What about non-square systems?

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?
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

\[ \begin{align*}
\alpha + \beta x_1^2 + \delta x_1^2 &= y_1 \\
\alpha + \beta x_2^2 + \delta x_2^2 &= y_2 \\
\alpha + \beta x_3^2 + \delta x_3^2 &= y_3 \\
\alpha + \beta x_4^2 + \delta x_4^2 &= y_4 \\
\end{align*} \]

\[ \| A_2 w - b \|_2 \leq \min \]
Rewriting Data Fitting

Rewrite in matrix form.

\[
\begin{pmatrix}
1 & x_1 & x_1^2 \\
1 & x_2 & x_2^2 \\
1 & x_n & x_n^2
\end{pmatrix}
\begin{pmatrix}
a \\
b \\
c
\end{pmatrix}
= \begin{pmatrix}
y_1 \\
y_2 \\
y_n
\end{pmatrix}
\rightarrow \text{min}
\]

Vandermonde matrices
Least Squares: The Problem In Matrix Form

\[ \|Ax - b\|_2^2 \rightarrow \text{min!} \]

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

\[ Ax \approx b \]

NOTE:
- Data Fitting is one example where LSQ problems arise.
- Many other applications lead to \( Ax \approx b \), with different matrices.
Data Fitting: Nonlinearity

Give an example of a nonlinear data fitting problem.

\[
\begin{align*}
\min & \quad \left| \exp(\alpha) + \beta x_1 + \gamma x_1^2 - y_1 \right|^2 \\
& \quad + \cdots + \\
& \quad \left| \exp(\alpha) + \beta x_n + \gamma x_n^2 - y_n \right|^2 \\
\end{align*}
\]

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

\[
\begin{align*}
\min & \quad \left| \alpha + \exp(\beta x_1 + \gamma x_1^2) - y_1 \right|^2 \\
& \quad + \cdots + \\
& \quad \left| \alpha + \exp(\beta x_n + \gamma x_n^2) - y_n \right|^2 \\
\end{align*}
\]

Demo: Interactive Polynomial Fit [cleared]
Properties of Least-Squares

Consider LSQ problem $A\mathbf{x} \approx \mathbf{b}$ and its associated objective function

$$
\varphi(x) = \|b - Ax\|_2^2.
$$

Does this always have a solution?

$$
\varphi(x) \geq \|A\mathbf{x}\|_2^2 - \|\mathbf{b}\|_2^2 \quad \text{A no null sp.}
$$

Is it always unique?

not unique if $A$ has null sp.

Examine the objective function, find its minimum.

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

(the normal eqn.)
Demo: Polynomial fitting with the normal equations [cleared]

What’s the shape of $A^T A$?

$$A_{m \times n} \rightarrow n \times n.$$  

Demo: Issues with the normal equations [cleared]
Least Squares, Viewed Geometrically

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

When $r \perp \text{span}(A)$, $\|r\|_2$ is minimized.
Phrase the Pythagoras observation as an equation.

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

\[
\Rightarrow A^T (b - Ax) = 0 \Rightarrow A^T A x = A^T b
\]

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?

\[ P^T P = P \iff \text{Im}(P) \perp \text{Ker}(P) \]

How do I make one projecting onto span\( \{ q_1, q_2, \ldots, q_\ell \} \) for orthogonal \( q_i \)?

\[ QQ^T \]
Least Squares and Orthogonal Projection

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

\[
P^2 = A(A^T A)^{-1} A^T A^T (A^T A)^{-1} A^T \\
= A(A^T A)^{-1} A^T = P
\]

Symmetric, \( \Rightarrow \) \( P \) is a projector.

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

\( A^T A \) is invertible.
Pseudoinverse

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?

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

$\infty$ if $A$ is not full rank.

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

$x = A^+ b$ solves $Ax \approx b$.
In-Class Activity: Least Squares

**In-class activity:** Least Squares
Sensitivity and Conditioning of Least Squares

What values of $\theta$ are bad?

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

$$\frac{\|Ax\|_2}{\|x\|_2} \leq \text{cond } A \cdot \frac{1}{\cos \theta} \cdot \frac{\|\Delta b\|}{\|b\|}$$

What values of $\theta$ are bad?

when $b + \cos(\text{span}(A)) \theta$
\[ \begin{align*}
x &= A^T b \\
\Delta x &= A^T \Delta b \\
\frac{\|\Delta x\|}{\|x\|} &\leq \frac{\|A^T\|}{\|A\|} \frac{\|\Delta b\|}{\|b\|} \\
&= \frac{\text{cond}(A)}{\|A\|} \cdot \frac{\|\Delta b\|}{\|b\|} \\
&= \frac{\|b\|}{\|A\| \|x\|} \cdot \frac{\|\Delta b\|}{\|b\|} \\
&\leq \frac{1}{\cos \theta} \text{cond}(A) \cdot \frac{\|\Delta b\|}{\|b\|}
\end{align*} \]
Sensitivity and Conditioning of Least Squares (II)

Any comments regarding dependencies?

Sensitivity bound depends on $b$

What about changes in the matrix?

$$\frac{\| \Delta x \|}{\| x \|} \leq (\text{cond}(A))^2 \tan \Theta + \text{cond}(A) \cdot \frac{\| \Delta A \|}{\| A \|}$$

Two behaviours:
- $\tan \Theta \approx 0$, then cond nr is approx cond($A$)
- Otherwise, cond($A$)$^2$. 
Recap: Orthogonal Matrices

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?

$$\|Qv\|_2^2 = (Qv)^T (Qv) = v^T Q^T Qv = v^T v = \|v\|_2^2.$$
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 $Ax \approx b$ to one with an upper triangular matrix?

$$
\begin{align*}
\min_x & \quad \|Ax - b\|_2 \\
= & \quad \|QRx - b\|_2 \\
= & \quad \|Q(Rx - b)\|_2 \\
& \quad \underbrace{\|Rx - Q^Tb\|_2}_{\text{new RHS}}
\end{align*}
$$
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.