What about non-square systems?

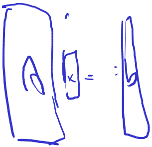
HW3 → next

Exercise 1.1 → going on, LU (without pivoting)

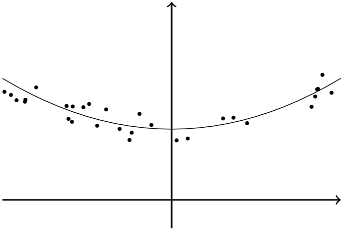
Specifically, what about linear systems with 'tall and skinny' matrices? (A: $m \times n$ with $m > n$) (aka *overdetermined* linear systems)

Specifically, any hope that we will solve those exactly?

hope



Example: Data Fitting



Have data: (x_i, y_i) and model:

$$y(x) = \alpha + \beta x + \gamma x^2$$

Find data that (best) fit model!

Data Fitting Continued

$$\begin{aligned}
\alpha + \beta x_1 + \gamma x_1^2 &= y_1\\
\alpha + \beta x_2 + \gamma x_2^2 &= y_2\\
\alpha + \beta x_3 + \gamma x_3^2 &= y_1\\
&\vdots\\
\alpha + \beta x_n + \gamma x_n^2 &= y_n
\end{aligned}$$

$$\begin{pmatrix} x_1 & y_1 \\ & \\ x_n & y_n \end{pmatrix}$$

$$\|Az - b\|_2^2 = \begin{aligned}
&|\alpha + \beta x_1 + \gamma x_1^2 - y_1|^2 +\\
&|\alpha + \beta x_2 + \gamma x_2^2 - y_2|^2 +\\
&|\alpha + \beta x_3 + \gamma x_3^2 - y_1|^2 +\\
&\quad\vdots\\
&|\alpha + \beta x_n + \gamma x_n^2 - y_n|^2
\end{aligned} \rightarrow \min$$

Rewriting Data Fitting

Rewrite in matrix form.

$$\left\| \underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \\ 1 & x_n & x_n^2 \end{pmatrix}}_{\text{Vandermonde matrices}} \underbrace{\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}}_{z} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\|^2 \to \min$$

Least Squares: The Problem In Matrix Form

$$\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 \to \text{min!}$$

is cumbersome to write.
Invent new notation, defined to be equivalent:

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

cong

NOTE:

- Data Fitting is *one example* where LSQ problems arise.
- Many other application lead to $A\boldsymbol{x} \cong \boldsymbol{b}$, with different matrices.

Data Fitting: Nonlinearity

Give an example of a nonlinear data fitting problem.

$$
\begin{aligned}
&\left|\overbrace{\exp(\alpha)}^{\alpha'} + \beta x_1 + \gamma x_1^2 - y_1\right|^2 \\
&\qquad\qquad + \cdots + \\
&\left|\exp(\alpha) + \beta x_n + \gamma x_n^2 - y_n\right|^2 \quad \to \quad \text{min!}
\end{aligned}
$$

But that would be easy to remedy: Do linear least squares with $\exp(\alpha)$ as the unknown. More difficult:

"genuine nonlinearity"

$$
\begin{aligned}
&\left|\alpha + \exp(\beta x_1 + \gamma x_1^2) - y_1\right|^2 \\
&\qquad\qquad + \cdots + \\
&\left|\alpha + \exp(\beta x_n + \gamma x_n^2) - y_n\right|^2 \quad \to \quad \text{min!}
\end{aligned}
$$

Demo: Interactive Polynomial Fit [cleared]

Properties of Least-Squares

Consider LSQ problem $A\boldsymbol{x} \cong \boldsymbol{b}$ and its associated *objective function* $\varphi(x) = \|\boldsymbol{b} - A\boldsymbol{x}\|_2^2$. Does this always have a solution?

$\varphi^{\frac{1}{2}}(x) \geq \big|\, \|Ax\|_2 - \|b\|_2 \,\big|$ A no null sp.

Is it always unique?

not unique if A has null sp.

Examine the objective function, find its minimum.

$$0 = \nabla\varphi(x) = \nabla(b - Ax)^T \cdot (b - Ax)$$
$$= \nabla(b^Tb - 2b^TAx + x^TA^TAx)$$
$$= -2b^TA + 2A^TAx \Rightarrow A^TAx = A^Tb.$$

(the normal eqn.)

Least squares: Demos

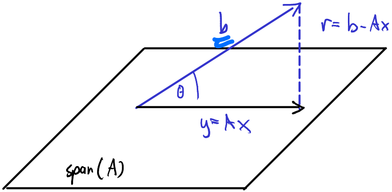
Demo: Polynomial fitting with the normal equations [cleared]

What's the shape of A^TA? $m \times n \rightarrow n \times n$.

square.

Demo: Issues with the normal equations [cleared]

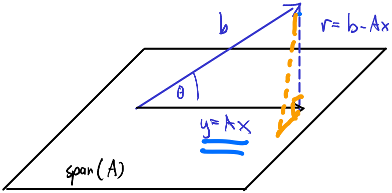
Least Squares, Viewed Geometrically



Why is $\boldsymbol{r} \perp \operatorname{span}(A)$ a good thing to require?

when $r \perp \operatorname{span}(A)$, $\|r\|_2$ is minimized,

Least Squares, Viewed Geometrically (II)



Phrase the Pythagoras observation as an equation.

$$\text{span}(A) \perp b - Ax$$

$$\Rightarrow A^T(b - Ax) = 0 \Rightarrow A^TAx = A^Tb$$

Write that with an orthogonal projection matrix P.

$$Ax = Pb$$

About Orthogonal Projectors

What is a *projector*?

$P: \quad P^2 = P$

What is an *orthogonal projector*?

$P^T = P \iff \mathrm{Im}(P) \perp \cancel{\mathrm{Range}}\ \mathrm{Ker}(P)$

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

QQ^T

Least Squares and Orthogonal Projection

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

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

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

symmetric. $\Rightarrow$ P is a projector.

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

A^TA is invertible.

Pseudoinverse

What is the pseudoinverse of A?

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

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

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

∞ if A is not full rank.

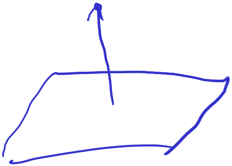
What does all this have to do with solving least squares problems?

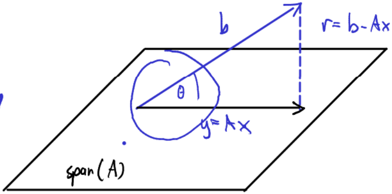
$x = A^{+}b$ solves $Ax \cong b$

In-Class Activity: Least Squares

In-class activity: Least Squares

Sensitivity and Conditioning of Least Squares





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

$$\frac{\|\Delta x\|_2}{\|x\|_2} \le \operatorname{cond} A \cdot \frac{1}{\cos\theta} \cdot \frac{\|\Delta b\|}{\|b\|}$$

What values of θ are bad?

when $b \perp \operatorname{colspan}(A)$ $\quad \theta \approx \frac{\pi}{2}$.

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

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

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

$$= \frac{\operatorname{cond}(A)}{\|A\|} \cdot \frac{\|\Delta b\|}{\|x\|}$$

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

$$\le \operatorname{cond}(A) \cdot \underbrace{\frac{\|b\|}{\|Ax\|}}_{\frac{1}{\cos\theta}} \cdot \frac{\|\Delta b\|}{\|b\|}$$

$$\|\cdot\| = \|\cdot\|_2$$

$$\operatorname{cond}(A) = \|A\| \, \|A^{+}\|$$

$$\|A^{+}\| = \frac{\operatorname{cond}(A)}{\|A\|}$$

Sensitivity and Conditioning of Least Squares (II)

Any comments regarding dependencies?

Sensitivity bound depends on b

What about changes in the matrix?

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

Two behaviors:

- $\tan\theta \approx 0$, then cond nr is approx. $\operatorname{cond}(A)$
- Otherwise $\operatorname{cond}(A)^2$.

Recap: Orthogonal Matrices

$Q = \begin{pmatrix} q_1 & \cdots & q_n \end{pmatrix}$ $\quad q_i^T q_j = 0 \quad i \neq j$

$q_i^T q_i = 1$

What's an *orthogonal* (*=orthonormal*) *matrix*?

One that satisfies $Q^T Q = I$ and $QQ^T = I$.

How do orthogonal matrices interact with the 2-norm?

$$\|Q\mathbf{v}\|_2^2 = (Q\mathbf{v})^T(Q\mathbf{v}) = \mathbf{v}^T Q^T Q\mathbf{v} = \mathbf{v}^T\mathbf{v} = \|\mathbf{v}\|_2^2.$$

Transforming Least Squares to Upper Triangular

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?

$$\min \leftarrow \|Ax - b\|_2 = \|QRx - b\|_2$$

$$= \|Q^T(QRx - b)\|_2 = \|Rx - \underbrace{Q^T b}_{\text{new RHS}}\|_2 \rightarrow \min$$

Simpler Problems: Triangular

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

How would we minimize the residual norm?

Computing QR

- Gram-Schmidt
- Householder Reflectors
- Givens Rotations

Demo: Gram-Schmidt–The Movie [cleared]
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.