3. Linear Least Squares

- Example
- QR factorization

Positive semi-definite matrix.

Definition:
A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semi-definite if $x^T A x \geq 0$ for all $x \in \mathbb{R}^n$.

Notation: $A \succcurlyeq 0$

Similarly, A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite if $x^T A x > 0$ for all non-zero $x \in \mathbb{R}^n$ $(A \succ 0)$

If x is a eigenvector of A, $\quad Ax = \lambda x$

$$\Rightarrow x^T A x = \lambda \underbrace{x^T x}_{} \geq 0$$

$$\Rightarrow \lambda \geq 0$$

Q. Which of the following statements are true.

i. If A is PSD, A is invertible.

ii. If A is PD, A^{-1} exists and A^{-1} is also PD.

iii. Let $f(x) = \|Ax - b\|_2^2$. For all x, the Hessian of f has non-negative eigenvalues.

pick one:

A) i

B) i, ii

C) ii, iii

Least squares

$$\underset{x}{\text{minimize}} \quad f(x) := \frac{1}{2}\|Ax - b\|_2^2 = \frac{1}{2}\sum_{i=1}^{m}(a_i^T x - b_i)^2$$

$$f(x) = \frac{1}{2}\langle Ax-b, Ax-b\rangle = \frac{1}{2}\langle Ax, Ax\rangle - \langle Ax, b\rangle + \langle b, b\rangle$$

$$= \frac{1}{2}\langle A^T A x, x\rangle - \langle A^T b, x\rangle + \|b\|^2$$

- Gradient

$$\nabla f(x) = A^T(Ax - b)$$

- Hessian

$$\nabla^2 f(x) = A^T A \qquad x^T A^T A x = \|Ax\|_2^2 \geq 0$$

- Is the Hessian positive definite? positive semidefinite?

column

Ans: Always positive semidefinite. Positive definite if A has full [redacted] rank.

- Conditions for $x = x^*$ to be a global minimizer?

$$A^T(Ax^* - b) = 0$$

These are the **normal equations** of the least squares problem.

QR factorization

Orthogonal and orthonormal vectors

Two vectors $x \in \mathbf{R}^n$, $y \in \mathbf{R}^n$

- Recall cosine identity

$$x^T y = \|x\|_2 \|y\|_2 \cos(\theta)$$

- x and y are orthogonal if they are "perpendicular"

$$x^T y = 0 \quad (\cos(\theta) = 0)$$

- x and y are orthonormal if they are orthogonal and normalized

$$x^T y = 0, \quad x^T x = 1, \quad y^T y = 1$$

Orthogonal matrix

$Q \in \mathbf{R}^{n \times n}$ is orthogonal if its columns are all pairwise orthonormal

$$Q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}, \qquad Q^T Q = Q Q^T = I$$

- Then the inverse is the transpose

$$Q^{-1} = Q^T$$

- Inner products are "invariant" under orthogonal transformations

$$(Qx)^T (Qy) = x^T Q^T Q y = x^T y$$

- 2-norm is also invariant

$$\|Qx\|_2 = \|x\|_2$$

- Determinant is either 1 or -1. (Why?)

Orthogonal matrix

$Q \in \mathbf{R}^{n \times n}$ is orthogonal if its columns are all pairwise orthonormal

$$Q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}, \qquad Q^T Q = QQ^T = I$$

- Then the inverse is the transpose

$$Q^{-1} = Q^T$$

- Inner products are "invariant" under orthogonal transformations

$$(Qx)^T (Qy) = x^T Q^T Q y = x^T y$$

- 2-norm is also invariant

$$\|Qx\|_2 = \|x\|_2$$

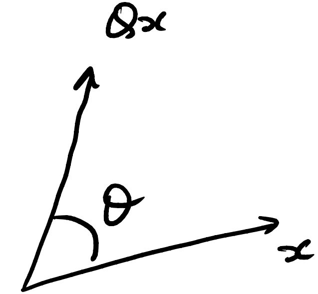
- Determinant is either 1 or -1. (Why?)

$$\det(I) = \det(Q^T Q) = \det(Q)^2 = 1$$

Example

The rotation matrix

$$Q = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$



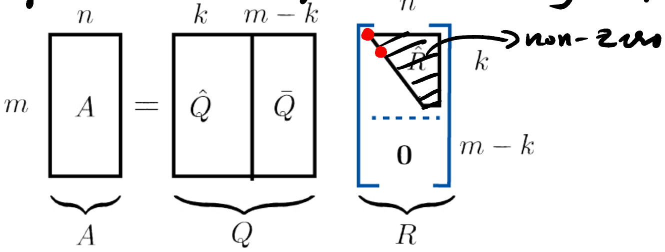
is orthogonal.

Check: $$Q^T Q = \begin{pmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \cos^2\theta + \sin^2\theta \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

QR Factorization

Assume the first k columns of A are linearly indep.



where

$\mathcal{R}(A) \oplus \mathcal{N}(A^T) = \mathbb{R}^m$

- Q is orthogonal ($Q^TQ = QQ^T = I$)
- R is upper triangular ($R_{ij} = 0$ whenever $i > j$)
- $\hat{Q}$ spans the range of A
- $\bar{Q}$ spans the nullspace of A^T,

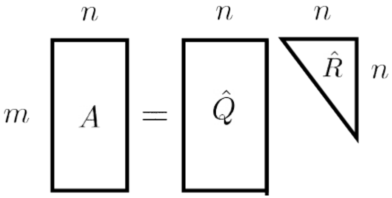
In Julia,

```
(Q,R) = qr(A)
```

Always exists for any $A \in \mathbf{R}^{m\times n}$

Reduced ("thin", "economode") QR factorization

For $A \in \mathbf{R}^{m\times n}$ full rank



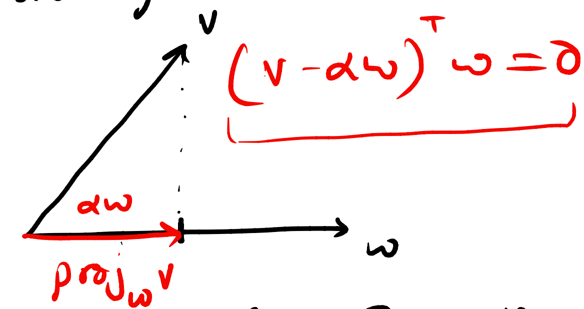
Equivalently,

$$
\begin{aligned}
a_1 &= r_{11}q_1 \\
a_2 &= r_{12}q_1 + r_{22}q_2 \\
a_3 &= r_{13}q_1 + r_{23}q_2 + r_{33}q_3 \\
&\vdots \\
a_n &= r_{1n}q_1 + r_{2n}q_2 + \cdots + r_{nn}q_n
\end{aligned}
$$

Using Projection to find Q.

Given vectors $v, w \in \mathbb{R}^n$, the projection of v onto w is

$$\text{proj}_w v = \frac{w^T v}{w^T w} w$$



Given a matrix $A = [a_1 \ a_2 \cdots a_n]$

we can find Q with $R(Q) = R(A)$ and Q orthonormal by:

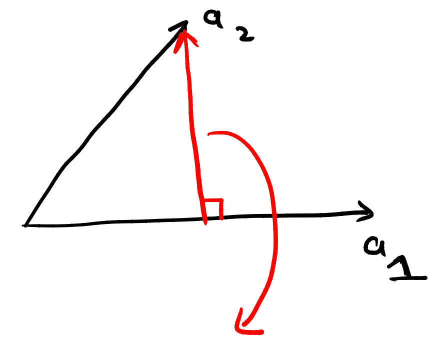
$$\tilde{q}_1 = a_1$$

$$\tilde{q}_2 = a_2 - \text{proj}_{\tilde{q}_1} a_2$$

$$\vdots$$

$$\tilde{q}_n = a_n - \text{proj}_{\tilde{q}_1} a_n - \cdots \text{proj}_{\tilde{q}_{n-1}} a_n$$

$$q_K = \tilde{q}_K / \|\tilde{q}_K\|_2$$



$$\tilde{q}_2 = a_2 - \text{proj}_{a_1} a_2$$

Solving least squares via QR

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Ax - b\|_2^2$$

$x = \begin{bmatrix} a \\ b \end{bmatrix}$

Because 2-norm is invariant under orthogonal transformation,

$\|x\|_2^2 = \|a\|_2^2 + \|b\|_2^2$

$$\begin{aligned}
\|Ax - b\|_2^2 &= (Ax-b)^T(Ax-b) \\
&= (Ax-b)^T Q^T Q (Ax-b) \\
&= \|Q^T(Ax-b)\|_2^2 \\
&= \left\| \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix} x - \begin{bmatrix} \hat{Q}^T \\ \bar{Q}^T \end{bmatrix} b \right\|_2^2 \\
&= \underbrace{\|\hat{R}x - \hat{Q}^T b\|_2^2}_{(1)} + \underbrace{\|\bar{Q}^T b\|_2^2}_{(2)}
\end{aligned}$$

$(Q(Ax-b))^T(-)$

$Q^T = \begin{bmatrix} \hat{Q}^T \\ \bar{Q}^T \end{bmatrix}$

(1) is minimized when $\hat{R}x = \hat{Q}^T b$, (2) is constant

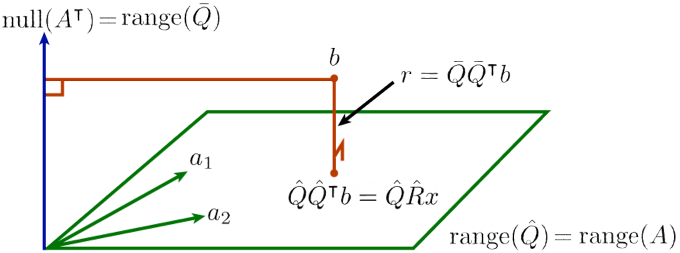
Geometric prespective

$P = \begin{bmatrix} 0 & 0 \\ \alpha & 1 \end{bmatrix}$

- Orthogonal projection matrix:
  1. A square matrix P is a projection matrix if $P^2 = P$.
  2. A projection matrix P is an orthogonal projection matrix if $P = P^T$.

$P^2 = P$

- Let $A = QR$.
  1. $\hat{Q}\hat{Q}^T$ and $\bar{Q}\bar{Q}^T$ are orthogonal projection matrices.
  2. For any $b \in \mathbf{R}^m$, $\hat{Q}\hat{Q}^T b$ is the closet point in **range**(A) w.r.t. 2-norm.
  3. For any $b \in \mathbf{R}^m$, $\bar{Q}\bar{Q}^T b$ is the closet point in **null**(A^T) w.r.t. 2-norm.



Solving least squares via QR

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Ax - b\|_2^2$$

Mathematically

$$\begin{aligned} A^T A x &= A^T b \\ R^T Q^T Q R x &= R^T Q^T b \\ R x &= Q^T b \\ x &= R^{-1} Q^T b \end{aligned}$$

Comptationally

1. Compute $A = \hat{Q}\hat{R}$
2. Compute $y = \hat{Q}^T b$
3. Solve $\hat{R}x = y$

More computationally stable than solving $A^T A x = A^T b$ by forming $A^T A$