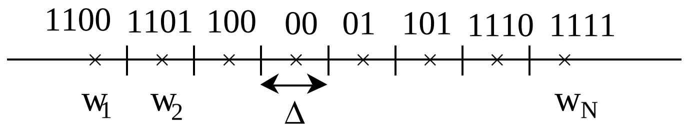
VARIABLE-RATE VQ (AKA VQ WITH ENTROPY CODING)

- Variable-Rate VQ = Quantization + Lossless Variable-Length Binary Coding
- A range of options -- from simple to complex

  a. Uniform scalar quantization with variable-length coding, one index at a time.



  b. Nonuniform scalar quantization with variable-length coding -- one index at a time.

  c. Scalar quantization with higher-order variable-length coding -- either block coding of n indices at a time or nth-order conditional coding of the indices.

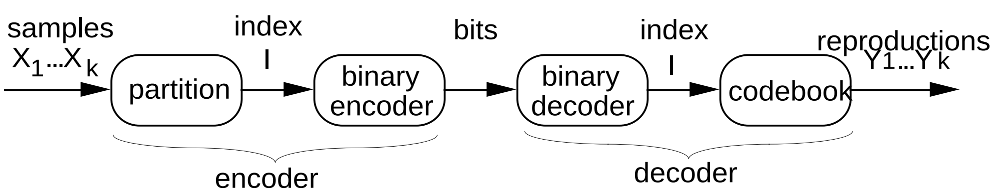
  d. k-dimen'l VQ with variable-length coding, one index at a time.

  e. k-dimen'l VQ with higher-order variable-length coding -- either block coding of n indices at a time or nth-order conditional coding of the indices.

  f. k-dimen'l VQ with other types of lossless coding

- We study e. with block coding. Conditional coding is just a slight variation. a.-d. are special cases of e.

D. VQ WITH 1ST-ORDER LOSSLESS CODING



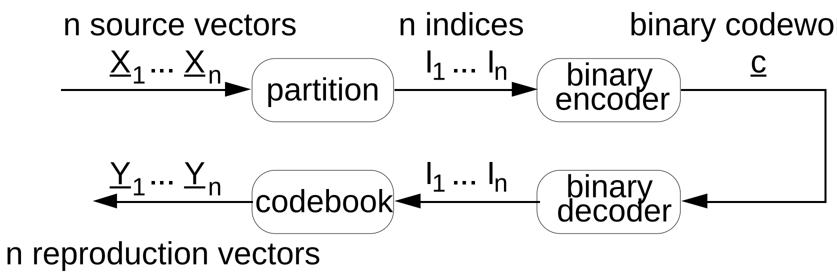
- Key characteristics: k-dimensional VQ with partition $S = \{S_1, ..., S_M\}$, codebook $C = \{\underline{w}_1, ..., \underline{w}_M\}$, quantization rule Q, and binary **prefix** codebook $B = C_b = \{\underline{c}_1, ..., \underline{c}_M\}$ with lengths $\{L_1,...,L_M\}$
- Decompose encoder into "partition" and "binary encoder"

  Given $\underline{x}$, the partition produces index i when $\underline{x} \in S_i$
  The binary encoder outputs binary codeword $\underline{c}_i$ with length l_i
- Decompose decoder into "binary decoder" and "codebook"

  The binary decoder decodes the bits into the index i.
  The codebook outputs $\underline{w}_i$.
- Rate = $R = \frac{1}{k} \sum_{\underline{x}} L(\underline{x})\, p(\underline{x}) = \frac{1}{k} \times$ rate of binary encoder

  we often assume $R = \frac{1}{k} H(I) = \frac{1}{K} H(Q(X))$ ("VQ with entropy coding")
- Distortion = $D = E \frac{1}{k} ||\underline{X}-Q(\underline{X})||^2$ (not affected by choice of lossless coder)

E. VQ WITH BLOCK LOSSLESS CODING

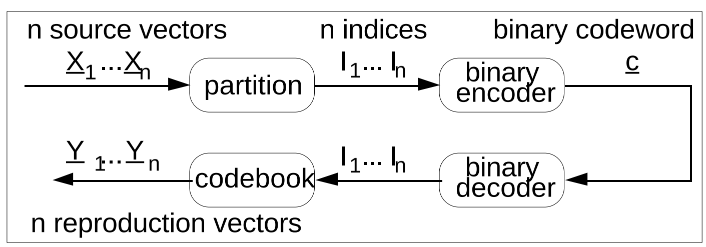


- k-dimen'l VQ with variable-length coding of indices in blocks of n.
- Decompose encoder into "partition" and "binary encoder":

  "Partition" n successive k-dim'l source vectors $\underline{X}_1,\ldots,\underline{X}_n$ into indices $I_1,\ldots,I_n$, where $\underline{X}_j = (X_{j,1},\ldots,X_{j,k})$. Losslessly encode n indices at once, $(I_1,\ldots,I_n)$, using FVL block lossless code with prefix codebook C_b containing M^n binary codewords.
- Decompose decoder into "binary decoder" and "codebook":

  Decode binary codeword into n indices $I_1,\ldots,I_n$. Output corresponding quantization vectors $\underline{w}_{I_1},\ldots,\underline{w}_{I_n}$ as reproductions of $\underline{X}_1,\ldots,\underline{X}_n$, respectively.
- "Block lossless binary coding" is an easy to analyze paradigm for studying the benefits of variable-length coding (a.k.a entropy coding).
- **We usually assume source is stationary**, so that $\underline{X}_j$ has the same pdf for all j, which will be denoted $f_{\underline{X}}(\underline{x})$, $f(\underline{x})$ or $f_k(\underline{x})$.

SUMMARY OF CHARACTERISTICS



Quantizer = k-dimensional VQ

- k = dimension
- M = size (not neces'ly power of 2, not so important, **may be infinite**)
- $S = \{S_1,\ldots,S_M\}$
  $= $ k-dimensional partition
- $C = \{\underline{w}_1,\ldots,\underline{w}_M\}$ = k-dimensional codebook
- $Q(\underline{x})$ = quantization rule

Binary Encoder:

- n = order of binary encoder (i.e. input blocklength)
- $C_b = \{\underline{c}_{\underline{i}} : \underline{i} \in \underline{I}\}$ = binary prefix codebook -- one cdwrd for each seq. of n indices

  where $\underline{I}$ = set of cell index n-tuples = $\{ \underline{i} = (i_1,\ldots,i_n) : 1 \le i_1 \le M, \ldots, 1 \le i_n \le M \}$

  $\underline{c}_{\underline{i}} = (c_{\underline{i},1},\ldots,c_{\underline{i},L_{\underline{i}}})$ = binary codeword of length $L_{\underline{i}}$ for $\underline{i}$

Derivative Characteristics:

- quantization rule: $Q(\underline{x}_j) = \underline{w}_i$ when $\underline{x}_j \in S_i$
- encoding rule: $e(\underline{x}_1,\ldots,\underline{x}_n) = \underline{c}_{\underline{i}}$, when $\underline{i} = (i_1,\ldots,i_n)$ and $\underline{x}_j \in S_{i_j}$, $j = 1,\ldots,n$
- decoding rule: $d(\underline{c}_{\underline{i}}) = (\underline{w}_{i_1}, \underline{w}_{i_1}, \ldots, \underline{w}_{i_1})$

Performance

Distortion (same as usual)

$$D = \frac{1}{k} E\,||\underline{X} - Q(\underline{X})||^2 = \frac{1}{k}\sum_{i=1}^{M} \int_{S_i} ||\underline{x}-\underline{y}_i||^2 f_k(\underline{x})\, d\underline{x}$$

where $\underline{X} = (X_1,\ldots,X_k)$ and $f_k(\underline{x})$ is its density. Dist'n depends on S and C but not C_b.

Rate

$$R = \frac{1}{kn}\overline{L} = \frac{1}{kn}\sum_{\underline{i}} P_{\underline{i}} L_{\underline{i}} \text{ bits/sample}$$

where $\overline{L}$ = average length of binary codewords

$P_{\underline{i}}$ = probability of binary codeword $\underline{c}_{\underline{i}}$ = $\Pr(\underline{X}_1 \in S_{i_1},\ldots,\underline{X}_n \in S_{i_n})$, $\underline{i} \in I$

From lossless coding theorem

$$H(\underline{I}) \le \overline{L}^*_n \le H(\underline{I}) + 1$$

where $\overline{L}^*_n$ = least avg. length of prefix code for given VQ & n

$$H(\underline{I}) = -\sum_{\underline{i}} P_{\underline{i}} \log_2 P_{\underline{i}} = \text{entropy of } \underline{I} \text{ (or of } (\underline{Y}_1,\ldots,\underline{Y}_n))$$

From now on we assume (unless otherwise stated) that

$$R = \frac{1}{kn} H(\underline{I}) = \frac{1}{k} H_n(I) = \frac{1}{kn} H(\underline{Y}_1,\ldots,\underline{Y}_n) = H_{kn}(Y)$$

We call this "**VQ with nth-order entropy coding (EC)**".

Implementation and Complexity

- Quantizer -- same issues as with fixed-rate coding
- Lossless Coder -- table lookup is the brute force method
  + Table stores M^n binary codewords of various lengths
  + $M = 2^{nkR_f}$ where $R_f = \frac{1}{k}\log_2 M$ is "fixed-rate" rate
  + Complexity of brute force implementation of entropy increases exponentially with nkR_f.

Optimal Performance

- OPTA functions we seek

$\delta(k,n,R) \triangleq$ least MSE of k-dim'l VQ w. nth-order entropy coding and rate R or less

$S(k,n,R) \triangleq$ max SNR of k-dim'l VQ's w. nth-order entropy coding and rate R or less

$\delta(R) \triangleq \inf_{k,n} \delta(k,n,R)$ = least MSE of VQ with EC and rate R or less (any k,n)

$S(R) \triangleq \sup_{k,n} \delta(k,n,R)$ = max SNR of VQ with EC and rate R or less (any k,n)

HIGH-RESOLUTION ANALYSIS

- As before, assume the VQ has mostly small cells, negligible overload distortion, large M, neighboring cells with similar sizes & shapes, point density approx'ly $\Lambda(\underline{x})$, inertial profile approx'y $m(\underline{x})$
- Since quantizer size is unimportant (e.g. its not related to rate), and can even be infinite, we use unnormalized point density, $\Lambda(\underline{x})$, which is a function such that

1. $\int_A \Lambda(\underline{x})\, d\underline{x} \cong$ number of codevectors (or cells) in region A
2. If A is small, but much larger than the cells in the vicinity of $\underline{x}$,

   $\Lambda(\underline{x})\,|A| \cong$ # points\cells in A
3. $\Lambda(\underline{x}) \geq 0$, $\int \Lambda(\underline{x})\, d\underline{x} = M =$ total number of quantization points (can be ∞)
4. Ordinarily $\Lambda(\underline{x})$ is a smooth or piecewise smooth function.
5. $\Lambda(\underline{x}) \cong \frac{1}{|S_i|}$ when $\underline{x} \in S_i$

DISTORTION: BENNETT'S INTEGRAL

Under high-resolution conditions, a derivation like that for the original Bennett shows

$$D \cong \int \frac{m(\underline{x})}{\Lambda^{2/k}(\underline{x})} f_k(\underline{x})\, d\underline{x}$$

RATE: ASYMPTOTIC ENTROPY FORMULA

Fact: If $\underline{X}_1,\ldots,\underline{X}_n$ are identically distributed, then under high-resolution conditions,

$$R = \frac{1}{kn} H(\underline{I}) \cong h_{kn} + \frac{1}{k} \int f_k(\underline{x}) \log_2 \Lambda(\underline{x})\, d\underline{x}$$

where

$\underline{x} = (x_1 \ldots x_k)$

$h_{kn} = \frac{1}{kn} h(X_1,\ldots,X_{kn}) =$ kn-th order differential entropy

$$= -\frac{1}{kn} \int f_{kn}(x_1 \ldots x_{kn}) \log_2 f_{kn}(x_1 \ldots x_{kn})\, dx_1 \ldots dx_{kn}$$

Most Important Example:

Uniform scalar quantizer with step size Δ, infinite support, and infinitely many levels

$$\Lambda(x) \cong \frac{1}{\Delta}$$

Then from the approximate rate formula

$$R \cong h_{kn} + \frac{1}{k}\int f_k(\underline{x}) \log_2 \Lambda(\underline{x})\, d\underline{x}$$

$$= h_n - \log\Delta = h_n - \frac{1}{2}\log 12\frac{\Delta^2}{12}$$

$$\cong h_n - \frac{1}{2}\log 12\, D$$

Equivalently

$$D \cong \frac{1}{12} 2^{2h_n} 2^{-2R}$$

DERIVATION OF ASYMPTOTIC FORMULA FOR $H(\underline{I})$

First case: $n = 1$ (for simplicity)

$$H(I) = -\sum_i P_i \log P_i\,, \quad \text{where } P_i = \Pr(\underline{X}_1 \in S_i) = \int_{S_i} f_k(\underline{x})\, d\underline{x}$$

$$= -\sum_i \left(\int_{S_i} f_k(\underline{x})\, d\underline{x}\right) \log\left(\int_{S_i} f_k(\underline{x})\, d\underline{x}\right)$$

$$\cong -\sum_i \left(f_k(\underline{w}_i)\int_{S_i} d\underline{x}\right) \log\left(f_k(\underline{w}_i)\int_{S_i} d\underline{x}\right) \quad \text{because cells are small}$$

$$= -\sum_i f_k(\underline{w}_i)\,|S_i| \log\left(f_k(\underline{w}_i)\,|S_i|\right)$$

$$= -\sum_i \left(f_k(\underline{w}_i) \log f_k(\underline{w}_i)\right)|S_i| - \sum_i \left(f_k(\underline{w}_i) \log \frac{1}{\Lambda(\underline{w}_i)}\right)|S_i|)$$

$$\cong -\int f_k(\underline{x}) \log f(\underline{x})\, d\underline{x} + \int f_k(\underline{x}) \log \Lambda(\underline{x})\, d\underline{x}$$

$$= h_k + \frac{1}{k}\int f_k(\underline{x}) \log \Lambda(\underline{x})\, d\underline{x}$$

General case: $n \geq 1$

$$H(\underline{I}) = -\sum_{\underline{i}} P_{\underline{i}} \log P_{\underline{i}}$$

$$= -\sum_{\underline{i}} \left(\int_{S_{\underline{i}}} f_{kn}(\underline{x})\, d\underline{x} \right) \log \left(\int_{S_{\underline{i}}} f_{kn}(\underline{x})\, d\underline{x} \right), \quad \text{where } \underline{x} = (x_1 \ldots x_{kn}),\ S_{\underline{i}} = (S_{i_1} \times S_{i_2} \times \ldots \times S_{i_n})$$

$$\cong -\sum_{\underline{i}} \left(f_{kn}(\underline{w}_{\underline{i}}) \int_{S_{\underline{i}}} d\underline{x} \right) \log \left(f_{kn}(\underline{w}_{\underline{i}}) \int_{S_{\underline{i}}} d\underline{x} \right), \quad \text{where } \underline{w}_{\underline{i}} = (\underline{w}_{i_1}, \underline{w}_{i_2}, \ldots, \underline{w}_{i_n})$$

$$= -\sum_{\underline{i}} f_{kn}(\underline{w}_{\underline{i}})\, |S_{\underline{i}}| \log \left(f_{kn}(\underline{w}_{\underline{i}})\, |S_{\underline{i}}| \right)$$

$$= -\sum_{\underline{i}} \left(f_{kn}(\underline{w}_{\underline{i}}) \log f_{kn}(\underline{w}_{\underline{i}}) \right) |S_{\underline{i}}| - \sum_{\underline{i}} \left(f_{kn}(\underline{w}_{\underline{i}}) \log |S_{\underline{i}}| \right) |S_{\underline{i}}|$$

The first summation above can be approximated by the integral

$$-\int f_{kn}(\underline{x}) \log f_{kn}(\underline{x})\, d\underline{x} = kn\, h_{kn} \qquad (*)$$

Before approximating the second sum, note that

$$\log |S_{\underline{i}}| = \log \left(|S_{i_1}|\, |S_{i_2}| \ldots |S_{i_n}| \right) = \sum_{j=1}^{n} \log |S_{i_j}| \cong -\sum_{j=1}^{n} \log \Lambda(\underline{w}_{i_j})$$

Substitute the above into the second summation:

$$-\sum_{\underline{i}} \left(f_{kn}(\underline{w}_{\underline{i}}) \log |S_{\underline{i}}| \right) |S_{\underline{i}}|$$

$$= -\sum_{\underline{i}} \left(f_{kn}(\underline{w}_{\underline{i}}) \left(-\sum_{j=1}^{n} \log \Lambda(\underline{w}_{i_j}) \right) \right) |S_{\underline{i}}|$$

$$\cong \int f_{kn}(\underline{x}_1 \ldots \underline{x}_n) \sum_{j=1}^{n} \log \Lambda(\underline{x}_j)\ d\underline{x}_1 \ldots d\underline{x}_k$$

$$= \sum_{j=1}^{n} \int f_{kn}(\underline{x}_1 \ldots \underline{x}_n) \log \Lambda(\underline{x}_j)\ d\underline{x}_1 \ldots d\underline{x}_k$$

$$= \sum_{j=1}^{n} \int f_k(\underline{x}_j) \log \Lambda(\underline{x}_j)\ d\underline{x}_j$$

$$= n \int f_k(\underline{x}_1) \log \Lambda(\underline{x})\, d\underline{x}_1 \quad \text{because } \underline{X}_1, \ldots, \underline{X}_n \text{ are identical} \qquad (**)$$

Substituting $(*)$ and $(**)$ into the expression for $H(\underline{I})$ gives

$$\frac{1}{kn} H(\underline{I}) \cong \frac{1}{kn} \left(kn\, h_{kn} + n \int f_k(\underline{x}_1) \log \Lambda(\underline{x})\, d\underline{x}_1 \right)$$

$$= h_{kn} + \frac{1}{k} \int f_k(\underline{x}) \log_2 \Lambda(\underline{x})\, d\underline{x}$$

Zador-Gersho Theorem for Variable-Rate VQ

For a stationary source and large R, the least distortion of k-dim'l VQ with nth-order entropy coding and rate R or less is

$$\delta(k,n,R) \cong m_k^* \sigma^2 \eta_{kn} 2^{-2R} \triangleq \mathcal{Z}(k,nR)$$

where

$\mathcal{Z}(k,n,R)$ = Zador-Gersho function for k-dim'l VQ with nth-order entropy coding

m_k^* = best inertial profile = least NMI of any tessel'g polytope (Gersho's conj.)

$\eta_{kn} = \frac{1}{\sigma^2} 2^{2h_{kn}}$

Equivalently, $S(k,n,R) \cong 6.02 R - 10 \log_{10} m_k^* \eta_{kn}$. (Again, 6 dB per bit.)

Notes:

- k-dimensional VQ-EC with n = 1 is at least as good as VQ-FR, because the latter is a special case of the former.

  Later we show directly that $\eta_k \le \beta_k$, which implies $\mathcal{Z}(k,1,R) \le \mathcal{Z}(k,R)$.
- The proof shows that an approximately optimal k-dimensional VQ with nth-order entropy coding can be constructed with a partition that is a tesselation of the best k-dimensional polytope, scaled to volume $2^{k(h_{kn}-R)}$. This has

  constant inertial profile $m(\underline{x}) = m_k^*$,

  constant point density $\Lambda(\underline{x}) = \Lambda_k^* = 2^{k(R-h_{kn})}$,

  distortion $D \cong m_k^* (\Lambda_k^*)^{-2/k} = \mathcal{Z}(k,nR)$.

Proof of Zador-Gersho Theorem

We begin with

$$\delta(k,n,R) \cong \min_{m(\underline{x}),\Lambda(\underline{x})} \int \frac{m(\underline{x})}{\Lambda^{2/k}(\underline{x})} f(\underline{x})\, d\underline{x}$$

where $\underline{x} = (x_1 \ldots x_k)$ and the minimization is over all inertial profiles $m(\underline{x})$ and all point densities $\Lambda(\underline{x})$ such that

$$h_{kn} + \frac{1}{k}\int f(\underline{x}) \log \Lambda(\underline{x})\, d\underline{x} \le R \qquad (\text{such a } \Lambda \Rightarrow \text{rate} \lesssim R)$$

- Best inertial profile:

  We assume Gersho's conjecture -- In the high rate, small distortion regime, most cells of an optimal quantizer are, approximately, congruent to the tesselating polytope with least NMI.

  We conclude: The best inertial profile is

  $$m_k^*(x) = m_k^* \triangleq \text{least NMI of k-dimen'l tess'ng polytopes}$$
- It follows that

  $$\delta(k,n,R) \cong m_k^* \min_{\Lambda(\underline{x})} \int \frac{1}{\Lambda^{2/k}(\underline{x})} f(\underline{x})\, d\underline{x}$$

  where the min is taken over functions $\Lambda(\underline{x})$ such that $\Lambda(\underline{x}) \ge 0$ and

  $$h_{kn} + \frac{1}{k}\int f(\underline{x}) \log \Lambda(\underline{x})\, d\underline{x} \le R$$

- Best point density:

Suppose $\Lambda(\underline{x})$ satisfies

$$h_{kn} + \frac{1}{k}\int f(\underline{x}) \log_2 \Lambda(\underline{x})\, d\underline{x} \le R \qquad (*)$$

Then by convexity of the logarithm and Jensen's inequality

$$\log_2 \int \frac{1}{\Lambda^{2/k}(\underline{x})} f(\underline{x})\, d\underline{x} \ \ge\ \int \log_2\left(\frac{1}{\Lambda^{2/k}(\underline{x})}\right) f(\underline{x})\, d\underline{x}$$ equality iff $\Lambda(\underline{x})$ is constant with probability one

$$= -\frac{2}{k}\int f(\underline{x}) \log \Lambda(\underline{x})\, d\underline{x} \ \ge\ 2h_{kn} - 2R$$

Hence,

$$\int \frac{1}{\Lambda^{2/k}(\underline{x})} f(\underline{x})\, d\underline{x} \ \ge\ 2^{2h_{kn}} 2^{-2R} \qquad (**)$$

with equality iff $\Lambda(\underline{x})$ is a constant with probability one.

$(*)$ and $(**)$ $\Rightarrow$ $\delta(k,n,R) \cong m_k^* 2^{2h_{kn}} 2^{-2R} = m_k^* \sigma^2 \eta_{kn}\, 2^{-2R}$

Moreover, we have shown that the optimal point density is a constant.
The constant must be such that $(*)$ holds with equality. Therefore,

$$\Lambda(\underline{x}) = 2^{k(R-h)} \triangleq \Lambda_k^*$$

We see from the proof that an approximately VQ can be constructed with a partition that is a tesselation of the best k-dimensional polytope, scaled to volume $2^{k(h_{kn}-R)}$. Indeed, the tesselation need only cover the region where $f(\underline{x})$ is not small.

SUMMARY OF FIXED- AND VARIABLE-RATE VQ

Let 0th-order entropy coding denote fixed-rate coding.

Then, given a stationary source and large R, the least distortion of VQ with rate R or less, dimension k, and nth-order order entropy coding is

$$\delta(k,n,R) \cong Z(k,n,R) = m_k^* \sigma^2 \eta_{k,n}\, 2^{-2R},$$

where

$$\eta_{k,n} = \begin{cases} \beta_k, & n=0 \\ 2^{2h_{kn}}/\sigma^2, & n\ge 1 \end{cases}$$

Notice the "," in $\eta_{k,n}$ but not in h_{kn}!