Input $A = [b_1 | b_2 | \cdots | b_p] \in \mathbb{R}^{m \times p}$

columns are independent

$W = \text{span}\{b_1, b_2, \ldots, b_p\} = \text{col } A$

$\|$

Output $Q = [q_1 | q_2 | \cdots | q_p] \in \mathbb{R}^{m \times p}$

$W = \text{span}\{q_1, \cdots q_p\} = \text{col } Q$

$\text{col } A = \text{col } Q$

Gram-Schmidt algorithm:

$z_1 = b_1$ $\qquad b_1 = z_1 = \|z_1\| q_1$ $\qquad q_1 = \dfrac{z_1}{\|z_1\|}$

$z_2 = b_2 - (q_1 \cdot b_2) q_1 = b_1 - q_1 q_1^T b_1$ $\qquad q_2 = \dfrac{z_2}{\|z_2\|}$

$b_2 = (q_1 \cdot b_2) q_1 + z_2 = (q_1 \cdot b_2) q_1 + \|z_2\| q_2$

$z_3 = b_3 - (q_1 \cdot b_3) q_1 - (q_2 \cdot b_3) q_2$ $\qquad q_3 = \dfrac{z_3}{\|z_3\|}$

$= b_3 - [q_1 | q_1] \begin{bmatrix} q_1^T \\ q_2^T \end{bmatrix} b_3$

projection

Note since the columns of A came from a basis the columns are independent therefore none of the denominators here are 0

$\vdots$

$b_3 = (q_1 \cdot b_3) q_1 + (q_2 \cdot b_3) q_2 + \|z_3\| q_3$

$z_p = b_p - (q_1 \cdot b_p) q_1 - \cdots - (q_{p-1} \cdot b_p) q_{p-1}$ $\qquad q_p = \dfrac{z_p}{\|z_p\|}$

$b_p = (q_1 \cdot b_p) q_1 + \cdots + (q_{p-1} \cdot b_p) q_{p-1} + \|z_p\| q_p$

Solving back for b's in terms of q's gives...

$b_1 = z_1 = \|z_1\| q_1$

$b_2 = (q_1 \cdot b_2) q_1 + z_2 = (q_1 \cdot b_2) q_1 + \|z_2\| q_2$

$b_3 = (q_1 \cdot b_3) q_1 + (q_2 \cdot b_3) q_2 + \|z_3\| q_3$

$\vdots$

$b_p = (q_1 \cdot b_p) q_1 + \cdots + (q_{p-1} \cdot b_p) q_{p-1} + \|z_p\| q_p$

$$\underbrace{\left[\, b_1 \,\middle|\, b_2 \,\middle|\, b_3 \,\middle|\, \cdots \,\middle|\, b_p \,\right]}_{\in \mathbb{R}^{m\times p}} = \underbrace{\left[\, q_1 \,\middle|\, q_2 \,\middle|\, q_3 \,\middle|\, \cdots \,\middle|\, q_p \,\right]}_{\in \mathbb{R}^{m\times p}} \overbrace{\begin{bmatrix} \|z_1\| & q_1\cdot b_2 & q_1\cdot b_3 & \cdots & q_1\cdot b_p \\ 0 & \|z_2\| & q_2\cdot b_3 & & q_2\cdot b_p \\ 0 & 0 & \|z_3\| & \ddots & \vdots \\ \vdots & \vdots & 0 & \ddots & q_{p-1}\cdot b_p \\ 0 & 0 & 0 & \cdots & \|z_p\| \end{bmatrix}}^{\in \mathbb{R}^{p\times p}}$$

R upper triangular matrix

Since $\|z_i\|$ are all non-zero then R is invertible.

Factorization $A = QR$

<u>Idea:</u> Solve $Ax = b$ ...

How

$$QRx = b$$
$$\cancel{Q^TQ}Rx = Q^Tb$$
$$Rx = Q^Tb$$

upper triangular so solve for x using substitutions...

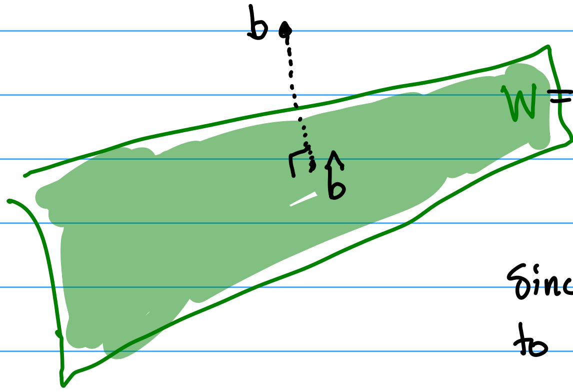
Know Q is a matrix with orthonormal columns... so
$$Q^TQ = I$$

Note: Even if $Ax=b$ didn't have a solution (but the columns of A are independent) then R is invertible and so $Rx = Q^Tb$ always has a solution:

Interpret what that solution means when there isn't one...

Recall $W = \operatorname{col} A = \operatorname{col} Q$

If $Ax=b$ doesn't have a solution, then $b \notin \operatorname{col} A$



$W = \operatorname{col} A = \operatorname{col} Q$

$$\hat{b} = \operatorname{proj}_W b = QQ^Tb$$

Since $\hat{b} \in W$ then it's possible to solve $Ax = \hat{b}$

We know $A = QR$ and $\hat{b} = QQ^Tb$. Then

$$QRx = QQ^Tb$$

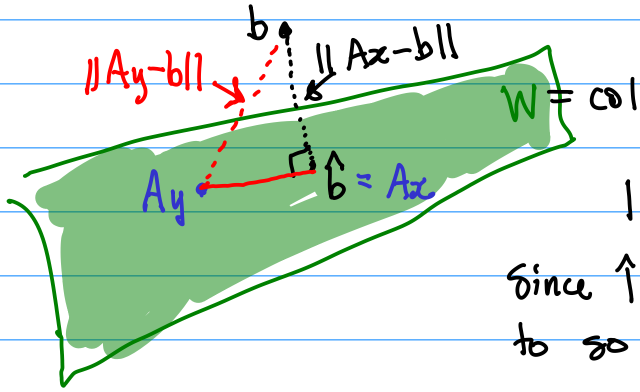
$$Q^TQRx = Q^TQQ^Tb$$

$$Rx = Q^Tb$$

So actually we we solving not $Ax = b$ but instead $Ax = \hat{b}$

projection of b onto col A.

Interpretation



Since the hypotenuse of a right triangle is bigger than the leg... It follows that

$$\|Ax-b\| \le \|Ay-b\|$$

for all $y \in \mathbb{R}^p$.

Since $\uparrow$ to so

So we have that the solution to $Rx = Q^Tb$ has the geometric interpretation of minimizing $\|Ax-b\|^2$.

$\boxed{\text{This is called the least squares solution to } Ax = b}$

Algebraic interpretation:

Interpret $Rx = Q^Tb$ in term of the original matrix A.

$$Ax = \hat{b}$$

$$Ax = QQ^Tb$$

$$A^TAx = A^TQQ^Tb$$

$$A^TAx = R^TQ^TQQ^Tb = R^TQ^Tb = A^Tb$$

$A = QR$

$A^T = R^TQ^T$

Therefore $A^TAx = A^Tb$ ← Normal equations...

If x is the minimizer for $\|Ax-b\|$ then $A^TAx = A^Tb$.

To show the other way around, it is enough to check whether A^TA is invertible...

$$A^TA = (R^TQ^T)(QR) = R^T\cancel{Q^TQ}R = R^TR$$

since R is invertible, so is R^T and consequently R^TR is also invertible.

Therefore A^TA is invertible...

Note: Finding the minimizer of $\|Ax-b\|$ using the normal equations $A^TAx = A^Tb$ is not practical for very big problems and can cause rounding errors on a computer solution..

If A has 2 column then $A^TA \in \mathbb{R}^{2\times 2}$ and maybe the normal equations are okay..

Applications to statistics for making a regression...

Applications to computers, since $Ax=b$ never has an exact solution on a computer due to rounding errors... and so minimizing $\|Ax-b\|$ is better...

The Gram–Schmidt Process

Given a basis $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p\cdot\mathbf{v}_{p-1}}{\mathbf{v}_{p-1}\cdot\mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \quad \text{for } 1 \le k \le p \qquad (1)$$

note the Gram-Schmidt in the book looks different because they don't introduce the unit vectors... it turns out to be the same — the only reason it's done this way is to avoid the square roots.. We used the unit vectors because it is easier to figure out what R is in that case...