CSC475 Music Information Retrieval

Discrete Fourier Transform

George Tzanetakis

University of Victoria

2014

Table of Contents I

Sampling

- Discretize continuous signal by taking regular measurements in time (discretization of the measurements is called quantization)
- Notation: f_s is sampling rate in Hz, ω_s is sampling rate in radians per second
- Sampling a sinusoid - only frequencies below half the sampling rate (Nyquist frequency) will be accurately reprsented after sampling
- For sinusoid at ω_0 then all frequencies $\omega_0 + k\omega_s$ are aliases

Phasor view of aliasing

University of Victoria Computer Science

Illustration of sampling at a high sampling rate compared to the phasor frequency, sampling at the Nyquist rate, and slightly above. Numbers indicate the discrete samples of the continuous phasor rotation. Each sample is a complex number.

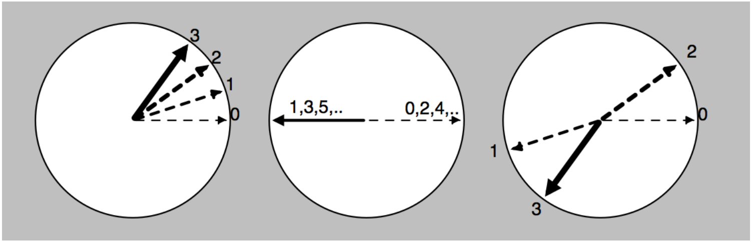


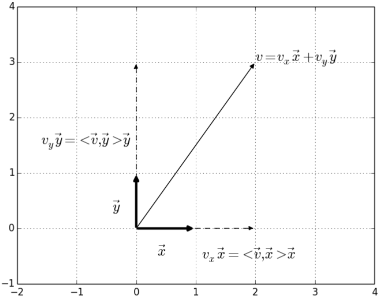
Table of Contents I

Frequency Domain

- Any periodic sound can be represented as a sum of sinusoids (or equivalently phasors)
- This representation is called a *frequency domain* representation and the linear combination coefficients are called the *spectrum*
- Commonly used variants: Fourier Series, Discrete Fourier Transform, the z-transform, and the classical continuous Fourier Transform
- These transforms provide procedures for obtaining the linear combination weights of the *frequency domain* from the signal in *time domain* (as well as the inverse direction)

2D coordinate system

A vector $\vec{v}$ in 2-dimensional space can be written as a combination of the 2 unit vectors in each coordinate direction. The inner product operation $< \hat{v}, \hat{w} >$ corresponds to the projection of $\vec{v}$ onto $\vec{w}$. It is the sum of the products of like coordinates $< v, w >= v_x w_x + v_y w_y = \sum_{i=0}^{N-1} v_i w_i$.



Inner product (projection) properties

- $< \vec{x}, \vec{y} >= 0$ then the vectors are orthogonal
- $< \vec{u} + \vec{v}, \vec{w} >=< \vec{u}, \vec{w} > + < \vec{v}, \vec{w} >$ distributive law
- Basis vectors are orthogonal and have length 1
- $< \vec{v}, \vec{v} >= v_x^2 + v_y^2$ is the square of the vector length
- In order to have the inner product with self be the square of length for vectors of complex numbers we have to slightly change the definition by using the complex conjugate.
- $< v, w >= v_x w_x^* + v_y w_y^* = \sum_{i=0}^{N-1} v_i w_i^*$. where $()^*$ denotes the complex conjugate of a number.

Hilbert Spaces

University of Victoria Computer Science

Key idea

Generalize the notion of a Euclidean space with finite dimensions to other types of spaces for which a suitable notion of an inner product can be defined. These space can have an infitite number of dimensions but as long as we have an appropriate definition of a projection operator/inner product we can reuse a lof the notation and concepts familiar from Euclidean space. For example a space we will investigate are all continuous functions that are periodic with an interval $[0, T]$.

Orthogonal Coordinate System

We need an orthogonal coordinate system i.e a projection (inner product) operator and an orthogonal basis for each space we are interested

The Fourier Series, the Discrete Fourier Transform, the z-transform and the continuous Fourier Transform can all be defined by specifying what projection operator to use and what basis elements to use

Table of Contents I

Generalizing to continuous periodic functions

We are interested in periodic functions of a continuous variable t in the interval $0 \leq t \geq T$. We can think of each particular value of t as a coordinate and then generalize the sum of the Euclidean inner product to an integral.

Inner product

$$< f, g >= \frac{1}{T} \int_0^T f(t) g^*(t)$$

Basis elements (surprise, surprise phasors)

$$e^{jk\omega_0 t}, k = \ldots, -1, 0, 1, 2, ..$$

where $\omega_0 = 2\pi / T$.

Fourier Series

Now that we have the definition of an inner product and a basis we can write any periodic function as a linear combination of the basis elements. The periodic signal $f(t)$ is expressed in a new coordinate system, with component c_K in the "direction" of the phasor $e^{jk\omega_0 t}$ with frequecy $k\omega$.

Definition

The **Fourier Series** of f(t) is:

$$f(t) = \sum_{-\infty}^{+\infty} c_k e^{jk\omega_0 t}$$

The sequence c_k is the **spectrum** and the c_k coefficients are complex numbers.

Obtaining the spectrum

University of Victoria Computer Science

Definition

To find the Fourier coefficients c_k we simply project using our definition of the inner product for continuous periodic functions the function $f(t)$ on the kth basis element:

$$c_k = < f(t), e^{jk\omega_t} > = \frac{1}{T} \int_0^T f(t) e^{-jk\omega_0 t}$$

Note 1: we have changed the coordinate system drastically from a continuously indexed coordinate (time) to one with a discretely indexed coordinate (the phasor basis) that is infinitely countable. There are some associated mathematical restrictions but remarkably it works.

Table of Contents I

Discrete Fourier Transform Introduction

- The **DFT** is an abstract mathematical transformation and the Fast Fourier Transform **FFT** is a very efficient algorithm for computing it
- The **FFT** is at the heart of digital signal processing and a lot of MIR systems utilize it one way or another
- It is applied on sequences of N samples of a digital signal
- Similarly to the Fourier Series we will define it using the components of an orthogonal coordinate system: an inner product and a set of basis elements

DFT

University of Victoria Computer Science

Our input is a finite, legnth N segment of a digital signal $x[0], \ldots, x[N-1]$.

Definition

The inner product is what one would expect:

$$< x, y >= \sum_{t=0}^{N-1} x[t]y^*[t]$$

Switch frequency interval from $[-\omega_s/2, +\omega_s/2]$ to $[0, \omega]$a as they are equivalent. One possibility would be all phasors in that frequency range: $e^{jt\omega}$ for $0 \leq \omega < \omega_s$. It turns out we just need N phasors in that range.

DFT basis elements

We need N frequencies spaced in the range from 0 to the sampling frequency. If we use radians per sample then we have $0, 2\pi/N, 2(2\pi/n), \ldots, (N-1)(2\pi/N)$. The corresonding basis is:

Definition

$$e^{jk2\pi/N} \quad \text{for} \quad 0 \leq k \leq N-1$$

Note: using the definition of the inner product above one can show that indeed these basis elements are orthogonal i.e the inner product between any two elements is zero.

The Discrete Fourier Transform (DFT)

University of Victoria Computer Science

Definition

The inverse DFT expresses the time domain signal as a complex weighted sum of N phasors with spectrum $X[k]$.

$$x[t] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{jtk2\pi/N}$$

Definition

The DFT can be obtained by projecting the signal to the basis elements using the inner product definition:

$$X[k] = < x, e^{jtk2\pi/N} > = \sum_{t=0}^{N-1} x[t] e^{-jtk2\pi/N}$$

Matrix-vector formulation

One can view the DFT as a way to transformation a sequence of N complex numbers to a different sequence of N complex numbers. The DFT can be expressed in matrix-vector notation. If we use x and $\mathbf{X}$ to denote the N dimensional vectors with component $x[t]$ and $X[k]$ respectively, and define the $N \times N$ matrix F by

$$[F]_{k,t} = e^{-jtk2\pi/N}$$

then we can write:

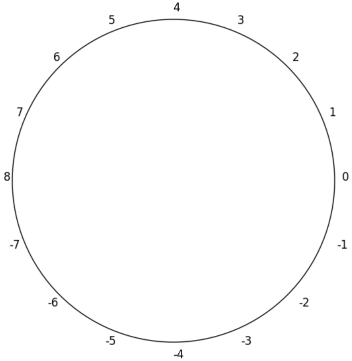
$$\mathbf{X} = \mathbf{F}\mathbf{x}$$

and

$$\mathbf{x} = \mathbf{F}^{-1}\mathbf{X}$$

Circular Domain

The bins of the DFT are numbered $0, N-1$ but correspond to frequencies between $[-\omega_s/2, +\omega_s/2]$.



The discrete frequency domain

The bins of the DFT are numbered $0, N-1$ but correspond to frequencies between $[-\omega_s/2, +\omega_s/2]$. Since N corresponds to the sampling rate, we need to divide by N to get the frequencies in terms of fractions of the sampling rate. So in the case shown in the figure we would have the following frequencies (fractions of sampling rate):

$$0, \frac{1}{16}, \frac{2}{16}, ... \frac{8}{16}, -\frac{7}{16}, \frac{6}{16}, \ldots, -\frac{1}{16}$$

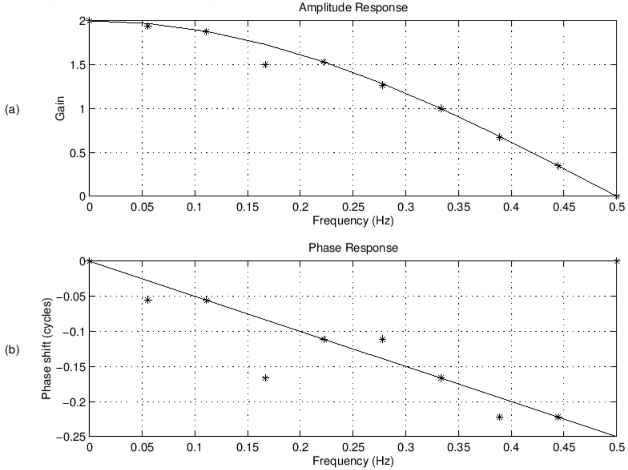
DFT frequency mapping example

Example of a 2048-point DFT with 44100 Hz sampling rate:

bin	frequency
0	0
1	21.5
2	43.1
...	...
1024	22050
1025	-220285
1026	22007

Magnitude and Phase Spectrum

Amplitude and phase spectrum of a low-pass filter response



Fast Fourier Transform

- Straight implementation of the DFT requires $O(N^2)$ arithmetic operations.
- Divide and conquer: do two $N/2$ DFTs and then merge the results
- $O(NlogN)$ much faster when N is not small.

Summary I

- Sampling a phasor introduces **aliasing** which means that multiple frequencies (the aliases) are indistinguishable from each other based on the samples
- We can extend the concept of an orthogonal coordinate system beyond Euclidean space vectors
- By appropriate definitions of a **projection operator** (inner product) and **basis elements** we can formulate transformations from the time domain to the frequency domain such as the Fourier Series and the Discrete Fourier Transform
- The FFT is a fast implementation of the DFT

Summary II

- Any signal of interest can be expressed as a weighted sum (with complex coefficients) of basis elements that are **phasors**.
- The complex coefficients that act as **coordinates** are called the **spectrum** of the signal
- We can obtain the **spectrum** by projecting the **time domain** signal to the phasor basis elemnts
- The DFT output contains frequency between $[-\omega_s/2, +\omega_s/2]$ using a circular domain. For real signal the coefficients of the negative frequencies are symmatric to the positive frequencie and carry not additional information.