Math 2270-004 Week 13 notes
We will not necessarily finish the material from a given day's notes on that day. We may also add or subtract some material as the week progresses, but these notes represent an in-depth outline of what we plan to cover. These notes cover material in 6.4-6.8

Mon Apr 9
- 6.4 Gram Schmidt and $A = QR$ decomposition. Orthogonal matrices

Announcements:

(Recall)
Warm-up Exercise: Can you still use Gram-Schmidt to convert $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \end{bmatrix} \right\}$ into an orthonormal basis for $\mathbb{R}^2$?

hint: $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|}$ ✓

$\vec{z}_2 = \vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1$

$\vec{u}_2 = \dfrac{\vec{z}_2}{\|\vec{z}_2\|}$

(page 3)

$$\left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$$

(if it was up to you, reorder $\left\{ \begin{bmatrix} 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \rightarrow \left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$)

$$\vec{z}_2 = \begin{bmatrix} 0 \\ 4 \end{bmatrix} - \left(\begin{bmatrix} 0 \\ 4 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ 4 \end{bmatrix} - \frac{1}{2}(4)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = 2\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\vec{u}_2 = \frac{\vec{z}_2}{\|\vec{z}_2\|} = \frac{1}{\sqrt{8}}\begin{bmatrix} -2 \\ 2 \end{bmatrix} = \frac{1}{2\sqrt{2}}\begin{bmatrix} -2 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

better way is to normalize $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$: $\frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$

(there is only one unit vector in direction of any $\vec{v}$, normalize any appropriate positive scalar mult of $\vec{v}$ to get it. This will help on HW)

We begin on Monday with a continuation of the discussion of Gram-Schmidt orthogonalization from 6.4. Keeping track of the G.S. process carefully yields the $A = QR$ matrix product decomposition theorem, where Q is an "orthogonal matrix" consisting of an orthonormal basis for the span of the columns of A and R is an upper triangular matrix with positive entries along the diagonal. This decomposition is one way to understand why matrix determinants correspond to $\pm$ Volumes, in $\mathbb{R}^n$, among other uses.

T Section 6.5, *Least square solutions* is about finding approximate solutions to inconsistent matrix equations, and relies on many of the ideas we've been studying in Chapter 6 up to this point.

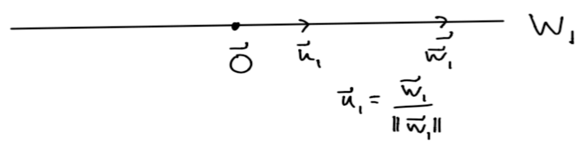
W Section 6.6, *Applications to linear models,* is an application of the least squares method to e.g. linear regression in statistics.

F Finally, sections 6.7 and 6.8 generalize our orthogonality discussions that began with the dot product, to *inner products* in other vector spaces such as function spaces. These ideas lie at the heart of physics applications that use Fourier series, and more recent applications such as image and audio compression.

Recall the Gram-Schmidt process from Friday:

Start with a basis $B = \{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_p\}$ for a subspace W of $\mathbb{R}^n$. How can you convert it into an orthonormal basis? Here's how! The inductive process is called Gram-Schmidt orthogonalization.

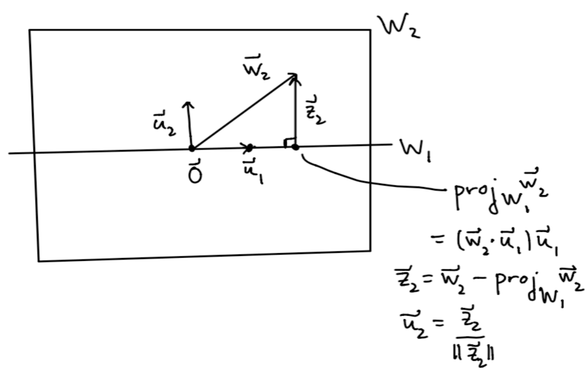
- Let $W_1 = span\{\underline{\boldsymbol{w}}_1\}$. Define $\underline{\boldsymbol{u}}_1 = \dfrac{\underline{\boldsymbol{w}}_1}{\|\underline{\boldsymbol{w}}_1\|}$. Then $\{\underline{\boldsymbol{u}}_1\}$ is an orthonormal basis for W_1.



Let $W_2 = span\{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2\} = span\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{w}}_2\}$.

Let $\underline{\boldsymbol{z}}_2 = \underline{\boldsymbol{w}}_2 - proj_{W_1} \underline{\boldsymbol{w}}_2 = \underline{\boldsymbol{w}}_2 - (\underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1$ so $\underline{\boldsymbol{z}}_2 \perp \underline{\boldsymbol{u}}_1$.

Define $\underline{\boldsymbol{u}}_2 = \dfrac{\underline{\boldsymbol{z}}_2}{\|\underline{\boldsymbol{z}}_2\|}$. So $\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2\}$ is an orthonormal basis for W_2.



Inductively,

Let $W_j = span\{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_j\} = span\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_{j-1}, \underline{\boldsymbol{w}}_j\}$.

Let $\underline{\boldsymbol{z}}_j = \underline{\boldsymbol{w}}_j - proj_{W_{j-1}} \underline{\boldsymbol{w}}_j = \underline{\boldsymbol{w}}_j - (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 - (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2 - \ldots (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_{j-1})\underline{\boldsymbol{u}}_{j-1}$.

...so $\underline{\boldsymbol{z}}_j \perp span\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_{j-1}\}$.

Define $\underline{\boldsymbol{u}}_j = \dfrac{\underline{\boldsymbol{z}}_j}{\|\underline{\boldsymbol{z}}_j\|}$. Then $\{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_j\}$ is an orthonormal basis for W_j.

Continue up to $j = p$.

We're denoting the original basis for W by $B = \{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_p\}$. Denote the orthonormal basis we've constructed with Gram-Schmidt by $O = \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_p\}$. Because O is orthonormal it's easy to express these two bases in terms of each other. Notice

$$W_j = span\ \{\underline{\boldsymbol{w}}_1, \underline{\boldsymbol{w}}_2, \ldots \underline{\boldsymbol{w}}_j\} = \underbrace{span\ \{\underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_2, \ldots \underline{\boldsymbol{u}}_j\}}_{\text{orthormal set}} \quad \text{for each } 1 \le j \le p.$$

$\vec{x} \in W_j$

$\vec{x} = (\vec{x}\cdot\vec{u}_1)\vec{u}_1 + (\vec{x}\cdot\vec{u}_2)\vec{u}_2 + \cdots (\vec{x}\cdot\vec{u}_j)u_j$

So,

1 $$\underline{\boldsymbol{w}}_1 = (\underline{\boldsymbol{w}}_1 \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1$$

2 $$\underline{\boldsymbol{w}}_2 = (\underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{w}}_2 \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2$$

$$\vdots$$

$$\underline{\boldsymbol{w}}_j = (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_1)\underline{\boldsymbol{u}}_1 + (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_2)\underline{\boldsymbol{u}}_2 + \ldots\ (\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j)\underline{\boldsymbol{u}}_j$$

$$\vdots$$

$$\underline{\boldsymbol{w}}_p = \sum_{l=1}^{p} (\underline{\boldsymbol{w}}_l \cdot \underline{\boldsymbol{u}}_l)\underline{\boldsymbol{u}}_l\ .$$

Notice that the coefficients of the last terms in the sums above, namely $(\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j)$ can be computed as

keep in mind. $$(\underline{\boldsymbol{w}}_j \cdot \underline{\boldsymbol{u}}_j) = \underline{\boldsymbol{z}}_j \cdot \frac{\underline{\boldsymbol{z}}_j}{\|\underline{\boldsymbol{z}}_j\|} = \|\underline{\boldsymbol{z}}_j\| \qquad \vec{u}_j \cdot \left[\ \vec{z}_j = \vec{w}_j - proj_{W_{j-1}} \vec{w}_j\ \right]$$

In matrix form (column by column) we have

$$* \quad \underbrace{\begin{array}{c} \begin{array}{ccc} 1 & 2 & p \end{array} \\ \begin{bmatrix} \vert & \vert & & \vert \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_p \\ \vert & \vert & & \vert \end{bmatrix} \end{array}}_{\text{"A"}} = \underbrace{\begin{bmatrix} \vert & \vert & & \vert \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_p \\ \vert & \vert & & \vert \end{bmatrix}}_{\text{"Q"}} \underbrace{\begin{array}{c} \begin{array}{cccc} 1 & 2 & & p \end{array} \\ \begin{bmatrix} \vec{w}_1\cdot\vec{u}_1 & \vec{w}_2\cdot\vec{u}_1 & \vec{w}_3\cdot\vec{u}_1 & \cdots & \vec{w}_p\cdot\vec{u}_1 \\ 0 & \vec{w}_2\cdot\vec{u}_2 & \vec{w}_3\cdot\vec{u}_2 & & \vec{w}_p\cdot\vec{u}_2 \\ 0 & 0 & \vec{w}_3\cdot\vec{u}_3 & \ddots & \vdots \\ \vdots & \vdots & 0 & & \\ 0 & 0 & 0 & & \vec{w}_p\cdot\vec{u}_p \end{bmatrix} \end{array}}_{\text{"R"}}$$

"A" columns are original basis for W

$A_{n\times p}$

"Q" columns are orthonormal

$Q_{n\times p}$

"R" upper Δ'ular, with positive diagonal entries

$\vec{w}_j \cdot \vec{u}_j = \|\vec{z}_j\|$

$R_{p\times p}$

R also known as $P_{O\leftarrow B}$ ☺

Thus any matrix with linearly independent columns may be written in factored form as above, ($W = Col\, A$),

$$A_{n\times p} = Q_{n\times p} R_{p\times p}.$$

This factorization contains geometric information and can simplify the computational work needed to solve matrix equations $A\,\underline{\boldsymbol{x}} = \underline{\boldsymbol{b}}$.

From previous page...

$$* \qquad A_{n\times p} = Q_{n\times p} R_{p\times p}$$

<u>shortcut</u> (or what to do if you forgot the formulas for the entries of R) If you just know Q you can recover R by multiplying both sides of the $*$ equation on the previous page by the transpose Q^T of the Q matrix:

$$\underset{Q^T}{\begin{bmatrix} \cdots & \vec{u}_1^T & \cdots \\ \cdots & \vec{u}_2^T & \cdots \\ & \vdots & \\ \cdots & \vec{u}_p^T & \cdots \end{bmatrix}} \underset{A}{\begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_p \\ \vdots & \vdots & & \vdots \end{bmatrix}} = \underbrace{\underset{Q^T}{\begin{bmatrix} \cdots & \vec{u}_1^T & \cdots \\ \cdots & \vec{u}_2^T & \cdots \\ & \vdots & \\ & \vec{u}_p^T & \end{bmatrix}} \underset{Q}{\begin{bmatrix} \vdots & \vdots & & \vdots \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_p \\ \vdots & \vdots & & \vdots \end{bmatrix}}} \; R = IR = R!$$

$$[\ u_i^T\ ]\,[\vec{u}_j] = \vec{u}_i \cdot \vec{u}_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

$$Q^T Q = \begin{bmatrix} 1 & 0 & \\ 0 & 1 & \\ 0 & 0 & \\ \vdots & 0 & \\ 0 & \vdots & \end{bmatrix}$$

$$A = Q\,R$$
$$Q^T A = Q^T Q\,R = I\,R = R.$$

<u>Example)</u> From last Friday,

$$B = \left\{ \underset{\vec{w}_1}{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}, \underset{\vec{w}_2}{\begin{bmatrix} 0 \\ 4 \end{bmatrix}} \right\}, \quad O = \left\{ \underset{\vec{u}_1}{\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}}, \underset{\vec{u}_2}{\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}} \right\}.$$

$\vec{w}_1 \cdot \vec{u}_1 = \frac{2}{\sqrt{2}} = \sqrt{2}$

$\vec{w}_2 \cdot \vec{u}_1 = \frac{4}{\sqrt{2}} = 2\sqrt{2}$

$\vec{w}_2 \cdot \vec{u}_2 = \frac{4}{\sqrt{2}} = 2\sqrt{2}$

$$\underset{B}{\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix}} = \underset{Q}{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}} \underset{R}{\begin{bmatrix} \boldsymbol{w}_1 \cdot \boldsymbol{u}_1 & \boldsymbol{w}_2 \cdot \boldsymbol{u}_1 \\ 0 & \boldsymbol{w}_2 \cdot \boldsymbol{u}_2 \end{bmatrix}} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix} = Q\,R.$$

<u>Exercise 1)</u> Verify that R could have been recovered via the formula

$$Q^T A = R$$

"A" $= B$

$$A = QR$$
$$Q^T A = Q^T Q R = IR.$$

$$\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix} \checkmark$$

From previous page ...

$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \underbrace{\begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix}}_{R}.$$

<u>Exercise 2)</u> Verify that the $A = Q\,R$ factorization in this example may be further factored as

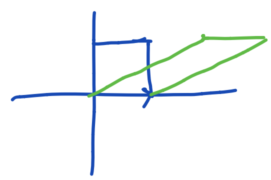
$$\begin{bmatrix} 1 & 0 \\ 1 & 4 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{Rot_{\pi/4}} \underbrace{\begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix}}_{\text{diag. stretch matrix}} \underbrace{\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}}_{\text{area preserving shear}}.$$

$$\begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 2\sqrt{2} \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \qquad \frac{2\sqrt{2}}{\sqrt{2}}$$

diag. stretch matrix
(x_1-dir by $\sqrt{2}$
x_2-dir by $2\sqrt{2}$)

area preserving shear

$$T(\vec{x}) = A\vec{x}$$



- So, the transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$ is a composition of (1) an area-preserving shear, followed by (2) a diagonal scaling that increases area by a factor of $\sqrt{2}\cdot 2\sqrt{2} = 4$, followed by a rotation of $\frac{\pi}{4}$, which does not effect area. Since determinants of products matrices are the products of determinants (we checked this back when we studied determinants), and area expansion factors of compositions are also the products of the area expansion factors, the generalization of this example explains why the determinant of A (or its absolute value in general) coincides with the area expansion factor, in the 2×2 case. You show in your homework that the only possible Q matrices in the 2×2 case are rotations as above, or reflections across lines through the origin. In the latter case, the determinant of Q is -1, and the determinant of A is negative.

$$Q_{2\times 2} \quad \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{Rot_\theta,\ \det = +1} \quad \text{or} \quad \underbrace{\begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}}_{\det = -1} \qquad HW$$

($Refl_L$ L is at angle $\frac{\theta}{2}$)

Example from last Friday.

$$B = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} \right\} \qquad O = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Exercise 3a Find the $A = Q\,R$ factorization based on the data above, for

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 4 & -2 \\ 0 & 0 & 3 \end{bmatrix}$$

$A = QR$

shortcut for R

$$Q^T A = \underbrace{Q^T Q}_{I} R = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 4 & -2 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & 2\sqrt{2} & -\frac{3}{\sqrt{2}} \\ 0 & 0 & 3 \end{bmatrix}$$

✓

$$\text{solution } A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} \sqrt{2} & 2\sqrt{2} & -\frac{1}{\sqrt{2}} \\ 0 & 2\sqrt{2} & -\frac{3}{\sqrt{2}} \\ 0 & 0 & 3 \end{bmatrix}}_{R}$$

Exercise 3b Further factor R into a diagonal matrix times a volume-preserving shear and interpret the transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$ as a composition of (1) a volume preserving shear, followed by (2) a coordinate scaling that increases volume by a factor of 12, followed by a rotation about the x_3 axis in $\mathbb{R}^3$, which preserves volume. The generalization of this example explains why the determinant of A (or its absolute value in general) is the volume expansion factor for the transformation $T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}}$.

$$\frac{-3}{\sqrt{2}\cdot 2\sqrt{2}}$$

$$R = \underbrace{\begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 2\sqrt{2} & 0 \\ 0 & 0 & 3 \end{bmatrix}}_{\text{vol} \times 12} \underbrace{\begin{bmatrix} 1 & 2 & -\frac{1}{2} \\ 0 & 1 & -3/4 \\ 0 & 0 & 1 \end{bmatrix}}_{\substack{\text{shear} \\ \text{vol pres.}}}$$

Q

Definition A square $n \times n$ matrix Q is called *orthogonal* if its columns are ortho-normal. (You can read more about orthogonal matrices at e.g. Wikipedia.)

Theorem. Let Q be an orthogonal matrix. Then

a) $Q^{-1} = Q^T$.

$$\text{entry}_{ij}\ Q^T Q = [\text{Row}_i\ Q^T] \begin{bmatrix} \text{Col}_j Q \end{bmatrix} = \text{Col}_i Q \cdot \text{col}_j Q = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

$$\Rightarrow Q^T Q = I$$

$$\Rightarrow Q Q^T = I \text{ also (prev. Chapter).}$$

$$\text{i.e. } Q^{-1} = Q^T$$

b) The rows of Q are also ortho-normal.

$$Q Q^T = I$$

$$ij \text{ entry: } \text{row}_i Q \cdot \underbrace{\text{col}_j Q^T}_{\text{row}_j (Q)} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

c) the transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ given by

$$T(\underline{x}) = Q\,\underline{x}$$

preserves dot products and magnitudes, (so also volumes, since cubes generated by perpendicular vectors will be transformed into equal-volume cubes). In other words, for all $\underline{x}, \underline{y} \in \mathbb{R}^n$,

$$T(\underline{x}) \cdot T(\underline{y}) = \underline{x} \cdot \underline{y}$$

$$\|T(\underline{x})\| = \|\underline{x}\|.$$

d) The only matrix transformations $T : \mathbb{R}^n \to \mathbb{R}^n$ that preserve dot products are orthogonal transformations. (These transformations are often referred to as *isometries*.)

Tues Apr 10 to $A\vec{x} = \vec{b}$

- 6.5 Least squares "solutions", and projection revisited.

Announcements:
- graded HW + quizzes
- ↳ 6.5 first, then go back to rest of Monday notes.

'til 12:56

Warm-up Exercise: What is the condition on $\vec{y}$ so that the system

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

is consistent?

low tech (Chptrs 1-2)

$$\left[\begin{array}{cc|c} 1 & 2 & y_1 \\ 0 & 1 & y_2 \\ 1 & 0 & y_3 \end{array}\right]$$

$R_3 \to R_1$, $R_1 \to R_3$:

$$\left[\begin{array}{cc|c} 1 & 0 & y_3 \\ 0 & 1 & y_2 \\ 1 & 2 & y_1 \end{array}\right]$$

$-R_1 + R_3 \to R_3$:

$$\left[\begin{array}{cc|c} 1 & 0 & y_3 \\ 0 & 1 & y_2 \\ 0 & 2 & y_1 - y_3 \end{array}\right]$$

$-2R_2 + R_3 \to R_3$:

$$\left[\begin{array}{cc|c} 1 & 0 & y_3 \\ 0 & 1 & y_2 \\ 0 & 0 & y_1 - 2y_2 - y_3 \end{array}\right]$$

$\vec{y}$ is on plane with implicit eqtn

$$\boxed{y_1 - 2y_2 - y_3 = 0}$$

high tech Chapter 6.

$(W^{\perp})^{\perp} = W.$

$W^{\perp}$ for span$\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}\right\}$

$$W^{\perp}: \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 \end{array}\right]$$

$-2R_1 + R_2 \to R_2$:

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & -2 & 0 \end{array}\right]$$

$W^{\perp}$: span$\left\{\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}\right\}$

$(W^{\perp})^{\perp} = W$

$$\| \quad \begin{bmatrix} -1 & 2 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = 0$$

equivalent

$$\boxed{-y_1 + 2y_2 + y_3 = 0}$$

(better in high dims)

Least squares solutions, section 6.5

In trying to fit experimental data to a linear model you must often find a "solution" to

$$A\,\underline{\mathbf{x}} = \underline{\boldsymbol{b}}$$

where no exact solution actually exists. Mathematically speaking, the issue is that $\underline{\boldsymbol{b}}$ is not in the range of the transformation

$$T(\underline{\mathbf{x}}) = A\,\underline{\mathbf{x}},$$

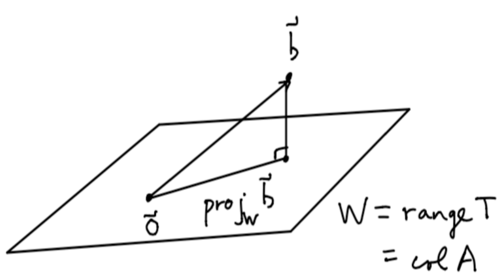
i.e.

$$\underline{\mathbf{x}} \notin Range\ T = Col\ A.$$

In such a case, the *least squares solution(s)* $\hat{\mathbf{x}}$ solve(s)

$$\boxed{A\,\hat{\mathbf{x}} = proj_{Col\,A}\,\underline{\boldsymbol{b}}\,.}$$

Thus, for the least squares solution(s), $A\,\hat{\mathbf{x}}$ is as close to $\underline{\mathbf{b}}$ as possible. Note that there will be a unique least squares solution $\hat{\mathbf{x}}$ if and only if $Nul\,A = \{\underline{\mathbf{0}}\}$, i.e. if and only if the columns of A are linearly independent. (Recall, any two solutions to the same nonhomogeneous matrix equation differ by a solution to the homogeneous equation.)



warmup

Exercise 1 Find the least squares solution to

$$\begin{array}{cc} \vec{w}_1 & \vec{w}_2 \end{array}$$
$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}.$$

Note that the implicit equation of the plane spanned by the two columns of A is

$$-y_1 + 2y_2 + y_3 = 0. \qquad -3+6+3 \neq 0$$

You know two ways to find that implicit equation (!) .....at least it's easy to check that the the two column vectors satisfy it. Since $[3\ 3\ 3]^T$ does not satisfy the implicit equation, there is no exact solution to this problem. If you wish, it could be instructive review the two ways.

You may use the Gram-Schmidt ortho-normal basis for *Col A*, namely

$$O = \left\{ \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\vec{u}_1}, \underbrace{\frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}}_{\vec{u}_2} \right\}.$$

Step 1 find $\text{proj}_{Col A} \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \hat{b}$

Step 2 Solve $A\hat{x} = \hat{b}$. $\hat{x}$ = "least squares soln"

$$\text{proj}_W \vec{b} = (\vec{b}\cdot\vec{u}_1)\vec{u}_1 + (\vec{b}\cdot\vec{u}_2)\vec{u}_2$$
$$= \left(\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}\cdot\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right)\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \left(\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}\cdot\frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\right)\frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$
$$= \frac{1}{2}(6)\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \frac{1}{3}(3)\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$$
$$= 3\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$

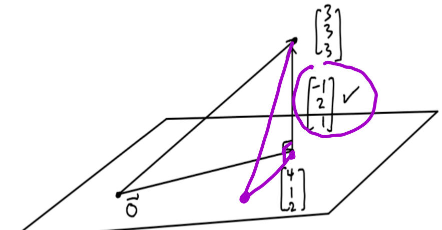
$$\left[\begin{array}{cc|c} 1 & 2 & 4 \\ 0 & 1 & 1 \\ 1 & 0 & 2 \end{array}\right]$$

$$\begin{array}{l} R_3 \to R_1 \\ \\ R_1 \to R_3 \end{array} \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 1 & 2 & 4 \end{array}\right]$$

$$\begin{array}{l} \\ \\ -R_1 - 2R_2 + R_3 \to R_3 \end{array} \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{array}\right]$$

$x_1 = 2$
$x_2 = 1$

Solution:



$$\begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$

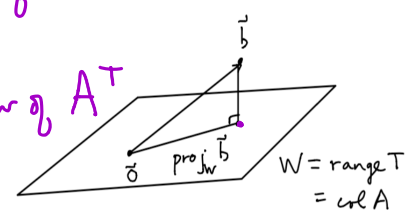
There's actually a smart way to find the least squares solutions that doesn't require an orthonormal basis for *Col A*. And as a result, it turns out that one can also compute projections onto a subspace without first constructing an orthonormal basis for the subspace !!! Consider the following chain of equivalent conditions on $\underline{\boldsymbol{x}}$:

$$\boxed{A\,\underline{\mathbf{x}} = proj_{Col\,A}\,\underline{\boldsymbol{b}}}$$

$$\vec{z} = b - A\,\underline{\mathbf{x}} \in (Col\,A)^{\perp}$$

$$b - A\vec{x} \perp \text{each col of } A$$

$$b - A\vec{x} \perp \text{each row of } A^T$$



$$A^T(\underline{\boldsymbol{b}} - A\,\underline{\mathbf{x}}) = \underline{\mathbf{0}}$$

$$A^T\underline{\boldsymbol{b}} - A^T A\,\underline{\mathbf{x}} = \underline{\mathbf{0}}$$

$$\boxed{A^T A\,\underline{\mathbf{x}} = A^T\underline{\boldsymbol{b}}\,.}$$

This last equation will always be consistent because projections exist. And if the columns of A are linearly independent the solutions to the top equation, and hence the final equation, will be unique. So the matrix $A^T A$ will be invertible in that case. The final matrix equation is called the *normal equation* for least squares solutions.

Exercise 2 Re-do Exercise 1 using the normal equation, i.e find the least squares solution $\hat{x}$ to

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}.$$

And then note that $A\,\hat{x}$ is $\overbrace{proj_{Col\,A}\underline{\boldsymbol{b}}}^{\hat{b}}$, i.e. you found the projection of $[3\ 3\ 3]^T$ without ever finding and using an ortho-normal basis!!!

$$A\hat{x} = \hat{b}$$

$$A^T A\,\hat{x} = A^T\vec{b}$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 5 \end{bmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 6 \\ 9 \end{bmatrix}$$

no Gram-Schmidt $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ 9 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 12 \\ 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

Same as in Exercise 1 !!

after find $\hat{x}$,

$$A\hat{x} = proj_{Col\,A}\vec{b}$$

$$A\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$

Exercise 3 In the case that $A^T A$ is invertible we may take the normal equation for finding the least squares solution to $A x = b$ and find $A\hat{x} = proj_{Col\,A} b$ directly:

$$\boxed{A^T A \underline{\mathbf{x}} = A^T \underline{\mathbf{b}}}$$

$$\underline{\mathbf{x}} = \left(A^T A\right)^{-1} A^T \underline{\mathbf{b}}$$

$$proj_{Col\,A} \underline{\mathbf{b}} = A\,\underline{\mathbf{x}} = A\left(A^T A\right)^{-1} A^T \underline{\mathbf{b}}.$$

Verify for the third time that for $W = span\left\{\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}2\\1\\0\end{bmatrix}\right\}$, $proj_W \begin{bmatrix}3\\3\\3\end{bmatrix} = \begin{bmatrix}4\\1\\2\end{bmatrix}$ by "plug and chug".

$$proj_W \begin{bmatrix}3\\3\\3\end{bmatrix} = \begin{bmatrix}1&2\\0&1\\1&0\end{bmatrix}\begin{bmatrix}2&2\\2&5\end{bmatrix}^{-1}\begin{bmatrix}1&0&1\\2&1&0\end{bmatrix}\begin{bmatrix}3\\3\\3\end{bmatrix}$$

$$= \begin{bmatrix}1&2\\0&1\\1&0\end{bmatrix}\frac{1}{6}\begin{bmatrix}5&-2\\-2&2\end{bmatrix}\begin{bmatrix}6\\9\end{bmatrix}$$

See previous page

$$= \begin{bmatrix}4\\1\\2\end{bmatrix}$$

Wed Apr 11

- 6.6 Fitting data to "linear" models.

Announcements: 6.6 HW: (1) (7) (and graph the points and best parabolic fit in #7, with technology)

+ another

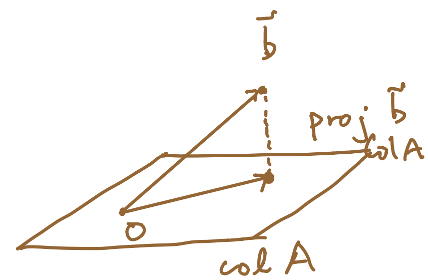
- handout

'til 12:56

Warm-up Exercise: a) Find the least squares sol'n to

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}}_{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

b) What is $\text{proj}_{col\,A}\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$?



$A\hat{x} = \text{proj}_{col\,A}\vec{b}$

$A\vec{x} = \vec{b}$

$A^T A\vec{x} = A^T\vec{b}$

gives least squares sol'n

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^T\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$2x_1 = 2$
$x_2 = 0$

$\hat{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$

b) $A\hat{x} = A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \text{proj}_{Col\,A}\vec{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

$Col\,A = \text{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}$

$= \text{span}\left\{\underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\vec{u}_1}, \underbrace{\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\vec{u}_2}\right\}$

Note Col's $\perp$ in this example.

$\text{proj}_{Col\,A}\vec{b} = (\vec{b}\cdot\vec{u}_1)\vec{u}_1 + (\vec{b}\cdot\vec{u}_2)\vec{u}_2$

$= \left(\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}\cdot\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right)\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

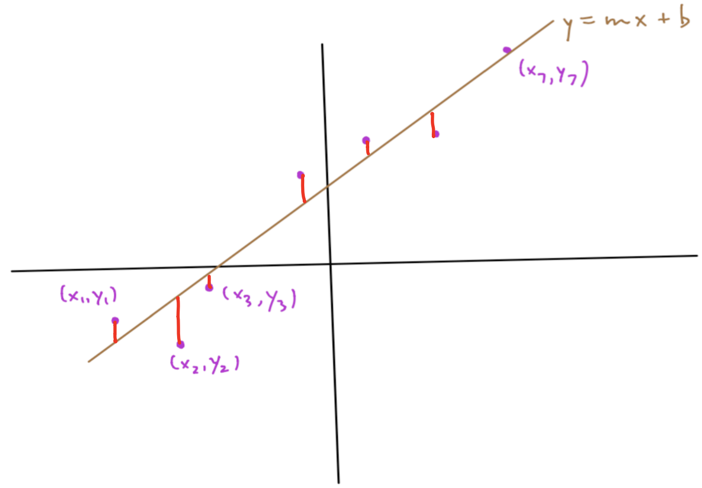
$+ \left(\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}\cdot\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \frac{2}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + \vec{0}$

Applications of least-squares to data fitting.

- Find the best line formula $y = m\,x + b$ to fit n data points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. We seek $\begin{bmatrix} m \\ b \end{bmatrix}$ so that

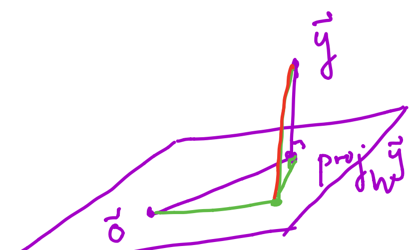
$y_i = m x_i + b$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = m \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$



In matrix form, find $\begin{bmatrix} m \\ b \end{bmatrix}$ so that

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}. \qquad A \begin{bmatrix} m \\ b \end{bmatrix} = \mathbf{y}.$$



$W = \mathrm{span}\{\vec{x}, \vec{1}\}$

There is no exact solution unless all the data points are actually on a single line!

$A \widehat{\begin{bmatrix} m \\ b \end{bmatrix}} = \mathrm{proj}_W \vec{y}$

<u>Least squares solution</u>:

$$A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \mathbf{y}.$$

$\|\vec{z}\|^2 = \| y - A \widehat{\begin{bmatrix} m \\ b \end{bmatrix}} \|^2$ is the smallest that $\|\vec{y} - A\vec{x}\|^2$ can be.

minimized the sum of the squared vertical deviations

$$\vec{y} - A \widehat{\begin{bmatrix} m \\ b \end{bmatrix}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - m \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - b \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} y_1 - m x_1 - b \\ y_2 - m x_2 - b \\ \vdots \\ y_n - m x_n - b \end{bmatrix}$$

$$A^T A \begin{bmatrix} m \\ b \end{bmatrix} = A^T \underline{\mathbf{y}}$$

As long as the columns of A are linearly independent (i.e.at least two different values for x_j) there is a unique solution $[m, b]^T$. Furthermore, you are actually solving

$$A\,\underline{\mathbf{x}} = proj_W \underline{\mathbf{y}}$$

where

$$W = span\left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\},$$

so

$$\left\| \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} - m \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - b \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\|^2$$

is as small as possible. In other words, you've minimized the sum of the *squared vertical deviations* from points on the line to the data points,

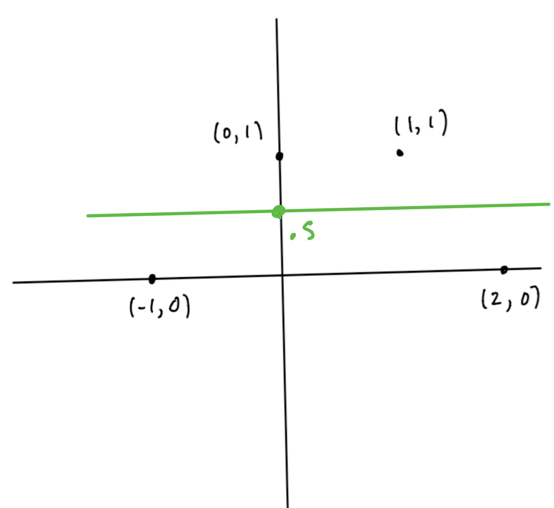
$$\sum_{i=1}^{n} \left(y_i - m x_i - b_i \right)^2.$$

x_1 x_2 x_3 x_4

Exercise 1 Find the least squares line fit for the 4 data points $\{(-1, 0), (0, 1), (1, 1), (2, 0)\}$. Sketch.

y_1 y_2 y_3 y_4

$$m \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$



$$\begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\boxed{\begin{aligned} m &= 0 \\ b &= .5 \\ y &= 0x + .5 \end{aligned}}$$

$$\begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} ; \quad \begin{bmatrix} m \\ b \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 4 & -2 \\ -2 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 0 \\ 10 \end{bmatrix} = \begin{bmatrix} 0 \\ .5 \end{bmatrix}$$

Example 2 Find the best quadratic fit to the same four data points. This is still a "linear" model!! In other words, we're looking for the best quadratic function

$$p(x) = c_0 + c_1 x + c_2 x^2$$

to fit to the four data points

$$\begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ (-1, 0), & (0, 1), & (1, 1), & (2, 0) \\ y_1 & y_2 & y_3 & y_4 \end{array}$$

$$\{(-1, 0), (0, 1), (1, 1), (2, 0)\}.$$

We want to solve

$$c_0 \begin{bmatrix} 1 \\ 1 \\ : \\ 1 \end{bmatrix} + c_1 \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_n \end{bmatrix} + c_2 \begin{bmatrix} x_1^2 \\ x_2^2 \\ : \\ x_n^2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ : \\ y_n \end{bmatrix}.$$

For our example this is the system

$$\begin{array}{cccc} & x_i & x_i^2 & y_i \end{array}$$

$$c_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + c_1 \begin{bmatrix} -1 \\ 0 \\ 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

$$A\vec{c} = \vec{y}$$
$$(A^T A)\vec{c} = A^T \vec{y}$$
$$\vec{c} = (A^T A)^{-1} A^T \vec{y}$$

I used technology (Maple, with which I write these notes), and the least squares normal equation , see next page...

$$A^T A\, \underline{\mathbf{c}} = A^T \underline{\mathbf{b}}\,.$$

> with(LinearAlgebra) :

$$C := \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} : \; b := \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} :$$

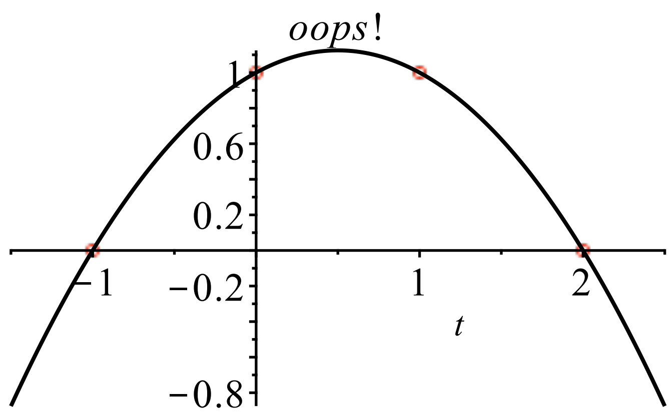
$c := (Transpose(C).C)^{-1}.Transpose(C).b;$

$$c := \begin{array}{c} \begin{bmatrix} 1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \end{array} \begin{array}{l} c_0 \\ c_1 \\ c_2 \end{array} \qquad p(x) = c_0 + c_1 x + c_2 x^2 \tag{1}$$

$p(x) = 1 + .5x - .5x^2$

> with(plots) :

$plot1 := plot\left(1 + .5 \cdot t - .5 \cdot t^2, t = -1.5 ..2.5, color = black\right) :$
$plot2 := pointplot(\{[-1, 0], [0, 1], [1, 1], [2, 0]\}, color = red, symbol = circle, symbolsize = 18) :$
$display(\{plot1, plot2\}, title = 'oops!');$



$BMI = C \cancel{\frac{height^2}{weight}}$

$= C \frac{weight}{height^2}$

Math 2270-004
April 11, 2018
Applying least squares linear regression to obtain power law fits

How do you test for power laws?

Suppose you have a collection of n data points

$$[[x_1, y_1], [x_2, y_2], [x_3, y_3], \ldots, [x_n, y_n]]$$

and you expect there may be a good power-law fit

$$\boxed{y = b\, x^m}$$

which approximately explains how the y_i's are related to the x_i's . You would like to find the "best possible" values for b and m to make this fit. It turns out, if you take the ln-ln data, your power law question is actually just a best-line fit question:
Taking (natural) logarithms of the proposed power law yields

$$\ln(y) = \ln(b) + m \ln(x).$$

So, if we write $\underline{Y = \ln(y)}$ and $\underline{X = \ln(x)}$, $B = \ln(b)$, this becomes the equation of a line in the new variables X and Y:

$$Y = m X + B \qquad e^Y = y \quad e^B = b$$

Thus, in order for there to be a power law for the original data, the ln-ln data should (approximately) satisfy the equation of a line, and vise verse. If we get a good line fit to the ln-ln data, then the slope m of this line is the power relating the original data, and the exponential e^B of the Y-intercept is the proportionality constant b in the original relation $y = b\, x^m$. With real data it is not too hard to see if the ln-ln data is well approximated by a line, in which case the original data is well-approximated by a power law.

Astronomical example As you may know, Isaac Newton was motivated by Kepler's (observed) Laws of planetary motion to discover the notions of velocity and acceleration, i.e. differential calculus and then integral calculus, along with the inverse square law of planetary acceleration around the sun.....from which he deduced the concepts of mass and force, and that the universal inverse square law for gravitatonal attraction was the ONLY force law depending only on distance between objects, which was consistent with Kepler's observations! Kepler's three observations were that

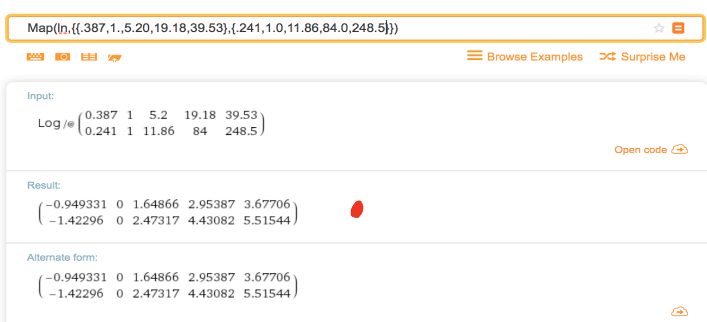
(1) Planets orbit the sun in ellipses, with the sun at one of the ellipse foci.
(2) A planet sweeps out equal areas from the sun, in equal time intervals, independently of where it is in its orbit.
(3) The square of the period of a planetary orbit is directly proportional to the cube of the orbit's semi-major axis.

$T^2 = kR^3$ $\qquad$ $T = kR^{3/2}$

So, for roughly circular orbits, Keplers third law translates to the statement that the period t is related to the radius r, by the equation $t = b\, r^{1.5}$, for some proportionality constant b. Let's see if that's consistent with the following data:

Planet	mean distance r from sun (in astronomical units where 1=dist to earth)	Orbital period t (in earth years)
Mercury	0.387	0.241
Earth	1.	1.
Jupiter	5.20	11.86
Uranus	19.18	84.0
Pluto	39.53	248.5

Taking the (natural) logarithm of the data points, as put into a matrix, using Wolfram alpha.



ln(radius)   ln(period)

We want the least squares solution to the ln-ln data, $Y = m\,X + B$

$$\begin{bmatrix} -.9493 & 1 \\ 0 & 1 \\ 1.64866 & 1 \\ 2.95387 & 1 \\ 3.67706 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} -1.42296 \\ 0 \\ 2.47317 \\ 4.43082 \\ 5.51544 \end{bmatrix}.$$

$$A\,\underline{\hat{x}} = \underline{\hat{b}}$$

$$A^T A\,\underline{x} = A^T\,\underline{b}$$

$$\widehat{\begin{bmatrix} m \\ b \end{bmatrix}} \qquad \underline{\hat{x}} = \left(A^T A\right)^{-1} A^T\,\underline{b}$$

I didn't have time (yet) to do these steps neatly at Wolfram alpha. In Maple:

> $with(LinearAlgebra):$

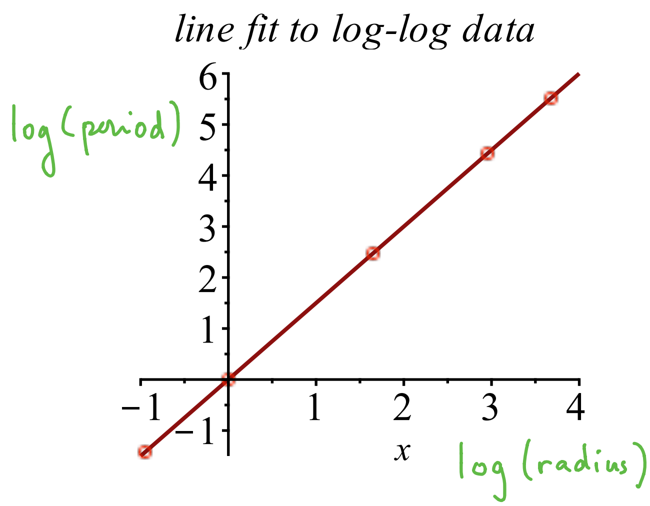
$$A := \begin{bmatrix} -.9493 & 1 \\ 0 & 1 \\ 1.64866 & 1 \\ 2.95387 & 1 \\ 3.67706 & 1 \end{bmatrix}:$$

$$\left(Transpose(A).A\right)^{-1}.Transpose(A).\begin{bmatrix} -1.42296 \\ 0 \\ 2.47317 \\ 4.43082 \\ 5.51544 \end{bmatrix};$$

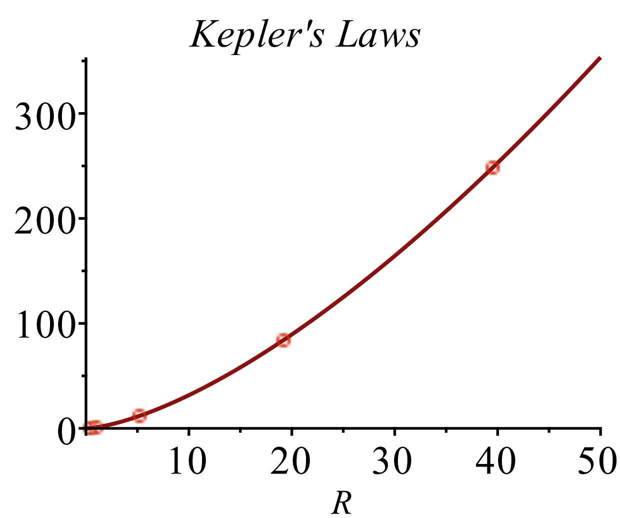
$$\begin{bmatrix} 1.49982355212829 \\ 0.000465682813906573 \end{bmatrix} \begin{matrix} \leftarrow m \\ \leftarrow b \end{matrix} \qquad 1.5 \tag{1}$$

So we get essentially the correct power.

> with(plots) :
plot1 := pointplot({[-.9493, -1.42296], [0, 0], [1.64866, 2.47317], [2.95387, 4.43082], [3.67706, 5.51544]}, color = red, symbol = circle, symbolsize = 18) :
plot2 := plot(1.4998·x + .0005, x = -1 ..4) :
display({plot1, plot2}, title = `line fit to log-log data`);



> plot3 := pointplot({[.387, .241], [1., 1.], [5.20, 11.86], [19.18, 84.0], [39.53, 248.5]}, color = red, symbol = circle, symbolsize = 18) :
plot4 := plot(exp(0.00046568)·$R^{1.49982}$, R = 0 ..50) :
display({plot3, plot4}, title = `Kepler's Laws`);



Fri Apr 13
- 6.7-6.8 Introduction to inner product spaces.

Announcements: .jpg show & tell on Monday I hope.

Warm-up Exercise: look over HW assignment

Flow chart of dot product development in $\mathbb{R}^n$

$$\vec{x}\cdot\vec{y} := \sum_{i=1}^{n} x_i y_i$$

algebra

a) $\vec{x}\cdot\vec{y} = \vec{y}\cdot\vec{x}$ (symmetry)

b) $\vec{x}\cdot(\vec{y}+\vec{z}) = \vec{x}\cdot\vec{y} + \vec{x}\cdot\vec{z}$
$\quad \vec{x}\cdot(k\vec{y}) = k\,\vec{x}\cdot\vec{y}$ (linear in each factor)

c) $\vec{x}\cdot\vec{x} \geq 0$; $\vec{x}\cdot\vec{x} = 0$ iff $\vec{x} = \vec{0}$ (positive)

from algebra...

$\|\vec{x}\|^2 = \vec{x}\cdot\vec{x}$

magnitude (norm)

$\vec{x}\cdot\vec{y} = 0$

orthogonal

$\mathrm{proj}_L \vec{x}$
$L = \mathrm{span}\{\vec{v}\}$

$\vec{x}\cdot\vec{y} = \|\vec{x}\|\|\vec{y}\|\cos\theta$ (angles)

distance from $\vec{x}$ to $\vec{y}$ $= \|\vec{y}-\vec{x}\|$

Pythagorean Theorem

orthonormal basis for $W \subset \mathbb{R}^n$, coords for such basis

Gram-Schmidt algorithm

$|\vec{x}\cdot\vec{y}| \leq \|\vec{x}\|\|\vec{y}\|$ Cauchy Schwarz inequality

$\|\vec{x}+\vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ triangle inequality (for estimates)

projection onto W
$\mathbb{R}^n = W \oplus W^\perp$

• i.e. $\vec{x} = \underbrace{\mathrm{proj}_W \vec{x}}_{\in W} + \underbrace{\vec{z}}_{\in W^\perp}$, uniquely

nearest point to $\vec{b} \in \mathbb{R}^n$ in W is $\mathrm{proj}_W \vec{b}$

Least squares solutions to $A\vec{x} = \vec{b}$

data fitting

linear regression

polynomial fits

power laws.

An inner product space is a (real scalar) vector space V together with an inner product $\langle\cdot,\cdot\rangle$ which gives a real number for each pair of vectors, s.t. the following axioms hold, $\forall f,g,h \in V$, $k \in \mathbb{R}$:

• a) $\langle f,g\rangle = \langle g,f\rangle$

• b) $\langle f, g+h\rangle = \langle f,g\rangle + \langle f,h\rangle$

• $\quad \langle f, kg\rangle = k\langle f,g\rangle$

• c) $\langle f,f\rangle \geq 0$. $\langle f,f\rangle = 0$ iff $f = 0$

From these algebra axioms the entire concept chart on the left also holds, for finite dimensional subspaces W.

Examples of function space inner products:

$$V = \{ f : [a, b] \rightarrow \mathbb{R} \ s.t. f \text{ is continuous} \} := \underline{C([a, b])}.$$

$$\langle f, g \rangle := \int_a^b \cancel{f(t)\, dt}\ f(t)g(t)dt \quad \text{(or some fixed positive multiple of this integral).}$$

Exercise 1) Check the algebra requirements a), b), c) for an inner product.

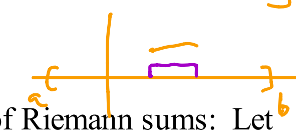
$\vec{x}\cdot\vec{y} = \vec{y}\cdot\vec{x}$ a) $\langle f, g \rangle = \langle g, f \rangle$ $\quad \int_a^b f(t)g(t)dt = \int_a^b g(t)f(t)dt$

$\vec{x}\cdot(\vec{y}+\vec{z}) = \vec{x}\cdot\vec{y} + \vec{x}\cdot\vec{z}$ b) $\langle f, g+h \rangle = \langle f, g \rangle + \langle f, h \rangle$ $\quad \int_a^b f(t)(g(t)+h(t))\,dt$

✓ $= \int_a^b f(t)g(t)\,dt + \int_a^b f(t)h(t)\,dt$

c) $\langle f, f \rangle \geq 0$, $\langle f, f \rangle = 0$ if only if $f = 0$ $\quad \int_a^b f(t)^2 dt \geq 0.$

if f is continuous & $\int_a^b f(t)^2 dt = 0$ $\Rightarrow f = 0$:



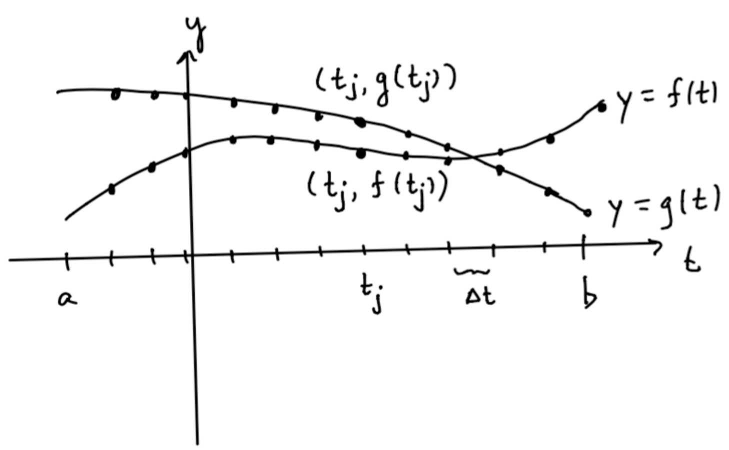
This inner product $\langle f, g \rangle$ is not so different from the $\mathbb{R}^n$ dot product if you think of Riemann sums: Let

$$\Delta t = \frac{b-a}{n}; \qquad t_j = a + j\,\Delta t, j = 1, 2, .. n.$$

Then

$$\langle f, g \rangle = \int_a^b \cancel{f(x)\, dx}\ f(t)g(t)dt = \lim_{n\to\infty} \sum_{j=1}^{n} f(t_j)\, g(t_j) \underbrace{\Delta t}_{= \frac{b-a}{n}}$$

$$= \lim_{n\to\infty} \left(\begin{bmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_n) \end{bmatrix} \cdot \begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_n) \end{bmatrix} \Delta t \right).$$



Prime examples:

<u>Example</u> For the inner product on $C[-1, 1]$ given by

$$\langle f, g\rangle := \int_{-1}^{1} f(t)g(t)\,\mathrm{d}t$$

If one applies Gram-Schmidt to the the set $\{\underset{w_1}{1}, \underset{w_2}{t}, \underset{w_3}{t^2}, t^3, \ldots\}$ one creates the (normalized) *Legendre polynomials* which have an interesting entry at Wikipedia. Projecting a continuous function f onto

$$W_n = span\{1, t, t^2, \ldots t^n\}$$

will create polynomical approximations, that improve in the sense that

$$\lim_{n\to\infty} \left\| f - proj_{W_n} f \right\|^2 = 0.$$

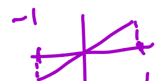
$\vec{u}_1 = \dfrac{\vec{w}_1}{\|\vec{w}_1\|}$ $\qquad \|\vec{w}_1\| = \langle \vec{w}_1, \vec{w}_1\rangle^{1/2} = \left(\int_{-1}^{1} (1)(1)\,dt\right)^{1/2} = \sqrt{2}$

$w_1 =$ "1"

$\vec{u}_1 =$ is the fcn
$u_1(t) = \dfrac{1}{\sqrt{2}}$

$\vec{w}_2 \perp \vec{w}_1$ or $\vec{u}_1$

$$\vec{z}_2 = \vec{w}_2 - \cancel{(\vec{w}_2\cdot\vec{u}_1)\vec{u}_1}$$



$$\vec{w}_2\cdot\vec{u}_1 = \int_{-1}^{1} t\,\frac{1}{\sqrt{2}}\,dt = \frac{t^2}{2\sqrt{2}}\Big]_{-1}^{1} = 0$$

$\vec{u}_2 = \dfrac{\text{"}t\text{"}}{\|\text{"}t\text{"}\|}$ $\quad\ldots$

Example for the inner product on $C[-\pi, \pi]$ given by

$$\langle f, g\rangle := \frac{1}{\pi}\int_{-\pi}^{\pi} \cancel{f(t)\,dt}\; f(t)\,g(t)\,dt$$

$$\left\langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\,dt = \frac{1}{\pi}\frac{1}{2}\int_{-\pi}^{\pi} 1\,dt = \frac{2\pi}{2\pi} = 1$$

The infinite set of functions

$$\left\{\frac{1}{\sqrt{2}}, \cos(t), \sin(t), \cos(2\,t), \sin(2\,t), \ldots, \sin(nt), \cos(nt), \ldots..\right\}$$

is already orthonormal! Thus begins the subject of *Fourier Series*. (See Wikipedia.)

To show the ortho-normality properties one applies the following trig identities, which follow from the addition angle formulas

$$\cos(m\,t)\,\cos(n\,t) = \frac{1}{2}\left[\cos((m+n)t) + \cos((m-n)\,t)\right]$$

$$\cos^2(n\,t) = \frac{1}{2}\left[\cos(2\,n\,t) + 1\right]$$

$$\sin(m\,t)\,\sin(n\,t) = \frac{1}{2}\left[-\cos((m+n)\,t) + \cos((m-n)t)\right]$$

$$\sin^2(n\,t) = \frac{1}{2}\left[-\cos(2\,n\,t) + 1\right]$$

$$\cos(m\,t)\,\sin(n\,t) = \frac{1}{2}\left[\sin((m+n)t) + \sin((-m+n)t)\right]$$

Exercise verify how ortho-normality follows from these identities.

$\langle \cos mt, \cos nt \rangle = 0 \quad m \neq n.$

$$\frac{1}{\pi}\int_{-\pi}^{\pi} \cos(mt)\cos(nt)\,dt = \frac{1}{\pi}\int_{-\pi}^{\pi} \frac{1}{2}\left(\cos(m+n)t + \cos(m-n)t\right)dt$$

$$= \frac{1}{\pi}\left[\frac{1}{2}\left(\frac{\sin(m+n)t}{m+n} + \frac{\sin(m-n)t}{m-n}\right)\right]_{-\pi}^{\pi}$$

$$= \frac{1}{\pi}(0-0) = 0$$

$$\|\cos nt\|^2 = \langle \cos nt, \cos nt\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \underbrace{\frac{1}{2}\left[\cos 2nt + 1\right]}_{\cos^2 nt}\,dt = \frac{1}{\pi}\frac{1}{2}\left(\frac{\sin 2nt}{2n} + t\right)\Big]_{-\pi}^{\pi}$$

$$= \frac{1}{\pi}\frac{1}{2}\left(0 + \pi - (0 - \pi)\right)$$

$$= \frac{2\pi}{2\pi} = 1$$

Let $V_n := span\left\{\frac{1}{\sqrt{2}}, \cos(t), \sin(t), \cos(2t), \sin(2t), ... \cos(nt), \sin(nt)\right\}$ be the $2n+1$ dimensional subspace spanned by the first $2n+1$ of these functions. A deep theorem says that if $f \in C(-\pi, \pi)$ (actually, f only needs to be piecewise continous), then

$$\lim_{n \to \infty} \left\| f - proj_{V_n} f \right\| = 0.$$

Because we have an orthonormal basis for V_n the projection formula is easy to write down:

$proj_W \vec{x} = (\vec{x}\cdot\vec{u}_1)\vec{u}_1 + (\vec{x}\cdot\vec{u}_2)\vec{u}_2 + \cdots + (\vec{x}\cdot u_N)\vec{u}_N$

$$proj_{V_n} f = \left\langle f, \frac{1}{\sqrt{2}} \right\rangle \frac{1}{\sqrt{2}} + \langle f, \cos(t)\rangle \cos(t) + \langle f, \sin(t)\rangle \sin(t) + ... + \langle f, \cos(nt)\rangle \cos(nt) + \langle f, \sin(nt)\rangle \sin(nt)$$

$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)g(t)\,dt$

We write

$$a_0 = \langle f, 1\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\,\mathrm{d}t$$

$$a_k = \langle f, \cos(kt)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)\,\mathrm{d}t$$

$$b_k = \langle f, \sin(kt)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)\,\mathrm{d}t.$$

Then

$$proj_{V_n} f = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos(kt) + \sum_{k=1}^{n} b_k \sin(kt).$$

The infinite series converges to $f(t)$ pointwise at places where f is differentiable, and to the average of right and left hand limits at jump discontinuities, so we also often consider the infinite Fourier series

$$f(t) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(kt) + \sum_{k=1}^{\infty} b_k \sin(kt).$$

$$f(t) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(kt) + \sum_{k=1}^{\infty} b_k \sin(kt).$$

$$a_0 = \langle f, 1 \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\, \mathrm{d}t \qquad a_k = \langle f, \cos(kt) \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)\, \mathrm{d}t$$

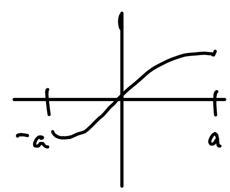
$$b_k = \langle f, \sin(kt) \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(kt)\, \mathrm{d}t.$$

<u>Exercise</u>: Define $f(t) = t$, on the interval $-\pi < t < \pi$. Show

$$t \sim 2\sum_{k=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(nt)$$

$$a_0 = \langle f, 1 \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} \underbrace{t \cdot 1}_{g(t) = t \text{ is odd}}\, dt = 0$$

$g(-t) = -g(t)$



$$\int_{-a}^{a} g(t)\, dt = 0.$$

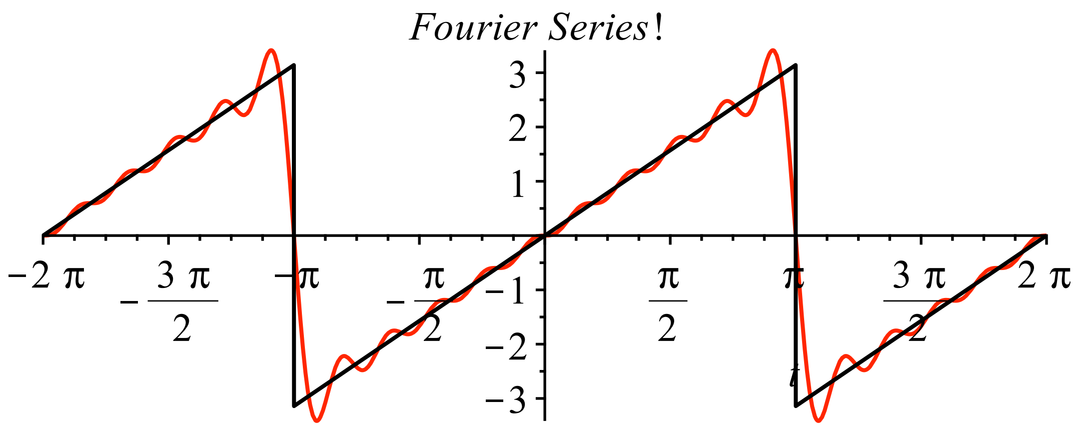
finish on Monday!

$proj_{V_{10}} f(t)$:

```
> with(plots) :
plot1 := plot(t + 2·π − 2·π·Heaviside(t + π) − 2·π·Heaviside(t − π), t=-2·π..2·π, color
    = black) :
```

$$plot2 := plot\left(2 \cdot \sum_{n=1}^{10} (-1)^{n+1} \cdot \frac{\sin(n \cdot t)}{n}, t=-2\cdot\pi..2\cdot\pi, color = red\right):$$

```
display({plot1, plot2}, title='Fourier Series!');
```




```
>
>
```