Recap:

INPUT: $X \in \mathbb{R}^{n \times d}$

$(n > d)$

OUTPUT: v, that $\max_{\|v\|=1} \|Xv\|$

POWER ITERATION

- v_0 is a random unit vector
- for $t = 1, \dots, T$:
  - $u = X^T \cdot (X \cdot v_{t-1})$
  - $v_t = \dfrac{u}{\|u\|}$

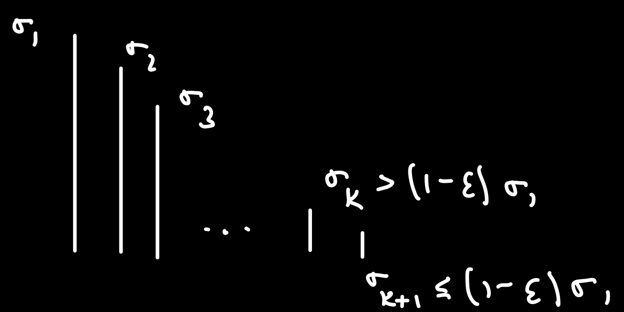
Theorem:

PI will converge to a top right s.v.

After $T = O\left(\frac{\log(d/\varepsilon)}{\varepsilon}\right)$ iterations, $\|X \cdot v_T\| \geq (1-\varepsilon)\sigma_1$.

$X = U \Sigma V^T$ be its SVD

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_d$$



Let k be the index such that

$$\sigma_k > (1-\varepsilon)\sigma_1$$

$$\sigma_{k+1} \leq (1-\varepsilon)\sigma_1$$

Let S_k = span of the first k singular vectors $v^{(1)}, v^{(2)}, \dots, v^{(k)}$

THEOREM:

After $T = O\left(\frac{\log(d/\delta)}{\varepsilon}\right)$ iterations

v_T is almost entirely within S_k

$$\| v_T - \text{Proj}_{S_k}(v_T) \| \leq \delta$$

ε such that:

$$\sigma_k > (1-\varepsilon)\sigma_1$$

$$\sigma_{k+1} \leq (1-\varepsilon)\sigma_1$$

THEOREM:

After T iterations

$$\| v_T - \text{Proj}_{S_k}(v_T) \| \leq \frac{(1-\varepsilon)^{2T}}{|\langle v_0, v^{(1)} \rangle|}$$

first right singular vector

Remark:

If $\sigma_2 \leq \sigma_1/2$

then $\| v_T - \text{Proj}_{S_1}(v_T) \| \leq \frac{\left(\frac{1}{2}\right)^{2T}}{|\langle v_0, v^{(1)} \rangle|}$

Corollary:

Suppose $\sigma_2 < (1-\varepsilon)\sigma_1$, then

$$\| v_T - \mathrm{Proj}_{S_1}(v_T) \| \leq \frac{(1-\varepsilon)^{2T}}{|\langle v_0, v^{(1)} \rangle|}$$

Proof:

$$v_1 = \frac{X^T \cdot (X \cdot (v_0))}{\| X^T \cdot (X \cdot (v_0)) \|}$$

$$v_2 = \frac{(X^T \cdot (X \cdot v_1))}{\| X^T \cdot (X \cdot v_1) \|}$$

$$\vdots$$

$$v_t = \frac{(X^T \cdot X)^t \cdot v_0}{\| (X^T \cdot X)^t \cdot v_0 \|}$$

$$X = \sum_{i=1}^{d} \sigma_i \, u^{(i)} v^{(i)T}$$

$$X^T X = \sum_{i=1}^{d} \sigma_i^2 \cdot v^{(i)} \cdot v^{(i)T}$$

$$(X^T X)^t = \sum_{i=1}^{d} \sigma_i^{2t} \cdot v^{(i)} \cdot v^{(i)T}$$

$$v_t = \frac{(X^T X)^t \cdot v_0}{\|(X^T X)^t \cdot v_0\|}$$

$$\begin{array}{|c|c|c|c|c|} \hline v^1 & \cdots & v^d & v^{d+1} & \cdots \\ \hline \end{array}$$

Form an orthonormal basis

$$v^1, v^2, \ldots, v^d, v^{d+1}, \ldots, v^n$$

$$\rightarrow v_0 = c_1 v^1 + c_2 v^2 + \cdots + c_d v^d + c_{d+1} v^{d+1} + \cdots + c_n \cdot v^n$$

$$(X^T X)^t \cdot v_0 = \left(\sum_{i=1}^{d} \sigma_i^{2t} \cdot v^{(i)} v^{(i)T} \right) \cdot (v_0)$$

$$= \sum_{i=1}^{d} \sigma_i^{2t} \cdot \langle v_0, v^{(i)} \rangle \cdot v^{(i)}$$

$$\mathrm{Proj}_{\{v^{(1)}\}} \left((X^T X)^t \cdot v_0 \right) = \langle (X^T X^t) \cdot v_0 , v^{(1)} \rangle$$

$$= \sigma_1^{2t} \cdot \langle v_0, v^{(1)} \rangle \cdot v^{(1)}$$

$$\| (X^T X)^t \cdot v_0 \|^2 = \sum_{i=1}^{d} \sigma_i^{4t} \cdot \langle v_0, v^{(i)} \rangle^2$$

$$\| v_t - \mathrm{Proj}_{\{v^{(1)}\}} (v_t) \|^2 = \frac{\| \sum_{i=2}^{d} \sigma_i^{2t} \cdot \langle v_0, v^{(i)} \rangle \cdot v^{(i)} \|^2}{\| (X^T X)^t \cdot v_0 \|^2}$$

$$= \frac{\sum_{i=2}^{d} \sigma_i^{4t} \cdot \langle v_0, v^{(i)} \rangle}{\sum_{i=1}^{d} \sigma_i^{4t} \cdot \langle v_0, v^{(i)} \rangle^2}$$

Numerator $\leq \sum_{i=2}^{d} \left(\sigma_1 (1-\varepsilon)\right)^{4t} \cdot \langle v_0, v^{(i)} \rangle^2$

Recall : $\sigma_2 \leq (1-\varepsilon)\sigma_1$

$$\leq \sigma_1^{4t} (1-\varepsilon)^{4t} \cdot \left(\sum_{i=2}^{d} \underbrace{\langle v_0, v^{(i)} \rangle^2} \right)$$

project a vector to orthonormal basis and take their sum of squares $\rightarrow$ length $\rightarrow$ unit

$$\leq \sigma_1^{4t} (1-\varepsilon)^{4t}$$

Denominator $\geq \sigma_1^{4t} \cdot \langle v_0, v^{(1)} \rangle^2$

$$\|v_t - \mathrm{Proj}_{\{v^{(1)}\}}(v_t)\|^2 \leq \frac{\sigma_1^{4t} \cdot (1-\varepsilon)^{4t}}{\sigma_1^{4t} |\langle v_0, v^{(1)} \rangle|^2}$$

$$= \frac{(1-\varepsilon)^{4t}}{\langle v_0, v^{(1)} \rangle^2}$$

$\Rightarrow$ Corollary!

PRACTICE:

1. X_1, X_2, X_3 : $X = X_1 \cdot X_2 \cdot X_3$

All we need is ability to compute matrix-vector products for X.

Example: To compute SVD of X_1, X_2, X_3

$$(X_1 \cdot (X_2 \cdot (X_3 \cdot v)))$$

LAPACK have in built support for such computations.

2. How to compute higher singular vectors?

a. (Recall: If $X = U \Sigma V^T$

$$\sigma_1 = \| X \cdot v^{(1)} \|$$

$\searrow$ first right sv

$$u^{(1)} = \frac{X \cdot v^{(1)}}{\sigma_1}$$

$$X = \sigma_1 u^{(1)} v^{(1)T} + \sigma_2 u^{(2)} v^{(2)T} + \dots + \sigma_d u^{(d)} v^{(d)T}$$

To compute 2^{ND}:

compute 1^{ST} of $(X - \sigma_1 u^{(1)} v^{(1)T})$

$\vdots$

so on.

b. GENERAL PI:

→ Pick k orthonormal vectors $y^{(1)}, y^{(2)}, \dots, y^{(k)}$.

→ For $t = 1, \dots, T$:

* Let $z^{(1)} = X^T \cdot X \cdot y^{(1)}_{t-1}$, $z^{(2)} = X^T \cdot X \cdot y^{(2)}_{t-1}$, ...

$$z^{(k)} = X^T \cdot X \cdot y^{(k)}_{t-1}$$

* Let $y_t^{(1)}, y_t^{(2)}, \ldots, y_t^{(k)}$ be an orthonormal basis for $\text{span}(\{z^{(1)}, z^{(2)}, \ldots, z^{(k)}\})$.

3. Let's say X had singular values

$\underbrace{1, 0.99}, 0.98, 0.97, 0.97, \ldots, 0.97$

convergence will be very slow!

→ Suppose $X = V \Sigma V^T$

* What are singular values of $X - \mathbb{I}/2$:

Has singular values

$0.5, 0.49, 0.48, 0.47, \ldots, 0.47$

* $X - 0.96\mathbb{I}$ has singular values

$\underbrace{0.04, 0.03}, 0.02, 0.01, \ldots, 0.01.$

multiplicative gap decreases.

4. PI gets convergence at the rate $1/\varepsilon$.

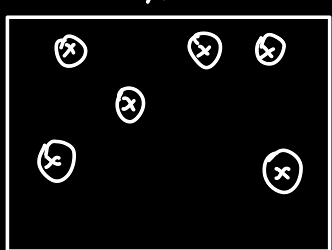
PI with momentum gets convergence at the rate $1/\sqrt{\varepsilon}$.

POWER ITERATION + "MOMENTUM":

- v_0 is a random unit vector
- for $t = 1, \dots, T$:
  - $u = X^T \cdot (X \cdot v_{t-1}) - \beta \cdot v_{t-2}$
  - $v_t = \dfrac{u}{\|u\|}$

MATRIX COMPLETION / NETFLIX CHALLENGE:

X



→ We only see x_{ij} for $(i,j) \in$ some set of observed entries $O \subseteq [n] \times [d]$

INPUT: Entries of X in O, k

OUTPUT: Find $\tilde{X}$ of rank k to

$$\min \sum_{(i,j) \in O} (x_{ij} - \tilde{x}_{ij})^2$$

$$L(\tilde{X}) = \sum_{i,j \in O} (x_{ij} - \tilde{x}_{ij})^2 \rightarrow \text{convex}$$

$$\boxed{\begin{array}{l} \min L(\tilde{X}) \\ \text{Rank}(\tilde{X}) \leq k \end{array}} \rightarrow \text{Non-Convex Optimization problem (NP-Hard)}$$

$$\boxed{\begin{array}{l} C = \{Y : \text{Rank}(Y) \leq k\} \\ \min L(\tilde{X}) \\ \tilde{X} \in C \end{array}}$$

SINGULAR VALUE PROJECTION : (JMD 09)

→ X_0

→ For $t = 1, \dots, T$:

$$Y_t = X_{t-1} - \eta \nabla L(X_{t-1})$$

$$X_t = (\text{Take top-}k\text{-SVD of } Y_t) \rightarrow \text{Proj}_C(Y_t)$$

↓

The closest rank k matrix.