Two books

- Numerial optimization by Nocedal & Wright
- Convex Optimization by Boyd & Vendenberghe

Grading

- 40 % Homework (4, coding)
- 30 % Midterm exam
- 30 % Final exam

Topics:

- Gradient descent
- Newton method
- Conjugate gradient method ····· $Ax=b$?
- Stochastic gradient descent
- Interior point method ··· constrained optimization
- Coordinate descent
- Linear programming ····
- ADMM (Augmented Lagrangian Methods).

0. Basic Linear Algebra

0.1 Vectors

$x \in R^d$, $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$, $x^T = [x_1, x_2 \cdots x_d]$

- vector norms:

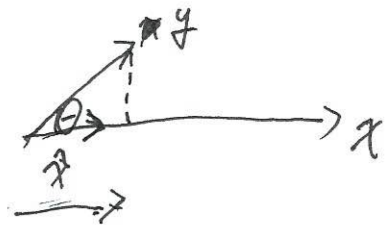
$\|x\|_2 = \sqrt{x_1^2 + \cdots + x_d^2}$ .... L_2-norm

$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_d|^p)^{1/p}$ .... L_p-norm

$\|x\|_1 = (|x_1| + |x_2| + \cdots + |x_d|)$ .... L_1-norm

- Given two vectors $x, y \in R^d$

inner product $x^T y = \sum_i x_i y_i$



unit vector: $\hat{x} = \frac{x}{\|x\|}$

$(\hat{x}^T y)\hat{x}$: projection of y on x

$\qquad \|$

$(\hat{x}\hat{x}^T) y \qquad \begin{array}{cc} & d \\ d & \left[\ \hat{x}\hat{x}^T\ \right] \end{array} y$

$\uparrow$

projection matrix

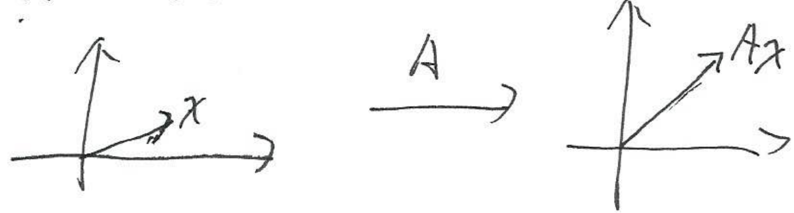
$\cos\theta = \frac{\hat{x}^T y}{\|y\|} = \frac{x^T y}{\|x\| \|y\|}$

$x \perp y \Leftrightarrow x^T y = 0$

0.2 Matrix

$A \in R^{m \times n}$ $A = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ A_{21} & \ddots & \vdots \\ \vdots & & \\ A_{m1} & \cdots & A_{mn} \end{bmatrix}$ — elementwise

$x \longmapsto Ax$



... operation view

Norm of matrix

- $\|A\|_F = \sqrt{\sum_{ij} A_{ij}^2}$
- Operation norm (induced norm)

$$\|A\|_p = \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sigma_1$$

0-3. Eigenvalue Decomposition

$A \in R^{n \times n}$

Def: If $Ax = \lambda x$, λ is an eigenvalue
x is an eigenvector

$\Longleftrightarrow (A - \lambda I)x = 0$

λ is an eigenvalue $\Longleftrightarrow A - \lambda I$ is singular

$\Longleftrightarrow \boxed{\det(A - \lambda I) = 0}$

degree-n polynomial on λ

A is symmetric $\Longleftrightarrow A = A^T$

$A = A^T$ and $A_{ij} \in R$

✓ $\Rightarrow$ All eigenvalues are $\boxed{\text{real}}$

✓ $\Rightarrow$ Eigenvectors corresponding to different eigenvalues are orthogonal

$\lambda_1 \neq \lambda_2$

$Av_1 = \lambda_1 v_1$, $Av_2 = \lambda_2 v_2$ $\Rightarrow$ $v_1 \perp v_2$ or $v_1^T v_2 = 0$

$\Rightarrow \exists\, v_1, \cdots v_n \in R^n,\ \lambda_1, \cdots \lambda_n \in R$

$$\left\{\begin{array}{l} Av_1 = v_1\lambda_1 \\ Av_2 = v_2\lambda_2 \\ \vdots \\ Av_n = v_n\lambda_n \end{array}\right\} \qquad \boxed{v_i^T v_j = 0} \quad v_i^T v_i = 1$$

$\Downarrow$

$$\Rightarrow A\underbrace{[v_1, v_2, \cdots, v_n]}_{V} = \underbrace{[v_1, v_2, \cdots, v_n]}_{V}\underbrace{\begin{bmatrix}\lambda_1 & & 0\\ & \lambda_2 & \\ 0 & & \lambda_n\end{bmatrix}}_{\Lambda}$$

$\rightarrow AV = V\Lambda$ or $\boxed{A = V\Lambda V^T}$, $V^TV = I$

eigen decomposition

$x \mapsto Ax$

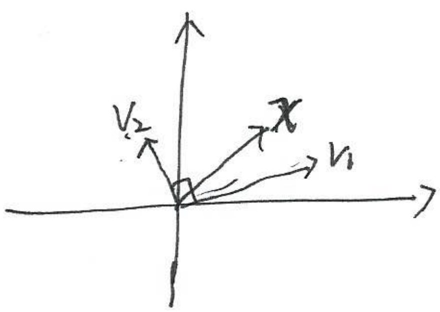
$v_1 \mapsto \lambda_1 v_1$

$v_2 \mapsto \lambda_2 v_2$

$\boxed{Ax} = V\Lambda V^T x$

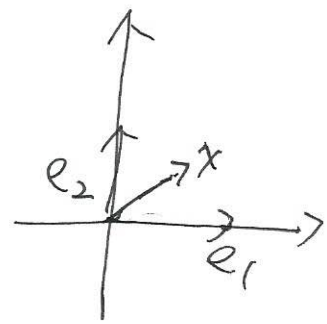
$n=2$

rotation



$\xrightarrow{V^T x}$

$v_1 \to e_1$
$v_2 \to e_2$
$x \to V^T x$

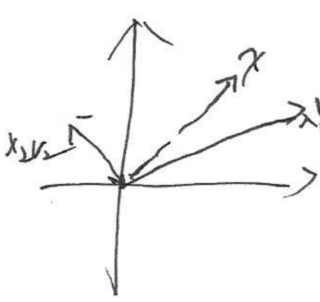


$v_1 \to \lambda_1 e_1$
$v_2 \to \lambda_2 e_2$
$x \to \Lambda(V^T x)$

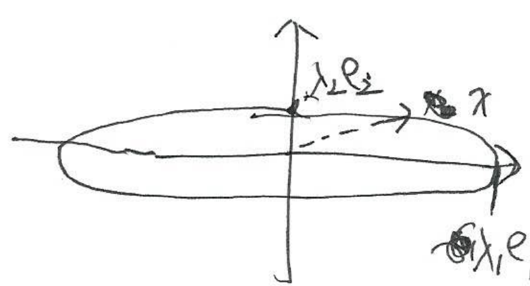
$\Lambda(V^T x)$

$\begin{bmatrix}100 & \\ & 1\end{bmatrix}$ scaling

$v_1 \to \lambda_1 v_1$
$v_2 \to \lambda_2 v_2$



$\xleftarrow{V(\Lambda V^T x)}$



Def : A is <u>postive semidefinite</u> ($A \succeq 0$)

$\Leftrightarrow \forall x,\ x^T A x \geq 0$

$\Leftrightarrow \lambda_i \geq 0 \quad \forall i = 1, \cdots, n$

$A = V \Lambda V^T$

$x^T A x = x^T V \Lambda V^T x$

$= \underbrace{\quad}_{} \begin{bmatrix} \ddots \end{bmatrix} \begin{bmatrix} \\ \\ \end{bmatrix}$

$V^T x$

~~zero~~ $\lambda_i = 0, \quad A v_i = \lambda_i v_i = 0$

$v_i^T A v_i = 0$

Def : A is <u>positive definite</u> ($A \succ 0$)

$\Leftrightarrow \forall x \neq 0,\ x^T A x > 0$

$\Leftrightarrow \lambda_i > 0 \quad \forall i = 1, \cdots, n$

0.4 Singular Value Decomposition (SVD)

$A \in R^{m \times n}$ $\quad A = m \overset{n}{\boxed{\ }}$

SVD of A : ($m \geq n$)

singular value

$$m \begin{bmatrix} \\ & \overset{n}{A} & \\ \\ \end{bmatrix} = m \begin{bmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \sigma_2 \ddots & \\ 0 & & \sigma_n \end{bmatrix} \overbrace{\begin{bmatrix} \cdots & v_1^T & \cdots \\ & v_2^T & \\ & \vdots & \\ & v_n^T & \end{bmatrix}}^{n}$$

$U \quad , \quad \Sigma \quad \cdot \quad V^T$

$U^T U = I, \ V^T V = I$

σ_i is a singular value

u_i is a left singular vector

v_i is a right singular vector

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$

Thm : SVD exists for any $m \times n$ matrices.

singular values are unique

physial meanings :

I. $A = U\Sigma V^T$ $\qquad A\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix}\begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix}$

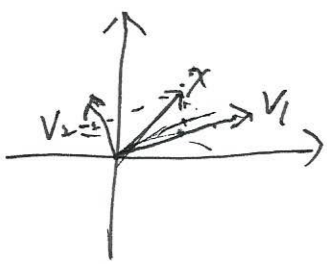
$\Rightarrow AV = U\Sigma$

$\Rightarrow Av_i = u_i \cdot \sigma_i \quad \forall i = 1, \cdots, n$

$v_i \xmapsto{A} \sigma_i u_i$ $\qquad$ (eigendecomp. $v_i \xrightarrow{A} \lambda_i v_i$)

II. $x \longmapsto Ax$ $\qquad m = n = 2$

$x \longmapsto U\Sigma V^T x$

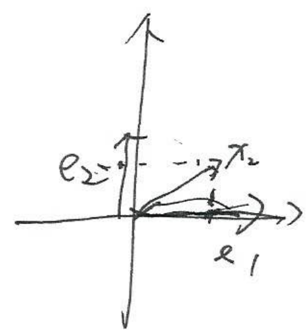


$V^T x$ rotation

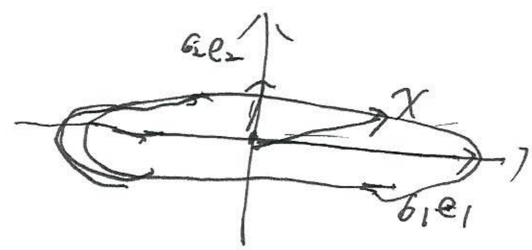
$v_1 \to e_1$

$v_2 \to e_2$

$\begin{bmatrix} v_1^T \\ v_2^T \end{bmatrix} x = \begin{bmatrix} v_1^T x \\ v_2^T x \end{bmatrix}$

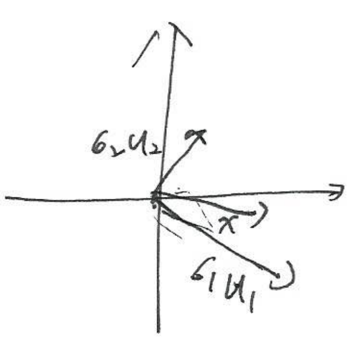


$\Sigma(V^T x)$ scaling



$U(\Sigma V^T x)$

rotation



$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_1$$

III. If A is low rank

A is rank r, $r < \min(m, n)$

$\sigma_1 \geq \sigma_2 \cdots \geq \sigma_r > 0$

$\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$

$$A = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} \sigma_1 \cdots \sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix}$$

$$\begin{cases} v_1 \xrightarrow{A} u_1 \sigma_1 \\ v_2 \xrightarrow{A} u_2 \sigma_2 \\ \vdots \\ v_r \xrightarrow{A} u_r \sigma_2 \\ v_{r+1} \xrightarrow{A} 0 \\ \vdots \\ v_n \xrightarrow{A} 0 \end{cases}$$

What's the nullspace of A.

$\{x \mid Ax = 0\} = \text{span}\langle v_{r+1} \cdots v_n \rangle$

The range space (column space) of A

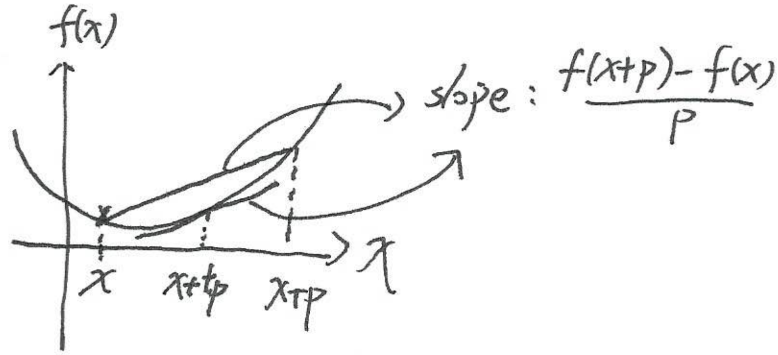
$\{y \mid y = Ax\} = \text{Span}\langle u_1 \cdots u_r \rangle$

0.5 Taylor expansion.

Thm: $f: \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, then

(Mean Value Theorem) $\quad f(x+p) = f(x) + \nabla f(x+tp)^T p$

for some $t \in (0,1)$



Thm: If f is twice continuous differentiable, then

$$\nabla_i f(x+p) = \nabla_i f(x) + (\nabla^2 f(x+tp) \cdot p)_i$$

for some $t \in [0,1)$, $\forall i$

Thm: If f is twice continuous differentiable, then

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p$$

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

Thm: If f is twice differentiable and

$MI \succeq \nabla^2 f(x) \succeq mI \quad \forall x \in \mathrm{dom}(f)$, then

$$f(x+p) \geq f(x) + \nabla f(x)^T p + \frac{1}{2} m \|p\|_2^2$$

and $f(x+p) \leq f(x) + \nabla f(x)^T p + \frac{1}{2} M \|p\|_2^2$.