October 1, 2024

Announcements

- Exam 1 grades
- Exam 2
  - HW4 due this week
    - EC for typed hw
    - numerics debug

↳ hypothesis generation!

assert (. . . ,)

bool

Goals

LSQ sensitivity

QR → LSQ

Computing QR

- Gram-Schmidt

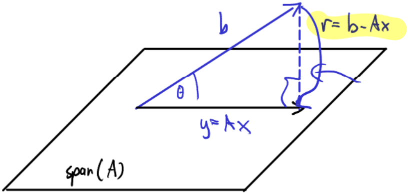
Review

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

Least Squares, Viewed Geometrically (II)

$A\tilde{x} \cong \tilde{b}$

$\min \|A\tilde{x} - \tilde{b}\|_2$



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

$\tilde{x} = (A^TA)^{-1}A^T\tilde{b}$

Phrase the Pythagoras observation as an equation.

" $\operatorname{span}(A) \cdot (b - Ax) = 0$ "

$A^T(b - Ax) = 0 \iff$ normal equations

Write that with an orthogonal projection matrix P.

$$Pb = Ax = A\underbrace{(A^TA)^{-1}A^T}\tilde{b}$$

About Orthogonal Projectors

What is a *projector*?

$$P^2 P = P$$

What is an *orthogonal projector*?

P symm. (to show!)

How do I make one projecting onto span$\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_\ell\}$ for orthonormal $\boldsymbol{q}_i$?

$$Q = [q_1, \ldots, q_\ell]$$

$$P = QQ^T$$

Least Squares and Orthogonal Projection

Check that $P = A(A^TA)^{-1}A^T$ is an orthogonal projector onto colspan(A).

$$P^2 = A(\cancel{A^TA})^{-1}\cancel{A^T A}(A^TA)^{-1}A^T = P$$

$$\cancel{((A)^{-1})^T}$$

$$P^T = \left(A(A^TA)^{-1}A^T \right)^T$$

$$= \left(A^{TT}(A^TA)^{-T}A^T \right) = P.$$

What assumptions do we need to define the P from the last question?

A full rank.

Pseudoinverse

$\vec{x} = A^+ \vec{b}$ $\quad$ 5 $\quad$ 3000

What is the pseudoinverse of A?

$$A^+ = (A^T A)^{-1} A^T$$

What can we say about the condition number in the case of a tall-and-skinny, full-rank matrix?

$\|A\| \, \|A^{-1}\|$

$$\mathrm{cond}_2(A) = \|A\|_2 \|A^+\|_2$$

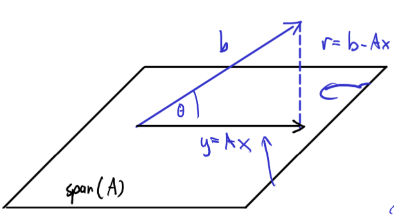
$\to \mathrm{cond}_2(A) = \infty$ if not full rank.

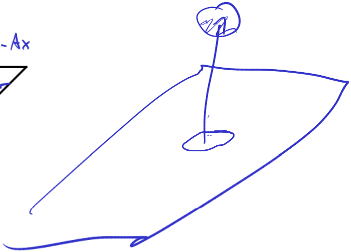
$\mathrm{cond}_2(A^T) = \infty$

What does all this have to do with solving least squares problems?

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

Sensitivity and Conditioning of Least Squares





Relate $\|A\boldsymbol{x}\|$ and $\|\boldsymbol{b}\|$ with θ via trig functions.

$$\cos(\theta) = \frac{\|Ax\|_2}{\|b\|_2}$$

Sensitivity and Conditioning of Least Squares (II)

Derive a conditioning bound for the least squares problem.

$$\vec{x} = A^+ \vec{b} \qquad \Delta\vec{x} = A^+ \Delta\vec{b}$$

$$\frac{\|\Delta \vec{x}\|}{\|x\|} \le \|A^+\| \cdot \frac{\|\Delta \vec{b}\|}{\|x\|}$$

$$= \frac{\kappa(A)}{\|A\|\,\cancel{\|A^+\|}}\,\cancel{\|A^+\|}\,\frac{\|b\|}{\|b\|}\,\frac{\|\Delta b\|}{\|x\|}$$

$$= \kappa(A) \cdot \frac{\|b\|}{\|A\|\,\|x\|}\,\frac{\|\Delta b\|}{\|b\|}$$

$$\le \kappa(A)\,\frac{\|b\|}{\|Ax\|}\,\frac{\|\Delta b\|}{\|b\|} = \kappa(A) \cdot \frac{1}{\cos(\theta)}\,\frac{\|\Delta b\|}{\|b\|}$$

What values of θ are bad?

$$\theta \approx \frac{\pi}{2} \qquad / \qquad \vec{b} \perp \operatorname{span}(A)$$

Sensitivity and Conditioning of Least Squares (III)

Any comments regarding dependencies?

sens. of $A\vec{x} \cong \vec{b}$ depends on A and $\vec{b}$

What about changes in the matrix?

$$\frac{\|\Delta \vec{x}\|_2}{\|\vec{x}\|_2} \leq \left(\kappa(A)^2 \tan(\theta) + \kappa(A) \right) \cdot \frac{\|\Delta A\|}{\|A\|}$$

Transforming Least Squares to Upper Triangular

$\|r\|_2 = \|(Q^T b)_{\text{bottom}}\|_2$

Suppose we have $A = QR$, with Q square and orthogonal, and R upper triangular. This is called a QR factorization.

How do we transform the least squares problem $A\boldsymbol{x} \cong \boldsymbol{b}$ to one with an upper triangular matrix?

$\|Qz\|_2 = \|z\|_2$

$$\|r\|_2 = \|b - Ax\|_2$$
$$= \|Q^T(b - Ax)\|_2$$
$$= \|Q^T(b - QRx)\|_2 = \|Q^T b - Rx\|_2$$

$A \in \mathbb{R}^{m \times n}$

$\vec{x}$ does not affect bottom of R

$(Rx)_{\text{top}} = (Q^T b)_{\text{top}}$

Simpler Problems: Triangular

What do we win from transforming a least-squares system to upper triangular form?

bw subst. on $(Rx)_{top} = (Q^T b)_{top}$

How would we minimize the residual norm?

Computing QR

$\varepsilon < \sqrt{\varepsilon_{mach}}$

- Gram-Schmidt
- Householder Reflectors $\Leftarrow$
- Givens Rotations

$\left\| \begin{pmatrix} 1 \\ \varepsilon \end{pmatrix} \right\|_2 = \sqrt{1^2 + \sqrt{\varepsilon_{mach}}^2}$

Demo: Gram-Schmidt–The Movie [cleared] (shows *modified G-S*)
Demo: Gram-Schmidt and Modified Gram-Schmidt [cleared]
Demo: Keeping track of coefficients in Gram-Schmidt [cleared]
Seen: Even modified Gram-Schmidt still unsatisfactory in finite precision arithmetic because of roundoff.

NOTE: Textbook makes further modification to 'modified' Gram-Schmidt:

- Orthogonalize *subsequent* rather than *preceding* vectors.
- Numerically: no difference, but sometimes algorithmically helpful.

Economical/Reduced QR

Is QR with square Q for $A \in \mathbb{R}^{m \times n}$ with $m > n$ efficient?