Math Background Review

Basic Analysis:

A. 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 $\underline{x} = 0$.
* (homogenity): $f(t\underline{x}) = |t| f(\underline{x})$, $\forall \underline{x} \in \mathbb{R}^n$, $t \in \mathbb{R}$.
* (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: $\|\underline{x}\|$.

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

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

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.): $\|\underline{x}\|_1 \triangleq |x_1| + \cdots + |x_n|$.
* l_∞-norm (Chebyshev): $\|\underline{x}\|_\infty = \max\{|x_1|, \cdots, |x_n|\}$
* l_p-norm: $\|\underline{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}$. $\|\underline{x}\|_\infty = \lim_{p\to\infty} \|\underline{x}\|_p$ (btw).
* Quadratic norm: For any positive semdef. matrix $\underline{P}$, $\in \mathbb{R}^{n\times n}$.

$$\|\underline{x}\|_{\underline{P}} = (\underline{x}^T \underline{P}\, \underline{x})^{\frac{1}{2}} = \|\underline{P}^{\frac{1}{2}} \underline{x}\|_2$$

$\underline{x}^T \underline{P}\, \underline{x} \geq 0$, $\forall \underline{x} \in \mathbb{R}^n$

4. Equivalence of Norms:

Suppose $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\mathbb{R}^n$. 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: $\|\underline{x}\|_2 \le \|\underline{x}\|_1 \le \sqrt{n}\, \|\underline{x}\|_2$.

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

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

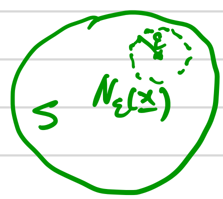
B. Sets and Sequences.

1. ε-neighborhood about a pt. $\underline{x}_0$

$N_\varepsilon(\underline{x}_0) = \{\underline{x} : \|\underline{x} - \underline{x}_0\| \underset{\le}{<} \varepsilon\}$.

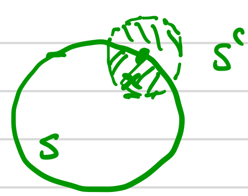
2. Interior of S:

$\text{int}(S) = \{\underline{x} \in S : \exists\, \varepsilon > 0, \text{ s.t. } N_\varepsilon(\underline{x}) \subseteq S\}$.



3. Boundary of S:

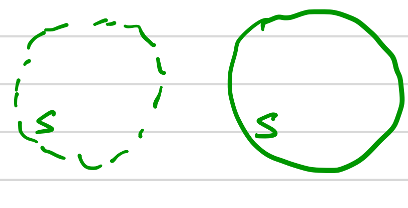
$\partial(S) = \{\underline{x} : \forall \varepsilon > 0,\ N_\varepsilon(\underline{x}) \cap S \ne \phi,\ N_\varepsilon(\underline{x}) \cap S^c \ne \phi\}$.

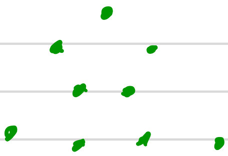


4. Open and Closed Sets.

Open set S: $S \cap \partial(S) = \phi \Leftrightarrow S = \text{int}(S)$.

closed set S: $\partial(S) \subseteq S \Leftrightarrow \{S^c \text{ is open}\}$.





closed b/c every pt. is bndy pt.

$S=\mathbb{R}^n$: both open and closed.

1. $S^c=\phi$, $\partial(S^c)=\phi=\partial(S)$, $S\cap\partial(S)=S\cap\phi=S\cap S^c=\phi$. Open.
2. $\partial(S)=\phi\subseteq S \Rightarrow S$ is closed.
   (empty set is a subset of any set).

open or closed? neither.

5. Closure of S: $cl(S)=\partial(S)\cup S$. (smallest closed set that contains S).
6. S is bounded if it can be contained within a ball of finite radius.
7. S is compact if it's closed and bound.

closed but unbnded

8. Convergent Sequence and Limits.

1° Def (Convergence): A seq. of vectors $\underline{x}_1, \underline{x}_2, \cdots$ are said to be convergent to a limit pt. $\bar{\underline{x}}$ if $\forall \varepsilon>0$, $\exists N_\varepsilon\in\mathbb{N}$ s.t. $\|\underline{x}_k-\bar{\underline{x}}\|<\varepsilon$, $\forall k\geq N_\varepsilon$. ($\{\underline{x}_k\}\to\bar{\underline{x}}$ as $k\to\infty$. $\lim\limits_{k\to\infty}\underline{x}_k=\bar{\underline{x}}$).

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

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

Ex: (p-series). $a_n = \frac{1}{n^2}$. Show $\{b_n\} = \left\{\sum_{k=1}^{n} a_k\right\}$ $(p=2)$ has a limit.

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

$$b_n - b_m = \sum_{k=1}^{n} \frac{1}{k^2} - \sum_{k=1}^{m} \frac{1}{k^2} = \sum_{k=m+1}^{n} \frac{1}{k^2} < \sum_{k=m+1}^{n} \frac{1}{k(k-1)}$$

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

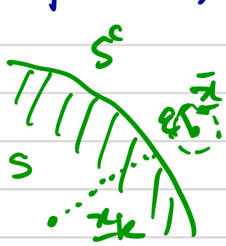
$$= \frac{1}{m} - \frac{1}{n} < \frac{1}{m} < \varepsilon,$$

can always find suff. large m s.t. $b_n - b_m < \varepsilon$. $\blacksquare$

9. Closedness & Compactness characterized convergent seq. & limits.

Thm: A set S is closed iff for any seq. $\{\underline{x}_k\} \to \bar{\underline{x}}$, s.t. $\underline{x}_k \in S$, we also have $\bar{\underline{x}} \in S$.

Proof: $(\Rightarrow)$ By contradiction: Suppose not: $\exists$ a $\{\underline{x}_k\} \to \bar{\underline{x}}$, $\underline{x}_k \in S, \forall k$



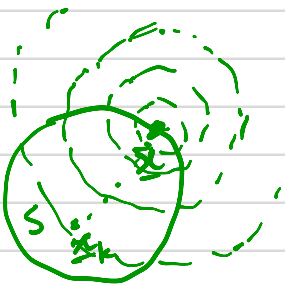
but $\bar{\underline{x}} \notin S$.

S closed $\Rightarrow$ S^c open $\Rightarrow$ $S^c = \text{int}(S^c)$. $(*)$

Since $\bar{\underline{x}} \in S^c \overset{(*)}{\Rightarrow} \bar{\underline{x}} \in \text{int}(S^c) \Rightarrow \exists N_\varepsilon(\bar{\underline{x}}) \subseteq S^c$

$\rightarrow\leftarrow$ convergence assumption.

($\Leftarrow$) By contra: If $S \neq$ closed: $\exists\, \bar{x} \in \partial(S)$, but $\bar{x} \notin S$.

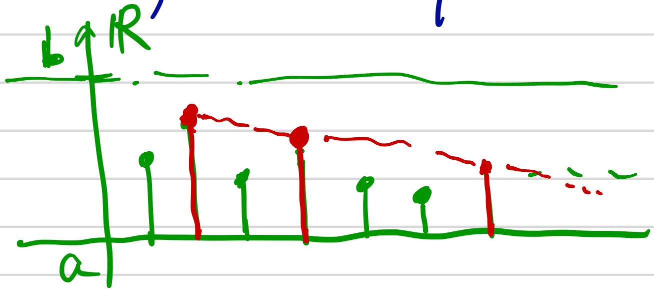


keep shrinking $\varepsilon \Rightarrow$ create a seq. $\{x_k\} \to \bar{x}$

$x_k \in S, \forall k \Rightarrow \bar{x} \in S$

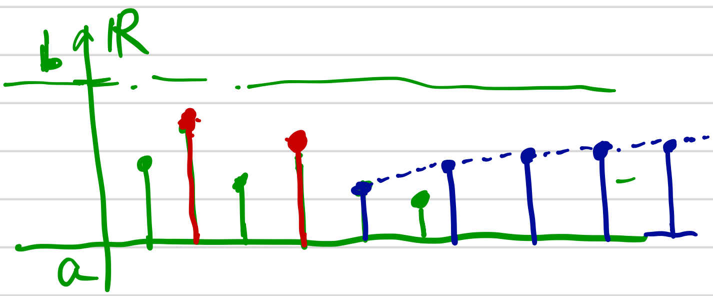
$\blacksquare$

<u>Thm</u> (Bolzano–Weierstrass): Every bnd seq. in $\mathbb{R}^n$ has a convergent subseq.



1. enlighted terms are infinite

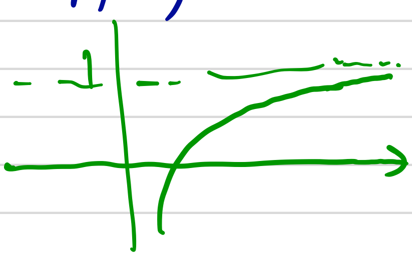
MCT: if $\{a_n\}$ is a mono seq. reals, then $\{a_n\}$ has a limit iff $\{a_n\}$ is bnded.



2. enlighted terms are finite.

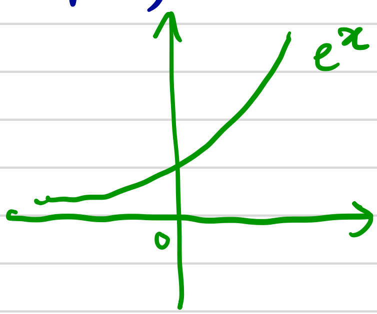
$\blacksquare$

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



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

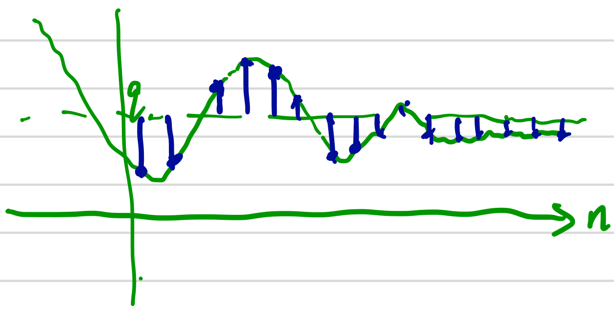
Infimum of S (Largest LB): Largest possible value α satisfying $\alpha \le x,\ \forall x \in S$.



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

Maximum, minimum (achievable).

* The limit superior $\limsup_{k\to\infty} x_k$ is the infimum of all $q \in \mathbb{R}$ for which all but a finite # of elements in $\{x_k\}$ exceed q. $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} \left\{ \sup_{m \ge n} x_m \right\}$ (HW).



$\{x_n\}$ $\{a_n\}$

$n_1 > n_2$ $\{x_{n_1}\} \subset \{x_{n_2}\}$

* The limit infimum $\liminf_{k\to\infty} x_k$ is the supremum of all $q \in \mathbb{R}$ for which all but a finite # of elements in $\{x_k\}$ less than q. $\liminf_{n\to\infty} x_n = \lim_{n\to\infty} \left\{ \inf_{m \ge n} x_m \right\}$.

* limsup and liminf always exist.

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

3. Functions.

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

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

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

2° Diff'ble fn:

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

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

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

$\beta(\underline{x}, \bar{\underline{x}}) \to 0$ as $\underline{x} \to \bar{\underline{x}}$, such that

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

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

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

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

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

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:

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

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

$$f(\underline{x}_2) = f(\underline{x}_1) + \nabla f(\underline{x})^T (\underline{x}_2 - \underline{x}_1), \text{ where } \underline{x} = \lambda \underline{x}_1 + (1-\lambda) \underline{x}_2$$

for some $\lambda \in (0,1)$.

5° Taylor's Thm: S non-empty, open, convex in $\mathbb{R}^n$.

$f: S \to \mathbb{R}$. twice diff'ble. For every $\underline{x}_1, \underline{x}_2 \in S$, we have,

$f(\underline{x}_2) = f(\underline{x}_1) + \nabla f(\underline{x}_1)^T(\underline{x}_2 - \underline{x}_1) + \frac{1}{2}(\underline{x}_2 - \underline{x}_1)^T \underline{\underline{H}}(\underline{x})(\underline{x}_2 - \underline{x}_1)$, where

$\underline{\underline{H}}(\underline{x})$ 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

$$\sum_{i=1}^{k} \lambda_i \underline{x}_i = \underline{0} \Rightarrow \lambda_i = 0, \ \forall i = 1, \cdots, k.$$

2. linear comb: $y \in \mathbb{R}^n$ is lin. comb. of $\underline{x}_1 \cdots \underline{x}_k \in \mathbb{R}^n$ if

$$y = \sum_{i=1}^{k} \lambda_i \underline{x}_i \text{ for some } \lambda_1 \cdots \lambda_k.$$

* $\sum_{i=1}^{k} \lambda_i = 1$: y is an <u>affine comb.</u> of $\underline{x}_1, \cdots, \underline{x}_k$.

* $\sum_{i=1}^{k} \lambda_i = 1$, $\lambda_i \geq 0$, $\forall i$: y is a <u>convex comb.</u> of $\underline{x}_1, \cdots, \underline{x}_k$.

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 : $\underline{x}_1, \cdots, \underline{x}_k \in \mathbb{R}^n$, $k \geq n$, said to be spanning $\mathbb{R}^n$ if <u>any</u> vector in $\mathbb{R}^n$ can be represented as a <u>lin.</u> <u>comb.</u> of $\underline{x}_1, \cdots, \underline{x}_k$.

The <u>cone</u> spanned by $\underline{x}_1 \cdots, \underline{x}_k$ is set of non-neg. lin. comb.

4. <u>Basis</u>: A set of $\underline{x}_1 \cdots \underline{x}_k \in \mathbb{R}^n$ spans $\mathbb{R}^n$ and if the deletion of any of $\underline{x}_1 \cdots \underline{x}_k$ prevents remaining vector from spanning $\mathbb{R}^n$ (Basis $\underline{x}_1, \cdots, \underline{x}_k$ spans $\mathbb{R}^n$ iff $k = n$).

5. Cauchy – Schwatz Ineq: $|\langle \underline{x}, \underline{y} \rangle| = |\underline{x}^T \underline{y}| \leq \|\underline{x}\|_2 \cdot \|\underline{y}\|_2$.

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

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

($\underline{x}$ & $\underline{y}$ are orthogonal, ($\underline{x} \perp \underline{y}$), if $\langle \underline{x}, \underline{y} \rangle = 0$).

6. Orthogonal matrix: $\underline{\underline{Q}} \in \mathbb{R}^{m \times n}$: $\underline{\underline{Q}}^T \underline{\underline{Q}} = \underline{\underline{I}}_n$ or $\underline{\underline{Q}}\,\underline{\underline{Q}}^T = \underline{\underline{I}}_m$

If $\underline{\underline{Q}}$ is square : $\underline{\underline{Q}}^{-1} = \underline{\underline{Q}}^T$.

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

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

8. Eigenvalues and eigenvectors : $\underset{=}{A} \in \mathbb{R}^{n\times n}$. If λ and $\underline{x} \neq \underline{0}$ satisfy $\underset{=}{A}\,\overset{\alpha}{\underline{x}} = \overset{\alpha}{\lambda}\,\underline{x}$ $\underline{x}' = \alpha \underline{x}$, then λ, $\underline{x}$ are eigenvalues & eigenvectors.

* λ can computed by solving $\det(\underset{=}{A} - \lambda \underset{=}{I}) = 0$ (characteristic eqn.)

* $\underset{=}{A}$ is symmetric $\Rightarrow$ n (possibly non-distinct) real eigenvalues

* Eigenvectors assoc. w/ distinct eigenvalues are orthogonal.

* Given symmetrix $\underset{=}{A}$ $\Rightarrow$ can construct an orthogonal basis $\underset{=}{B} \in \mathbb{R}^{n\times n}$ where each col in $\underset{=}{B}$ is an eigenvector of $\underset{=}{A}$.

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

* Let $\lambda_1 \cdots \lambda_n$ be eigenvalues of symmetric $\underset{=}{A}$. Let $\underset{=}{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

Note $\underset{=}{A}\,\underset{=}{B} = \underset{=}{B}\,\underset{=}{\Lambda}$

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

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

(eigenvalue decomp).

10. 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 $i \neq j$.

$\underbrace{(\underline{\underline{\Sigma}})_{ij} \geq 0}_{\in \mathbb{R}}$, for $i=j$.

* Cols of $\underline{\underline{U}}$: Normalized eigenvectors of $\underline{\underline{A}}\,\underline{\underline{A}}^T$.
* Cols of $\underline{\underline{V}}$: – – – – – – – – – – of $\underline{\underline{A}}^T\underline{\underline{A}}$
* $(\underline{\underline{\Sigma}})_{ij}$, $i=j$: Are 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.

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

	PD		pos.	
$\underline{\underline{A}}$ is	PSD	if eigenvalues are	non-neg.	, resp.
	ND		neg.	
	NSD		non-pos.	

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

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

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