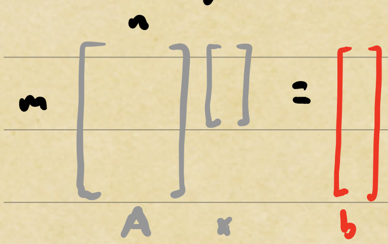
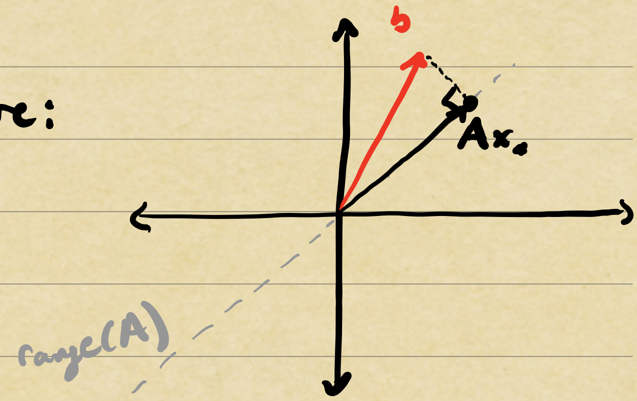
Linear Systems of Eq.'s

Part 5: Orthogonalization & triangularization

Recap:

"Least squares" in a picture:





Solve $\quad x_* = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax - b\|_2$

To find x_*, we use $A = QR$:

1) Compute $d = Q^* b$ $\quad$ | $\Rightarrow$ Project b on range(A) with $QQ^* b$

2) Solve $Rx = d$ $\quad$ | $\Rightarrow$ Solve $(\underbrace{QR}_{A})x = QQ^* b$

Last time we found $A = QR$ with classical Gram-Schmidt (CGS):

$$q_1 = \frac{a_1}{r_{11}}, \ldots, \quad q_n = \frac{1}{r_{nn}}\Big[a_n - \sum_{i=1}^{n-1} \underbrace{(q_i^* a_n)}_{r_{in}} q_i\Big]$$

normalization

"Orthonormalize columns of A"

However, CGS is numerically unstable. (See Experiment 2 in Lecture 9 of LNT).

Modified Gram-Schmidt (MGS)

MGS rearranges the order of operations to improve stability.

Initialize: $v_k = a_k$ for $k = 1, \ldots, n$

CGS

for $j = 1$ to n
    for $i = 1$ to $j-1$
        $r_{ij} = q_i^* a_j$
        $v_j = v_j - r_{ij} q_i$
    $r_{jj} = \|v_j\|$
    $q_j = v_j / r_{jj}$

MGS

for $i = 1$ to n
    $r_{ii} = \|v_i\|$
    $q_i = v_i / r_{ii}$
    for $j = i+1$ to n
        $r_{ij} = q_i^* v_j$
        $v_j = v_j - r_{ij} q_i$

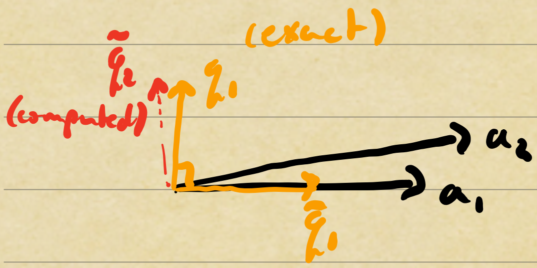
CGS: Each column a_j is orthogonalized against all previous columns at once.

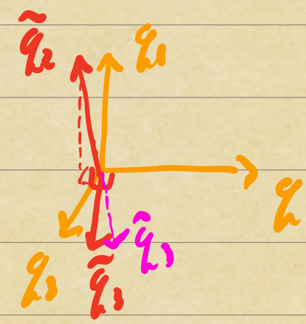
MGS: Each column $(j>i)$ is orthogonalized against q_i as soon as q_i becomes available.

<u>Loss of Orthogonality</u> (both CGS & MGS)

=> In floating point, computed $\tilde{Q}$ may <u>not</u> have columns close to orthogonal.

=> The intuition is that cancellation errors can make $q_2, q_3, \ldots$ inaccurate and then orthogonalizing against these inaccurate approximations $\tilde{q}_2, \tilde{q}_3, \ldots$ compounds the error later in the sequence.





For MGS, loss of orthogonality is controlled

$$\|Q^*Q - I\| \leq C_m \left[\frac{\sigma_1}{\sigma_m}\right] \epsilon_{mach}$$

singular vals of A

For CGS, rule of thumb is

$$\|Q^*Q - I\| \approx \tilde{C}_m \left[\frac{\sigma_1}{\sigma_m}\right]^2 \varepsilon_{mach}$$

Remark: ratio σ_1/σ_m is the condition # of A, indicating sensitivity of A's column space to perturbations. It is large when $\sigma_m \approx 0$ and columns of A are almost linearly dependent (i.e., A is "near" a singular matrix). More on this in Lecture 10 (notes).

Householder Triangularization

So far, we have not had much success finding backward stable algorithms. In some sense, we have been working (numerically) with the wrong type of transformations - both elimination matrices and GS orthogonalization matrices have been "triangular" transformations, which could amplify perturbations & lead to large cancellation errors.

A fundamental shift in the history

of NLA and in this course is the idea of working with orthogonal/unitary transformations whenever possible.
These avoid amplification and cancellation because they only rotate reflect inputs!

<u>Idea:</u> instead of orthogonalizing columns of A with triangular transformations, triangularize A with orthogonal transformations.

Householder reflections

$$\underset{A}{\begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \\ x & x & x \end{bmatrix}} \xrightarrow{Q_1} \underset{Q_1 A}{\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{bmatrix}} \xrightarrow{Q_2} \underset{Q_2 Q_1 A}{\begin{bmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & x \end{bmatrix}}$$

$$\underbrace{Q_n \cdots Q_2 Q_1}_{Q^{-1} = Q^*} A = R \quad \text{↑ upper triangular}$$

$$\Rightarrow A = QR$$

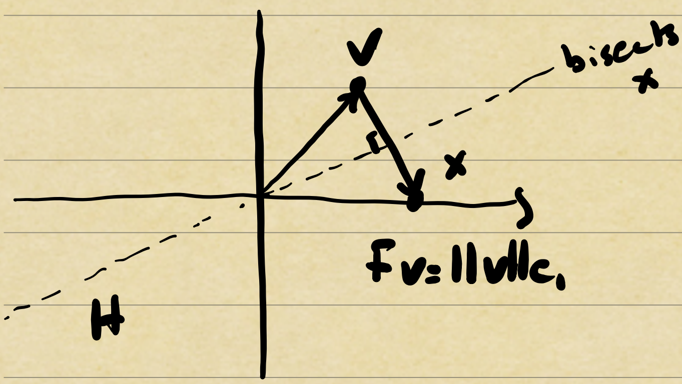
Notice, now Q is $m \times m$ and R is $m \times n$:

$$m \underset{A}{\overset{n}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}} = \underset{Q}{\overset{m}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}} \underset{R}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

We can drop zero rows of R and corresponding columns of Q to recover "thin" QR factorization from Lecture 8 (Check!).

We'll construct the matrices $Q_1, \dots, Q_m$ using <u>reflections</u> F_x.

$$v = \begin{bmatrix} x \\ x \\ x \\ \vdots \\ x \end{bmatrix} \xrightarrow{F} \begin{bmatrix} \|x\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \|x\| e_1$$



$$x = \|v\| e_1 - v$$

We need to reflect v across the

"hyperplane" H orthogonal to x.

To project v onto H, we need

$$P_H v = v - \frac{(x^* v)}{(x^* x)} x = \left[I - \frac{x x^*}{x^* x} \right] v$$

To reflect over v, we need to go twice as far

$$F v = v - 2 \frac{(x^* v)}{(x^* x)} x = \left[I - 2 \frac{x x^*}{x^* x} \right] v$$

This is called a Householder reflection.

In practice, we take $x = \text{sign}(v_1) \|v\| e_1 + v$ to avoid cancellation when v and $\|v\| e_1$ are very close.

Using reflectors we can triangularize A using orthogonal matrices

$$Q_1 = F_1, \quad Q_2 = \begin{bmatrix} 1 & \\ & F_2 \end{bmatrix}, \ldots, \quad Q_j = \begin{bmatrix} I_{j-1} & \\ & F_k \end{bmatrix}$$