High-Resolution Functional Quantization

Vinith Misra Vivek K Goyal Lav R. Varshney*

Abstract

Suppose a function of N real source variables $X_1^N = (X_1, X_2, \ldots, X_N)$ is desired at a destination constrained to receive a limited number of bits. If the result of evaluating the function, $Y = G(X_1^N)$, can be itself encoded, this is the optimal strategy—the origin of Y becomes irrelevant to the communication problem. We consider two alternative scenarios: distributed quantization, in which each X_i must be separately encoded; and linear transform coding of X_1^N. Optimal fixed- and variable-rate scalar quantizers are derived under the conventional assumptions of high-resolution quantization theory, and we find optimal transforms for transform coding. For certain classes of functions, examples demonstrate large improvements over using quantizers designed to minimize distortion of the X_is.

1 Introduction

Traditional modes of lossy compression seek to approximately recreate every letter of a source vector $X_1^N = (X_1, X_2, \ldots, X_N) \in \mathbb{R}^N$; attaining small mean-squared error (MSE) is the implicit aim. We consider here the class of problems for which the goal is to estimate at the destination a function $G(X_1^N)$ of the source data. We call this *(lossy) functional source coding* (FSC).[1] A wide variety of practical situations fit into the FSC framework; consider, for instance, computation on digitized analog data.

Many interesting variations are possible depending on the allowed functions G, the distortion measure, the separability of the joint distribution of X_1^N, the support of this distribution, and characteristics of the encoding process (distributed operation, fixed rate, variable rate, etc.). In addition, one may fruitfully focus on Shannon-theoretic limits or practical code constructions.

If G is unconstrained, we have generalized the "ordinary" source coding problem. In particular, taking G to be the identity function recovers the ordinary source coding problem, and a distortion measure d_G that applies to the range of G induces a distortion measure on the source vector through $d(X_1^N, \widehat{X}_1^N) = d_G(G(X_1^N), \widehat{G}(\widehat{X}_1^N))$, where $\widehat{X}_1^N$ is the compressed representation of X_1^N, and $\widehat{G}$ estimates $G(X_1^N)$ from $\widehat{X}_1^N$. If G is continuous, the intermediate value theorem predicts the existence of a point Z_1^N such that $\widehat{G}(\widehat{X}_1^N) = G(Z_1^N)$; therefore we may restrict $\widehat{G} = G$ without loss of optimality.

Our interest in this paper is not in generalizing but rather in exploring the potential gains that come from optimizing for a particular destination function G. We restrict our attention to $G : \mathbb{R}^N \to \mathbb{R}$ and $d_G : \mathbb{R}^2 \to \mathbb{R}$ given by $d_G(G, \widehat{G}) = (G - \widehat{G})^2$. We primarily address the design of scalar quantizers for the X_is with the goal of minimizing $D_G = E[d_G(G(X_1^N), G(\widehat{X}_1^N)]$, i.e., the MSE of the function. This is to be seen in comparison to allowing each X_i to be recovered with low MSE. In particular, we generate examples with clear scaling behavior with respect to N.

*The authors are with the Signal Transformation and Information Representation Group in the Research Laboratory of Electronics at MIT (`http://www.rle.mit.edu/stir/`). Work supported in part by NSF Grant 0729069 and an NSF Graduate Research Fellowship.

[1]Our usage follows [1, 2]. An alternative—not considered here—is for *functional quantization* to be quantization of an (uncountably infinite) alphabet of functions [3].

DOI 10.1109/DCC.2008.100

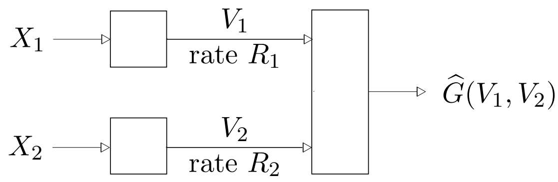

Figure 1: Generic functional source coding scenario shown with two variables.

General Formulation. While several papers have explicitly addressed FSC problems, many more have done so obliquely. Before developing our results, we briefly mention connections between FSC and some other problems in multiterminal information theory.

Consider the network shown in Fig. 1, where the number of source variables is $N = 2$ for simplicity. The variables X_1 and X_2 are known to distributed encoders which produce codewords V_1 and V_2 at rates R_1 and R_2. A joint decoder estimates $G(X_1, X_2)$ from (V_1, V_2) at distortion level D_G. Several relationships to other problems emerge:

- If G is the identity function, we have a general distributed source coding problem that is well-known in the lossless setting [4] and recently solved in the quadratic Gaussian case [5]. In this situation, the correlation of X_1 and X_2 is of primary interest.
- If $G(X_1, X_2) = X_1$ and R_2 is unconstrained, then X_2 can be viewed as receiver side information available at the decoder. The trade-off between R_1 and D_G is given by the Wyner-Ziv rate-distortion function [6, 7].
- The Wyner-Ziv scenario has been examined at high rate by Rebollo-Monedero et al. [8]. It has been shown that providing the receiver side information to the encoder yields no improvement in performance.
- For general G and R_2 unconstrained, the problem has been studied by Feng et al. [1].
- Let $Y = G(X_1, X_2)$. Then Y may be interpreted as a *remote source* that is observed only through X_1 and X_2 and we have the remote source multiterminal source coding problem [9].
- There has been significant work in the related notion of "task-oriented" quantization. Quantizing for classification [10], estimation [11], and detection [12] has been of particular interest.
- Rather than having a single function G, one may consider a set of functions $\{G_i\}_{i \in I}$ and define $D_G = E[d(G_\alpha(X_1^N), G_\alpha(\widehat{X}_1^N)]$, where α is a random variable taking values in index set I. In this setting, fixed- and variable-rate quantization to minimize MSE was studied by Bucklew [13]. This is perhaps the most closely related work, but it uses centralized (non-distributed) encoding.

Scope and Organization of the Paper. FSC is closely related to distributed coding, which has a mantra of "quantize and bin." In particular, it seems that quantization may as well be uniform or lattice-based since the interesting gains come from the binning process or other discrete operations; see [2] for a recent formulation using graph coloring.

In contrast, we take a more quantization-centric view: we look for gains from optimal (potentially nonuniform) quantization in situations in which binning plays no role. Using standard assumptions from high-resolution quantization theory, we develop optimal quantizers for FSC. The high-resolution approach applies only to the design of regular quantizers, and we discuss conditions under which optimal quantizers are regular. Note that binning yields irregular quantizers, so we are considering different types of FSC problems than commonly emphasized in the literature.

While functions of a single variable do not yield interesting examples, their consideration in Section 2 allows us to review standard results from high-resolution quantization and to plainly see that some conditions on G are necessary for optimal quantizers to be regular. Section 3 presents our main theoretical results for fixed- and variable-rate distributed functional quantization. Section 4 develops a (non-distributed) functional transform coding theory. We conclude with examples given in Section 5.

2 Functions of One Variable

Consider the quantization of X for estimation of $Y = G(X)$ with minimum MSE at the destination. The p.d.f. of Y can be derived from G and the p.d.f. of X, and optimal quantizers for Y can then be derived from the p.d.f. of Y. An optimal quantizer for X is given by $q_X(x) = q_Y(G(x))$ where q_Y is an optimal quantizer for Y. Thus the functional aspect of functional quantization does not add anything interesting for functions of one variable.

Fixed-Rate (Codebook-Constrained) Quantization. Consider the design of a K-level quantizer for Y; this yields a quantizer for X through $q_X(x) = q_Y(G(x))$. Standard assumptions for high-resolution analysis include that K is large, distortion due to quantizer overload is negligible, neighboring quantizer cells have similar sizes, and the p.d.f. of Y is reasonably smooth; see [14] for details. Quantizer analysis and design are then based on the use of point density functions. A *point density function* $\lambda(y)$ is a smooth function such that $\Delta\lambda(y)$ is approximately the fraction of quantizer reproduction points in an interval of length Δ centered at y. Standard analysis yields

$$D_G = E[(Y - q_Y(Y))^2] \approx \frac{1}{12K^2} \int_{\mathbb{R}} \frac{f_Y(y)}{\lambda^2(y)}\, dy, \tag{1}$$

where $f_Y(y)$ denotes the p.d.f. of Y. By minimizing D_G for fixed K, one shows that the optimal point density and resulting performance are

$$\lambda(y) = f_Y^{1/3}(y) \Big/ \left(\int f_Y^{1/3}(y)\, dy \right) \qquad \text{and} \qquad D_G \approx \frac{1}{12K^2} \left(\int_{\mathbb{R}} f_Y^{1/3}(y)\, dy \right)^3.$$

Variable-Rate (Entropy-Constrained) Quantization. The design of an entropy-constrained quantizer can be optimized similarly. Formally, an unnormalized point density should be used because the number of quantizer levels may be infinite. By minimizing D_G given in (1) subject to a constraint on $H(q_Y(Y))$, one sees that the optimal point density is constant.

Regularity. Description of a quantizer by means of a point density carries with it a regularity constraint. This is an innocuous restriction since the optimal high-rate quantizer for Y is regular. This regularity does not carry over to the partition induced on X. For instance, the optimal quantization cells for the function $G(x) = |x|$ are unions of disjoint positive and negative intervals.

If we constrain G to be monotonic, a regular cell in Y will transform to a regular cell in X. Thus an optimal quantizer for X is represented unambiguously by a point density function.

3 Optimal Regular Quantization

We work with three distinct variants of functional quantization with regular quantization, each applicable to different situations.

Fixed-Rate Distributed Quantization: Separated encoders are mandated, and (potentially nonuniform) fixed-rate quantizers are designed. All codewords from the ith encoder are of identical length R_i bits.

Variable-Rate Distributed Quantization: Separated encoders are again mandated, and (potentially nonuniform) quantizers are designed. Unlike in fixed-rate coding, block entropy coding is permitted to reduce the rate of the quantization indices.

Uniform Transform Coding: Violating distributed operation, a linear transformation is applied to X_1^N prior to scalar quantization. This will be discussed in Section 4, with the scalar quantizers constrained to be uniform for simplicity.

As in Section 2, we employ high-resolution analysis [14]. On one hand, this technique can directly yield asymptotically optimal quantizers. On the other hand, for it to be valid we must constrain our functions and distributions in the following manner:

- **C1.** $G(X_1^N)$ is smooth and monotonic in each of the source variables.
- **C2.** $\partial G/\partial X_i$ is both defined and bounded, for each i.
- **C3.** The joint probability distribution of the source variables, $f_{X_1^N}(x_1^N)$, is smooth and supported within a compact subset of $\mathbb{R}^N$.

Bucklew [13] has used a similar construction to deal with the single-encoder high-rate vector quantization of X_1^N when the exact form of G is only known to the decoder. By assuming that the statistical distribution of G is available; that each possible G adheres to **C1** and **C2**; and that **C3** is valid for the source distribution, several approximations are justified:

- **A1.** Within any quantizer cell, the distribution $f_{X_1^N}$ is roughly uniform (follows from **C3**).
- **A2.** Within any quantizer cell, G_i is roughly affine in X_1^N (follows from **C1** and **C2**).
- **A3.** All quantizer regions are regular (follows from **C1**).

We supplement these assumptions as the need arises.

3.1 Fixed-Rate (Codebook-Constrained) Quantization

We consider the distributed quantization problem depicted in Fig. 2; the scalar quantizers may be nonuniform. Specifically:

1. The ith encoder quantizes X_i into one of 2^{R_i} regular quantization cells.
2. The sum of rates cannot exceed a maximum, $\sum_i R_i \leq R_0$.
3. We allow the bit allocations, R_i, to be real valued. This is valid in two respects: (1) By means of block coding, real-valued R_i can be approximated, and (2) even in the absence of block coding, optimal integer-valued R_i can be obtained from optimal real-valued R_i [15].
4. The quantized variables, $\widehat{X}_i$, are sent to the decoder with no form of entropy coding.
5. The decoder performs the estimation $\widehat{G}(\widehat{X}_1^N) \triangleq G(\widehat{X}_1^N)$.

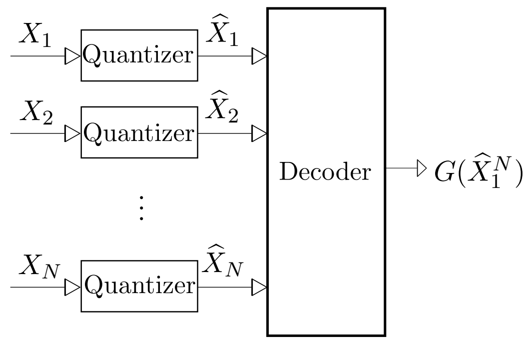

Figure 2: Fixed-Rate Distributed Quantization: Independent scalar quantization is performed at each source.

Our interest lies in minimizing the mean squared error of G. To describe the nonuniform quantizers, we define quantizer point density functions, $\lambda_i(x_i)$, as the limiting density of quantizer cells in the vicinity of x_i. Utilizing the high-rate approximation, we state that $\lambda_i(x_i)$ is approximately constant over any quantization interval, for large enough R_i.

Assume for convenience that f_{X_i} is supported over the interval $[0, 1]$, and that λ_i is nonzero almost everywhere. Any quantization cell, $Q \subset \mathbb{R}^N$, then takes the form of a box of length $2^{-R_i}/\lambda_i(x_{i0})$ in the ith dimension, where x_{i0} is an arbitrary element of Q. The distortion contribution from Q is then approximately

$$p(Q)E[d_G|Q] = p(Q)\sum_i \left|\frac{\partial G}{dx_i}(x^*)\right|^2 \frac{2^{-2R_i}}{12\lambda_i^2(x_i^*)},$$

where $p(Q)$ is the probability that x_1^N is an element of Q, and x^* is an arbitrary coordinate found within the box. Since the derivative term is (approximated as) constant throughout the box, this can be rewritten as an integral over Q:

$$p(Q)E[d_G|Q] = \sum_i \int_{x_1^N \in Q} f_{X_1^N}(x^N)\left|\frac{\partial G(x_1^N)}{\partial x_i}(x^*)\right|^2 \frac{2^{-2R_i}}{12\lambda_i^2(x_i^*)} dx_1^N. \tag{2}$$

Note that we have used the fact that the derivatives $\partial G(x_1^N)/\partial x_i$ and point densities $\lambda_i(x_i)$ are (approximately) constant throughout the box to replace the single point x^* with the variable x_1^N. Defining

$$g_i(x) = \left(E\left[\left|\partial G(x_1^N)/\partial x_i\right|^2 \mid X_i = x_i\right]\right)^{1/2},$$

the expected distortion over all intervals takes the form

$$D_G = \sum_i 2^{-2R_i} E\left[\frac{g_i^2}{12\lambda_i^2(x_i)}\right]. \tag{3}$$

The problem has been decoupled; there are N variables being separately scalar quantized, effectively companded with slope $g_i(x_i)$. Each can be optimized separately. An application of Hölder's inequality in one dimension (see [14]) shows that distortion is minimized by choice of $\lambda_i = C(g_i^2(x_i)f_{X_i}(x_i))^{1/3}$, where C is a normalization constant.

We can then minimize the rates. First, we discard all X_is whose companding expressions are 0, since this implies that the function is independent of them almost everywhere. Once this is done,

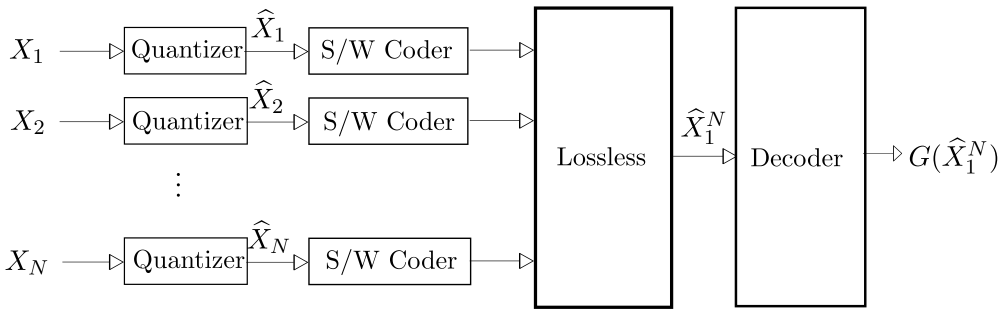

Figure 3: Variable Rate Quantization: Scalar quantization is now followed by Slepian-Wolf (S/W) coding.

the rates may be adjusted to attain a lower bound over the remaining N encoders. An application of the arithmetic-geometric mean inequality establishes this lower bound as

$$D_G \geq \tfrac{1}{12} N 2^{-2R_0/N} \prod_i \left\| f_{X_i}(x_i) g_i^2(x_i) \right\|_{1/3}^{1/N}. \tag{4}$$

This solution, despite its optimality within the fixed-rate construction, does not exploit dependencies among the X_is. By permitting block entropy coding, we may transmit at a total rate near the joint entropy.

3.2 Variable-Rate (Entropy-Constrained) Quantization

We introduce block entropy coding to our distributed scenario by splitting the decoder and the encoders into two pieces (see Fig. 3).

1. The ith scalar quantizer continues to scalar quantize each sample of X_i independently of other samples of X_i and independently of the happenings at other encoders. Quantization is performed at resolution R_i.
2. The second piece, a block entropy coder, losslessly encodes together M sequential outputs from the scalar quantizer, $\widehat{X}_{i1} \ldots \widehat{X}_{iM}$, for transmission. The rate of transmission of the ith encoder is denoted $\tilde{R}_i$.
3. $\sum_i \tilde{R}_i \leq \tilde{R}_0$.
4. The block decoder recreates $\widehat{X}_{i1} \ldots \widehat{X}_{iM}$ for each i.
5. Finally, the decoder computes an estimate of the values of G for each time instance: $\widehat{G}_j = G(\widehat{X}_{1j}, \ldots, \widehat{X}_{Nj})$.

While the high-rate distortion takes the same form as before (3), the optimization performed in Section 3.1 is no longer valid; the rate constraint is now on the sum rate with entropy coding, $\sum_i \tilde{R}_i$. In general, this sum rate is constrained to the rate region described by Doshi et al. [2]. However, due to the monotonicity requirement on our function, $\widehat{X}_i$ must be losslessly encoded. Slepian and Wolf [4] established $H(\widehat{X}_1^N)$ as an achievable lower bound on this performance. Assuming $f_{X_1^N}$ is constant in each quantizer region,

$$H(\widehat{X}_1^N) \approx h(X_1^N) + \sum_i E\left[\log \lambda_i(x_i)\right] + R,$$

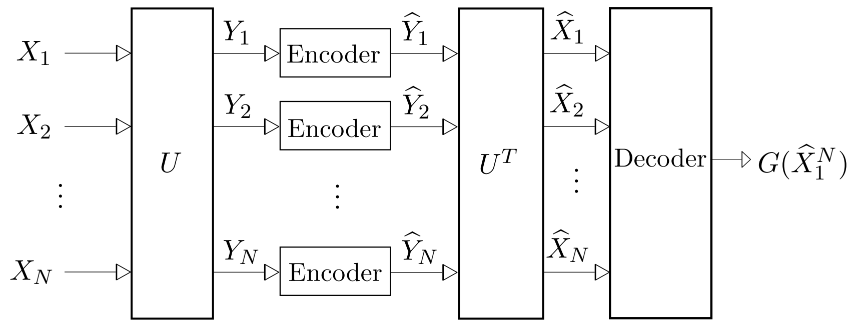

Figure 4: Uniform transform coding

where $h(X_1^N)$ is the joint differential entropy of the source variables. We will assume this rate without considering the details of its implementation.

The constraint $H(\widehat{X}_1^N) < \tilde{R}_0$ translates into a similar constraint on the resolution, $R < R_0$. As with fixed-rate, the bit allocations R_i are optimized for this sum-resolution constraint:

$$D_G = \frac{1}{12} N \exp_2 \left(-2\tilde{R}_0/N + 2h(X_1^N)/N + 2\sum_i E[\log \lambda_i]/N \right) \left(\prod_i E\left[g_i^2/\lambda_i^2\right] \right)^{1/N}. \quad (5)$$

To minimize this, we operate roughly in the footsteps of Gersho [16]. First, we rephrase:

$$D_G = \frac{1}{12} N \exp_2 \left(-2\tilde{R}_0/N + 2h(X_1^N)/N + 2\sum_i E[\log \lambda_i]/N \right) \left(\prod_i E\left[2^{\log g_i^2 - \log(\lambda_i^2)} \right] \right)^{1/N}.$$

Using Jensen's inequality, this may be lower bounded as follows:

$$\begin{aligned} D_G &\geq \frac{1}{12} N \exp_2 \left(-2\tilde{R}_0/N + 2h(X_1^N)/N + 2\sum_i E[\log \lambda_i]/N \right) \left(\prod_i 2^{E[\log g_i^2 - \log \lambda_i^2]} \right)^{1/N} \\ &= \frac{1}{12} N \exp_2 (-2\tilde{R}_0/N + 2h(X_1^N)/N + \sum_i E[\log g_i^2]/N). \end{aligned} \quad (6)$$

This lower bound is independent of λ_i, and achievable if we make the minimizing choice $\lambda_i \propto g_i$. Note that this optimal density is independent of the source distribution, analogous to the optimality of uniform quantization in ordinary (non-functional) variable-rate scalar quantization [14].

4 Functional Transform Coding

We now leave the distributed setting. The setup common to Feng et al. [1] and Bucklew [13] both require vector quantization—a computationally expensive premise if the form of the quantizer is to be left arbitrary. By constraining the quantizer to the form of a uniform transform code, we can reduce this cost while still exploiting redundancies between sources and properties of G.

Under the transform constraint (Fig. 4), an N vector of source variables, $X_1^N \in \mathbb{R}^N$ is first presented to an encoder. We constrain X_1^N so that its distribution, $f_{X_1^N}(x_1^N)$ is supported entirely within the N-sphere of radius 1. Following this,

1. An invertible linear transformation, U, is applied to X_1^N to yield vector $Y_1^N = UX_1^N$. For convenience, we constrain U to be of unity determinant.
2. The components of Y_1^N are uniformly scalar quantized into the vector $\widehat{Y}_1^N$. $\widehat{Y}_i$ has fixed rate R_i and resulting interval size 2^{-R_i}.

3. There is a total rate constraint: $\sum R_i \leq R$.
4. The decoder inverts the transformation $\widehat{X}_1^N = U^{-1}\widehat{Y}^N$, and computes an estimate of the function $\widehat{G}(X_1^N) \triangleq G(\widehat{X}_1^N)$.

U and the R_i are to be chosen to minimize the mean-squared error of G.

Uniform quantization entails high-rate point density functions over Y_1^N of the form $\lambda_i = 1$. Referring to (3), we may write the total distortion as

$$D_G = \tfrac{1}{12}\sum_i 2^{-2R_i} E\left[\left|\frac{\partial G_Y(y_1^N)}{\partial y_i}\right|^2\right]$$

where $G_Y(y_1^N) = G(U^{-1}y_1^N)$. Optimizing the rates R_i subject to the sum-rate condition,

$$D_G \geq \tfrac{1}{12} N 2^{-2R/N}\left[\prod_i E\left[\left|\frac{\partial G_Y(y_1^N)}{\partial y_i}\right|^2\right]\right]^{1/N}. \tag{7}$$

Note that this bound is achievable for finite R if and only if

$$E[|\partial G/\partial y_i|^2] > 0 \text{ for all } i = 1, \ldots, N. \tag{8}$$

Define the derivative vector, $\gamma_X(x_1^N)$ such that $\gamma_{Xi} = \partial G/\partial x_i$. Similarly, define $\gamma_{Yi} = \partial G_Y/\partial y_i$. Note that $\gamma_Y = U\gamma_X$. Define the matrix over X_1^N, $\Gamma_X(x_1^N) = \gamma_X\gamma_X^T$ and the matrix over Y^N, $\Gamma(y^N) = \gamma_Y\gamma_Y^T$. We then have the following properties:

1. Γ_X and Γ_Y are positive semidefinite, real, and symmetric.
2. $E[\Gamma_X]$ and $E[\Gamma_Y]$ inherit the three properties of Γ_X and Γ_Y.
3. $E[\Gamma_Y] = E[\gamma_Y\gamma_Y^T] = E[U\gamma_X\gamma_X^T U^{-1}] = UE[\Gamma_X]U^{-1}$.
4. The multiplicative trace of $E[\Gamma_Y]$ is the bracketed expression within (7).

By the Hadamard inequality, we can simultaneously minimize both the distortion and the multiplicative trace by choosing U to diagonalize $E[\Gamma_X]$. This yields total distortion

$$D_G = \tfrac{1}{12} N 2^{-2R/N} \det(E[\Gamma_X])^{1/N}. \tag{9}$$

If $E[\Gamma_Y]$ has any zero eigenvalues, $\det(E[\Gamma_X]) = 0$ and (8) has been violated. To correct for this, the zero components may be discarded, for G is independent of them with probability 1.

In general, use of the transform U reduces distortion by a factor $2^{-2R\overline{N}} \det(\overline{E[\Gamma_Y]})/E[\prod \gamma_{Xi}^2]$, where $\overline{N}$ denotes the number of null components, and $\overline{E[\Gamma_Y]}$ is the dimension-reduced $E[\Gamma_Y]$. Unlike the cases in Section 3, this distortion improvement is rate dependent when $\overline{N} > 0$.

Notice the similarity of the optimal transformation to that of the KLT for non-functional transform coding. The KLT diagonalizes the covariance of the random vector, X_1^N. U diagonalizes $E[\Gamma_X]$, which may be written as the sum of the covariance of the random vector γ_X and the matrix $A_{ij} = E[\partial G/\partial x_i]E[\partial G/\partial x_j]$.

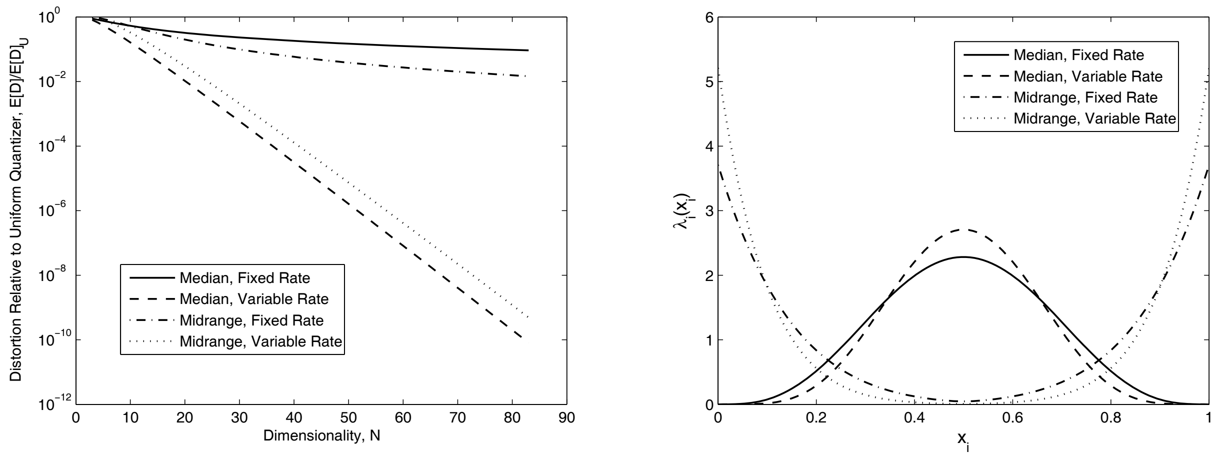

Figure 5: (Left) Distortion improvement against uniform quantizer as a function of number of variables N, on logarithmic scale; (Right) Optimal quantization point densities.

5 Examples

Intuitively, it should improve performance to account for the function G in the design of quantizers for $\{X_i\}_{i=1}^N$. We are interested in evaluating this improvement on a few sequences of functions, $G_N(X_1^N)$, $N \in \mathbb{Z}^+$. Specifically, suppose the source variables are i.i.d. uniform on $[0,1]$ and the decoder hopes to extract statistical properties from multiple samples of $f_X(x)$. We examine two such statistical properties: the median and the midrange. For a quantitative comparison, we calculate the ratio $D_G^{(\mathrm{F})}(\tilde{R})/D_G^{(\mathrm{O})}(\tilde{R})$, where $D_G^{(\mathrm{F})}(\tilde{R})$ and $D_G^{(\mathrm{O})}(\tilde{R})$ are the expected functional distortion of the functionally-optimized (F) or ordinary (O) quantizer at transmission rate $\tilde{R}$.

Ordinary Quantization. Optimal fixed- or variable-rate quantization for a uniform $[0,1]$ source X_i is uniform. The resulting performance in either case is $E[(X_i - \widehat{X}_i)^2] = \frac{1}{12}2^{-2R_i}$, where R_i is the rate allocated to X_i.

Fixed-Rate Quantization for the Median. For simplicity, restrict N to be odd valued with $N = 2M+1$. Given this, x_i is either itself the median of x_1^N or (differentially or locally) it has no bearing on the value of the median. Therefore, we have that

$$g_i^2(x) = \Pr[G(X_1^N) = x \mid X_i = x] = \binom{2M}{M} x^M (1-x)^M.$$

This yields an optimal point density $\lambda_i \propto g_i^2/3$ that reflects our intuition that more quantizer intervals should be assigned to the more important middle ground. This becomes increasingly true as the dimensionality is increased: the median's variance and $D_G^{(\mathrm{F})}(\tilde{R})$ fall as $1/N$. In contrast, since the median simply takes on one of the source values, $D_G^{(\mathrm{O})}(\tilde{R})$ does not depend on N. The ratio between these is depicted along with the optimal point density (for $N = 21$) in Fig. 5.

Fixed-Rate Quantization for the Midrange. The midrange is defined as the average of the minimum and the maximum components of x_1^N. No parity restriction on N is necessary to obtain clear results. Any x_i contributes to the midrange with slope $1/2$ if and only if it is the minimum or the maximum. As these events are disjoint, we may sum their contributions to write

$$g_i^2(x) = \tfrac{1}{4}\left(\Pr[\max(X_1^n = x) \mid X_i = x] + \Pr[\min(X_1^n = x) \mid X_i = x]\right).$$

Substituting in the probabilities,

$$g_i^2(x) = \frac{1}{4}(x^{N-1} + (1-x)^{N-1}).$$

One may obtain both the point density ($N = 10$) and the distortion ratio as before (Fig. 5). The optimal profile, weighed towards the extrema, is found to reduce distortion by a factor of $1/N^2$.

Variable-Rate Quantization for the Median and Midrange. Distortion for the functional quantizer is given by (6). Using this, we may find the ratio in distortion between the two quantizer's $D(\tilde{R})$ functions:

$$2^{E[\log \prod g_i^2]/N} / \|g_i^2\|_1. \quad (10)$$

For both median and midrange, this ratio is found to decay geometrically. This is illustrated by the linear curves in Fig. 5.

References

[1] H. Feng, M. Effros, and S. A. Savari, "Functional source coding for networks with receiver side information," in *Proc. Allerton Conf. Commun. Control Comput.*, Sep. 2004.

[2] V. Doshi, D. Shah, M. Medard, and S. Jaggi, "Distributed functional compression through graph coloring," in *Proc. Data Compression Conf. (DCC 2007)*, Snowbird, Utah, Mar. 2007, pp. 93–102.

[3] H. Luschgy and G. Pagès, "Functional quantization of Gaussian processes," *J. Funct. Anal.*, 193(2):486–531, 2002.

[4] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE T. Inf. Theory*, IT-19(4):471–480, 1973.

[5] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate region of the quadratic Gaussian two-terminal source-coding problem," *IEEE T. Inf. Theory*, submitted.

[6] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE T. Inf. Theory*, IT-22(1):1–10, 1976.

[7] R. Zamir, "The rate loss in the Wyner-Ziv problem," *IEEE T. Inf. Theory*, 42(6):2073–2084, 1996.

[8] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder," *Signal Process.*, 86(11):3160–3179, 2006.

[9] H. Yamamoto and K. Itoh, "Source coding theory for multiterminal communication systems with a remote source," *T. IECE Japan*, E63(10):700–706, 1980.

[10] H. Xie and A. Ortega, "Entropy- and complexity-constrained classified quantizer design for distributed image classification," in *IEEE Workshop on Multimedia Sig. Process*, Dec. 2002, pp. 77–80.

[11] L. Vasudevan, A. Ortega, and U. Mitra, "Application-specific compression for time delay estimation in sensor networks," in *Proc. 1st Int. Conf. Embedded Netw. Sensor Syst.*, Nov. 2003, pp. 243–254.

[12] S. Kassam, "Optimum quantization for signal detection," *IEEE T. Commun.*, 25(5):479–484, 1977.

[13] J. A. Bucklew, "Multidimensional digitization of data followed by a mapping," *IEEE T. Inf. Theory*, IT-30(1):107–110, 1984.

[14] R. M. Gray and D. L. Nehoff, "Quantization," *IEEE T. Inf. Theory*, 44(6):2325–2383, 1998.

[15] B. Farber and K. Zeger, "Quantization of multiple sources using nonnegative integer bit allocation," *IEEE T. Inf. Theory*, 52(11):4945–4964, 2006.

[16] A. Gersho, "Asymptotically optimal block quantization," *IEEE T. Inf. Theory*, 25(4):373–380, 1979.