Digital Signal Processing

ESE 547

Part 1

System Overview

Sampled-Signal Processing Basics

Frequency-Domain and Transforms

Signal Sampling and Reconstruction

1.1 System Overview

- Applications
- Drawbacks
- Advantages
- System Structure

1.1.1 Applications

- Consumer:
  - CD
  - DAT
  - HDTV (Future)
  - Speech Recognition
- Military:
  - Radar
  - Sonar
  - Guidance
  - Fundamental Technology

Applications (Continued)

- Industrial:
  - Control
  - Medical Imaging
  - Medical Signals (EEG, ECG, EMG)
  - Geophysical Exploration
  - Nondestructive Testing
  - Mechanical Monitoring and Diagnosis
- COMMUNICATIONS

1.1.2 Drawbacks

- Processing Bottlenecks
  - A/D
  - Processor
- High Initial Cost; justified by
  - High performance
  - High complexity
  - Leverage
  - Flexibility

1.1.3 Advantages

- Precision
- Flexibility
- Low incremental Cost
- High capability

1.1.4 Digital Signal Processing System Structure

- Anti-Aliasing Filter
- Sample-and-Hold (or Track-and-Hold)
- A-to-D Converter
- Digital Processor
- D/A converter
- Smoothing Filter

Filters and conversion: extremely important; will cover later.

Conversion process sets a limit on the signal bandwidth possible – usually $< 1/2$ sampling frequency.

Present focus on processing algorithms.

1.2 Sampled-Signal Processing Basics

- Input-Output Description and Impulse Response
- FIR & IIR Systems
- Processing Elements and Difference Equations

1.2.1 Input-Output Description

Input: a sequence $u = (u_n)$; output $y = (y_n)$

$$y = B(u)$$

Linear:

$$\begin{aligned} B(u+v) &= B(u) + B(v) \\ B(\alpha u) &= \alpha B(u) \end{aligned}$$

for all constants α.

Time-Invariant:

If

$$(y_n) = B((u_n))$$

then

$$(y_{n+k}) = B((u_{n+k}))$$

for all k.

Input-Output Description (continued):

Fundamental fact: If the processing is *linear* and *time-invariant* then it must be a *convolution* – so convolution is the main task of DSP chips.

Convolution is defined by

$$(y_n) = (u_n) * (h_n)$$

where

$$y_n = \sum_{k=-\infty}^{\infty} h_k u_{n-k}$$

and h_n is the *impulse response* given by:

$$(h_n) = B((\delta_n))$$

where (δ_n) is the discrete-time unit (im)pulse:

$$\delta_n = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

Causal: $h_n = 0$ for $n < 0$.

1.2.2 FIR & IIR Systems

Special convolution cases:

1. FIR filtering – Finite Impulse Response, usually causal, simplest; a direct convolution whose impulse response is given by the filter coefficients:

$$\begin{aligned} y_n &= h_0 u_n + h_1 u_{n-1} + h_2 u_{n-2} \cdots + h_M u_{n-M} \\ &= \sum_{i=0}^{M} h_i u_{n-i} \end{aligned}$$

2. IIR filtering – Infinite Impulse Response, causal, more complex than FIR; uses past outputs:

$$\begin{aligned} y_n &= b_0 u_n + b_1 u_{n-1} + \cdots + b_M u_{n-M} \\ &\quad -a_1 y_{n-1} - a_2 y_{n-2} - \cdots - a_N y_{n-N} \\ &= \sum_{i=0}^{M} b_i u_{n-i} - \sum_{k=1}^{N} a_k y_{n-k} \end{aligned}$$

3. FFT – Fast Fourier Transform: usually noncausal, can be used for causal, usually FIR

FIR & IIR Systems (continued):

FIR Terminology:

$$y_n = b_0 u_n + b_1 u_{n-1} + \cdots + b_M u_{n-M} = \sum_{i=0}^{M} b_i u_{n-i}$$

— Finite Discrete Convolution

— Transversal Filter

— Tapped Delay Line

Other Algorithms:

1. FFT is also widely used for spectral analysis and spectral estimation
2. Adaptive processing is nonlinear/time-varying
3. Many coding/modulation and decoding/demodulation and detection techniques are nonlinear.

Note: a DSP chip's architecture must have fast multiply-adds, efficient shifting (circular buffers), and zero-overhead (non-branching) looping; also usually has bit-reversed addressing.

1.2.3 Processing Elements and Difference Equations

For Linear Time-Invariant processing, the basic processing elements are:

- Unit Delays
- Constant gains
- Adders

Mathematically, interconnections of these elements give difference equations.

Schematically, we get block diagrams or signal flow graphs.

Note: z^{-1} is used to denote a unit delay, even in a time-domain block diagram or signal flow graph!

Examples: FIR: moving average filter:

$$y_n = 0.5u_n + 0.5u_{n-1}$$

IIR: (digital) integrator:

$$y_n = u_n + y_{n-1}$$

Block Diagrams: note that the FIR has no (directed) loops; if there are loops, they must have at least one delay.

Elementary sampled signals:

(Recall) unit (im)pulse:

$$\delta_n = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

Note: the unit impulse is an ordinary function in DSP. When simulating continuous-time systems, be careful about its normalization!

Unit step:

$$U_n = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases}$$

Check unit step is impulse response of (digital) integrator.

1.3 Frequency-Domain and Transforms

- Sampled Sinusoids
- Frequency, Amplitude, and Phase Responses
- DTFT, Z-transforms and Properties
- Poles and Zeros

1.3.1 Sampled Sinusoids

Continuous-time:

$$x(t) = A\cos(2\pi f t + \phi)$$

where A is the amplitude, and ϕ is the phase.

If signal is sampled every T seconds, the sampling instants are $t = nT$, and so the sampled (discrete-time) sinusoid is:

$$x_n = A\cos(2\pi f nT + \phi)$$

where A is the amplitude, and ϕ is the phase.

Amplitude: A (note: not necessarily peak value!)

Phase: ϕ

Sampling period: T

Sampling frequency: $f_S = 1/T$

Angular sampling frequency: $\omega_S = 2\pi/T$

Note: the sequence (x_n) is unchanged if the frequency f is changed to $f + k/T$, for any integer k (aliasing).

Sampled Sinusoids (continued):

Sinusoid in complex form:

$$\begin{aligned} g_n &= Ae^{j(2\pi f_0 nT + \phi)} \\ &= Ae^{j\phi}\left(e^{j2\pi f_0 T}\right)^n \end{aligned}$$

where A is the amplitude, and ϕ is the phase.

For a decaying sinusoid:

Continuous-time:

$$g(t) = Ae^{\sigma_0 t} e^{j(2\pi f_0 t + \phi)}$$

Sampled:

$$\begin{aligned} g_n &= Ae^{\sigma_0 nT} e^{j(2\pi f_0 nT + \phi)} \\ &= Ae^{j\phi} z_0^n \end{aligned}$$

where

$$\begin{aligned} z_0 &= e^{(\sigma_0 + j\omega_0)T} \\ &= e^{s_0 T} \end{aligned}$$

Sampled Sinusoids (continued):

Consequence:

The relationship between the (continuous-time) s-domain and the (discrete-time) z-domain is given by:

$$z = e^{sT}$$

Restrict to frequency domain:

$$z = e^{j2\pi fT} = e^{j\theta}$$

So the discrete-time *angular frequency* is given by

$$\theta = 2\pi fT = \omega T$$

and the frequency information is on the unit circle.

Also, stability is equivalent to the absence of poles *on or outside the unit circle*.

1.3.2 Frequency, Amplitude, and Phase Responses

Suppose we have an FIR system with input $u_n = e^{j2\pi fnT}$. Then, with $\omega = 2\pi f$:

$$\begin{aligned} y_n &= b_0 e^{j\omega nT} + b_1 e^{j\omega(n-1)T} + \cdots + b_M e^{j\omega(n-M)T} \\ &= e^{j\omega nT}\left(b_0 + b_1 e^{-j\omega T} + \cdots + b_M e^{-Mj\omega T}\right) \\ &= e^{j\omega nT} K(\omega T) \end{aligned}$$

So if u_n is a (complex) sampled sinusoid of angular frequency ω,

$$\begin{aligned} y_n &= K(\omega T) e^{j\omega nT} \\ &= K(\omega T) u_n \end{aligned}$$

where the complex number $K(\omega T)$ is given by

$$K(\omega T) = b_0 + b_1 e^{-j\omega T} + \cdots + b_M e^{-Mj\omega T}$$

Frequency Response (continued)

For an IIR system, if we also assume that the output y_n is given by $y_n = Ke^{j2\pi fnT}$, where K is a complex constant:

$$Ke^{j\omega nT} + a_1 Ke^{j\omega(n-1)T} + \cdots + a_N Ke^{j\omega(n-N)T}$$
$$= b_0 e^{j\omega nT} + b_1 e^{j\omega(n-1)T} + \cdots + b_M e^{j\omega(n-M)T}$$

or

$$Ke^{j\omega nT}\left(1 + a_1 e^{-j\omega T} + \cdots + a_N e^{-Nj\omega T}\right)$$
$$= e^{j\omega nT}\left(b_0 + b_1 e^{-j\omega T} + \cdots + b_M e^{-Mj\omega T}\right)$$

so that

$$\begin{aligned} K(\omega T) &= H(e^{j\omega T}) \\ &= \frac{b_0 + b_1 e^{-j\omega T} + \cdots + b_M e^{-Mj\omega T}}{1 + a_1 e^{-j\omega T} + \cdots + a_N e^{-Nj\omega T}} \end{aligned}$$

Frequency Response (continued)

For a general convolution, (with $\omega = 2\pi f$)

$$\begin{aligned} y_n &= \sum_{k=0}^{\infty} h_k e^{j\omega(n-k)T} \\ &= e^{j\omega nT} \sum_{k=0}^{\infty} h_k e^{-j\omega kT} \\ &= e^{j\omega nT} H(e^{j\omega T}) \end{aligned}$$

The function $H(e^{j\omega T})$ is called the *frequency* response of the system;

The function $A(\omega T) = \left|H(e^{j\omega T})\right|$ is called the *amplitude* response;

The function $\phi(\omega T) = \angle H(e^{j\omega T})$ is called the *phase* response.

Note:

- These are periodic with period $2\pi/T = 2\pi f_S = \omega_S$
- They depend only on the *ratio* $f/f_S = \omega/\omega_S$

For this reason, digital signal processing often works with a normalized frequency f/f_S.

1.3.3 DTFT and Z-transforms

Discrete-Time Fourier Transform (DTFT):

Based on the formula for the frequency response, the Discrete-Time Fourier Transform of a (bounded) sequence (g_n), is defined by:

$$G(\theta) = \sum_{k=-\infty}^{\infty} g_k e^{-jk\theta}$$

The inverse transform is given by:

$$g_n = 1/2\pi \int_{-\pi}^{\pi} G(\theta) e^{jn\theta} d\theta$$

Note 1: *Boundednes* means that there is some constant M such that $|g_n| \leq M$ for all n.

Note 2: If (g_n) is the impulse response of a system, the frequency response of the system is given in terms of the DTFT of (g_n) by

$$G(e^{j2\pi fT}) = G(e^{j\theta})|_{\theta=2\pi fT}$$

DTFT and Z-transforms (continued)

Two-sided Z-transform:

The *two-sided* Z-transform of a bounded sequence (g_n) is defined by

$$\begin{aligned} G(z) &= \mathcal{Z}\{g_n\} \\ &= \sum_{n=-\infty}^{\infty} g_n z^{-n} \\ &= \cdots + g_{-1}z + g_0 + g_1 z^{-1} + g_2 z^{-2} + \cdots \end{aligned}$$

for all complex numbers z for which the series converges – assumed to include the unit circle.

Two-Sided Z-transform properties:

- Linearity
- Shifting: $\mathcal{Z}\{g_{n-k}\} = z^{-k}G(z)$ for all k
- Convolution: If f_n and g_n are two bounded sequences, then $\mathcal{Z}\{(f_n) * (g_n)\} = F(z)G(z)$

— *The Z-transform transforms convolution into multiplication.*

Examples:

1. $a^n U_n$ with $|a| < 1$
2. $-a^n U_{-n-1}$ with $|a| > 1$
3. sinusoids

DTFT and Z-transforms (continued)

Relation to frequency response and DTFT:

Frequency response is given by

$$K(2\pi fT) = \sum_{k=-\infty}^{\infty} h_k e^{-j2\pi kfT}$$

so that

$$\begin{aligned} K(2\pi fT) &= G(e^{j2\pi fT}) \\ &= G(e^{j\theta})|_{\theta=2\pi f/f_S} \\ &= G(z)|_{z=e^{j2\pi f/f_S}} \end{aligned}$$

The main use of the two-sided Z-transform is in calculating the inverse DTFT of functions which are rational in z.

The inverse DTFT of some other types of functions of θ is found by the direct inverse formula (will see later).

DTFT and Z-transforms (continued)

The *one-sided* Z-transform of a sequence (g_n) is defined by

$$\begin{aligned} G(z) &= \mathcal{Z}\{g_n\} \\ &= \sum_{n=0}^{\infty} g_n z^{-n} \\ &= g_0 + g_1 z^{-1} + g_2 z^{-2} + \cdots \end{aligned}$$

for all complex numbers z for which the series converges (region of convergence, or ROC).

Differs from the two-sided Z-transform and DTFT:

- assumes that (g_n) is one-sided
- does not assume that (g_n) is bounded
- therefore does not necessarily converge on unit circle
- converges in a region of the form $\{z \mid \; |z| > R\}$ for some R.
- handles initial conditions and transients.

DTFT and Z-transforms (continued)

Relation to frequency response, for a causal, *stable* system :

As before, frequency response is given by

$$K(2\pi fT) = \sum_{k=0}^{\infty} h_k e^{-j2\pi fTk}$$

so that

$$\begin{aligned} K(2\pi fT) &= H(e^{j2\pi fT}) \\ &= H(z)|_{z=e^{j2\pi fT}} \end{aligned}$$

Here, stable means BIBO stable, which is equivalent to h_n being absolutely summable, i.e., $\sum_{k=0}^{\infty} |h_k|$ is finite (stronger than bounded).

DTFT and Z-transforms (continued)

If g_n is the impulse response, h_n, of a causal system, its one-sided Z-transform

$$H(z) = \mathcal{Z}\{h_n\} = \sum_{n=0}^{\infty} h_n z^{-n}$$

is called the *transfer function* of the system

Examples:

1. $a^n U_n$
2. sinusoids
3. exponentially decaying and growing sinusoids

DTFT and Z-transforms (continued)

(One-sided) Z-transform properties:

- Linearity
- Shifting: $\mathcal{Z}\{g_{n-k}\} = z^{-k}G(z)$ for $k \geq 0$
- Non-Causal Shift: $\mathcal{Z}\{g_{n+1}\} = zG(z) - zg_0$
- Convolution: If $f_n = 0$ and $g_n = 0$ for $n < 0$, then $\mathcal{Z}\{(f_n) * (g_n)\} = F(z)G(z)$

— *The Z-transform transforms convolution into multiplication.*

Note 1: The noncausal shifting property enables the Z-transform to handle initial conditions and transients.

Note 2: Because the signals are one-sided, the convolution here is given by:

$$p_n = \sum_{k=0}^{n} f_k g_{n-k}$$

DTFT and Z-transforms (continued)

Inverse one-sided Z-transform:

Theoretical formula (rarely used, actually a contour integral):

$$g_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(Re^{j\theta})(Re^{j\theta})^n \, d\theta$$

where R is such that all poles of $G(z)$ are inside the circle of radius R.

The normal method is to use partial fractions, but the calculations are simpler if $G(z)/z$, instead of $G(z)$, is written in terms of z (not z^{-1}) and is then expanded in partial fractions.

Examples:

1. $1/(1 - az^{-1})$ with $|a| < 1$
2. $1/(1 - az^{-1})$ with $|a| > 1$
3. real poles – stable
4. real poles – unstable
5. complex poles – stable and unstable
6. comples poles on unit circle – second-order oscillator; (see p.394)

1.3.4 Poles and Zeros

Terminology (for rational functions with no common factors between numerator and denominator):

1. Pole: a value of z where $H(z)$ is infinite, i.e., where the denominator is zero.
2. Simple Pole: a is a simple pole if there is a factor $(z - a)$ in the denominator, and no factor of $(z - a)^k$ with $k > 1$.
3. Multiple Pole: a is a multiple pole if there is a factor $(z - a)^k$ in the denominator, with $k > 1$.
4. Zero: a value of z where $H(z)$ is zero
5. Residue: the coefficient A in the term $A/(z - a)$ in the partial fraction expansion.

Poles and Zeros (continued)

Notes:

1. All rational functions can be expanded as product of first-order factors with complex coefficients.
2. All rational functions can be expanded as product of first- and second-order factors with real coefficients.
3. To have a partial fraction expansion, we must have $H(z) \to 0$ as $z \to \infty$
4. Partial fraction expansions get more complicated with multiple poles; the basic formulas can be found by differentiating the equation

$$\frac{1}{1 - az^{-1}} = \sum_{n=0}^{\infty} a^n z^{-n}$$

5. Knowing the poles gives us a good idea of what the impulse response will look like.
6. Knowing the poles and using the initial- and final-value theorems gives a good idea of the step response.
7. The poles and zeros (product expansion) can be used to give a good idea of the frequency response.

1.4 Signal Sampling and Reconstruction

- Theory
- Antialiasing and Smoothing Filters
- Signal Conversion

1.4.1 Sampling and Reconstruction Theory

Theoretically, sampling is a multiplication operation:

Give a sampl*ing* signal $s(t)$ which is periodic with period T, and a continuous-time signal $g(t)$, define the sampl*ed* signal $g_S(t)$ by

$$g_S(t) = s(t)g(t)$$

T is called the *sampling period* and $f_S = 1/T$ is the *sampling frequency.*

Normally, $s(t)$ will approximate a periodic sequence of δ-functions, and a further integration step is used to obtain a specific value.

Note that the sampled function is still a continuous-time function.

Sampling and Reconstruction Theory (continued)

For *ideal* sampling, we assume that $s(t)$ is a periodic sequence of δ-functions:

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$$

and the sampling operation then has two steps:

- Multiply by the sequence of δ-functions:

$$\begin{aligned} g_S(t) &= \sum_{n=-\infty}^{\infty} \delta(t - nT) g(t) \\ &= \sum_{n=-\infty}^{\infty} g(nT) \delta(t - nT) \end{aligned}$$

- Pick off numbers representing samples

$$g_n = g(nT)$$

Sampling and Reconstruction Theory (continued)

Effect of sampling in frequency domain:

$$G_S(2\pi jf) = \sum_{k=-\infty}^{\infty} c_k G(2\pi j(f - kf_S))$$

where c_k are the complex Fourier coefficients of the $s(t)$:

$$c_k = \frac{1}{T}\int_{-T/2}^{T/2} s(t)e^{-j2\pi kt/T}\ dt$$

Here $2\pi f_S = 2\pi/T = \omega_S$ is called the *sampling angular frequency.*

Also, $f_S/2$ is sometimes called the *folding frequency*.

For ideal sampling, we get

$$G_S(2\pi jf) = 1/T \sum_{k=-\infty}^{\infty} G(2\pi j(f - kf_S))$$

Sampling and Reconstruction Theory (continued)

Classical Sampling Criterion:

Must Sample at a rate higher than the *Nyquist rate* $= 2f_{MAX}$

WARNING!1! Nonlinear operations internal to the signal processing can increase the bandwidth, and require much higher sampling rates.

WARNING!2! Aliased components are not just a mathematical abstraction, but very real; you can see them on an oscilloscope, and hear them.

WARNING!3! Higher sampling rates can also be needed to reduce phase-shift, for example, in a feedback loop.

Sampling and Reconstruction Theory (continued)

To reconstruct: use lowpass filter

Reconstruction is also possible in other cases:

- Bandpass
- Periodic

In time-domain:

$$g_R(t) = \sum_{n=-\infty}^{\infty} g(nT)l(t - nT)$$

where $l(t)$ is the impulse response of the reconstruction (smoothing) filter.

Sampling and Reconstruction Theory (continued):

“Ideal” reconstruction: ideal low-pass filter with theoretical gain T; unrealizable since

$$\begin{aligned} l(t) &= \frac{\sin(\pi t/T)}{\pi t/T} \\ &= \operatorname{sinc}(t/T) \end{aligned}$$

(Note that this vanishes at all sample points except $t = 0$, and so gives the correct value at the sample points.)

Actual D/A conversion: zero-order hold followed by smoothing filter. Frequency response of zero-order hold is given by

$$\begin{aligned} H(f) &= e^{-j\pi f/f_S}\frac{\sin(\pi f/f_S)}{\pi f/f_S} \\ &= e^{-j\pi x}\frac{\sin(\pi x)}{\pi x} \end{aligned}$$

where x is the normalized frequency.

It follows that the zero-order hold has a group delay of $T/2$, and has a rough lowpass characteristic with infinite attenuation at non-zero multiples of the sampling frequency, and an attenuation of $2/\pi$ or about 3.9 dB at $f_S/2$.

1.4.2 Antialiasing (Input) and Smoothing (Output) Filters:

- Needed from sampling theorem
- To reject spurious signals and noise at input
- To reject aliased components at output

Problem: Need high order *analog* filter, especially at output

Can alleviate requirements at output:

- Upconvert, filter digitally, high-rate D/A converter,
- Upconvert, high-rate D/A converter, switched-capacitor filter, final simple smoothing filter.

Antialiasing and Smoothing Filters (continued)

Can also alleviate requirements at input:

- Gentle analog filter, very high sampling rate – low accuracy
- Use high-frequency, low word-width Digital lowpass filter
- Use DSP to Down-convert to low sampling rate, high word-width – noise-shaping
- Σ-Δ converters.

1.4.3 Signal Conversion

Track and Hold Amplifier:

- Tracks signal until strobed, and holds constant value until released.
- Stores charge on capacitor
- Properties
  - Aperture
  - Acquisition time
  - Step
  - Droop
  - Jitter

Signal Conversion (continued)

A/D Converters:

- Successive Approximation
- Flash
- Sigma-Delta
- Dual-Slope, Counting, usually too slow

Signal Conversion (continued)

D/A Converters:

Switched current source; numerous other types.

Virtually always use zero-order hold.

Zero-order hold amplitude response: “$\sin(x)/x$”

Zero-order hold delay: 1/2 sample period.