Orthogonal Bases

L13

Last time: we found the best approximate solution(s) of $Ax=b$ (in the sense of least squares) by solving $A^TA\hat{x}=A^Tb$.

Now we turn to computational considerations. The goal is the QR decomposition. This plays the role of an LU decomposition for least squares (among other things):

LU makes solving $Ax=b$ fast	QR makes least-□s solving $Ax=b$ fast

("fast" means: only substitution / $O(n^2)$ flops)

Idea: projections are much easier to compute in the presence of a basis of orthogonal vectors.

Recall: two vectors v, w are orthogonal if $v \cdot w = 0$.

Here's what it means for more vectors to be orthogonal.

Def: A set of nonzero vectors $\{u_1, u_2, \ldots, u_n\}$ is:

(1) orthogonal if $u_i \cdot u_j = 0$ for $i \neq j$ (pairwise orthogonal)

(2) orthonormal if they're orthogonal and $u_i \cdot u_i = 1$ for all i (unit vectors)

So orthonormal means $u_i \cdot u_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$

Matrix version: let $Q = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{pmatrix} \rightsquigarrow Q^TQ = \begin{pmatrix} u_1\cdot u_1 & u_1\cdot u_2 & \cdots \\ u_2\cdot u_1 & u_2\cdot u_2 & \\ \vdots & & \ddots \end{pmatrix}$

(1) $\{u_1, u_2, \ldots, u_n\}$ is orthogonal

$\iff Q^TQ$ is diagonal & invertible

(the diagonal entries $u_i \cdot u_i$ are all nonzero)

(2) $\{u_1, u_2, \ldots, u_n\}$ is orthonormal $\iff Q^TQ = I_n$

Q: Wait! Doesn't $Q^TQ = I_n$ mean $Q^T = Q^{-1}$?

$\longrightarrow$ Only if Q is square.

Eg: $u_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$ $u_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}$ $u_3 = \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$ (nonzero vectors)

(1) $u_1 \cdot u_2 = 0$ $u_1 \cdot u_3 = 0$ $u_2 \cdot u_3 = 0$

$\Rightarrow \{u_1, u_2, u_3\}$ is orthogonal

(2) $u_1 \cdot u_1 = 4$ $u_2 \cdot u_2 = 4$ $u_3 \cdot u_3 = 4$

$\Rightarrow \{u_1, u_2, u_3\}$ is not orthonormal

Matrix version:

$$Q = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \rightsquigarrow Q^TQ = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$ diagonal but $\neq I_3$

NB: If $\{u_1, u_2, \ldots, u_n\}$ is orthogonal then you can make it orthonormal by dividing by the lengths:

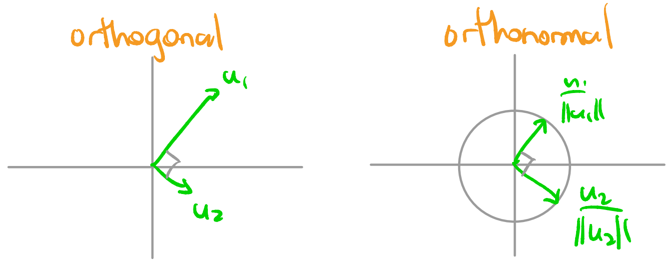
$$\left\{ \frac{u_1}{\|u_1\|}, \frac{u_2}{\|u_2\|}, \ldots, \frac{u_n}{\|u_n\|} \right\}$$ is orthonormal

Eg: $\left\|\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\right\| = 2 \qquad \left\|\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\right\| = 2 \qquad \left\|\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\right\| = 2$

$\rightsquigarrow \left\{ \frac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \frac{1}{2}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}, \frac{1}{2}\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix} \right\}$ is orthonormal

$$Q = \frac{1}{2}\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & -1\\ 1 & -1 & 1\\ 1 & -1 & -1\end{pmatrix} \rightsquigarrow Q^TQ = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}$$

Picture in $\mathbb{R}^2$:



Fact: If $\{u_1, u_2, \ldots, u_n\}$ is orthogonal then it is LI.

$\rightsquigarrow$ so it's a basis for $\mathrm{Span}\{u_1, u_2, \ldots, u_n\}$.

We will use the following trick several times.

Proof: Suppose $x_1u_1 + x_2u_2 + \cdots + x_nu_n = 0$. We need to show $x_1 = x_2 = \cdots = x_n = 0$ (trivial solution).

Trick: take the dot product of both sides with u_1:

$$u_1 \cdot (x_1u_1 + x_2u_2 + \cdots + x_nu_n)$$
$$= x_1(u_1\cdot u_1) + x_2\overbrace{\cancel{(u_1\cdot u_2)}}^{0} + \cdots + x_n\overbrace{\cancel{(u_1\cdot u_n)}}^{0} = u_1 \cdot 0$$

$$\Longrightarrow x_1\overbrace{(u_1\cdot u_1)}^{\text{nonzero}} = 0 \Rightarrow x_1 = 0.$$

Now take $u_2 \cdot (-)$, etc. //

Geometric Facts About Matrices with Orthonormal Columns:

Suppose that Q has orthonormal columns, so $Q^TQ = I_n$.

(1) $(Qx)\cdot(Qy) = x\cdot y$ for all $x, y \in \mathbb{R}^n$

(2) $\|Qx\| = \|x\|$ for all $x \in \mathbb{R}^n$

This says that multiplication by Q does not change lengths or angles:

$$\begin{pmatrix}\text{angle} \\ \text{from } x \\ \text{to } y\end{pmatrix} = \cos^{-1}\left(\frac{x\cdot y}{\|x\|\,\|y\|}\right) = \cos^{-1}\left(\frac{(Qx)\cdot(Qy)}{\|Qx\|\,\|Qy\|}\right) = \begin{pmatrix}\text{angle} \\ \text{from } Qx \\ \text{to } Qy\end{pmatrix}$$

Proof: (1) $(Qx)\cdot(Qy) \overset{v\cdot w = v^Tw}{=} (Qx)^T(Qy) = x^TQ^TQy$

$= x^TI_ny = x^Ty = x\cdot y$ ✓

(2) $\|Qx\| = \sqrt{(Qx)\cdot(Qx)} \overset{(1)}{=} \sqrt{x\cdot x} = \|x\|$ ✓

Def: A square matrix with orthonormal columns is called orthogonal.

$\boxed{\text{Beware the strange terminology!!}}$

Eg: $\frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$ is orthogonal

Eg: $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}$ is not orthogonal

The Projection Formula

Thm: Let $\{u_1, u_2, \ldots, u_n\}$ be orthogonal and let $V = \text{Span}\{u_1, u_2, \ldots, u_n\}$. For any vector b,

Projection Formula:

$$\boxed{b_V = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{b \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{b \cdot u_n}{u_n \cdot u_n} u_n}$$

NB: You're only taking dot products – no elimination needed!

> If you have an orthogonal basis for V, the projection formula is way faster than solving $A^TA\hat{x} = A^Tb$!

NB: If $n=1$ this says $b_V = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1$, which recovers the formula for projection onto a line.

NB: If $\{u_1, u_2, \ldots, u_n\}$ is orthonormal then $u_i \cdot u_i = 1$, so the projection formula becomes

$$b_V = (b \cdot u_1) u_1 + (b \cdot u_2) u_2 + \cdots + (b \cdot u_n) u_n.$$

NB: The projection formula only works if your basis is orthogonal! Otherwise you just don't get b_V.

Proof: Let $b' = \dfrac{b\cdot u_1}{u_1\cdot u_1}u_1 + \dfrac{b\cdot u_2}{u_2\cdot u_2}u_2 + \cdots + \dfrac{b\cdot u_n}{u_n\cdot u_n}u_n$. We want to show that $b' = b_V$. Recall that b_V is characterized by $b - b_V \in V^\perp$. So we need to show $b - b' \in V^\perp$ (then $b' = b_V$). Since V is the span of $u_1, u_2, \ldots, u_n$, we need to prove $u_i \cdot (b - b') = 0$ for each i. This uses the trick from before:

$$i=1:\quad u_1\cdot(b-b') = u_1\cdot b - u_1\cdot b'$$

$$= u_1\cdot b - u_1\cdot\left(\frac{b\cdot u_1}{u_1\cdot u_1}u_1 + \frac{b\cdot u_2}{u_2\cdot u_2}u_2 + \cdots + \frac{b\cdot u_n}{u_n\cdot u_n}u_n\right)$$

$$= u_1\cdot b - \frac{b\cdot u_1}{\cancel{u_1\cdot u_1}}\cancel{u_1\cdot u_1} - \frac{b\cdot u_2}{u_2\cdot u_2}\underbrace{\cancel{u_1\cdot u_2}}_{0} - \cdots - \frac{b\cdot u_n}{u_n\cdot u_n}\underbrace{\cancel{u_1\cdot u_n}}_{0}$$

$$= u_1\cdot b - b\cdot u_1 = 0 \quad \checkmark$$

The same works for $u_2, u_3, \ldots, u_n$. //

Eg: Compute b_V for $V = \mathrm{Span}\left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\right\}$ $\quad b = \begin{pmatrix}1\\2\\3\\2\end{pmatrix}$

The spanning vectors are orthogonal, so we use the projection formula:

$$b_V = \frac{\begin{pmatrix}1\\2\\3\\2\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}}{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \frac{\begin{pmatrix}1\\2\\3\\2\end{pmatrix}\cdot\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}}{\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} + \frac{\begin{pmatrix}1\\2\\3\\2\end{pmatrix}\cdot\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}}{\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}}\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}$$

$$= \frac{8}{4}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} + \frac{-2}{4}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix} + \frac{0}{4}\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix} = \begin{pmatrix}3/2\\3/2\\5/2\\5/2\end{pmatrix}$$

easy ✓

Here's a version of the projection formula that computes the projection matrix:

Thm: Let $\{u_1, u_2, \ldots, u_n\}$ be orthogonal and let $V = \mathrm{Span}\{u_1, u_2, \ldots, u_n\}$. The projection matrix onto V is

Outer Product Formula

$$\boxed{P_V = \frac{u_1 u_1^T}{u_1 \cdot u_1} + \frac{u_2 u_2^T}{u_2 \cdot u_2} + \cdots + \frac{u_n u_n^T}{u_n \cdot u_n}}$$

NB: If $n=1$ this says $P_V = \frac{u_1 u_1^T}{u_1 \cdot u_1}$, which recovers the formula for the projection matrix onto a line.

NB: If $\{u_1, u_2, \ldots, u_n\}$ is orthonormal then $u_i \cdot u_i = 1$, so the outer product formula becomes

$$P_V = u_1 u_1^T + u_2 u_2^T + \cdots + u_n u_n^T.$$

Fast-forward: this is the SVD of P_V.

Proof: The projection matrix is defined by $P_V b = b_V$ for all vectors b. So let's check:

$$\left(\frac{u_1 u_1^T}{u_1 \cdot u_1} + \frac{u_2 u_2^T}{u_2 \cdot u_2} + \cdots + \frac{u_n u_n^T}{u_n \cdot u_n}\right) b$$

$$= \frac{u_1 (u_1^T b)}{u_1 \cdot u_1} + \frac{u_2 (u_2^T b)}{u_2 \cdot u_2} + \cdots + \frac{u_n (u_n^T b)}{u_n \cdot u_n}$$

$$\overset{(u_i^T b = u_i \cdot b)}{=} \frac{u_1 \cdot b}{u_1 \cdot u_1} u_1 + \frac{u_2 \cdot b}{u_2 \cdot u_2} u_2 + \cdots + \frac{u_n \cdot b}{u_n \cdot u_n} u_n \xrightarrow[\text{formula}]{\text{projection}} b_V \quad //$$

Eg: Compute P_V for $V = \mathrm{Span}\left\{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\right\}$.

$$P_V = \frac{1}{\begin{pmatrix}1\\1\\1\\1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\1\\1\end{pmatrix}}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}(1\ 1\ 1\ 1) + \frac{1}{\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}}\begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}(1\ 1\ -1\ -1) + \frac{1}{\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\cdot\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}}\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}(1\ -1\ 1\ -1)$$

$$= \frac{1}{4}\begin{pmatrix}1&1&1&1\\1&1&1&1\\1&1&1&1\\1&1&1&1\end{pmatrix} + \frac{1}{4}\begin{pmatrix}1&1&-1&-1\\1&1&-1&-1\\-1&-1&1&1\\-1&-1&1&1\end{pmatrix} + \frac{1}{4}\begin{pmatrix}1&-1&1&-1\\-1&1&-1&1\\1&-1&1&-1\\-1&1&-1&1\end{pmatrix}$$

$$= \frac{1}{4}\begin{pmatrix}3&1&1&-1\\1&3&-1&1\\1&-1&3&1\\-1&1&1&3\end{pmatrix}$$

The outer product formula has this matrix version:

Thm: Suppose that Q has orthonormal columns. Set $V = \mathrm{Col}(Q)$. (In other words, the columns of Q are an orthonormal basis for V.) Then

$$\boxed{P_V = QQ^T}$$

This comes from the outer product form of matrix mult:

$$QQ^T = \begin{pmatrix}| & & |\\ u_1 & \cdots & u_n\\ | & & |\end{pmatrix}\begin{pmatrix}— u_1^T —\\ \vdots\\ — u_n^T —\end{pmatrix} = u_1u_1^T + \cdots + u_nu_n^T = P_V.$$

NB: Orthonormal columns means $Q^TQ = I_n$.

Eg: If Q is square then $Q^TQ = I_n \Longrightarrow Q^T = Q^{-1}$

$$\Longrightarrow QQ^T = I_n.$$

This makes sense: if Q is invertible then $V = \mathrm{Col}(Q) = \mathbb{R}^n$, so $P_V = P_{\mathbb{R}^n} = I_n$. ✓

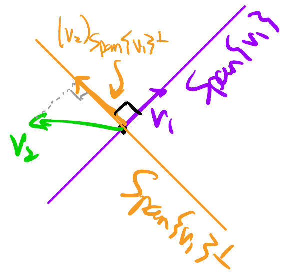
The Gram-Schmidt Procedure

We like orthogonal bases because they make orthogonal projections easier. How do we produce one?

Idea: Start with some basis $\{v_1, v_2, \dots, v_n\}$ for V.

- Force v_2 to be $\perp v_1$ by replacing it with

  $$(v_2)_{\text{Span}\{v_1\}^\perp} = \text{projection onto } \text{Span}\{v_1\}^\perp$$



- Force v_3 to be $\perp v_1, v_2$ by replacing it with $(v_3)_{\text{Span}\{v_1, v_2\}^\perp}$.
  - → Since $\{v_1, v_2\}$ is now orthogonal, you can compute this easily with the projection formula!
- etc.

So this "straightens out" the basis vectors, one at a time.

Gram-Schmidt Procedure: Let $\{v_1, v_2, \dots, v_n\}$ be LI.

(1) $u_1 = v_1$

(2) $u_2 = v_2 - \boxed{\dfrac{u_1 \cdot v_2}{u_1 \cdot u_1} u_1}$ $\leftarrow (v_2)_{\text{Span}\{u_1\}}$

$u_2 = (v_2)_{\text{Span}\{u_1\}^\perp}$

(3) $u_3 = v_3 - \boxed{\dfrac{u_1 \cdot v_3}{u_1 \cdot u_1} u_1 - \dfrac{u_2 \cdot v_3}{u_2 \cdot v_2} u_2}$ $\leftarrow (v_3)_{\text{Span}\{u_1, u_2\}}$

$u_3 = (v_3)_{\text{Span}\{u_1, u_2\}^\perp}$

$\vdots$

(n) $u_n = v_n - \dfrac{u_1 \cdot v_n}{u_1 \cdot u_1} u_1 - \dfrac{u_2 \cdot v_n}{u_2 \cdot u_2} u_2 - \cdots - \dfrac{u_{n-1} \cdot v_n}{u_{n-1} \cdot u_{n-1}} u_n$

Result: $\{u_1, u_2, \dots, u_n\}$ is orthogonal, and for $i = 1, 2, \dots, n$ we have

$$\text{Span}\{u_1, u_2, \dots, u_i\} = \text{Span}\{v_1, v_2, \dots, v_i\}$$

In particular, if $\{v_1, v_2, \dots, v_n\}$ is a basis for a subspace V, then $\{u_1, u_2, \dots, u_n\}$ is an orthogonal basis for V: it's a way to describe V as the Span of an orthogonal set of vectors.

basis for V $\xrightarrow[\text{SCHMIDT}]{\text{GRAM}}$ orthogonal basis for V

Eg: $v_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} \quad v_3 = \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix}$

(1) $u_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$

(2) $$u_2 = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} - \frac{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}}{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} - \frac{2}{2}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

(3) $$u_3 = \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix} - \frac{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix}}{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \frac{\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

$$= \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix} - \frac{6}{2}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \frac{-6}{6}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

check: $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = 0 \quad \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0 \quad \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0$ ✓

Q: What happens if you start with vectors that are LD?

Eventually you'll have

$$v_{i+1} \in \mathrm{Span}\{v_1, v_2, \ldots, v_i\} = \mathrm{Span}\{u_1, u_2, \ldots, u_i\}$$

$$\Rightarrow u_{i+1} = (v_{i+1})_{\mathrm{Span}\{v_1, v_2, \ldots, v_i\}^\perp} = 0$$

Gram-Schmidt detected that $v_{i+1} \in \mathrm{Span}\{v_1, v_2, \ldots, v_i\}$.
So you can discard v_{i+1} & keep going!

QR Decomposition

This "keeps track" of the dot products in the Gram-Schmidt procedure in the same way that LU "keeps track" of the row operations you performed.

Procedure: Run Gram-Schmidt on $\{v_1, v_2, \dots, v_n\}$:

$$\{v_1, v_2, \dots, v_n\} \xrightarrow{G-S} \{u_1, u_2, \dots, u_n\}$$

Solve for the v's in terms of the u's:

$$v_1 = u_1$$

$$v_2 = \frac{u_1 \cdot v_2}{u_1 \cdot u_1} u_1 + u_2$$

$$v_3 = \frac{u_1 \cdot v_3}{u_1 \cdot u_1} u_1 + \frac{u_2 \cdot v_3}{u_2 \cdot u_2} u_2 + u_3$$

$$v_4 = \frac{u_1 \cdot v_4}{u_1 \cdot u_1} u_1 + \frac{u_2 \cdot v_4}{u_2 \cdot u_2} u_2 + \frac{u_3 \cdot v_4}{u_3 \cdot u_3} u_3 + u_4$$

Express these 4 equations as equalities of the columns of two matrices:

$$\begin{pmatrix} | & | & | & | \\ v_1 & v_2 & v_3 & v_4 \\ | & | & | & | \end{pmatrix} = \begin{pmatrix} | & | & | & | \\ u_1 & u_2 & u_3 & u_4 \\ | & | & | & | \end{pmatrix} \begin{pmatrix} 1 & \frac{u_1 \cdot v_2}{u_1 \cdot u_1} & \frac{u_1 \cdot v_3}{u_1 \cdot u_1} & \frac{u_1 \cdot v_4}{u_1 \cdot u_1} \\ 0 & 1 & \frac{u_2 \cdot v_3}{u_2 \cdot u_2} & \frac{u_2 \cdot v_4}{u_2 \cdot u_2} \\ 0 & 0 & 1 & \frac{u_3 \cdot v_4}{u_3 \cdot u_3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

This is almost the QR decomposition, but we have to scale the u's to be unit vectors.

We can divide the columns of the first matrix by their lengths, but then we have to multiply the rows of the second matrix by the same thing so we don't change the product:

cancels!

$$\underbrace{\begin{pmatrix} | & | & | & | \\ v_1 & v_2 & v_3 & v_4 \\ | & | & | & | \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} | & | & | & | \\ \frac{u_1}{\|u_1\|} & \frac{u_2}{\|u_2\|} & \frac{u_3}{\|u_3\|} & \frac{u_4}{\|u_4\|} \\ | & | & | & | \end{pmatrix}}_{Q} \underbrace{\begin{pmatrix} \|u_1\| & \frac{u_1\cdot v_2}{u_1\cdot u_1}\|u_1\| & \frac{u_1\cdot v_3}{u_1\cdot u_1}\|u_1\| & \frac{u_1\cdot v_4}{u_1\cdot u_1}\|u_1\| \\ 0 & \|u_2\| & \frac{u_2\cdot v_3}{u_2\cdot u_2}\|u_2\| & \frac{u_2\cdot v_4}{u_2\cdot u_2}\|u_2\| \\ 0 & 0 & \|u_3\| & \frac{u_3\cdot v_4}{u_3\cdot u_3}\|u_3\| \\ 0 & 0 & 0 & \|u_4\| \end{pmatrix}}_{R}$$

This is the QR decomposition.

QR Decomposition:

Let A be an $m\times n$ matrix with FCR (LI columns). Then

$$A = QR$$

where

Q is an $m\times n$ matrix with orthonormal columns

R is an upper-triangular $n\times n$ matrix with positive diagonal entries.

The Procedure is explained above. It says:

Q: The columns form an orthonormal basis for Col(A).
They are the vectors you get by applying G-S to the columns of A & dividing by lengths.

R: This is filled with the dot products & lengths you computed when running G-S & rescaling.

Analogy to LU Decompositions: in $A=LU$,

U: a REF of A

L: row operations to get to REF

NB: Since Q has orthonormal columns $\Rightarrow Q^TQ=I_n$.

So $\quad A=QR \iff Q^TA=Q^TQR=I_nR=R$

$$\boxed{R=Q^TA}$$

But you'd never compute R this way. You never have to "compute" R — finding R is just bookkeeping + Gram-Schmidt.

Eg: $A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{pmatrix} \rightsquigarrow v_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad v_2 = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} \quad v_3 = \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix}$

We did Gram-Schmidt before:

$$u_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad \|u_1\| = \sqrt{2}$$

$$u_2 = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} - \frac{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}}{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix} - \frac{2}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \qquad \|u_2\| = \sqrt{6}$$

$$u_3 = \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix} - \frac{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix}}{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \frac{\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix}}{\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

$$= \begin{pmatrix} 3 \\ -3 \\ 3 \end{pmatrix} - \frac{6}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} - \frac{-6}{6} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \|u_3\| = \sqrt{3}$$

$$Q = \begin{pmatrix} | & | & | \\ \frac{u_1}{\|u_1\|} & \frac{u_2}{\|u_2\|} & \frac{u_3}{\|u_3\|} \\ | & | & | \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \end{pmatrix} \qquad R = \begin{pmatrix} \sqrt{2} & 1\sqrt{2} & 3\sqrt{2} \\ 0 & \sqrt{6} & -1\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{pmatrix}$$

QR decompositions have many important applications. Here is how to use it to speed up least-squares.

Suppose that A has FCR. Finding the least-squares solution of $Ax=b$ means solving $A^TA\hat{x} = A^Tb$. If we have a QR decomposition then we substitute $A=QR$:

$$A^TA\hat{x} = (QR)^T(QR)\hat{x} = R^TQ^TQR\hat{x} = R^TI_nR\hat{x} = R^TR\hat{x}$$

$$A^Tb = (QR)^Tb = R^TQ^Tb$$

Now R is invertible (it's in REF & it has n pivots), so R^T is invertible too.

So we can multiply both sides of $R^TR\hat{x} = R^TQ^Tb$ by $(R^T)^{-1}$:

$$A^TA\hat{x} = A^Tb \iff R^TR\hat{x} = R^TQ^Tb$$

$$\iff (R^T)^{-1}R^TR\hat{x} = (R^T)^{-1}R^TQ^Tb$$

$$\iff R\hat{x} = Q^Tb.$$

But R is in REF, so you can solve this by substitution!

How to Solve $Ax=b$ by Least Squares Using $A=QR$:

Solve $\boxed{R\hat{x} = Q^Tb}$ by substitution.

Computational Complexity:

Computing $A=QR$ takes $\approx \frac{4}{3}n^3$ flops if A is $n\times n$ (using a much more clever algorithm).

Then you need $\approx 2n^2$ flops to find the least-squares solution of $Ax=b$ for any number of values of b (multiply Q^Tb then substitute $R\hat{x} = Q^Tb$).

So it's the same speed-up as an LU decomposition.

Wait! Why not just compute a $PA=LU$ decomposition for A^TA instead?

$\to$ It turns out QR is usually more accurate (less rounding error).

Eg: Find the least-squares solution of $Ax=b$ for

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

using $A=QR$, $\quad Q=\begin{pmatrix}1/\sqrt{2} & 1/\sqrt{6}\\ -1/\sqrt{2} & 1/\sqrt{6}\\ 0 & -2/\sqrt{6}\end{pmatrix} \quad R=\begin{pmatrix}\sqrt{2} & \sqrt{2}\\ 0 & \sqrt{6}\end{pmatrix}$.

We need to solve $R\hat{x}=Q^Tb$.

$$Q^Tb=\begin{pmatrix}1/\sqrt{2} & -1/\sqrt{2} & 0\\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6}\end{pmatrix}\begin{pmatrix}1\\ 1\\ -1\end{pmatrix}=\begin{pmatrix}0\\ 4/\sqrt{6}\end{pmatrix}$$

$$(R\mid Q^Tb)=\left(\begin{array}{cc|c}\sqrt{2} & \sqrt{2} & 0\\ 0 & \sqrt{6} & 4/\sqrt{6}\end{array}\right)\xrightarrow{R_2\div=\sqrt{6}}\left(\begin{array}{cc|c}\sqrt{2} & \sqrt{2} & 0\\ 0 & 1 & 2/3\end{array}\right)$$

$$\xrightarrow{R_1-=\sqrt{2}R_2}\left(\begin{array}{cc|c}\sqrt{2} & 0 & -2\sqrt{2}/3\\ 0 & 1 & 2/3\end{array}\right)\xrightarrow{R_1\div=\sqrt{2}}\left(\begin{array}{cc|c}1 & 0 & -2/3\\ 0 & 1 & 2/3\end{array}\right)$$

$$\Longrightarrow \hat{x}=\frac{2}{3}\begin{pmatrix}-1\\ 1\end{pmatrix}$$