The story so far...

- Quantifying errors in computation $\begin{cases} \text{conditioning of problems} \\ \text{stability of algorithms} \end{cases}$
- Floating-Point numbers (cancellation)
- The problem of solving $A\underline{x} = \underline{b}$ for $\underline{x}$, where $A \in \mathbb{R}^{n \times n}$.

(End of) Today... Linear Least-Squares Problem:

$A\underline{x} \approx \underline{b}$ $A \in \mathbb{R}^{m \times n}$, $m > n$. — an overdetermined system

we generally won't have $A\underline{x} = \underline{b}$

or $A\underline{x} - \underline{b} = \underline{0}$

but we can minimize $\| A\underline{x} - \underline{b} \|_2$

We'll propose a solution to $A\underline{x} \approx \underline{b}$ $A \in \mathbb{R}^{m \times n}$

to a special case of $A'\underline{x} = \underline{b}'$ $A' \in \mathbb{R}^{m \times n}$

Agenda: - $A\underline{x} = \underline{b}$ $A \in \mathbb{R}^{n \times n}$, symmetric positive definite.

↳ Cholesky Factorization.

- The problem $A\underline{x} \approx \underline{b}$, $A \in \mathbb{R}^{m \times n}$

↳ Normal Equation Method.

Def: An $n \times n$ matrix A is symmetric if $A^T = A$.

or $a_{ij} = a_{ji}$ for all i, j.

Def: An $n \times n$ matrix A is positive definite if $\underline{x}^T A \underline{x} > 0$.

$y = A\underline{x}$ $\underline{x}^T A \underline{x} = \underline{x}^T y = \|\underline{x}\|_2 \|y\|_2 \cos(\theta)$ for $\underline{x} > \underline{0}$

where θ is the angle between $\underline{x}$ and y

$$x^T y > 0 \quad \Rightarrow \quad \cos(\theta) > 0$$
$$\Rightarrow \quad \theta < \frac{\pi}{2} \quad \text{or} \quad 90^\circ$$

Fact: Let M be an $m \times n$ matrix with $m > n$, and $\text{rank}(M) = n$. Then $M^T M$ is symmetric and positive definite.

Pf: Homework 3, Q5(d).

Symmetric Positive Definite matrices are "nice" because if $A \in \mathbb{R}^{n\times n}$ is symmetric pos. def, we can solve $A\underline{x} = \underline{b}$ in a "better" way than G.E.

More specifically, the LU factorization of $A = LU$ can be arranged so that $U = L^T$ so $A = LL^T$

The factorization $A = LL^T$ is called the Cholesky Factorization of A.

Computation of Cholesky Factorization.

$$\underset{A}{\begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ a_{31} & a_{32} & & \ddots \\ \vdots & \vdots & & \end{bmatrix}} = \underset{L}{\begin{bmatrix} l_{11} & 0 & \cdots & \\ l_{21} & l_{22} & 0 & \cdots \\ l_{31} & l_{32} & l_{33} & 0 \cdots \\ & & & \end{bmatrix}} \underset{L^T}{\begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ \vdots & 0 & l_{33} \\ \vdots & \vdots & 0 \\ & & \vdots \end{bmatrix}}$$

$$a_{11} = l_{11}^2 \quad \Rightarrow \quad l_{11} = \sqrt{a_{11}}$$
$$a_{21} = l_{11} \cdot l_{21} \quad \Rightarrow \quad l_{21} = a_{21}/l_{11}$$
$$a_{31} = l_{11} \cdot l_{31} \quad \Rightarrow \quad l_{31} = a_{31}/l_{11}$$
$$\vdots$$

$$a_{22} = l_{21}^2 + l_{22}^2 \implies l_{22} = \sqrt{a_{22} - l_{21}^2}$$

$$a_{23} = l_{21} \cdot l_{31} + l_{22} \cdot l_{32} \implies l_{32} = \frac{a_{23} - l_{21} \cdot l_{31}}{l_{22}}.$$

$$\vdots$$

properties

1. Square roots are all positive values (because A is pos. def).
2. We don't need to pivot for numerical stability
3. Half as much work as L·U factorization because $L = U^T$ $(U = L^T)$

Linear Least Squares Example ($A\underline{x} \approx \underline{b}$)

We want to predict a student's Assignment 2 mark given their midterm grade per question.

i.e. we want to build a model (linear model).

$$x_1 q_1 + x_2 q_2 + x_3 q_3 + x_4 q_4 + x_5 q_5 = h$$

q_1: midterm Q_1 mark; q_2: Q2 mark; x_3, x_4, x_5: coefficients; h: A2 mark.

How do we find the value of $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_5 \end{bmatrix}$

We have data from students who wrote the midterm and submitted A2.

We have these constraints for each student.

$$A = \begin{bmatrix} q_1^{(1)} & q_2^{(1)} & \cdots & q_5^{(1)} \\ q_1^{(2)} & q_2^{(2)} & \cdots & q_5^{(2)} \\ & & & \\ & & & \end{bmatrix} \qquad b = \begin{bmatrix} h^{(1)} \\ h^{(2)} \\ \\ \\ \end{bmatrix} \begin{array}{l} \text{— student 1.} \\ \text{— student 2.} \\ \\ \\ \end{array}$$

we want an $\underline{x}$ such that $\underset{m\times n}{A}\, \underset{n\times 1}{\underline{x}} \approx \underset{m\times 1}{\underline{b}}$.

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

$x \in \mathbb{R}^n$

$b \in \mathbb{R}^m$.

m = number of students

n = number of features. $= 5$

Since $m > n$ we have overdetermined system

Unless $\underline{b} \in \text{span}(A)$ we won't have an $\underline{x}$ such that $A\underline{x} = \underline{b}$, or $A\underline{x} - \underline{b} = 0$

Instead, we minimize the 2-norm of the residual

$$\min_{\underline{x}} \| A\underline{x} - \underline{b} \|_2$$

The existence of $\underline{x}$ that minimize $\| A\underline{x} - \underline{b} \|_2$ is guaranteed, but $\underline{x}$ may not be unique if $\text{rank}(A) < n$.

Normal Equation (calculus)

$$\|A\underline{x}-\underline{b}\|_2 = \sqrt{(A\underline{x}-\underline{b})^T\cdot(A\underline{x}-\underline{b})}$$

minimizing $\|A\underline{x}-\underline{b}\|_2$ is the same as

minimizing $(A\underline{x}-\underline{b})^T(A\underline{x}-\underline{b})$

So, let $\phi(\underline{x}) = (A\underline{x}-b)^T(A\underline{x}-\underline{b})$

$$\underbrace{\nabla\phi(\underline{x})}_{n\times 1.} = \underline{0} = 2\underbrace{\overset{n\times m}{A^T}\overset{m\times n}{A}\overset{n\times 1}{\underline{x}}}_{n\times 1} - 2\underbrace{\overset{n\times m}{A^T}\overset{m\times 1}{\underline{b}}}_{n\times 1}$$

$$\left(2\cdot A^T(Ax-b)\right)$$

Set: $A^TA\underline{x} - A^T\underline{b} = 0$

$$\Rightarrow \boxed{\underbrace{A^TA}_{n\times n.}\underbrace{\underline{x}}_{n\times 1} = \underbrace{A^T\underline{b}}_{n\times 1}}$$ — The Normal Equation

If A is full rank, i.e. rank$(A) = n$.

then A^TA is symmetric positive definite.

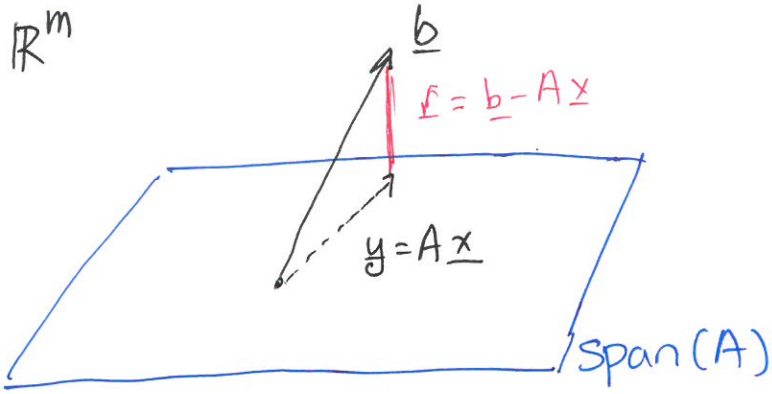
So we can solve $A^TA\underline{x} = A^T\underline{b}$ using Cholesky ~~Factorizat~~ Factorization

Geometric Derivation of Normal Equation

Def: Two vectors $\underline{u}, \underline{v} \in \mathbb{R}^m$ are orthogonal (perpendicular, normal) if $\underline{u}^T\underline{v} = 0$.

Recall $\underline{u}^T\underline{v} = \|u\|_2\cdot\|v\|_2\cos(\theta)$, (angle between $\vec{u}, \vec{v}$) so $\underline{u}^T\underline{v} = 0 \iff \cos(\theta) = 0$

$\iff \theta = \frac{\pi}{2}$ or $90°$.



~~Assume rank(A) = n.~~

we want $\vec{x}$ s.t.
$\| \underbrace{A\underline{x}}_{y} - \underline{b} \|_2$ is minimized.

- $\underline{b}$ does not lie in span(A)
- $y = A\underline{x}$ closest vector in span(A) to $\underline{b}$.

(to minimize $\|A\underline{x} - \underline{b}\|_2$)

$\Leftrightarrow$. if $\underline{r} \perp \text{span}(A)$.

i.e. $\underline{r}$ is perpendicular to every column of A.

$$\Rightarrow \quad A^T \underline{r} = \underline{0}$$

$$\Rightarrow \quad A^T(\underline{b} - A\underline{x}) = 0$$

$$\Rightarrow \quad \boxed{A^T A \underline{x} = A^T \underline{b}}$$

The normal equation.

Normal Equation Method

To solve $\min_{\underline{x}} \|A\underline{x} - \underline{b}\|_2$.

We solve $A^T A \underline{x} = A^T b$ using Cholesky Factorization we solve $LL^Tx = A^Tb$ using forward & backward substitution.

Rectangular System $\rightarrow$ Square System $\rightarrow$ Triangular System

$A\underline{x} \approx \underline{b}$ $\quad$ $(A^TA)x = A^Tb$ $\quad$ $LL^Tx = A^Tb$

In practice, the normal equation method is problematic because ~~we turn~~ the conditioning of the new problem $(A^TA)x = A^Tb$ is worse than the original problem.

Specifically, $\text{cond}(A^TA) = \underline{\underline{\text{cond}(A)^2}}$

???