Orthogonality

Definition: An inner product \( \langle \cdot, \cdot \rangle \) is a function from \( V \times V \rightarrow \text{scalar} \) (where \( V \) is a vector space) with the following properties:

- \( \langle x, y \rangle = \langle y, x \rangle \) for all \( x, y \in V \) \( \Rightarrow \) symmetry
- \( \langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \) for all \( x, y, z \in V \) \( \Rightarrow \) linearity in first arg
- \( \langle x, y \rangle = \langle y, x \rangle \) for all \( x, y \in V \) \( \Rightarrow \) conjugate symmetry
- \( \langle x, x \rangle \geq 0 \) for all \( x \in V \) \( \Rightarrow \) positive definiteness
  with \( \langle x, x \rangle = 0 \Leftrightarrow x = 0 \)

0. Linear in second argument? (yes: flip, use linearity in first, flip again)

Example: \( \langle x, y \rangle = x_1y_1 + \cdots + x_ny_n \)
  for coordinate vectors \( x \) and \( y \)

Specifically, this IP is called the dot product: \( x \cdot y \)

Facts:
- \( x \) and \( y \) called "orthogonal" iff \( \langle x, y \rangle = 0 \). "\( x \perp y \)"
  \( \perp \) aka perpendicular
- Inner products make norms: \( \|x\| = \sqrt{\langle x, x \rangle} \)

Also: 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 $a, b \in \mathbb{R}$ st.

$$x = a' + b' v$$

with

$$a = a' v$$

$$b = b' v$$

$$\Rightarrow (b, v) = 0$$

$$\Rightarrow (x - a, v) = 0$$

$$\Rightarrow (x - a v, v) = 0$$

$$\Rightarrow (x, v) - a (v, v) = 0$$

$$\Rightarrow a = \frac{(x, v)}{(v, v)}$$

$$\Rightarrow x \perp v = x - \frac{(x, v)}{(v, v)} v$$

$$x^v = x - x \perp v$$  

\[\text{Denote}\]

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

What is the closest point to $x$ on the line $\beta v$ (for all $\beta$)?

Why? (Python) How can we compute it? (see above)

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

  in general, a direction vector (in $n$-space)

  $$x = a + \lambda b \Rightarrow \text{find } \lambda \text{ with } (n, b) = 0$$  

  Example: line

  $$(n, x) = (n, a) + \lambda (n, b) = 0$$

  $$(n, x) = (n, a)$$  

  defines the plane

  If $\|n\| = 1$, $(n, x) - (n, a)$

  compute distance from plane  

  $n$ called the normal
- Orthogonal basis \((b_i) \subset V: b_i \perp b_j \text{ if } i \neq j\)
  
  (and needs to be a basis)

- Orthonormal basis (ONB): orthogonal basis and \(\|b_i\| = 1\).

  How do I find coordinates with respect to an ONB? Very easy:
  \[
  x = (x, b_1)b_1 + (x, b_2)b_2 + \cdots + (x, b_n)b_n.
  \]

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

- Orthogonal matrix: orthornormal basis (of subspace), for columns

  Suppose \(Q\) is orth. \(QQ^\top = ?\) \(\text{ (identity)}\)

  Is \(Q\) square? \((\text{yes, because of "basis" requirement})\)

  Is \(Q^\top\) orthogonal as well? \((\text{is } Q^\top Q = I?)\)

  \(Q^\top = ?\) \((Q^\top)\)

  What if the orthornormal vectors in \(Q\) are not a full basis?

  \[
  Q = \mathbb{I} \implies QQ^\top = \mathbb{I} \implies Q^\top = Q
  \]

  \[
  P^2 = QQ^\top QQ^\top = Q \implies Q^\top = P
  \]

  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>
Orthogonalization/ Gram-Schmidt

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

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

Demos Orthogonalizing three vectors

Movie Gram-Schmidt

Demos Gram-Schmidt and Modified Gram-Schmidt

Demos Keeping track of the coefficients $\rightarrow$ QR factorization

Application of QR: Solving linear systems $\square$ How?

$A\mathbf{x} = \mathbf{b} \rightarrow QR\mathbf{x} = \mathbf{b} \rightarrow Q\mathbf{y} = \mathbf{b} \rightarrow R\mathbf{x} = \mathbf{y}$ [How?] [How?]

$Q^T\mathbf{b} = \mathbf{y}$ back-substitute

Asymptotic cost? $O(n^3)$, but practically more expensive than LU
What about QR of non-square matrices? (→ hw)

$A: \text{m} \times \text{n}, \text{m} \geq \text{n}$

Either:

- $Q: \text{m} \times \text{m}$ square \text{“full”}
- $R: \text{m} \times \text{n}$

Or:

- $Q: \text{m} \times \text{n}$ \text{“thin”}
- $R: \text{n} \times \text{n}$ square

[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?

\[ \|Qx\|_2^2 = (Qx)^T (Qx) = (Q^T Q) x = x^T x = \|x\|_2^2 \]

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 system

\[ Ax = b \quad \text{with} \quad A \in \mathbb{R}^{m \times n}, \quad m > n \]

- Bad news: *many* more equations than unknowns
  - Won't find an \( x \) that works *exactly*
  - Unless we're very lucky
  - Best hope in general:
    \[ \text{minimize residual } \|Ax - b\| \]

\[ \min_{x} \|Ax - b\|_2^2 = (Ax - b)_1^2 + (Ax - b)_2^2 + \cdots + (Ax - b)_n^2 \]

"Least-squares problem"
How do we solve least squares problems?

Just learned QR — maybe that can help.

\[ \| Q R x - b \|_2^2 = \| Q (Q R x - b) \|_2^2 \]

\[ = \| Q^T Q R x - Q^T b \|_2^2 \]

\[ = \| R x - Q^T b \|_2^2 \]

\[ \overset{\text{why?}}{=} \]

[Full QR? Reduced QR?]

\[ \begin{bmatrix} x \\ z \end{bmatrix} \]

\[ \overset{\text{By solving } R x = (Q^T b)_{\text{upper}}}{=} \]

\[ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]

[How do we minimize that?]
Data Fitting

Have: Data points

\[(x_i, y_i)\] for \(i = 1, 2, \ldots, n\)

Have: Idea for a Model \(\Rightarrow\) How does \(y\) depend on \(x\)?

\[y = f(x) = a + bx\]

but: don't know

Want: Values for \(a, b\) that lead to small errors.

\[\|a + bx - y\|_2\] = 2-norm residual

\[a + bx \approx y,\]

\[
\begin{pmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{pmatrix}
\begin{pmatrix}
  a \\
  b
\end{pmatrix}
\approx
\begin{pmatrix}
  y_1 \\
  \vdots \\
  y_n
\end{pmatrix}
\]

to matrix form

\[a + bx \approx y_n\]

Demo: Data Fitting with Least Squares

General form of model that least squares can handle:

\[f(x) = \alpha_1 \psi_1(x) + \alpha_2 \psi_2(x) + \cdots + \alpha_n \psi_n(x)\]

\(\exists\) coefficient function

\[\begin{array}{c}
\alpha_1 \\
\vdots \\
\alpha_n
\end{array}\]

unknown \(\Rightarrow\) given

\(<\text{in class: least squares}\>\)
10:35 AM

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