Step 1(a): Given k, T and $R_1,\ldots,R_k$ such that $\frac{1}{k}\Sigma_{i=1}^{k} R_i = R$, choose $Q_1,\ldots,Q_k$ to have rates $R_1,\ldots,R_k$, respectively, and to minimize distortion.

Since

$$D = \frac{1}{k}\sum_{i=1}^{k} D_{U_i,Q_i},$$

for each i we choose Q_i to minimize D_{U_i,Q_i} subject to $R(Q_i) \le R_i$. Then

$$D_{U_i,Q_i} = \delta_{sq,U_i}(R_i) \triangleq \text{OPTA for scalar quantization of } U_i.$$

The OPTA for U_i is either the fixed-rate or the variable-rate OPTA, depending on whether we want to optimize transform coding for fixed or variable rate coding.

It follows that the least distortion attainable for the given k, T, and $R_1,\ldots,R_k$ is

$$D = \frac{1}{k}\sum_{i=1}^{k} \delta_{sq,U_i}(R_i)$$

Step 1(b): Bit/rate allocation

We now seek to find the $R_1,\ldots,R_k$ that minimizes

$$\frac{1}{k}\sum_{i=1}^{k} \delta_{sq,U_i}(R_i)$$

subject to the constraint

$$\frac{1}{k}\sum_{i=1}^{k} R_i \le R.$$

This is called a bit or rate allocation, i.e. an allocation of kR bits among the scalar quantizers for the individual coefficients. If some are given rate larger than R, then others must be given rate less than R.

Recall the original idea of transform coding, namely, that some coefficients will be smaller on the average than others, and that these will be encoded at lower rates than those that are larger on the average. If there is to be a net gain, then it will depend on a successful bit allocation.

For fixed-rate coding, the values of $\delta_{sq,U_i}(R_i)$ can be tabulated for the various values of R_i and a numerical optimization algorithm can be applied to find the best choice of the R_i's. See Gersho and Gray, pp. 226-235. Similar methods may be used for variable-rate coding.

We now consider the high-resolution case. In this case, one may solve analytically for the best bit allocation.

HIGH-RESOLUTION ANALYSIS

To simplify discussion assume X comes from zero-mean, stationary source.

Assume R is so large that the optimal R_i's are so large that we may use the approximations

$$\delta_{sq,U_i}(R_i) \cong \frac{1}{12}\,\sigma_i^2 \alpha_i\, 2^{-2R_i},$$

where σ_i^2 = variance of U_i , and
for fixed-rate coding

$$\alpha_i = \beta_{U_i} = \frac{1}{\sigma_i^2}\left(\int f_{U_i}^{1/3}(u)\,du\right)^3$$

and for variable-rate coding

$$\alpha_i = \frac{1}{\sigma_i^2}\,2^{2h(U_i)} \quad , \quad \text{where } h(U_i) = -\int_{-\infty}^{\infty} f_{U_i}(u)\,\log_2 f_{U_i}(u)\,du$$

We now have

$$D \cong \frac{1}{k}\sum_{i=1}^{k} \frac{1}{12}\,\sigma_i^2 \alpha_i\, 2^{-2R_i}$$

And the least dist'n mfor dimen. k, transform T, and rate R or less is approx'y

$$\min_{R_1\ge 0,\ldots,R_k\ge 0:\ \frac{1}{k}\sum_{i=1}^{k} R_i \le R} \frac{1}{k}\sum_{i=1}^{k} \frac{1}{12}\,\sigma_i^2 \alpha_i\, 2^{-2R_i}$$

A USEFUL OPTIMIZATION RESULT

Lemma: Let $g_1(R), \ldots, g_k(R)$ be positive valued, strictly decreasing, continuous functions defined on $[0,\infty)$. If $R_1,\ldots,R_k$ minimize

$$\sum_{i=1}^{k} g_i(R_i) \text{ subject to } \frac{1}{k}\sum_{i=1}^{k} R_i \le R.$$

are zero, then

$$g_i'(R_i) = g_j'(R_j) \text{ for all } i,j \text{ s.t. } R_i > 0,\ R_j > 0$$

and

$$|g_i'(R_i)| \le |g_j'(R_j)| \text{ for all } i,j \text{ s.t. } R_i = 0,\ R_j > 0$$

That is, for the optimum choice of $R_1,\ldots,R_k$, the slopes of the g_i functions are the same.

Notes:

The slopes are negative.

The above is the basis of the Lagrange multiplier method.

Proof by Contradiction:

Suppose $R_1,\ldots,R_k$ minimize $\frac{1}{k}\sum_{i=1}^{k} g_i(R_i)$ with $\frac{1}{k}\sum_{i=1}^{k} R_i \le R$,

and for some i,j, $R_i > 0$, $R_j > 0$, and

$$g_i'(R_i) < g_j'(R_j) < 0$$

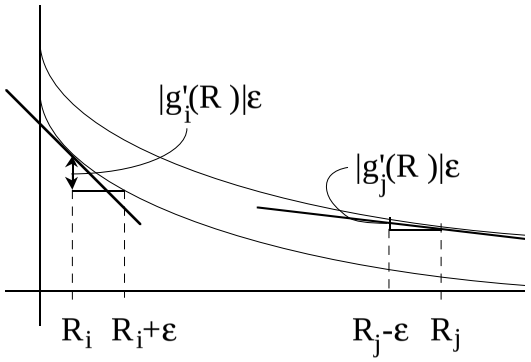
Then for some small ε, replace

R_i by $R_i+\varepsilon$ and R_j by $R_j-\varepsilon$.

This has no effect on $\frac{1}{k}\sum_{i=1}^{k} R_i$.

It decreases $g_i(R_i)$ to $g_i(R_i+\varepsilon) \cong g_i(R_i)+\varepsilon g_i'(R_i)$.

It increases $g_j(R_j)$ to approximately $g_j(R_j-\varepsilon) \cong g_j(R_j) - \varepsilon g_j'(R_j)$.



Since $g_i'(R_i) < g_j'(R_j)$, the new rate allocation decreases overall distortion:

$$g_i(R_i+\varepsilon) + g_j(R_j-\varepsilon) \cong g_i(R_i)+g_j(R_j)+\varepsilon(g_i'(R_i)-g_j'(R_j)) < g_i(R_i)+g_j(R_j).$$

(Despite the "≅", the overall inequality will be strict if ε is sufficiently small.)
This contradicts the original assumption that slopes are not the same.
Thus, they must be the same.

Similar argument, if $R_i = 0$, $R_j > 0$, and $|g_i'(R_i)| > |g_j'(R_j)|$

Optimal Rate Allocation in the High-Resolution Case

This lemma shows that if $R_1,\ldots,R_k > 0$ minimize

$$\frac{1}{k}\sum_{i=1}^{k} \frac{1}{12}\sigma_i^2\alpha_i 2^{-2R_i} \quad \text{subject to} \quad \frac{1}{k}\sum_{i=1}^{k} R_i = R,$$

then there is a constant c such that for each i,

$$c = \frac{d}{dR_i}\left(\frac{1}{12}\sigma_i^2\alpha_i 2^{-2R_i}\right) = \frac{1}{12}\sigma_i^2\alpha_i 2^{-2R_i}(-2\ln 2)$$

Solving the above yields

$$R_i = \frac{1}{2}\log_2 \sigma_i^2\alpha_i - \frac{1}{2}\log_2\left(\frac{-6c}{\ln 2}\right).$$

Since the R_i's average to R, equating the avg. of the RHS terms to R gives

$$-\frac{1}{2}\log_2\left(\frac{-6c}{\ln 2}\right) = R - \frac{1}{k}\sum_{j=1}^{k}\frac{1}{2}\log_2 \sigma_j^2\alpha_j = R - \frac{1}{2}\log_2\left(\prod_{j=1}^{k}\sigma_j^2\alpha_j\right)^{1/k}$$

Therefore the optimal rate allocation is

$$R_i = \frac{1}{2}\log_2 \sigma_i^2\alpha_i + R - \frac{1}{2}\log_2\left(\prod_{j=1}^{k}\sigma_j^2\alpha_j\right)^{1/k} = R + \frac{1}{2}\log_2 \frac{\sigma_i^2\alpha_i}{\left(\prod_{j=1}^{k}\sigma_j^2\alpha_j\right)^{1/k}}$$

Substituting the optimal allocation shown above into the expression for distortion yields

$$D \cong \frac{1}{12}\left(\prod_{j=1}^{k}\sigma_j^2\alpha_j\right)^{1/k} 2^{-2R}$$

Summary: Given a k-dimensional transform T and large R, the optimal rate allocation is

$$R_i = R + \frac{1}{2}\log_2 \frac{\sigma_i^2 \alpha_i}{\left(\prod_{j=1}^{k} \sigma_j^2 \alpha_j\right)^{1/k}} \qquad (*)$$

and the minimal distortion is

$$D \cong \frac{1}{12}\left(\prod_{j=1}^{k} \sigma_j^2 \alpha_j\right)^{1/k} 2^{-2R} \qquad (**)$$

Notes:

- Compare $(**)$ to the high-rate approx to the OPTA of scalar quantization

  $$\delta_{sq,X}(R) \cong \frac{1}{12}\sigma_X^2 \alpha_X^2 2^{-2R}$$

  The SNR gain of transform coding with a particular transform over scalar quantization is

  $$10\log_{10}\frac{\delta_{sq,X}(R)}{D_{transform}} \cong 10\log_{10}\frac{\sigma_X^2 \alpha_X}{\left(\prod_{j=1}^{k} \sigma_j^2 \alpha_j\right)^{1/k}}$$

- A good transform is one that makes $\prod_{j=1}^{k} \sigma_j^2 \alpha_j$ small.
- Coefficient correlations do not enter into formula for D.
- We won't compare to k-dimensional VQ until we optimize T.

- If $\underline{X}$ is Gaussian, then each U_i is Gaussian, and

  $\alpha_1 = \ldots = \alpha_k = \alpha_X = 32.6$ for FRC and 17.08 for VRC.

  Therfore, gain of transform coding with transform T over scalar quantization is

  $$10\log_{10}\frac{\sigma_X^2}{\left(\prod_{j=1}^{k} \sigma_j^2\right)^{1/k}} = 10\log_{\text{ratio of variance to geometric mean of coef variances}}$$

  Fact: geometric mean $\leq$ ar\ithmetic mean; equality iff "averagees" are identical

  $$\Rightarrow \quad \left(\prod_{j=1}^{k} \sigma_j^2\right)^{1/k} \leq \frac{1}{k}\sum_{j=1}^{k} \sigma_j^2 = \frac{1}{k} E||\underline{U}||^2 = \frac{1}{k} E||\underline{X}||^2 = \sigma_X^2, \quad \text{equality iff } \sigma_j^2 = \sigma_X^2 \text{ all } j$$

  Consequences: For Gaussian case,

  - Gain in dB ≥ 0.
  - A good transform does "energy compaction" i.e. it reduces

    $$\Gamma \triangleq \left(\prod_{j=1}^{k} \sigma_j^2\right)^{1/k},$$

    by making some σ_j^2's as small as is possible with orthogonal transform.
    Since transform is orthogonal, some σ_j^2's must be large.
    Summary: σ_j^2's should be as different as possible.

- If $\underline{X}$ is not Gaussian, there's no simple relation among the α_i's. Nevertheless, minimizing Γ is a good rule of thumb.

- Comments on optimal rate allocation:

$$R_i = R + \frac{1}{2}\log_2 \frac{\sigma_i^2 \alpha_i}{\left(\prod_{j=1}^{k} \sigma_j^2 \alpha_j\right)^{1/k}} \qquad (*)$$

  The rate R_i allocated to U_i is R plus one half the log of the ratio of $\sigma_i^2 \alpha_i$ to the geometric mean of such.

  This addition to R quantity is positive if $\sigma_i^2 \alpha_i$ greater than the geometric mean, and negative if less.

- Substituting the optimal R_i into the high-resolution expression for $\delta_{sq,U_i}(R_i)$ shows that when the quantizers are optimized, the dist'n of Q_i on U_i is

$$D_i \cong \frac{1}{12}\sigma_i^2 \alpha_i 2^{-2R_i} = \frac{1}{12}\left(\prod_{j=1}^{k} \sigma_j^2 \alpha_j\right)^{1/k} 2^{-2R} = \text{same for all } i\,.$$

  Thus with the optimal allocation, each coefficient is quantized with the <u>same distortion</u>. It isn't that we allow larger distortion for some coefficients, rather we need to assign more rate to the coefficients that have larger $\sigma_j^2 \alpha_j$ in order that they be quantized with the same distortion. (Compare to JPEG.)

- What to do if $(*)$ yields a negative R_i for some i? (This happens when $\sigma_i^2 \alpha_i$ is sufficiently small.) In this case, the high-rate assumption is invalid, and we should not use the high-resolution analysis. Actually, we need each $R_i \gtrsim 3$ for each i for this high-resolution analysis to be accurate.

- Suppose we were to give the same rate allocation to all coefficients, i.e. $R_i = R$, then in Gaussian case

$$D \cong \frac{1}{k}\sum_{i=1}^{k} \frac{1}{12}\sigma_i^2 \alpha_X 2^{-2R_i} = \frac{1}{12}\alpha_X \frac{1}{k}\sum_{i=1}^{k} \sigma_i^2 2^{-2R} = \frac{1}{12}\alpha_X \sigma_X^2 2^{-2R} = \delta_{sq,X}(R).$$

  That is, with equal bit allocation, the transform code gives the same performance as direct scalar quantization of the X_i's, no matter what orthogonal transform is chosen. We conclude that the proper bit allocation is crucial.

- Recall zero mean assumption. If the data does not have zero mean, then the mean could be subtracted from the data before transform coding, and added back to the reconstructions produced by the decoder. But actually, the analysis applies as is.

Step 2: The Optimal Transform

We now ask how to choose the orthogonal transform T to minimize

$$\Gamma \triangleq \left(\prod_{j=1}^{k} \sigma_j^2 \right)^{1/k}$$

This will be the optimal transform for the Gaussian case, and a "good" transform more generally.

Main Fact: The transform T that minimizes Γ is a Karhunen-Loeve Transform (KLT). A KLT is any transform whose rows are an orthonormal set of eigenvectors of the covariance matrix $K_{\underline{X}}$ of $\underline{X}$. For a KLT, the coefficient variances σ_i^2 are the eigenvalues of $K_{\underline{X}}$, denoted $\lambda_1,\ldots,\lambda_k$ and

$$\Gamma = \left(\prod_{j=1}^{k} \lambda_j \right)^{1/k} = |K_{\underline{X}}|^{1/k}$$

where $|K_{\underline{X}}|$ denotes the determinant of $K_{\underline{X}}$.

The proof is based on:

More Facts from Linear Algebra

Definition: If K is a $k \times k$ matrix, λ is a real or complex number and $\underline{v}$ is a k-dimensional vector such that $K\underline{v} = \lambda\underline{v}$, then λ is said to be an eigenvalue of K, $\underline{v}$ is said to be an eigenvector of K, and $(\lambda,\underline{v})$ are an eigenpair for K.

Fact 1: Every $k \times k$ matrix has k eigenvalues, though they need not be distinct.

Fact 2: The eigenvalues of a $k \times k$ diagonal matrix K are the diagonal elements.

Proof: One may directly verify that the $(K(i,i),\underline{v})$ is an eigenpair, where $\underline{v} = (0 \ldots 0\ 1\ 0 \ldots 0)$ with a 1 in the ith place.

Fact 3: Real symmetric matricies have real (as opposed to complex) eigenvalues.

The eigenvectors associated with distinct eigenvalues are orthogonal.

There is an orthonormal set of eigenvectors.

If the matrix is also positive (respectively, nonnegative) definite, i.e. $\underline{x}^t K \underline{x} > 0$ for all $\underline{x}$, (respectively, $\underline{x}^t K \underline{x} \geq 0$), then the eigenvalues are positive (respectively, nonnegative).

Fact 4: The determinant of a square matrix is the product of its eigenvalues. i.e.e if a $k \times k$ matrix has eigenvalues $\lambda_1,\ldots,\lambda_k$, then

$$|K| = \prod_{j=1}^{k} \lambda_j$$

Fact 5: The determinant of an orthogonal matrix T is $|T| = \pm 1$.

Proof: Suppose λ is an eigenvalue of T and $\underline{v}$ is a corresponding eigenvector, i.e. $T\underline{v} = \lambda\underline{v}$. Then

$$||\underline{v}|| = ||T\underline{v}|| = ||\lambda\underline{v}|| = |\lambda| \, ||\underline{v}||$$

which implies $|\lambda| = 1$. Since all the eigenvalues have magnitude one, Fact 1 implies $|T| = \pm 1$.

Facts about covariance matrices

Fact 6: For any $k \times k$ covariance matrix K, $\prod_{i=1}^{k} K_{i,i} \geq |K|$, equality iff K is diag'l.

Proof: See Gersho and Gray, pp. 241-242

Fact 7: If $\underline{U} = T\underline{X}$, then $K_{\underline{U}} = T K_{\underline{X}} T^t$

Proof: $K_{\underline{U}} = E\,\underline{U}\underline{U}^t = E\,T\underline{X}\,(T\underline{X})^t = E\,T\underline{X}\underline{X}^t T^t = T\,E\underline{X}\underline{X}^t T^t = T\,K_{\underline{X}}\,T^t$.

Fact 8: If T is orthogonal and $\underline{U} = T\underline{X}$, then

(a) $K_{\underline{X}}$ and $K_{\underline{U}}$ have the same eigenvalues.

(b) $|K_{\underline{U}}| = |K_{\underline{X}}|$

Proof: (a) Let $(\lambda,\underline{v})$ be an eigenpair for $K_{\underline{X}}$. Then $K_{\underline{U}}T\underline{v} = TK_{\underline{X}}T^tT\underline{v} = T\lambda\underline{v} = \lambda T\underline{v}$. Hence, $(\lambda T\underline{v})$ is an eigenpair for $K_{\underline{U}}$.

Thus every eigenvalue of $K_{\underline{X}}$ is also an eigenvalue of $K_{\underline{U}}$.

The same argument applied to $K_{\underline{U}}$ and T^{-1} shows that every eigenvalue of $K_{\underline{U}}$ is also an eigenvalue of $K_{\underline{X}}$. Thus $K_{\underline{X}}$ and $K_{\underline{U}}$ have the same eigenvalues.

(b) Since $K_{\underline{X}}$ and $K_{\underline{U}}$ have the same eigenvalues, Fact 1 implies they have the same determinants.

Fact 9: For any covariance matrix there is a set of k orthonormal eigenvectors.

Proof: Because covariance matrices are real and symmetric, and Fact 3.

Proof of Optimality of KLT

Lemma: If T is orthogonal and makes $K_{\underline{U}}$ is diagonal, then for any other orthogonal matrix $\tilde{T}$, $\tilde{\Gamma}^k \geq \Gamma^k$, with equality if and only if $K_{\underline{\tilde{U}}}$ is diagonal.

Proof:

$\tilde{\Gamma}^k = \prod_{i=1}^{k} \tilde{\sigma}_i^2$ by definition of $\tilde{\Gamma}$, where $\tilde{\sigma}_i^2 = E\tilde{U}_i^2$ & $\underline{\tilde{U}} = \tilde{T}\underline{X}$.

$\geq |K_{\underline{\tilde{U}}}|$ Fact 6 fact that $\tilde{\sigma}_i^2$'s are diag elements of $K_{\underline{\tilde{U}}}$

$= |K_{\underline{X}}|$ Fact 8(b)

$= |K_{\underline{U}}|$ Fact 8(b)

$= \prod_{i=1}^{k} \sigma_i^2$ Fact 6 & fact that σ_i^2's are diag elements of $K_{\underline{U}}$

$= \Gamma^k$ for T

By Fact 6, equality holds iff $K_{\underline{\tilde{U}}}$ is diagonal.

We see from this lemma that one can do no better than to make the transform coefficients have a diagonal covariance matrix.

When this is done, the coef variances are the diagonal elements, which by Fact 2, are the eigenvalues of $K_{\underline{U}}$. By Fact 8a, these are also the eigenvalues of $K_{\underline{X}}$ as claimed in the Main Fact.

It remains only to show that there is a choice of T that makes $K_{\underline{U}}$ diagonal.

Accordingly, let T be the matrix mentioned in the Main Fact, i.e. its rows are an orthonormal set of eigenvectors $\underline{t}_1,\ldots,\underline{t}_k$

Then, by Fact 5, and the fact that $\underline{t}_i$ is an eigenvector with eigenvalue λ_i

$$K_{\underline{U}} = T K_{\underline{X}} T^t = \begin{bmatrix} --\underline{t}_1^t-- \\ --\underline{t}_2^t-- \\ \cdots \\ --\underline{t}_k^t-- \end{bmatrix} K_{\underline{X}} \begin{bmatrix} | & | & & | \\ \underline{t}_1 & \underline{t}_2 & \cdots & \underline{t}_k \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} --\underline{t}_1^t-- \\ --\underline{t}_2^t-- \\ \cdots \\ --\underline{t}_k^t-- \end{bmatrix} \begin{bmatrix} | & | & & | \\ \lambda_1\underline{t}_1 & \lambda_2\underline{t}_2 & \cdots & \lambda_k\underline{t}_k \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 \\ & & & \\ 0 & \cdots & 0 & \lambda_k \end{bmatrix}.$$

So as we hoped, $K_{\underline{U}}$ is diagonal, which completes the proof of the main result.

Substituting the value $\Gamma = |K_{\underline{X}}|^{1/k}$ into the expression (**) for distortion gives:

The OPTA Function for k-dimensional Transform Coding applied to a Zero-Mean Stationary, Gaussian Source:

For large R,

$$\delta_{tr}(k,R) \cong \frac{1}{12} |K_{\underline{X}}|^{1/k} \alpha\, 2^{-2R}, \quad \text{where } \alpha = 32.6 \text{ for FRC, 17.08 for VRC.}$$

SNR Gain:

$$10 \log_{10} \frac{\delta_{sq}(R)}{\delta_{tr}(k,R)} \cong 10 \log_{10} \frac{\sigma_X^2}{|K_{\underline{X}}|^{1/k}}$$

The best transform is the KLT, i.e. rows are orthonormal eigenvectors for $K_{\underline{X}}$.

The resulting coefficients $U_1,\ldots,U_k$ are uncorrelated (indeed, independent),

Their variances $\sigma_1^2,\ldots,\sigma_k^2$ equal the eigenvalues $\lambda_1,\ldots,\lambda_k$ of $K_{\underline{X}}$.

The rate allocated to the ith coefficient is: $R_i = R + \frac{1}{2} \log_2 \frac{\lambda_i}{|K_{\underline{X}}|^{1/k}}$

The resulting coeff. distortions are all the same and equal to $\delta_{tr}(k,R)$.