**Example: Polar Quantization**

Quantize $X = (X_1, X_2)$ by independently scalar quantizing its magnitude and phase:

![Diagram of polar quantization](image)

**Magnitude:**

$A = ||X|| = \sqrt{X_1^2 + X_2^2}$ = magnitude/amplitude,  $A \geq 0$

**Phase:**

$\Phi = \angle X = \text{angle or phase of } X = (X_1, X_2), \quad 0 \leq \Phi < 2\pi$

$\Phi = \begin{cases} 
\tan^{-1}\frac{X_2}{X_1}, & x_2 \geq 0 \\
\tan^{-1}\frac{X_2}{X_1} + \pi, & x_2 < 0
\end{cases}$, (assuming $0 \leq \tan^{-1} z < \pi$)

**Magnitude quantizer:** $\hat{A} = Q_A(A), \quad C_A = \{w_1, \ldots, w_{M_A}\}, \quad S_A = \{S_1, \ldots, S_{M_A}\}$

**Phase quantizer:** $\hat{\Phi} = Q_\Phi(\Phi), \quad C_\Phi = \{v_1, \ldots, v_{M_\Phi}\}, \quad T_\Phi = \{T_1, \ldots, T_{M_\Phi}\}$

---

**Polar Quantization is a kind of 2-dimensional VQ.**

**Dimension:** $k = 2$

**Size:** $M = M_A \times M_\Phi$

**Codebook:** $C = \{w_{i,j}\}$, where $w_{i,j} = (w_i \cos v_j, w_i \sin v_j), \quad i = 1, \ldots, M_A, \quad j = 1, \ldots, M_\Phi$

**Partition:** $S = \{S_{i,j}\}$, where $S_{i,j} = \{X: ||X|| \in S_i, \angle X \in T_j\}, \quad i = 1, \ldots, M_A, \quad j = 1, \ldots, M_\Phi$

**Quantization rule:** $Q(X) = \left(Q_A(||X|| \cos Q_\Phi(\angle X)), Q_A(||X|| \sin Q_\Phi(\angle X)) \right)$
Distortion analysis via Bennett's integral

Assumptions: $M_A$, $M_{\Phi}$ are large, magnitude quantizer has point density $\lambda_A(a)$, phase quantizer is uniform with step size $\Delta = 2\pi M_{\Phi}$

Consequence: cells of polar quant cells are, approximately, small rectangles:

$$W \approx 2\pi w_i M_{\Phi}$$

$$H \approx \frac{1}{M_A}$$

Point density of the VQ:

$$\lambda(\Delta) = \frac{1}{MHW} = \frac{\lambda_A(|x||)}{2\pi |x||}$$

Inertial profile of the VQ:

$$m(x) = \frac{1}{12} \left( \frac{1}{H^2} + \frac{1}{W^2} \right)$$

Substituting $\lambda(\Delta)$ and $m(x)$ into Bennett's integral and simplifying gives:

$$D = \int \frac{1}{M} \frac{m(x)}{\lambda(\Delta)} f_X(x) \, dx$$

$$= M \int \frac{1}{\lambda_A(|x||)} f_A(a) \, da$$

$$= \frac{1}{M_A} \int \frac{M_A^2}{M_{\Phi}} f_A(a) \, da + \frac{1}{M_{\Phi}} \int \frac{M_{\Phi}^2}{M_A} f_A(a) \, da$$

$$= \frac{1}{M_A} \int \frac{M_A^2}{M_{\Phi}} f_A(a) \, da + \frac{1}{M_{\Phi}} \int \frac{M_{\Phi}^2}{M_A} f_A(a) \, da$$

$$= \frac{1}{M_A} \int \frac{M_A^2}{M_{\Phi}} f_A(a) \, da + \frac{1}{M_{\Phi}} \int \frac{M_{\Phi}^2}{M_A} f_A(a) \, da$$

Notice that $D$ does not depend on the probability distribution of the angle. This is because the angle quantizer is uniform. So it does not favor some angles over others.
Optimizing Polar Quantization

Given \( M \) (large), choose \( L \) and \( \lambda_A \) (the key characteristics) to minimize distortion

\[
D \equiv \frac{1}{M} \int_0^\infty \frac{1}{\lambda_A(a)^2} f_A(a) \, da + \frac{1}{M} \frac{\pi^2}{6} L^2 \int_0^\infty a^2 f_A(a) \, da
\]

Approach 1: For given choice of \( \lambda_A \), find best \( L \) by equating to zero the derivative of \( D \) with respect to \( L \). Then find best \( \lambda_A \) by calculus of variations.

Approach 2: For given choice of \( L \), find best \( \lambda_A \) by calculus of variations or Holder’s inequality. Then find best \( L \) by equating to zero the derivative wrt \( L \) of the resulting expression for distortion.

The result:

\[
\lambda_A(a) = c \, p_A(a)^{1/3} \quad (c \text{ chosen to make } \lambda_A(a) \text{ integrate to one})
\]

\[
L^2 = \frac{M_A}{M_\Phi} = \frac{1}{2\pi} \left( EA^2 \int_0^\infty f_A^{1/3}(a) \, da \right)^{1/2}
\]

\[
D \equiv \frac{1}{12} 2\pi \sqrt{E A^2 \left( \int_0^\infty f_A^{1/3}(a) \, da \right)^{3/2} \frac{1}{M}}
\]

We'll review calculus of variations shortly.

Prime Example: \( X_1, X_2 \) IID Gaussian

\[
f_X(x) = \frac{1}{2\pi} e^{-||x||^2/2} \quad \text{and} \quad f_A(a) = a \, e^{-a^2/2}, \quad a \geq 0 \quad (\text{Rayleigh density})
\]

\[
EA = \sqrt{\pi/2}, \quad EA^2 = E(X_1^2 + X_2^2) = 2 \, E X^2 = 2, \quad \sigma_A^2 = 2 - \frac{\pi}{2} = .429
\]

For optimized polar quantization

\[
\lambda_A(a) = c \, p_A(a)^{1/3} = \frac{a}{\sqrt{3}} \, e^{-a^2/6}, \quad a \geq 0
\]

\[
L^2 = \frac{M_A}{M_\Phi} = \frac{1}{2\pi} \left( EA^2 \int_0^\infty f_A^{1/3}(a) \, da \right)^{1/2} = .376
\]

\[
\frac{M_A}{M_\Phi} = .613 \quad (\text{more phase levels than amplitude levels})
\]

\[
D \equiv \frac{1}{12} \sigma_X^2 29.7 \frac{1}{M}
\]

Later we'll show that for optimal scalar quantization applies directly to \( X_1, X_2 \)

\[
D \equiv \frac{1}{12} \sigma_X^2 32.6 \frac{1}{M}
\]

Gain of polar quantization over conventional optimal scalar quantization

\[
10 \log_{10} \frac{32.6}{29.7} = .41 \text{ dB}
\]

Gain of optimal two-dimensional VQ over scalar quantization = 1.30 dB
Comments on Polar Quantization

- Polar quantization would seem to be "especially well suited" to quantizing $X$ when $f_X(x)$ is circularly symmetric; i.e. when $f_X(x)$ depends only on $||x||$. equivalently,

  $A$ and $\Phi$ are indep., and $\Phi$ is uniformly distributed between 0 and $2\pi$.

  equivalently,

  $f_X(x) = \frac{1}{2\pi} \frac{1}{||x||} f_A(||x||)$, where $f_A(a)$ is pdf of amplitude $a$

  because

  $f_A(||x||) \Delta \equiv \Pr(||x|| \leq A \leq ||x|| + \Delta) = \Pr(||x|| \leq ||X|| \leq ||x|| + \Delta) \equiv 2\pi \Delta f_X(x)$

- On the other hand, we found that distortion does not depend on the angle distribution. The explanation: The opta function for a noncircularly symmetric density is less than the opta function for a circularly symmetric density with the same magnitude density. Therefore, it's not that polar quantization attains less distortion for a density that is circularly symmetric than for a density that is not. Rather, for a circularly symmetric density, polar quantization is closer to being an optimal 2-dimensional VQ than for a noncircularly symmetric pdf.

Calculus of Variations

Fix $M$ and $L$. Let $J(\lambda_A)$ be the functional defined by the formula for the distortion of a polar quantizer. Calculus of variations find an equation that $\lambda_A$ must solve.

If $\lambda_A$ is the optimal point density, i.e. the one that makes $J(\lambda_A)$ smallest among all nonnegative functions $\lambda$ that integrate to one, then for any function $g$ such that $\int_0^\infty g(a) da = 0$ and any $\varepsilon > 0$, $\lambda_A(a) + \varepsilon g(a)$ cannot be better than $J(\lambda_A)$, i.e.

$J(\lambda_A + \varepsilon g)$ has a local minimum with respect to $\varepsilon$ when $\varepsilon = 0$

Therefore,

$\frac{d}{d\varepsilon} J(\lambda_A + \varepsilon g)|_{\varepsilon=0} = 0$

and this must be true for every function $g$ such that $\int_0^\infty g(a) da = 0$.

Therefore, we take the derivative of $J(\lambda_A + \varepsilon g)$ with respect to $\varepsilon$. We set $\varepsilon = 0$ and we equate the derivative at $\varepsilon=0$ to zero. This gives an equation that $\lambda_A$ must satisfy for every function $g$ such that $\int_0^\infty g(a) da = 0$. Solving this equation gives the answer. Typically, the equation is something like

$\int g(a)(\lambda_A(a) - h(a)) da = 0$ for all functions $g$ such that $\int g(a) da = 0$

which implies that $\lambda_A(a) = h(a) + c$, where $c$ is some constant.

---

5To be perfectly rigorous, we need to restrict attention to functions $g$ such that $g(a) = 0$ whenever $\lambda_A(a) = 0$.

6One should check that $g(a) = 0$ whenever $\lambda_A(a) = 0$. 

Bennett-36
Generalizing Bennett's Integral to Quantizers Whose Neighboring Cells Do Not Have Similar Sizes or Shapes

Consider the two-dimensional quantizer below in which cells come in identical groups of three. Notice that, although neighboring one another, the cells within each group do not all have the same, or even similar, sizes and shapes. This invalidates two of the basic approximations made in deriving Bennett's integral.

To see the problem, recall the three approximations made in deriving Bennett's integral:

1. \( D = \frac{1}{k} \sum_{i=1}^{M} \int_{S_i} ||x-w_i||^2 f_X(x) \, dx \)
   
   This approximation is OK; it requires only that the cells be small

2. \( D = \sum_{i=1}^{M} f_X(w_i) \frac{1}{k} M(S_i,w_i) \)
   
   This approximation requires \( |S_i| \approx \frac{1}{M \lambda(w_i)} \) and \( m(S_i,w_i) \approx m(w_i) \) (***)

3. \( D = \frac{1}{M^{2/3}} \int \frac{m(x)}{\lambda^{2/3}(x)} f_X(x) \, dx \)
   
   This approximation is OK; it requires only that \( \{S_1, \ldots, S_M\} \) be a partition with small cells and \( w_i \in S_i \)

The problem is that the approximations (*** ) are not valid because \( \lambda(x) \) and \( m(x) \) are, generally, slowly varying smooth functions, but cell size and shape change abruptly from cell to cell.

Bennett-39
To rectify this problem we need to consider groups of three cells as one unit. Previously, we approximated distortion one cell at a time:

$$\frac{1}{k} \int_{S_i} \| x - w_i \|^2 f_x(x) \, dx \equiv f_x(w_i) \frac{1}{k} M(S_i, w_i) = f_x(w_i) m(S_i, w_i) |S_i|^{1+2/k} \equiv f_x(w_i) \frac{m(w_i)}{M^{2/k} \lambda^{2/k}(w_i)} |S_i|$$

Now, instead, we compute the distortion of the three cells as a group.

Let $S_{n,1}, S_{n,2}, S_{n,3}$ denote the cells in the $n$th group, let $w_{n,1}, w_{n,2}, w_{n,3}$ denote the corresponding codewords, let $\bar{S}_n = S_{n,1} \cup S_{n,2} \cup S_{n,3}$, let $\bar{w}_n$ be some point in $\bar{S}_n$, and let $\bar{m}(S_{n,1}, w_{n,1}, S_{n,2}, w_{n,2}, S_{n,3}, w_{n,3})$ denote the following average normalized moment of inertia of the group of three cells:

$$\bar{m}(S_{n,1}, w_{n,1}, S_{n,2}, w_{n,2}, S_{n,3}, w_{n,3}) = \frac{1}{M} (M(S_{n,1}, w_{n,1}) + M(S_{n,2}, w_{n,2}) + M(S_{n,3}, w_{n,3}))$$

and let $\bar{m}(x)$ denote the corresponding inertial profile:

$$\bar{m}(x) \equiv \bar{m}(S_{n,1}, w_{n,1}, S_{n,2}, w_{n,2}, S_{n,3}, w_{n,3}) \text{ when } x \in \bar{S}_n$$

One can straightforwardly check that $\bar{m}$ is invariant to the scale and position of the group of cells. Thus, it depends only on the shapes and relative sizes of the cells in the group.

If in the example shown previously the squares have sides of length 1, then

$$\bar{m} = \frac{1}{2} \left( \frac{1}{3} \frac{5}{1+2+2} \right) = \frac{13}{128} = 0.102$$

which lies between the nmi of a square .083, and that of a $1 \times 2$ rectangle .104.

Bennett-40

Then the distortion due to the group of three cells is

$$\frac{1}{k} \sum_{i=1}^{3} \int_{S_{n,i}} \| x - w_i \|^2 f_x(x) \, dx \equiv f_x(\bar{w}_n) \sum_{i=1}^{3} \frac{1}{k} M(S_{n,i}, w_{n,i}) \text{ since } \bar{S}_n \text{ is also small}$$

$$= f_x(\bar{w}_n) 3\bar{m}(S_{n,1}, w_{n,1}, S_{n,2}, w_{n,2}, S_{n,3}, w_{n,3}) \left(\frac{1}{3}\right)^{1+2/k} |\bar{S}_n|^{1+2/k}$$

$$= f_x(\bar{w}_n) \bar{m}(S_{n,1}, w_{n,1}, S_{n,2}, w_{n,2}, S_{n,3}, w_{n,3}) \left(\frac{1}{3}|\bar{S}_n|\right)^{2/k} |\bar{S}_n|$$

$$\equiv f_x(\bar{w}_n) \frac{\bar{m}(\bar{w}_n)}{(M/3)^{2/k} \lambda^{2/k}(\bar{w}_n)} |\bar{S}_n|$$

where we have used the approximation $\frac{1}{3} |\bar{S}_n| \equiv \frac{1}{M \lambda(\bar{w}_n)}$.

Now, summing the approximate distortions of all groups gives

$$D = \sum_{n=1}^{M/3} \frac{1}{k} \sum_{i=1}^{3} \int_{S_{n,i}} \| x - w_i \|^2 f_x(x) \, dx \equiv \sum_{n=1}^{M/3} f_x(\bar{w}_n) \frac{\bar{m}(\bar{w}_n)}{M^{2/k} \lambda^{2/k}(\bar{w}_n)} |\bar{S}_n|$$

$$\equiv \frac{1}{M^{2/k}} \int \frac{\bar{m}(x)}{\lambda^{2/k}(x)} f_x(x) \, dx$$

This is just like the original Bennett's integral, but with a definition of inertial profile that averages over cells in a group.

Bennett-41
More generally if the partition comes in groups of $M_0$ cells, then the Bennett integral approximation is:

$$D \cong \frac{1}{M^{2/k}} \int \frac{\tilde{m}(x)}{\lambda^{2/k}(x)} f_X(x) \, dx$$

where

$$\tilde{m}(x) \equiv \tilde{m}(S_{n,1}, w_{n,1}, ..., S_{n,M_0}, w_{n,M_0}) \quad \text{when} \quad x \in S_{n,1} \cup ... \cup S_{n,M_0}$$

$$\tilde{m}(S_{n,1}, w_{n,1}, ..., S_{n,M_0}, w_{n,M_0}) = \frac{1}{k} \frac{M_0}{M_0} \sum_{i=1}^{M_0} M(S_n, w_n, i)$$

$$(\frac{1}{M_0(|S_{n,1}| + ... + |S_{n,M_0}|)^{1+2/k}}$$

**Distortion Density and Local Distortion**

Let $G$ be some subset of $\mathbb{R}^k$. If one repeats the derivation of Bennett's integral, but for the distortion computed only $X$ in the set $G$, then one obtains

$$\int_G \lVert x - Q(x) \rVert^2 f_X(x) \, dx \equiv \int_G \frac{m(x)}{M^{2/k} \lambda^{2/k}(x)} f_X(x) \, dx = \int_G \rho(x) \, dx$$

where

$$\rho(x) \equiv \frac{m(x)}{M^{2/k} \lambda^{2/k}(x)} f_X(x)$$

Since $\rho$ is a function that when integrated over a region $G$ yields the distortion in region $G$, it can be called the **distortion density** of the quantizer.

One may also rewrite the distortion in $G$ as

$$\int_G \lVert x - Q(x) \rVert^2 f_X(x) \, dx \equiv \int_G d(x) \, f(x) \, dx$$

where

$$d(x) \equiv \frac{m(x)}{M^{2/k} \lambda^{2/k}(x)}$$

is considered the **local distortion** of the quantizer at $x$. 
Notice that the distortion in region $G$ is not the same as the conditional distortion given region $G$, which is

$$E \left[ ||X-Q(X)||^2 \mid X \in G \right] = \int_G ||x-Q(x)||^2 \frac{f_X(x)}{Pr(X \in G)} \, dx$$

$$\approx \int_G \frac{\rho(x)}{Pr(X \in G)} \, dx$$

$$= \int_G d(x) \frac{f_X(x)}{Pr(X \in G)} \, dx$$

This shows that conditional distortion is the average of local distortion.

When $G$ is small and $x \in G$, approximating $d$ and $f$ as being constant on $G$ yields

$$E \left[ ||X-Q(X)||^2 \mid X \in G \right] \approx \frac{\rho(x)}{Pr(X \in G)} |G| \approx \frac{\rho(x)}{f_X(x)} = d(x).$$

Thus for a small region, the conditional distortion approximately equals the local distortion in that region.