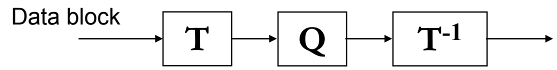
ECE 8873

Data Compression and Modeling

Lecture 10: Transform & Subband Coding

School of Electrical and Computer Engineering
Georgia Institute of Technology
Spring, 2004

Transform – What and Why



- Structural component of the information may be made more explicit after transformation than the original observation, e.g.,
  - Rotation, translation
  - Frequency domain
- Transform may decorrelate or decompose the observation into compact components for more efficient coding – method of principle components
- Issue of perception may be more easily addressed in the transform domain (e.g., masking).

Linear Transform – General Expression

$$\boldsymbol{\theta} = \mathbf{A}\mathbf{x} \qquad \theta_n = \sum_{i=0}^{N-1} x_i a_{ni} \qquad \mathbf{A} = \left[a_{ij}\right]_{i,j=0}^{N-1}$$

$$\mathbf{x} = \mathbf{B}\boldsymbol{\theta} \qquad x_n = \sum_{i=0}^{N-1} \theta_i b_{ni} \qquad \mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

If $\mathbf{x}$ is two-dimensional,

$$\boldsymbol{\Theta} = \mathbf{A}\,\mathbf{X}\,\mathbf{A}^t \qquad \Theta_{kl} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} x_{ij} a_{ijk} = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} a_{ki} x_{ij} a_{lj}$$

$$\mathbf{X} = \mathbf{B}\,\boldsymbol{\Theta}\,\mathbf{B}^t$$

Use separable transform to simplify

Orthonormal transform $\mathbf{B} = \mathbf{A}^{-1} = \mathbf{A}^t$

$$\mathbf{X} = \mathbf{A}^t\,\boldsymbol{\Theta}\,\mathbf{A} \qquad \sum_{n=0}^{N-1} x_n^2 = \sum_{n=0}^{N-1} \theta_n^2$$

Discrete Karhunen-Loéve Transform

- Also known as Hotelling transform; maximum coding gain in the sense of variance minimization
- Rows are eigenvectors of the autocorrelation matrix of data – again recall structural component of information

$$R = U\Lambda U^{-1}$$

- If signal is changing with time, or short-time spectral analysis is involved, the autocorrelation matrix and the eigenvectors will also change with time. Sending the corresponding side information may be costly.

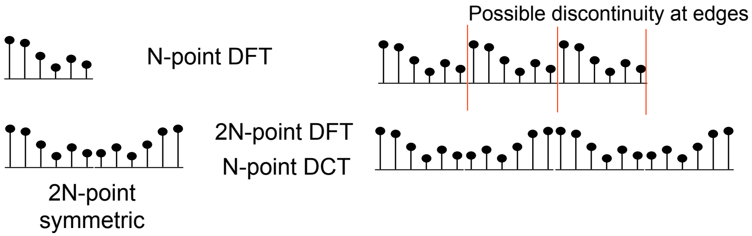
DFT and Discrete Cosine Transform

$$\mathbf{C} = \left[c_{ij}\right]_{i,j=0}^{N-1}$$

Extensively used in audio, image and video

$$c_{ij} = \begin{cases} \sqrt{\frac{1}{N}} \cos\frac{(2j+1)i\pi}{2N}, & i = 0,\ j = 0,1,2,\cdots,N-1 \\ \sqrt{\frac{2}{N}} \cos\frac{(2j+1)i\pi}{2N}, & i = 1,2,\cdots,N-1,\ j = 0,1,2,\cdots,N-1 \end{cases}$$

Relates to real part of 2N-point DFT

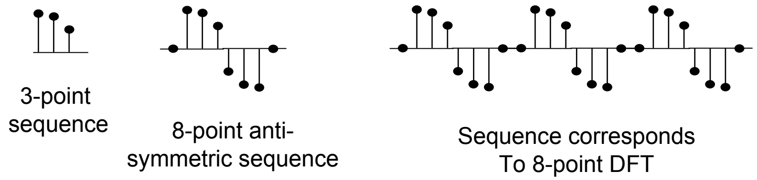


Discrete Sine Transform

$$\mathbf{S} = \left[s_{ij}\right]_{i,j=0}^{N-1}$$

$$s_{ij} = \sqrt{\frac{2}{N+1}} \sin\frac{(j+1)(i+1)\pi}{N+1}, \quad i,j = 0,1,2,\cdots,N-1$$

Relates to imaginary part of ~2N-point DFT



Discrete Walsh-Hadamard Transform

- Discrete Hadamard matrix of order *N*,

  $$H: \quad HH^t = NI$$

- Hadamard matrix of order 2N can be generated by

$$(H)_{2N}: \quad (H)_{2N} = \begin{bmatrix} (H)_N & (H)_N \\ (H)_N & -(H)_N \end{bmatrix}$$

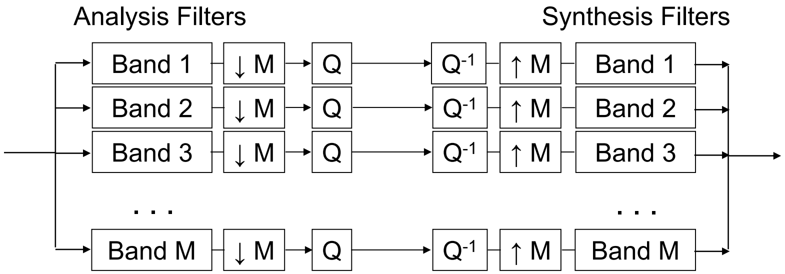
$$(H)_1 = 1 \qquad (H)_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

- Suitable in bit manipulation but not much gain in compression

Transform Coding - Issues

- Changing signal characteristics may require adaptive transforms; may require extensive side information.
- Any quantization error in the transform domain, even with one single coefficient, will spread throughout the entire signal block.
- Coding artifacts likely at block boundaries or edges.

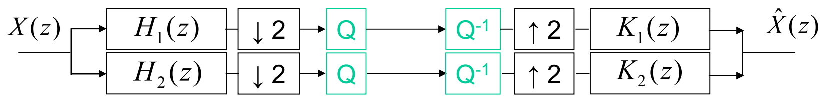
Subband Coding

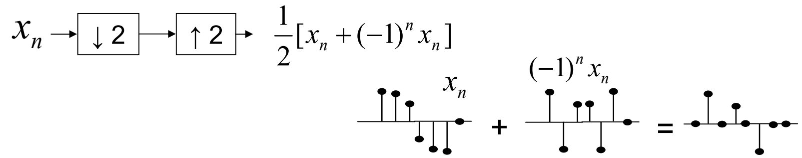


- As in vector quantization, dividing signals into separate frequency bands leads to sub-optimal coding performance – omitting the gain due to correlation between bands; but
- Different bands may represent different significance in the signal and dividing them would allow separate treatment for various purposes; e.g. non-uniform bit allocation to take advantage of masking, or to incorporate non-uniform error protection, or to create coding diversity.

Filter Design & Decimation

- Basic methodology covered in DSP course.
- Of interest here is “perfect reconstruction” filterbanks.

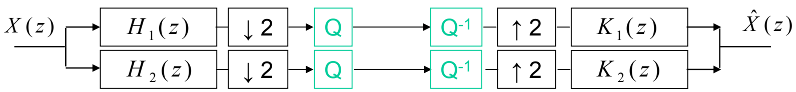




Without filtering, only with down- and up-sampling by 2

$$X(z) \Rightarrow \frac{1}{2}\left[X(z) + X(-z)\right]$$

Conditions for Perfect Reconstruction



$$\hat{X}(z) = \frac{1}{2}[X(z)H_1(z) + X(-z)H_1(-z)]K_1(z) + \frac{1}{2}[X(z)H_2(z) + X(-z)H_2(-z)]K_2(z)$$

$$= \frac{1}{2}\{[H_1(z)K_1(z) + H_2(z)K_2(z)]X(z) + [H_1(-z)K_1(z) + H_2(-z)K_2(z)]X(-z)\}$$

$$= \frac{1}{2}\begin{bmatrix}K_1(z) & K_2(z)\end{bmatrix}\begin{bmatrix}H_1(z) & H_1(-z)\\ H_2(z) & H_2(-z)\end{bmatrix}\begin{bmatrix}X(z)\\ X(-z)\end{bmatrix}$$

$\hat{X}(z) = cX(z)z^{-n_0}$ = scaled and delayed version of X

$$\begin{bmatrix}K_1(z) & K_2(z)\end{bmatrix}\begin{bmatrix}H_1(z) & H_1(-z)\\ H_2(z) & H_2(-z)\end{bmatrix} = \begin{bmatrix}cz^{-n_0} & 0\end{bmatrix}$$

$$\begin{bmatrix}K_1(z) & K_2(z)\end{bmatrix} = \frac{cz^{-n_0}}{\det[\mathbf{H}(z)]}\begin{bmatrix}H_2(-z) & -H_1(-z)\end{bmatrix} \qquad \mathbf{H}(z) = \begin{bmatrix}H_1(z) & H_1(-z)\\ H_2(z) & H_2(-z)\end{bmatrix}$$

$$\det[\mathbf{H}(z)] = H_1(z)H_2(-z) - H_1(-z)H_2(z) = P(z) - P(-z) = \gamma z^{-n_1}$$

where $P(z) = H_1(z)H_2(-z)$

Quadrature Mirror Filters

$$\hat{X}(z) = \frac{1}{2}\{[H_1(z)K_1(z) + H_2(z)K_2(z)]X(z) + [H_1(-z)K_1(z) + H_2(-z)K_2(z)]X(-z)\}$$
$$= T(z)X(z) + S(z)X(-z)$$

$S(z) = 0$ if $K_1(z) = H_2(-z)$ and $K_2(z) = -H_1(-z)$

$T(z) = cz^{-n_0} \quad \Rightarrow \quad |T(e^{j\omega})| = |c|$ and $\arg[T(e^{j\omega})] = k\omega$ (linear phase)

$$T(z) = \frac{1}{2}\left[H_1(z)H_2(-z) - H_1(-z)H_2(z)\right]$$

select lowpass filter $H_1(z)$

then choose $H_2(z) = H_1(-z)$ the mirror version of $H_1(z)$

$$\Rightarrow \quad T(z) = \frac{1}{2}\left[H_1^2(z) - H_1^2(-z)\right]$$

To achieve perfect reconstruction:

$$H_1(z) = h_0 z^{-2k_0} + h_1 z^{-(2k_1+1)} \quad \text{and} \quad T(z) = 2h_0 h_1 z^{-(2k_0+2k_1+1)}$$

Issues with Perfect Reconstruction

- To achieve perfect reconstruction (no amplitude distortion and phase distortion), very limited choices of filter response.
- The motivation for the subband approach is to address different significance in each band and thus being able to localize the band (compact bandwidth) is important.
- Filterbank design is thus result of trade-off between perfect reconstruction and bandwidth concentration in a joint criterion:

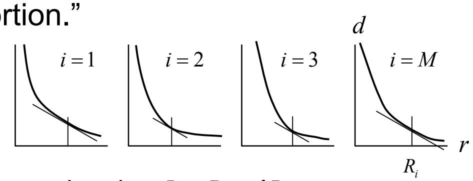
$$J = \alpha \underbrace{\int_{\omega_s}^{\pi} |H_1(e^{j\omega})|^2 \, d\omega}_{\text{Stop-band leakage}} + (1-\alpha) \underbrace{\int_0^{\pi} (1 - |T(e^{j\omega})|^2) d\omega}_{\text{Deviation from PR}}$$

- When quantization (a non-linear operation) is involved, harmonic distortions are inevitable and perfect reconstruction cannot be accomplished anyway.

Bit Allocation

- Distribute the bits available among M bands so as to minimize "total distortion."

$$R = \frac{1}{M} \sum_{i=1}^{M} R_i$$

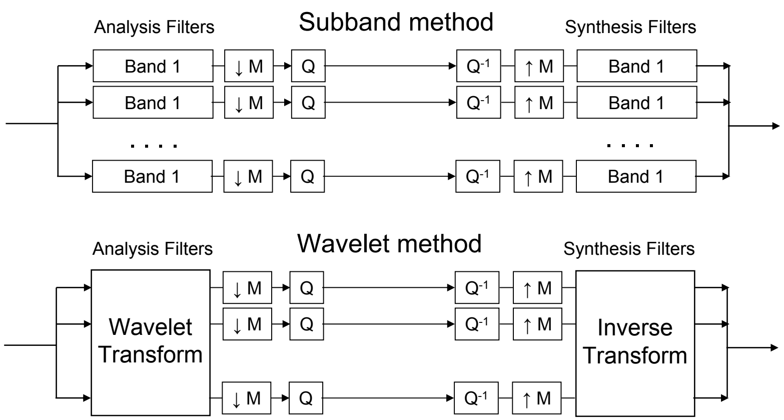


Use variational method across bands: $J_i = D_i + \lambda R_i$

Result: operating points in R-D curves have same slope; otherwise, can always trade bit in bands of small slope with bit in bands of high slope. (D may involve perceptual judgment.)

- Examine the rate-distortion function in each band;
- Find operating slope and the corresponding rate in each band;
- Check if the total rate satisfies requirement;
- If not, adjust the operating slope and repeat.

Wavelet-Based Methods



What is in the WT block? What is the difference?

Concept of Time-Frequency Resolution

- When interested in the changing characteristics of the signal, short time Fourier transform or short time spectral analysis is used.

$$F(\omega, t) = \int_{-\infty}^{\infty} f(\tau) g(t-\tau) e^{-j\omega\tau} d\tau$$

$g(t)$ is the window function providing focus in time.

Temporal spread: $\sigma_t^2 = \int_{-\infty}^{\infty} (t-t_c)^2 |g(t)|^2 dt \bullet \left[\int_{-\infty}^{\infty} |g(t)|^2 dt \right]^{-1}$

Frequency spread: $\sigma_\omega^2 = \int_{-\infty}^{\infty} (\omega-\omega_c)^2 |G(\omega)|^2 d\omega \bullet \left[\int_{-\infty}^{\infty} |G(\omega)|^2 d\omega \right]^{-1}$

If $g(t)$ is Gaussian, it achieves the minimum time-bandwidth product $\sigma_t \sigma_\omega = 0.5.$

Continuous prolate spheroidal wave function: a bandlimited signal that have the highest energy concentration in a specified time interval.

- To maintain a similar level of uncertainty, the window function should be short to attain time resolution for high frequency components and long to attain frequency resolution for low frequency components.

Wavelet Transform

- Built on set of expansion (or basis) functions that are derived from a mother wavelet through scaling and translation:

Let $\psi(t)$ be a mother wavelet : $\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right)$

Continuous Wavelet Transform:

$$F_w(a,\tau) = \int_{-\infty}^{\infty} f(t)\psi^*_{a,\tau}(t)dt = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi^*\left(\frac{t-\tau}{a}\right)dt = \frac{1}{\sqrt{a}} f(t) * \psi^*\left(-\frac{\tau}{a}\right)$$

$$F_w(a,\tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(at)\psi^*\left(t - \frac{\tau}{a}\right)dt$$ stretch or compress the signal for analysis

$|F_w(a,\tau)|^2 =$ scalogram; $|F(\omega,\tau)|^2 =$ spectrogram

Let $\Psi(\omega) = \mathrm{F}\{\psi(t)\}$ and $\Psi_{a,\tau}(\omega) = \mathrm{F}\{\psi_{a,\tau}(t)\}$

Inverse transform: $f(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_w(a,\tau)\psi_{a,\tau}(t) \frac{da\,d\tau}{a^2}$

Admissibility condition: $C_\psi = \int_0^\infty \frac{|\Psi_0(\omega)|^2}{\omega} d\omega < \infty$ ⟹ $\Psi(0) = 0$

Scaling and Translation

- For discrete wavelet transform,

$a = a_0^{-m}$ and $\tau = n\tau_0 a_0^{-m}$ m, n are integers

$\psi_{a,\tau}(t) = \psi_{m,n}(t) = a_0^{m/2}\psi(a_0^m t - n\tau_0)$ $m,n \in Z$

If $a_0 = 2$ and $\tau_0 = 1$, ← dyadic or octave sampling

$\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n)$ $m,n \in Z$

If the set $\{\psi_{m,n}(t)\}$ is complete, it is called affine wavelets.

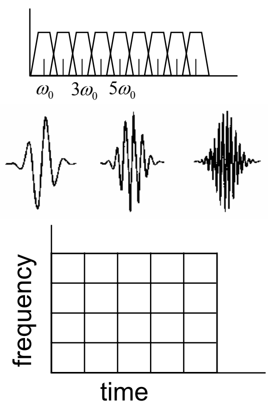
Orthogonality condition : $\int_{-\infty}^{\infty} \psi_{m,n}(t)\psi^*_{p,q}(t)dt = \delta(m-p)\delta(n-q)$

$$F_w(a,\tau) = w_{m,n} = a_0^{m/2}\int_{-\infty}^{\infty} f(t)\psi(a_0^m t - n\tau_0)dt$$

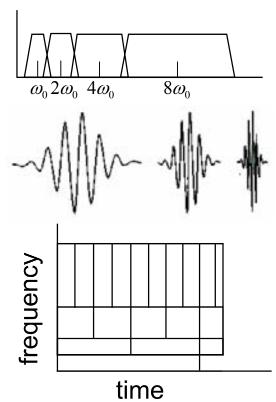
and $f(t) = \sum_m \sum_n w_{m,n}\psi_{m,n}(t)$

The orthogonality condition may not be easy to satisfy for a general set of "wavelets;" but it is still possible to invert discrete wavelet transform, sometimes.

STFT and CWT



Constant bandwidth



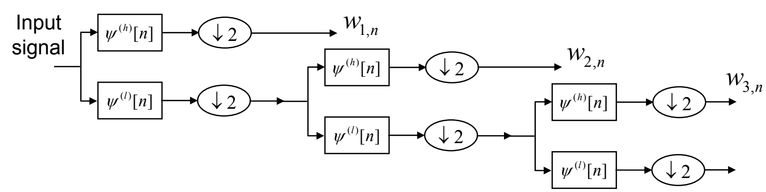
Constant-Q bandwidth

Iterative Filtering Implementation of DWT

$\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n)$ $m,n \in Z$

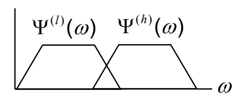
$\psi[n] \Leftrightarrow \psi(t)$ discrete-time sampled version

Let $\psi^{(h)}[n] = \psi[n]$ and $\psi^{(l)}[n] = (-1)^n\psi[-n+1]$.

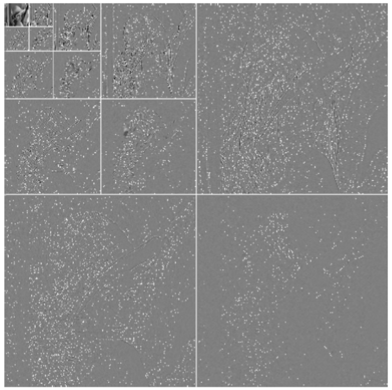


$\Psi(\omega) = \mathrm{F}\{\psi(t)\}$ and $\Psi_{a,\tau}(\omega) = \mathrm{F}\{\psi_{a,\tau}(t)\}$

$\Psi^{(h)}(\omega) = \mathrm{F}\{\psi^{(h)}[n]\}$ and $\Psi^{(l)}(\omega) = \mathrm{F}\{\psi^{(l)}[n]\}$



Multiresolution Analysis & Coding



13-band discrete wavelet transform