Last time Projection as Best Approximation

With an inner product space V and a finite-dimensional subspace W, if v is a vector in V, then the "best approximation to v in W" is the vector $\boxed{\text{proj}_W v}$.

This means $\|\text{proj}_W v - v\| < \|u - v\|$ for any other vector u in W.

We saw how to use this to get approximate solutions — called "least squares solutions" — to linear systems $Ax=b$ using normal equations $\boxed{A^T A x = A^T b}$.

Fourier Series — an important application of "Projection as best approx."

approximating functions

Set $V = F((-\infty,\infty))$ — vector space of all functions on $\mathbb{R}$

We're given f in V.

We want to approximate f using functions from some "nice" collection of functions e.g. by a function in say

(1) polynomials of degree ≤ 2

(2) functions of form $a_0 + a_1 e^x + a_2 e^{2x} + a_3 e^{3x}$

(3) sums of sine and cosine waves

(1) is span of $\{1, x, x^2\}$

(2) is span of $\{1, e^x, e^{2x}, e^{3x}\}$

(3) is (say) span of $\{1, \sin(x), \cos(x), \sin(2x), \cos(2x), \ldots, \sin(mx), \cos(mx)\}$ for some natural # m.

In each case the subspace given would play the role of W in our setup.

Goal would be: find $\operatorname{proj}_W f$, the best approximation to f in W.

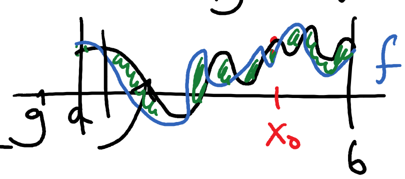
That is $\operatorname{proj}_W f = g$ is the function in W which minimizes

$$\|g-f\| = \sqrt{\frac{1}{b-a}\int_a^b \big((g-f)(x)\big)^2\,dx}$$

OK so we have to be careful and work on some interval $[a,b]$ i.e. start with f in $F([a,b])$.

The square of this $\|g-f\|^2$ is called the "<u>mean square error</u>"

Notice: we are <u>not</u> trying to minimize $|g(x_0)-f(x_0)|$ for some fixed x_0 in $[a,b]$



instead looking at area

– "average" difference between f and g across $[a,b]$.

Let's look more closely at Example (3) above.

i.e. $W_m = \operatorname{span}\{1, \sin(x), \cos(x), \dots, \sin(mx), \cos(mx)\}$.

i.e. linear combinations

$a_0 + a_1\sin(x) + b_1\cos(x) + \dots + a_m\sin(mx) + b_m\cos(mx)$

↳ These are called a <u>trigonometric sum</u> / trigonometric polynomial

... of order m if not both a_m and b_m are zero

<u>Remark</u> These trig. sums are linear combinations of sine and cosine waves of different frequencies ($\sin(2x)$, $\underset{[nx]}{\sin(3x)\text{ etc.}}$) and amplitudes ($5\cos(2x)$, $6\cos(2x)$) $\left[\begin{smallmatrix} a_n, \\ b_n \end{smallmatrix}\right]$

When working with signals $f(x)$, if we understand its "frequency components" (individual terms in trig. sum approximating f) we can e.g. remove noise, simplify transmission (eg for data compression/reconstitution).

Also can replace need to solve differential equations in <u>time domain</u> (x) with linear equations in frequency domain.

If we want to approximate f in $C([0,2\pi])$ with a trig. sum of order m, i.e. a function in W_m — we look for $\text{proj}_{W_m} f$.

To do so, we need an orthogonal basis for W_m.

It turns out that $\{1, \sin(x), \cos(x), \ldots, \sin(mx), \cos(mx)\}$ is in fact orthogonal! Why? (Below.)

So the set is also linearly independent (by earlier result) hence basis for its spanning set W_n.

We need to check

$$\langle 1, \sin(nx)\rangle = \underset{(n\neq 0)}{\langle 1, \cos(nx)\rangle} = \langle \sin(nx), \cos(kx)\rangle =$$

$$\underset{(n\neq k)}{\langle \sin(nx), \sin(kx)\rangle} = \underset{(n\neq k)}{\langle \cos(nx), \overset{\cos}{\cancel{\sin}}(kx)\rangle} = 0$$

$$\langle 1, \sin(nx)\rangle = \frac{1}{2\pi}\int_0^{2\pi} \sin(nx)\,dx = \frac{-1}{2\pi}\left[\frac{1}{n}\cos(nx)\right]_0^{2\pi} = 0$$

$$\langle 1, \cos(nx)\rangle = \frac{1}{2\pi}\int_0^{2\pi} \cos(nx)\,dx = \frac{1}{2\pi}\left[\frac{1}{n}\sin(nx)\right]_0^{2\pi} = 0.$$

For the other 3 types, recall that

$$\sin A \cos B = \frac{1}{2}(\sin(A-B) + \sin(A+B))$$
$$\sin A \sin B = \frac{1}{2}(\cos(A-B) - \cos(A+B))$$
$$\cos A \cos B = \frac{1}{2}(\cos(A-B) + \cos(A+B))$$

So we can use what we just found i.e.

$$0 = \underbrace{\int_0^{2\pi} \sin(nx)\,dx}_{\text{for any } n} = \underbrace{\int_0^{2\pi} \cos(nx)\,dx}_{\text{for } n \neq 0}$$ to see that the rest are 0

$(A+B = (n_1+n_2)x$ and $A-B = (n_1-n_2)x$, $A \neq B)$.

Then by our earlier formula (can you write it down?):

$$\text{proj}_{W_m} f = \frac{\langle f, 1\rangle}{\|1\|^2} 1 + \frac{\langle f, \sin(x)\rangle}{\|\sin(x)\|^2}\sin(x) + \frac{\langle f, \cos x\rangle}{\|\cos(x)\|^2}\cos(x) + \dots + \frac{\langle f, \sin(mx)\rangle}{\|\sin(mx)\|^2}\sin(mx) + \frac{\langle f, \cos(mx)\rangle}{\|\cos(mx)\|^2}\cos(mx).$$

useful to know the denominators:

$$\| 1 \|^2 = \frac{1}{2\pi}\int_0^{2\pi} 1 \, dx = \frac{2\pi}{2\pi} = 1. \leftarrow$$ This is why we work with this version of the integral inner product: $\langle f, g \rangle = \frac{1}{b-a}\int_a^b f(x)g(x)\,dx$ – so $\|1\| = 1$.

and
$$\| \sin(nx) \|^2 = \frac{1}{2\pi}\int_0^{2\pi} (\sin(nx))^2 dx = \frac{1}{4\pi}\int_0^{2\pi} \underset{\cos(0)}{\underset{\parallel}{1}} - \cos(2nx)\,dx$$
$$= \frac{1}{4\pi}\int_0^{2\pi} 1\,dx = 2\pi/4\pi = 1/2.$$

$\|\cos(nx)\|^2 = \frac{1}{2}$ similarly.

So we get:

$$\mathrm{proj}_{W_m}(f) = \overbrace{\langle f, 1\rangle}^{a_0} + \overbrace{2\langle f, \sin(x)\rangle}^{a_1} \sin(x) + \overbrace{2\langle f, \cos(x)\rangle}^{b_1}\cos(x)$$
$$+ \dots + \underbrace{2\langle f, \sin(mx)\rangle}_{a_m} \sin(mx) + \underbrace{2\langle f, \cos(mx)\rangle}_{b_m}\cos(mx)$$

In particular $a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx$, $a_n = \frac{\cancel{2}\,1}{\cancel{2}\pi}\int_0^{2\pi} f(x)\sin(nx)\,dx$,

$b_n = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos(nx)\,dx$

These are called the Fourier coefficients of f

The mean square error $\|f - \mathrm{proj}_{W_m} f\|^2 \longrightarrow 0$ as $m \longrightarrow \infty$ i.e. we get a better & better approx. to f the more terms in the trig. sum we take.

The limit $\lim_{m\to\infty}\left(\mathrm{proj}_{W_m} f\right)$ is a series:

$$\boxed{a_0 + \sum_{n=1}^{\infty}\left(a_n \sin(nx) + b_n \cos(nx)\right) \quad \text{A series of this form is called a } \underline{\text{Fourier series}}.}$$

Notice: if we set W_∞ = the subspace of all Fourier series, then $W_\infty^\perp = \{0\}$.
(By the limit above.)

But $\{0\}^\perp = C([0,2\pi])$ ("everything")

However $C([0,2\pi]) \neq W_\infty$ (e.g. $f(x) = x$ not a Fourier series)
So $(W_\infty^\perp)^\perp = \{0\}^\perp \neq W_\infty$

Why does this not contradict the earlier theorem? Because W_∞ is NOT finite-dimensional.

(It could be that $(W^\perp)^\perp = W$ and W is infinite dimensional, though – e.g. take any infinite dim. vector space V; then $V^\perp = \{0\}$ and $\{0\}^\perp = V$ so $(V^\perp)^\perp = V$.)

Linear Transformations

Key Example If A is an $m\times n$-matrix over $\mathbb{R}$ then it transforms vectors in $\mathbb{R}^n$ into vectors in $\mathbb{R}^m$ by multiplication: $y = Ax$

We always have ① $A(x+u) = Ax + Au$ for x, u in $\mathbb{R}^n$

② $A(kx) = kAx$ for x in $\mathbb{R}^n$, k in $\mathbb{R}$

But also if we have any transformation $T: \mathbb{R}^n \to \mathbb{R}^m$ satisfying ① and ② i.e. $T(x+u) = T(x) + T(u)$ and $T(kx) = kT(x)$

then there is an $m\times n$ matrix A with the property that

$T(x) = Ax$. Why? :

If ① & ② hold for T, then if $x = x_1e_1 + \ldots + x_ne_n$ (for the standard basis vectors $e_1 = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \ldots, e_n = \begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix}$),

$$\begin{aligned} \text{then } T(x) &= T(x_1e_1 + \ldots + x_ne_n) \\ &= T(x_1e_1) + \ldots + T(x_ne_n) \quad \text{by ①} \\ &= x_1T(e_1) + \ldots + x_nT(e_n) \quad \text{by ②} \end{aligned}$$

Recall if $M = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$ is an $m\times n$-matrix, then $Me_i = v_i$ for each i

← This just means that $v_1, \ldots, v_n$ are the columns of M.

& $Mx = x_1v_1 + \ldots + x_nv_n$

Let $A = \begin{pmatrix} | & & | \\ T(e_1) & \cdots & T(e_n) \\ | & & | \end{pmatrix}$. Then $Ax = T(x)$.

i.e. let $T(e_1), \dots, T(e_n)$ be the column vectors of A.

So there is a one-to-one correspondence between $m \times n$ matrices over $\mathbb{R}$ and functions $T: \mathbb{R}^n \to \mathbb{R}^m$ satisfying ① + ②. We sometimes write $T_A : \mathbb{R}^n \to \mathbb{R}^m$ to emphasise $T_A(x) = Ax$.

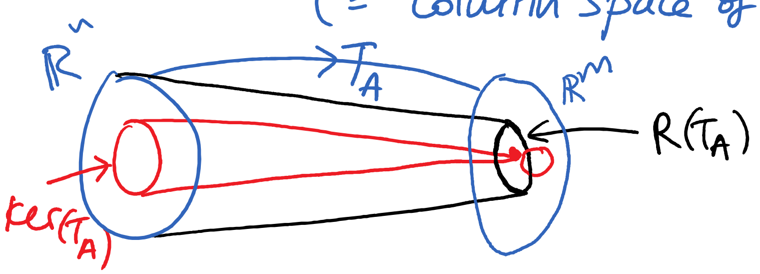
For such T_A we have

$\ker(T_A) = \{x \text{ with } T_A(x) = 0\}$, the <u>kernel</u> of T_A

(= nullspace of corresponding matrix A)

$R(T_A) = \{y \text{ with } T_A(x) = y \text{ for some } x\}$, the <u>range</u> of T_A

(= column space of corresponding A)



We say that T_A is one-to-one (1-1) if it maps distinct vectors in $\mathbb{R}^n$ to distinct vectors in $\mathbb{R}^m$.

<u>Theorem</u> If A is $M_{m\times n}(\mathbb{R})$, $T_A(x) = Ax$ then $\ker(T_A) = \{0\}$ exactly when T_A is one-to-one.

i.e. A is invertible: the only solution to the homogeneous system $Ax=0$ is the zero vector $x=0$.

Inverses If it makes sense to talk about T_A^{-1}, What is B with $Bx = T_A^{-1}$?

Well, we want $x = B(T_A(x)) = B(Ax)$

so $B = A^{-1}$. (i.e. $(T_A)^{-1} = T_{A^{-1}}$).

In other words, T_A is invertible (as a transformation) if A is (as a matrix)!

(General) Linear Transformations

Functions between vector spaces that preserve addition & scalar multiplication

Definition A linear transformation $T: U \longrightarrow V$ between vector spaces U and V is a function satisfying (1) $T(u \underset{\text{+ in } U}{+} w) = T(u) \underset{\text{+ in } V}{+} T(w)$ (2) $T(\underset{\substack{\text{scalar}\\\text{mult.}\\\text{in } U}}{ku}) = \underset{\substack{\text{scalar}\\\text{mult.}\\\text{in } V}}{kT(u)}$

for all vectors u, w in U and scalars k.

(If $U = V$, we sometimes call T a linear operator.)

Basic Properties (1) $T(0) = 0$ (2) $T(u-w) = T(u) - T(w)$
(3) $T(-u) = -T(u)$.

Examples (A) Matrix transformations (see above: T_A).

(B) Identity operator $I: U \to U$, $I(u) = u$.

(C) Zero transformation $T: U \to V$, $T(u) = 0$ ← zero in V

(D) Dilation: $T: \mathbb{R}^n \to \mathbb{R}^n$, $T(x) = cx$ for some fixed c in $\mathbb{R}$.

(E) Transpose : $T: M_{m\times n}(\mathbb{R}) \to M_{n\times m}(\mathbb{R})$, $T(A) = A^T$

(F) If $U = P_3$ and $V = P_5$, $T: U \to V$ given by

$$T(p) = x^2 p(x)$$

(say $p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$, then $T(p) = a_0 x^2 + a_1 x^3 + a_2 x^4 + a_3 x^5$)

(G) Orthogonal projection in $\mathbb{R}^3$ $P: \mathbb{R}^3 \to \mathbb{R}^3$ onto (xy)-plane

$P(x,y,z) = (x,y,0)$ (Notice, by the above, there is a 3x3 matrix A with $T(x,y,z) = (x,y,0) = A\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. Can you find A ?)

(H) Differential transformation on space $C'((-\infty,\infty))$

(Space of continuously differentiable functions on $\mathbb{R}$)

$D: C'((-\infty,\infty)) \longrightarrow F((-\infty,\infty))$, $D(f) = \frac{df}{dx}$ $(=f')$

We have $D(f+g) = (f+g)' = f' + g' = D(f) + D(g)$

& $D(kf) = (kf)' = kf' = kD(f)$

(I) Integral transformation on space $C((-\infty,\infty))$ of continuous functions on $\mathbb{R}$ $J: C((-\infty,\infty)) \longrightarrow C'((-\infty,\infty))$

$$J(f) = \int_0^x f(t)\,dt = F(x) - F(0) \quad (\text{where } F' = f).$$

to determine the constant $+C$

We need a well-defined function i.e. 1 output per input f. So we need to specify a choice of constant.

Non-examples

(J) Any translation operator $T: U \to U$, $T(x) = x + x_0$ for some fixed vector x_0 in U is NOT linear if $x_0 \neq 0$ (as $0 = T(0) = x_0$).

(K) Determinants : $\det: M_n(\mathbb{R}) \to \mathbb{R}$ is NOT linear as in general $\det(A+B) \neq \det(A) + \det(B)$

As in the matrix transformation case, the action of a linear transformation $T: U \to V$ on a basis $S = \{u_1, \dots, u_n\}$ for U (if U is finite dimensional) completely determines how T acts on U (i.e. enough to look at basis vectors).

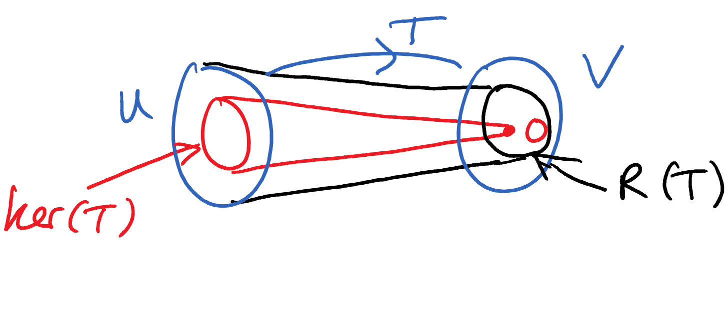
If $u \in U$, then $u = a_1 u_1 + \dots + a_n u_n$ for scalars a_i

So $T(u) = a_1 T(u_1) + \dots + a_n T(u_n)$.

So we find $T(u)$ by taking the "same" linear combination of images of u_i under T. (i.e. same coefficients)

We also have the kernel of a linear transformation $T: U \to V$: $\ker(T) = \{u \text{ in } U \text{ with } T(u) = 0\}$

& the range of T : $R(T) = \{v \text{ in } V \text{ with } T(u) = v \text{ for some } u \text{ in } U\}$



Examples (A) $\ker(T_A)$ = null space of A

$R(T_A)$ = column space of A

(B) $\ker(I) = \{0\}$, $R(I) = U$

(C) $T: u \mapsto 0$, $\ker(T) = U$, $R(T) = \{0\}$

(G) $T(x,y,z) = (x,y,0)$; $\ker(T) = z$-axis

$R(T) = (x,y)$-plane

(H) $D(f) = f'$: $D(f) = 0 \to \ker(D) = \{$constant functions$\}$

& $R(D) = \{$integrable functions$\}$

(I) $J(f) = \int_0^x f(t)\,dt$ has $J(f) = 0$ exactly when

$F(x) = F(0)$ – F constant

$0 = F' = f$ so $\ker(J) = \{0\}$.

& $R(J) = \{$differentiable functions which are 0 at 0$\}$