This can be written as

$$\frac{\sigma^2}{D} = 2^{+2R}$$

where $\frac{\sigma^2}{D}$ is the Signal-To-Quantization-Noise Rati

Changin into decibel, we get

$$SQNR = 10 \log \frac{\sigma^2}{D} = 10 \log 2^{+2R} = +20R \log 2$$

$$SQNR = +6R \quad dB$$

This means that for each 1 bit increase in resolution, we gain 6 dB (4 times improvement in SQNR.

X 11th lecture, Nov. 11, 2003

Simultaneous Quantization (Description) of Independent Gaussian Random variables:

Take normal sources $X_1, X_2, \ldots, X_m$ where $X_i \sim N(0, \sigma_i^2)$. We would like to find

$$R(D) = \min_{f(\hat{x}^m | x^m) : Ed(x^m, \hat{x}^m) \leq D} I(X^m; \hat{X}^m)$$

We have:

$$I(X^m; \hat{X}^m) = h(X^m) - h(X^m | \hat{X}_m)$$

$$= \sum_{i=1}^{m} h(X_i) - \sum_{i=1}^{m} h(X_i | X^{i-1}, \hat{X}^m)$$

$$(a) \quad \geq \sum_{i=1}^{m} h(X_i) - \sum_{i=1}^{m} h(X_i | \hat{X}_i)$$

$$= \sum_{i=1}^{m} I(X_i; \hat{X}_i)$$

$$\text{(b)} \quad \geq \sum_{i=1}^{m} R(D_i) = \sum_{i=1}^{m} \left(\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right)^+$$

We can achieve equality in (a) by letting

$$f(x^m|\hat{x}^m) = \prod_{i=1}^{m} f(x_i|\hat{x}_i)$$

and we get equality in (b) by letting

$$\hat{X}_i \sim N(0, \sigma_i^2 - D_i).$$

So, the problem is to minimize

$$\sum_{i=1}^{m} \max\left\{\frac{1}{2}\log\frac{\sigma_i^2}{D_i}, 0\right\}$$

w.r.t. the constraint $\sum_{i=1} D_i = D$

let

$$J(D) = \sum_{i=1}^{m} \frac{1}{2}\log\frac{\sigma_i^2}{D_i} + \mu\sum_{i=1}^{m} D_i$$

$$\frac{\partial J(D)}{\partial D_i} = -\frac{1}{2}\frac{1}{D_i} \times \frac{1}{\ln 2} + \mu = 0$$

or $D_i = \theta$ where $\theta = (2\mu \ln 2)^{-1}$

we ~~use~~ select θ to satisfy the Kuhn-Tucker Condition: ~~to get~~

$$\frac{\partial J}{\partial D_i} = \begin{cases} 0 & \text{if } D_i < \sigma_i^2 \\ \leq 0 & \text{if } D_i \geq \sigma_i^2 \end{cases}$$

$D_i = \theta$ means that we allocate equal distortion to each source, i.e., we assign the number of bits such that the distortion is the same. But, this is possible if θ is less than all σ_i^2's (Since $\frac{1}{2}\log\frac{\sigma_i^2}{D_i} = \frac{1}{2}\log\frac{\sigma_i^2}{\theta}$ will be $-\infty$ if $\sigma_i^2 < \theta$ for some i). When θ is greater than some of the σ_i^2's we set the distortion for those i's equal to σ_i^2, i.e., we don't encode these and let the decoder use its best guess.

Therefore,

$$R(D) = \sum_{i=1}^{m} \frac{1}{2}\log\frac{\sigma_i^2}{D_i} = \sum_{i=1}^{m}\left(\frac{1}{2}\log\frac{\sigma_i^2}{\theta}\right)^+$$

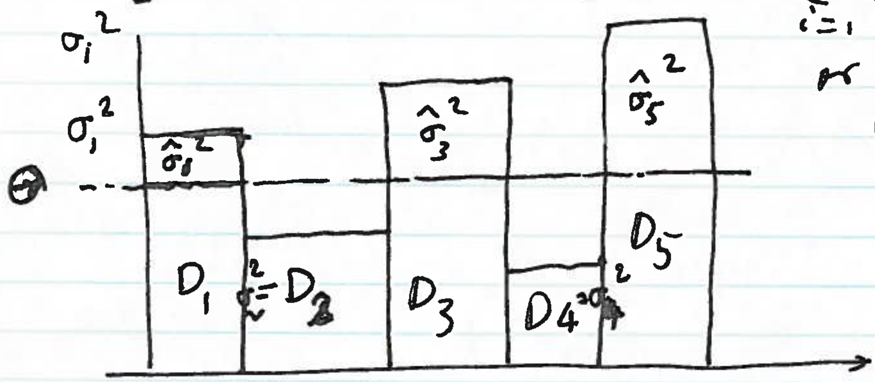
where

$$D_i = \begin{cases} \theta & \text{if } \theta < \sigma_i^2 \\ \sigma_i^2 & \text{if } \theta \geq \sigma_i^2 \end{cases} = \min(\sigma_i^2, \theta)$$

where θ is selected so that $\sum_{i=1}^{m} D_i = D$.

or $D = \sum_{i=1}^{n} \min(\sigma_i^2, \theta)$

Note: for per symbol $R(D)$ expressions for R & D need to be divided by m.



Example of JPEG and MPEG.

Correlated Gaussian Source

~~Assume that $(X_1, \cdots, X_n) = \underline{X}^n \sim N(0, K_n)$~~

~~Since~~ K_n can be written as

$$K_n = U \Lambda_n U^{-1}$$

where $\Lambda_n = \text{diag}[\lambda_1, \cdots, \lambda_n]$

$$\underline{Y}^n = U^{-1} \underline{x}_n$$

also,

$$\hat{\underline{Y}}^n = U^{-1} \hat{\underline{x}}_n$$

Since U^{-1} is invertible $\underline{I(\underline{X}^n; \hat{X}^n)} = \underline{I(Y^n; \hat{Y}^n)}$

It is also easy to show that $d(Y^n, \hat{Y}_n) = d(X; \hat{X}^n)$

So, the rate distortion function for X is the same as the rate distortion for Y.

$$R(D) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} \log \frac{\lambda_i}{\theta}\right)^+ = \frac{1}{n} \sum \max\left(\frac{1}{2} \log \frac{\lambda_i}{\theta}, 0\right)$$

$$D = \frac{1}{n} \sum_{i=1}^{n} \min(\lambda_i, \theta)$$

Toeplitz distribution Theorem

Let K_∞ be an infinite Toeplitz matrix with entry K_m on the mth diagonal. The eigenvalues of K_∞ are contained in the interval $\delta \leq \lambda \leq \Delta$

where δ and Δ denote the essential infimum and supremum, respectively of the function

$$S(\omega) = \sum_{m=-\infty}^{\infty} K_m e^{-jm\omega}$$

Moreover, if both δ and Δ are finite and $G(\lambda)$ is any continuous function of $\lambda \in [\delta, \Delta]$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} G(\lambda_m^{(n)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G[S(\omega)] d\omega$$

where $\lambda_m^{(n)}$ are the eigenvalues of the n-th order matrix K_n centred about the main diagonal of K_∞.

(See Grenander & Szegö: Toeplitz forms and their applications. See also T. Berger: Rate-Distortion Theory).

Applying this theorem to the Equations in the previous page, we find the rate-distortion of a discrete-time stationary source as:

$$D_\theta = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min[\theta, S(\omega)] \, d\omega$$

$$R(D_\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \max\left[0, \log \frac{S(\omega)}{\theta}\right] d\omega$$

where θ traverses the interval $0 \leq \theta \leq \Delta$.

<u>Example</u>: $R(D)$ for a first-order Gauss–Markov source.

$$X_n = \rho X_{n-1} + Z_n$$

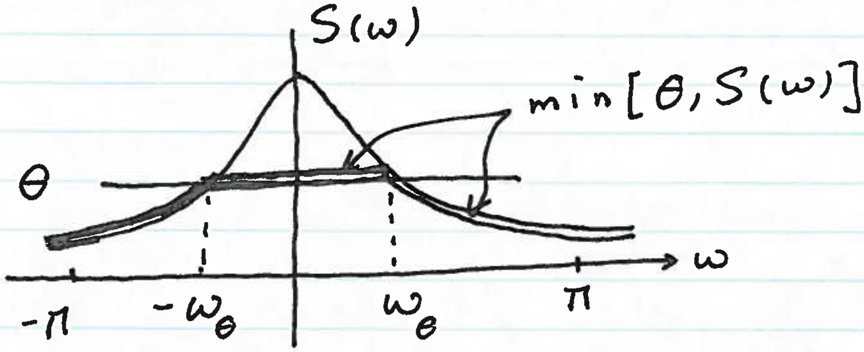
where Z_n is zero-mean Gaussian

$$K_m = \sigma^2 |\rho|^m$$

So:

$$S(\omega) = \sigma^2 \sum_{m=-\infty}^{\infty} |\rho|^m e^{-jm\omega} = \frac{\sigma^2(1-\rho^2)}{1 - 2\rho \cos\omega + \rho^2}$$

$$\int_0^{\pi} \ln\left[\frac{(1 - 2a \cos\omega + a^2)}{(1 - 2b\cos\omega + b^2)}\right] d\omega = 2\pi \ln \frac{(1-ab)}{1-b^2}$$



For small values of D, when;

$$\theta \leq S(\pi) = \frac{\sigma^2(1-\rho)}{1+\rho}$$

we have,

$$D = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta \, d\omega = \theta$$

and

$$R(D) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \log \frac{\sigma^2(1-\rho^2)}{\theta(1+\rho^2 - 2\rho\cos\omega)} d\omega = \frac{1}{2} \log \frac{(1-\rho^2)}{D} \sigma^2$$

$$\int_0^{\pi} (1 + \rho^2 - 2\rho\cos\omega)\, d\omega = 2\pi \ln 1 = 0$$

So, for a Gauss-Markov source

$$R(D) = \frac{1}{2} \log \frac{(1-\rho^2)\sigma^2}{D} \qquad D \leq \frac{1-\rho}{1+\rho}\sigma^2$$

for larger values of D, we need to vary θ from zero θ to maximum of $S(\omega)$ and derive the $R(D_\theta)$ parametrically.

<u>SQNR</u>:

$$\frac{\sigma^2}{D} = \frac{2^{2R}}{(1-\rho^2)}$$

$$SQNR = 10 \log \frac{\sigma^2}{D} = 20 R \log 2 - 10 \log(1-\rho^2)$$
$$= 6R - 10 \log(1-\rho^2)$$

<u>Example</u>: $\rho = 0.9$ and $R = 1$ bit/sample

$SQNR = 6 + 7.2 = 13.2\,dB$

results using Vector Quantization and Finite State (Feedback) VQ.

k	VQ (SQNR)	FSVQ
1	4.4	10
2	7.9	10.8
3	9.2	11.4
4	10.2	12.1
5	10.6	
6	10.9	
7	11.2	

Characterization of the Rate-Distortion Function

The problem is to minimize

$$I(X;\hat{X}) \quad \text{subject to} \quad \sum_{x}\sum_{\hat{x}} p(x)q(\hat{x}|x)d(x,\hat{x}) \le D$$

by proper choice of $\{q(\hat{x}|x)\}$. We use the method of Lagrange Multipliers.

$$\frac{q(\hat{x}|x)}{q(\hat{x})}$$

$$J(q) = \sum_{x}\sum_{\hat{x}} p(x)\, q(\hat{x}|x) \log \frac{q(\hat{x}|x)}{\sum_{x} p(x)q(\hat{x}|x)}$$

$$+ \lambda \sum_{x}\sum_{\hat{x}} p(x)\, q(\hat{x}|x)\, d(x,\hat{x})$$

$$+ \sum_{x} \nu(x) \sum_{\hat{x}} q(\hat{x}|x).$$

$$\frac{\partial J}{\partial q(\hat{x}|x)} = p(x) \log \frac{q(\hat{x}|x)}{q(\hat{x})} + p(x) - \sum_{x'} p(x')\, q(\hat{x}|x') \frac{1}{q(\hat{x})} p(x$$

$$+ \lambda p(x)\, d(x,\hat{x}) + \nu(x) = 0.$$

Let $\nu(x) = p(x) \log \mu(x)$ to get

$$p(x)\left[\log \frac{q(\hat{x}|x)}{q(\hat{x})} + \lambda d(x,\hat{x}) + \log \mu(x)\right] = 0$$

or

$$q(\hat{x}|x) = \frac{q(\hat{x})\, e^{-\lambda d(x,\hat{x})}}{\mu(x)} \qquad (A)$$

Summing over $\hat{x}$, we get

$$\sum_{\hat{x}} q(\hat{x}|x) = 1 = \frac{\sum_{\hat{x}} q(\hat{x}) e^{-\lambda d(x,\hat{x})}}{\mu(x)}$$

or

$$\boxed{\mu(x) = \sum_{\hat{x}} q(\hat{x}) e^{-\lambda d(x,\hat{x})}}$$

or,

$$q(\hat{x}|x) = \frac{q(\hat{x}) e^{-\lambda(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x}) e^{-\lambda d(x,\hat{x})}}$$

Multiplying (A) by $p(x)$ and summing over x, we get

$$q(\hat{x}) = \sum_{x} p(x) q(\hat{x}|x) = q(\hat{x}) \sum_{x} \frac{p(x) e^{-\lambda d(x,\hat{x})}}{\mu(x)}$$

if $q(\hat{x}) \neq 0$, we get

$$\boxed{\sum_{x} \frac{p(x)}{\mu(x)} e^{-\lambda d(x,\hat{x})} = 1} \rightarrow \boxed{\sum_{x} r(x) e^{-\lambda d(x,\hat{x})} = 1 \quad \text{for all } \hat{x}}$$

in general, when som $q(\hat{x}) = 0$, we get

$$\sum_{x} \frac{p(x)}{\mu(x)} e^{-\lambda d(x,\hat{x})} \begin{cases} = 1 & q(\hat{x}) > 0 \\ \leq 1 & q(\hat{x}) = 0 \end{cases}$$

Parametric form of $R(D)$:

$$R(D) = \sum_{x}\sum_{\hat{x}} P(x)\,q(\hat{x}|x)\log\frac{q(\hat{x}|x)}{q(\hat{x})}$$

$$= \sum_{x}\sum_{\hat{x}} p(x)\,q(\hat{x}|x)\log\frac{e^{-\lambda d(x,\hat{x})}}{\mu(x)}$$

$$= \sum_{x}\sum_{\hat{x}} P(x)\,q(\hat{x}|x)\log\frac{r(x)}{p(x)}\,e^{-\lambda d(x,\hat{x})}$$

$$= +\sum_{x}\sum_{\hat{x}} p(x)\,q(\hat{x}|x)\log r(x)$$

$$- \sum_{x}\sum_{\hat{x}} p(x)\,q(\hat{x}|x)\log p(x)$$

$$- \lambda\sum_{\hat{x}}\sum_{\hat{x}} p(x)\,q(\hat{x}|x)\,d(x,\hat{x})$$

$$\boxed{= \sum_{x}\sum_{\hat{x}} p(x)\,q(\hat{x}|x)\log r(x) + H(X) - \lambda D} = \sum_{x} p(x)\log r(x) + H(X) - \lambda D$$

and

$$D = \sum_{x}\sum_{\hat{x}} p(x)\,q(\hat{x}|x)\,d(x,\hat{x})$$

$$= \sum_{x}\sum_{\hat{x}} \cancel{p(x)}\frac{q(\hat{x})\,r(x)}{\cancel{p(x)}}\,e^{-\lambda d(x,\hat{x})}\,d(x,\hat{x})$$

$$\boxed{D = \sum_{x}\sum_{\hat{x}} q(\hat{x})\,r(x)\,e^{-\rho d(x,\hat{x})}\,\rho(x,\hat{x})}$$

$$\boxed{R = \sum_{x} p(x)\log r(x) + H(X) - \lambda D}$$

let $\frac{P(x)}{\mu(x)} = r(x)$

then, when all $q(\hat{x}) \neq 0$, we get

$$\sum_x r(x) e^{-\lambda d(x,\hat{x})} = 1$$

define a matrix A with elements $e^{-\lambda d(x,\hat{x})}$

(1) $A^T \cdot \underline{r} = \mathbb{1}$ where $\mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

and

(2) $\underline{A} \cdot \underline{Q} = \underline{M}$ $\quad M = \begin{bmatrix} \mu(1) \\ \vdots \\ \mu(n) \end{bmatrix}$

$$Q = \begin{bmatrix} q(1) \\ \vdots \\ q(n) \end{bmatrix}$$

From (1), we can find $\{r(x)\}$ and from $r(x) = \frac{P(x)}{\mu(x)}$ find $\mu(x)$ and then from (2) find $q(1), \ldots, q(n)$.

Example: Binary memoryless source with $P(0) = p$, $P(1) = 1-p$ and $d(x,\hat{x}) = \delta_{x,\hat{x}}$.

$$\underline{d} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$A = \{e^{-\lambda d(x,\hat{x})}\} = \begin{bmatrix} 1 & e^{-\lambda} \\ e^{-\lambda} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

$$A^T \cdot \underline{r} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} r(0) \\ r(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \frac{1}{1-\alpha^2} \begin{bmatrix} 1 & -\alpha \\ -\alpha & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{1+\alpha} \\ \frac{1}{1+\alpha} \end{bmatrix}$$

$$r(0) = \frac{1}{1+\alpha}, \quad r(1) = \frac{1}{1+\alpha}$$

$$r(0) = \frac{p(0)}{\mu(0)} \Rightarrow \mu(0) = p(1+\alpha)$$

$$r(1) = \frac{p(1)}{\mu(1)} \Rightarrow \mu(1) = (1-p)(1+\alpha)$$

$$\begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \begin{bmatrix} q(0) \\ q(1) \end{bmatrix} = \begin{bmatrix} p(1+\alpha) \\ (1-p)(1+\alpha) \end{bmatrix}$$

$$q(0) = \frac{p - \alpha(1-p)}{1-\alpha} \qquad q(1) = \frac{-p(\alpha) + (1-p)}{1-\alpha}$$

Now,

$$D = r(0)q(0)\cancel{\times 1}\times 0 + r(0)\,q(1)\,\alpha + r(1)q(0)\times\alpha$$
$$+ r(1)q(1)\cancel{\times 1}\times 0 = \frac{1}{1+\alpha} \times \frac{-p\alpha + (1-p)}{1-\alpha} + \frac{1}{1+\alpha} \times \frac{p - \alpha(1-}{1-\alpha}$$

$$D = \frac{\alpha}{1+\alpha} \Rightarrow \alpha = \frac{D}{1-D}$$

$$R(D) = \sum_x p(x) \log r(x) + H(X) - \lambda D$$

$$= p(0) \log r(0) + p(1) \log r(1) + H_b(p) - \lambda D$$

$$= -p \log(1+\alpha) - (1-p) \log(1+\alpha) + H_b(p) - \lambda D$$

$$= -\log(1+\alpha) + H_b(p) - \lambda D$$

$$= -\log \frac{1}{1-D} + H_b(p) - \lambda D$$

$$= \log(1-D) + H_b(p) + \log \alpha \, D$$

$$= \log(1-D) + H_b(p) + D \log \frac{D}{1-D}$$

$$= \log(1-D) + H_b(p) + D \log D - D \log(1-D)$$

$$\boxed{= H_b(p) - H_b(D)}$$

Exercise: Consider a source with four letters with probability vector $\underline{p} = \frac{1}{2}\begin{bmatrix} p \\ 1-p \\ 1-p \\ p \end{bmatrix}$

and the distortion measure

$$\underline{d} = \begin{bmatrix} 0 & 1/2 & 1/2 & 1 \\ 1/2 & 0 & 1 & 1/2 \\ 1/2 & 1 & 0 & 1/2 \\ 1 & 1/2 & 1/2 & 0 \end{bmatrix}$$

Show that:

$$R(D) = \log 2 - H_b(p) - 2 H_b(D)$$

$$0 \leq D \leq D_1 = \frac{1}{2}\left(1 - \sqrt{1-2p}\right)$$

and

$$R(D) = \eta(1-p) - \frac{1}{2}[\eta(2D-p) + \eta(2(1-D)-p)]$$

for

$$D_1 \leq D \leq D_{max} = \frac{1}{2}$$

and

$$\eta(x) \triangleq -x \log x .$$

The above formulation provides us with a method for finding a lower bound on $R(D)$

Assume that $\{r(x)\}$, $x \in \mathcal{X}$ satisfy

$$\sum_x r(x) e^{-\lambda d(x,\hat{x})} \leq 1 \qquad \text{all } \hat{x}$$

Then

$$R(D) \geq H(X) + \sum_x p(x) \log[r(x)] - \lambda D = R_L(D)$$