Transform Coding: Background and Motivation*

Phil Schniter

Abstract

Transform coding is described and an analysis is performed for the simple 2-dimensional case, including a comparison to PCM.

- In transform coding (TC), blocks of N input samples are transformed to N transform coefficients which are then quantized and transmitted. At the decoder, an inverse transform is applied to the quantized coefficients, yielding a reconstruction of the original waveform. By designing individual quantizers in accordance with the statistics of their inputs, it is possible to allocate bits in a more optimal manner, e.g., encoding the "more important" coefficients at a higher bit rate.

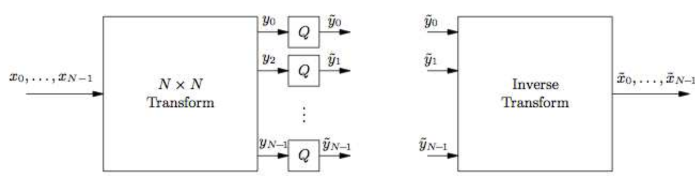

Figure 1: $N \times N$ Transform Coder/Decoder with Scalar Quantization

- *Orthogonal Transforms:* From our perspective, an $N \times N$ "transform" will be any real-valued linear operation taking N input samples to N output samples, or transform coefficients. This operation can always be written in matrix form

$$\mathbf{y}(m) = \mathbf{T}\mathbf{x}(m), \quad \mathbf{T} \in \mathbb{R}^{N \times N} \tag{1}$$

where $\mathbf{x}(m)$ and $\mathbf{y}(m)$ are vectors representing $N \times 1$ blocks of input/output elements:

$$\begin{aligned} \mathbf{x}(m) &= (x(mN),\ x(mN-1),\ \cdots,\ x(mN-N+1))^t \\ \mathbf{y}(m) &= (y(mN),\ y(mN-1),\ \cdots,\ y(mN-N+1))^t. \end{aligned} \tag{2}$$

*Version 1.2: Sep 23, 2009 9:45 am -0500
†http://creativecommons.org/licenses/by/3.0/

Intuition comes from considering the transform's *basis vectors*$\{\mathbf{t}_k\}$ defined by the rows of the matrix

$$\mathbf{T} = \begin{pmatrix} - \mathbf{t}_0^t - \\ - \mathbf{t}_1^t - \\ \vdots \\ - \mathbf{t}_{N-1}^t - \end{pmatrix} \quad (3)$$

since the coefficient $y_k = \mathbf{t}_k^t \mathbf{x}$ can be thought of as the result of a "comparison" between the k^{th} basis vector and the input $\mathbf{x}$. These comparisons are defined by the inner product $<\mathbf{t}_k, \mathbf{x}> = \mathbf{t}_k^t \mathbf{x}$ which has a geometrical interpretation involving the angle θ_k between vectors t_k and $\mathbf{x}$.

$$<\mathbf{t}_k, \mathbf{x}> \;=\; cos(\theta_k) \;\| \mathbf{t}_k\|_2 \| \mathbf{x} \|_2. \quad (4)$$

When the vectors $\{\mathbf{t}_k\}$ are mutually orthogonal, i.e., $\mathbf{t}_k^t \mathbf{t}_\ell = 0$ for $k \neq \ell$, the transform coefficients represent separate, unrelated features of the input. This property is convenient if the transform coefficients are independently quantized, as is typical in TC schemes.

Example 1: 2×2 **Transform Coder**

Say that stationary zero-mean Gaussian source $x(m)$ has autocorrelation $r_x(0) = 1$, $r_x(1) = \rho$, and $r_x(k) = 0$ for $k > 1$. For a bit rate of R bits per sample, uniformly-quantized PCM implies a mean-squared reconstruction error of

$$\sigma_r^2|_{\text{PCM}} = \frac{\Delta^2}{12}\Big|_{\substack{\Delta = 2x_{max}/L \\ L = 2^R}} \;=\; \underbrace{\frac{1}{3}\frac{x_{max}^2}{\sigma_x^2}}_{\gamma_x}\sigma_x^2 2^{-2R} \;=\; \gamma_x \sigma_x^2 2^{-2R}. \quad (5)$$

For transform coding, say we choose linear transform

$$\mathbf{T} \;=\; \begin{pmatrix} \mathbf{t}_0^t \\ \mathbf{t}_1^t \end{pmatrix} \;=\; \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (6)$$

Setting $\mathbf{x}(m) = \begin{pmatrix} x(2m) & x(2m-1) \end{pmatrix}^t$ and $\mathbf{y}(m) = \mathbf{T}\mathbf{x}(m)$, we find that the transformed coefficients have variance

$$\sigma_{y_0}^2 \;=\; E\{|\mathbf{t}_0^t \mathbf{x}(m)|^2\} \;=\; \frac{1}{2}E\{|x(2m) + x(2m-1)|^2\} \;=\; \frac{1}{2}(2r_x(0) + 2r_x(1)) \;=\; 1+\rho \quad (7)$$

$$\sigma_{y_1}^2 = E\{|\mathbf{t}_1^t \mathbf{x}(m)|^2\} \;=\; \frac{1}{2}E\{|x(2m) - x(2m-1)|^2\} \;=\; \frac{1}{2}(2r_x(0) - 2r_x(1)) \;=\; 1-\rho \quad (8)$$

and using uniformly-quantized PCM on each coefficient we get mean-squared reconstruction errors

$$\sigma_{q_0}^2 \;=\; (1+\rho)\,\gamma_x 2^{-2R_0} \quad (9)$$

$$\sigma_{q_1}^2 \;=\; (1-\rho)\,\gamma_x 2^{-2R_1}. \quad (10)$$

We use the same quantizer performance factor γ_x as before since linear operations preserve Gaussianity.

For orthogonal matrices $\mathbf{T}$, i.e., $\mathbf{T}^{-1} = \mathbf{T}^t$, we can show that the mean-squared reconstruction error ${\sigma_r}^2$ equals the mean-squared quantization error:

$$\begin{array}{rcl} \sigma_r^2 & := & \frac{1}{N}\sum_{k=0}^{N-1} E\{(\tilde{x}\,(Nm-k) - x\,(Nm-k))^2\} \qquad \text{(here } N=2\text{)} \\ & = & \frac{1}{N}E\{\parallel \tilde{\mathbf{x}}(m) - \mathbf{x}\,(m) \parallel^2\} \\ & = & \frac{1}{N}E\{\parallel \mathbf{T}^{-1}\tilde{\mathbf{y}}(m) - \mathbf{x}\,(m) \parallel^2\} \\ & = & \frac{1}{N}E\{\parallel \mathbf{T}^{-1}\,(\mathbf{y}\,(m) + \mathbf{q}\,(m)) - \mathbf{x}\,(m) \parallel^2\} \\ & = & \frac{1}{N}E\{\parallel \mathbf{T}^{-1}\mathbf{T}\mathbf{x}\,(m) + \mathbf{T}^{-1}\mathbf{q}\,(m) - \mathbf{x}\,(m) \parallel^2\} \\ & = & \frac{1}{N}E\{\parallel \mathbf{T}^{-1}\mathbf{q}\,(m) \parallel^2\} \\ & = & \frac{1}{N}E\{\mathbf{q}^t\,(m)\underbrace{\left(\mathbf{T}^{-1}\right)^t\mathbf{T}^{-1}}_{\mathbf{I}}\mathbf{q}\,(m)\} \\ & = & \frac{1}{N}E\{\parallel \mathbf{q}\,(m) \parallel^2\} \\ & = & \frac{1}{N}\sum_{k=0}^{N-1}\sigma_{q_k}^2. \end{array} \tag{11}$$

Since our 2×2 matrix is indeed orthogonal, we have mean-squared reconstruction error

$$\sigma_r^2|_{\mathrm{TC}} = \frac{1}{2}\left((1+\rho)\,\gamma_x 2^{-2R_0} + (1-\rho)\,\gamma_x 2^{-2R_1}\right) \tag{12}$$

at bit rate of $R_0 + R_1$ bits per two samples. Comparing TC to PCM at equal bit rates (i.e. $R_0 + R_1 = 2R$),

$$\frac{\sigma_r^2|_{\mathrm{TC}}}{\sigma_r^2|_{\mathrm{PCM}}} = \frac{1}{2}\frac{(1+\rho)\,\gamma_x 2^{-2R_0} + (1-\rho)\,\gamma_x 2^{-2(2R-R_0)}}{\gamma_x 2^{-2R}} = (1+\rho)\,2^{2(R-R_0)-1} + (1-\rho)\,2^{2(R_0-R)-1}. \tag{13}$$

Figure 2 shows that (i) allocating a higher bit rate to the quantizer with stronger input signal can reduce the average reconstruction error relative to PCM, and (ii) the gain over PCM is higher when the input signal exhibits stronger correlation ρ. Also note that when $R_0 = R_1 = R$, there is no gain over PCM—a verification of the fact that $\sigma_r^2 = \sigma_q^2$ when $\mathbf{T}$ is orthogonal.

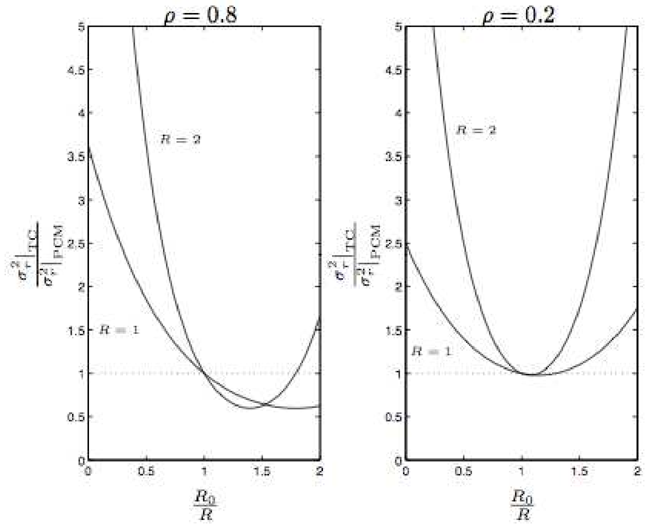

Figure 2: Ratio of TC to PCM mean-squared reconstruction errors versus bit rate R_0 for two values of ρ.