LECTURE 21

A REFRESHER IN LINEAR ALGEBRA.

Def/Prop | SKIP | An $n\times n$ matrix U is orthogonal (denoted $U\in O(n)$) if one of the following equivalent properties hold:

(a) U is invertible and $U^{-1}=U^T$

(b) $U^TU=I$

(c) $UU^T=I$

(d) the columns of U are orthonormal ($\Rightarrow$ o.n. basis)

(e) the rows of U are orthonormal ($\Rightarrow$ o.n. basis)

SPECTRAL DECOMPOSITION

THM (Spectral thm for symmetric matrices)

$\forall$ symmetric $n\times n$ matrix A,

(a) all eigenvalues $\lambda_1 > \lambda_2 > \dots > \lambda_n$ are real;

(b) the unit eigenvectors $u_1,\dots,u_n$ are orthonormal ($\Rightarrow$ o.n. basis)

Cor (Spectral decomposition) $\forall$ symmetric $n\times n$ matrix A:

$$A=\sum_{i=1}^{r}\lambda_i u_i u_i^T \quad \text{where } r=\operatorname{rank}(A)$$

and (λ_i, u_i) are eigenvalues & unit eigenvectors of A.

$\forall x\in\mathbb{R}^n$: $\quad x=\sum_{i=1}^{n}\langle u_i,x\rangle u_i \quad$ (o.n. basis decomposition)

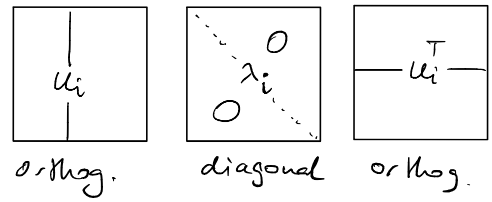
$$\Rightarrow Ax=\sum_{i=1}^{n}\langle u_i,x\rangle \underbrace{Au_i}_{=\lambda_i u_i} = \underbrace{\left(\sum \lambda_i u_i u_i^T\right)}_{=B} x$$

$\Rightarrow Ax = Bx \quad \forall x \overset{\text{WHY?}}{\Longrightarrow} A=B.$

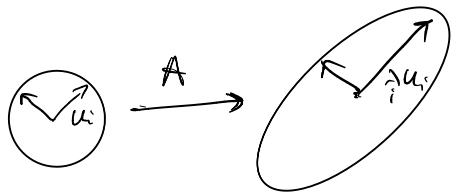
REMARKS (a) Matrix form:

$$A = U \Lambda U^T$$

$A = U \cdot \Lambda \cdot U^T$



(b) Geometric form:



stretches directions u_i
λ_i times

(c) Optimization form:

$$\lambda_1 = \langle \overset{\lambda_1 u_1}{\overset{\|}{A u_1}}, u_1 \rangle = \max_{x \text{ unit}} x^T A x, \quad u_1 = \operatorname{argmax}$$

$$\lambda_2 = \langle A u_2, u_2 \rangle = \max_{\substack{x \perp u_1 \\ \text{unit}}} x^T A x, \quad u_2 = \operatorname{argmax}$$

$$\lambda_3 = \langle A u_3, u_3 \rangle = \max_{\substack{x \perp \{u_1, u_2\} \\ \text{unit}}} x^T A x, \quad u_3 = \operatorname{argmax}$$

....

$$\lambda_n = \langle A u_n, u_n \rangle = \min_{x \text{ unit}} x^T A x, \quad u_n = \operatorname{argmin}$$

SINGULAR VALUE DECOMPOSITION

THM (SVD) $\forall$ $m \times n$ matrix A can be expressed as

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \qquad (*)$$

where $r = \text{rank}(A)$ and:

$\sigma_1 \geq \sigma_2 \geq \dots > 0$ "singular values":

$u_1, \dots, u_m \in \mathbb{R}^m$ are orthonormal "left-singular vectors"

$v_1, \dots, v_n \in \mathbb{R}^n$ are orthonormal "right-singular vectors"

(informal) Let's guess how to find σ_i, u_i, v_i.

$$AA^T = \left(\sum_i \sigma_i u_i v_i^T\right)\left(\sum_j \sigma_j v_j u_j^T\right) = \sum_{ij} \sigma_i \sigma_j \, u_i \underbrace{v_i^T v_j}_{\langle v_i, v_j\rangle = \begin{cases} 1 & \text{if } i=j \\ 0 & \end{cases}} u_j^T = \sum_i \sigma_i^2 u_i u_i^T$$

Similarly,

$$A^T A = \sum_i \sigma_i^2 v_i v_i^T. \quad \text{Thus:}$$

$$\sigma_i(A) = \sqrt{\lambda_i(AA^T)} = \sqrt{\lambda_i(A^T A)}.$$

$u_i(A)$ = eigenvectors of AA^T

$v_i(A)$ = eigenvectors of $A^T A$.

SKIP

(More formal, for $m=n=r$):
$AA^T \succeq 0 \Rightarrow$ by Spectral Thm, $AA^T = U \Lambda U^T$.
For simplicity assume $r = m = n$. Set $\Sigma := \Lambda^{1/2}$, $V := A^T U \Sigma^{-1}$.
and check (DIY)

SKIP

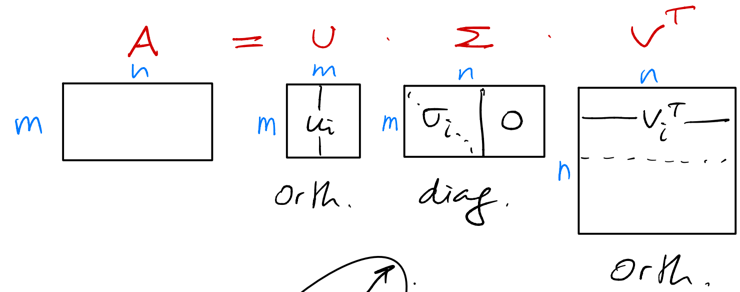
Lem $\boxed{A \succeq 0 \text{ (PSD)} \iff \lambda_i \geq 0 \;\forall i}$

"$\Rightarrow$" if $A \succeq 0$, $0 \leq u_i^T A u_i = u_i^T \cdot \lambda u_i = \lambda_i \|u_i\|_2^2 = \lambda_i \;\forall i$.

"$\Leftarrow$": If all $\lambda_i \geq 0$, $x^T A x = \sum_{i=1}^{n} \lambda_i \langle u_i, x\rangle^2 \geq 0$.

REMARKS (a) Matrix form:

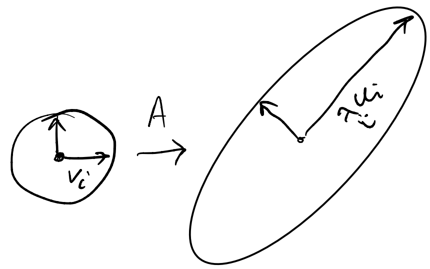
$$A = U \Sigma V^T$$



(b) Geometric form:

rotate $v_i \mapsto u_i$

& stretch u_i $\times \lambda_i$ times



(c) Optimization form:

$$\sigma_1 = \max_{x \text{ unit}} \|Ax\|_2 , \quad v_1 = \text{argmax}$$

$$\sigma_2 = \max_{\substack{x \perp v_1 \\ \text{unit}}} \|Ax\|_2 , \quad v_2 = \text{argmax} \ldots$$

$$\sigma_3 = \max_{\substack{x \perp \{v_1, v_2\} \\ \text{unit}}} \|Ax\|_2 , \quad v_3 = \text{argmax} \ldots$$

$$\ldots \qquad \ldots$$

$$\sigma_n = \min_{x \text{ unit}} \|Ax\|_2 \quad v_n = \text{argmin}$$

EXTREME VALUES

In particular, $\sigma_1(A) = \max_x \dfrac{\|Ax\|_2}{\|x\|_2}$, $\sigma_n(A) = \min_x \dfrac{\|Ax\|_2}{\|x\|_2}$

- Maximal distortion of geometry, like in JL lemma:

$$\sigma_n(A)\|x-y\|_2 \le \underbrace{\|Ax - Ay\|_2}_{= \|A(x-y)\|_2} \le \sigma_1(A)\|x-y\|_2 \qquad \forall x, y$$

GEOMETRIC MEANING

SKIP

<u>Prop</u> For a symmetric $n\times n$ matrix A:

$$\lambda_1(A) = \max_{\|x\|_2=1} x^T A x, \quad \lambda_n(A) = \min_{\|x\|_2=1} x^T A x$$

$u_1 = \arg\max \qquad u_n = \arg\min$

$A = U\Lambda U^T \Rightarrow$

$$\max_x x^T A x = \max_x x^T U \Lambda U^T x = \max_x (\underbrace{U x}_{=y})^T \Lambda (U x) \qquad U\in O(n) \Rightarrow$$

$$= \max_{y \text{ unit}} y^T \underbrace{\Lambda}_{\operatorname{diag}(\lambda_1,\dots,\lambda_n)} y = \max_{\sum y_i^2 = 1} \sum_i \lambda_i y_i^2 \qquad y_i^2 =: z_i$$

$$= \max_{\substack{\sum z_i = 1 \\ z_i \geq 0}} \sum \lambda_i z_i = \max \{ u \in \operatorname{Conv}(\lambda_i)\} = \max_i \lambda_i$$

$$= \lambda_1(A).$$

- <u>Remark</u>: quadratic optimization problem over sphere: tractable! Spectral

SKIP

- $\forall$ $m\times n$ matrix A, apply Prop for A^TA

$$\lambda_k(A^TA)=\sigma_k(A)^2, \quad x^T(A^TA)x=\langle Ax, Ax\rangle = \|Ax\|_2^2 \Rightarrow$$

$$\sigma_1(A)=\max_{\|x\|_2=1}\|Ax\|_2=\max_x \frac{\|Ax\|_2}{\|x\|_2}; \qquad \sigma_n(A)=\min_{\|x\|_2=1}\|Ax\|_2=\min_x\frac{\|Ax\|_2}{\|x\|_2}$$

- Maximal distortion of geometry, like in JL lemma:

$$\|Ax-Ay\|_2=\|A(x-y)\|_2 \begin{cases}\le \sigma_1(A)\|x-y\|_2 \\ \ge \sigma_n(A)\|x-y\|_2\end{cases} \Rightarrow$$

Prop We have

$$m\|x-y\|_2 \le \|Ax-Ay\|_2 \le M\|x-y\|_2 \quad \forall x,y$$

where $m=\sigma_n(A)$, $M=\sigma_1(A)$. These are best possible m, M.

- Intermediate singular values:

$$\sigma_1(A)=\max_{x \text{ unit}}\|Ax\|_2, \quad v_1=\operatorname{argmax}\ldots$$

$$\sigma_2(A)=\max_{x\perp v_1}\|Ax\|_2, \quad v_2=\operatorname{argmax}\ldots$$

$$\sigma_3(A)=\max_{x\perp\{v_1,v_2\}}\|Ax\|_2, \quad v_3=\operatorname{argmax}\ldots$$

$$\ldots$$

$$\sigma_n(A)=\min_{x \text{ unit}}\|Ax\|_2 \quad v_n=\operatorname{argmin}$$

- Can be derived from min-max thm for A^TA

$\downarrow$

THM (Min-max thm) $\forall$ symmetric $n\times n$ A, $\forall k$:

$$\lambda_k(A)=\max_{\dim E=k}\ \min_{\substack{x\in E\\ \text{unit}}} x^TAx=\min_{\dim F=n-k+1}\ \max_{\substack{x\in F\\ \text{unit}}} x^TAx$$

Moreover, the optimal E, F are spanned by eigenvectors of A:

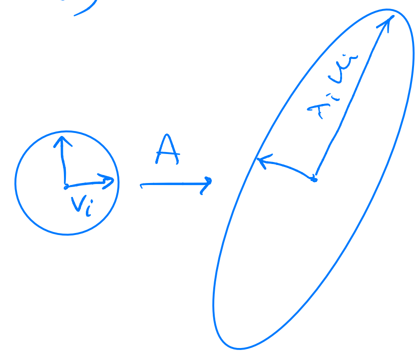
$E=\operatorname{span}\{u_1,\ldots,u_k\}$, $F=\operatorname{span}\{u_k,\ldots,u_n\}$

The optimal x is the eigenvector u_k

SKIP

Prop (Geometric meaning of singular values & vectors)

$$\boxed{\begin{array}{llll}
\sigma_1 = \max\limits_{x \text{ unit}} \|Ax\|_2, & & v_1 = \operatorname{argmax}\limits_{x \text{ unit}} \|Ax\|_2 \\
\sigma_2 = \max\limits_{\substack{x \perp v_1 \\ \text{unit}}} \|Ax\|_2, & & v_2 = \operatorname{argmax} \ldots \\
\sigma_3 = \max\limits_{\substack{x \perp \{v_1, v_2\} \\ \text{unit}}} \|Ax\|_2, & & v_3 = \operatorname{argmax} \ldots \\
\ldots & & \ldots \\
\sigma_n = \min\limits_{x \text{ unit}} \|Ax\|_2 & & v_n = \operatorname{argmin}
\end{array}}$$



In particular, $\sigma_1(A) = \max\limits_{x} \dfrac{\|Ax\|_2}{\|x\|_2}$, $\sigma_n(A) = \min\limits_{x} \dfrac{\|Ax\|_2}{\|x\|_2}$

- Maximal distortion of geometry, like in JL Lemma:

$$\sigma_n(A)\|x-y\|_2 \le \underbrace{\|Ax - Ay\|_2}_{= \|A(x-y)\|_2} \le \sigma_1(A)\|x-y\|_2 \qquad \forall x, y$$

REMARK | If A is symmetric, a version of Prop for eigenvalues/vectors:

$$\lambda_1 = \max_{x \text{ unit}} x^T A x, \qquad u_1 = \operatorname{argmax} \ldots$$

$$\lambda_2 = \max_{\substack{x \perp u_1 \\ \text{unit}}} x^T A x \qquad u_2 = \operatorname{argmax} \ldots$$

MATRIX NORMS.

A : $m \times n$ matrix

Def Frobenius (a.k.a. Hilbert-Schmidt) norm:

$$\boxed{\|A\|_F^2 = \sum_{i,j} A_{ij}^2}$$

$=$ Euclidean norm on $\mathbb{R}^{m \times n}$ $\Leftarrow$ Hilbert space $\Rightarrow$ True norm

$$\langle A, B \rangle = \sum_{i,j} A_{ij} B_{ij} = \mathrm{tr}(A^T B). \qquad (*)$$

$$\Rightarrow \langle A, A \rangle = \|A\|_F^2.$$

Prop (Orthogonal invariance) $\forall$ orthogonal U, V:

$$\boxed{\|UA\|_F = \|AV\|_F = \|A\|_F}$$

$$\left[\ \|UA\|_F = \langle UA, UA \rangle \overset{(*)}{=} \mathrm{tr}(A^T \underbrace{U^T U}_{= I} A) = \mathrm{tr}(A^T A) = \langle A, A \rangle = \|A\|_F^2\ \right]$$

Prop

$$\boxed{\|A\|_F^2 = \sum_i \sigma_i(A)^2}$$

singular values

$$\left[\ \|A\|_F^2 \overset{SVD}{=} \|U^T \Sigma V\|_F^2 = \|\underset{\mathrm{diag}(\sigma_i)}{\Sigma}\|_F^2 = \sum \sigma_i\ \right]$$

Def The spectral (a.k.a. operator) norm:

$$\|A\| = \max_{\|x\|_2=1} \|Ax\|_2 = \max_{\|x\|_2\le 1} \|Ax\|_2 = \max_{x} \frac{\|Ax\|_2}{\|x\|_2}.$$

TRUE NORM.

We showed (p.3) that $\boxed{\|A\| = \sigma_1(A)}$ $(*)$

Facts

(a) Orthogonal invariance: $\|UA\| = \|AV\| = \|A\|$ $\forall$ orthog U, V ⌈DIY⌋

(a) $\|A^{-1}\| = \dfrac{1}{\sigma_n(A)}$ ⌈DIY⌋

(b) $\|A^T\| = \|A\|$ ⌈Eigenvalues of a matrix & its transpose are equal (since char. polynomial is the same) $\Rightarrow$ sing. values, too⌋

(c) $\|A+B\| \le \|A\| + \|B\|$ (Δ ineq.)

(d) $\|AB\| \le \|A\|\,\|B\|$ ⌈$\|ABx\|_2 \le \|A\|\cdot\|Bx\|_2 \le \|A\|\cdot\|B\|\cdot\|x\|_2$⌋

(e) $\|A\| \le \|A\|_F \le \sqrt{r}\,\|A\|$ (where $r = \operatorname{rank}(A)$)

⌈$\sigma_1(A)^2 \le \sum_{i=1}^{r} \sigma_i(A)^2 \le r\cdot\sigma_1(A)^2$⌋

(f) $\|A\| = \max_{x,y \text{ unit}} y^T A x.$ ⌈$\|z\|_2 = \max_{\|y\|_2=1} y^T z$, use for $z = Ax$⌋

(g) A symmetric $\Rightarrow$ $\|A\| = \max_{\|x\|_2=1} |x^T A x|$ ⌈(DIY)⌋

SKIP

Low-rank approximations

THM $\forall$ $m\times n$ matrix A, $\exists$ matrix A_k of rank k s.t.

$$\|A-A_k\| = \sigma_{k+1}(A)$$

$\ulcorner A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T \Rightarrow A - A_k = \sum_{i>k} \sigma_i u_i v_i^T.$

$\|A-A_k\| = \max$ singular value $(A-A_k) = \sigma_{k+1}. \lrcorner$

• **Remark** Moreover, A_k is the best rank-k approximation

• $\forall$ B with $\mathrm{rank}(B)\le k$, let P be the orthogonal projection onto the image of B. $\Rightarrow$

$$\|A-B\| = \|A-PB\| \geq \|(1-P)(A-PB)\|$$

$$= \|(1-P)A\|$$

$$= \|A^T(1-P)\| = \max_{\|x\|_2\le 1} \|A^T \underbrace{(1-P)x}_{=y}\|$$

$$= \max_{\substack{y\in F \\ \|y\|_2\le 1}} \|A^T x\| \qquad \text{where } F = (\mathrm{Im}\, B)^\perp \Rightarrow \dim E \le n-k$$

$$= \sigma_{k+1}(A^T) \qquad \text{(Min-MAX Thm)}$$

$$= \sigma_{k+1}(A). \lrcorner$$

Remark: Same for Frobenius norm.