R}$, $d(x, \hat{x}) = (x - \hat{x})^2$, $\mu : \mathscr{Y} \to \mathbb{R}$ measurable, and $\sigma^2 > 0$. Define $\mathcal{D}_\infty = \mathrm{E}_{YZ} \operatorname{Var}[X|\, Y, Z]$. Suppose that*

(i) *for a. e. $y \in \mathscr{Y}$, $Z|\, y \sim \mathcal{N}(\mu(y), \sigma^2)$, and*

(ii) *there exist $f : \mathscr{Y} \to \mathbb{R}$ measurable and $\alpha \in \mathbb{R}$ such that $\mathrm{E}[X|\, y, z] = f(y) + \alpha\, z$ for a. e. $y \in \mathscr{Y}$, $z \in \mathbb{R}$.*

Then, for all $\mathcal{D} > \mathcal{D}_\infty$,

$$\mathcal{R}^{\mathrm{NWZ}}_{XZ|Y}(\mathcal{D}) = \mathcal{R}^{\mathrm{N}}_{XZ|Y}(\mathcal{D}) = \frac{1}{2} \log^+ \frac{\alpha^2\, \sigma^2}{\mathcal{D} - \mathcal{D}_\infty}. \tag{3}$$

Proof: In both the WZ and the conditional case, the modified distortion function of Proposition 4 is given by

$$\begin{aligned}\tilde{d}(z, \hat{x}, y) &= \mathrm{E}[(X - \hat{x})^2|\, y, z] = \operatorname{Var}[X|\, y, z] + (\mathrm{E}[X|\, y, z] - \hat{x})^2 = \\ &= \operatorname{Var}[X|\, y, z] + (\alpha\, z - \hat{x} + f(y))^2 \quad \text{for all } \hat{x} \in \hat{\mathscr{X}},\ \text{for a.e. } y \in \mathscr{Y},\ z \in \mathscr{Z},\end{aligned}$$

thus the expected distortion is $\mathcal{D} = \mathcal{D}_\infty + \mathrm{E}\,(\alpha\, Z - \hat{X} + f(Y))^2$. The term $\mathcal{D}_\infty$ is a constant completely determined by the joint statistics of X, Y and Z, and consequently independent of any code we may choose. For all $\mathcal{D} > \mathcal{D}_\infty$ define $\bar{\mathcal{D}} = \mathcal{D} - \mathcal{D}_\infty > 0$ and consider the equivalent coding problem with distortion function $\bar{d}(z, \hat{x}, y) = (\alpha\, z - \hat{x} + f(y))^2$ and $\bar{\mathcal{D}} = \mathrm{E}\, \bar{d}(Z, \hat{X}, Y)$. This is precisely the coding problem of Proposition 3, where now Z plays the role of the clean source data X. Denote by $\mathcal{R}^{\mathrm{WZ}}_{Z|Y}$ the rate-distortion function corresponding to WZ coding of Z regarded as a clean source, with quadratic distortion. Since by assumption $Z|\, y \sim \mathcal{N}(\mu(y), \sigma^2)$, then, by Theorem 2, if $\alpha \neq 0$,

$$\mathcal{R}^{\mathrm{NWZ}}_{XZ|Y}(\mathcal{D}) = \mathcal{R}^{\mathrm{WZ}}_{Z|Y}\left(\frac{\bar{\mathcal{D}}}{\alpha^2}\right) = \frac{1}{2} \log^+ \frac{\alpha^2\, \sigma^2}{\bar{\mathcal{D}}},$$

and similarly for $\mathcal{R}^{\mathrm{N}}_{XZ|Y}$. Finally, if $\alpha = 0$, again by Proposition 3, since $\log^+ 0 = 0$ by definition, the formula for the rate-distortion function still holds. ■

At this point we confirm that Theorem 5 is consistent with the theory of high-rate WZ quantization of noisy sources presented in [10]. Define $\bar{x}(y, z) = \mathrm{E}[X|\, y, z]$ and $\bar{X} = \bar{x}(Y, Z)$. Clearly, $\bar{X}|y \sim \mathcal{N}(f(y) + \alpha\, \mu(y),\ \alpha^2\, \sigma^2)$ for a.e. y, thus $\mathrm{h}(\bar{X}|Y) = \frac{1}{2} \log(2\pi\mathrm{e}\, \alpha^2\, \sigma^2)$. Since $\bar{x}(y, z)$ is additively separable, the theorem of high-rate WZ quantization of a noisy source [10, Theorem 3] can be applied to the coding case in Theorem 5. Using the fact that $((X_i, Y_i, Z_i))_{i=1}^n$ is an independent and identically distributed random sequence, and that the normalized moment of inertia $M_n \to 1/2\pi\mathrm{e}$ as the block length $n \to \infty$, for high rates $\mathcal{R}$,

$$\mathcal{D}^{\mathrm{NWZ}}_{XZ|Y}(\mathcal{R}) \simeq \mathcal{D}^{\mathrm{N}}_{XZ|Y}(\mathcal{R}) \simeq \mathcal{D}_\infty + \frac{1}{2\pi\mathrm{e}}\, 2^{2\, \mathrm{h}(\bar{X}|Y)}\, 2^{-2\mathcal{R}} = \mathcal{D}_\infty + \alpha^2\, \sigma^2\, 2^{-2\mathcal{R}},$$

which is consistent with (3).

Observe that Theorem 5 gives the exact rate-distortion function for all rates, and guarantees that there is no rate loss due to the unavailability of the side information at the encoder, also for all rates. However, the hypotheses are more restrictive than those in [10, Theorem 3]. First, $Z|y$ must be Gaussian with constant variance for a. e. y. Secondly, further to the additive separability of $\bar{x}(y,z) \overset{\text{a.e.}}{=} \bar{x}_Y(y) + \bar{x}_Z(z)$ required by the high-rate theory, we must have $\bar{x}_Z(z) \overset{\text{a.e.}}{=} \alpha\, z$.

If X, Y and Z are jointly Gaussian with positive definite covariance matrix, then Theorem 5 implies

$$\mathcal{D} = \mathcal{D}_\infty + \sigma^2_{\bar{X}|Y}\, 2^{-2\mathcal{R}} = \sigma^2_{X|YZ} + (\sigma^2_{X|Y} - \sigma^2_{X|YZ})\, 2^{-2\mathcal{R}},$$

as found in [4,11], and we have $\alpha = 0$ if and only if $X \leftrightarrow Y \leftrightarrow Z$.

4 Experimental Results

We illustrate the previous theoretic analysis with a simple, intuitive example of noisy WZ coding. The side information Y is a discrete random state, uniformly distributed on $\mathscr{Y} = \{-1, 1\}$, and only known by the decoder. The encoder observes a noisy version $Z = X + Y$ of the state, where $X \sim \mathcal{N}(0, \sigma^2)$ is independent from Y, and $\sigma = 1$. Z can also be regarded as a Gaussian mixture with two components of equal weight and variance. At the decoder we are interested in estimating the value X of the additive noise itself, or the component of the mixture, given both the coded observation and the state information. Consider codes operating on blocks of n samples of (X, Y, Z), minimizing $\mathcal{D} = \frac{1}{n}\,\mathrm{E}\,\|X^n - \hat{X}^n\|^2$, given a rate constraint $\mathcal{R}$.

Clearly, if the state Y were available at the encoder, the value $X = Z - Y$ of the noise could be encoded directly, and the resulting conditional distortion-rate function would be $\mathcal{D}^{\mathrm{N}}_{XZ|Y}(\mathcal{R}) = \sigma^2\, 2^{-2\mathcal{R}}$, in other words, for each value $y \in \mathscr{Y}$, the Gaussian r. v. $Z|y = X + y \sim \mathcal{N}(y, \sigma^2)$ is optimally encoded. Suppose now that Z is conventionally coded, i. e., the side information is ignored in the design of both the encoder and the decoder, and suppose further that the estimate $\hat{X}$ is obtained from the reconstruction $\hat{Z}$ as $\hat{X} = \hat{Z} - Y$. Then, $X - \hat{X} = Z - \hat{Z}$ (the distortions of X and Z are the same), and the high-rate approximation of the distortion-rate function is $\mathcal{D}_Z(\mathcal{R}) \simeq \frac{1}{2\pi \mathrm{e}}\, 2^{2(\mathrm{h}(Z) - \mathcal{R})} \simeq 1.96\, 2^{-2\mathcal{R}}$, which represents a distortion increase of approximately 2.93 dB with respect to optimal conditional coding ($\mathrm{h}(Z) \simeq 2.53$ bit was computed by numerical integration for $\sigma = 1$).

We return to the case of noisy WZ coding. $\mathrm{E}[X|y,z] = z - y$ is additively separable, hence the high-rate approximation results for WZ coding of noisy sources [10] (set $\bar{x}_Z(z) = z$, $\bar{x}_Y(y) = -y$) guarantee that for *each* fixed dimension, an optimal lattice quantizer followed by an ideal Slepian-Wolf coder can approach the rate-distortion performance of an optimal system allowed to use the side information at the encoder so long as the rate is sufficiently large. On the other hand, since this example also satisfies the hypotheses of Theorem 5, there will not be any rate loss either if a fixed-rate vector quantizer of dimension large enough is used, regardless of whether the rate is high or low. In this case, dimension is a trade-off for rate.

For dimensions $n = 1, 2, 3$, noisy WZ quantizers $q(z^n)$ and reconstruction functions $\hat{x}^n(q, y^n)$ were designed using an extension of the Lloyd algorithm for distributed source coding, described in [12], to noisy sources [13]. Such quantizers were designed assuming that the quantization index $Q = q(Z^n)$ is losslessly coded with an ideal Slepian-Wolf coder at rate $\mathcal{R} = \frac{1}{n}\,\mathrm{H}(Q|Y^n)$. The experimental distortion-rate functions obtained are plotted in Fig. 3, along with the distortion-rate function derived using Theorem 5. The direct application of the theory developed in [10] pro-

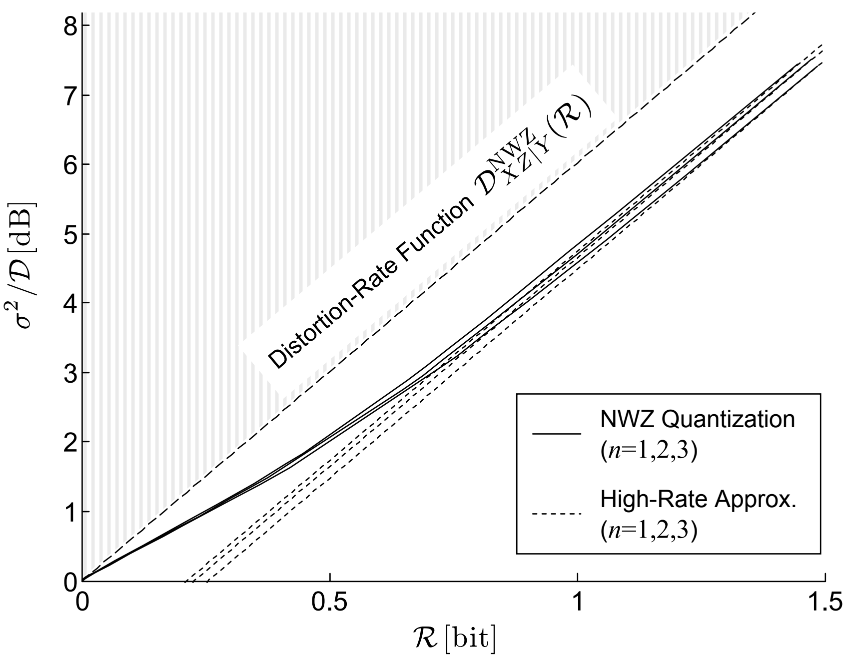

Figure 3: Distortion-rate function $\mathcal{D}^{\mathrm{NWZ}}_{XZ|Y}$ for the example of noisy WZ coding in the text ($\sigma = 1$), compared with distortion-rate performances of optimized noisy WZ quantizers for dimensions $n = 1, 2, 3$, and their high-rate approximations. Rates and distortions are normalized by n.

vides the high-rate approximation of the distortion-rate function of such quantizers, $\mathcal{D}(\mathcal{R}) \simeq 2\pi\mathrm{e}\, M_n\, \sigma^2\, 2^{-2\mathcal{R}}$, for fixed dimension n. At high rates, the slope approaches 6.02 dB/bit, and the distortion gap with respect to the distortion-rate function in Theorem 5 is $2\pi\mathrm{e}\, M_n$, approximately equal to 1.53, 1.37 and 1.28 dB for dimensions n equal to 1, 2 and 3, respectively, consistently with the results shown in Fig. 3.

5 Conclusions

In order to ensure that no rate loss is incurred in the WZ coding problem with quadratic distortion, jointly Gaussian statistics are *not* necessary. It suffices to require that the source data X be the sum of *any* (measurable) *function* of the side information Y and independent Gaussian noise, even less restrictively than it was established in [3]. More generally, in the noisy WZ problem, the same condition, applied to the noisy observation Z in place of X, suffices, while X must be the sum of a function of Y, a linear function of Z, and a r. v. N satisfying $\mathrm{E}[N|y, z] \equiv 0$.

Furthermore, the side information Y may be arbitrarily distributed in any alphabet, discrete or continuous.

The condition on X in the noisy case, albeit slightly stronger, is similar to the additive separability condition required in the study of high-rate WZ quantization of noisy sources in [10]. However, it leads to an exact closed formula for the rate-distortion function, valid also at low rates. Finally, observe that in both the above-mentioned high-rate approximation study and the present one, the fulfillment of the sufficient conditions for absence of rate loss is determined by the conditional joint distribution of (X, Z) given Y, regardless of the marginal distribution of Y.

References

[1] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.

[2] A. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inform., Contr.*, vol. 38, no. 1, pp. 60–80, July 1978.

[3] S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1181–1203, May 2003.

[4] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *Trans. IECE Japan*, vol. E63, pp. 700–706, Oct. 1980.

[5] T. Flynn and R. Gray, "Encoding of correlated observations," *IEEE Trans. Inform. Theory*, vol. 33, no. 6, pp. 773–787, Nov. 1987.

[6] S. C. Draper, "Successive structuring of source coding algorithms for data fusion, buffering, and distribution in networks," Ph.D. dissertation, MIT, June 2002.

[7] H. S. Witsenhausen, "Indirect rate-distortion problems," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 518–521, Sept. 1980.

[8] I. Csiszár and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*. New York: Academic, 1981.

[9] T. Linder, R. Zamir, and K. Zeger, "On source coding with side-information-dependent distortion measures," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2697–2704, Nov. 2000.

[10] D. Rebollo-Monedero, S. Rane, and B. Girod, "Wyner-Ziv quantization and transform coding of noisy sources at high rates," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2004.

[11] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 966–976, Aug. 2004.

[12] D. Rebollo-Monedero, R. Zhang, and B. Girod, "Design of optimal quantizers for distributed source coding," in *Proc. IEEE Data Compression Conf. (DCC)*, Snowbird, UT, Mar. 2003, pp. 13–22.

[13] D. Rebollo-Monedero and B. Girod, "Design of optimal quantizers for distributed coding of noisy sources," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Philadelphia, PA, Mar. 2005, invited paper.