Midterm next week!

Recap

SVD : $A = \sum_{i=1}^{r} \sigma_i w_i v_i^*$

Rank-k approximation : $A_k = \sum_{i=1}^{k} \sigma_i w_i v_i^*$

$\forall B$, rank$(B) \le k$. $\|A-B\|_2 \geq \sigma_{k+1} = \|A - A_k\|_2$

Least squares approximation : $a_1 \dots a_n \in \mathbb{R}^d$, $\dim(S) \le k$
(PCA)

$$\sum_{i=1}^{n} \mathrm{dist}(a_i, S) \geq \sigma_1^2 + \dots + \sigma_k^2$$

$$A = \begin{bmatrix} — & a_1^T & — \\ & \vdots & \\ — & a_n^T & — \end{bmatrix}$$

Opt = Span of top k right singular vectors

Gershgorin Disc Theorem

Given $M \in \mathbb{C}^{n \times n}$

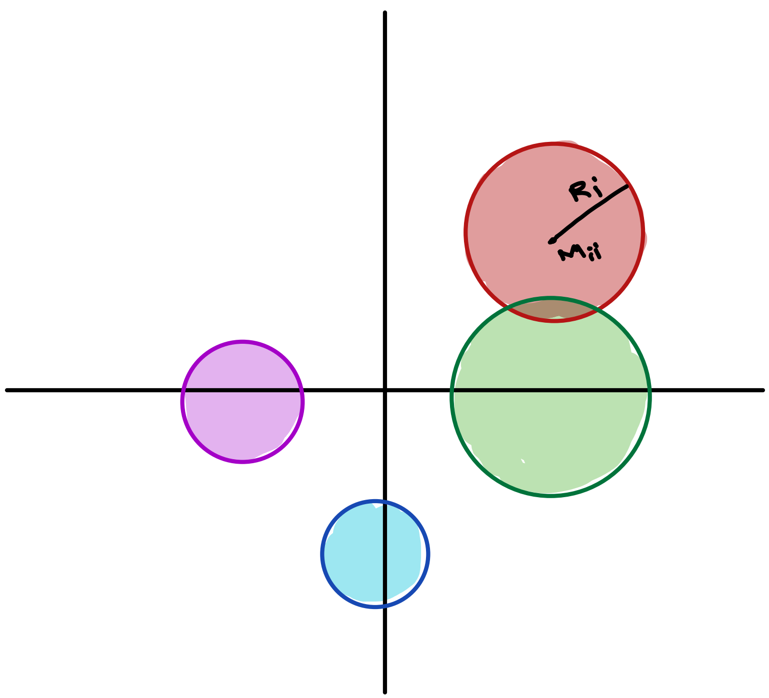
$$M = \begin{pmatrix} a_1 & & & \\ & a_2 & & 0 \\ & & \ddots & \\ 0 & & & a_n \end{pmatrix}$$

eigenvalues of M are $a_1, \ldots\ldots, a_n$

What about general $M \in \mathbb{C}^{n \times n}$

- Let $M \in \mathbb{C}^{n\times n}$. Let $R_i = \sum_{j\neq i} |M_{ij}|$ and $\mathrm{Disc}(M_{ii}, R_i) = \{z \mid |z - M_{ii}| < R_i\}$

Then, for any eigenvalue λ of M, $\lambda \in \bigcup_i \mathrm{Disc}(M_{ii}, R_i)$



Proof: Let x be an eigenvector with eigenvalue $\lambda \in \mathbb{C}$.
Let $i_0 = \underset{i\in[n]}{\mathrm{argmax}}\ |x_i|$

$$Mx = \lambda x \Rightarrow (Mx)_{i_0} = \lambda x_{i_0}$$

$$\Rightarrow \sum M_{i_0 j}\, x_j = \lambda x_{i_0}$$

$$\Rightarrow M_{i_0 i_0} + \sum_{j\neq i_0} M_{i_0 j} \frac{x_j}{x_{i_0}} = \lambda$$

$$\therefore |\lambda - M_{i_0 i_0}| \leq \sum_{j\neq i_0} |M_{i_0 j}| \cdot \overbrace{\frac{|x_j|}{|x_{i_0}|}}^{\leq 1}$$

$$\leq R_{i_0}$$

An application to Compressed Sensing

- $A \in \mathbb{R}^{k \times n}$ is said to have restricted isometry property with parameters (s, δ) if

$$(1-\delta)\|x\|^2 \leq \|Ax\|^2 \leq (1+\delta)\|x\|^2$$

for all s-sparse vectors x s.t. $|\{i \mid x_i \neq 0\}| \leq s$

useful condition. NP-hard to check.

- $A \in \mathbb{R}^{k \times n}$

$$A = \begin{bmatrix} | & & & | \\ a_1 & \cdots & \cdots & a_n \\ | & & & | \end{bmatrix} \quad \text{with } \|a_i\| = 1 \quad \forall i$$

Coherence $\mu(A) = \max_{i \neq j} |\langle a_i, a_j \rangle|$

Incoherence implies restricted isometry

- For $A \in \mathbb{R}^{k\times n}$ as before with coherence $\mu(A)$.

A has restricted isometry property with parameters $(s, (s-1)\cdot\mu(A))$ i.e. $\forall x$ s.t. $|\{i \mid x_i \neq 0\}| \le s$

$$(1-(s-1)\cdot\mu(A))\cdot\|x\|^2 \le \|Ax\|^2 \le (1+\underbrace{(s-1)\cdot\mu(A)}_{\delta})\cdot\|x\|^2$$

<u>Proof:</u> Say $s=1$ $\quad \|Ax\|^2 = x_i^2 = \|x\|^2$

If $S=\{i \mid x_i\neq 0\}$ $\quad \|Ax\|^2 = \|A_S x_S\|^2$ only keep columns for $i\in S$

$$= \langle x_S, A_S^* A_S x_S\rangle$$

$|\langle a_i, a_j\rangle| \le \mu(A)$

$$A_S^*A_S = \begin{bmatrix}1 & & \\ & \ddots & \\ & & 1\end{bmatrix}$$

$\therefore$ by Gershgorin, $|\lambda-1| \le (s-1)\mu(A) = \delta$ $\forall$ eigenvalues λ

$$\therefore \frac{\langle x_S, A_S^*A_S x_S\rangle}{\|x_S\|^2} \in [1-\delta, 1+\delta]$$

Solving linear equations

$$Ax = b$$

$$[A \mid b]$$

$$\begin{aligned} x + y + z &= 10 \\ x + 2y + 3z &= 15 \\ 2x + 3y + z &= 19 \end{aligned}$$

$$\begin{bmatrix} 1 & 1 & 1 & 10 \\ 1 & 2 & 3 & 15 \\ 2 & 3 & 1 & 19 \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} 1 & 1 & 1 & 10 \\ & 1 & 2 & 5 \\ & 1 & -1 & -1 \end{bmatrix}$$

$$\downarrow$$

$$\vdots$$

$$\begin{aligned} x + y + z &= 10 \\ y + 2z &= 5 \\ y - z &= -1 \end{aligned}$$

$x = 7$

$$x + y + z = 10$$

$y = 1$

$$y + 2z = 5$$

$z = 2$

$$-3z = -6$$

Row-reduced form

$$M = \begin{bmatrix} 1 & 1 & 1 & 10 \\ 0 & 1 & 2 & 15 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

- First non-zero entry of each row is 1. (leading entry)
- If M_{ij} is leading entry then M is 0 "below" & "left" $(M_{i'j'} = 0 \;\forall\; i' \geq i,\; j' \leq j)$
- Rows containing all 0s at bottom.

$$\begin{bmatrix} 1 & 1 & 1 & 10 \\ 1 & 2 & 3 & 15 \\ 2 & 3 & 1 & 19 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 & 10 \\ 0 & 1 & 2 & 5 \\ 0 & 1 & -1 & -1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 & 10 \\ 0 & 1 & 2 & 5 \\ 0 & 0 & -3 & -6 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 & 1 & 10 \\ 0 & 1 & 2 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

Elementary row-operations:

- $M_i \leftarrow M_i + c \cdot M_j$
- $M_i \leftarrow c \cdot M_i \quad (c \neq 0)$
- Swap M_i, M_j

Properties

Ex: Matrix in row-reduced form "obviously" has row-rank = column-rank

e.g. $\begin{bmatrix} 1 & 1 & 1 & 10 \\ 0 & 1 & 2 & 15 \\ 0 & 0 & 1 & 2 \end{bmatrix}$

Ex: Elementary row-operations do not change row-rank or column-rank

Ex: Above is true over any field $\mathbb{F}$.

- For any matrix $M \in \mathbb{F}^{m \times n}$, row-rank = column-rank

Solving (sparse) linear systems fast

Given $A_0 x = b_0 \longrightarrow \underbrace{A_0^T A_0}_{A} x = \underbrace{A_0^T b_0}_{b}$

(say s non-zero entries in A_0)

$Ax = b$

$\ker(A_0) = \{0\} \iff A$ is positive definite

$\Updownarrow$ $\longrightarrow$ Ex.

$\exists$ unique solution, say u

Goal: Compute approximate solution fast.

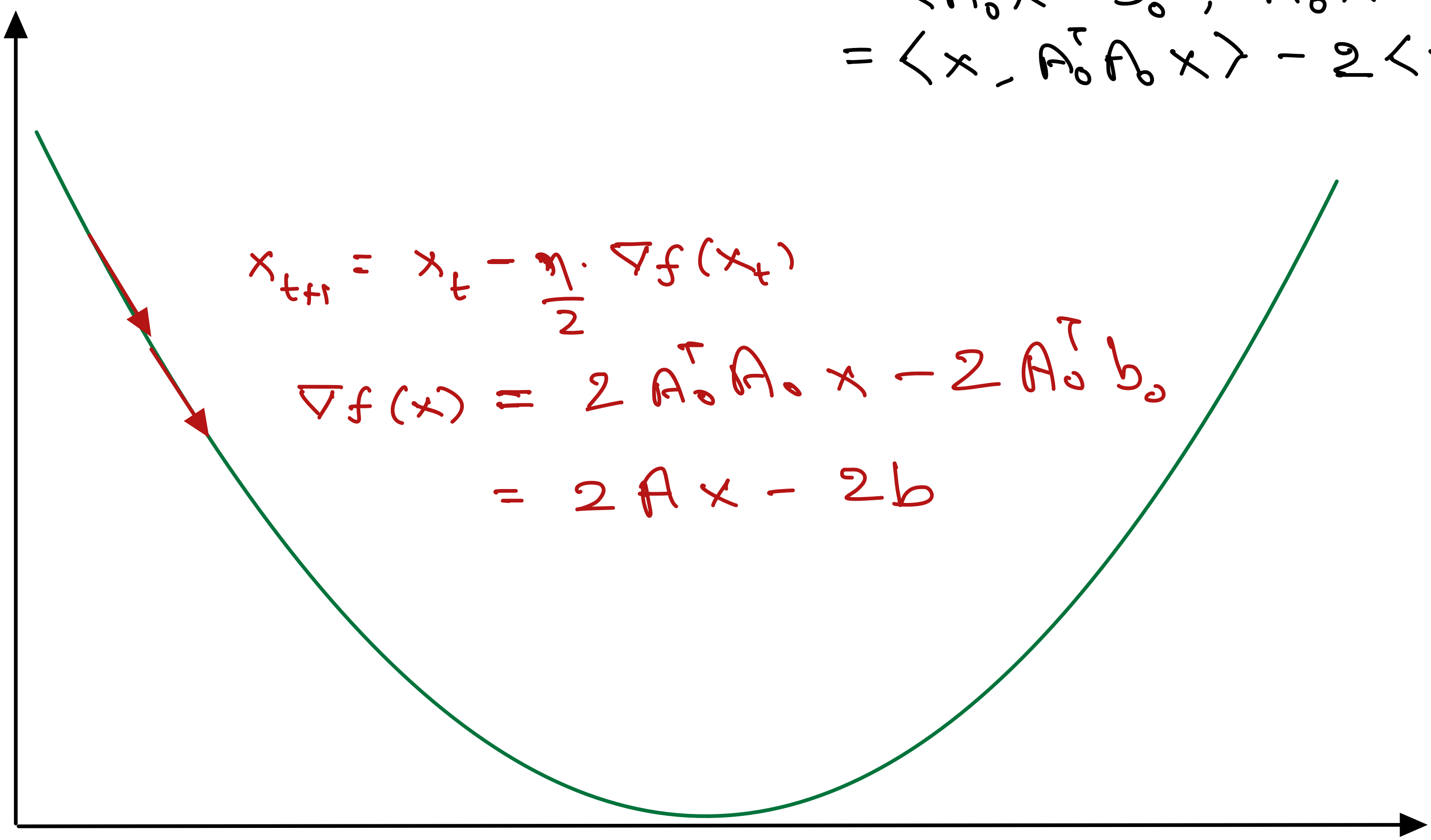
Gold standard: Time to check solution
= matrix-vector product
= ?

Solvers from optimization

Consider: $$\underset{x}{\text{minimize}} \; f(x) = \|A_0 x - b_0\|^2$$

$$= \langle A_0 x - b_0 , A_0 x - b_0 \rangle$$

$$= \langle x , A_0^T A_0 x \rangle - 2 \langle x, A_0^T b_0 \rangle + \|b_0\|^2$$



$$x_{t+1} = x_t - \frac{\eta}{2} \cdot \nabla f(x_t)$$

$$\nabla f(x) = 2 A_0^T A_0 x - 2 A_0^T b_0$$

$$= 2Ax - 2b$$

$$x_{t+1} = x_t - \eta (A x_t - b)$$

Steepest descent iteration

- Let u be the solution to $Ax=b$. Then

$$(x_t - u) = (I - \eta A)^t (x_0 - u)$$

Proof:

$$x_{t+1} = x_t - \eta (Ax_t - Au)$$

$$x_{t+1} - u = x_t - u - \eta A (x_t - u)$$

$$x_{t+1} - u = (I - \eta A)(x_t - u)$$

Convergence analysis : choosing η

$$(x_t - u) = (I - \eta A)^t \cdot (x_0 - u) \Rightarrow \|x_t - u\| \leq \|(I - \eta A)^t\|_2 \cdot \|x_0 - u\|$$

Eigenvalues

A : $\lambda_1 \geq \cdots \geq \lambda_n > 0$ (say)

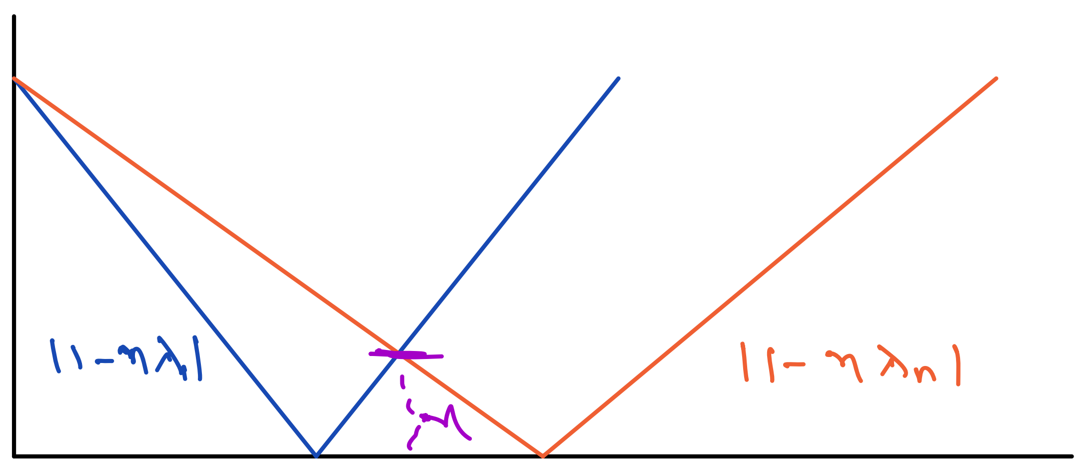
$I - \eta A$: $1 - \eta\lambda_1 \leq \cdots \leq 1 - \eta\lambda_n < 1$

$(I - \eta A)^t$: $(1 - \eta\lambda_1)^t \cdots (1 - \eta\lambda_n)^t$

$$\|(I - \eta A)^t\|_2 = \max\{|1 - \eta\lambda_1|^t, |1 - \eta\lambda_n|^t\}$$

$\eta\lambda_1 - 1 = 1 - \eta\lambda_n$

$$\eta = \frac{2}{\lambda_1 + \lambda_n}$$



Convergence analysis

$$\|(I - \eta A)^t\|_2 = (1 - \eta\lambda_n)^t = \left(1 - \frac{2\lambda_n}{\lambda_1 + \lambda_n}\right)^t = \left(1 - \frac{2}{\underbrace{(\lambda_1/\lambda_n)}_{\kappa} + 1}\right)^t$$

Condition number

$$\|(I - \eta A)^t\|_2 \le \epsilon \quad \text{after} \quad t = O\left(\kappa \log \frac{1}{\epsilon}\right)$$

Another distance: $\|x_t - u\|_A = \sqrt{\underbrace{(x_t - u)^T A (x_t - u)}}$

inner product?

Ex: Show that same convergence analysis also works for $\|x - u\|_A$