Transform coding

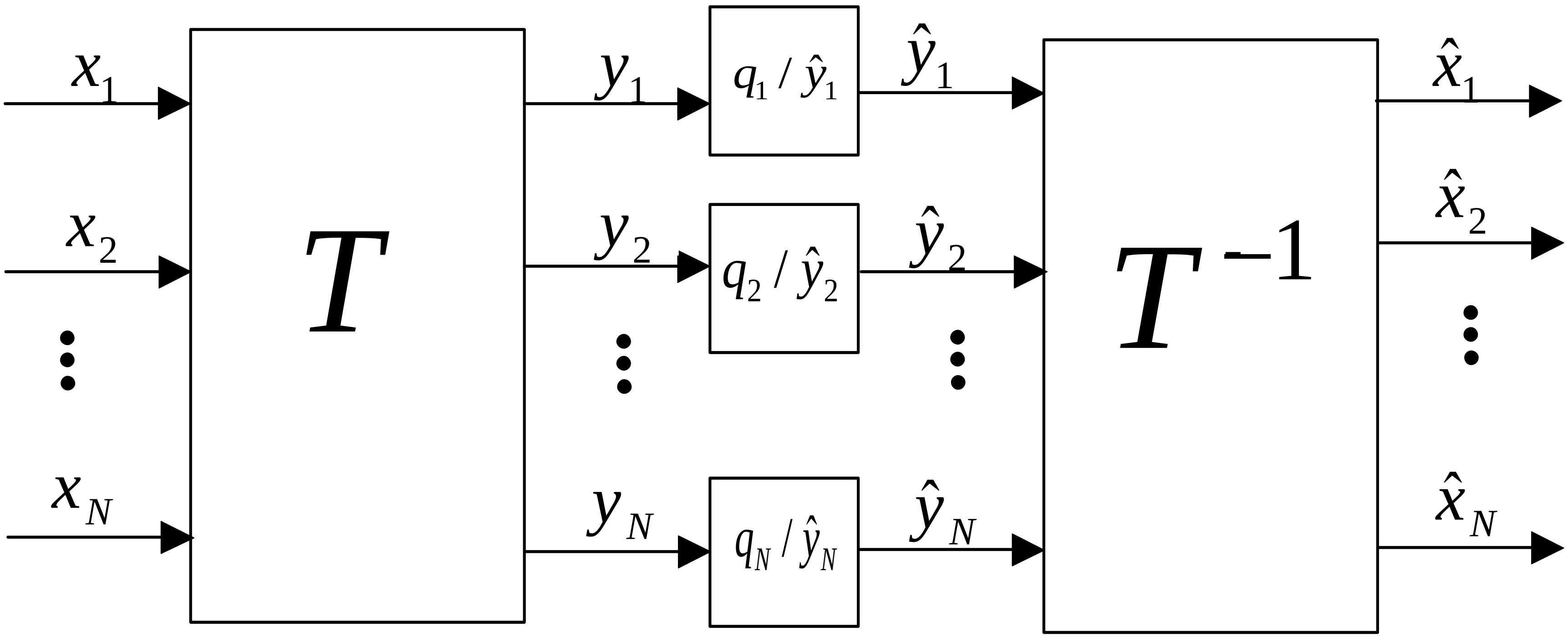
Transform coding is the second approach to exploiting redundancy by using scalar quantization with linear preprocessing.

The source samples are collected into a vector that is linearly transformed (multiplied by a transform matrix) and the resulting coefficients are scalar quantized.

Notice that each coefficient can be quantized by different quantizer. The method was introduced in 1956 by Kramer and Mathews and then popularized in 1962-1963 by Huang and Schultheiss.

It was developed for coding images and video. More recently, transform coding has also been widely used in high-fidelity audio coding.

Transform coding



Transform coding

A typical discrete linear transform is a decomposition of the input discrete-time signal over a system of basis functions.

Harmonic functions are often used as basis functions. In this case coefficients represent intensities of the corresponding harmonics.

Transformations of this type are called conversion to the frequency domain.

Let $\mathbf{x}$ be an input column vector of dimension N then linear transform of $\mathbf{x}$ can be expressed as follows

$$\mathbf{y} = T\mathbf{x},$$

where T is $N \times N$ transform matrix.

Properties of transforms

Usually it is required that transform would have the following properties:

Localization of the essential part of the signal energy in a small number of coefficients. After quantization we can exclude from consideration the least informative coefficients .

•Coefficients should be uncorrelated. In this case scalar quantization followed by symbol-by-symbol variable-length coding provides $R(D)$ close to $H(D)$.

•Transform should be orthonormal. If this property holds then the MSE introduced by quantizing the transform coefficients coincides with MSE occurring in the input vector. Preserves H(D)!

•Low computational complexity. It is desirable to use the separable transforms.

Properties of transforms

A transform is called the orthonormal if

$$T^{-1} = T^{*}, \tag{6.1}$$

where T is a matrix with complex elements. If T is real then (6.1) reduces to the condition

$$T^{-1} = T^{T}. \tag{6.2}$$

(6.2) is equivalent to the following relation for rows of $T = \{\mathbf{t}_i\},\ i = 1,2,\ldots,N$

$$\mathbf{t}_i \mathbf{t}_j^T = \delta_{ij},$$

where $\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ is Kronecker's delta function.

Properties of transforms

Vectors $\mathbf{t}_i$ are the orthonormal basis vectors of transform T.

For orthonormal transform

$$\mathbf{y} = T\mathbf{x}, \quad \mathbf{x} = T^{-1}\mathbf{y} = T^{T}\mathbf{y}, \quad \mathbf{x} = \sum_{i=1}^{N} y_i \mathbf{t}_i^{T}.$$

The input vector $\mathbf{x}$ can be represented as a weighted sum of basis vectors, y_i are transform coefficients.

The orthonormal transform preserves the signal energy

$$\sum_{i=1}^{N} x_i^2 = \sum_{j=1}^{N} \left| y_j \right|^2.$$

This property is known as the discrete Parseval's theorem

$$\sum_{i=1}^{N} x_i^2 = \mathbf{x}^T\mathbf{x} = (T^T\mathbf{y})^T(T^T\mathbf{y}) = \mathbf{y}^T T T^T \mathbf{y} = \mathbf{y}^T T T^{-1}\mathbf{y} = \mathbf{y}^T\mathbf{y} = \sum_{j=1}^{N} \left| y_j \right|^2$$

Properties of transforms

Noncorrelatedness of transform coefficients.

This property implies that the transform coefficients y_i, $i = 1,2,\dots,N$ satisfy the condition

$$\mathrm{E}\left\{\left(y_i - \mathrm{E}\{y_i\}\right)\left(y_j - \mathrm{E}\{y_j\}\right)\right\} = \alpha_i \delta_{ij}, \quad \forall i, j$$

where α_i is a variance of y_i, $\mathrm{E}\{\cdot\}$ denotes mathematical expectation. For simplicity we assume $\mathrm{E}\{\mathbf{x}\} = \mathrm{E}\{\mathbf{y}\} = \mathbf{0}$.

Localization of the essential part of signal energy in a small number of transform coefficients.

Let $y_1 \dots, y_N$ be sorted in such a manner that

$$\alpha_1 \ge \alpha_2 \ge \dots \ge \alpha_N .$$

Properties of transforms

Assume that only first $pN,\ 0<p<1$ coefficients are transmitted. The receiver uses truncated vector $\hat{\mathbf{y}}=(y_1,\dots,y_{pN},0,\dots,0)^T$ to reconstruct $\hat{\mathbf{x}}=T^{-1}\hat{\mathbf{y}}=T^T\hat{\mathbf{y}}$.

The MSE occurred when we replace $\mathbf{x}$ by $\hat{\mathbf{x}}$ is

$$\mathrm{E}\left\{\frac{1}{N}\sum_{i=1}^{N}(x_i-\hat{x}_i)^2\right\}=\frac{1}{N}\mathrm{E}\left\{(\mathbf{x}-\hat{\mathbf{x}})^T(\mathbf{x}-\hat{\mathbf{x}})\right\}=$$

$$=\frac{1}{N}\mathrm{E}\left\{(\mathbf{y}^TT-\hat{\mathbf{y}}^TT)(T^T\mathbf{y}-T^T\hat{\mathbf{y}})\right\}=\frac{1}{N}\mathrm{E}\left\{\sum_{j=1}^{N}(y_j-\hat{y}_j)^2\right\}=$$

$$=\frac{1}{N}E\left\{\sum_{j=pN+1}^{N}y_j^2\right\}=\frac{1}{N}\sum_{j=pN+1}^{N}\alpha_j. \qquad (6.3)$$

We would like to find the orthonormal transform which minimizes the error (6.3)

Properties of transforms

Any orthonormal transform preserves the achievable rate-distortion function $H(D)$!!!

Low computational complexity.

To apply nonseparable 2-D transform we rearrange input matrix X into a vector $\mathbf{x}$ with N^2 components and multiply it by the transform matrix T_2 of size $N^2 \times N^2$

$$\mathbf{y} = T_2 \mathbf{x}$$

For separable transform the $N \times N$ matrix of the transform coefficients can be obtained as $Y = TXT^T$

The $N^2 \times N^2$ matrix T_2 is the Kronecker product of two $N \times N$ matrices T of a 1-D transform

We reduce the computational complexity to $2N^3$ instead of N^4

The Karhunen-Loeve transform

The most efficient in terms of listed properties is the Karhunen-Loeve transform.

- This transform is orthonormal.
- Its coefficients are uncorrelated.
- The KL transform minimizes (among all orthonormal transforms) the MSE (6.3) occured because of rejecting transform coefficients with small variances.

KL transform is optimal in terms of localization signal energy and it maximizes the number of transform coefficients which are insignificant and might be quantized to 0

The Karhunen-Loeve transform

Let T_{KL} be the matrix of KL transform.

The covariance matrix of the input vector $\mathbf{x}$ is expressed via covariance matrix of transform coefficients as

$$R = \mathrm{E}\{\mathbf{x}\mathbf{x}^T\} = T_{KL}^T \mathrm{E}\{\mathbf{y}\mathbf{y}^T\} T_{KL}.$$

Multiplying by T_{KL}^T we obtain

$$R T_{KL}^T = T_{KL}^T \mathrm{E}\{\mathbf{y}\mathbf{y}^T\}$$

Since coefficients are uncorrelated $\mathrm{E}\{\mathbf{y}\mathbf{y}^T\} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & \lambda_N \end{pmatrix}$,

where λ_i is the variance of y_i.

The Karhunen-Loeve transform

$$\boxed{R\mathbf{t}_{KLi}^{T} = \lambda_i \mathbf{t}_{KLi}^{T},} \quad i = 1,\ldots, N$$

Thus the basis vectors of the Karhunen-Loeve transform are eigenvectors of the covariance matrix R normalized to satisfy $\mathbf{t}_{KLi}\mathbf{t}_{KLj}^{T} = \delta_{ij}$.

The variances $\{\lambda_i\}$ of the transform coefficients are eigenvalues of R.

Since R is symmetric and positive definite matrix ($\mathbf{x}A\mathbf{x}^T > 0$ for any nonzero $\mathbf{x}$) then eigenvalues are real and positive. As the result the basis vectors and the transform coefficients of KL transform are real.

The Karhunen-Loeve transform

It can be shown that among all possible orthonormal transforms applied to stationary vectors of dimension N the KL transform minimizes the MSE (6.3) occurred due to truncation, that is, the KL transform is optimal in terms of localization signal energy.

Equation (6.3) for the KL transform has the form

$$\frac{1}{N} E\left\{ \sum_{j=pN+1}^{N} y_j^2 \right\} = \frac{1}{N} \sum_{j=pN+1}^{N} \lambda_j .$$

The main shortcoming of the KL transform is that its basis functions depend on the transformed signal. We have to store not only quantized transform coefficients but also the basis functions which can require many more bits for storing than the quantized coefficients.

The discrete Fourier transform

The DFT is the counterpart of the continuous Fourier transform. It is defined for the discrete-time signals. The transformed signal represents samples of the signal spectrum.

Let $x(nT_s)$ be input sequence then the DFT of $x(nT_s)$ is

$$X(k\Omega) = \sum_{n=0}^{N-1} x(nT_s)e^{-jkn\Omega T_s}, \quad 0 \le k \le N-1. \qquad (6.4)$$

The inverse transform is defined as follows

$$x(nT_s) = \frac{1}{N}\sum_{k=0}^{N-1} X(k\Omega)e^{jkn\Omega T_s}, \quad 0 \le n \le N-1, \qquad (6.5)$$

where $\Omega = 2\pi / NT_s$ is the base frequency of the transform

or the distance between samples of the signal spectrum.

The discrete Fourier transform

Notice that (6.4) determines a periodical sequence of numbers with period N.

Expressions (6.4),(6.5) can be rewritten in the form

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn}, \quad 0 \le k \le N-1$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W_N^{-kn}, \; 0 \le n \le N-1,$$

where $W_N = e^{-j2\pi/N}$.

In the matrix form (6.4), (6.5) can be rewritten as

$$\mathbf{X} = T_F \mathbf{x}, \quad \mathbf{x} = T_F^{-1} \mathbf{X} = \frac{1}{N} T_F^{*} \mathbf{X},$$

The discrete Fourier transform

$$\mathbf{X} = (X(0), X(1), \ldots, X(N-1))^T,$$

$$\mathbf{x} = (x(0), x(1), \ldots, x(N-1))^T,$$

$$T_F = \begin{pmatrix} 1 & 1 & \ldots & 1 & 1 \\ 1 & W_N & \ldots & W_N^{N-2} & W_N^{N-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & W_N^{N-2} & \ldots & W_N^{(N-2)(N-2)} & W_N^{(N-2)(N-1)} \\ 1 & W_N^{N-1} & \ldots & W_N^{(N-1)(N-2)} & W_N^{(N-1)(N-1)} \end{pmatrix}$$

is the transform matrix. The basis functions are powers of

$$W_N = e^{-j2\pi/N}.$$

The discrete Fourier transform

Since $T_F^{-1} = \frac{1}{N} T_F^*$ the DFT is orthogonal transform. It is easy to normalize the DFT in order to obtain the orthonormal transform. For this purpose we should use factors $1/\sqrt{N}$ in the forward and in the inverse transform instead of using factor $1/N$ in the inverse transform.

The 2-dimensional DFT is defined as follows

$$\mathbf{X}(k,l) = \sum_{n=0}^{N-1}\sum_{m=0}^{M-1} \mathbf{x}(n,m) e^{-j\left(kn\frac{2\pi}{N} + lm\frac{2\pi}{M}\right)} \quad 0 \le k \le N-1, \quad 0 \le l \le M-1,$$

$\mathbf{x}(n,m)$ is the element of the input matrix $\mathbf{x}$, $\mathbf{X}(k,l)$ is the element of $\mathbf{X}$.

The discrete Fourier transform

Since

$$\mathbf{X}(k,l)=\sum_{m=0}^{M-1}\left\{\sum_{n=0}^{N-1}\mathbf{x}(n,m)e^{-jkn\frac{2\pi}{N}}\right\}e^{-jlm\frac{2\pi}{M}}=\sum_{n=0}^{N-1}\left\{\sum_{m=0}^{M-1}\mathbf{x}(n,m)e^{-jlm\frac{2\pi}{M}}\right\}e^{-jkn\frac{2\pi}{N}}$$

the 2-dimensional DFT can be split into two 1-dimensional transforms, that is, the DFT is a separable transform.

Linearity: $DFT\{ax(n)+by(n)\}=aDFT\{x(n)\}+bDFT\{y(n)\}$

Circular convolution: Let $X(k)$ and $Y(k)$ be DFTs of $x(n)$ and $y(n)$, respectively. Then the inverse transform of the product $X(k)Y(k)$ is $v(l)=\frac{1}{N}\sum_{k=0}^{N-1}X(k)Y(k)e^{j\frac{2\pi}{N}kl}$.

The discrete Fourier transform

By inserting definitions for $X(k)$ and $Y(k)$ we obtain

$$v(l) = \frac{1}{N}\sum_{k=0}^{N-1}\left[\sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn}\right]\times\left[\sum_{m=0}^{N-1} y(m)e^{-j\frac{2\pi}{N}km}\right]e^{j\frac{2\pi}{N}kl}.$$

Changing the order of summation we get

$$v(l) = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} x(n)y(m)\left[\sum_{k=0}^{N-1} e^{j\frac{2\pi k}{N}(l-n-m)}\right].$$

The sum in brackets is equal to 0 for all m and n except $m = (l-n) \bmod N$ for which it is equal to N.

$$v(l) = \sum_{n=0}^{N-1} x(n)y((l-n) \bmod N) = \sum_{n=0}^{N-1} y(n)x((l-n) \bmod N)$$

is circular or periodical convolution.

Example

Let $x(n) = (2,\ 3,\ 1)$ and $y(n) = (1,\ 2,\ 4)$ We extend $x(n)$ periodically $x(n) =$ (…, 2, 3, 1, 2, 3, 1, 2, 3, 1,…)

Then $v(l)$ is determined as follows

$$v(0) = y(0)x(0) + y(1)x(2) + y(2)x(1) = 1 \cdot 2 + 2 \cdot 1 + 4 \cdot 3 = 16$$

$$v(1) = y(0)x(1) + y(1)x(0) + y(2)x(2) = 1 \cdot 3 + 2 \cdot 2 + 4 \cdot 1 = 11$$

$$v(2) = y(0)x(2) + y(1)x(1) + y(2)x(0) = 1 \cdot 1 + 2 \cdot 3 + 4 \cdot 2 = 15$$

The periodical sequence $x(n)$ is reversed in time and multiplied by the corresponding term in $y(n)$. Then the reversed $x(n)$ is shifted by 1 sample to the right and again multiplied by $y(n)$ generating the new sample of $v(l)$.

Example

The periodical reversed $x(n)$

1 3 2 1 3 2 1 3 2 1 3 2

$y(n)$ 1 2 4

The reversed $x(n)$ shifted to the right

2 1 3 2 1 3 2 1 3 2 1 3 2

3 2 1 3 2 1 3 2 1 3 2 1 3

The discrete Fourier transform

$$DFT(x((n-l) \bmod N)) = X(k)W_N^{lk}$$

That is displacement of l samples from the end to the beginning of $x(n)$ is equivalent to multiplying DFT of this sequence by $\exp\{-j2\pi lk/N\}$

Example.

Let $x(n)$ be equal to 1 2 3 4 5 3 7 4 and its DFT be equal to $X(k)$ then $x(n-3) \bmod 8$ is 3 7 4 1 2 3 4 5 and its DFT is determined as $X(k)\exp\{-j2\pi \times 3k/8\}$

The discrete Fourier transform

In general case the DFT and IDFT require approximately

N^2 additions and N^2 multiplications of complex numbers.

There exist the so-called fast algorithms which allow to reduce the computational complexity of DFT to $N \log_2 N$ operations.