Sequences

$\{x_n\}_{n=-\infty}^{\infty}$ is a discrete sequence.

Can derive a sequence from a function $x(t)$ by having $x_n = x(t_s n) = x(\frac{n}{f_s})$.

Absolute summability: $\sum_{n=-\infty}^{\infty} |x_n| < \infty$

Square summability: $\sum_{n=-\infty}^{\infty} |x_n|^2 < \infty$ (aka energy signal)

Periodic: $\exists k > 0 : \forall n \in \mathbb{Z} : x_n = x_{n+k}$

$u_n = \begin{cases} 0 & n < 0 \\ 1 & n \geq 0 \end{cases}$ is the unit-step sequence

$\delta_n = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$ is the impulse sequence

Systems

A discrete system T transforms a sequence: $\{y_n\} = T\{x_n\}$.

Causal systems cannot look into the future: $y_n = f(x_n, x_{n-1}, x_{n-2}, \ldots)$

Memoryless systems depend only on the current input: $y_n = f(x_n)$

Delay systems shift sequences in time: $y_n = x_{n-d}$

A system T is time invariant if for all d: $\{y_n\} = T\{x_n\} \iff \{y_{n-d}\} = T\{x_{n-d}\}$

A system T is linear if for any sequences: $T\{ax_n + bx'_n\} = aT\{x_n\} + bT\{x'_n\}$

Causal linear time-invariant systems are of the form $\sum_{k=0}^{N} a_k y_{n-k} = \sum_{m=0}^{M} b_m x_{n-m}$, but all of them can be represented with only one set of coefficients, i.e. $b_i = \delta_i$. In this case the sequence $\{a_n\}$ is called the impulse response of T since $\{a_n\} = T\{\delta_n\}$.

Convolution

$\{p_n\} * \{q_n\} = \{r_n\} \iff \forall n \in \mathbb{Z} : r_n = \sum_{k=-\infty}^{\infty} p_k q_{n-k}$

Associativity: $(\{p_n\} * \{q_n\}) * \{r_n\} = \{p_n\} * (\{q_n\} * \{r_n\})$

Commutativity: $\{p_n\} * \{q_n\} = \{q_n\} * \{p_n\}$

Linearity: $\{p_n\} * \{aq_n + br_n\} = a(\{p_n\} * \{q_n\}) + b(\{p_n\} * \{r_n\})$

Identity: $\{p_n\} * \{\delta_n\} = \{p_n\}$

Shifting: $\{p_{n-d}\} = \{p_n\} * \{\delta_{n-d}\}$

Sine-wave and exponential sequences form a family of discrete sequences that is closed under convolution with arbitrary sequences.

Dirac's Delta Function

$\delta(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases}$, $\int_{-\infty}^{\infty} \delta(x) dx = 1$

Sampling: $\int_{-\infty}^{\infty} f(x)\delta(x-a)dx = f(a)$

Dirac comb: $s(t) = t_s \sum_{n=-\infty}^{\infty} \delta(t - t_s n)$, $\hat{x}(t) = x(t)s(t)$

Fourier Transform

Discrete Sejong's of the form $x_n = e^{j\omega n}$ are eigensequences with respect to a system T because for any ω, there is some $H(\omega)$ such that $T\{x_n\} = H(\omega)\{x_n\}$

$\mathcal{F}\{g(t)\}(\omega) = G(\omega) = \int_{-\infty}^{\infty} g(t)e^{\mp j\omega t}dt$

$\mathcal{F}^{-1}\{G(\omega)\}(t) = g(t) = \int_{-\infty}^{\infty} G(\omega)e^{\pm j\omega t}dt$

Linearity: $ax(t) + by(t) \rightleftharpoons aX(f) + bY(f)$

Time scaling: $x(at) \rightleftharpoons \frac{1}{|a|}X(\frac{f}{a})$

Frequency scaling: $\frac{1}{|a|}x(\frac{t}{a}) \rightleftharpoons X(af)$

Time shifting: $x(t - \Delta t) \rightleftharpoons X(f)e^{-2\pi jf\Delta t}$

Frequency shifting: $x(t)e^{2\pi j\Delta ft} \rightleftharpoons X(f - \Delta f)$

Parseval's theorem: $\int_{\infty}^{\infty} |x(t)|^2 dt = \int_{-\infty}^{\infty} |X(f)|^2 df$

Continuous convolution: $\mathcal{F}\{(f * g)(t)\} = \mathcal{F}\{f(t)\}\mathcal{F}\{g(t)\}$

Discrete convolution: $\{x_n\} * \{y_n\} = \{z_n\} \iff X(e^{j\omega})Y(e^{j\omega}) = Z(e^{j\omega})$

Sample Transforms

The Fourier transform preserves function even/oddness but flips real/imaginaryness iff $x(t)$ is odd.

$\mathcal{F}\{cos(2\pi f_0 t)\}(f) = \frac{1}{2}\delta(f - f_0) + \frac{1}{2}\delta(f + f_0)$

$\mathcal{F}\{sin(2\pi f_0 t)\}(f) = -\frac{j}{2}\delta(f - f_0) + \frac{j}{2}\delta(f + f_0)$

$\mathcal{F}\{t_s \sum_{n=-\infty}^{\infty} \delta(t - t_s n)\}(f) = \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{t_s})$

Aliasing

Due to overlapping in the frequency domain, the Nyquist limit stipulates that frequencies in the signal should be limited so that $|f| \leq \frac{f_s}{2}$. Bandpass sampled signals can also be reconstructed if their spectral component lie within the interval $n\frac{f_s}{2} < |f| < (n+1)\frac{f_s}{2}$

An ideal filter for enforcing the Nyquist criterion is $rect(t_s f) = \begin{cases} 1 & |f| < \frac{f_s}{2} \\ 0 & |f| > \frac{f_s}{2} \end{cases}$ with$\mathcal{F}^{-1}\{rect(f)\}(t) = f_s \frac{\sin \pi t f_s}{\pi t f_s}$.

Typically the sampling frequency is chosen so there is a transition band between the Nyquist limit and the highest frequency expected to be sampled.

Discrete Fourier Transform

$X_k = \sum_{i=0}^{n-1} x_i e^{-2\pi j \frac{ik}{n}}$

$x_k = \frac{1}{n}\sum_{i=0}^{n-1} X_i e^{2\pi j \frac{ik}{n}}$

The elements $X_0$to $X_{\frac{n}{2}}$contain the frequency components 0, $\frac{f_s}{n}$, $\frac{2f_s}{n}$, …, $\frac{f_s}{2}$

We can make use of the fast Fourier transform: $F_n\{x_i\}_{i=0}^{n-1} = \sum_{i=0}^{n-1} x_i e^{-2\pi j \frac{ik}{n}}$, so $= \sum_{i=0}^{\frac{n}{2}-1} x_{2i} e^{-2\pi j \frac{2ik}{n}} + e^{-2\pi j \frac{k}{n}} \sum_{i=0}^{\frac{n}{2}-1} x_{2i+1} e^{-2\pi j \frac{2ik}{n}}$. This repeated subdivision has $\log_2 n$ rounds and $n \log_2 n$ additions and multiplications overall, compared to n^2 for the equivalent matrix multiplication.

By the symmetry properties of the FFT, $\forall i : x_i = \mathcal{R}(x_i) \iff \forall i : X_{n-i} = X_i^\star$ and $\forall i : x_i = j\mathcal{I}(x_i) \iff \forall i : X_{n-i} = -X_i^\star$. We can therefore compute the FFT of two real valued sequences by computing that of $x_i' + jx_i''$ and saying $X_i' = \frac{1}{2}(X_i + X_{n-1}^\star)$ and $X_i' = \frac{1}{2}(X_i - X_{n-1}^\star)$.

We can also compute complex multiplication with 3 multiplications and 5 additions rather than 4 multiplications and 2 additions that are required naively if we have the numbers in Cartesian format.

Application to Convolution

We can compute convolution using FFT with $O(m \log m + n \log n)$ rather than mn multiplications if we multiply in the Fourier domain instead. Note however that the sequences being convolved must be periodic with equal period lengths. If this condition is not fulfilled the sequences must be zero-padded to a length of at least $m + n - 1$ to ensure the start and end of the resulting sequence do not overlap.

We can use this result to perform deconvolution by dividing the Fourier representation: $u = \mathcal{F}^{-1}\{\frac{\mathcal{F}\{s\}}{\mathcal{F}\{h\}}\}$

Spectral Estimation

If the input to the DFT is sampled but not periodic, the DFT may still be used to calculate an approximation. However, we see *leakage* of energy to frequency bins that are not strictly present in the input signal but adjacent to expressed frequencies. The peak amplitude also changes as the frequency of a tone changes from one bin to the next, reaching its lowest amplitude of 62% halfway between the two. This is known as *scalloping*. It occurs since the effective window introduced by the DFT is rectangular, which corresponds to convolution with *sinc* in the frequency domain.

A non-periodic signal can have a windowing function applied to try and reduce scalloping and leakage. Possibilities include:

Triangular: $w_i = 1 - \left|1 - \frac{2i}{n}\right|$

Hanning: $w_i = 0.5 - 0.5\cos(2\pi \frac{i}{n-1})$

Hamming: $w_i = 0.54 - 0.46\cos(2\pi \frac{i}{n-1})$

Since a windowed signal has been forced to 0 outside the sampled region, it may be zero-padded at either end without further distortion. However, that does increase the frequency resolution of the DFT (albeit without adding any new information!).

Filter Design

We can turn a spectrum $X(f)$ into an inverted spectrum $X'(f) = X(\frac{f_s}{2} - f)$ we shift the spectrum by $\frac{f_s}{2}$. This can be done by multiplying with the sequence with $y_i = \cos(\pi i)$. This can turn a low-pass filter into a high-pass filter.

The ideal low-pass filter has two problems in that it is not causal and has an infinitely long impulse response. This is solved by applying a windowing function and a delay to make it causal, such as:

$h_i = 2\frac{f_c}{f_s}\frac{sin(2\pi(i-\frac{n}{2})\frac{f_c}{f_s})}{2\pi(i-\frac{n}{2})\frac{f_c}{f_s}} w_i$ for an nth order low pass filter.

Z-Transform

If $X(z) = \sum_{n=-\infty}^{\infty} x_n z^{-n}$, which defines a complex-valued surface over $\mathbb{C}$. For finite sequences, this surface is defined everywhere, else it converges in the region $\lim_{n\to\infty}\left|\frac{x_{n+1}}{x_n}\right| < |z| < \lim_{n\to-\infty}\left|\frac{x_{n+1}}{x_n}\right|$.

For an LTI system defined by $\sum_{l=0}^{k} a_l y_{n-l} = \sum_{l=0}^{m} b_l x_{n-l}$, $H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_m z^{-m}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_k z^{-k}}$ so that $\{y_n\} = \{h_n\} * \{x_n\}$. This has m zeroes and k poles at non-zero locations in the z plane, plus $k - m$ zeroes (if $k > m$) or $m - k$ poles (if $m > k$) at $z = 0$.

The *order* of a filter is the number of zeroes it has. Possible filter types are Butterworth, Chebyshev Type I/Type II and Elliptic.

Random Sequences

In such a sequence every value is the outcome of a one random variable from a corresponding sequence. This collection of random variables is called a *random process*. Such a process is stationary if $P_{x_{n_1+l},\ldots,x_{n_k+l}}(a_1,\ldots,a_k) = P_{x_{n_1},\ldots,x_{n_k}}(a_1,\ldots,a_k)$ where $P_{x_n}(a) = Prob(x_n \le a)$.

Expected value: $m_x = \varepsilon(x_n) = \int a p_{x_n}(a) da$

Variance: $Var(x_n) = \varepsilon(|x_n - \varepsilon(x_n)|^2) = \varepsilon(|x_n|^2) - |\varepsilon(x_n)|^2$

Correlation: $Cor(x_n, x_m) = \varepsilon(x_n x_m^\star) = \phi_{xx}(0)$

Cross-correlation: $\phi_{xy}(k) = \varepsilon(x_{n+k} y_n^\star)$, Autocorrelation: $\phi_{xx}(k)$

Cross-covariance: $\gamma_{xy}(k) = \varepsilon[(x_{n+k} - m_x)(y_n - m_y)^\star] = \phi_{xy}(k) - m_x m_y^\star$, Autocovariance: $\gamma_{xx}(k)$

Deterministic cross-correlation: $c_{xy}(k) = \sum_{i=-\infty}^{\infty} x_{i+k} y_i$, so that $\{c_{xy}(k)\} = \{x_k\} * \{y_{-k}\}$. This implies that the Fourier transform $C_{xx}(f)$ is identical to the power spectrum.

If $\{y_n\} = \{h_n\} * \{x_n\}$ then $m_y = m_x \sum_{k=-\infty}^{\infty} h_k$, $\{\phi_{yy}(n)\} = \{c_{hh}(n)\} * \{\phi_{xx}(n)\}$ and $\{\phi_{yx}(n)\} = \{h_n\} * \{\phi_{xx}(n)\}$

White noise is characterized by $m_x = 0$ and $\phi_{xx}(k) = \sigma_x^2 \delta_k$

A DFT can be averaged by cutting a signal into a number of windows, taking the DFTs of these and plotting the average of their absolute values: this is known as incoherent averaging. Coherent averaging is done by taking the absolute value after averaging the complex numbers, which suppresses anything which is not a periodic waveform with a period that divides the number of elements per window.

Compression

A compression system will typically involve several changes:

1. Transducer: converts input into voltage
2. A-to-D converter: samples and quantizes the signal
3. Transformation into a perceptual domain
4. Quantization based on perceptual model
5. Decorrelation transform to reduce entropy
6. Entropy coding for transmission

Coding

Coding techniques which can be used include Huffman (iterated low-probability tree building), arithmetic (successive weighted numeric interval refinement) and run-length. There is also a class of predictor codings, where only the difference between the prediction and the reality is transmitted. Simple cases are delta coding ($P(x) = x$) and linear predictive coding ($P(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i x_i$).

Decorrelation

If $P(X = x \wedge Y = y) \neq P(X = x)P(Y = y)$ for any x and y then $H(X|Y) < H(X) \wedge H(Y|X) < H(Y)$. This cannot be exploited practically since there are too many conditional probabilities. However, we can approximate it with correlation, which captures most dependence relationships.

The Pearson correlation coefficient: $\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

The covariance matrix: $Cov(\vec{X}) = (Cov(X_i, X_j))_{i,j}$ for $\vec{X} \in \mathbb{R}^n$. Note that $Cov(\vec{X}) = Cov^T(\vec{X})$

If $\vec{X} \in \mathbb{R}^n$ and $\vec{Y} \in \mathbb{R}^n$ so that $\vec{Y} = A\vec{X} + b$ with $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$ then $E(\vec{Y}) = AE(\vec{X}) + b$ and $Cov(\vec{Y}) = ACov(\vec{X})A^T$

Recall that $\vec{x}$ is an eigenvector if for some $\lambda \in \mathbb{R}$ $A\vec{x} = \lambda x$. Now any symmetric matrix A can be represented as $U \Lambda U^T$ where $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ of increasing λ and the columns of U are the corresponding orthonormal eigenvectors.

The Karhunen-Loeve transform finds a matrix A so that $Cov(AX) = ACov(X)A^T$ is diagonalized. It can be shown that since that $UU^T = I$ the U^T from above is just such an A. This can be applied straightforwardly to decorrelate color planes, but for spatial decorrelation we work on a matrix where the columns are entire monochrome images!

It turns out that in practice (for most sample images used to calculate decorrelation) the eigenvector matrix is almost indistinguishable from that of the Discrete Cosine Transform:

$S(u) = \frac{C(u)}{\sqrt{N/2}} \sum_{x=0}^{N-1} s(x) \cos \frac{(2x+1)u\pi}{2N}$

$s(u) = \sum_{x=0}^{N-1} \frac{C(u)}{\sqrt{N/2}} S(u) \cos \frac{(2x+1)u\pi}{2N}$

Where $C(u) = \begin{cases} \frac{1}{\sqrt{2}} & u = 0 \\ 1 & u > 0 \end{cases}$, ensuring orthogonality.

The Discrete Wavelet Transform ensures that the bandwidth of each output signal is proportional to the highest input frequency that it contains. It is defined by the combination of a low pass and high pass filter. For an n-point DFT the vector is convolved separately with a low pass and a high pass filter. The results each have n numbers, but as the resolution of each has been halved, half those numbers can be discarded. The output values of the high pass filter are retained and the low pass information is recursively treated the same way until only a single value remains (the average).

Psychophysics

Weber's law gives the difference limit (smallest stimulus perceivable at an intensity level): $\Delta\phi = c(\phi + a)$

Fechner's scale gives defines a perception intensity scale with the sensation limit of ϕ_0 as the origin and the respective different limit $\Delta\phi$ as a unit step: $\psi = \log_c \frac{\phi}{\phi_0}$

Stevens' law is like Fechner's scale, but rational so that it reflects subjective relations: $\psi = k(\phi - \phi_0)^a$

The human eye processes color and luminosity at different resolutions. To exploit this, some TV signals are transmitted as luminance and chrominance channels (YUV) rather than RGB. Typically $Y = 0.3R + 0.6G + 0.1B$, $V = R - Y$ and $U = B - Y$. There also exist normalized channels $Cb = \frac{U}{2.0} + 0.5$ and $Cr = \frac{V}{1.6} + 0.5$

A pair of pure tones cannot be distinguished as two frequencies if both are in the same critical band. The human ear has 24 critical bands, which are expressed on the Bark scale. This can be mapped to from frequency by the approximation $b \approx \frac{26.81}{1 + \frac{1960Hz}{f}} - 0.53$

Louder tones increase the sensation limit for nearby critical bands asymmetrically. This extends both backwards and forwards from the time the masking signal is introduced, peaking at its edges.

Due to these sensation limit effects, non-uniform quantization can reduce perceived quantization noise. Typically this will be done with a logarithmic scale, of which μ-law and A-law are the competing flavors.

JPEG

The JPEG algorithm proceeds as follows:

1. 8 bit RGB input image transformed into 8 bit YCrCb
2. Chrominance channel resolution reduced by a factor of 2
3. For each channel:
   (a) Split channel into 8×8 blocks
   (b) Perform forward DCT on each block
   (c) Quantize DCT coefficients by perceptual matrix
   (d) Apply delta coding to coefficients
   (e) Read remaining values from DCT in a zigzag
   (f) Apply run-length coding to zeroes
   (g) Apply Huffman coding

MPEG

This is a video coding scheme very similar to JPEG. However, it adds:

- Spatially scalable coding (for progressive rendering of a kind)
- Predictive coding with motion compensation
- Interframe coding with I (independent), P (differences relative to previous frame) and B (interpolation between neighboring) frames