

ECE 6270 - Advanced DSP Notes

Ben Rotker

September 10, 2025

Contents

Introduction	2
About	2
Disclaimer	2
Github Information	2
Week 1	3
Lecture 1: Sampling pt 1	3
Lecture 2: Sampling pt 2	6
Lecture 3: Sampling pt 3	8
Week 2	11
Lecture 4: Basis Expansions	11
Lecture 5: Vector Spaces	12
Lecture 6: Linear Subspaces	13
Lecture 7: Linear Combinations and Spans	14
Lecture 8: Norms and Inner Products	16
Week 3	21
Lecture 9: Approximations in Hilbert Spaces pt 1	21
Lecture 10: Approximations in Hilbert Spaces pt 2	24
Lecture 11: Approximation Example	25
Lecture 12: Orthobases	27
Lecture 13: Orthobasis Expansions	29
Lecture 14: Parseval's Theorem++	30
Week 4	30

Introduction

About

The goal of this document is to provide detailed notes that supplement the lecture material and provide quick lookups for formulas and concepts. It was inspired by Shuvo Newaz's lecture notes for ECE 6271 - Adaptive Filtering.

Disclaimer

I did my best to closely follow the lectures when creating these notes; however, they may not be 100% perfect or error free. Please treat them as a supplement to the course and double-check anything you're unsure about using the official course material.

Github Information

These notes are open source and hosted on GitHub. You can view the latest version, make corrections, or suggest improvements at:

<https://github.com/rotkstar/ECE6250-Advanced-DSP-Notes>

If you'd like to contribute, please open an issue or submit a pull request with your correction.

Week 1

Lecture 1: Sampling pt 1

Shannon-Nyquist

Sampling is simply converting a continuous signal into discrete samples.

$$x[n] = x_c(nT) \quad (1.1)$$

Where T is the sampling period or interval. In this course, square brackets denote discrete signals and parentheses denote continuous signals.

Questions:

1) Can we reconstruct $x_c(t)$ from $x[n]$?

Short answer: sometimes. When $x_c(t)$ is band limited, that is, the frequencies are zero beyond a certain point, that is,

$$X_c(j\Omega) = 0 \quad \forall \quad |\Omega| \geq \frac{\pi}{T} \quad (1.2)$$

Where

$$X_c(j\Omega) = \int_{-\infty}^{\infty} X_c(t) e^{-j\Omega t} dt \quad (1.3)$$

and

$$e^{-j\Omega t} = \cos(\Omega t) - j\sin(\Omega t)$$

We're really projecting our time signal onto a frequency basis. We get a representation of our original function in frequency (where Ω is the frequency in radians). When x is bandlimited, the frequency of x is zero above a certain point. Thus we can reconstruct $x_c(t)$ from $x[n]$ as long as the sampling rate is \geq twice the highest frequency in $x_c(T)$. This is the Nyquist Criterion:

$$\frac{2\pi}{T} > 2\Omega_{max} \quad (1.4)$$

2) How exactly do we reconstruct the signal?

Answer: We reconstruct using a sinc interpolator.

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n] \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)/T} \quad (1.5)$$

$$= \sum_{n=-\infty}^{\infty} x[n] \underbrace{g_T(t - nT)}_{\text{shifted sinc}} \quad (1.6)$$

Issue: the sinc function dies out as $\frac{1}{T}$, meaning it never dies out until T reaches infinity. So we're adding an infinite number of functions with infinite supports.

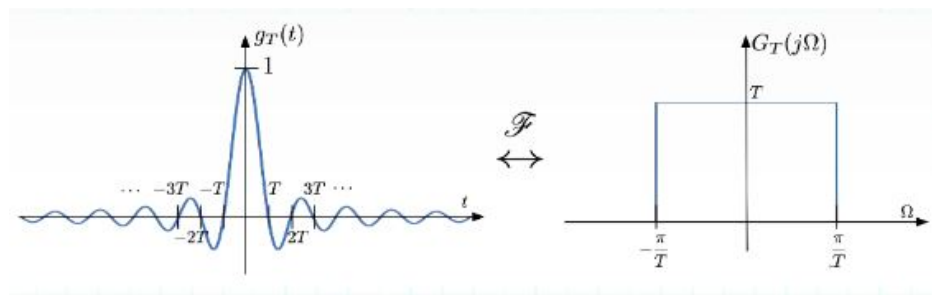


Figure 1.1: Sinc Function

The Fourier transform of the Sinc function is a rectangular box. We're basically taking our samples and filtering with an ideal LPF to ensure that we get our bandlimited signal back when we reconstruct it.

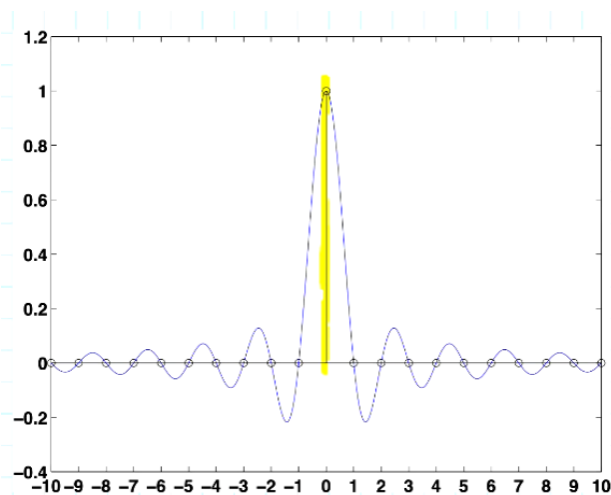


Figure 1.2: Unit Sample Reconstruction

Figure 1.2 Shows the continuous signal reconstructed from the discrete unit sample using sinc interpolation. The sinc perfectly passes through zero at all the points equal to zero in the original unit sample (which is all the points except $n=0$). If we were to resample this continuous signal once again, we would get the unit sample at $n=0$.

When we have multiple samples in our discrete signal and then reconstruct the continuous signal with sinc interpolation,

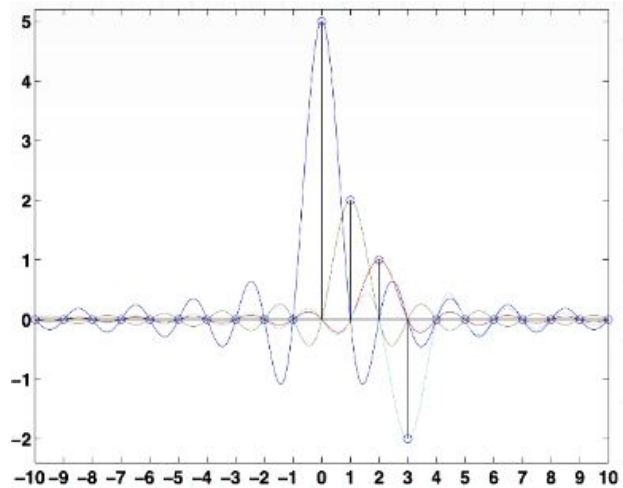


Figure 1.3: Multiple Samples Reconstruction

We sum the colored sinc functions (one for each unit sample) to get our reconstructed continuous signal. This may seem strange at first because we get little bumps added in like at $n = -2.5$. Intuitively, it would make more sense to just interpolate between each one of the unit samples; however, this version of the signal is bandlimited. While this sinc interpolation method isn't ideal for something like an image, it is best for a bandlimited signal, which is what we're typically concerned with.

Fundamental Theorem of DSP

If $x_c(t)$ is bandlimited to B ($X_c(j\Omega) = 0$ for $|\Omega| > B$), then it can be perfectly reconstructed from samples spaced $T \leq \pi/B$ apart.

$$x_c(t) = \sum_{n=-\infty}^{\infty} x[n]g_T(t - nT) \quad (1.7)$$

Where

$$x[n] = x_c(nT) \quad g_T(t) = \frac{\sin(\pi t/T)}{\pi t/T}$$

Lecture 2: Sampling pt 2

Frequency Domain Interpretation

So far, we have said that samples are simply the continuous-time signal evaluated at periodic points, separated by a distance T . We need to discuss some intermediate steps here.

Step 1: Multiply continuous-time signal by a delta function.

There are two types of delta functions:

- The Kronecker delta (discrete time) is 0 everywhere except $n=0$, where it is equal to one.
- The Dirac delta function (continuous time) integrates to 1.

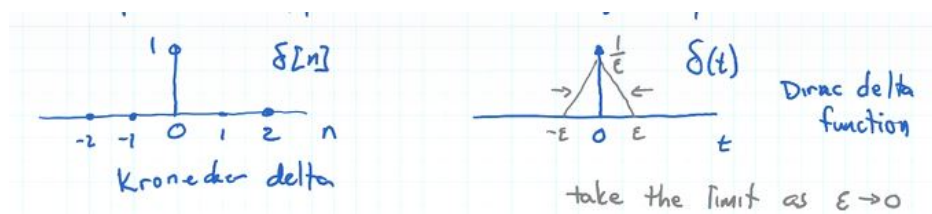


Figure 2.1: Kronecker and Dirac Delta Functions

The Dirac delta function has some interesting properties:

- $\int_{-\infty}^{\infty} \delta(t) dt = 1$
- $\delta(t) = 0$ if $t \neq 0$
- $x(t)\delta(t - t_0) = x(t_0)\delta(t - t_0)$

We define the DTFT as:

$$X(e^{j\omega}) \triangleq \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (2.1)$$

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x_c(nT)e^{-j\omega n}$$

We can express x_c in terms of its Fourier transform.

$$= \sum_{n=-\infty}^{\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} X_c(j\Omega)e^{-j\Omega nT} d\Omega \right) e^{-j\omega n}$$

Rearranging terms, we get

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_c(j\Omega) \left(\sum_{n=-\infty}^{\infty} e^{jn(\Omega T - \omega)} \right) d\Omega$$

Unfortunately, we hit a roadblock here in that we have an infinite sum. Luckily, we can leverage the Poisson summation formula:

$$\sum_{n=-\infty}^{\infty} e^{jn\omega} = 2\pi \sum_{k=-\infty}^{\infty} \delta(\omega - 2\pi k) \quad (2.2)$$

Plugging this in, we get

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} X_c(j\Omega) \left(\sum_{k=-\infty}^{\infty} \delta(\Omega T - \omega - 2\pi k) \right) d\Omega$$

Switching the order of summation and integration:

$$= \sum_{k=-\infty}^{\infty} \int_{\Omega} X_c(j\Omega) \delta(\Omega T - \omega - 2\pi k) d\Omega$$

We can swap $\Omega = (\omega + 2\pi k)/T$ because this causes the argument of the delta function to be zero, which is the only point where the delta function is nonzero.

$$= \frac{1}{T} \sum_{k=-\infty}^{\infty} X_c \left(j \frac{\omega + 2\pi k}{T} \right) \quad (2.3)$$

We're taking our original continuous time signal spectrum and summing infinitely many copies of it together, each one spaced by $2\pi k$.

The $1/T$ in the argument of X_c dilates the spectrum (stretches it out).

Lecture 3: Sampling pt 3

Frequency Domain Interpretation (continued)

B is the continuous-time frequency in radians. We get aliasing if $B > \pi/T$, i.e. $f_{max} < f_s/2$. In words, this is corruption of the spectrum by one of the copies of the period.

Last lecture, we discussed reconstructing signals using infinite sums. Regarding computational complexity, infinite sums are obviously not ideal, so we need an alternative.

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n]g_T(t - nT) \quad (3.1)$$

$$\Rightarrow X_r(j\Omega) = \sum_n x[n]G_T(j\Omega)e^{-j\Omega nT} \quad (3.2)$$

The Fourier transform is not a function of n , so it can come out front.

$$= G_T(j\Omega) \sum_n x[n]e^{-j\Omega nT}$$

If $\omega = \Omega T$, then this is the DTFT spectrum of $x[n]$. ΩT dilates the spectrum.

$$ssX_r(j\Omega) = G_T(j\Omega)X(e^{j\Omega T}) \quad (3.3)$$

The figures below outline the process of creating a continuous-time signal from discrete samples, viewed through the frequency domain.

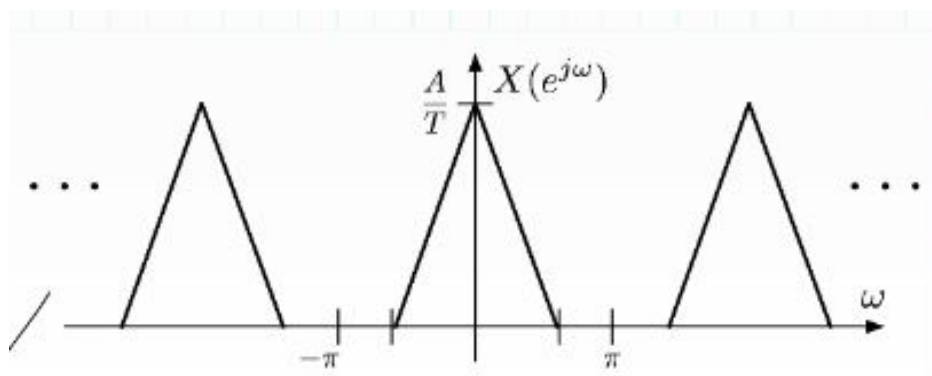
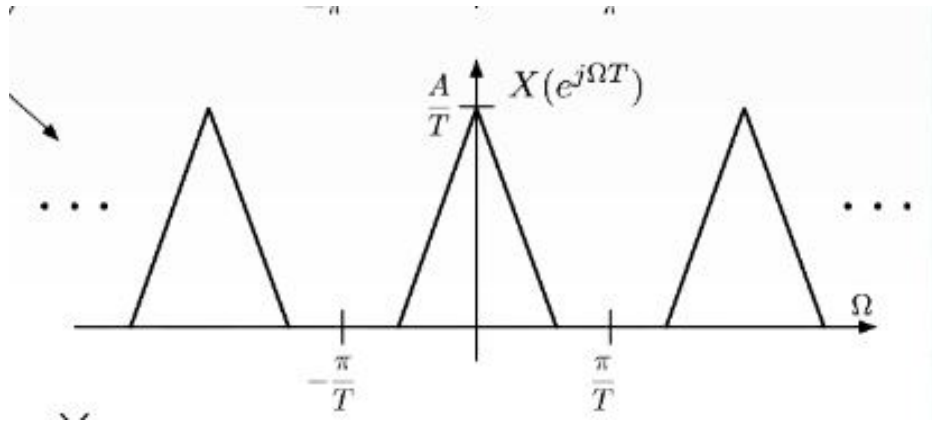


Figure 3.1: DTFT Spectrum

Figure 3.1 shows the DTFT spectrum of a sampled signal.

Figure 3.2: Spectrum of $x(nT)$

The first step in converting this signal to continuous-time is by turning the scaled and shifted Kronecker delta functions (1 per sample) into scaled and shifted Dirac delta functions (1 per sample).

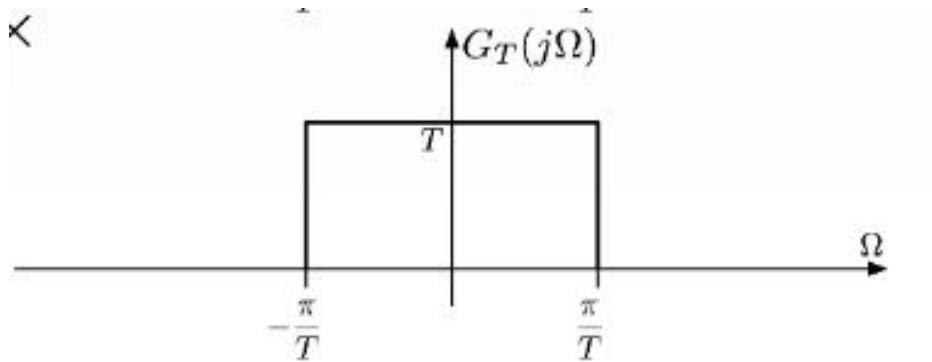


Figure 3.3: Interpolation Function Spectrum

Next, we filter these Dirac delta functions using our filter, i.e. we convolve them with our g_T function.

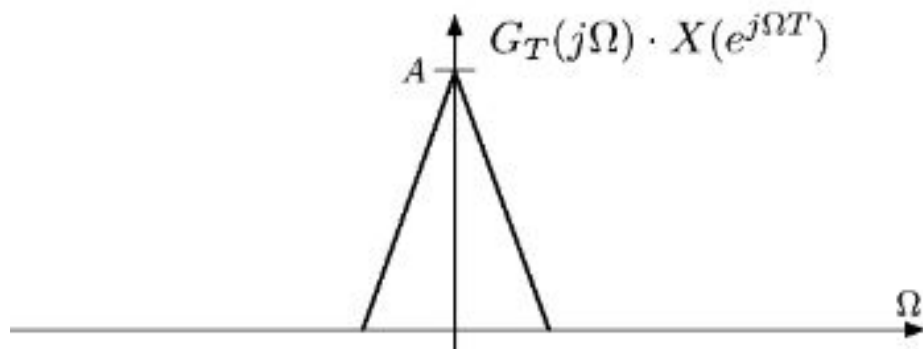


Figure 3.4: Continuous-Time Signal Spectrum

This gives us our continuous-time signal, the spectrum of which is shown in *Figure 3.4*.

As we have stated several times, the sinc function is infinite in length. There are a couple approximate reconstruction techniques we can use instead.

The first is "sample and hold". Draw a horizontal line at the height of the current sample, from the current sample to the next. This creates a piecewise continuous function, and yields a $G_T(j\Omega)$ that is a sinc in the frequency domain. This is not ideal - the frequency content of the signal becomes distorted because the passband isn't flat, and because we've added in frequency content from neighboring replications.

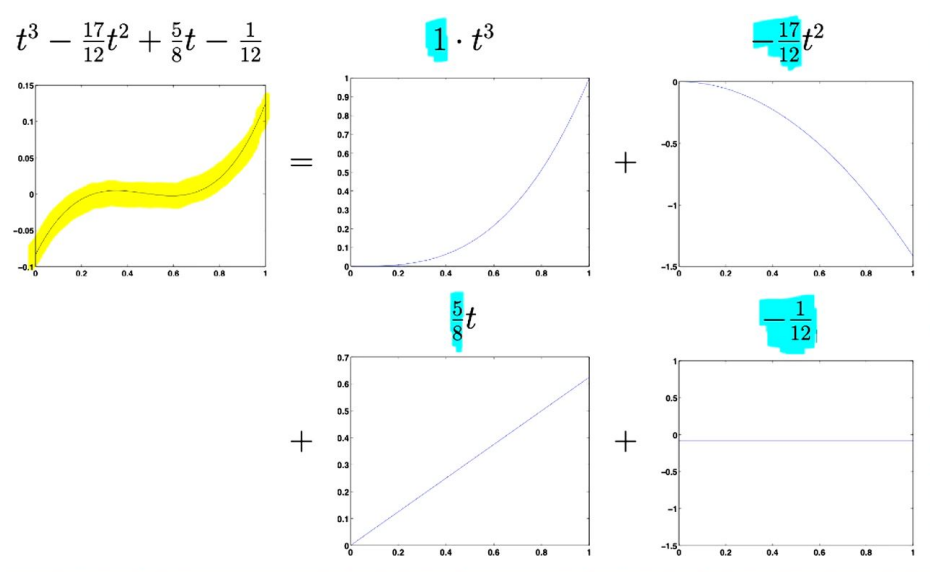
A better alternative is linear interpolation. The frequency domain transform of a linear interpolator is a triangle function. We're still distorting the spectrum, but less so than with sample and hold interpolation.

Whatever we choose as our interpolator, we can think of it as an impulse response of some filter in the frequency domain. One thing we can do to make it work better is ensure that we're sufficiently sampled.

Week 2

Lecture 4: Basis Expansions

Last time we found that we could sample a continuous time signal into a list of numbers $x[n]$, and perfectly reconstruct that continuous time signal, at least theoretically. This is really remarkable because we can map a sequence of infinite points to a countable number of points indexed by n , $x[n]$. But this only works because $x(t)$ is band-limited. So with that extra constraint, we know that we can take some continuous signal and we can turn it into a list of numbers that we can then analyze and operate on.



There are other ways that we can take a continuous signal and come up with a list of numbers. For example, in the leftmost block above, we have a continuous signal on the interval $[0, 1]$ represented by a polynomial. This polynomial signal is really just a list of four numbers, which are the coefficients for each of the powers of t . Although we're only using up to the power of three in this example, we could go up to infinite order. If we take our coefficients to be the k derivatives evaluated at zero all over k factorial, we come up with the Taylor series:

$$x(t) = \sum_{k=0}^{\infty} \alpha_k t^k \quad \text{if} \quad \alpha_k = \frac{x^{(k)}(0)}{k!} \quad (4.1)$$

Alternatively, we could use Lagrange Polynomials or B-Splines:

$$x(t) = \sum_{k=0}^M \alpha_k P_k(t) \quad (4.2)$$

We cannot take an arbitrary continuous signal and turn it into a list of numbers until we put some constraints on it, because we initially have an uncountable, infinite number of degrees of freedom. In the polynomial example, the constraint is a differentiable signal that is limited to the interval $[0,1]$.

Ultimate goal: Represent continuous signals in a discrete way to enable the use of linear algebra for processing

Lecture 5: Vector Spaces

S : a vector space composed of a set of elements, called vectors, and members of a field \mathbb{F} called scalars. Rules for addition and multiplication by scalars:

- Vector addition (+) combines two vectors into a third
- Scalar multiplication combines a scalar and vector to produce another vector

Addition: The '+' operation obeys four rules for all $\mathbf{x}, \mathbf{y} \in \mathbf{S}$

1. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ (commutative)
2. $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$ (associative)
3. There is a unique zero vector, $\mathbf{0}$, s.t. $\mathbf{x} + \mathbf{0} = \mathbf{x} \quad \forall \mathbf{x} \in \mathbf{S}$
4. For each $\mathbf{x} \in \mathbf{S}$, there is a unique vector called $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$

Scalar multiplication: For $a, b \in \mathbb{F}$ and $\mathbf{x}, \mathbf{y} \in \mathbf{S}$, scalar multiplication must obey

1. $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$ and $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$ (distributive)
2. $(ab)\mathbf{x} = a(b\mathbf{x})$ (associative)
3. for the multiplicative identity in \mathbb{F} , 1, we have $1\mathbf{x} = \mathbf{x} \quad \forall \mathbf{x} \in \mathbf{S}$
4. for the additive identity in \mathbb{F} , 0, we have $0\mathbf{x} = \mathbf{0}$

\mathbf{S} is closed under vector addition and scalar multiplication.

$$\boxed{\mathbf{x}, \mathbf{y} \in \mathbf{S} \Rightarrow a\mathbf{x} + b\mathbf{y} \in \mathbf{S} \quad \forall a, b \in \mathbb{F}} \quad (5.1)$$

This means that using the above operations, we will always get something back that is within our vector space. Some examples of vector spaces:

1. \mathbb{R}^N

$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad \text{where } x_i \text{ are real}$$

with scalars as real numbers, using standard mult. and addition

2. \mathbb{C}^N , same as above but with complex x_i and scalars.

3. Galois Field $GF(2)^N$, the scalar field is 0, 1, so vectors are lists of N bits. Addition is modulo 2.

$$0 + 0 = 0$$

$$0 + 1 = 1 + 0 = 1$$

$$1 + 1 = 0$$

$$\mathbf{x} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} \Rightarrow_{\text{Example}} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

4. Bounded, continuous functions $f(t)$ on the interval $[a, b]$ that are real-valued.

Vector addition \Rightarrow pointwise addition of functions

Scalar mult. \Rightarrow multiplying by $a \in \mathbb{R}$ pointwise

Example of **not** a vector space: bounded, continuous functions $f(t)$ on $[a, b]$ such that $|f(t)| < 2$. (Not closed)

Lecture 6: Linear Subspaces

A (non-empty) subset \mathbf{T} of \mathbf{S} is called a linear subspace of \mathbf{S} if

$$\forall a, b \in \mathbb{F}, \quad \mathbf{x}, \mathbf{y} \in \mathbf{T} \Rightarrow a\mathbf{x} + b\mathbf{y} \in \mathbf{T} \quad (6.1)$$

This implies $\mathbf{0} \in \mathbf{T}$. \mathbf{T} is also a vector space by itself.

Lecture 7: Linear Combinations and Spans

Linear Combinations

Suppose we have some linear combination of vectors

$$a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \cdots + a_m \mathbf{v}_m \quad \text{for some} \quad a_1, \dots, a_m \in \mathbb{F}$$

The span of a set of vectors in \mathbf{S} is the set of all vectors that are linear combinations of that set. We write it as

$$\text{Span}(\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\})$$

Suppose

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{S} \in \mathbb{R}^3, \quad \mathbb{F} = \mathbb{R}$$

All linear combinations of \mathbf{v}_1 and \mathbf{v}_2 can be written as

$$a \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix} \quad \text{for any} \quad x_1, x_2 \in \mathbb{R}$$

Linear Dependence

A set of vectors $\{\mathbf{v}_j\}_{j=1}^N$ is linearly dependent if $\sum_{n=1}^N a_n \mathbf{v}_n = 0$ for some set a_1, \dots, a_n (where not all a_n 's are zero)

Linear Independence

$\sum_{n=1}^N a_n \mathbf{v}_n = 0$ only for $a_n = 0 \quad \forall n$

Example 7.1

$$\mathbf{S} = \mathbb{R}^3, \quad \mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

Find a_1, a_2, a_3 such that $a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + a_3 \mathbf{v}_3 = 0$

Note, any two of these are linearly independent.

$$\text{span}(\{\mathbf{v}_1, \mathbf{v}_2\}) = \text{span}(\{\mathbf{v}_2, \mathbf{v}_3\}) = \text{span}(\{\mathbf{v}_1, \mathbf{v}_3\})$$

Example 7.2

Let \mathbf{S} be the set of continuous signals on $[0, 1]$

$$\mathbf{S} = \mathbf{C}([0, 1]) \quad \mathbf{v}_1 = \cos(2\pi t) \quad \mathbf{v}_2 = \sin(2\pi t) \quad \mathbf{v}_3 = 2\cos(2\pi t + \frac{\pi}{3})$$

Can we find a set $a_1, a_2, a_3 \neq 0$ such that the linear combination goes to zero? (i.e. are they linearly dependent)?

Suppose we have some set $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ are dependent. Then

$$\sum_n a_n \mathbf{v}_n = 0 \Rightarrow \mathbf{v}_k = \frac{1}{a_k} \sum_{n \neq k} a_n \mathbf{v}_n \quad \text{for } a_k \neq 0 \quad (7.1)$$

at least one vector may be removed without changing the span. We can repeat until we have a linearly independent set.

Basis

A basis of a finite-dimensional vector space \mathbf{S} is a set of vectors \mathbf{B} such that

1. $\text{span}(\mathbf{B}) = \mathbf{S}$
2. \mathbf{B} is linearly independent

The second condition ensures bases of \mathbf{S} will have the same number of elements.

The dimension of \mathbf{S} is the number of elements required in a basis for \mathbf{S} .

Example 7.3

$$\mathbf{S} = \{\text{polynomials of degree at most } p\}$$

A basis for \mathbf{S} is $\mathbf{B} = \{1, t, t^2, \dots, t^p\}$.

The dimension of \mathbf{S} is $p + 1$.

Lecture 8: Norms and Inner Products

Norms

A norm $\|\cdot\|$ on a vector space \mathbf{S} is a mapping

$$\|\cdot\| : \mathbf{S} \rightarrow \mathbb{R}$$

with the following properties for all $\mathbf{x}, \mathbf{y} \in \mathbf{S}$

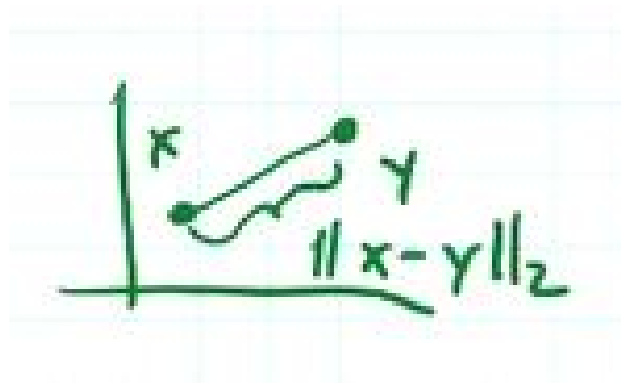
1. $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$
2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (triangle inequality)
3. $\|a\mathbf{x}\| = |a| \cdot \|\mathbf{x}\|$ for any scalar a (homogeneity)

Other definitions

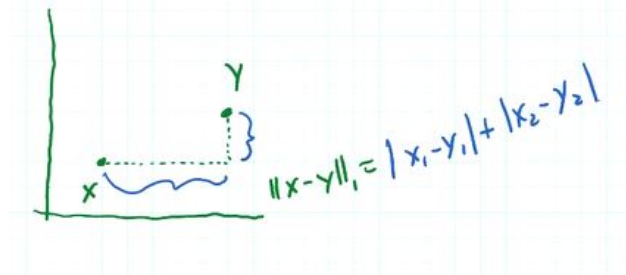
- The length of $\mathbf{x} \in \mathbf{S}$ is simply $\|\mathbf{x}\|$
- The distance between \mathbf{x} and \mathbf{y} is $\|\mathbf{x} - \mathbf{y}\|$
- A linear vector space equipped with a norm is called a normed vector space

Examples

$$\mathbf{S} = \mathbb{R}^N \quad \|\mathbf{x}\|_2 = \left(\sum_{n=1}^N |\mathbf{x}_n|^2 \right)^{1/2} \quad (\text{Euclidean or } \ell_2\text{-norm})$$

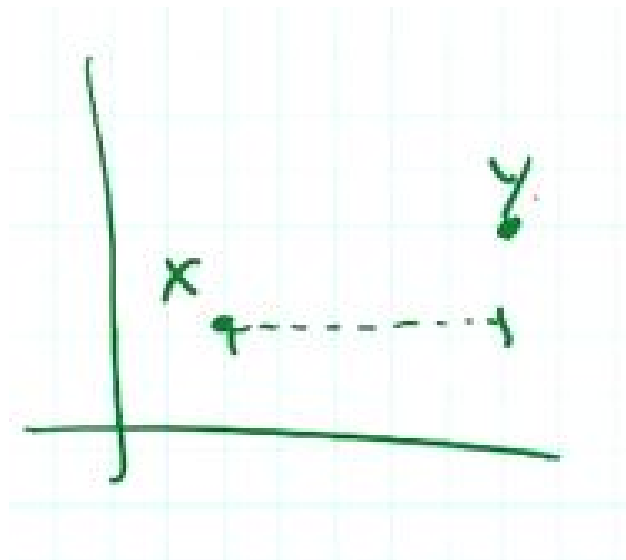


$$\mathbf{S} = \mathbb{R}^N \quad \|\mathbf{x}\|_1 = \sum_{n=1}^N |\mathbf{x}_n| \quad (\text{Manhattan norm})$$

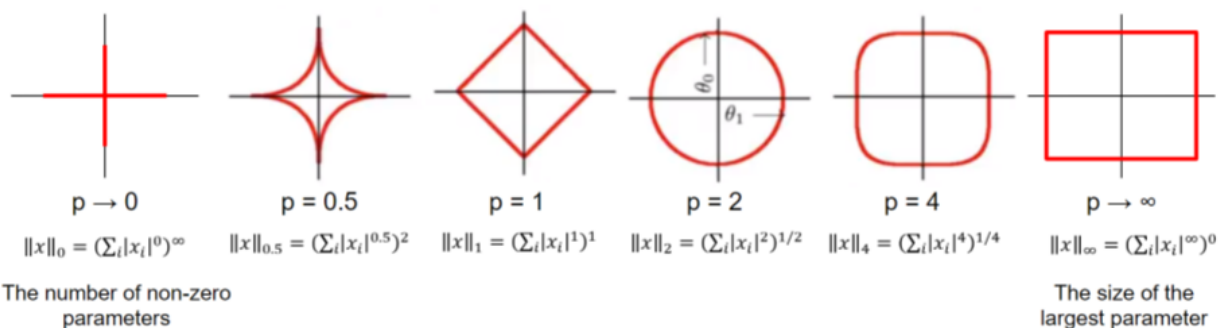


$$\mathbf{S} = \mathbb{R}^N \quad \|\mathbf{x}\|_p = \left(\sum_{n=1}^N |\mathbf{x}_n|^p \right)^{1/p} \text{ for some } 1 \leq p \leq \infty \quad (\ell_p \text{ norm})$$

$$\mathbf{S} = \mathbb{R}^N \quad \|\mathbf{x}\|_\infty = \max_{1, \dots, N} |\mathbf{x}_n| \quad (\ell_\infty \text{ norm})$$



We can plot the unit ball of each norm as follows:



What if $p=0$? ℓ_0 is not a norm, it counts the number of nonzero elements.

$$\text{Ex. let } \mathbf{x} = \begin{bmatrix} 3 \\ 2 \\ 0 \\ 0 \end{bmatrix} \text{ and } a = 2. \quad ||a\mathbf{x}||_0 = 2 \neq |a| \cdot ||\mathbf{x}||_0 = 2 \cdot 2 = 4$$

Inner Products

An inner product on a real or complex-valued vector space \mathbf{S} is a mapping

$$\langle \cdot, \cdot \rangle : \mathbf{S} \times \mathbf{S} \rightarrow \mathbb{C}$$

That obeys

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$ (Conjugate symmetry)
2. $\langle a\mathbf{x} + b\mathbf{y}, \mathbf{z} \rangle = a\langle \mathbf{x}, \mathbf{z} \rangle + b\langle \mathbf{y}, \mathbf{z} \rangle$ for any $a, b \in \mathbb{C}$ (Linearity)
3. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0}$ (Positive definiteness)

Common examples

- $\mathbf{S} = \mathbb{R}^N$ $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^N x_n y_n = \mathbf{y}^T \mathbf{x}$
(All vectors in this class are column vectors unless otherwise noted)
- $\mathbf{S} = \mathbb{C}^N$ $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^N x_n \overline{y_n} = \mathbf{y}^H \mathbf{x}$
- $\mathbf{S} = L_2[a, b]$ $\langle \mathbf{x}, \mathbf{y} \rangle = \int_a^b x(t) \overline{y(t)} dt$

Other (less common) examples

- \mathbf{S} = Gaussian RVs with zero-mean and finite variance

$$\langle \mathbf{x}, \mathbf{y} \rangle = E[\mathbf{xy}]$$

- \mathbf{S} = Differentiable, real-valued continuous signals on \mathbb{R}

$$\langle \mathbf{x}, \mathbf{y} \rangle = \int x(t)y(t)dt + \int x'(t)y'(t)dt$$

Induced norms A valid inner product induces a valid norm as

$$||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Induced norms are norms (they obey all the properties we introduced previously). They also have additional properties.

1. Cauchy-Schwarz Inequality

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$$

Equality is only achieved when \mathbf{x} and \mathbf{y} are colinear, meaning
 $(\forall a \in \mathbb{C}, a \neq 0, \text{ such that } \mathbf{y} = a\mathbf{x})$

2. Pythagorean Theorem

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0 \Rightarrow \|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

Side note:

$$\langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$$

And we can also prove that

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

3. Parallelogram Law

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2$$

4. Polarization Identity

$$\operatorname{Re}\{\langle \mathbf{x}, \mathbf{y} \rangle\} = \frac{\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2}{4}$$

Aside: The angle between two vectors

$$|\langle \mathbf{x}, \mathbf{y} \rangle| = \|\mathbf{x}\| \cdot \|\mathbf{y}\| \cos(\Theta)$$

We can define the angle between two vectors \mathbf{x} and \mathbf{y} in any inner product space as

$$\cos(\Theta) = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} \text{ if } \langle \cdot, \cdot \rangle \text{ induces } \|\cdot\|$$

We can go even further and define orthogonality : Vectors \mathbf{x} and \mathbf{y} in an inner product space are orthogonal to one another if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$

Bonus: Plotting a Unit Ball

We define some inner product $\langle \cdot, \cdot \rangle_R$ as:

$$\mathbf{x} \in \mathbb{R}^2 \quad \langle \mathbf{x}, \mathbf{y} \rangle_A = \mathbf{y}^T A \mathbf{x} \text{ where } A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

With norm

$$||\mathbf{x}||_A = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_A}$$

The question is: how do we plot the unit ball?

$$||\mathbf{x}||_A = (x_1 \ x_2) \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = ax_1^2 + 2bx_1x_2 + cx_2^2 = 1$$

Then, we need to plot this equation. This will either be an ellipse or hyperbola.

We get the former if A is positive definite.

For $A > 0$, A has unique, orthogonal eigenvectors.

Let $\mathbf{x} = a\mathbf{v}_1 + b\mathbf{v}_2 \Rightarrow$ solve $\mathbf{x}^T A \mathbf{x} = 1$

$$\begin{aligned} (a\mathbf{v}_1 + b\mathbf{v}_2)^T A (a\mathbf{v}_1 + b\mathbf{v}_2) &= (a\mathbf{v}_1 + b\mathbf{v}_2)^T (a\lambda_1\mathbf{v}_1 + b\lambda_2\mathbf{v}_2) \\ &= a^2\lambda_1 \cancel{\mathbf{v}_1^T \mathbf{v}_1} + \cancel{ab\lambda_2 \mathbf{v}_1^T \mathbf{v}_2} + \cancel{ba\lambda_1 \mathbf{v}_2^T \mathbf{v}_1} + b^2\lambda_2 \cancel{\mathbf{v}_2^T \mathbf{v}_2} \\ &= \boxed{a^2\lambda_1 + b^2\lambda_2 = 1} \end{aligned}$$

A solution to this is to let $a = \frac{\cos(\Theta)}{\sqrt{\lambda_1}}$ and $b = \frac{\sin(\Theta)}{\sqrt{\lambda_2}}$

$$a^2\lambda_1 = \cos^2(\Theta), \quad b^2\lambda_2 = \sin^2(\Theta), \quad a^2\lambda_1 + b^2\lambda_2 = \cos^2(\Theta) + \sin^2(\Theta) = 1$$

$$\mathbf{x} = a\mathbf{v}_1 + b\mathbf{v}_2 = \frac{\cos(\Theta)}{\sqrt{\lambda_1}}\mathbf{v}_1 + \frac{\sin(\Theta)}{\sqrt{\lambda_2}}\mathbf{v}_2$$

Let $\Theta \in [0, 2\pi]$

Finally, we can plot the unit ball using code.

```

1 % parametric ellipse plot
2 omega = -pi:0.01:pi;
3 w0=cos(omega);
4 w1=sin(omega);
5
6 A=[2.0 0.5;
7 0.5 2.0]
8 [v,lambda]=eig(A)
9 c=v(:,1)*cos(omega)/sqrt(lambda(1,1))+v(:,2)*sin(omega)/sqrt(lambda
    (2,2));
10 plot(c(1,:),c(2,:), 'r')
11 hold on
12 grid on;
13 fill(c(1,:),c(2,:), 'b')
```

Week 3

Lecture 9: Approximations in Hilbert Spaces pt 1

Hilbert and Banach Spaces

We talk about completeness of a vector space \rightarrow no missing points. If an infinite sequence of vectors, increasingly close together exists in \mathbf{S} , it will converge to something in \mathbf{S} .

Ex.

$$[0, 1) \leftarrow \text{not complete.} \quad 1 - \frac{1}{x} \xrightarrow{x \rightarrow \infty} 1$$

We do not have a complete space because vectors can get arbitrarily close to 1, but 1 is still not in the space.

$$[0, 1] \leftarrow \text{may be complete}$$

A normed linear space is complete if every Cauchy sequence is a convergent sequence; that is, for every sequence $\mathbf{x}_1, \mathbf{x}_2, \dots \in \mathbf{S}$ for which

$$\lim_{\min(m,n) \rightarrow \infty} \|\mathbf{x}_m - \mathbf{x}_n\| = 0$$

will also have

$$\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}^* \in \mathbf{S}$$

where $\|\cdot\|$ is the norm induced by the inner product in the case that \mathbf{S} is an inner product space.

In a nutshell, this is saying that as m, n get really large, the \mathbf{x} vectors get closer together. The limit of these vectors squishing together is \mathbf{x}^* , and we are saying that \mathbf{x}^* is also in our vector space \mathbf{S} .

- A normed linear space which is also complete is called a Banach Space.
- An inner product space which is also complete is called a Hilbert Space.

We want to know that it makes sense to write expressions like

$$x(t) = \sum_{n=1}^{\infty} \alpha_n \psi_n(t)$$

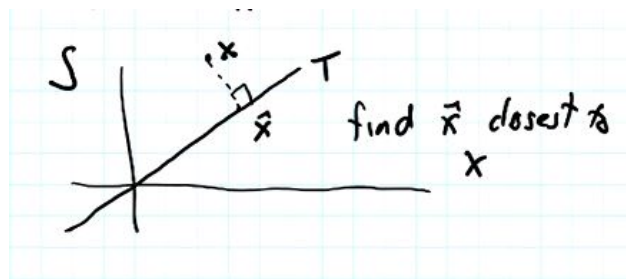
an infinite dimensional space with basis $\psi(n)$.

Approximating Functions in Hilbert Spaces

Theorem

Let S be a Hilbert space, and let T be a finite-dimensional subspace. Given an arbitrary $\mathbf{x} \in S$,

1. There is exactly one $\hat{\mathbf{x}} \in T$ such that $\mathbf{x} - \hat{\mathbf{x}} \perp T$, meaning $\langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{y} \rangle = 0$ for all $\mathbf{y} \in T$
2. This $\hat{\mathbf{x}}$ is the closest point in T to \mathbf{x} ; that is, $\hat{\mathbf{x}}$ is the unique minimizer to $\|\mathbf{x} - \hat{\mathbf{x}}\|$

Proof

Start with 2.:

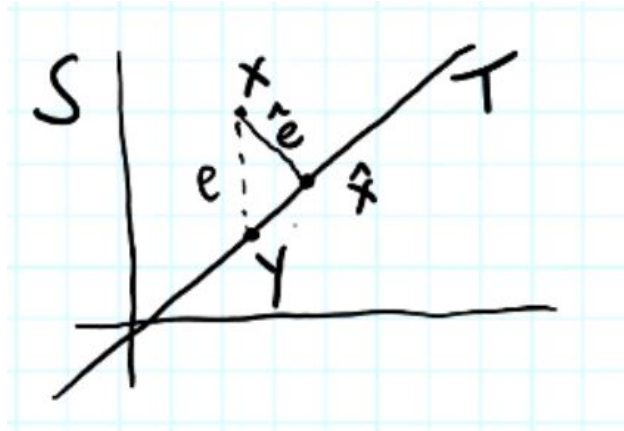
Let $\hat{\mathbf{x}}$ be the vector that satisfies $\hat{\mathbf{e}} = \mathbf{x} - \hat{\mathbf{x}} \perp T$

Let \mathbf{y} be any other vector in T , $\mathbf{e} := \mathbf{x} - \mathbf{y}$

We will show that $\|\mathbf{e}\| > \|\hat{\mathbf{e}}\|$

$$\begin{aligned}
 \|\mathbf{e}\| &= \|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{y} + \hat{\mathbf{x}} - \hat{\mathbf{x}}\|^2 = \|\mathbf{x} - \mathbf{y} + \hat{\mathbf{x}} + \hat{\mathbf{e}} - \mathbf{x}\|^2 = \|\hat{\mathbf{e}} - \mathbf{y} + \hat{\mathbf{x}}\|^2 = \|\hat{\mathbf{e}} - (\mathbf{y} - \hat{\mathbf{x}})\|^2 \\
 &= \langle \hat{\mathbf{e}} - (\mathbf{y} - \hat{\mathbf{x}}), \hat{\mathbf{e}} - (\mathbf{y} - \hat{\mathbf{x}}) \rangle \\
 &= \|\hat{\mathbf{e}}\|^2 + \underbrace{\|\mathbf{y} - \hat{\mathbf{x}}\|^2}_{\text{not zero}} - \cancel{\langle \hat{\mathbf{e}}, \mathbf{y} - \hat{\mathbf{x}} \rangle} - \cancel{\langle \mathbf{y} - \hat{\mathbf{x}}, \hat{\mathbf{e}} \rangle} \xrightarrow{0}
 \end{aligned}$$

$\therefore \|\mathbf{e}\| > \|\hat{\mathbf{e}}\|$. Geometrically:



Lecture 10: Approximations in Hilbert Spaces pt 2

Best Approximation

We proved there is a unique minimizer (clause 2 from the theorem) in the previous lecture. Now we will prove the first clause.

We compute $\hat{\mathbf{x}}$, the optimal point.

Let N be the dimension of the subspace T , and let $\mathbf{v}_1, \dots, \mathbf{v}_N$ be a basis for T .

We want to find the set of coefficients $a_1, \dots, a_N \in \mathbb{C}$ such that

$$\hat{\mathbf{x}} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_N \mathbf{v}_N$$

From the orthogonality principle:

$$\langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{v}_n \rangle = 0 \quad \text{for } n = 1, \dots, N$$

so

$$\underbrace{\langle \mathbf{x} - \sum_{k=1}^N a_k \mathbf{v}_k, \mathbf{v}_n \rangle}_{\hat{\mathbf{x}}} = 0 \quad \text{for } n = 1, \dots, N$$

then

$$\langle \mathbf{x}, \mathbf{v}_n \rangle = \sum_{k=1}^N a_k \langle \mathbf{v}_k, \mathbf{v}_n \rangle \quad \text{for } n = 1, \dots, N$$

which gives us n linear equations.

$$\underbrace{\begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_N, \mathbf{v}_1 \rangle \\ \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_N, \mathbf{v}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{v}_1, \mathbf{v}_N \rangle & \langle \mathbf{v}_2, \mathbf{v}_N \rangle & \cdots & \langle \mathbf{v}_N, \mathbf{v}_N \rangle \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}}_{\mathbf{a}} = \underbrace{\begin{bmatrix} \langle \mathbf{x}, \mathbf{v}_1 \rangle \\ \langle \mathbf{x}, \mathbf{v}_2 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{v}_N \rangle \end{bmatrix}}_{\mathbf{b}}$$

Where \mathbf{G} is the Gram matrix or Grammian of the basis \mathbf{v}_n .

$$\mathbf{G}\mathbf{a} = \mathbf{b} \Rightarrow \mathbf{a} = \mathbf{G}^{-1}\mathbf{b}$$

- \mathbf{G} is invertible because \mathbf{v}_n are linearly independent
- \mathbf{G} is conjugate symmetric (Hermitian) i.e. $\mathbf{G} = \mathbf{G}^H$

There is a unique \mathbf{a} therefore, only on $\hat{\mathbf{x}}$ that satisfies $\mathbf{x} - \hat{\mathbf{x}} \perp T$

Lecture 11: Approximation Example

Approximation Example

Let $S = \mathbb{R}^3$, with standard inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=1}^3 x_n y_n$ and the induced norm $\|\cdot\|_2$

$$T = \text{Span} \left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right), \quad \mathbf{x} = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix}$$

Find $\hat{\mathbf{x}} \in T$ that is closest to \mathbf{x} .

We start by building the Gram matrix:

$$\mathbf{G} = \begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_1 \rangle & \cdots & \langle \mathbf{v}_N, \mathbf{v}_1 \rangle \\ \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle & \cdots & \langle \mathbf{v}_N, \mathbf{v}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{v}_1, \mathbf{v}_N \rangle & \langle \mathbf{v}_2, \mathbf{v}_N \rangle & \cdots & \langle \mathbf{v}_N, \mathbf{v}_N \rangle \end{bmatrix} = \begin{bmatrix} \langle \mathbf{v}_1, \mathbf{v}_1 \rangle & \langle \mathbf{v}_2, \mathbf{v}_1 \rangle \\ \langle \mathbf{v}_1, \mathbf{v}_2 \rangle & \langle \mathbf{v}_2, \mathbf{v}_2 \rangle \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \langle \mathbf{x}, \mathbf{v}_1 \rangle \\ \langle \mathbf{x}, \mathbf{v}_2 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{v}_N \rangle \end{bmatrix} = \begin{bmatrix} \langle \mathbf{x}, \mathbf{v}_1 \rangle \\ \langle \mathbf{x}, \mathbf{v}_2 \rangle \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\mathbf{a} = \mathbf{G}^{-1} \mathbf{b}$$

Refresher: The inverse of a 2x2 matrix is

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Here

$$\begin{aligned} \mathbf{G}^{-1} &= \frac{1}{8} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \\ \mathbf{G}^{-1} \mathbf{b} &= \frac{1}{8} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 3 \\ -1 \end{bmatrix} \\ \hat{\mathbf{x}} &= \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix} \end{aligned}$$

$\hat{\mathbf{x}}$ is our solution to find the best approximation in the subspace.

We can double check:

$$\mathbf{x} - \hat{\mathbf{x}} = \begin{bmatrix} -2.5 \\ 0 \\ 2.5 \end{bmatrix}$$

$$\langle \mathbf{x} - \hat{\mathbf{x}}, \hat{\mathbf{x}} \rangle = \begin{bmatrix} -2.5 & 0 & 2.5 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{2} \\ 1 \\ \frac{1}{2} \end{bmatrix} = 0.$$

Lecture 12: Orthobases

Orthogonal Bases

A collection of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ in a finite-dimensional vector space S is called an orthogonal basis if

1. $\text{span}(\{\mathbf{v}_1, \dots, \mathbf{v}_n\}) = S$
2. $\mathbf{v}_j \perp \mathbf{v}_k$ (i.e. $\langle \mathbf{v}_j, \mathbf{v}_k \rangle = 0$ for all $j \neq k$)

in addition, if the vectors are normalized

$$\|\mathbf{v}_n\| = 1 \text{ for } n = 1, \dots, N$$

we have an orthonormal basis or orthobasis.

Examples

1. $S = \mathbb{R}^2$ with standard inner product

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

the $\frac{1}{\sqrt{2}}$ normalizes the vectors.

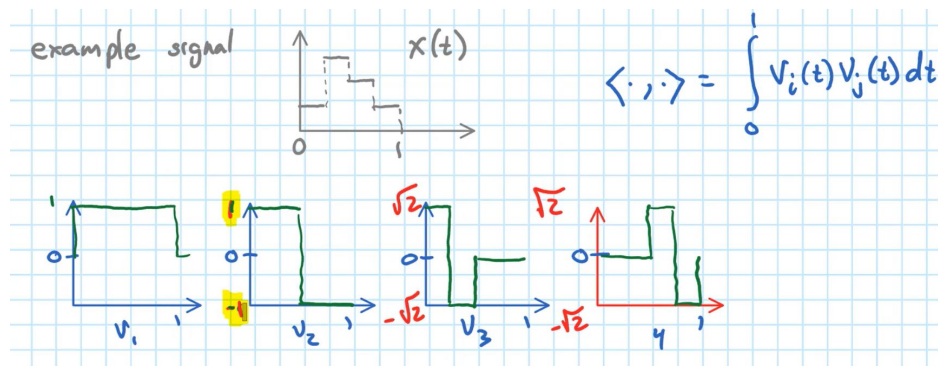
Alternative example:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

2. $S =$ space of piecewise constant functions on $[0, 1/4), [1/4, 1/2), [1/2, 3/4), [3/4, 1]$.

We define the inner product as

$$\langle \cdot, \cdot \rangle = \int_0^1 v_i(t) v_j(t) dt$$



One way to construct a basis for the example signal in grey is to just use one basis signal for every level $x(t)$. An alternative is to use the basis functions at the bottom of the image.

3. $\{\mathbf{v}_k(t) = \frac{1}{\sqrt{2\pi}}e^{jkt}, k \in \mathbb{Z}\}$ is an orthobasis for $L_2([0, 2\pi])$
(standard inner product)

$$\langle \mathbf{v}_k(t), \mathbf{v}_\ell(t) \rangle = \frac{1}{2\pi} \int_0^{2\pi} e^{jt(k-\ell)} dt = \begin{cases} 1 & k = \ell \\ 0 & k \neq \ell \end{cases}$$

When $k \neq \ell$, we integrate over the entire period of a sinusoid, which cancels out to zero.

4. $B_{\frac{\pi}{T}}(\mathbb{R}) =$ real-valued functions which are bandlimited to $\frac{\pi}{T}$ (standard inner product) The set of functions

$$\left\{ v_n(t) = \sqrt{T} \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)}, n \in \mathbb{Z} \right\}$$

is an orthobasis for $B_{\frac{\pi}{T}}$

This says that

$$x(t) = \sum_n \underbrace{x[n]}_{a_n} \sqrt{T} \frac{\sin(\pi(t - nT)/T)}{\pi(t - nT)}$$

(reconstruction of a signal from samples)

Lecture 13: Orthobasis Expansions

Linear Approximation with Orthobases

To approximate $\mathbf{x} \in S$ with $\hat{\mathbf{x}} \in T$ where S is a Hilbert space with subspace T : if $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis for T then $\hat{\mathbf{x}} = a_1\mathbf{v}_1 + \dots + a_n\mathbf{v}_n$ and $\mathbf{a} = G^{-1}\mathbf{b}$ as defined in lecture 10.

If $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is orthonormal then $G = I$

$$\underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}}_{\mathbf{a}} = \underbrace{\begin{bmatrix} \langle \mathbf{x}, \mathbf{v}_1 \rangle \\ \langle \mathbf{x}, \mathbf{v}_2 \rangle \\ \vdots \\ \langle \mathbf{x}, \mathbf{v}_N \rangle \end{bmatrix}}_{\mathbf{b}} \quad \text{i.e. } a_i = \langle \mathbf{x}, \mathbf{v}_i \rangle \quad \underline{\text{Only for an orthobasis!}}$$

$$\hat{\mathbf{x}} = \sum_{k=1}^n \langle \mathbf{x}, \mathbf{v}_k \rangle \mathbf{v}_k$$

Orthobasis expansion

If T is the whole space S , given $\mathbf{x} \in S$, the closest point in T is the point itself.

$$\mathbf{x} = \sum_{n=1}^N \langle \mathbf{x}, \mathbf{v}_n \rangle \mathbf{v}_n \quad \text{for all } \mathbf{x} \in S$$

also works for infinite-dimensional spaces as long as

$$\sum_{n=-\infty}^{\infty} |\langle \mathbf{x}, \mathbf{v}_n \rangle|^2 < \infty \Rightarrow \mathbf{x} = \sum_{n=-\infty}^{\infty} \langle \mathbf{x}, \mathbf{v}_n \rangle \mathbf{v}_n$$

Lecture 14: Parseval's Theorem++

Let S be a Hilbert space with an inner product $\langle \cdot, \cdot \rangle_S$ which induces a norm $\|\cdot\|_S$. Let $\{\mathbf{v}_k\}_k$ be an orthobasis for S .

Then for every $\mathbf{x}, \mathbf{y} \in S$,

$$\langle \mathbf{x}, \mathbf{y} \rangle_S = \sum_k \alpha_k \bar{\beta}_k = \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle_{\ell_2}$$

Where $\alpha_k = \langle \mathbf{x}, \mathbf{v}_k \rangle_S$ and $\beta_k = \langle \mathbf{y}, \mathbf{v}_k \rangle_S$. We call them the transform coefficients. The significance of this is that

$$\langle \mathbf{x}, \mathbf{y} \rangle_S = \langle \boldsymbol{\alpha}, \boldsymbol{\beta} \rangle_{\ell_2}$$

An orthobasis can make every Hilbert space equivalent to ℓ_2 . With Parseval's theorem, we are used to seeing:

$$\|\mathbf{x}\|_S^2 = \|\boldsymbol{\alpha}\|_2^2$$

Proof sketch:

$$\begin{aligned} \mathbf{x} &= \sum_k \alpha_k \mathbf{v}_k & \mathbf{y} &= \sum_k \beta_k \mathbf{v}_k \\ \langle \mathbf{x}, \mathbf{y} \rangle_S &= \left\langle \sum_k \alpha_k \mathbf{v}_k, \sum_\ell \beta_\ell \mathbf{v}_\ell \right\rangle = \sum_k \alpha_k \left\langle \mathbf{v}_k, \sum_\ell \beta_\ell \mathbf{v}_\ell \right\rangle \\ &= \sum_k \alpha_k \sum_\ell \beta_\ell \underbrace{\langle \mathbf{v}_k, \mathbf{v}_\ell \rangle}_a \end{aligned}$$

Where the inner product denoted by $a = 0$ for $k \neq \ell$ and $a = 1$ for $k = \ell$

$$\boxed{\langle \mathbf{x}, \mathbf{y} \rangle_S = \sum_k \alpha_k \bar{\beta}_k}$$

Week 4