Source Coding With Distortion Side Information At The Encoder

Emin Martinian, Gregory W. Wornell
Massachusetts Institute of Technology
Cambridge, MA USA
Email: {emin,gww}@allegro.mit.edu

Ram Zamir*
Tel Aviv University
Tel Aviv, Israel
Email: {zamir}@eng.tau.ac.il

Abstract: We consider lossy source coding when side information affecting the distortion measure may be available at the encoder, decoder, both, or neither. For example, such distortion side information can model reliabilities for noisy measurements, sensor calibration information, or perceptual effects like masking and sensitivity to context. When the distortion side information is statistically independent of the source, we show that in many cases (*e.g.*, for additive or multiplicative distortion side information) there is no penalty for knowing the side information only at the encoder, and there is no advantage to knowing it at the decoder. Furthermore, for quadratic distortion measures scaled by the distortion side information, we evaluate the penalty for lack of encoder knowledge and show that it can be arbitrarily large. In this scenario, we also sketch transform based quantizers constructions which efficiently exploit encoder side information in the high-resolution limit.

1 Introduction

In many large systems such as sensor networks, communication networks, and biological systems different parts of the system may each have limited or imperfect information but must somehow cooperate. Key issues in such scenarios include the penalty incurred due to the lack of shared information, possible approaches for combining information from different sources, and the more general question of how different kinds of information can be partitioned based on the role of each system component.

One example of this scenario is when an observer records a signal $\mathbf{x}$ to be conveyed to a receiver who also has some additional signal side information $\mathbf{w}$ which is correlated with $\mathbf{x}$. As demonstrated by various researchers, in many cases the observer and receiver can obtain the full benefit of the signal side information even if it is known only by the receiver [1] [2] [3].

In this paper we consider a different scenario where instead the observer has some distortion side information $\mathbf{q}$ which describes what components of the data are more sensitive to distortion than others, but the receiver may not have access to $\mathbf{q}$. Specifically, let us model the differing importance of different signal components by measuring the distortion between the ith source sample, $x[i]$, and its quantized value, $\hat{x}[i]$, by a distortion function which depends on the side information $q[i]$: $d(x[i], \hat{x}[i], q[i])$.

*This work was conducted in part while R. Zamir was visiting the Digital Signal Processing Group at MIT.

Proceedings of the Data Compression Conference (DCC'04)

In principle, one could treat the source-side information pair $(\mathbf{q}, \mathbf{x})$ as an "effective composite source", and apply conventional techniques to quantize it. Such an approach, however, ignores the different effect $\mathbf{q}$ and $\mathbf{x}$ have on the distortion. And as often happens in lossy compression, good understanding of the distortion measure may lead to better designs.

For example, a sensor may have side information corresponding to reliability estimates for measured data (which may or may not be available at the receiver). This may occur if the sensor can calibrate its accuracy to changing conditions (*e.g.*, the amount of light, background noise, or other interference present), if the sensor averages data for a variety of measurements (*e.g.*, combining results from a number of sub-sensors) or if some external signal indicates important events (*e.g.*, an accelerometer indicating movement).

Alternatively, certain components of the signal may be more or less sensitive to distortion due to masking effects or context [4]. For example errors in audio samples following a loud sound, or errors in pixels spatially or temporally near bright spots are perceptually less relevant. Similarly, accurately preserving certain edges or textures in an image or human voices in audio may be more important than preserving background patterns/sounds. Masking, sensitivity to context, etc., is usually a complicated function of the entire signal. Yet often there is no need to explicitly convey information about this function to the encoder. Hence, from the point of view of quantizing a given sample, it is reasonable to model such effects as side information.

Clearly in performing data compression with distortion side information, the encoder should weight matching the more important data more than matching the less important data. The importance of exploiting the different sensitivities of the human perceptual system are widely recognized by engineers involved in the construction and evaluation of practical compression algorithms *when distortion side information is available at both observer and receiver*. In contrast, the value and use of distortion side information known only at either the encoder or decoder but not both has received relatively little attention in the information theory and quantizer design community. The rate-distortion function with decoder-only side information, relative to side information dependent distortion measures (as an extension of the Wyner-Ziv setting [3]), is given in [2]. A high resolution approximation for this rate-distortion function for locally quadratic weighted distortion measures is given in [5].

We are not aware of an information-theoretic treatment of encoder-only side information with such distortion measures. In fact, the mistaken notion that encoder only side information is never useful is common folklore. This may be due to a misunderstanding of Berger's result that side information *which does not affect the distortion measure* is never useful when known only at the encoder [6].

In this paper we study the rate-distortion trade-off when side information about the distortion sensitivity is available. We show that such distortion side information can provide an arbitrarily large advantage (relative to no side information) even when the distortion side information is known only at the encoder. Furthermore, we show that just as knowledge of signal side information is often only required at the decoder, knowledge of distortion side information is often only required at the encoder. Beyond the theoretical results, these observations serve as a useful guide for designing quantizers with distortion side information.

We first illustrate how distortion side information can be used even when known only by the observer with some examples in Section 2. Next, in Section 3, we precisely define a problem model and state the relevant rate-distortion trade-offs. In Section 4, we present our main results characterizing when knowledge of distortion side information is sufficient at only the encoder and sketch one practical construction.

2 Examples

2.1 Discrete Uniform Source

Consider the case where the source, $\mathsf{x}[i]$, corresponds to n samples each uniformly and independently drawn from the finite alphabet $\mathcal{X}$ with cardinality $|\mathcal{X}| \geq n$. Let $\mathsf{q}[i]$ correspond to n binary variables indicating which source samples are relevant. Specifically, let the distortion measure be of the form $d(q, x, \hat{x}) = 0$ if and only if either $q = 0$ or $x = \hat{x}$. Finally, let the sequence $\mathsf{q}[i]$ be statistically independent of the source with $\mathsf{q}[i]$ drawn uniformly from the n choose k subsets with exactly k ones.

If the side information were unavailable or ignored, then losslessly communicating the source would require exactly $n \cdot \log |\mathcal{X}|$ bits. A better (though still sub-optimal) approach when encoder side information is available would be for the encoder to first tell the decoder which samples are relevant and then send only those samples. This would require $n \cdot H_b(k/n) + k \cdot \log |\mathcal{X}|$ bits where $H_b(\cdot)$ denotes the binary entropy function. Note that if the side information were also known at the decoder, then the overhead required in telling the decoder which samples are relevant could be avoided and the total rate required would only be $k \cdot \log |\mathcal{X}|$. We will show that this overhead can in fact be avoided even without decoder side information.

Pretend that the source samples $\mathsf{x}[0]$, $\mathsf{x}[1]$, ..., $\mathsf{x}[n-1]$, are a codeword of an (n, k) Reed-Solomon (RS) code (or more generally any MDS[1] code) with $\mathsf{q}[i] = 0$ indicating an erasure at sample i. Use the RS *decoding* algorithm to "correct" the erasures and determine the k corresponding information symbols which are sent to the receiver. To reconstruct the signal, the receiver *encodes* the k information symbols using the encoder for the (n, k) RS code to produce the reconstruction $\hat{\mathsf{x}}[0]$, $\hat{\mathsf{x}}[1]$, ..., $\hat{\mathsf{x}}[n-1]$. Only symbols with $\mathsf{q}[i] = 0$ could have changed, hence $\hat{\mathsf{x}}[i] = \mathsf{x}[i]$ whenever $\mathsf{q}[i] = 1$ and the relevant samples are losslessly communicated using only $k \cdot \log |\mathcal{X}|$ bits.

As illustrated in Fig. 1, RS decoding can be viewed as curve-fitting and RS encoding can be viewed as interpolation. Hence this source coding approach can be viewed as fitting a curve of degree $k - 1$ to the points of $\mathsf{x}[i]$ where $\mathsf{q}[i] = 1$. The resulting curve can be specified using just k elements. It perfectly reproduces $\mathsf{x}[i]$ where $\mathsf{q}[i] = 1$ and interpolates the remaining points.

[1] The desired MDS code always exists since we assumed $|\mathcal{X}| \geq n$. For $|\mathcal{X}| < n$, near MDS codes exist which give asymptotically similar performance with an overhead that goes to zero as $n \to \infty$.

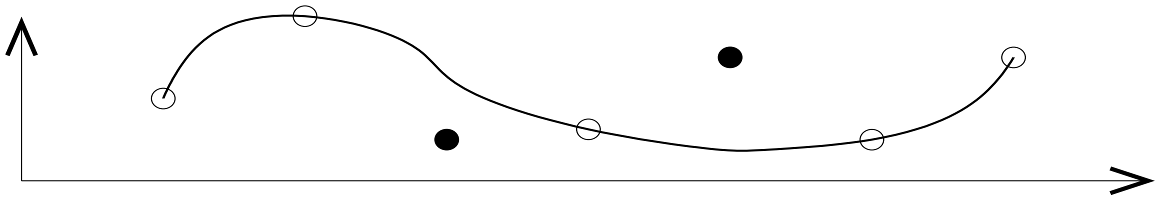


Figure 1: Losslessly encoding a source with $n = 7$ points where only $k = 5$ points are relevant (*i.e.*, the unshaded ones), can be done by fitting a fourth degree curve to the relevant points. The resulting curve will require k elements (yielding a compression ratio of k/n) and will exactly reproduce the desired points.

2.2 Gaussian Source

A similar approach can be used to quantize a zero mean, unit variance, complex Gaussian source relative to quadratic distortion using the Discrete Fourier Transform (DFT). Specifically, to encode the source samples $\mathsf{x}[0]$, $\mathsf{x}[1]$, ..., $\mathsf{x}[n-1]$, pretend that they are samples of a complex, periodic, Gaussian, sequence with period n, which is band-limited in the sense that only its first k DFT coefficients are non-zero. Using periodic, band-limited, interpolation we can use only the k samples for which $\mathsf{q}[i] = 1$ to find the corresponding k DFT coefficients, $\mathsf{X}[0]$, $\mathsf{X}[1]$, ..., $\mathsf{X}[k-1]$.

The relationship between the k relevant source samples and the k interpolated DFT coefficients has a number of special properties. In particular this $k \times k$ transformation is unitary. Hence, the DFT coefficients are Gaussian with unit variance and zero mean. Thus, the k DFT coefficients can be quantized with average distortion D per coefficient and $k \cdot R(D)$ bits where $R(D)$ represents the rate-distortion trade-off for the quantizer. To reconstruct the signal, the decoder simply transforms the quantized DFT coefficients back to the time domain. Since the DFT coefficients and the relevant source samples are related by a unitary transformation, the average error per coefficient for these source samples is exactly D.

Note if the side information were unavailable or ignored, then at least $n \cdot R(D)$ bits would be required. If the side information were losslessly sent to the decoder, then $n \cdot H_b(k/n) + k \cdot R(D)$ would be required. Finally, even if the decoder had knowledge of the side information, at least $k \cdot R(D)$ bits would be needed. Hence, the DFT scheme achieves the same performance as when the side information is available at both the encoder and decoder, and is strictly better than ignoring the side information or losslessly communicating it.

3 Problem Model

Vectors and sequences are denoted in bold (*e.g.*, $\mathbf{x}$) with the ith element denoted as $x[i]$. Random variables are denoted using the sans serif font (*e.g.*, x) while random vectors and sequences are denoted with bold sans serif (*e.g.*, $\boldsymbol{\mathsf{x}}$). We denote mutual information, entropy, and expectation as $I(\mathsf{x};\mathsf{y})$, $H(\mathsf{x})$, $E[\mathsf{x}]$. Calligraphic letters denote sets (*e.g.*, $x \in \mathcal{X}$).

We are primarily interested in a particular type of side information (which we call "distortion side information") that is statistically independent of the source but affects the distortion measure. Specifically, we consider the source coding with distortion side information problem defined as the tuple

$$(\mathcal{X}, \hat{\mathcal{X}}, \mathcal{Q}, p_{\mathsf{x}}(x), p_{\mathsf{q}}(q), d(\cdot,\cdot,\cdot)). \tag{1}$$

A source $\mathbf{x}$ consists of the n samples $x[1]$, $x[2]$, ..., $x[n]$ drawn from the alphabet $\mathcal{X}$. The distortion side information $\mathbf{q}$ likewise consists of n samples drawn from the alphabet $\mathcal{Q}$. These random variables are generated according to the distribution

$$p_{\mathbf{x},\mathbf{q}}(\mathbf{x}, \mathbf{q}) = \prod_{i=1}^{n} p_{\mathsf{x}}(x[i]) \cdot p_{\mathsf{q}}(q[i]).$$

A rate R encoder, $f(\cdot)$, maps a source as well as possible side information to an index $i \in \{1, 2, \ldots, 2^{nR}\}$. The corresponding decoder, $g(\cdot)$, maps the resulting index as well as possible decoder side information to a reconstruction of the source. Distortion for a source $\mathbf{x}$ which is quantized and reconstructed to the sequence $\hat{\mathbf{x}}$ taking values in the alphabet $\hat{\mathcal{X}}$ is measured via

$$d(\mathbf{x}, \hat{\mathbf{x}}, \mathbf{q}) = \frac{1}{n}\sum_{i=1}^{n} d(x[i], \hat{x}[i], q[i]). \tag{2}$$

As usual, the rate-distortion function is the minimum rate such that there exists a system where the distortion is at most D with probability approaching 1 as $n \to \infty$.

The four scenarios where $\mathbf{q}$ is available at the encoder, decoder, both, or neither are illustrated in Fig. 2 along with the symbol denoting each rate-distortion function.

Proposition 1. *The rate-distortion functions for the scenarios in Fig. 2 are*

$$R_{\text{NONE}}(D) = \inf_{p_{\hat{x}|x}(\hat{x}|x):E[d(x,\hat{x},q)]\le D} I(x;\hat{x}) \tag{3a}$$

$$R_{\text{DEC}}(D) = \inf_{p_{u|x}(u|x),v(\cdot,\cdot):E[d(x,v(u,q),q)]\le D} I(x;u) - I(u;q) \tag{3b}$$

$$R_{\text{ENC}}(D) = \inf_{p_{\hat{x}|x,q}(\hat{x}|x,q):E[d(x,\hat{x},q)]\le D} I(x,q;\hat{x}) = I(x;\hat{x}|q) + I(\hat{x};q) \tag{3c}$$

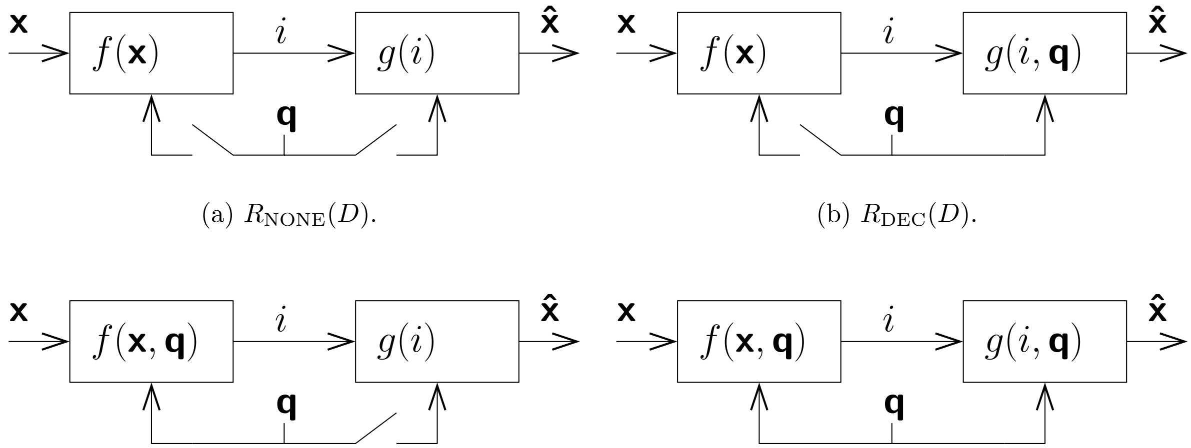
$$R_{\text{BOTH}}(D) = \inf_{p_{\hat{x}|x,q}(\hat{x}|x,q):E[d(x,\hat{x},q)]\le D} I(x;\hat{x}|q). \tag{3d}$$

The rate-distortion functions in (3a), (3b), and (3d) follow from standard results (*e.g.*, [6] [1] [2] [7] [3]). To obtain (3c) we can apply the classical rate-distortion theorem to the "super source" $\mathbf{x}' = (\mathbf{x}, \mathbf{q})$ as suggested by Berger [8]. In the sequel we characterize the penalty or rate-loss incurred by having side information available only at the encoder, only at the decoder, or neither compared to full side information.

4 Main Results

A system with encoder only side information corresponds to a system with a fixed codebook but a variable partition which depends upon $\mathbf{q}$.[2] As an almost trivial

[2]This structure also appears in the study of robust codebooks [9].



(a) $R_{\text{NONE}}(D)$. (b) $R_{\text{DEC}}(D)$.

(c) $R_{\text{ENC}}(D)$. (d) $R_{\text{BOTH}}(D)$.

Figure 2: Possible scenarios and rate-distortion functions with distortion side information at the decoder (b), encoder (c), both (d), and neither (a). The terms in (b) and (d) are also known as the Wyner-Ziv and conditional rate-distortion functions.

example, consider an encoder which observes $\mathbf{x} = \mathbf{z} + \mathbf{q}$ where $\mathbf{z}$ represents the true signal and $\mathbf{q}$ represents observation noise, *i.e.*, $d(z - \hat{x}) = d(x - q - \hat{x}) = d'(x, \hat{x}, q)$. By shifting the partition by $\mathbf{q}$ to quantize $\mathbf{x} - \mathbf{q}$, the encoder achieves optimal performance. By contrast, systems with decoder only side information correspond to fixed partitions with variable codebooks and often can not exploit distortion side information as easily. In the following, we make these notions precise for more general distortion measures.

4.1 Rate-Distortion Trade-Offs

We begin with the following theorems (proved in Appendix A) which show when side information at the encoder can be optimally used even though such side information may be useless if known only at the decoder.

Theorem 1. *Let distortion side information $\mathbf{q}$ be statistically independent of the source $\mathbf{x}$ and let $\mathbf{x}$ be uniformly distributed over a group with distortion measured via $d(x, \hat{x}, q) = d(x \ominus \hat{x}, q)$ where $\ominus$ represents a binary group operation. Then the rate-distortion function when $\mathbf{q}$ is available at the encoder is the same as when it is available at both encoder and decoder,* i.e., $R_{\text{ENC}}(D) = R_{\text{BOTH}}(D)$.

To state a similar result for continuous sources, we require various technical conditions describing a "smooth" source and distortion measure. Essentially, all that is required is that the source have a density and finite differential entropy and that an entropy maximizing distribution exists for the distortion measure of interest. For example, any vector source and distortion measure with

$$-\infty < h(\mathsf{x}) < \infty \text{ and } E[||\mathsf{x}||^{\gamma_q}] < \infty \text{ and } d(x, \hat{x}, q) = \alpha_q + \beta_q \cdot ||x - \hat{x}||^{\gamma_q} \quad \forall q \tag{4}$$

will satisfy the required conditions provided α_{q}, β_{q}, γ_{q} are non-negative. See [10] or [11] for a more detailed discussion of the necessary technical conditions.

Theorem 2. *Let* $\mathbf{q}$ *be statistically independent of the source* $\mathbf{x}$ *and consider any "smooth" source and distortion measure satisfying the conditions in [10, Theorem 1] for each* $q \in \mathcal{Q}$*. Then the rate-distortion function when* $\mathbf{q}$ *is available only at the encoder is asymptotically the same as when it is available at both encoder and decoder,* i.e., $\lim_{D \to D_{\min}} R_{\text{ENC}}(D) - R_{\text{BOTH}}(D) = 0$.[3]

Finally, in addition to the previous theorems showing when only the encoder requires $\mathbf{q}$, we have the following result stating when $\mathbf{q}$ is useless to the decoder.

Theorem 3. *Let the distortion side information* $\mathbf{q}$ *be statistically independent of the source* $\mathbf{x}$ *and consider scaled distortion measures of the form* $d(x, \hat{x}, q) = d_0(q) \cdot d_1(x, \hat{x})$*. Then the rate-distortion function for* $\mathbf{q}$ *available at the decoder is the same as when* $\mathbf{q}$ *is available at neither encoder nor decoder,* i.e., $R_{\text{DEC}}(D) = R_{\text{NONE}}(D)$.

Combining our results shows that in many cases knowledge of $\mathbf{q}$ is optimal at the encoder and useless at the decoder.

Corollary 1. *For sources and side information weighted difference distortion measures satisfying the conditions in Theorems 1 and 3 (or respectively in Theorems 2 and 3),* $R_{\text{ENC}}(D) - R_{\text{BOTH}}(D) = 0$ *and* $R_{\text{NONE}}(D) - R_{\text{DEC}}(D) = 0$*, or respectively,* $\lim_{D \to D_{\min}} R_{\text{ENC}}(D) - R_{\text{BOTH}}(D) = 0$ *and* $\lim_{D \to D_{\min}} R_{\text{NONE}}(D) - R_{\text{DEC}}(D) = 0$.

4.2 The Penalty for Lack of Encoder Knowledge

Consider generalizing the commonly used quadratic distortion model by scaling the distortion as a function of the side information as in [5]. Specifically, let $d(\mathsf{q}, \mathsf{x}, \hat{\mathsf{x}}) = \mathsf{q} \cdot (\mathsf{x} - \hat{\mathsf{x}})^2$. For this scenario, [5] implies that $R_{\text{BOTH}}(D) = h(\mathsf{x}) - (1/2)\ln(2\pi e D) + (1/2)E[\ln \mathsf{q}]$ while $R_{\text{DEC}}(D) = h(\mathsf{x}) - (1/2)\ln(2\pi e D) + (1/2)\ln E[\mathsf{q}]$. Combining this with Corollary 1 shows that the asymptotic penalty for lack of encoder knowledge of q is $(1/2) \cdot (\ln E[\mathsf{q}] - E[\ln \mathsf{q}])$ nats per sample. Table 1 evaluates this penalty for various distributions of q. Note that in many cases, the rate loss can be made arbitrarily large by choosing the appropriate shape parameter to place more probability near $\mathsf{q} = 0$. Intuitively, this occurs because when $\mathsf{q} \approx 0$, the informed encoder can transmit almost zero rate while the uninformed encoder must transmit a large rate to achieve high resolution. Furthermore, all but one of these distributions would require infinite rate to losslessly communicate the side information.

4.3 Quantizer Design

As discussed in Section 2, for distortion side information indicating that a given source sample is relevant or completely irrelevant, a transform followed by a scalar quantizer[4]

[3]Usually $D_{\min} = 0$, but to allow for more general distortion measures we define $D_{\min}$ as the minimum achievable distortion when arbitrarily high rates are allowed.

[4]Entropy coding the scalar quantizers output is also possible without changing this result.

Table 1: The rate-penalty (in nats) for not knowing side-information with the given distribution at the encoder. Euler's constant is denoted by γ.

Distribution Name	Density for q	Rate Gap in nats
Exponential	$\tau \exp(-q\tau)$	$-\frac{1}{2}\ln\gamma \approx 0.2748$
Uniform	$1_{q\in[0,1]}$	$\frac{1}{2}(1-\ln 2) \approx 0.1534$
Lognormal	$\frac{1}{q\sqrt{2\pi Q^2}} \exp\left[-\frac{(\ln q - M)^2}{2Q^2}\right]$	$\frac{Q^2}{4}$
Pareto	$\frac{a^b}{q^{a+1}}, q \geq b > 0, a > 1$	$\frac{1}{2}\left[\ln\frac{a}{a-1} - 1/a\right]$
Gamma	$\frac{b(bq)^{a-1}\exp(-bq)}{\Gamma(a)}$	$\frac{1}{2}\left\{\ln a - \frac{d}{dx}[\ln\Gamma(x)]_{x=a}\right\} \approx \frac{1}{2a}$
Pathological	$(1-\epsilon)\delta(q-\epsilon) + \epsilon\delta(q-1/\epsilon)$	$\frac{1}{2}\ln(1+\epsilon-\epsilon^2) - \frac{1-2\epsilon}{2}\ln\epsilon \approx \frac{1}{2}\ln\frac{1}{\epsilon}$
Positive Cauchy	$\frac{2/\pi}{1+q^2}, q \geq 0$	∞

efficiently exploits encoder side information. To generalize this transform coding construction, consider two-level side information with the alphabet $\mathcal{Q} = \{q_0, q_1\}$ where $q_1 \geq q_0 \geq 0$ and distortion is measured via $d(q, x, \hat{x}) = q \cdot (x - \hat{x})^2$. Furthermore, let a random k out of n samples of $\mathbf{q}$ take the value q_1 while the other $n-k$ samples take the value q_0. If $\mathbf{q}$ is known at both encoder and decoder then the optimal strategy is to use a rate R_0 quantizer for samples when $q = q_0$ and a rate $R_1 \geq R_0$ quantizer when $q = q_1$ such that the overall rate or distortion constraint is satisfied.

To asymptotically achieve the same performance via transform coding when $\mathbf{q}$ is known only at the encoder, we can use the following procedure. First, quantize the k more important source samples where $q[i] = q_1$ with a rate R_0 quantizer to produce $\hat{\mathbf{x}}_1$. Define the first stage error signal as $e[i] = x[i] - \hat{x}_1[i]$ where we assume $\hat{x}_1[i] = 0$ when $q[i] = q_0$ since these less important samples have not yet been quantized. Next use band-limited interpolation to find the k DFT coefficients $E[i]$ such that the IDFT of $E[i]$ accurately reproduces $e[i]$ when $q[i] = q_1$. Quantize these coefficients using a rate $R_1 - R_0$ quantizer. Define the second stage error signal as $e'[i] = x[i] - \hat{x}_1[i] - \hat{e}[i]$ where $\hat{e}[i]$ represents the IDFT of the quantized $E[i]$. Finally, quantize the $n-k$ samples of $e'[i]$ where $q[i] = q_0$ using a rate R_0 quantizer to produce $\hat{\mathbf{x}}_2$.

The receiver obtains the reconstruction $\hat{x}[i] = \hat{x}_1[i] + \hat{x}_2[i] + \hat{e}[i]$ consisting of a rate R_0 scalar quantization of each source sample and a quantized shift $\hat{e}[i]$. Although the receiver can not deduce from $\hat{e}[i]$ which samples were more important, $\hat{e}[i]$ was chosen by the encoder to make the quantization of the more important samples more accurate. As illustrated in Fig. 3 for $(n, k) = (2, 1)$, this type of system corresponds to a quantization lattice where the encoder can choose the partition to shape the error based on the side information. It is possible to show that in high resolution this system approaches the performance of a fully informed system (*i.e.*, using a rate R_0 quantizer when $q[i] = q_0$ and a rate R_1 quantizer when $q[i] = q_1$) [11]. Conceptually, in the high resolution limit, edge effects become negligible and the shape of each cell in Fig. 3 approaches a rectangle. This system specializes to the one in Section 2.2

when $q_0 = 0$ and can be further generalized to larger side information alphabets [11].

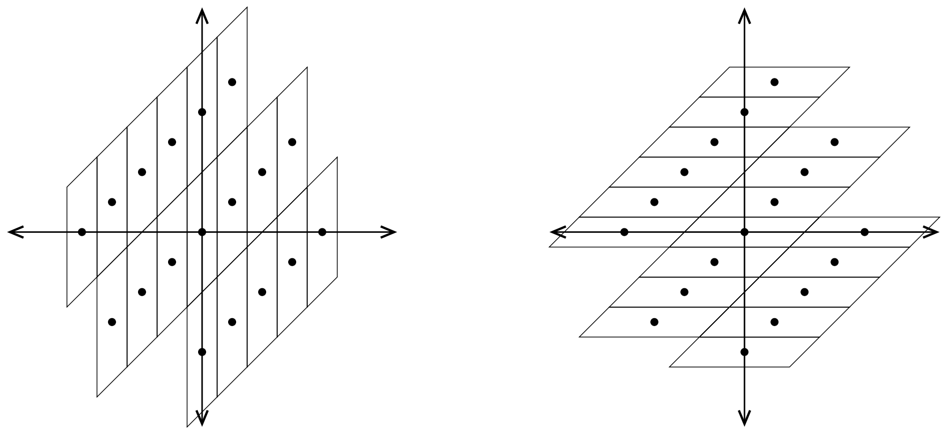


Figure 3: The quantization points and possible partitions for a transform coder. If the encoder knows the horizontal error (respectively, vertical error) is more important, it can use the partition on the left to increase horizontal accuracy (resp., vertical accuracy). The decoder only needs to know the quantization point not the partition.

A Proofs

Proof of Theorem 1: For a finite group, choosing z^* to maximize $H(z|q)$ subject to the constraint $E[d(z,q)] \leq D$ yields the following lower bound on $R_{\mathrm{ENC}}(D)$:

$$\begin{aligned} I(\hat{x}; x, q) &= H(x) + H(q) - H(x, q|\hat{x}) & (5)\\ &= \log|\mathcal{X}| + H(q) - H(q|\hat{x}) - H(\hat{x} - x|x, q) & (6)\\ &\geq \log|\mathcal{X}| - H(\hat{x} - x|q) & (7)\\ &\geq \log|\mathcal{X}| - H(z^*|q) & (8) \end{aligned}$$

where (7) follows since conditioning reduces entropy. Choosing the test-channel distribution $\hat{x} = z^* + x$ achieves this bound with equality and must therefore be optimal. Furthermore, since $\hat{x}$ and q are statistically independent for this test-channel distribution, $I(\hat{x}; q) = 0$ and thus comparing (3c) and to (3d) shows $R_{\mathrm{ENC}}(D) = R_{\mathrm{BOTH}}(D)$ for finite groups. The same argument holds for continuous groups with entropy replaced by differential entropy and $|\mathcal{X}|$ replaced by the Lebesgue measure of $\mathcal{X}$. For more general groups (*e.g.*, mixed groups with both discrete and continuous components), a more complicated convexity argument is required [11]. □

Proof Sketch For Theorem 2: Due to space constraints we only sketch the ideas behind the proof. As for Theorem 1, we can develop a lower bound for $R_{\mathrm{BOTH}}(D)$ using an entropy maximizing distribution and the Shannon lower bound [10]. Then by using the resulting test-channel distribution for $R_{\mathrm{ENC}}(D)$ we can show that $I(\hat{x}; q)$ goes to zero in the high resolution limit and therefore $R_{\mathrm{ENC}}(D) \to R_{\mathrm{BOTH}}(D)$. □

Proof of Theorem 3: When side information is available only at the decoder, Wyner-Ziv coding is optimal [3]. First we compute the optimal reconstruction function $v(\cdot,\cdot)$:

$$v(u,q) = \arg\min_{\hat{x}} E[d(\hat{x},\mathsf{x},\mathsf{q})|\mathsf{q}=q,\mathsf{u}=u] \quad (9)$$

$$= \arg\min_{\hat{x}} d_0(q) E[d_1(\hat{x},\mathsf{x})|\mathsf{q}=q,\mathsf{u}=u] \quad (10)$$

$$= \arg\min_{\hat{x}} E[d_1(\hat{x},\mathsf{x})|\mathsf{q}=q,\mathsf{u}=u] \quad (11)$$

$$= \arg\min_{\hat{x}} E[d_1(\hat{x},\mathsf{x})|\mathsf{u}=u] \quad (12)$$

where (10) follows by the assumption that we have a separable distortion measure and (12) follows because q is statistically independent of x (by assumption) and independent of u (since u is generated at the encoder from x). Thus since neither the optimal reconstruction function, $v(\cdot,\cdot)$ nor the auxiliary random variable, u, depend on q, knowing q at only the decoder provides no advantage. □

References

[1] T. M. Cover and M. Chiang, "Duality between channel capacity and rate distortion with two-sided state information," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1629–1638, June 2002.

[2] I. Csiszár and J. Körner, *Information Theory: Coding Theorems For Discrete Memoryless Systems*. Academic Press, 1981.

[3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.

[4] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on models of human perception," *Proceedings of the IEEE*, vol. 81, pp. 1385–1422, Oct. 1993.

[5] T. Linder, R. Zamir, and K. Zeger, "On source coding with side-information-dependent distortion measures," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2697–2704, Nov. 2000.

[6] T. Berger, *Rate Distortion Theory: A Mathematical Basis For Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[7] R. Gray, "A new class of lower bounds to information rates of stationary sources via conditional rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 19, pp. 480–489, July 1973.

[8] T. Berger. Private Communication.

[9] R. Zamir, "The half a bit loss of robust source/channel codebooks," in *Information Theory Workshop*, (Bangalore, India), pp. 123–126, Oct. 2002.

[10] T. Linder and R. Zamir, "On the asymptotic tightness of the Shannon lower bound," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2026–2031, Nov. 1994.

[11] E. Martinian, G. W. Wornell, and R. Zamir, "Compression and coordination with distributed side information." in preparation.