A Zador-Like Formula for Quantizers Based on Periodic Tilings

Neil J. A. Sloane, *Fellow, IEEE,* and
Vinay A. Vaishampayan, *Member, IEEE*

***Abstract*—We consider Zador's asymptotic formula for the distortion-rate function for a variable-rate vector quantizer in the high-rate case. This formula involves the differential entropy of the source, the rate of the quantizer in bits per sample, and a coefficient G which depends on the geometry of the quantizer but is independent of the source. We give an explicit formula for G in the case when the quantizing regions form a periodic tiling of n-dimensional space, in terms of the volumes and second moments of the Voronoi cells. As an application we show, extending earlier work of Kashyap and Neuhoff, that even a variable-rate three-dimensional quantizer based on the "A15" structure is still inferior to a quantizer based on the body-centered cubic lattice. We also determine the smallest covering radius of such a structure.**

***Index Terms*—A15 structure, distortion-rate function, honeycomb, optimal covering, optimal quantizer, vector quantizer, Voronoi cell, Zador bound.**

I. Introduction

Zador gave two asymptotic formulas for the distortion-rate function for a vector quantizer in $\mathbb{R}^n$, depending on whether a fixed- or variable-rate code is used to transmit which cell the sample point belongs to. These are treated as [15, cases (A) and (B)]. See also [7, eqs. (19) and (20)]. In the present note we are concerned with the variable-rate case. Zador's formula for this case states that the distortion-rate function has the form

$$\delta(R) \cong G 2^{2h(X)} 2^{-2R} \tag{1}$$

where $\delta(R)$ is the average squared error per symbol, $h(X)$ is the differential entropy per dimension of the source X, R bits per symbol is the rate of the quantizer, and G depends on the quantizer but not on the source. This is from [15, p. 4] and [7, eq. (20)]. Note that, as pointed out in [7], the version of [16, eq. (1)] is incorrect. For further information see also [14] and [6, Ch. 5]. The value of G can therefore be used to compare different quantizers. To calculate G we may choose any convenient distribution for the source X.

Suppose first that the quantizer points form a lattice $\Lambda \subseteq \mathbb{R}^n$, with Voronoi cell $\mathcal{V}$ of volume $V = \sqrt{\det \Lambda}$. To calculate G we assume that X is uniformly distributed over $\mathcal{V}$. Then G becomes

$$G = \frac{1}{n} \frac{\int_{\mathcal{V}} \|x\|^2 \, dx}{V^{1+2/n}} \tag{2}$$

the familiar expression for the average mean-squared error per dimension of a lattice quantizer [3], [4, Ch. 2], [5, Ch. 10], [6, Ch. 5]. In this case, all the quantizing cells have the same volume, and the fixed-rate and variable-rate versions of Zador's formula coincide.

The main purpose of this note is to put on record an analog of (2) (see (8) and (9)) for the variable-rate Zador formula which applies to the case when the quantizing regions form a periodic tiling of $\mathbb{R}^n$, such as that formed by the Voronoi cells for a union of a finite number of translates of a lattice. This question arose in a recent investigation of quantizers that are based on writing a lattice as an intersection of several simpler lattices [11].

A second motivation for our work was a recent paper of Kashyap and Neuhoff [8], which makes use of a fixed-rate analog of (2) for periodic lattice quantizers (see (10) and (14) later). One of the goals of [8] was to see if the A15 arrangement of points in $\mathbb{R}^3$ that has recently arisen in several different contexts [9], [10], [12], [13] could produce a better three-dimensional quantizer than the body-centered cubic (bcc) lattice. The latter is known to be the best *lattice* quantizer in $\mathbb{R}^3$ [1], but the question of the existence of a better nonlattice quantizer remains open.

Kashyap and Neuhoff [8] found that with their figure of merit the best quantizer based on the A15 arrangement is inferior to the bcc lattice. In Section III, we repeat the comparison using our figure of merit. Now a different version of the A15 quantizer is best, but is still inferior to the bcc lattice.

Other applications of our formula will be found in [11].

Another open problem in three-dimensional geometry is to determine the best covering of $\mathbb{R}^3$ by equal (overlapping) spheres. In Section III, we determine the smallest covering radius that can be achieved with the A15 structure. This is also (slightly) worse than that of the bcc lattice.

II. Periodic Quantizers

Consider a vector quantizer in $\mathbb{R}^n$ in which the quantizing regions form a periodic tiling. Let $\mathcal{V}$ be a minimal periodic unit or tile for the tiling, and let $\mathcal{P}_1, \ldots, \mathcal{P}_k$ be a list of the different polytopes occurring among the quantizing regions. Let c_i be the centroid of $\mathcal{P}_i$. In order to determine G, we assume the source X is uniformly distributed over $\mathcal{V}$, and let p_i $(i = 1, \ldots, k)$ be the probability that X is in a cell of type $\mathcal{P}_i$. Also, let $N_i = p_i V / V_i$, where $V_i = \operatorname{vol} \mathcal{P}_i$ and $V = \operatorname{vol} \mathcal{V}$. Then N_i is the number of cells of type $\mathcal{P}_i$ per copy of $\mathcal{V}$, and

$$V = N_1 V_1 + \cdots + N_k V_k. \tag{3}$$

For example, let $\Lambda \subseteq \mathbb{R}^n$ be a lattice and let $a_1 + \Lambda, \ldots, a_r + \Lambda$ $(a_i \in \mathbb{R}^n)$ be distinct translates. Then the Voronoi cells for the union of the points $a_i + \Lambda$ $(i = 1, \ldots, r)$ form a quantizer of the type considered here.

We now apply (1). The left-hand side is the normalized mean-squared error per dimension $U/(nV)$, where

$$U = \sum_{i=1}^{k} N_i U_i = V \sum_{i=1}^{k} \frac{p_i U_i}{V_i} \tag{4}$$

$$U_i = \int_{\mathcal{P}_i} \|x - c_i\|^2 \, dx \tag{5}$$

is the unnormalized mean-squared error over a cell of type $\mathcal{P}_i$.

The differential entropy per dimension is

$$h(X) = \frac{1}{n} \log_2 V, \qquad \text{so} \quad 2^{2h(X)} = V^{2/n}. \tag{6}$$

It remains to calculate the rate R of the quantizer. Observe that we need

$$H(p_1, \ldots, p_k) = -\sum_{i=1}^{k} p_i \log_2 p_i$$

bits to specify the type of cell to which the quantized point belongs, and a further

$$\sum_{i=1}^{k} p_i \log_2 N_i = \sum_{i=1}^{k} p_i \log_2 (p_i V / V_i)$$

Manuscript received September 4, 2001; revised August 15, 2002.

The authors are with the Information Sciences Research Center, AT&T Shannon Labs, Florham Park, NJ 07932-0971 USA (e-mail: njas@research.att.com; vinay@research.att.com).

Communicated by R. Zamir, Associate Editor for Source Coding.
Digital Object Identifier 10.1109/TIT.2002.805086

bits to specify the particular one of the N_i cells of that type. This requires a total of $\log_2 V - \sum_{i=1}^k p_i \log_2 V_i$ bits, and then R is this quantity divided by n, so that

$$2^{-2R} = V^{-2/n} \prod_{i=1}^{k} V_i^{2p_i/n}. \tag{7}$$

We substitute (4), (6), and (7) into (1) to get our expression

$$G = \frac{\sum_{i=1}^{k} \frac{p_i U_i}{V_i}}{n \left(\prod_{i=1}^{k} V_i^{p_i} \right)^{\frac{2}{n}}} \tag{8}$$

for the average mean-squared error per dimension using variable-rate coding. The numerator of (8) is equal to U/V (see (4)), so we may rewrite (8) as

$$G = \frac{U}{nV \prod_{i=1}^{k} V_i^{2p_i/n}}. \tag{9}$$

In contrast, the expression given by Kashyap and Neuhoff [8] for fixed-rate coding is the following. Suppose the basic tile $\mathcal{V}$ contains L cells, $Q_1, \ldots, Q_L$, not assumed to be distinct, where Q_i has volume V_i and unnormalized second moment U_i (as in (5)). Then their expression is

$$G = \frac{L^{2/n} \sum_{i=1}^{L} U_i}{nV^{1+2/n}} \tag{10}$$

where $V = \operatorname{vol} \mathcal{V}$. The following justification is equivalent to the one in [8], but clarifies the difference between our two approaches. Consider a large region of space, B, which is partitioned into λ copies of $\mathcal{V}$, and let X be uniformly distributed over B. The left-hand side of (1) is

$$\frac{\lambda \sum_{i=1}^{L} U_i}{n \operatorname{vol} B} = \frac{\sum_{i=1}^{L} U_i}{nV}. \tag{11}$$

The differential entropy per dimension is

$$h(X) = \frac{1}{n} \log_2(\operatorname{vol} B), \qquad \text{so} \quad 2^{2h(X)} = (\lambda V)^{2/n}. \tag{12}$$

To compute the rate, it may be argued that $\log_2 \lambda$ bits are required to specify the copy of $\mathcal{V}$, and $\log_2 L$ bits to specify which of the Q_i's the point belongs to. This requires a total of $\log_2 \lambda L$ bits, so $R = (1/n) \log_2 \lambda L$, and

$$2^{-2R} = (\lambda L)^{-2/n}. \tag{13}$$

Arguing as before, we substitute (11)–(13) into (1), which gives (10).

If there are k different types of cells among the Q_i, with the jth cell occurring N_j times, then (10) can be written as

$$G = \frac{\left(\sum_{i=1}^{k} N_i \right)^{2/n} U}{nV^{1+2/n}}. \tag{14}$$

The ratio of the two expressions, (9) divided by (14), can be written as

$$2^{-\frac{2}{n}\left(\log_2 L - \sum_{i=i}^{k} p_i \log_2 (p_i/N_i)\right)}. \tag{15}$$

Both formulas, (9) and (14), depend only on the geometry of the quantizer. The difference between the two expressions arises because, as long as the cells do not all have the same volume, variable-length coding can take advantage of the different cell probabilities to reduce the overall rate.

If all cells have the same volume then (9) and (14) coincide, and if there is only type of cell then they both reduce to (2).

In general, the fact that (9) is less than or equal to (14) can be shown directly. After canceling some common factors and rearranging, we must show that

$$\prod_{j=1}^{k} V_j^{p_j} \geq \frac{1}{\sum_{i=1}^{k} \frac{p_i}{V_i}}$$

and this follows from the geometric-mean harmonic-mean inequality.

III. Quantizers Based on the A15 Structure

The A15 structure was discovered by Kasper *et al.* [9] in the clathrate compound Na_8Si_{46}. It was used by Weaire and Phelan ([13]; see also [12]) to construct a counterexample to Kelvin's conjecture on minimal surface soap films. It was also used by Lagarias and Shor [10] to construct counterexamples to Keller's conjecture in dimensions 10 and above. Since the soap film problem attempts to find a partition of space into cells which are good approximations to spheres, it was therefore natural to ask if the A15 structure could also lead to a record-breaking quantizer in three dimensions.

The A15 structure may be defined as the union of eight translates of the cubic lattice $4\mathbb{Z}^3$ by the vectors $(0, 0, 0)$, $(2, 2, 2)$, $(0 \pm 1, 2)$, $(2, 0, \pm 1)$, $(\pm 1, 2, 0)$. (The first two translates alone give the bcc lattice.)

There are two types of points in this structure, the *even* points in which all coordinates are even, and the *odd* points in which some coordinate is odd. There are isometries of $\mathbb{R}^3$ which map A15 to itself and act transitively on the even points and on the odd points. However, even points are not equivalent to odd points.

Weaire and Phelan consider a weighted Voronoi decomposition of $\mathbb{R}^3$: walls between points of the same type occur along perpendicular bisectors, but a wall between an even point E and an odd point D is defined by the plane

$$(X - E) \cdot (D - E) = \mu (D - E) \cdot (D - E)$$

where μ is a weighting factor to be determined. For $0 < \mu < 3/5$ there are two types of Voronoi cells: 12-sided polyhedra $\mathcal{P}_1$ centered at the even points and 14-sided polyhedra $\mathcal{P}_2$ centered at the odd points. Weaire and Phelan use the "Surface Evolver" computer program of Brakke [2] to perturb the polyhedra so that they have equal volumes and minimal total surface area.

Kashyap and Neuhoff [8] use the weighted Voronoi decomposition based on the A15 structure, but adjust μ to give the smallest value of (10). The formula are simpler if we set $\mu = 2\alpha/5$. Then Kashyap and Neuhoff find (and we have confirmed) that

$$V_1 = \operatorname{vol} \mathcal{P}_1 = 4\alpha^3$$

$$V_2 = \operatorname{vol} \mathcal{P}_2 = \frac{4}{3}(8 - \alpha^3)$$

$$U_1 = \frac{71}{30}\alpha^5$$

$$U_2 = \frac{1}{90}(1200 - 600\alpha^3 + 360\alpha^4 - 71\alpha^5).$$

The fundamental tile $\mathcal{V}$ has volume $V = 64$, and $L = 8$, $N_1 = 2$, $N_2 = 6$, $p_1 = \alpha^3/8$, $p_2 = 1 - \alpha^3/8$. Then (10) becomes

$$\frac{8^{2/3}(2U_1 + 6U_2)}{3 \cdot 64^{5/3}} = \frac{1}{96}(10 - 5\alpha^3 + 3\alpha^4). \tag{16}$$

This has a minimal value of $0.07873535\ldots$ at $\alpha = 5/4$. For comparison, the values of G for the bcc and face-centered cubic (fcc) lattices are $19/(192 \cdot 2^{1/3}) = 0.07854328\ldots$ and $2^{-11/3} = 0.07874507\ldots$, respectively [1], [3], [4].

Using our formula (8) we find that

$$G = \frac{2U_1 + 6U_2}{192\left(V_1^{\alpha^3/8} V_2^{1-\alpha^3/8}\right)^{2/3}} \tag{17}$$

which has a minimal value of $0.07872741\ldots$ at $\alpha = 1.2401\ldots$. This is slightly better, but still inferior to the bcc lattice.

Another unsolved problem is to find the best covering of $\mathbb{R}^3$ by overlapping spheres (cf. [4, Ch. 2, Table 2.1]). The bcc lattice is the best *lattice* covering, with thickness $1.4635\ldots$, but the question of the existence of a better nonlattice covering remains open. We find that the covering radius of the weighted Voronoi decomposition of A15 is minimized at $\alpha = 5/4$, which gives a thickness of

$$\frac{125\sqrt{(3)}\pi}{432} = 1.5745\ldots$$

just slightly worse than that of the bcc lattice (but again better than the fcc lattice, which has thickness $2.0944\ldots$).

Acknowledgment

The authors thank the referees for some helpful comments.

References

[1] E. S. Barnes and N. J. A. Sloane, "The optimal lattice quantizer in three dimensions," *SIAM J. Algebraic Methods*, vol. 4, pp. 30–41, 1983.
[2] K. Brakke, "The surface evolver," *Exper. Math.*, vol. 1, pp. 141–165, 1992.
[3] J. H. Conway and N. J. A. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 211–226, Mar. 1982.
[4] ——, *Sphere Packings, Lattices and Groups*, 3rd ed. New York: Springer-Verlag, 1998.
[5] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Boston, MA: Kluwer, 1992.
[6] R. M. Gray, *Source Coding Theory*. Boston, MA: Kluwer, 1990.
[7] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2325–2383, Oct. 1998.
[8] N. Kashyap and D. L. Neuhoff, "On quantization with the Weaire–Phelan partition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2538–2543, Sept. 2001.
[9] J. Kasper, P. Hagenmuller, M. Pouchard, and C. Cros, "Clathrate structure of silicon Na_8Si_{46} and Na_xSi_{136} ($x < 11$)," *Science*, vol. 150, pp. 1713–1714, 1965.
[10] J. C. Lagarias and P. W. Shor, "Cube-tilings of $\mathbb{R}^n$ and nonlinear codes," *Discr. Comput. Geom.*, vol. 11, pp. 359–391, 1994.
[11] N. J. A. Sloane and B. Beferull-Lozano, "Quantizing by lattice intersections," *Discr. Comput. Geom.*, to be published.
[12] D. Weaire, Ed., *The Kelvin Problem*. London, U.K.: Taylor and Francis, 1996.
[13] D. Weaire and R. Phelan, "A counter-example to Kelvin's conjecture on minimal surfaces," *Phil. Mag. Lett.*, vol. 69, pp. 107–110, 1994.
[14] P. L. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1963.
[15] ——, "Topics in the asymptotic quantization of continuous random variables," AT&T Bell Labs Tech. Memo., 51 pp., Murray Hill, NJ, Feb. 10, 1966.
[16] ——, "Asymptotic quantization error of continuous signals and their quantization dimension," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 139–148, Mar. 1982.

A PAC-Bayesian Margin Bound for Linear Classifiers

Ralf Herbrich and Thore Graepel

***Abstract*—We present a bound on the generalization error of linear classifiers in terms of a refined margin quantity on the training sample. The result is obtained in a probably approximately correct (PAC)-Bayesian framework and is based on geometrical arguments in the space of linear classifiers. The new bound constitutes an exponential improvement of the so far tightest margin bound, which was developed in the luckiness framework, and scales logarithmically in the inverse margin. Even in the case of less training examples than input dimensions sufficiently large margins lead to nontrivial bound values and—for maximum margins—to a vanishing complexity term. In contrast to previous results, however, the new bound does depend on the dimensionality of feature space. The analysis shows that the classical margin is too coarse a measure for the essential quantity that controls the generalization error: the fraction of hypothesis space consistent with the training sample. The practical relevance of the result lies in the fact that the well-known support vector machine is optimal with respect to the new bound only if the feature vectors in the training sample are all of the same length. As a consequence, we recommend to use support vector machines (SVMs) on normalized feature vectors only. Numerical simulations support this recommendation and demonstrate that the new error bound can be used for the purpose of model selection.**

***Index Terms*—Bayes classification strategy, computational learning theory, generalization error bound, Gibbs classification strategy, linear classifiers, margin, model selection, probably approximately correct (PAC)-Bayesian framework, support vector machine (SVM), volume ratios.**

I. Introduction

Linear classifiers are popular in the machine learning and statistics communities due to their straightforward applicability and high flexibility that has been greatly improved by the so-called kernel method [1]. A natural and popular framework for the theoretical analysis of classifiers is the *probably approximately correct* (PAC) framework [2] which is closely related to Vapnik's [3] work on the generalization error. For binary classifiers it turned out that the *growth function* is an appropriate measure of "complexity" and can tightly be upper-bounded by the Vapnik–Chervonenkis (VC) dimension [4]. *Structural risk minimization* [3] was suggested for directly minimizing the VC dimension based on a training sample and an *a priori* structuring of the hypothesis space.

In practice, for example in the case of linear classifiers, often a thresholded *real-valued* function is used for classification. In 1993, Kearns and Schapire [5] demonstrated that considerably tighter bounds can be obtained by considering a scale-sensitive complexity measure known as the *fat shattering dimension*. Further results [6] provided bounds on the growth function similar to those proved by Vapnik and others [4], [7]. The popularity of the theory greatly increased by the invention of the *support vector machine* (SVM) [1] which aims at directly minimizing the complexity as suggested by theory.

Manuscript received February 16, 2001; revised June 20, 2002. The material in this correspondence was presented in part at the Conference on Neural Information Processing Systems (NIPS 13), Denver, CO, December 2000.

R. Herbrich was with the Technical University of Berlin, Statistics and Business Mathematics, Sekr. FR 6-9, 10587 Berlin, Germany. He is now with Microsoft Research Cambridge, Cambridge CB3 0FB, U.K. (e-mail: rherb@microsoft.com).

T. Graepel was with the Technical University of Berlin, Statistics and Business Mathematics, 10587 Berlin, Germany. He is now with the Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, U.K. (e-mail: thore@cs.rhul.ac.uk).

Communicated by G. Lugosi, Associate Editor for Nonparametric Estimation, Classification, and Neural Networks.

Digital Object Identifier 10.1109/TIT.2002.805090