6 Orthogonality

Definition An inner product $(\cdot,\cdot)$ is a function from $V \times V \to$ scalar (where V is a vector space) with the following properties:

- $(\alpha x, y) = \alpha (x, y)$ for all $x, y \in V$ } linearity in first arg
- $(x+y, z) = (x,z) + (y,z)$ for all $x,y,z \in V$
- $(x,y) = (y,x)$ } symmetry
- $(x,x) \geq 0$ for all $x \in V$ } pos. definiteness
  with $(x,x) = 0 \iff x = 0$

Q [linear in second argument? (yes: flip, use linearity in first, flip again

Example $(x,y) = x_1 y_1 + \cdots + x_n y_n$
for coordinate vectors x and y
specifically this IP is called the dot product: $x \cdot y$

lec 12 review inner prod. props hw2 due in a week
review quiz

Facts:
- x and y called "orthogonal" iff $(x,y) = 0$. "$x \perp y$"
  aka perpendicular

- Inner products make norms: $\|x\| := \sqrt{|(x,x)|}$

[Abs value needed?
Dot product makes which norm?

- Pythagorean theorem

$x \perp y \Rightarrow \|x+y\|^2 = \|x\|^2 + \|y\|^2$

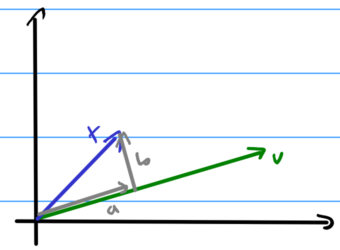
- For $x, v \in V$, there exist a, b s.t.

$$x = \underbrace{a}_{x^{\parallel v}} + \underbrace{b}_{x^{\perp v}}$$

with

$a = \alpha v$

$b \perp v$



$$\rightsquigarrow \quad (b, v) = 0$$

$$\overset{b = x - a}{\Leftrightarrow} \quad (x - a, v) \overset{!}{=} 0$$

$$\Leftrightarrow \quad (x - \alpha v, v) \overset{!}{=} 0$$

$$\Leftrightarrow \quad (x, v) - \alpha (v, v) \overset{!}{=} 0$$

$$\Leftrightarrow \quad \alpha = \frac{(x, v)}{(v, v)}$$

$$\rightsquigarrow \quad x^{\perp v} = x - \frac{(x, v)}{(v, v)} v$$

$$x^{\parallel v} = x - x^{\perp v}$$

Demo

[What happens to this if $\|v\| = 1$? denominator drops out

What is the closest point to x on the line βv (for all β)? a

Why? (Pyth.) How can we compute it? (see above)]

- Can define a line/ plane/ hyperplane this way.

$x = a + \alpha b$ (in general: $n-1$ direction vectors (in nD space)) $\rightsquigarrow$ find n with $(n, b) = 0$ [How? Nullsp. finding

Example: line

$\rightsquigarrow \quad (n, x) = (n, a) + \alpha (n, b) = 0$

$\rightsquigarrow \quad (n, x) = (n, a)$ defines the plane

$\rightsquigarrow$ If $\|n\| = 1$, $(n, x) - (n, a)$ computes distance from plane [Why? (Plug $a - x$ into norm.)

n called the normal

- <u>Orthogonal basis</u> $(b_i) \subset V$: $b_i \perp b_j$ if $i \neq j$
  ("pairwise $\perp$")

  (and needs to be a basis)

- <u>Orthonormal basis</u> (ONB): orthogonal basis and $\|b_i\| = 1$.

  How do I find coordinates with respect to an ONB? Very easy!
  $\rightsquigarrow \; x = (x,b_1)b_1 + (x,b_2)b_2 + \dots + (x,b_n)b_n$.

  How would I build a matrix that finds these coordinates?
  (assuming the dot product as the inner product)

- <u>Orthogonal matrix</u>: orthonormal basis (of entire space) for columns

  Suppose Q is orth. $QQ^T = ?$ (Identity)

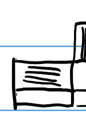
  Is Q square? (yes, because of "basis" requirement)

  Is Q^T orthogonal as well? (ie is $Q^TQ = I$?)

  $Q^{-1} = ?$ (Q^T)

  What if the orthonormal vectors in Q are not a full basis?

$Q =$  $\rightsquigarrow QQ^T =$  $? \; =: P$

$P^2 = Q\underbrace{Q^TQ}Q^T = Q$   $Q^T = P$

Identity restricted to colspace(Q).

A function with $f(f(x)) = f(x)$ is called a projection.

QQ^T is an orthogonal projection onto colspace (Q).

Demo Orthogonal projection

<inclass-orthogonality>

lec 13

CF comments:

- multiple sessions: fixed
- too much clicking: will fix
- hw grace period: will fix
- code area resizable: will fix
- email-based log-in...
- don't combine poll & quiz

Review quiz
Outline!
Show G-S movie
Mod G-S

Orthogonalization/ Gram-Schmidt

Orthonormal vectors seem useful. How do we make many of them?

- Know how to make two vectors orthogonal
- Need pairwise orthogonality beyond that

Demo Orthogonalizing three vectors

Movie Gram-Schmidt

Demo Gram-Schmidt and Modified Gram-Schmidt

Demo Keeping track of the coefficients → QR factorization

Application of QR: Solving linear systems [How?

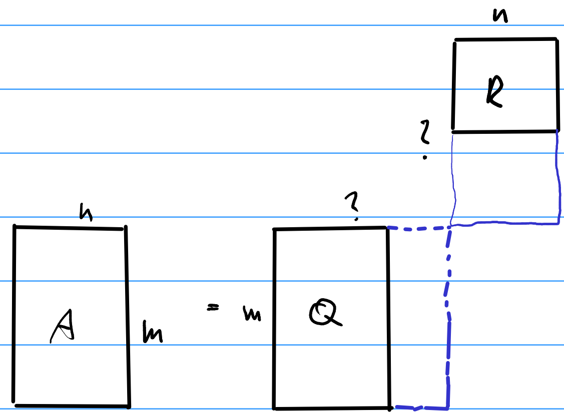
$$Ax=b \rightsquigarrow Q\underbrace{Rx}_{y}=b \rightsquigarrow Qy=b \rightsquigarrow Rx=y$$

[How? $Q^Tb=y$

[How? back-subst

[Asymptotic cost? $O(n^3)$, but practically more expensive than LU

What about QR of non-square matrices? ($\to$ hw)



A: $m \times n$, $m \geq n$

Either:

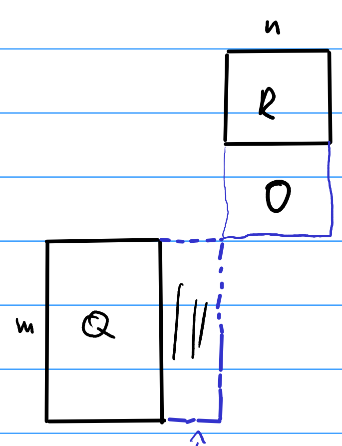
Q $m \times m$ square
R: $m \times n$ "full"

Or:

Q: $m \times n$
R: $n \times n$ square "thin"

[What does Gram-Schmidt produce? "thin"

How can we turn thin into full?



some more orthogonal vectors
(can start G-S from random ones: why?)

Demo

What happens to 2-norms before/after an orthogonal matrix is applied?

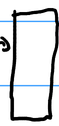
$$\|Qx\|_2^2 = (Qx)\cdot(Qx) = (Qx)^T(Qx) = x^TQ^TQx = x^Tx = \|x\|_2^2 \,!$$

In other words: Orthogonal matrices preserve the 2-norm.

[Can Q be non-square here? No.

[Which step breaks if Q is non-square?

"Solving" overdetermined linear systems

$Ax=b$ with $A \rightarrow$  tall and skinny

$\rightsquigarrow$ bad news:
- many more equations than unknowns
- won't find an x that works exactly unless we're very lucky
- best hope in general: minimize residual $\|Ax-b\|$

$$\min_x \|Ax-b\|_2^2 = (Ax-b)_1^2 + (Ax-b)_2^2 + \cdots + (Ax-b)_n^2$$

"least-squares problem"

How do we solve least squares problems?

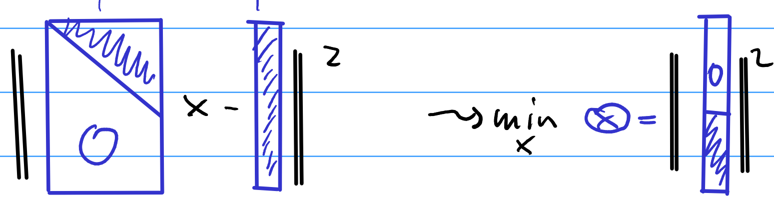
Just learned QR — maybe that can help.

$$\otimes = \|QRx - b\|_2^2 \overset{\text{why?}}{=} \|Q^T(QRx - b)\|_2^2$$

[Full QR? Reduced QR?

$$= \|\underbrace{Q^TQ}_{\mathrm{Id}}Rx - Q^Tb\|_2^2$$

$$= \|Rx - \underbrace{Q^Tb}\|_2^2$$



[How do we minimize that? By solving $Rx = (Q^Tb)_{\text{(upper)}}$.

Lec 14

review QR for LSQ

review questions below

review data fitting

review quiz

outline

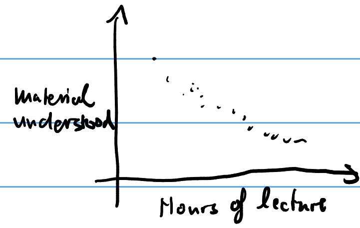
[What's the norm of the residual going to be? $\|(Q^Tb)_{\text{(lower)}}\|$

[What's the relationship between the residual and the columnspace of $A^\perp$?

Data Fitting

Have: Data points

$$(x_i, \; y_i)_i$$



Have: Idea for a Model → How does y depend on x?

$$y = f(x) = a + bx$$

for example

but: don't know

Want: values for a, b that lead to small ~~error~~ 2-norm residual

$$a + bx_1 \overset{\approx}{\cancel{=}} y_1$$

$$\vdots$$

$$a + bx_n \overset{\approx}{\cancel{=}} y_n$$

$\rightsquigarrow$ to matrix form

$$\begin{pmatrix} 1 & x_1 \\ 1 & \vdots \\ 1 & \\ 1 & \\ 1 & \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \overset{\approx}{\cancel{=}} \begin{pmatrix} y_1 \\ \vdots \\ \\ y_n \end{pmatrix}$$

Demo: Data Fitting with Least Squares

General form of model that least squares can handle:

$$f(x) = a_1 \cdot \varphi_1(x) + a_2 \varphi_2(x) + \cdots + a_n \varphi_n(x)$$

Σ coefficient · function

unknown given

<in class-least-squares-2>

lec15

outline

review quiz

hw3 out later today

questions:

svd numerical example

Focus on coding aspects?

what's the svd good for?

how do you actually compute it?

Another application of least squares:

Machine learning Demo: Predicting Breast Cancer