February 18, 2003

# High Rate Quantization Theory
# or
# The Other Theory of Source Coding

Robert M. Gray
Information Systems Laboratory
Department of Electrical Engineering
Stanford, CA 94305
rmgray@stanford.edu

Recent results describe joint work with Tamás Linder. This work was partially supported by the National Science Foundation, Natural Sciences and Engineering Research Council (NSERC) of Canada, and by Hewlett Packard. Slides available at http://ee.stanford.edu:∼gray/highrate.pdf

# Introduction

$X$ a $k$-dimensional real random vector with a distribution $P_f$, absolutely continuous wrt Lebesgue measure $V$ $\Leftrightarrow$ pdf $f = dP_f/dV$ for which

$$\Pr(X \in F) = \int_F f(x)\, dV(x) \stackrel{\Delta}{=} \int_F f(x)\, dx$$

$$\text{Volume}(F) = V(F) = \int_F dx$$

Differential entropy $h(f) \stackrel{\Delta}{=} -\int f(x) \log f(x)\, dx$ exists and is finite.

log may be base 2 or $e$

## Vector quantizer $Q$:

- *encoder* $\alpha : \Re^k \to \mathcal{I}$ (index set)

  $\mathcal{S} = \{S_i = \{x : \alpha(x) = i\};\ i \in \mathcal{I}\}$ (encoder partition)

  $p_i = P_f(S_i) > 0$ (useful technical condition)

- *decoder* $\beta : \mathcal{I} \to \Re^k$

  reproduction codebook $\mathcal{C} = \{\beta(i);\ i \in \mathcal{I}\}$.
  assume codevectors are all distinct.

- *length function* $\ell$ in nats satisfies the Kraft inequality

$$\sum_i e^{-\ell(i)} \leq 1 \text{ (admissible)} \qquad (1)$$

Length $\ell$ is *instantaneous rate*

If all $\ell(i)$ equal $\Rightarrow$ *fixed rate* or *fixed length* code. Then

$$\ell(i) = \log ||\mathcal{C}||$$

Otherwise variable rate or variable length code.

Idea underlying $\ell$: uniquely decodable lossless code for indices.

---

Average rate (length) $R_f(Q) = \sum_i p_i \ell(i)$

Quantizer entropy: $H_f(Q) = H_f(\alpha) = -\sum_i p_i \ln p_i$, For admissible $\ell$, divergence inequality $\Rightarrow$

$$R_f(Q) \geq H_f(Q) \text{ with equality iff } \ell(i) = -\ln p_i.$$

Distortion measure $d(x, \hat{x}) \geq 0$,

Here focus on squared error: $d(x, \hat{x}) = ||x - \hat{x}||^2$

Average distortion $D_f(Q) = Ed(X, \beta(\alpha(X)))$.

Most ideas generalize to weighted quadratic distortion measures such as

$d(x, \hat{x}) = (x - \hat{x})^t B_x (x - \hat{x})$ and $d(x, \hat{x}) = (x - \hat{x})^t B_{\hat{x}} (x - \hat{x})$,

where $B_x$ or $B_{\hat{x}}$ is positive definite, and many others.

Optimal Performance: (operational distortion-rate function)

$$\delta_f(R) = \inf_{Q: R_f(Q) \leq R} D_f(Q)$$

# Theories of Quantization

Three classical approaches:

**Shannon Rate Distortion Theory** (Shannon, Gallager)
Shannon distortion-rate function $D(R)$ provides lower bound to $\delta_f(R)$, achievable in asymptopia of large dimension $k$ and fixed rate $R$ under suitable stationarity assumptions

**High rate theory** (Bennett, Lloyd, Zador, Gersho)
Optimal performance in asymptopia of fixed dimension $k$ and large rate $R$.

**Nonasymptotic (Exact) results** (Shannon, Steinhaus, Lloyd)
Necessary conditions for optimal codes ($\Rightarrow$ iterative design algorithms).

Shannon theory well known and highly developed.

High rate theory less well known in information theory and signal processing theory, but classic results ubiquitous in practice e.g.,

- 6 dB/bit quantization approximations,
- fact that uniform quantization + optimal lossless coding $\approx$ optimal for memoryless sources, "quarter bit" off [Gish and Pierce (1968)]

High rate theory took much longer to be made rigorous, still many unproved folk theorems.

Here focus on high rate theory + relevant exact results

Aside: There is also a body of work considering asymptotically large $k$ and $R$ together (e.g., consistency of Shannon and Zador bounds; high rate and high dimension optimality of lattice codes [Zamir and Feder, 1996]

*Goal:* Survey old and new results, conjectures, some open problems.

## Variations of quantization, e.g., [Gruber (2002)]

- Fejes Tóth's minimum sums of moments:
  What partition $\{S_i\}$ minimizes $\sum_i E[||X - E(X|X \in S_i)||^2]$?

- best approximation of probability measures by discrete measures and support sets of best approximating discrete measures

- the minimum error of numerical integration formulas for classes of Hölder continuous functions and optimum sets of nodes

- best volume approximation of convex bodies by circumscribed convex polytopes and the form of best approximating polytopes

- theories of $k$-means and of “principal points” in statistics

# Zador/Gersho High Rate Approximations

As rate $R$ goes to $\infty$

**Fixed rate:**

$$\delta_k(R) \cong a_k ||f||_{k/k+2} 2^{-2R/k}, \tag{2}$$

where $a_k$ is a universal constant depending only on $k$ and

$$||f||_{k/k+2} = \left( \int f(x)^{k/k+2} \, dx \right)^{\frac{k+2}{k}}$$

**Variable rate:**

$$\delta_k(R) \cong b_k 2^{-\frac{2}{k}[R-h(f)]}, \tag{3}$$

where $b_k$ is a universal constant depending only on $k$.

Zador's original proofs highly technical and limited generality. Also contained several errors, some serious.

Gersho developed heuristic approach and popularized results.

*Gersho's Conjecture* High rate results achieved by quantizers that locally resemble lattice or tessellating quantizers.

Conjecture leads to intuitive "proofs," but neither conjecture nor many of its implications have been proved.

If conjecture true, then $a_k = b_k$, all $k$.

Currently only $b_1 = a_1 = 1/12$ and $a_2$ known, and limits as $k \to \infty$ known. There are many bounds.

Gersho's proof (like Lloyd's and, implicitly, Bennett's) based on idea of *quantizer point density* $\lambda(x)$ of a sequence of quantizers $Q_n$: for every measurable set $F$:

$$\frac{\#\text{ of quantizer reproduction points of } Q_n \in F}{\text{Total }\#\text{ of quantizer reproduction points of } Q_n} \overset{\to}{n \to \infty} \int_F \lambda(x)\,dx$$

# Lagrangian Approach:

Necessary conditions (and design algorithms) long known for fixed rate codes (Shannon, Lloyd). Similar results for variable rate require Lagrangian approach.

Berger (1972), Farvardin & Modestino (1984) for $k = 1$, Chou et al. (1989) for general case.

Lagrangian methods for Shannon approach, i.e., for evaluating rate-distortion functions, older [Shannon, Gallager]

For each value of a Lagrangian multiplier $\lambda > 0$ define a Lagrangian distortion

$$\rho_\lambda(x, i) = d(x, \beta(i)) + \lambda \ell(i)$$

with average distortion

$$\rho(f, \lambda, Q) = E_f \rho_\lambda(X, \alpha(X)) = D_f(Q) + \lambda R_f(Q)$$

Optimal performance $\rho(f, \lambda) = \inf_{Q:\ \text{admissible } \ell} \rho(f, \lambda, Q)$

Each $\lambda \Rightarrow (D, R)$ pair on the operational distortion-rate function curve.

Small $\lambda$ corresponds to high rate (small distortion) and large $\lambda$ corresponds to small rate (large distortion).

As sweep $\lambda$ from 0 to $\infty$, traces convex hull of operational distortion-rate function

As $\lambda \to \infty$, put all cost on rate. Optimal code tends to zero rate code, suffer whatever distortion is necessary.

As $\lambda \to 0$, cost concentrates on distortion. If distribution were discrete, optimal code would be lossless, zero distortion with rate $H(f)$.

High rate theory considers distortion-rate tradeoff as $\lambda \to 0$ for continuous distributions.

# Lloyd Quantizer Optimality Properties:

$$Q = (\alpha, \beta, \ell)$$

**Encoder** For a given $\beta, \ell$, optimal encoder is the minimum Lagrangian distortion mapping

$$\alpha(x) = \operatorname*{argmin}_i \left( d(x, \beta(i)) + \lambda \ell(i) \right)$$

**Decoder** For a given $\alpha, \ell$ the optimal decoder is the Lloyd centroid $\beta(i) = \operatorname*{argmin}_y E(d(X, y) | \alpha(X) = i)$

**Length Function** For a given $\alpha, \beta$, optimal length function is the

Shannon codelength $\ell(i) = -\ln p_i$, where $p_i \triangleq P_X(\alpha(X) = i)$:

$$E\ell(\alpha(X)) = H(\alpha(X)) = -\sum_i p_i \ln p_i \qquad \text{(ECVQ)}$$

# Lloyd Algorithm

$\Rightarrow$ Clustering algorithm: Given an initial code, improve by iteratively optimizing each component for the others.

For fixed rate: Lloyd (1957), Steinhaus (1956), $k$-means, MacQueen (1967)

For variable rate: Chou et al. (1989)

Early example of "grouped coordinate descent algorithm"

Descent algorithm, so converges.

Can initialize in many ways, including splitting to grow codebook.

# Precise Statement of Zador's High Rate Results

Zador's fixed rate results corrected and generalized by Bucklew and Wise (1982) and Graf and Luschgy (2000). Basic result is

**Theorem.** *If* $E(||X||^{2+\epsilon}) < \infty$ *for some* $\epsilon > 0$, *then*

$$\lim_{R \to \infty} \delta_f(R) 2^{R/k} = a_k ||f||_{k/k+2}$$

Zador's variable rate results only recently repaired and generalized.

The traditional form of Zador and Gersho was that under certain conditions,

$$\lim_{R \to \infty} 2^{\frac{2}{k}R} \delta_f(R) = b(2,k) 2^{\frac{2}{k}h(f)} \tag{4}$$

but no rigorous proof of this result existed until recently.

**Theorem. [Gray, Linder, Li (2002)]** *Assume that $f$ is absolutely continuous wrt Lebesgue measure, that $h(f)$ is finite, and for some $\Delta$ a partition into cubes of side $\Delta$ has finite entropy, then*

$$\lim_{\lambda\to 0}\left(\inf_{Q}\left(\frac{E_f[d(X,\beta(\alpha(X)))]}{\lambda}+E_f\ell(\alpha(X))\right)+\frac{k}{2}\ln\lambda\right)=h(f)+\theta_k \tag{5}$$

*where*

$$\theta_k=\theta([0,1)^k)\stackrel{\Delta}{=}\inf_{\lambda>0}\left(\frac{\rho(u_1,\lambda)}{\lambda}+\frac{k}{2}\ln\lambda\right) \tag{6}$$

*and $u_1$ is the uniform pdf on the $k$-dimensional unit cube*

Analogous to the approximate interpretation of the traditional Zador result, the interpretation here is that for small $\lambda$,

$$\rho(f, \lambda) \approx \lambda\theta_k + \lambda h(f) - \frac{k}{2}\lambda \ln \lambda.$$

The following relates the traditional and Lagrangian forms of Zador's results for variable rate vector quantization.

**Lemma 1.** *The conclusions of Theorem 1 hold under the stated conditions if and only if the limit of (4) exists, in which case*

$$\theta_k = \frac{k}{2} \ln \frac{2e}{k} b_k. \tag{7}$$

Thus in particular Zador's formula holds under the conditions given in the theorem.

As an example of the conditions, the divergence inequality $\Rightarrow$ if the random vector $X$ has a finite second moment, then $H_f(Q_1) < \infty$ and $h(f) < \infty$. Thus the theorem holds for pdfs with finite second moment and if $h(f) > -\infty$, e.g., for Gaussian pdfs.

*Proof of theorem:*

- Uniform pdfs on cubes
- Piecewise constant pdfs on cubes
- General distributions on the unit cube
- General distributions.

Does not use idea of quantizer point density.

Does make use of "composite codes," quantizers designed for disjoint mixtures by designing separate quantizers for each piece and then using union codebook.

# Asymptotically Optimal Quantizers:

The theorem guarantees that for a source with pdf $f$, there is an *asymptotically optimal* sequence of quantizers: for any decreasing sequence $\lambda_n$ converging to 0 there exists a sequence of quantizers $Q_n = (\alpha_n, \beta_n, \ell_n)$ such that

$$\lim_{n\to\infty} \left( \left( \frac{E[d(X, \beta_n(\alpha_n(X)))]}{\lambda_n} + E\ell_n(\alpha_n(X)) \right) + \frac{k}{2} \ln \lambda_n \right) = h(f)\cdots$$

Note: Evaluation of $b_k$ can be accomplished by finding asymptotically optimal quantizers for *any* pdf, e.g., uniform pdf on a cube.

# Worst Cases

If pdf has finite support $\Omega$, then *worst case* is uniform pdf on $\Omega$:

$$f(x) = \frac{1}{V(\Omega)}, x \in \Omega \quad , \quad h(f) = \log V(\Omega)$$

If know mean $\mu = EX$ and covariance $K = E[(X-\mu)(X-\mu)^t]$ of the source, then *worst case* is Gaussian pdf:

$$f(x) = \mathcal{N}(x, \mu, K) = \frac{1}{(2\pi)^{\frac{k}{2}}|K|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^t K^{-1}(x-\mu)\right)$$

$$h(f) = \frac{1}{2}\ln(2\pi e)^k |K|,$$

# High rate analog to Sakrison's Shannon rate-distortion result

What if we design asymptotically optimal sequence of quantizers $Q_n$ for pdf $g$, but apply it to $f$? $\Rightarrow$ *mismatch*

# Mismatch

Optimize a code for a distribution $P_g$ on $\Re^k$, but apply the code to a distribution $P_f$

*Classic example:*

- Lossless source code
- Distributions are discrete, described by pmfs $g$ and $f$.

Uniquely decodable lossless code must have a collection of codeword lengths $\ell(i)$ in nats that satisfies the Kraft inequality (1)

If a discrete source has pmf $g = \{g_i\}$ with Shannon entropy

$$H(g) = -\sum_i g_i \ln g_i,$$

divergence inequality $\Rightarrow$ If $\ell$ admissible,

$$E_g \ell = \sum_i g_i \ell(i) \geq H(g), \text{ with equality if}$$

$\ell(i) = -\ln p_i$ (Ignore constraint of integer lengths)

Apply optimal code for pmf $g$ instead to pmf $f$:

$$\begin{aligned} E_f \ell &= \sum_i \ell(i) f_i = -\sum_i f_i \ln g_i \\ &= H(f) + \sum_i f_i \ln \frac{f_i}{g_i} \triangleq H(f) + I(f||g), \end{aligned}$$

$I(f||g)$ is the *relative entropy* or *Kullback-Leibler divergence*

Extend mismatch idea to fixed dimension high rate vector quantization.

For fixed rate codes, done by Bucklew (1984)

**Theorem.** *Suppose that $Q_n$ is asymptotically optimal for a pdf $g$ and that $f$ is pdf satisfying an horrendously complicated condition given by Bucklew, the only simple version of which is that $f/g$ is bounded. Then*

$$\lim_{R\to\infty} 2^{R/k} D_f(Q_R) = a_k \left( \int \frac{f(x)}{g(x)^{2/2+k}} \, dx \right) \left( \int g(x)^{k/k+2} \, dx \right)^{2/k}$$

Intuition: This is the result that follows from Gersho's conjecture using approximations to integrals and a quantizer point density.

In fact, as part of the proof, Bucklew demonstrated the existence of a quantizer point density.

Variable rate case:

**Theorem.** *The mismatch theorem:[Gray, Linder (2002)] Suppose that $Q_n$ is asymptotically optimal for $\lambda_n \to 0$ for a pdf $g$ and that $f$ is pdf for which $f/g$ is bounded, then*

$$\lim_{n\to\infty} \frac{D_f(Q_n)}{\lambda_n} + E_f \ell_n(\alpha_n(X)) + \frac{k}{2}\ln\lambda_n = \theta_k - \int dx\, f(x) \ln g(x)$$

*or, equivalently,*

$$\lim_{n\to\infty} \frac{D_f(Q_n)}{\lambda_n} + E_f \ell_n(\alpha_n(X)) + \frac{k}{2} \ln \lambda_n - h(f) = \theta_k + I(f||g),$$

*where*

$$I(f||g) = \int dx f(x) \ln \frac{f(x)}{g(x)}$$

*is the relative entropy or Kullback-Leibler divergence between $f$ and $g$.*

The mismatch theorem implies that if $Q_n$ asymptotically optimal for $g$, then when applied to $f$ it will yield *the asymptotically optimal performance for $f$ plus $I(f||g)$.*

Also *robust* in the sense of Sakrison's and Lapidoth's Shannon rate distortion results

For Gaussian (or uniform) $g$, $I(f||g) = h(g) - h(f)$ and hence

$$D_f(Q_R) \approx b_k 2^{-(2/k)[R-h(g)]} \tag{8}$$

which is the best asymptotic performance at rate $R$ for $g$.

So indeed robust quantizers in the Lapidoth sense

# Mismatch as Distortion between pdfs

The relative entropy quantifies the high rate mismatch from optimal performance of a quantizer optimized for a "model" pdf and then applied to a "true"

$\Rightarrow$ adds motivation for $I(f||g)$ as a "distance" or "distortion measure" on pdfs in order to "quantize" the space of pdf's to fit models to observed data.

# High Rate Variable Rate Universal Coding

**Corollary 1. [Gray and Linder (2002)]** *Suppose that* $Q_n = (\alpha_n, \beta_n, \ell_n)$ *is a sequence of variable rate quantizers that is asymptotically optimal for a pdf* $g$ *for some decreasing sequence* $\lambda_n \to 0$. *Assume also that* $f$ *is a pdf that meets the condition of the mismatch theorem. Define* $\ell'_n$ *to be the optimal length function for* $\alpha_n$ *and* $P_f$. *Then* $Q'_n = (\alpha_n, \beta_n, \ell'_n)$ *is asymptotically optimal for* $P_f$, *i.e.,* $\lim_{n\to\infty} \theta(f, \lambda_n, Q'_n) = \theta_k$.

The length function of the quantizer matched to the true source, but encoder not optimized for the new length function. Thus there remains a mismatch in the code sequence, which nonetheless is asymptotically optimal!

# Shortcoming

Constraint that $f/g$ be bounded too strong, e.g., eliminates Gaussian $g$ and a Laplacian $f$ and two Gaussians.

Gersho's conjecture suggests only need $I(f||g) < \infty$

Can show directly for some cases violating bounded condition, e.g., for $k = 1$ a sequence of uniform quantizers with optimal lossless codes for $g$ work for any $f$ for which $I(f||g) < \infty$

# Some Open Questions

- Does quantizer point density function exist for variable length case?
- If so, is it uniform as Gersho's conjecture and known $k = 1$ case implies?
- Does $a_k = b_k$ for $k > 1$?
- Does Gersho's conjecture hold?
- Can the mismatch theorem be generalized to unbounded $f/g$?

- Only case optimal high rate variable rate quantizers known is for $k = 1$ where uniform quantizers are best. Common conjecture is lattice or tessellating quantizers asymptotically optimal (implied by Gersho's conjecture). Is this conjecture true? (Known asymptotically true as $k \to \infty$ [Zamir and Feder (1996)])

- There may be other asymptotically optimal quantizers, mismatch theorem holds for any such sequence. Some are better than others, e.g., know for pdfs with bounded support or subGaussian tails, asymptotically optimal quantizers need only a finite number of quantization points.

- Do the results hold for more general distortions, especially those of the form

$$d(x, y) = (x - y)' B_x (x - y)?$$

  Results of Gardner et al., Li et al. using Gersho's conjecture and similar results for Shannon rate-distortion functions of Linder, Zamir, and Zeger suggest the answer is yes.

- Clustering pdfs with relative entropy distortion, applications to Gauss mixture modeling, compression, and classification. (Aiyer, Young, Pyun)

- Nonrigorous approach of Gersho yields simpler "proofs." Can the highly technical rigorous proofs be made more intuitive?

# Selected References


J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with $r$th power distortion measures," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 239–247, Mar. 1982.

J. A. Bucklew, "Two results on the asymptotic performance of quantizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 341–348, Mar. 1984.

P. A. Chou and B. Betts, "When optimal entropy-constrained quantizers have only a finite number of codewords," IEEE International Symposium on Information Theory. Full paper available at *http://research.microsoft.com/ pachou/publications.htm*.

P.A. Chou, T. Lookabaugh, and R.M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Acoust. Speech and Signal Proc.*, Vol. 37, pp. 31–42, Jan. 1989.

R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, NY, 1968.

W. R. Gardner and B. D. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters," *IEEE Trans. Speech and Audio Processing*, Vol 3, pp. 367-381, September 1995.

A. Gersho, "Asymptotically optimal block quantization," *IEEE Trans. Inform. Theory.*, vol. 25, pp. 373–380, July 1979.

H. Gish and J. N. Pierce, "Asymptotically efficient quantizing," *IEEE Trans. Inform. Theory*, vol. 14, pp. 676–683, Sep. 1968.

S. Graf and H. Luschgy, *Foundations of Quantization for Probability Distributions*, Springer, Lecture Notes in Mathematics, 1730, Berlin, 2000.

R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, pp. 2325–2384, Oct. 1998.

R.M. Gray, T. Linder, and J. Li, "A Lagrangian formulation of Zador's entropy-constrained quantization theorem," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 695–707, Mar. 2002.

R.M. Gray and T. Linder, "Mismatch in high rate entropy constrained vector quantization,"' *IEEE Trans. Inform. Theory*, to appear, `http://ee.stanford.edu/~gray/mismatch.pdf`

A. György, T. Linder, P. A. Chou, and B. J. Betts "When optimal entropy-constrained quantizers have a finite or infinite number of codewords," submitted for possible publication. Preprint available at `http://magenta.mast.queensu.ca/~linder/psfiles/GyLiChBe02.ps`

A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inform. Theory*, vol. 43, pp. 38–47, Jan. 1997.

J. Li, N. Chaddha, and R.M. Gray, "Asymptotic performance of vector quantizers with a perceptual distortion measure," *IEEE Trans. on Information Theory*, Vol. 45, pp. 1082–1091, May 1999.

J. MacQueen, "Some methods for classification and analysis of multivariate observations," *Proc. of the Fifth Berkeley Symp. on Math. Stat. and Prob.*, vol. 1, pp. 281–296, 1967.

D. J. Sakrison, "Worst sources and robust codes for difference distortion measures," *IEEE Trans. Inform. Theory*, vol. 21, pp. 301–309, May 1975.

C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE National Convention Record, Part 4*, pages 142–163, 1959.

H. Steinhaus, "Sur la division des corp materiels en parties," *Bull. Acad. Polon. Sci.*, C1. III, vol IV, 801–804, 1956.

P. L. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," Ph.D. Dissertation, Stanford University, 1963. (Also Stanford University Department of Statistics Technical Report.)

P. L. Zador, "Topics in the asymptotic quantization of continuous random variables," Bell Laboratories Technical Memorandum, 1966.

P. L. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inform. Theory*, vol. 28, pp. 139–148, Mar. 1982.

R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Info. Thy.*, pp. 1152–1159, July 1996.