$\boxed{\text{Power Method}}$ for Tensor decomposition.

- $T = \sum_{i=1}^{K} \lambda_i \, \vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$

symm tensor with $[\vec{v}_1, \vec{v}_2, \dots, \vec{v}_K]$

orthogonal array of vectors $\vec{v}_i \in \mathbb{R}^D$

$(K \leq D)$.

Power Method allows to find λ_i's & $\vec{v}_i$'s.

- For Matrices $M = \sum_{i=1}^{K} \lambda_i \, \vec{v}_i \otimes \vec{v}_i$.

standard power method $\rightarrow$ Discuss first.

a) Power Method for $M \in \mathbb{R}^{N \times N}$ real sym

$$M \vec{v}_i = \lambda_i \vec{v}_i \qquad \lambda_i \in \mathbb{R} \qquad \text{and}$$

$\|\vec{v}_i\| = 1$ $\qquad [\vec{v}_1 \; \ldots \; \vec{v}_N]$ orthogonal array.

$$\vec{v}_i^T \cdot \vec{v}_j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N .$$

$$\Rightarrow \qquad M = \sum_{i=1}^{N} \lambda_i \vec{v}_i \vec{v}_i^T$$

$$= \sum_{i=1}^{N} \lambda_i \vec{v}_i \otimes \vec{v}_i$$

Power Method: find the top eigenvalue $\lambda_1, \vec{v}_1$ when $\underline{\lambda_1 > \lambda_2}$.

.... find the next eigenvalue $\lambda_2, \vec{v}_2$ when $\lambda_2 > \lambda_3$.

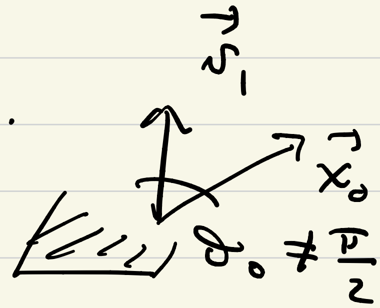
... and so on.

Assume that $\lambda_1 > \lambda_2$.

$\|\vec{x}^{(0)}\|_2 = 1$

1) Choose initial vector $\vec{x}^{(0)}$ at time $t=0$

s.t $\vec{x}^{(0)}$ NOT $\perp$ to $\vec{v}_1$.



2) $$\vec{x}^{(t)} = \frac{M\vec{x}^{(t-1)}}{\|M\vec{x}^{(t-1)}\|_2}$$

Lemma: if $\lambda_1 > \lambda_2$ Then as $t \to +\infty$

$$\vec{x}^{(t)} \to \vec{v}_1 \ ; \quad \vec{x}^{(t)T} \cdot M\vec{x}^{(t)} \to \lambda_1 .$$

with rate:

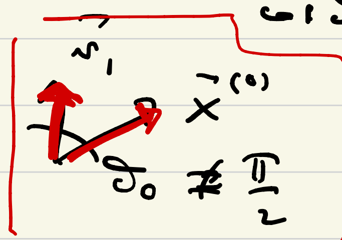
$$\|\vec{x}^{(t)} - \vec{v}_1\|_2 \leq (\tan\theta_0)\left(\boxed{\frac{\lambda_2}{\lambda_1}}\right)^t$$

ratio could be very close to 1 as $N \nearrow$.

similar bound for

$$|\vec{x}^{(t)T} \cdot M\vec{x}^{(t)} - \lambda_1|$$

$\tan\theta_0 = \dfrac{\sin\theta_0}{\cos\theta_0}.$



Remark: High Dim $N \nearrow$

$$\vec{v}_1^T \cdot \vec{x}^{(0)} = \sum_{j=1}^{N} \underbrace{v_1^j}_{\frac{1}{\sqrt{N}}} \underbrace{x^{(0)j}}_{\frac{1}{\sqrt{N}}} \approx \frac{1}{\sqrt{N}} \underbrace{\frac{1}{\sqrt{N}} \sum_j \text{rand \# of } O(1)}_{O(1)} \approx \frac{1}{\sqrt{N}} \text{ with high prob.}$$

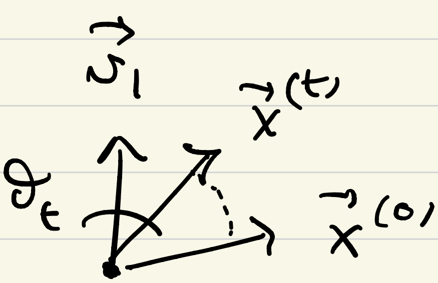
Main idea of proof of lemma

$$x^{(t)} = \frac{M x^{(t-1)}}{\|M x^{(t-1)}\|_2} \implies x^{(t)} = \frac{M^t x^{(0)}}{\|M^t x^{(0)}\|_2} .$$

(you can check)

$$M^t = \sum_{i=1}^{N} \lambda_i^t \, \vec{v}_i \vec{v}_i^T \qquad \boxed{\vec{v}_1^T \cdot \vec{v}_i = \delta_{1i}}$$

$$\vec{x}^t = \frac{\sum_{i=1}^{N} \lambda_i^t \, \vec{v}_i (\vec{v}_i^T \cdot x^{(0)})}{\left\{ \sum_{i=1}^{N} \lambda_i^{2t} (\vec{v}_i^T \cdot x^{(0)})^2 \right\}^{1/2}}$$



$$(\cos\vartheta_t)^2 = (\vec{v}_1^T \cdot \vec{x}^{(t)})^2$$

$$(\cos\vartheta_t)^2 = \frac{\lambda_1^{2t} (\vec{v}_1^T \cdot \vec{x}^{(0)})^2}{\sum_{i=1}^{N} \lambda_i^{2t} (\vec{v}_i^T \cdot x^{(0)})^2}$$

$$(\cos\vartheta_t)^2 = \frac{1}{1 + \sum_{i=2}^{N} \left(\frac{\lambda_i}{\lambda_1}\right)^{2t} \left(\frac{\vec{v}_i^T \cdot x^{(0)}}{v_1^T \cdot x^{(0)}}\right)^2}$$

use that for all $i \geq 2$ $\left(\frac{\lambda_i}{\lambda_1}\right)^{2t} \leq \underline{\underline{\left(\frac{\lambda_2}{\lambda_1}\right)^{2t}}}$

$$\sum_{i=2}^{N} (\vec{v}_i^{\,T} \cdot x^{(0)})^2 = 1 - (\vec{v}_1 \cdot x^{(0)})^2 = 1 - (\cos\vartheta_0)^2$$

because $\|\vec{x}^{(0)}\|_2^2 = 1$

At the end you get (exercise).

$$(\cos\vartheta_t)^2 \geq \frac{1}{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^{2t} \frac{1-\cos\vartheta_0^2}{\cos\vartheta_0}.}$$

$$\Rightarrow \underbrace{1 - (\cos\vartheta_t)^2} \leq \underline{\underline{\left(\frac{\lambda_2}{\lambda_1}\right)^{2t} (tg\,\vartheta_0)^2}} \cdot \underbrace{\frac{1}{1 + \left(\frac{\lambda_2}{\lambda_1}\right)^{2t} (tg\,\vartheta_0)^2}}_{< 1.}$$

$1 - \cos\vartheta_t \leq (1+\cos\vartheta_t)(1-\cos\vartheta_t)$

Now:

$$\|\vec{x}^{(t)} - \vec{v}_1\|_2^2 = \|\vec{x}^{t}\|_2^2 + \|\vec{v}_1\|^2 - 2\cos\vartheta_t = \underline{\underline{2(1-\cos\vartheta_t)}}$$

$$\Rightarrow \|x^t - \vec{v}_1\|_2^2 \leq C \cdot \left(\frac{\lambda_2}{\lambda_1}\right)^{2t} (tg\,\vartheta_0)^2. \qquad \blacksquare$$

__exercise__: deduce the bound for $|x^{(t)} \cdot M x^{(t)} - \lambda_1|$ $\to 0$.

• To obtain λ_2 & $\vec{v}_2$ <u>assuming that</u> $\lambda_2 > \lambda_3$.

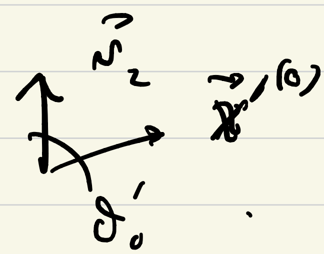
use a "deflation of the matrix M":

$$M - \lambda_1 \vec{v}_1 \vec{v}_1^{\,T} \longrightarrow M'$$

and apply power method iterates to M'

$\Longrightarrow$ you get λ_2 & $\vec{v}_2$ (rate of convergence is again

$$\left(\frac{\lambda_3}{\lambda_2}\right)^t (\tan \theta_0').$$



• Ect ...

FL.

b) Power Iteration Method for Tensors.

$$T = \sum_{i=1}^{K} \lambda_i \, \vec{v}_i \otimes \underbrace{\vec{v}_i \otimes \vec{v}_i}$$

$\vec{v}_i$'s are mutually orth. in $\mathbb{R}^D$.

($K \leq D$ & Jennrich Thm applies: decomp is unique up to rescalings.).

Main idea of Power Method:

$$[T(I, x^t, x^t)]^\alpha = \sum_{\beta\gamma} T^{\alpha\beta\gamma} x^{t\beta} x^{t\gamma}$$

in practice you know $T^{\alpha\beta\gamma}$.

1) $x^{(0)}$ at random initial vector.

2) $$x^t = \frac{T(I, x^{t-1}, x^{t-1})}{\| T(I, x^{t-1}, x^{t-1}) \|_2}$$

3) Under suitable condition $x^t \to$ some $\vec{v}_*$

$T(x^t, x^t, x^t) \to$ some λ_*

$* \in \{1 \ldots K\}$.

$\uparrow$ depend on $x^{(0)}$ see later

4) $T' \leftarrow T - \lambda_* v_* \otimes v_* \otimes v_*$

deflation.

$$T(I, x^t, x^t) = \sum_{i=1}^{K} \lambda_i \vec{v}_i (\vec{v}_i^T \cdot \vec{x}^{(t)})^2 \Big\} \; T(x^t, x^t, x^t) = \sum_i \lambda_i (\vec{v}_i^T \cdot x^t)^3$$

Theorem on tensor power method.

Suppose that

$$\left|\lambda_1 \vec{v}_1^T \cdot \vec{x}^{(0)}\right| > \left|\lambda_2 \vec{v}_2^T \cdot \vec{x}^{(0)}\right| \geq \dots \geq \left|\lambda_N \vec{v}_N^T \cdot \vec{x}^{(0)}\right|$$

strict inequality

Then $$x^{(t)} = \frac{T(I, x^t, x^t)}{\| T(I, x^t, x^t)\|_2} \; ; \; t \geq 1.$$

$$\begin{cases} \| x^{(t)} - \vec{v}_1 \|_2^2 \leq \left(2\lambda_1^2 \sum_{i=2}^{K} \lambda_i^{-2}\right) \left(\dfrac{\lambda_2 \vec{v}_2^T \cdot \vec{x}^{(0)}}{\lambda_1 \vec{v}_1^T \cdot \vec{x}^{(0)}}\right)^{2^{t+1}} \\ \left| T(x^t, x^t, x^t) - \lambda_1 \right| \to 0 \text{ at some rate.} \end{cases}$$

Remark: in very high dim $\vec{v}_2^T$, $\vec{x}^{(0)}$ & $\vec{v}_1^T \cdot \vec{x}^{(0)} \sim \frac{1}{\sqrt{N}}$ with high probability.

ratio might be close to one for $D \nearrow$ ...

Remark: Add noise to the tensor then this affect the ratios → research problems here // "effect of noise"

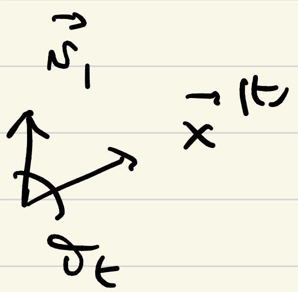
Proof of Theorem.

By def $\quad T(I, x^{t-1}, x^{t-1}) = \sum_{i=1}^{k} \lambda_i \vec{v}_i \, (\vec{v}_i^{T} . x^{t-1})^2$

$$\underline{x}^t = \frac{\sum_{i=1}^{k} \lambda_i (\vec{v}_i^{T} . \vec{x}^{t-1})^{2} \vec{v}_i}{\left\{ \sum_{i=1}^{K} \lambda_i^2 (\vec{v}_i^{T} . x^{t})^4 \right\}^{1/2}}$$; $\vec{v}_i$'s $\perp$.

small calculation allows to check (as in matrices)

$$\vec{x}^{(t)} = \frac{\sum_{i=1}^{k} \lambda_i^{2^t - 1} (v_i^{T} . x^{(0)})^{2^t} \vec{v}_i}{\left\{ \sum_{i=1}^{k} \lambda_i^{2^{t+1} - 2} (\vec{v}_i^{T} . x^{(0)})^{2^{t+1}} \right\}^{1/2}}$$



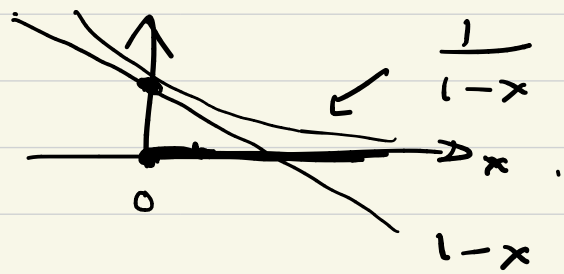
$\cos \vartheta_t = \vec{v}_1^{T} . x^{(t)}$ $\quad$ use $\vec{v}_1^{T} . \vec{v}_i = \delta_{1i}$

This gives

$$(\cos \vartheta_t)^2 = \frac{\lambda_i^{2^{t+1} - 2} (\vec{v}_i^{T} . x^{(0)})^{2^{t+1}}}{\sum_{i=1}^{k} \lambda_i^{2^{t+1} - 2} (\vec{v}_i^{T} . x^{(0)})^{2^{t+1}}}$$

$$\Rightarrow (\cos\theta_t)^2 = \frac{1}{1+\displaystyle\sum_{i=2}^{k}\left(\frac{\lambda_i \vec{v}_i^{\,T}\cdot x^{(0)}}{\lambda_1 v_1^T\cdot x^{(0)}}\right)^{2^{t+1}} \underline{\left(\frac{\lambda_1}{\lambda_i}\right)^2}}$$

$$\frac{1}{1+x} \geq 1-x \quad ; \quad x \geq 0$$



$$(\cos\theta_t)^2 \geq 1 - \sum_{i=2}^{k} \cdots$$

$$1-(\cos\theta_t)^2 \leq \sum_{i=2}^{k} \cdots$$

using $\displaystyle\frac{\lambda_i \vec{v}_i^{\,T}\cdot x^0}{\lambda_1 v_1^T\cdot x^0} \leq \frac{\lambda_2 \vec{v}_2^{\,T}\cdot x^{(0)}}{\lambda_1 \vec{v}_1^{\,T}\cdot x^{(0)}} \;;\; \underline{i \geq 2}$

<u>Finally you get</u>:

$$1-(\cos\theta_t)^2 \leq 2\lambda_1^2\left(\sum_{i\geq 2}\lambda_i^{-2}\right)\left(\frac{\lambda_2 v_2^T\cdot x^{(0)}}{\lambda_1 v_1^T\cdot x^{(0)}}\right)^{2^{t+1}}$$

V/

$2(1-\cos\theta_t) = \underline{\underline{\|x^t - v_1\|_2^2}}$.

$\blacksquare$

$\boxed{\text{exercise}}$: $T(x^t, x^t, x^t) \longrightarrow \lambda_1$ at some rate.

ends the power method.