Lecture 12 — Numerical Analysis — 03/01/18

Last time:

The SVD:

- existence (any matrix)
- construction

find $A^T A \underline{v}_j = \lambda_j \underline{v}_j$

set $\underline{u}_i = \frac{1}{\sqrt{\lambda_i}} A \underline{v}_i$

$$U = (\underline{u}_1 \dots \underline{u}_n)$$

$$V = (\underline{v}_1, \dots, \underline{v}_n)$$

$$S = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix}$$

Intro to Least Squares:

For $A \in \mathbb{R}^{m \times n}$, minimize $\| A\underline{x} - \underline{b} \|_2$

$$\Rightarrow \min_{\underline{x}} \left(\sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2 \right)$$

$\Rightarrow A^T A \underline{x} = A^T \underline{b}$ ← why does this have a solution?

<u>Note:</u> If $A \in \mathbb{R}^{m \times n}$, ~~Assume A has rank n~~

then $A^T A \in \mathbb{R}^{n \times n}$, and $A^T \underline{b} \in \mathbb{R}^n$.

- If A has rank n, then $A^T A$ is invertible.

  <u>Proof:</u> Use SVD.

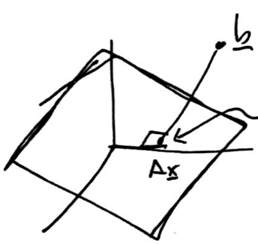
- If ~~A~~ rank $A < n$, note that $A^T \underline{b} \in \text{col}(A^T A)$

  so there exists at <u>least</u> one solution.

<u>However</u> <u>never</u> form $A^T A$ → this ~~to~~ squares the condition number.

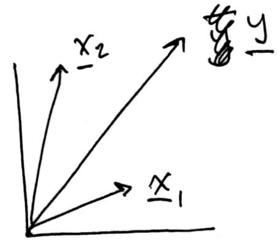
How do we ~~solve~~ minimize $\| A\underline{x} - \underline{b} \|$ without the normal equation?

<u>Idea:</u> We need to make sure that the linear system we solve is <u>consistent</u>: i.e. that it has <u>at least</u> one solution.



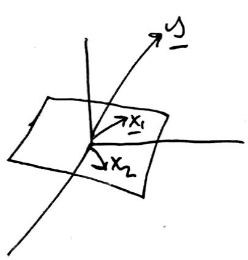
~~proj~~ $\text{proj}_A \underline{b}$ ← this is the projection of $\underline{b}$ onto the column space of A – how do we compute this?

How do we compute a projection of one vector onto another?



$$\text{proj}_{\underline{x}_1} \underline{y} = \frac{(\underline{y}, \underline{x}_1)}{\|\underline{x}_1\|} \cdot \underbrace{\frac{\underline{x}_1}{\|\underline{x}_1\|}}_{\text{unit vector}} = \left(\underline{y}, \frac{\underline{x}_1}{\|\underline{x}_1\|}\right) \cdot \frac{\underline{x}_1}{\|\underline{x}_1\|}$$

For 3D



$$\text{proj}_{X} \underline{y} = (\underline{y}, \hat{\underline{x}}_1)\, \hat{\underline{x}}_1 + \Big(\underbrace{\underline{y} - (\underline{y}, \hat{\underline{x}}_1)\hat{\underline{x}}_1}_{}, \hat{\underline{x}}_2\Big)\, \hat{\underline{x}}_2$$

the remaining part of $\underline{y}$ after projecting onto $\hat{\underline{x}}_1$.

If $(\hat{\underline{x}}_1, \hat{\underline{x}}_2) = 0$, then

$$\text{proj}_X \underline{y} = (\underline{y}, \hat{\underline{x}}_1)\, \hat{\underline{x}}_1 + (\underline{y}, \hat{\underline{x}}_2)\, \hat{\underline{x}}_2 = (\hat{\underline{x}}_1 \ \hat{\underline{x}}_2) \begin{pmatrix} \hat{\underline{x}}_1^T \\ \hat{\underline{x}}_2^T \end{pmatrix} \underline{y}$$

So this means we want to find an orthonormal basis for col A, $U = (\underline{u}_1 \ldots \underline{u}_n)$ and then solve: $A\underline{x} = \underbrace{UU^T\underline{b}}_{\text{projection onto col} A}$.

Such a U can be computed via Gram-Schmidt on the columns of A.

<u>G-S algorithm</u>: Let $q_1 = a_1 / \|a_1\|$

$$\underline{q_2'} = a_2 - (a_2, q_1) q_1 \quad \leftarrow \text{subtract proj onto } q_1$$

$$q_2 = \frac{q_2'}{\|q_2'\|} \quad \leftarrow \text{normalize.}$$

In the end, we have that

$$\underline{a}_1 = r_{11} q_1$$
$$\underline{a}_2 = r_{12} q_1 + r_{22} q_2$$
$$\underline{a}_3 = r_{13} q_1 + r_{23} q_2 + r_{33} q_3$$
$$\vdots$$

$$\Rightarrow \quad A = QR$$
$$= (q_1 \cdots q_n) \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & & \\ 0 & & \ddots & r_{nn} \end{pmatrix}$$

this is the QR factorization of A.

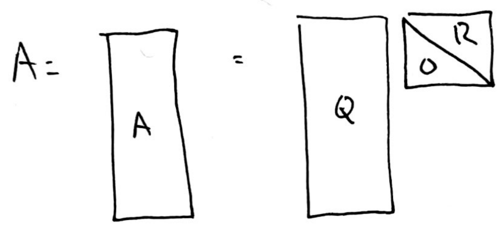
So we can solve, instead, $A\underline{x} = QQ^T \underline{b}$

<u>or</u> $\Rightarrow$

$$A\underline{x} = \underline{b}$$
$$QR\underline{x} = \underline{b}$$
$$R\underline{x} = Q^T \underline{b} \quad \leftarrow \text{the preferred method.}$$

Thm If $A \in \mathbb{R}^{m \times n}$, $m \geq n$, then A can be written as: $A = QR$, with $Q^T Q = I_n$ and R upper triangular. If $\mathrm{rank}(A) = n$, then R is invertible.

Proof: By construction using Gram-Schmidt.



Example $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 2 \end{pmatrix}$

Least squares application: Regression

Model: $y = a + bx_1 + cx_2 + \epsilon$ ← $N(0,1)$ error

Observe data: (y_i, x_{1i}, x_{2i}) (noisy, y_i contains measurement errors)

~~Find a, b, c to minimize squared~~

The residuals are:

$$r_i = y_i - \hat{b}x_{1i} - \hat{c}x_{2i} - \hat{a}$$

estimated a, b, c

Minimize $\|\underline{r}\|_2 \iff$ solve least squares for a, b, c:

$$\underbrace{\begin{pmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{pmatrix}}_{\text{"design matrix"}} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Least squares and the SVD:

If $A \in \mathbb{R}^{n\times n}$ is invertible and $A = USV^T$, then

$$A^{-1} = VS^{-1}U^T.$$

For A not invertible, and $m > n$, rank n, we have that

$$A_{m\times n} = U_{m\times n} S_{n\times n} V^T_{n\times n}$$

Define the pseudo-inverse of A to be

$$A^+ = V_{n\times n} S^{-1}_{n\times n} U^T_{n\times m} \implies A^+A = (VS^{-1}U^T)(USV^T)$$

$$= VS^{-1}\underbrace{U^TU}_{I_n}SV^T$$

$$= VV^T = I_n.$$