Digital Signal Processing

Notes on Spoken Language Processing

Chia-Ping Chen

Department of Computer Science and Engineering
National Sun Yat-Sen University
Kaohsiung, Taiwan ROC

Introduction

- Digital signal processing (DSP) plays a fundamental role in spoken language processing.
- We need to process the speech “signal”, i.e. acoustic waveform, to a convenient representation for subsequent usage.
- The topics in DSP include digital signals, systems, filters, spectral-domain transforms, and so on.
- Like every subject we have treated so far, we are condensing a volume into a chapter: YOU NEED TO READ RELATED MATERIALS.

Discrete-Time Signals

- An analog signal $x_a(t)$ can be seen as a function varying continuously in time.
- A discrete-time signal $x[n]$ is a sequence of numbers indexed by integers.
- A discrete-time signal can be obtained by sampling a continuous-time signal at periodic time stamps,

$$x[n] = x_a(nT).$$

- T is called sampling period. $1/T$ is called sampling frequency.
- In DSP we manipulate $x[n]$ in digital processors.

Common Signals I

- sinusoidal

$$x[n] = A\cos(\omega n + \phi).$$

- impulse

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & \text{otherwise} \end{cases}$$

- step

$$u[n] = \begin{cases} 1, & n \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Common Signals II

- rectangular

$$\text{rect}_N[n] = \begin{cases} 1, & 0 \leq n < N \\ 0, & \text{otherwise} \end{cases}$$

- real exponential

$$x[n] = a^n u[n]$$

- complex exponential

$$x[n] = a^n u[n] = r^n e^{i\omega n} u[n] = r^n (\cos \omega n + j \sin \omega n) u[n]$$

Discrete-Time Systems

- A discrete-time system, given an input signal $x[n]$, outputs a signal $y[n]$

$$y[n] = T\{x[n]\}.$$

- Note that both $x[n]$ and $y[n]$ are discrete-time.

Linear and Time-Invariant

- A system is linear if

$$T\{x_1[n] + x_2[n]\} = T\{x_1[n]\} + T\{x_2[n]\}$$

$$T\{ax[n]\} = aT\{x[n]\}$$

- A system is time-invariant if

$$T\{x[n]\} = y[n] \Rightarrow T\{x[n - n_0]\} = y[n - n_0].$$

- A linear time-invariant (LTI) system is both linear and time-invariant.

Impulse Response

- The impulse response function of a system is the output signal when the input signal is $\delta[n]$,

$$h[n] = T\{\delta[n]\}.$$

Convolution

- The convolution of two discrete-time signals is defined by

$$x[n] * h[n] = \sum_{m=-\infty}^{\infty} x[m]h[n-m].$$

- The operation of convolution is commutative, associative and distributive.

Theorem

- (theorem) The output of an LTI system is the convolution of the input signal and the system's impulse response.
- (proof) A signal $x[n]$ can be written as a sum of impulse signals

$$x[n] = \sum_{m=-\infty}^{\infty} x[m]\delta[n-m]$$

$$\Rightarrow \; y[n] = T\{x[n]\} = T\left\{\sum_m x[m]\delta[n-m]\right\}$$

$$= \sum_m x[m]T\{\delta[n-m]\} = \sum_m x[m]h[n-m].$$

Eigenvector of LTI System

- If we input a complex exponential $x[n] = e^{j\omega n}$ to an LTI system with impulse response $h[n]$, the output is

$$y[n] = \sum_m h[m] e^{j\omega(n-m)} = e^{j\omega n} H(e^{j\omega}) = x[n] H(e^{j\omega}).$$

- $x[n] = e^{j\omega n}$ is an eigenvector of an LTI system with eigenvalue $H(e^{j\omega})$, and

$$H(e^{j\omega}) = \sum_m h[m] e^{-j\omega m}.$$

Fourier Transform

- $H(e^{j\omega})$ is called the Fourier transform of $h[n]$.
- Generally, the Fourier transform of a discrete-time signal $x[n]$ is defined by

$$X(e^{j\omega}) = \sum_n x[n]e^{-j\omega n}.$$

- It is a periodic function of ω, with period 2π: we only need the values in a period to characterize a Fourier transform.

Frequency Response

- $H(e^{j\omega})$ is also called the frequency response of the system.
- Suppose the input signal is composed of sinusoids of different frequencies.
- filtering effects: the components with frequencies of large $|H(e^{j\omega})|$ are amplified; the components with frequencies of small $|H(e^{j\omega})|$ are attenuated.
- Note that the normalized frequency $f = \frac{\omega}{2\pi}$ and the linear frequency f_l are related by

$$f_l = f F_s, \quad F_s = \text{sampling frequency}.$$

Power Spectrum

- The squared amplitude of Fourier transform is the power spectrum

$$S(e^{j\omega}) = |X(e^{j\omega})|^2$$

- It is the Fourier transform of the auto-correlation of signal $x[n]$ defined by

$$R_{xx}[n] = \sum_m x[m+n]x[m].$$

Inverse Fourier Transform

- The inverse Discrete-Time Fourier transform is given by

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega.$$

- $x[n]$ can be seen as being "synthesized" by exponential signal components.
- One can see that $X(e^{j\omega})$ is the spectrum of $x[n]$.

z-Transform

- The z-transform of a discrete-time signal $x[n]$ is defined by

$$X(z) = \sum_{n=-\infty}^{\infty} x[n] z^{-n}.$$

- Note that the Fourier transform is a special case of the z-transform: on the unit circle of the complex z-plane.

Inverse z-Transform

- The inverse z-transform is

$$x[n] = \frac{1}{2\pi j} \oint X(z) z^{n-1} dz.$$

- This equality can be proved by the residue theorem of complex analysis.

Region of Convergence

- The region in the complex z-plane where $X(z)$ is defined, i.e. the sum converges, is called the region of convergence, a.k.a. ROC, for $x[n]$.
- In stating the z-transform, ROC is crucial.

Examples of z-Transform

- delayed impulse

$$h[n] = \delta[n - N] \overset{z}{\Rightarrow} H(z) = z^{-N},\ z \neq 0$$

- rectangular

$$h[n] = u[n] - u[n - N] \overset{z}{\Rightarrow} H(z) = \frac{1 - z^{-N}}{1 - z^{-1}},\ z \neq 0$$

- exponential

$$h_3[n] = a^n u[n] \overset{z}{\Rightarrow} H_3(z) = \frac{1}{1 - az^{-1}},\ |z| > |a|$$

Linear Difference Equations

- An LTI system can also be characterized by a linear difference equation

$$\sum_{k=0}^{N} a_k y[n-k] = \sum_{k=0}^{M} b_k x[n-k].$$

- Taking the z-transform, one has

$$\sum_{k=0}^{N} a_k z^{-k} Y(z) = \sum_{k=0}^{M} b_k z^{-k} X(z)$$

$$\Rightarrow H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}}$$

Zeros and Poles

- A pole of a function is where the value of the function is singular.
- A zero of a function is where the value of the function is 0.
- For example, $z = a$ is a pole for $H_3[z]$.
- When a z-transform is expressed as the ratio of two polynomials, as in linear difference equations, the roots of the numerator are zeros, and the roots of the denominator are poles.

Causal and Stable

- A system is *causal* if the impulse response satisfies

$$h[n] = 0, \quad \forall n < 0.$$

- A system is *stable* if bounded inputs produces bounded outputs, a.k.a. BIBO. Equivalently,

$$\sum_n |h[n]| < \infty.$$

- If a system is causal and stable, then all its poles must be inside the unit circle.
  - stable $\Rightarrow$ unit circle in ROC.
  - causal $\Rightarrow$ ROC extends to infinity.

Convolution Theorem

- (theorem)

$$y[n] = x[n] * h[n] \;\Rightarrow\; Y(z) = X(z)H(z).$$

- (proof)

$$\begin{aligned} Y(z) &= \sum_n y[n]z^{-n} = \sum_n \sum_m x[m]h[n-m]z^{-n} \\ &= \sum_n \sum_m x[m]h[n-m]z^{-(n-m)}z^{-m} \\ &= \sum_m x[m]z^{-m} \sum_n h[n-m]z^{-(n-m)} \\ &= X(z)H(z). \end{aligned}$$

Discrete Fourier Transform

- The discrete Fourier transform (DFT) of a sequence of finite duration N or periodic with period N, is defined by

$$X[k] = \sum_{n=0}^{N-1} x[n]W^{nk},\ 0 \leq k < N$$

  where $W = e^{-j(2\pi/N)}$.

- If $x[n]$ is of finite duration N, then $X[k]$'s are exactly the N equally spaced samples at points $\omega_k = \frac{2\pi k}{N}$ of the Fourier transform of $x[n]$.

Inverse Discrete Fourier Transform

- The inverse discrete Fourier transform (IDFT) is

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] W^{-kn},\ 0 \leq n < N.$$

- The above equality can be proved with the help of the following equality

$$\sum_{k=0}^{N-1} W^{n'k} W^{-nk} = N \delta_{n'n}.$$

- Note that $x[n]$ is periodic.

FT of Periodic Signals

- A periodic signal has infinite energy, and the Fourier transform does not converge.
- One can make use of the Dirac delta function to establish the following transformation pairs.
  - complex exponential

$$e^{j\omega_0 n} \leftrightarrow 2\pi\delta(\omega - \omega_0)$$

  - impulse train

$$p_N[n] = \sum_k \delta[n - kN] \leftrightarrow \frac{2\pi}{N} \sum_{k=0}^{N-1} \delta(\omega - 2\pi k/N)$$

General Periodic Signals

- For a general periodic signals $x_N[n]$, we define

$$x[n] = \begin{cases} x_N[n], & 0 \le n < N \\ 0, & \text{otherwise} \end{cases}$$

- Clearly,

$$x_N[n] = \sum_k x[n - kN] = x[n] * \sum_k \delta[n - kN] = x[n] * p_N[n].$$

- Using the convolution theorem, one has

$$X_N(e^{j\omega}) = \frac{2\pi}{N} \sum_k X(e^{j2\pi k/N}) \delta(\omega - 2\pi k/N).$$

Radix-2 Fast Fourier Transform

- We can express an N-point DFT of $x[n]$ by

$$X[k] = F[k] + W_N^k G[k],$$

where $F[k]$ is the $N/2$-point DFT of $f[n] = x[2n]$ and $G[k]$ is the $N/2$-point DFT of $g[n] = x[2n+1]$.

- Recursively apply the same idea until the 2-point DFTs are to be computed by

$$X[0] = x[0] + x[1]; X[1] = x[0] - x[1].$$

- The time complexity of FFT is $N \log N$.

Circular Convolution

- The circular convolution of two signals $x_1[n]$ and $x_2[n]$ is defined by

$$x_1[n] \otimes x_2[n] = \sum_{m=0}^{N-1} \tilde{x}_1[m]\tilde{x}_2[n-m].$$

where $\tilde{x}_i[n]$ is the periodic extension of $x_i[n]$, i.e.,

$$\tilde{x}[n] = x[n \,\%\, N].$$

- Note if $x[n]$ is periodic, then

$$x[n] = \tilde{x}[n].$$

Convolution Theorem with DFT

$$y[n] = x_1[n] \otimes x_2[n] \;\Rightarrow\; Y[k] = X_1[k]X_2[k].$$

$$\begin{aligned}
\sum_{n=0}^{N-1} y[n]W^{nk} &= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} \tilde{x}_1[m]\tilde{x}_2[n-m]W^{nk} = \sum_{m=0}^{N-1} \tilde{x}_1[m] \sum_{n=0}^{N-1} \tilde{x}_2[n-m]W^{nk} \\
&= \sum_{m=0}^{N-1} \tilde{x}_1[m]W^{km} \sum_{n=0}^{N-1} \tilde{x}_2[n-m]W^{k(n-m)} \\
&= \cdots \sum_{r=-m}^{N-1-m} \tilde{x}_2[r]W^{kr} = \ldots \Big(\sum_{r=-m}^{-1} + \sum_{r=0}^{N-1-m} \Big) \tilde{x}_2[r]W^{kr} \\
&= \ldots \Big(\sum_{r=N-m}^{N-1} + \sum_{r=0}^{N-1-m} \Big) \tilde{x}_2[r]W^{kr} = \cdots \sum_{r=0}^{N-1} x_2[r]W^{kr} \\
&= \sum_{m=0}^{N-1} \tilde{x}_1[m]W^{km}X_2[k] = \sum_{m=0}^{N-1} x_1[m]W^{km}X_2[k] \\
&= X_1[k]X_2[k].
\end{aligned}$$

Discrete Cosine Transform

- DCT-II is defined by

$$C[k] = \sum_{n=0}^{N-1} x[n] \cos(\pi k(n + \frac{1}{2})/N),\ 0 \leq k < N.$$

- The inverse is given by

$$x[n] = \frac{1}{N} \left\{ C[0] + 2 \sum_{k=1}^{N-1} C[k] \cos(\pi k(n + \frac{1}{2})/N) \right\},\ 0 \leq n < N.$$

- N-point DCT can be computed by $2N$-point FFT.
- DCT is widely used in speech processing due to its energy compaction.

Filters

- In signal processing we often want to extract some components out of the signal.
- The filters are used to get rid of unwanted part and retain or even enhance the wanted part.
- A filter is basically a system, often an LTI one.
- Filters can be characterized in the frequency domain or in the time domain.

Ideal Low-Pass Filters

- frequency response

$$H(e^{j\omega}) = \begin{cases} 1, & |\omega| < \omega_0 \\ 0, & \omega_0 < |\omega| < \pi \end{cases}$$

- impulse response

$$h[n] = \frac{1}{2\pi}\int_{-\omega_0}^{\omega_0} e^{j\omega n} d\omega = \frac{\sin \omega_0 n}{\pi n} = \frac{\omega_0}{\pi}\text{sinc}(\omega_0 n/\pi)$$

- $h[n]$ is non-causal and has infinite duration.

Rectangular Window

- impulse response

$$h_\pi[n] = u[n] - u[n-N]$$

- frequency response

$$\begin{aligned} H_\pi(e^{i\omega}) &= \sum_{n=0}^{N-1} e^{-j\omega n} = \frac{1-e^{-j\omega N}}{1-e^{-j\omega}} \\ &= \frac{\sin \omega N/2}{\sin \omega/2} e^{-j\omega(N-1)/2} \\ &= A(\omega) e^{-j\omega(N-1)/2} \end{aligned}$$

$A(\omega)$ is 0 when $\omega N = 2k\pi$.

Generalized Hamming Window

- impulse response

$$h[n] = \begin{cases} (1-\alpha) - \alpha\cos(2\pi n/N), & 0 \le n < N \\ 0, & \text{otherwise} \end{cases}$$

Equivalently,

$$h[n] = (1-\alpha)h_\pi[n] + \alpha h_\pi[n]\cos(2\pi n/N)$$

- frequency response

$$H(e^{i\omega}) = (1-\alpha)H_\pi(e^{j\omega}) - \frac{\alpha}{2}H_\pi(e^{j(\omega-2\pi/N)}) - \frac{\alpha}{2}H_\pi(e^{j(\omega+2\pi/N)})$$

Finite Impulse Response

- If the impulse response of a filter has a finite duration, then the filter is called a finite-impulse response (FIR) filter.
- For example, the rectangular window and the generalized Hamming window are FIR filters.
- FIR filters are always stable, as there is only a finite number of terms in the sum of convolution,

$$y[n] = \sum_{m} x[n-m]h[m] = \sum_{m=M_1}^{M_2} x[n-m]h[m].$$

Infinite Impulse Response

- If the impulse response of a filter has an infinite duration, then the filter is called an infinite-impulse response (IIR) filter.
- For example, the ideal low-pass filter is an IIR filter.

$$h[n] = \frac{\omega_0}{\pi}\text{sinc}(\omega_0 n/\pi)$$

- It is not true that any IIR filter cannot be realized. In fact, a linear difference equation where the output signal has a regression is equivalent to an IIR filter.

Sampling Theorem

- A continuous-time signal $x(t)$ can be recovered from its discrete-time samples $x[nT]$ if the sampling rate is no less than twice the bandwidth $x(t)$.
- Suppose we sample $x(t)$ with the Diac delta function $p(t) = \sum_n \delta(t - nT)$. Then

$$X_p(\Omega) = \frac{1}{2\pi} X(\Omega) \ * \ P(\Omega) = \frac{1}{T} \sum_k X(\Omega - k\Omega_s),$$

where $\Omega_s = 2\pi/T$. There is no overlapping in $X_p(\Omega)$ if $X(\Omega) = 0$ for $\Omega > \Omega_s/2$, and by extracting one period of $X_p(\Omega)$ one gets $X(\Omega)$ to reconstruct $x(t)$.

Stochastic Processes

- To describe real-world signals, such as noises, we need to use the stochastic processes for their unpredictable nature.
- In a stochastic process, $\mathbf{x}[n]$ is a random variable for every n.
- The quantities that concern us most would be the auto-correlation function $R_{xx}[n_1, n_2]$ between the random variables $x[n_1], x[n_2]$ or the auto-correlation coefficients $r_{xx}[n_1, n_2]$.

Stationary Processes

- We focus on the stationary processes as the most general stochastic process is difficult to characterize.
- In a wide-sense stationary (WSS) process, we only require the mean and the auto-correlation to be invariant with respect to the time origin.
- In a strict-sense stationary (SSS) process, the probability distribution is invariant with respect to the time origin.
- In stationary processes, the auto-correlation (coefficients) reduce to

$$R_{xx}[n_1, n_2] \rightarrow R_{xx}[n_1 - n_2], \;\; r_{xx}[n_1, n_2] \rightarrow r_{xx}[n_1 - n_2]$$

Ergodic Processes

- ensemble average vs. time average
- The ensemble average is over all sample paths, while the time average only needs one sample path.
- Ergodicity is defined by the equality of the two types of averages.
- It allows us to compute mean and covariance of a random process by the time average.

LTI System with Stochastic Inputs

- Consider a WSS signal $\mathbf{x}[n]$ and an LTI system with impulse response $h[n]$.
- The system output is $y[n] = \sum x[n-m]h[m]$, and

$$\mu_y[n] = E\{y[n]\} = E\{\sum_m x[n-m]h[m]\} = \mu_x \sum_m h[m]$$

$$R_{xy}[l] = E\{x[n+l]y[n]\} = E\{x[n+l]\sum_m x[n-m]h[m]\}$$

$$= \sum_m h[m]R_{xx}[l+m] = h[-l] \ * \ R_{xx}[l]$$

$$R_{yy}[l] = E\{y[n+l]y[n]\} = E\{\sum_m h[m]x[n+l-m]y[n]\}$$

$$= \sum_m h[m]R_{xy}[l-m] = h[l] \ * \ R_{xy}[l] = h[l] \ * \ h[-l] \ * \ R_{xx}[l]$$

Power Spectral Density

- The power spectral density $S_{xx}(\omega) = |X(\omega)^2|$ of a WSS $\mathbf{x}[n]$ is the expectation value of the squared magnitude of the Fourier transform of $\mathbf{x}[n]$.
- It turns out that $S_{xx}(\omega)$ is the Fourier transform of the auto-correlation function $R_{xx}[n]$

$$E\{X(\omega+u)X^*(\omega)\} = E\{\sum_m x[m]e^{-j(\omega+u)m} \sum_n x[n]e^{j\omega n}\}$$

$$\Rightarrow E\{X(\omega)X^*(\omega)\} = \sum_l e^{-j\omega l} R_{xx}[l] \quad (\text{use } l = m-n)$$

Noises

- $\mathbf{x}[n]$ is a white noise if $\mu_x[n] = 0$ and

$$R_{xx}[n_1, n_2] = R[n_1]\delta[n_1 - n_2].$$

- If $\mathbf{x}[n]$ is WSS, then

$$R_{xx}[n] = q\delta[n],$$

and

$$S_{xx}(\omega) = q = const.$$