Lecture 5 Symmetic Matrices and their Eigenvalues

Note by Samion Suwito on 2/4/25

Last Time

Any $A \in \mathbb{R}^{m \times n}$ w/ full column rank can be written as

$$A = QR,\ Q \in \mathbb{R}^{m \times n},\ Q^\top Q = I, R \in \mathbb{R}^{n \times n}$$

Where Q has orthonormal columns and R is invertible upper triangular

If A *does not* have full column rank,

$$A = Q \begin{bmatrix} * & \dots & & & & & \\ 0 & 0 & * & \dots & & & \\ 0 & 0 & 0 & 0 & 0 & * & \dots \end{bmatrix}$$

$Q \in \mathbb{R}^{m \times r}$ where $r = \text{rank}(A)$.
(Rearrange the columns of A s.t. the first columns are LI. and do QR you will get the short and fat triangular matrix)
Gram Schmidt is sequential and greedy.

How to interpret Q?

Claim: $QQ^\top$ is projection onto $R(A)$. ($= \text{Sp}\{q^{(1)} \dots q^{(r)}\}$)
Check: need to show that

$$\begin{aligned} &\langle x - QQ^\top x, Ay \rangle = 0\ \forall x, y \\ \iff &\langle x - QQ^\top x, Qz \rangle = 0\ \forall x, z \\ &= x^\top Qz - x^\top QQ^\top Qz = 0 \\ &= x^\top Qz - x^\top QIz \end{aligned}$$

Example:

$$\min_x \|Ax - b\|_2^2 = \min_{a \in R(A)} \|a - b\|_2^2 = \|\pi_{R(A)}(b) - b\|_2^2$$
$$= \|QQ^\top b - b\|_2^2 = \|\,(I - QQ^\top)b\|_2^2$$

Symmetric matrices

Def: $A \in \mathbb{R}^{n\times n}$ w/ $A = A^\intercal$ $[A]_{ij} = [A]_{ji}$ (Sometimes written as $\mathbb{S}^n$)
Most important class of matrix for optimisation (In his opinion).

Examples

Example 1: $f : \mathbb{R}^n \to \mathbb{R}$

$$\text{Hess}(f) = \nabla^2 f = \mathbb{R}^{n\times n}$$
$$[\nabla^2 f(x)]_{ij} := \partial_{x_i}\partial_{x_j} f(x) = \partial_{x_j}\partial_{x_i} = [\nabla^2 f(x)]_{ji}$$
$$\Rightarrow \text{Hessian is always symmetric}$$

Example 2: Given set of data points $x^{(1)} \dots x^{(m)} \in \mathbb{R}^n$ Often useful to consider their sample correlation/covariant matrix:

$$C := \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \bar{x})(x^{(i)} - \bar{x})^\intercal$$

where $\bar{x} = \frac{1}{m}\sum x^{(i)}$ and is symmetric

Example 3: X a given matrix, $X^\intercal X$ (sometimes called kernel matrix) is symmetric

Example 4: Quadratic functions $q : \mathbb{R}^n \to \mathbb{R}$ is quadratic if it is polynomial of degree ≤ 2.
Claim: Every quadratic can be written as

$$q(x) = \frac{1}{2}x^\intercal hx + c^\intercal x + d$$

Where $H \in \mathbb{S}^n,\ c \in \mathbb{R}^n,\ d \in \mathbb{R}$
Why?

$$q(x) = \sum_{i\leq j} q_{ij}x_i x_j + \sum c_i x_i + d$$
$$\frac{1}{2}x^\intercal \begin{pmatrix} 2q_{11} & q_{21}\dots & q_{m1} \\ q_{12} & \ddots & \vdots \\ q_{1n} & \dots & 2q_{mn} \end{pmatrix} x$$

q_{ij} is the coefficients of the polynomial. Which is the hessian of q

Eigenvectors of symmetric matrices are orthogonal

Claim: Symmetric matrices have real eigenvalues.

Spectral Theorem for symmetric matrices

Let $A \in \mathbb{S}^n$ and $\lambda_1, \dots, \lambda_n$ be its eigenvalues. There exists orthogonal matrix*

$U = [u^{(1)} \cdots u^{(n)}]$ $UU^\top = U^\top U = I$

s.t. A=$U \Lambda U^\top$ where $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$

$$= \sum_{i=1} \lambda_i u^{(i)} u^{(i)\top}$$

$$= \sum_{i=1} \lambda_i \pi_{\mathrm{Sp}\{u^{(i)}\}}$$

Variational Characterisation of Eigenvalues

$x^\top A x = \sum \lambda_i \|\pi_{\mathrm{Sp}\{u^{(i)}\}}(x)\|_2^2 \leq \lambda_{\max}(A) \|x\|_2^2$

$$\implies \lambda_{\max}(A) = \max_{x : \|x\|_2 = 1} \frac{(x^\top A x)}{\|x\|_2^2}$$

Rayleigh Quotient

One approach to spectral decomposition.

Solve

$$\lambda^* = \max_{x \neq 0} \frac{x^\top A x}{\|x\|_2^2}$$

then Write $A' = A - \lambda^* \frac{x^* x^{*\top}}{\|x^*\|}{}^2_2$

Quiz: How to find

$$\lambda_{\min}(A) = \min_{x \neq 0} \frac{x^\top A x}{\|x\|_2^2}$$

Example in terms of matrix norms

Recall

$$\begin{aligned}
\|A\|_2^2 := \max_{x:\|x\|_2\leq 1} \|Ax\|_2^2 &= \max_{x\neq 0} \frac{\|Ax\|_2^2}{\|x\|_2^2} \\
&= \max_{x\neq 0} \frac{x^\intercal\left(A^\intercal A\right)x}{\|x\|_2^2} \\
&= \lambda_{\max}\left(A^\intercal A\right)
\end{aligned}$$

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

Which is also called the "**Spectral norm**"

PSD

Def: A symmetric matrix is

Positive Semidefinite if $x^\intercal Ax \geq 0 \ \ \forall x$, (psd, $\mathbb{S}_+^n$, $A \geq 0$)

Positive Definite if $x^\intercal Ax > 0 \ \ \forall x$, (pd, $A > 0$)

Remarks: For any matrix $A \in \mathbb{R}^{m\times n}$ both $A^\intercal A$ and $AA^\intercal$ are symmetric

1. $A^\intercal A \geq 0$ and $AA^\intercal \geq 0$
   Why: $x^\intercal A^\intercal Ax = \|Ax\|_2^2 \geq 0$
2. $A^\intercal A > 0 \iff N(A) = \{0\} \iff \operatorname{rank}(A) = n$
   (full column rank)
3. $AA^\intercal > 0 \iff N\left(A^\intercal\right) = \{0\} \iff R(A) = \mathbb{R}^n$
   (full row rank)
4. In context of quadratic functions, (Hessian) $H > 0 \implies q$ looks like an upright "bowl" (typical quadratic picture).
   If $H \geq 0 \implies q$ looks like an upward trough, looks curved in one way and flat in another.

Decompositions of PSD Matrices:

If $A > 0$, then there $\exists$ a unique pd matrix B s.t. $B^2 = A$

In this case, we call B the square root of A, and denoted by $A^{\frac{1}{2}}$.

If $A \geq 0$, $\exists$ unique psd $B \geq 0$, s.t. $B^2 = A$ and we call $B = A^{\frac{1}{2}}$

(Both pd and psd are symmetric!)

Proof (w/o uniqueness): Decompose $A = U\Lambda U^\top$ (is spectral decomp).
Λ diagonal w non-neg matrix so $\Lambda^{\frac{1}{2}} = \operatorname{diag}\left(\lambda_1^{\frac{1}{2}}, \ldots, \lambda_n^{\frac{1}{2}}\right)$

$$A = U\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}U^\top = (U\Lambda^{\frac{1}{2}}U^\top)(U\Lambda^{\frac{1}{2}}U^\top)$$

Therefore $A^{\frac{1}{2}} = \left(U\Lambda^{\frac{1}{2}}U^\top\right)$

another example of a decomposition
Cholesky Decomp of $A > 0$ where $A = LL^\top$ and L is lower triangular matrix
Proof: $A^{\frac{1}{2}} = QR \Rightarrow A^{\frac{1}{2}}A^{\frac{1}{2}} = R^\top Q^\top QR = R^\top R$

PD matrices and inner products

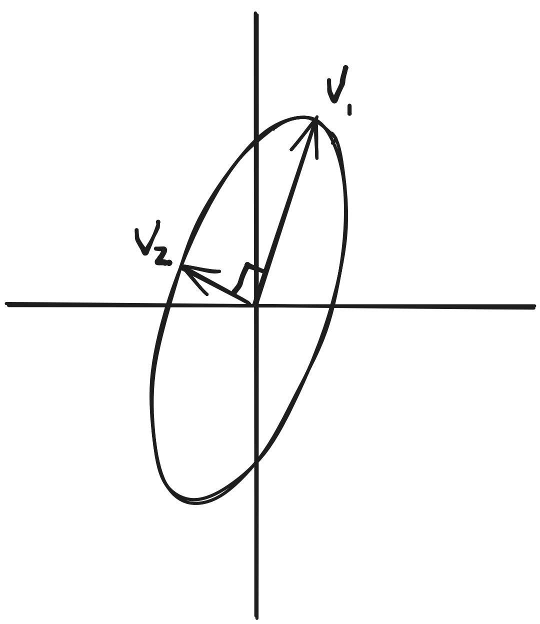
Claim: Every inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ can be written as $\langle x, y \rangle = x^\top Ay$ for some suitable $A > 0$.

$$\langle x, y \rangle = \left\langle \sum_i x_i e_i, \sum y_j e_j \right\rangle = \sum_{i,j} x_i y_j \langle e_i, e_j \rangle =: a_{ij} = a_{ji}$$

$$\implies x^\top Ay \;\; [A]_{ij} = a_{ij} \;\; A \in \mathbb{S}^n$$

$x^\top Ax = \langle x, x \rangle = \|x\|^2 \geq 0$ w/ = $\iff x = 0 \implies$ A > 0

Norm ball of the Hilbert's norm: $\{x : \|x\| \leq 1\}$
$= \{x : x^\top Ax \leq 1\} =$ an elipsoid



Where the directions are eigenvectors and lengths are related to eigenvalues ($\frac{1}{\sqrt{\lambda_i(A)}}$)