Upcoming

Final exam is Wednesday, May 12, from 8:00am to 10:30am

1 QR Factorization

We left off last lecture talking about the Gram–Schmidt Process. How does GS provide a factorization of A?

We let $W = \text{Col } A$ and construct an orthonormal basis for W.

From Thm. 11 (Gram–Schmidt)

$$\text{Span}\{\vec{x_1}, \cdots, \vec{x_k}\} = \text{Span}\{\vec{u_1}, \cdots, \vec{u_k}\}$$

We can write $\vec{x_k}$ as

$$\vec{x_k} = r_{1k}\vec{u_1} + \cdots + r_{kk}\vec{u_k} + 0\vec{u_{k+1}} + \cdots + 0\vec{u_n}$$

Then $\vec{x_k}$ is a linear combination of the columns Q, where $Q = \begin{bmatrix}\vec{u_1} & \cdots & \vec{u_n}\end{bmatrix}$ and weights

$$\vec{r_k} = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

That is, $\vec{x_k} = Q\vec{r_k}$ for $k = 1, \cdots, n$

Theorem 1 (QR Factorization) *If $A \in \mathbb{R}^{m\times n}$ (note: not square) with linearly independent columns, then A can be factored as $A = QR$, where $Q \in \mathbb{R}^{m\times n}$ whose columns form an orthonormal basis for Col A and $R \in \mathbb{R}^{n\times n}$ is invertible and upper trianglular with positive entries on its diagonal.*

Example

Find a QR factorization for $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

From last lecture, we have an orthogonal basis $\{\vec{u_1}, \vec{u_2}, \vec{u_3}\}$

$$\vec{u_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{u_2} = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \vec{u_3} = \begin{bmatrix} 0 \\ -2/3 \\ 1/2 \\ 1/2 \end{bmatrix}$$

We need an orthonormal basis though, so we'll need to normalize these basis vectors (Lecture 29).

$$Q = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}$$

Now that we've found Q, how do we find R? We know that for an orthogonal matrix (Q is), we can say

$$Q^T Q = I$$

Then from the equation $A = QR$, we get

$$Q^T A = Q^T (QR) = IR = R$$

$$R = Q^T A$$

We know both Q^T and A, so we can find R.

$$R = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{bmatrix}$$

This doesn't seem useful now but it will be later.

2 Least Squares Solutions

A lot of systems we come across in real life will be inconsistent. We still want to find approximate solutions to inconsistent systems.

Example

$$\begin{aligned} c_1 x &= b_1 \\ c_2 x &= b_2 \\ c_3 x &= b_3 \end{aligned}$$

How can we measure some error in this system? We want our error to be 0.

$$E^2 = (c_1 x - b_1)^2 + (c_2 x - b_2)^2 + (c_3 x - b_3)^2$$

If we found an exact solution, our error would be 0. If the error is not, then we have a numerical quantifier of how "off" our solution is. We can take the derivative of this equation

$$\frac{dE}{dx} = 2\left[c_1(c_1 x - b_1) + c_2(c_2 x - b_2) + c_3(c_3 x - b_3)\right] = 0$$

If we take that and solve for x, we get

$$x = \frac{c_1 b_1 + c_2 b_2 + c_3 b_3}{c_1^2 + c_2^2 + c_3^2}$$

If you look closely, you can see that this is the same as

$$\vec{a} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

$$x = \frac{\vec{a} \cdot \vec{b}}{\vec{a} \cdot \vec{a}}$$

which is the component in our projection formula. If x is an exact solution, then $||A\vec{x} - \vec{b}|| = 0$.

If we're approximating, we're looking for some $\hat{x}$ such that

$$||\vec{b} - A\hat{x}|| \leq ||\vec{b} - A\vec{x}||$$

If $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$, a **least–squares** solution of $A\vec{x} = \vec{b}$ is an $\hat{x} \in \mathbb{R}^n$ such that

$$||\vec{b} - A\hat{x}|| \leq ||\vec{b} - A\vec{x}||$$

for all $\vec{x} \in \mathbb{R}^n$.

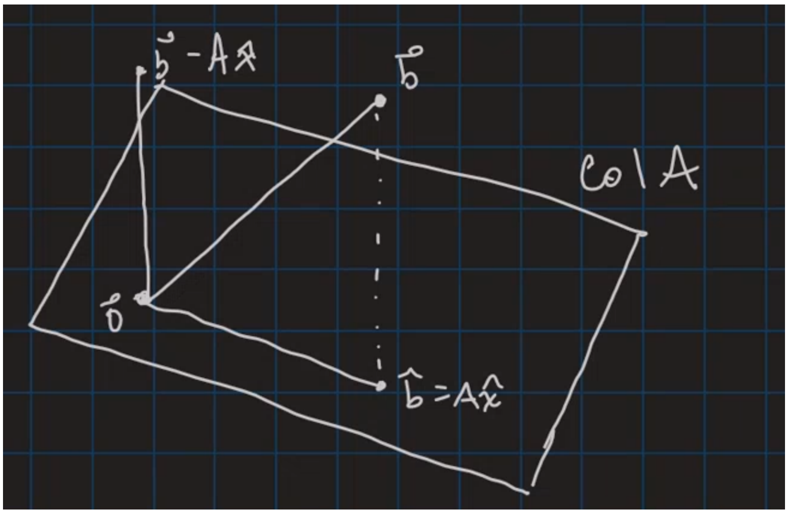
We need to tie in QR factorization and the Gram–Schmidt process into all of this.

Note that $\vec{x} \in$ Col A, because we can write $A\vec{x} = \vec{b}$. If $\vec{b} \in A\vec{x}$, then we'd have a solution. This means that $\vec{b}$ is not in Col A, and we need an approximate solution.

If we project the value $\vec{b}$ onto Col A, we can find the shortest distance between $\vec{b}$ and Col A. In other words, we're finding the closest point possible in Col A to $\vec{b}$ in order to approximate it best.

$$\hat{b} = proj_{\text{Col } A}\vec{b}$$

This means that $\hat{b} \in$ Col A and $A\vec{x} = \hat{b}$ is consistent. Then we solve $A\vec{x} = \hat{b}$.



A few things to note: $\vec{b} - A\hat{x}$ (or $\vec{b} - \hat{b}$) is orthogonal to Col A. We can choose some column of A called $\vec{a_j}$ and say that

$$\vec{a_j} \cdot (\vec{b} - A\hat{x}) = 0$$

Since each $\vec{a_j}$ is orthogonal to $\vec{b} - A\hat{x}$, we can write

$$A^T(\vec{b} - A\hat{x}) = 0 \qquad\qquad (\vec{u}^T\vec{v} = \vec{u}\cdot\vec{v})$$

We can use the left distributive law to write

$$A^T\vec{b} - A^T A\hat{x}$$

This means that each least-squares solution of $A\vec{x} = \vec{b}$ satisfies this linear system

$$A^T A\vec{x} = A^T\vec{b}$$

This linear system is named the **normal equations** for $A\vec{x} = \vec{b}$

Theorem 2 *The set of least-squares solutions of $A\vec{x} = \vec{b}$ coincides with the nonempty set of the solutions of the normal equations*

$$A^T A\vec{x} = A^T\vec{b}$$

Example

Find a least-squares solution of the inconsistent system $A\vec{x} = \vec{b}$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 2 & 1 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

First, compute the normal equations

$$A^T A = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T\vec{b} = \begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$A^T A\vec{x} = A^T\vec{b} \implies \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$(A^T A)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

Then,

$$\hat{x} = (A^T A)^{-1} A^T\vec{b}$$
$$= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$