Lecture 4 Orthogonality and QR Decomp

Note by Samion Suwito on 1/30/25

Matrices

Matrices $\equiv$ Linear transformations
Claim: If $\mathcal{S}$ is a subspace, then $\pi_{\mathcal{S}}$ (projection on to s) is a linear operator.
Proof: Claim that $\pi_{\mathcal{S}}(\alpha x + \beta y) = \alpha\pi_{\mathcal{S}}(x) + \beta\pi_{\mathcal{S}}(y)$
Check:

$$\begin{gathered}\langle \pi_{\mathcal{S}}(\alpha x + \beta y) - (\alpha x + \beta y), s\rangle \\ = \langle \alpha\pi_{\mathcal{S}}(x) + \beta\pi_{\mathcal{S}}(y) - (\alpha x + \beta y), s\rangle \\ = \alpha\langle \pi_{\mathcal{S}}(x) - x, s\rangle + \beta\langle \pi_{\mathcal{S}}(y) - y, s\rangle \\ = 0 \ \ \forall s \in \mathcal{S} \\ \implies \alpha\pi_{\mathcal{S}}(x) + \beta\pi_{\mathcal{S}}(y) = \pi_{\mathcal{S}}(\alpha x + \beta y) \text{ by HPT}\end{gathered}$$

Hpt is hilbert's projection theorem
Show the error vector (first arg of inner product) is perpendicular to the space

Connect to Least Square: A has full column rank

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_2^2 = \min_{y\in R(A)} \|y - b\|_2^2 = \|\pi_{R(A)}(b) - b\|_2^2$$

Claim: $\pi_{R(A)}(b) = A(A^\intercal A)^{-1}A^\intercal b$
Check: $\left(A(A^\intercal A)^{-1}A^\intercal b - b\right)^\intercal y = 0 \, \forall y \in R(A)$

$$\begin{gathered}\iff \left(A(A^\intercal A)^{-1}A^\intercal b - b\right)^\intercal Ax = 0 \, \forall x \in \mathbb{R}^n \\ b^\intercal A(A^\intercal A)^{-1}(A^\intercal A)x - b^\intercal Ax = 0\end{gathered}$$

As transpose reverses the order of multiplication.

Matrix Norms

For linear functions, operator norm was natural as a norm on vectors.
Can *extend* this idea to matrices.

Example: Induced p-Norm

$$(p \geq 1) : \|A\|_p := \max_{x:\|x\|_p \leq 1} \|Ax\|_p$$

is the operator norm when we view:$A : (\mathbb{R}^n, \|\cdot\|_p) \to (\mathbb{R}^m, \|\cdot\|_p)$
$\|\cdot\|_p$ is submultiplicative : $\|AB\|_p \leq \|A\|_p \|B\|_p$

$$\begin{aligned} \max_{x:\|x\|_p \leq 1} \|ABx\|_p &\leq \max_{x:\|x\|_p \leq 1} \|A\|_p \|Bx\|_p \\ &= \|A\|_p \|B\|_p \end{aligned}$$

Example:

$$\|A\|_1 = \max_{j=1...n} \sum_{i=1}^{m} |a_{ij}|$$

= max l^i norm over columns
Spectral norm of A

$$\|A\|_2 = \sqrt{\lambda_{\max}\left(A^\intercal A\right)}$$

$$\|A\|_\infty = \max_{i=1...m} \sum \|a_{ij}\|$$

max l^1 norm over rows

Frobianias Norm

$$\|A\|_F = \sqrt{\operatorname{Tr}\left(A^\intercal A\right)} = \sqrt{\sum_{i,j} (a_{ij})^2}$$

$= l^2$ norm of matrix A viewed as a $n \times m$ dim vector.

Orthogonalisation

Given vectors $\{x^{(1)} \dots x^{(k)}\}$ produce vectors $\{v^{(1)} \dots v^{(k)}\}$ s.t.

- $v^{(i)}$'s are orthogonal
- $\operatorname{Sp}\{v^{(1)} \dots v^{(m)}\} = \operatorname{Sp}\{x^{(1)} \dots x^{(k)}\} \forall 1 \leq m \leq k$
  Procedure: Gram-Schmidt:

$$v^{(m)} = x^{(m)} - \pi_{\mathrm{Sp}\{v^{(1)}\dots v^{(m-1)}\}}(x^m)$$

where $m \geq 1$
Proof: Induct on m. $m = 1 =$ base case $v^{(1)} = x^{(1)}$.
Suppose true for $m - 1$,

$$v^{(m)} = x^{(m)} - \pi_{\mathrm{Sp}\{v^{(1)}\dots v^{(m-1)}\}}(x^{(m)}) \perp \mathrm{Sp}(v^{(1)}\dots v^{(m-1)})$$

Due to the orthogonality principle of projection according to Hilbert project theorem. Difference between a vector and a projection to the subspace is orthogonal to the subspace.

$$x^{(m)} = v^{(m)} + \pi_{\mathrm{Sp}(v^{(1)}\dots v^{(m-1)})} =$$

LC of $\{v^{(1)}\dots v^{(m)} \in \mathrm{Sp}(v^{(1)}\dots v^{(m)})\}$
Can give 0 but its ok because 0 is orthogonal to every other vector

Sometimes in GS, $v^{(i)} = 0$. This is no problem:
Example:

$$\{x^{(1)}\dots x^{(3)}\} = \{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\-1\end{pmatrix}\}$$

Therefore is LD but it's ok because GS will match the span by having $v^{(3)} = 0$ as the span doesn't expand with $x^{(3)}$.

How to find an **orthonormal basis** of these $x^{(i)}$. We can just throw out all the 0 vectors and normalise all the rest.

Application: QR Decomposition

$A \in \mathbb{R}^{m\times n}$ matrix full column-rank $\mathrm{rank}(A) = n \iff \mathcal{N}(0) = \{0\}$

$$A = QR,$$

$Q \in \mathbb{R}^{m\times n}$ $\mathrm{w/\ ON\ columns}$
$R \in \mathbb{R}^{n\times n}$ matrix upper triangular and full-rank therefore invertible
$R_{ij} = 0$ for $i > j$
Note: Some call this "thin" or "reduced" QR as some people insist that Q should be an OG matrix that's square and ON columns which can be done by putting tons of 0s.

Return to Least Square

$$\min_x \|Ax - b\|_2^2$$

x^* earlier we saw

$$\begin{aligned} x^* &= (A^\intercal A)^{-1} A^\intercal b \\ &= (R^\intercal Q^\intercal Q R)^{-1} R^T Q^t b && Q^\intercal Q = I \\ &= (R^\intercal R)^{-1} R^\intercal Q^\intercal b \\ &= R^{-1} R^{-\intercal} R^\intercal Q^\intercal b = R^{-1} Q^\intercal b \end{aligned}$$

$$R x^* = Q^\intercal b$$

This is useful because R is upper-triangular which means we can solve this by back substitution. complexity of inverse is $dim^{2.\text{something}}$ is $\max(n^2, n \times n)$.
Useful for when the matrix doesn't change but b changes meaning that the cost of QR factorisation is more efficient than reminimising
For example, when having the same physical system but then gaining new data from the system therefore having less cost in the long run.

How to compute QR Decomposition of A (full col. rank)

Let $q^{(1)} \ldots q^{(n)}$ be *orthonormal* vectors obtained by running GS on cols $\{a^{(1)} \ldots a^{(n)}\}$ of A (+ normalise).
Let $\tilde{q}^{(k)}$ be the normalised $q^{(k)}$ (obtained from gps before normalising).
$\tilde{q}^{(k)} = a^{(k)} - \pi_{\text{Sp}(q^{(1)} \ldots q^{(k-1)})}(a^{(k)})$

$$= a^{(k)} - \sum_{=1}^{k-1} \langle a^{(i)}, q^{(i)} \rangle q^{(i)}$$

$$\implies a^{(k)} = \sum_{i=1}^{k} \langle a^{(i)}, q^{(i)} \rangle q^{(i)} + \|\tilde{q}^{(k)}\|_2 q^{(k)}$$

$$= \left(q^{(i)} \dots q^{(k)} \dots q^{(n)}\right) \begin{pmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \end{pmatrix}$$

On the first line the inner product is r_{ik} and $\|\tilde{q}^k\|_2$ is r_{kk}
The last vector is $r^{(k)}$. The array before is Q

Question: What if A does not have full column rank?
First suppose $A = [A_1, A_2]$ where A_1 has full column rank $R(A_2) \leq R(A_1)$
(I.e. $\operatorname{rank}(A_1) = rank(A) = r$) where $A \in \mathbb{R}^{m \times r}$

Can write $A_1 = QR_1$
$A_2 = QR_2$ for some other matrix R^2

$$A = Q[R_1, R_2]$$
$$= \begin{bmatrix} x & x & x & x & \\ & & & & \\ 0 & \ddots & & & \end{bmatrix}$$

If A does not have that form, start by permuting the column of A.
$AP = QR$ where P is a permutation matrix.