Elements of Digital Signal Processing

Notes on Speech and Audio Processing

Chia-Ping Chen

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

Introduction

- The theoretical foundation for automatic speech recognition is digital signal processing and statistical pattern recognition.
  - The digital signal processing is used to extract features from the speech signal.
  - The statistical pattern recognition is used to recognize the linguistic patterns hidden in the speech features.
- We will study the digital signal processing first.

Discrete-Time Signals

- A discrete-time signal is a sequence of numbers (amplitudes) indexed by the integers.
- More often than not, a discrete-time signal is obtained by sampling a continuous-time signal at periodic time stamps,

$$x[n] = x_c(nT),$$

where T is the sampling period.

Discrete-Time Systems

- A discrete-time system transforms an input discrete-time signal to an output discrete-time signal. This transformation can be denoted by

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

- 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

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

Impulse Response

- The discrete-time impulse sequence is defined by

$$\delta[n] = \delta_{n0}.$$

- Any sequence can be written as

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

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

z-Transform

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

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

  when the sum on the right side converges. Here z is a complex variable.

- The region where $X(z)$ is defined is called the region of convergence (ROC) for $x[n]$.

Inverse z-Transform

- From the residue theorem of complex integral, we have
$$x[n] = \frac{1}{2\pi j} \oint X(z) z^{n-1} dz.$$
The above equation is also known as the inverse z-transform.
- We can see $x[n]$ being “synthesized” by $X(z)$.

Convolution

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

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

- For an LTI system with impulse response $h[n]$, the output signal for input $x[n]$ is $x[n] * h[n]$.

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

Convolution Theorem

- If $y[n] = x[n] * h[n]$, then $Y(z) = X(z)H(z)$.
- Proof:

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

Sinusoidal Signal

- We will introduce the Fourier transform. We first introduce a few preliminaries.
- The sinusoidal signal is defined by

$$x[n] = e^{j\omega n}.$$

- If we input a sinusoidal signal to an LTI system, the output is

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

- $e^{j\omega n}$ is an eigenvector of an LTI system with eigenvalue $H(e^{j\omega})$.

The Fourier Transform

- 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 the z-transform when z is on the unit circle. That is,

$$z = e^{j\omega}, \quad |z| = 1.$$

The Inverse Fourier Transform

- The inverse Fourier transform is given by

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

  which follows from the inverse z-transform with the unit circle being the contour of integration.
- From the above formula, $x[n]$ can be seen as being "synthesized" by sinusoidal signals. The amplitude for the sinusoidal signal with frequency ω is $X(e^{j\omega})d\omega$. Put in another way, $X(e^{j\omega})$ is the spectrum of $x[n]$.

Spectral Domain Interpretation

- From the convolution theorem, when z is on the unit circle, we have

$$Y(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega}).$$

- This equation has the interpretation that an LTI system amplifies the spectrum of the signal $X(e^{j\omega})$ by $H(e^{j\omega})$. That is, we can think that the input $x[n]$ is decomposed into its component sinusoids, each sinusoid is amplified according to $H(e^{j\omega})$, and the result is recombined to be the output $y[n]$.
- For this reason, the LTI systems are also called filters.

Sampling

- Let's look at the spectrum of a discrete-time signal obtained from periodic sampling,

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

- The main result is stated below.

If $x_c(t)$ *is band-limited to* $\frac{1}{2T}$*, then the spectrum of* $x[n]$ *is periodic with period* $\frac{1}{T}$*. Furthermore, each period is a replica of the spectrum of* $x_c(t)$*, so* $x_c(t)$ *is uniquely determined by* $x[n]$.

- See, for example, Oppenheim and Schafer.

Linear Difference Equations

- A linear difference equation is defined by

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

- Taking the z-transform, one has

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

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

Poles and Zeros

- Note that a linear difference equation defines an LTI discrete-time system.
- $H(z)$ is called the transfer function. In fact, it is the z-transform of the output signal when the input is the impulse signal.
- A pole is a value z where $H(z)$ is singular. A zero is a value z where $H(z)$ is 0.
- For example, for the system defined by

$$y[n] = Ky[n-1] + x[n],$$

$z = K$ is a pole and $z = 0$ is a zero.

Resonances

- A resonance frequency is the angular frequency ω where $|H(e^{j\omega})|$ is a maximum when we look at $H(z)$ along the unit circle.
- A second-order difference equation with transfer function
$$\frac{1}{1 - Az^{-1} - Bz^{-2}}$$
has a pair of conjugate poles whose locations depend on A and B.
- One way to create resonance is to cascade a second-order system with another system having the poles of the first system as zeros.

Discrete Fourier Transform

- The discrete Fourier transform (DFT) of a finite-duration sequence $x[n], 0 \leq n \leq N-1$, is defined by

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

  where $W = e^{-j(2\pi/N)}$.
- The $X[k]$'s are exactly the N equally spaced samples of the spectrum of $x[n]$ at the points $\omega_k = \frac{2\pi k}{N}$.

Inverse Discrete Fourier Transform

- From $X[k]$, we can obtain $x[n]$ by

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

since

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

- This is called the inverse discrete Fourier transform (IDFT). Note that we get more than we put in, as $x[n]$ is non-zero beyond the original finite duration.

Circular Convolution

- The convolution theorem using DFT representation is subtle. To get things right, we need the circular convolution. The circular convolution of $x_1[n]$ 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.,

$$\tilde{x}_i[n] = x_i[n \mod N].$$

Convolution Theorem with DFT

- Let $x_1[n]$ and $x_2[n]$ be N-point finite-duration sequences with DFTs $X_1[k]$ and $X_2[k]$. Then the IDFT of the product $X_1[k]X_2[k]$ is the circular convolution $x_1[n] \otimes x_2[n]$.
- You should notice the difference between this and the earlier convolution theorem.

Proof

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

FIR Filter Implementation via DFT

- If we multiply the DFTs of two finite-duration sequences, then apply IDFT, we get the circular convolution of the sequences.
- To implement an LTI system, we want the linear convolution rather than the circular convolution.
- Can we implement an LTI system by DFT? Yes.
  - Suppose two finite-duration sequences are of lengths N_1, N_2.
  - We first augment the original sequences to a length $N \geq N_1 + N_2 - 1$ by padding zeros.
  - Then we compute the (N-point) IDFT of the product of DFTs.

Fast Fourier Transform

- Counting only multiplication, the DFT requires N^2 operations while FFT requires $N \log N$ operations.
  - We can express an N-point DFT of $x[n]$ by

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

    where $G[k]$ is the DFT of even-numbered points and $H[k]$ is the DFT of odd-numbered points.
  - Recursively apply the same idea until the 2-point DFTs are to be computed

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