Lecture 2-1. Math Background Review.

1. Basic Analysis:

A. 1° Norm: A fn $f: \mathbb{R}^n \to \mathbb{R}$ is called norm if:

* (non-neg.): $f(\underline{x}) \geq 0$, $\forall \underline{x} \in \mathbb{R}^n$, $f(\underline{x}) = 0$ iff $\underline{x} = 0$.
* (homogeneity): $f(t\underline{x}) = |t| f(\underline{x})$, $\forall \underline{x} \in \mathbb{R}^n$, $t \in \mathbb{R}$.
* (triangle ineq.): $f(\underline{x}+\underline{y}) \leq f(\underline{x}) + f(\underline{y})$, $\forall \underline{x}, \underline{y} \in \mathbb{R}^n$.

If $f(\underline{x})$ is a norm, we denote it as $\|\underline{x}\|$.

2° Norm $\|\underline{x}\|$'s meaning:

* $\|\underline{x}\|$: length of $\underline{x}$
* $\|\underline{x}-\underline{y}\|$: dist. btwn $\underline{x}$ and $\underline{y}$.

3°. Unit Ball: set of vectors with $\|\underline{x}\| \leq 1$.

$$\mathcal{B} = \{\underline{x} \in \mathbb{R}^n : \|\underline{x}\| \leq 1\}.$$

Ex: * l_2-norm (Euclidean norm): $\|\underline{x}\|_2 \triangleq (\underline{x}^T\underline{x})^{\frac{1}{2}} = (x_1^2 + \cdots + x_n^2)^{\frac{1}{2}}$.

* l_1-norm (sum_abs_val): $\|\underline{x}\|_1 \triangleq |x_1| + |x_2| + \cdots + |x_n|$.
* l_∞-norm (Chebyshev): $\|\underline{x}\|_\infty \triangleq \max\{|x_1|, \cdots |x_n|\}$.
* l_p-norm $(p \geq 1)$: $\|\underline{x}\|_p \triangleq (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}$

$$\|\underline{x}\|_\infty = \lim_{p \to \infty} \|\underline{x}\|_p$$

Proof: $\|\underline{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} = \left(\frac{|x_1|^p}{\|\underline{x}\|_\infty^p} + \cdots + \frac{|x_n|^p}{\|\underline{x}\|_\infty^p}\right)^{\frac{1}{p}} \|\underline{x}\|_\infty$

$\leq (1 + \cdots + 1)^{\frac{1}{p}} \|\underline{x}\|_\infty = n^{\frac{1}{p}} \cdot \|\underline{x}\|_\infty \to \|\underline{x}\|_\infty$ as $p \to \infty$.

Let $i^* \in \operatorname{argmax}\{|x_i|, \forall i\}$.

$\|\underline{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}} \geq (|x_{i^*}|^p)^{\frac{1}{p}} = |x_{i^*}| = \|\underline{x}\|_\infty$.

Note $n^{\frac{1}{p}} \to 1$ as $p \to \infty$. ∎

4° Equivalence of Norms.

Suppose $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms of $\mathbb{R}^n$. Then $\exists\ \alpha, \beta > 0$ s.t. $\forall \underline{x} \in \mathbb{R}^n$, $\alpha \|\underline{x}\|_a \le \|\underline{x}\|_b \le \beta \|\underline{x}\|_a$.

Ex: $\|\underline{x}\|_2 \le \|\underline{x}\|_1 \le \sqrt{n}\, \|\underline{x}\|_2$.

$\|\underline{x}\|_\infty \le \|\underline{x}\|_2 \le \sqrt{n}\, \|\underline{x}\|_\infty$

$\|\underline{x}\|_\infty \le \|\underline{x}\|_1 \le n \|\underline{x}\|_\infty$.

2. Convergent Sequences and Limits:

1° Def (Convergence): A seq. of vectors $\underline{x}_1 \cdots \underline{x}_n \cdots$ are said to be convergent to a limit pt. $\bar{\underline{x}}$ if $\forall \varepsilon > 0$, $\exists N_\varepsilon \in \mathbb{N}$ s.t. $\|\underline{x}_k - \bar{\underline{x}}\| < \varepsilon$, $\forall k \ge N_\varepsilon$. ($\{\underline{x}_k\} \to \bar{\underline{x}}$ as $k \to \infty$, $\lim_{k\to\infty} \underline{x}_k = \bar{\underline{x}}$).

2°. Def (Cauchy Seq.): A seq. $\{\underline{x}_k\}$ is Cauchy if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ s.t. $\|\underline{x}_m - \underline{x}_n\| < \varepsilon$, $\forall m, n \ge N$.

Thm: A seq. in $\mathbb{R}^n$ has a limit iff it's Cauchy.

Ex: (p-series). $a_n = \frac{1}{n^p}$. show $\{b_n\} \triangleq \left\{ \sum_{k=1}^{n} a_k \right\}$ has a limit for $p=2$, but doesn't converge for $p=1$.

Proof: W.l.o.g., let $m, n \in \mathbb{N}$ and $m<n$.

For $p=2$, $b_n - b_m = \sum_{k=1}^{n} \frac{1}{k^2} - \sum_{k=1}^{m} \frac{1}{k^2} = \sum_{k=m+1}^{n} \frac{1}{k^2} \le \sum_{k=m+1}^{n} \frac{1}{k} \cdot \frac{1}{k-1}$

$$= \sum_{k=m+1}^{n} \left(\frac{1}{k-1} - \frac{1}{k} \right) = \frac{1}{m} - \cancel{\frac{1}{m+1}} + \cancel{\frac{1}{m+1}} - \cancel{\frac{1}{m+2}} + \cdots + \cancel{\frac{1}{n-1}} - \frac{1}{n}$$

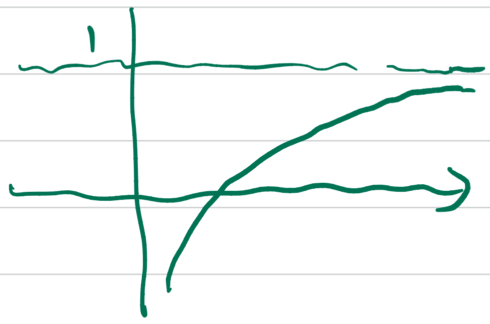
$$= \frac{1}{m} - \frac{1}{n} < \frac{1}{m} < \varepsilon$$, if m is suff. large.

For $p=1$, $b_n - b_m = \sum_{k=1}^{n} \frac{1}{k} - \sum_{k=1}^{m} \frac{1}{k} = \sum_{k=m+1}^{n} \frac{1}{k} = \underset{\geq \frac{1}{n}}{\frac{1}{m+1}} + \cdots + \underset{\geq \frac{1}{n}}{\frac{1}{n}}$

$\geq \frac{n-m}{n} = 1 - \frac{m}{n}$.

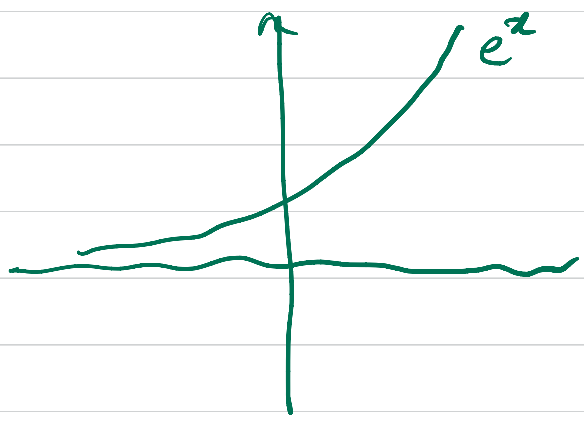
Consider any $\varepsilon > 0$, no matter how large m is, can choose $n \geq \left\lceil \frac{m}{1-\varepsilon} \right\rceil$, so that $b_n - b_m \geq \varepsilon$. $\boxtimes$

3°. Supremum of S (least UB): Smallest possible α satisfying $\alpha \geq x$, $\forall x \in S$.



$1 - \frac{1}{x}$, $\underbrace{x > 0}_{S}$.

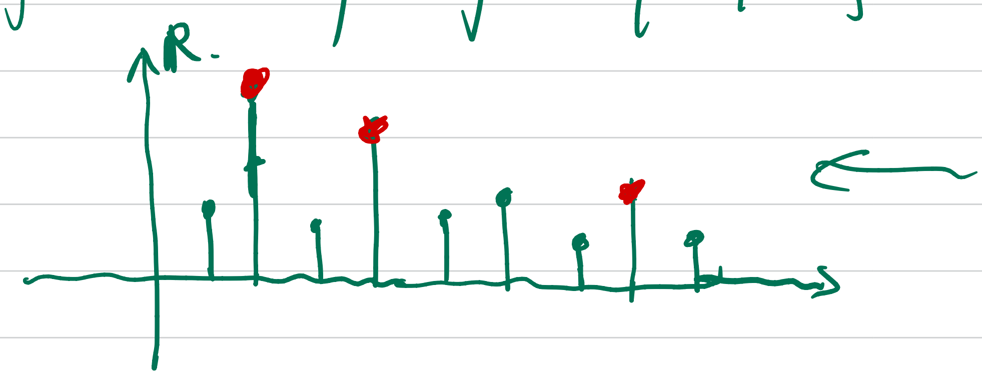
4° Infimum of S (largest LB): largest ------ ------ $\alpha \leq x$, $\forall x \in S$.

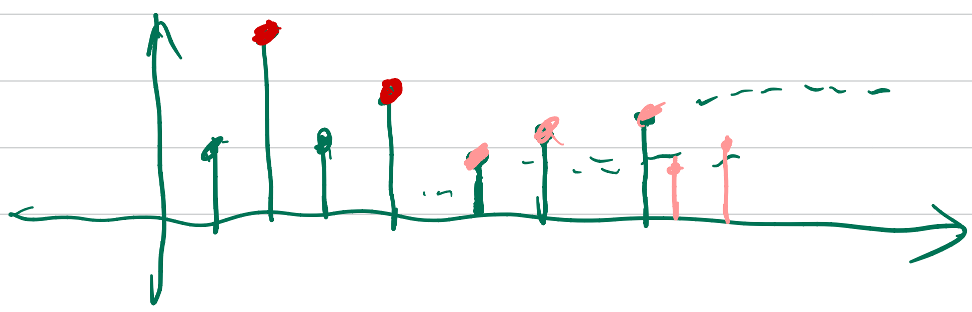


e^x, $\underbrace{x \in \mathbb{R}}_{S}$: Infimum : 0.

Thm: (Bolzano–Weierstrass): Every bdded seq. in $\mathbb{R}^n$ has a convergent subseq.

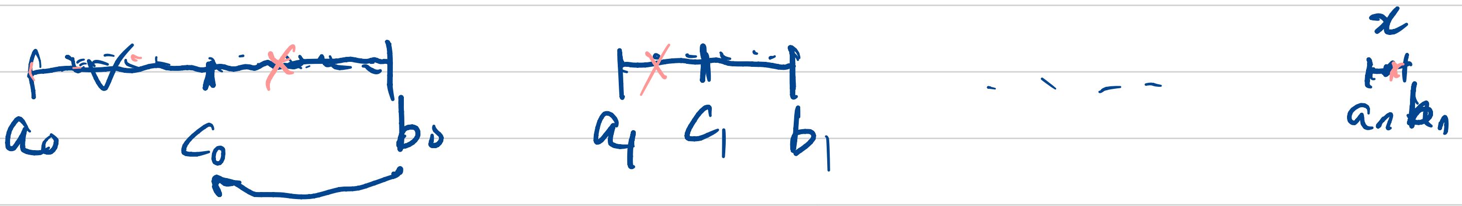
Proof: 1° 1. Every inf. seq. $\{x_n\}$ in $\mathbb{R}^1$ has mono. subseq.





2. <u>MCT</u>: If $\{a_n\}$ is mono seq. reals, then $\{a_n\}$ has a limit iff $\{a_n\}$ is bnded.

3. $\mathbb{R}^n$: Extract subseq. each dim. sequentially. ☒

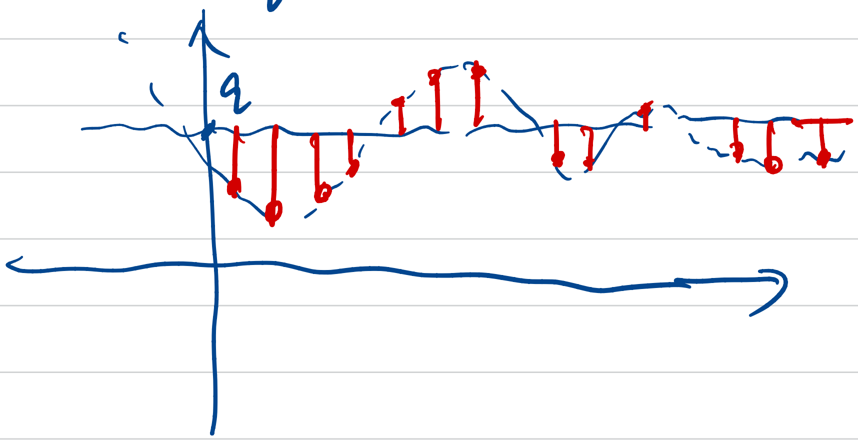
2°. 

☒

3° Maximum, Minimum. (achievable).

4° limsup. liminf:

* The limit supremum $\limsup_{k\to\infty} x_k$, is infimum of all $q \in \mathbb{R}$ for which all but a finite # of elements in $\{x_k\}$ exceed q.



$$\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \{\underbrace{\sup_{m\ge n} x_m}_{a_n}\}.$$

* ... limit infimum ... $\liminf_{k\to\infty} x_k$ ... supremum ... q.

$$\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \{\inf_{m\ge n} x_m\}.$$

* limsup and liminf always exist.

$\{x_n\}$ converge iff $\limsup_{n\to\infty} x_n = \liminf_{n\to\infty} x_n$.

3. Functions:

1° <u>Cont. fn</u>: A fn $f: S \to \mathbb{R}$ is cont. at $\bar{x} \in S$ if $\forall \varepsilon > 0$, $\exists \delta > 0$, s.t. $x \in S$ with $\|x - \bar{x}\| < \delta \to |f(x) - f(\bar{x})| < \varepsilon$.

write: $f(x) \to f(\bar{x})$, as $x \to \bar{x}$.

<u>Fact</u>: Cont. fn achieves both maximum & minimum over a non-empty $\underbrace{\text{compact}}_{\text{closed \& bnded.}}$ set.

2° Diff'ble fn:

(1) S non-empty set in $\mathbb{R}^n$, $x \in \text{int}\, S$, and $f: S \to \mathbb{R}$.

f is diff'ble at $\bar{x}$ if $\exists$ a vector (called gradient)

$$\nabla f(\bar{x}) \triangleq \left[\frac{\partial f(\bar{x})}{\partial x_1}, \cdots, \frac{\partial f(\bar{x})}{\partial x_n}\right]^T$$ at $\bar{x}$ and fn $\beta(x, \bar{x}) \to 0$.

as $x \to \bar{x}$, such that:

$$f(x) = \underbrace{f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})}_{\text{FO-approx.}} + \underbrace{\|x - \bar{x}\| \beta(x, \bar{x})}_{o(\|x - \bar{x}\|)}, \quad \forall x \in S$$

(2). f is called twice diff'ble at $\bar{x}$ if, in addition to grad, $\exists$ symmetric matrix $H(\bar{x})$ (called Hessian matrix) of f at $\bar{x}$, and $\beta(x, \bar{x}) \to 0$ as $x \to \bar{x}$, such that:

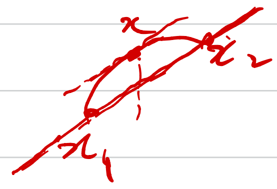
$$f(x) = \underbrace{f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}) + \frac{1}{2}(x - \bar{x})^T H(\bar{x})(x - \bar{x})}_{\text{SO-approx}} + \underbrace{\|x - \bar{x}\|^2 \beta(x, \bar{x})}_{o(\|x - \bar{x}\|^2).}$$

$$H(x) \triangleq \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

3° A vector-valued fn f is diff'ble (twice) if each component is diff'ble (twice).

A diff'ble vector-valued fn $h: \mathbb{R}^m \to \mathbb{R}^n$, the Jacobian, denoted by $\nabla \underline{h}(\underline{x})$, is given by the $n \times m$ matrix.

$$\underline{\underline{J}}(\underline{x}) = \nabla \underline{h}(\underline{x}) = \begin{bmatrix} \nabla h_1(\underline{x})^T \\ \vdots \\ \nabla h_n(\underline{x})^T \end{bmatrix}_{n \times m}.$$



4° (MVT): S non-empty open convex set in $\mathbb{R}^n$, let $f: S \to \mathbb{R}$ be diff'ble. For every $\underline{x}_1, \underline{x}_2 \in S$, we have:

$f(\underline{x}_2) = f(\underline{x}_1) + \nabla f(\underline{x})^T (\underline{x}_2 - \underline{x}_1)$, where $\underline{x} = \lambda \underline{x}_1 + (1-\lambda)\underline{x}_2$ for some $\lambda \in (0,1)$.

5° Taylor's Thm: S non-empty, open, convex set in $\mathbb{R}^n$. $f: S \to \mathbb{R}$, twice diff'ble. For every $\underline{x}_1, \underline{x}_2 \in S$, we have:

$f(\underline{x}_2) = f(\underline{x}_1) + \nabla f(\underline{x}_2)^T (\underline{x}_2 - \underline{x}_1) + \frac{1}{2}(\underline{x}_2 - \underline{x}_1)^T \underline{H}(\underline{x})(\underline{x}_2 - \underline{x}_1)$, where $\underline{H}(\underline{x})$ is Hessian at $\underline{x}$, and $\underline{x} = \lambda \underline{x}_1 + (1-\lambda)\underline{x}_2$, for some $\lambda \in (0,1)$.

Linear Algebra:

1. Linear indep.: $\underline{x}_1 \cdots \underline{x}_k \in \mathbb{R}^n$ are lin. indep. if
$$\sum_{i=1}^{k} \alpha_i \underline{x}_i = \underline{0} \Rightarrow \lambda_i = 0, \ \forall i = 1, \ldots, k.$$

2. linear comb.: $y \in \mathbb{R}^n$ is a lin. comb. of $x_1, \cdots, x_k \in \mathbb{R}^n$ if

$$y = \sum_{i=1}^{k} \lambda_i x_i \quad \text{for some } \lambda_1, \cdots, \lambda_k.$$

* $\sum_{i=1}^{k} \lambda_i = 1$: y is an <u>affine comb.</u> of $x_1, \cdots, x_k$.
* $\sum_{i=1}^{k} \lambda_i = 1$, $\lambda_i \geq 0$, $\forall i$: y is a <u>convex comb.</u> of $x_1, \cdots, x_k$.

The linear, affine, convex hull of $S \in \mathbb{R}^n$ are, resp., the set of <u>all</u> lin., affine, convex comb. of pts. in S.

3. Spanning vectors: $x_1 \cdots x_k \in \mathbb{R}^n$, $k \geq n$, said to be spanning $\mathbb{R}^n$ if <u>any</u> vector in $\mathbb{R}^n$ can be represented as lin. comb. of $x_1, \cdots, x_k$.

The <u>cone</u> spanned by $x_1 \cdots x_k$ is the set of non-neg. lin. comb.

4. <u>Basis</u>: A set of $x_1 \cdots x_k \in \mathbb{R}^n$ spans $\mathbb{R}^n$ and if the deletion of any of $x_1 \cdots x_k$ prevents remaining vectors from spanning $\mathbb{R}^n$. (Basis $x_1 \cdots x_k$ spans $\mathbb{R}^n$ iff $k = n$).

5. Cauchy–Schwartz Ineq.: $|\langle x, y\rangle| = |x^T y| \leq \|x\|_2 \cdot \|y\|_2$, with equality achieved iff x, y are lin dep.

(unsigned) angle btwn $x, y \in \mathbb{R}^n$.

$$\angle(x, y) \triangleq \cos^{-1}\left(\frac{x^T y}{\|x\|_2 \cdot \|y\|_2}\right) \in [0, \pi]$$

(x, y are orthogonal, $x \perp y$, if $\langle x, y\rangle = 0$)

6. Young's Ineq: $a>0$, $b>0$, and any $p,q>0$ s.t. $\frac{1}{p}+\frac{1}{q}=1$.
we have $ab \le \frac{a^p}{p}+\frac{b^q}{q}$, with eq. achieved iff $a^p=b^q$.
(special case, $p=q=2$).

7. Holder's Ineq: For any pair of vectors $\underline{x}$ and $\underline{y} \in \mathbb{C}^n$, and for any p, q satisfying $\frac{1}{p}+\frac{1}{q}=1$, we have:

$$\sum_{i=1}^{n} |x_i y_i| \le \|\underline{x}\|_p \cdot \|\underline{y}\|_q .$$

(special case: $p=q=2$, $p=1, q=\infty$).

8. Othogonal matrix: $\underline{Q} \in \mathbb{R}^{m\times n}$: $\underline{Q}^T\underline{Q} = \underline{I}_n$ or $\underline{Q}\underline{Q}^T = \underline{I}_m$.
If $\underline{Q}$ is square: $\underline{Q}^{-1} = \underline{Q}^T$.

9. Rank of matrix: For $\underline{A} \in \mathbb{R}^{m\times n}$, $\operatorname{rank}(\underline{A}) \triangleq$ max # of lin. indep. rows (or equivalently cols) of $\underline{A}$.

If $\operatorname{rank}(\underline{A}) = \min\{m,n\}$, $\underline{A}$ is full row/col rank.

10. Eigenvalues and eigenvectors: $\underline{A} \in \mathbb{R}^{n\times n}$. If λ and $\underline{x} \ne \underline{0}$ satisfy $\underline{A}\underline{x} = \lambda \underline{x}$, then λ and $\underline{x}$ are eigenvalue & eigenvector.

* λ can be computed by solving $\det(\underline{A} - \lambda \underline{I}) = 0$ (characteristic eqn.).
* $\underline{A}$ is symmetric $\Rightarrow$ n (possibly non-distinct) real eigenvalues.
* Eigenvectors assoc. with distinct eigenvalues are orthogonal.
* Given some symmetric $\underline{A}$ $\Rightarrow$ can construct a orthogonal basis $\underline{B} \in \mathbb{R}^{n\times r}$ ← rank, where each col in $\underline{B}$ is an eigenvector of $\underline{A}$.

* Normalize $\underline{\underline{B}}$ to have unit l_2 norm, s.t. $\underline{\underline{B}}^T\underline{\underline{B}} = \underline{\underline{I}}$ ($\underline{\underline{B}}^T = \underline{\underline{B}}^{-1}$).
Then, $\underline{\underline{B}}$ is called orthonormal matrix.

*. Let $\lambda_1 \cdots \lambda_n$ be real eigenvalues of symmetric $\underline{\underline{A}}$

Let $\underline{\underline{\Lambda}} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$. $\underline{\underline{A}}\,\underline{\underline{B}} = \underline{\underline{B}}\,\underline{\underline{\Lambda}}$

$$\underline{\underline{A}} \begin{bmatrix} \vdots & & \vdots \\ \underline{v}_1 & \cdots & \underline{v}_n \\ \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & & \vdots \\ \underline{v}_1 & \cdots & \underline{v}_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$\overset{\underline{\underline{B}}^T = \underline{\underline{B}}^{-1}}{\Longrightarrow} \quad \underline{\underline{A}} = \underline{\underline{B}}\,\underline{\underline{\Lambda}}\,\underline{\underline{B}}^T = \sum_{i=1}^{n} \lambda_i \underline{b}_i \underline{b}_i^T$$

11. Singular - Value Decomp. (SVD).

Let $\underline{\underline{A}} \in \mathbb{R}^{m \times n}$. Then $\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T$, where $\underline{\underline{U}} \in \mathbb{R}^{m \times m}$ orthonormal, $\underline{\underline{V}} \in \mathbb{R}^{n \times n}$ orthonormal, and $\underline{\underline{\Sigma}} \in \mathbb{R}^{m \times n}$, $(\underline{\underline{\Sigma}})_{ij} = 0$ for $i \neq j$.
$(\underline{\underline{\Sigma}})_{ii} \geq 0. \in \mathbb{R}$

* Cols of $\underline{\underline{U}}$: Normalized eigenvectors of $\underline{\underline{A}}\,\underline{\underline{A}}^T$
* Cols of $\underline{\underline{V}}$: — — — of $\underline{\underline{A}}^T\underline{\underline{A}}$
* $(\underline{\underline{\Sigma}})_{ii}$: Abs square root of eigenvalues of $\underline{\underline{A}}^T\underline{\underline{A}}$ if $m \leq n$
or $\underline{\underline{A}}\,\underline{\underline{A}}^T$ if $m \geq n$.

12. Definite & Semidefinite Matrices: $\underline{\underline{A}} \in \mathbb{R}^{n\times n}$ symmetric.

$\underline{\underline{A}}$ is

PD	if	$\underline{x}^T \underline{\underline{A}}\, \underline{x} > 0$	$\forall \underline{x} \neq 0,\ \underline{x} \in \mathbb{R}^n$
PSD		≥ 0	$\forall \underline{x} \in \mathbb{R}^n$
ND		< 0	$\forall \underline{x} \neq 0,\ \underline{x} \in \mathbb{R}^n$
NSD		≤ 0	$\forall \underline{x} \in \mathbb{R}^n$.

$\underline{\underline{A}}$ is indef. if neither PSD nor NSD.

$\underline{\underline{A}}$ is

PD	if eigenvalues are	pos.
PSD		non-neg., resp.
ND		neg.
NSD		non-pos.

13. If $\underline{\underline{A}}$ is PSD, $\underline{\underline{A}}^{\frac{1}{2}}$ is the matrix satisfying

$$\underline{\underline{A}}^{\frac{1}{2}} \cdot \underline{\underline{A}}^{\frac{1}{2}} = \underline{\underline{A}}, \quad \text{and} \quad \underline{\underline{A}}^{\frac{1}{2}} = \underline{\underline{B}}\, \underline{\underline{\Lambda}}^{\frac{1}{2}}\, \underline{\underline{B}}^T$$

$$\uparrow$$

$$\begin{bmatrix} \lambda_1^{\frac{1}{2}} & & 0 \\ & \ddots & \\ 0 & & \lambda_n^{\frac{1}{2}} \end{bmatrix}$$