# III. Tensor PCA

**Model** [Montanari, Richard '14]

$$T = \lambda v^{\otimes p} + W \quad \text{or } Z$$

$$v^{\otimes p} = \underbrace{v \otimes v \otimes \cdots \otimes v}_{p \text{ times}}$$

$\|v\| = \sqrt{n}$

(unif. on sphere)

$$Z \overset{iid}{\sim} N(0,1), \qquad \overbrace{n \times n \times \cdots \times n}^{p \text{ times}}$$

$$W = \frac{1}{\sqrt{p!}} \sum_{\pi \in S_p} \pi \cdot Z$$

$$(\pi \cdot Z)_{i_1 i_2 i_3} = Z_{i_{\pi(1)} i_{\pi(2)} i_{\pi(3)}}$$

$W$ symmetric : $W_{ijk} = W_{ikj} = \cdots$

$W_{1,2,3} \sim N(0,1)$

$W_{1,1,2} \sim N(0,2)$

$W_{1,1,1} \sim N(0,6)$

Fixed $R$, orthog.

$W$ rotationally invariant : $\mathrm{Law}(W) = \mathrm{Law}(R \cdot W)$

Goal: estimate $v$

$\to$ $p$ even: $v$ vs $-v$

$$v^{\otimes p} = (-v)^{\otimes p}$$

$$\to \frac{|\langle \hat{v}, v \rangle|}{\|\hat{v}\| \cdot \|v\|} = 1 - o(1) \quad (n \to \infty)$$

* Problem is rotationally invariant, algorithm should be equivariant $\Rightarrow$ tensor net

<u>p=2</u> $T = \lambda v v^T + W$ spiked Wigner matrix

Alg: $v_{max}(T) \Rightarrow$ succeeds when $\lambda \gg n^{-1/2}$

$\uparrow$ leading eigenvector

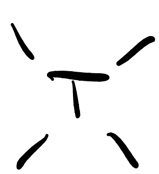
Compare <u>spectral</u> <u>norm</u>: $\|\lambda v v^T\| = \lambda \cdot n$, $\|W\| \simeq 2\sqrt{n}$

$\hookrightarrow \|M\| = \max_i |\lambda_i(M)|$

<u>p=4</u> $T = \lambda v^{\otimes 4} + Z$

Tensor unfolding [MR '14]

$$\underset{n \times n \times n \times n}{T} \xmapsto{\text{"flatten"}} \underset{n^2 \times n^2}{\boxed{M}}$$



(rows) $=$ T $=$ (cols)

$$M_{ij,k\ell} = T_{ijk\ell}$$

$$M = \lambda (v^{\otimes 2})(v^{\otimes 2})^T + \tilde{Z}$$

$\tilde{Z}$ ← iid $N(0,1)$

$n^2$ $\quad$ $n^2$

(1,1)
(1,2)
⋮
(2,1)
⋮

Then symmetrize $M \leftarrow \frac{1}{\sqrt{2}}(M + M^T)$

Alg: $v_{max}(M)$

Hope: $v_{max} \approx \pm v^{\otimes 2}$

Succeeds when $\lambda \gg n^{-1}$ $\qquad \frac{\lambda}{n^{-1}} \to \infty$

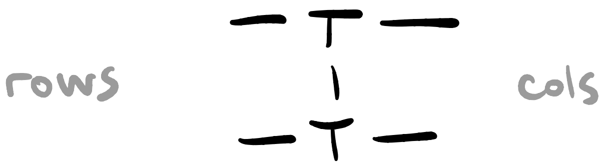
↳ Spectral norms: $\|\lambda (v^{\otimes 2})(v^{\otimes 2})^T\| = \lambda n^2$

$\|\tilde{Z}\| \approx \sqrt{n^2} \approx n$

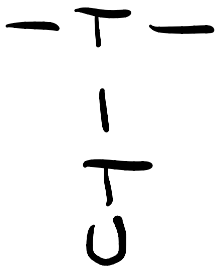
p=3 "Spectral methods from tensor networks"

[Hopkins, Shi, Steurer '15]
[Hopkins, Schramm, Shi, Steurer '15]
[Moitra, W '18]

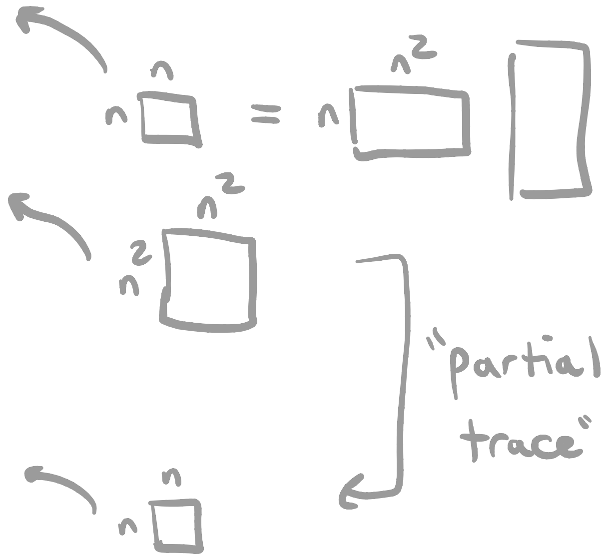
[HSS '15]




[HSS '15]



[HSSS '15]



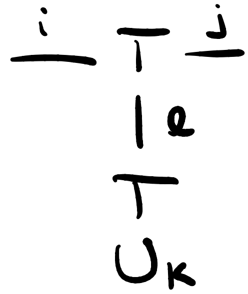

$T \mapsto M$, then $v_{max}(M)$

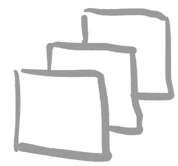
[Anandkumar, Deng, Ge, Mobahi '16]

↳ Doesn't work well on it's own, but "warm start"

* Many methods, all get $\lambda >> n^{-3/4}$
* General p: $\lambda >> n^{-p/4}$

Analysis



"slices"

$$M_{ij} = \sum_{kl} T_{ijl} T_{kkl} = \sum_{l} T_{ijl} \underbrace{\sum_{k} T_{kkl}}_{\mathrm{Tr}(T^{(l)})}$$

Slice $l$: $T^{(l)}$

$$(T^{(l)})_{ij} = T_{ijl}$$

$$M = \sum_{l} \underset{\substack{\uparrow \\ \text{matrix}}}{T^{(l)}} \cdot \mathrm{Tr}\underset{\substack{\uparrow \\ \text{scalar}}}{(T^{(l)})}$$

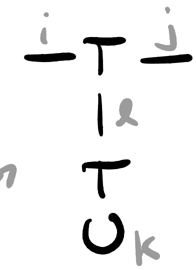
For noise $Z$ : slices are independent
($v$ fixed, $Z$ random)

* $M$ is the sum of independent random matrices $\Rightarrow$ standard matrix Chernoff bound

Recap...

Tensor PCA: $T = \lambda v^{\otimes 3} + W$ $\qquad \|v\| = \sqrt{n}$

Algorithm [HSSS '15]: $v_{max}$ of matrix



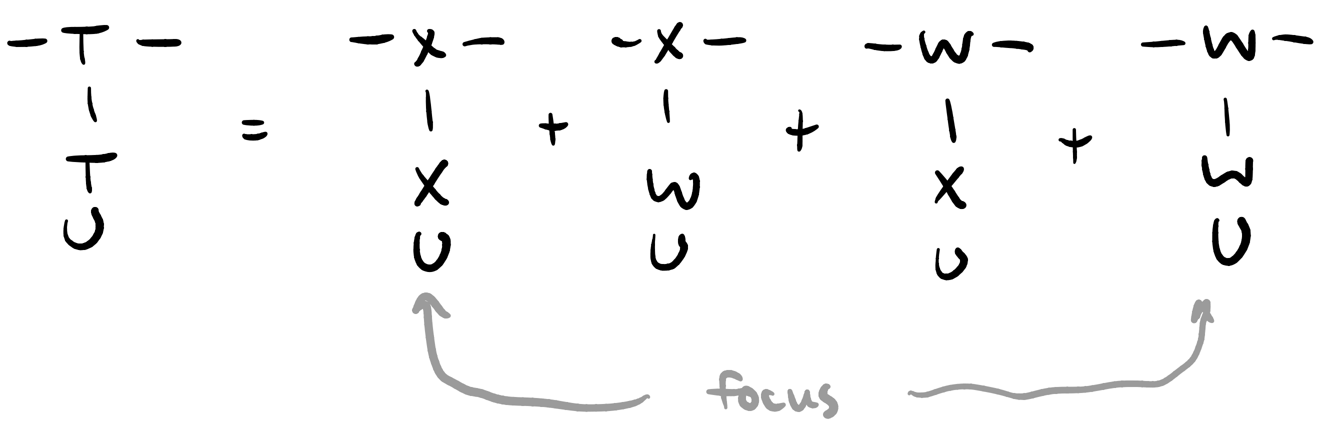

↳ succeeds when $\lambda \gg n^{-3/4}$

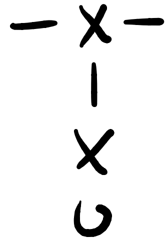
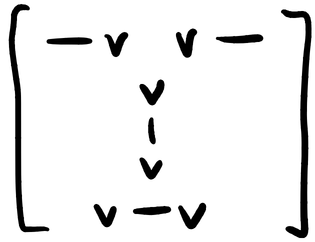
ij entry: $\sum_{kl} T_{ijl} T_{kkl}$

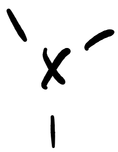
* Previously: direct analysis via matrix Chernoff (for simple networks only)

* Next: "diagrammatic" explanation for why it works

$$T = \underbrace{X}_{\lambda v^{\otimes 3}} + W$$



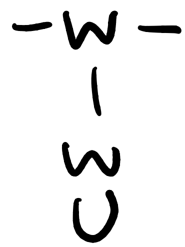
Signal term:  $= \lambda^2$ 

 $= \lambda$ 

$= \lambda^2 \|v\|^4 \cdot vv^T$

$\Rightarrow$ rank-1, spectral norm $\left\| \lambda^2 \|v\|^4 \cdot vv^T \right\|$

$= \lambda^2 n^2 \cdot n = \lambda^2 n^3$

Noise term: $M =$ 

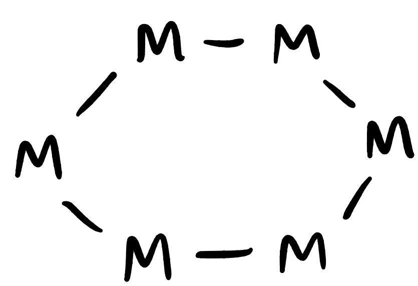
M symmetric

Want $\|M\| \le \dots$

Trace moment method:

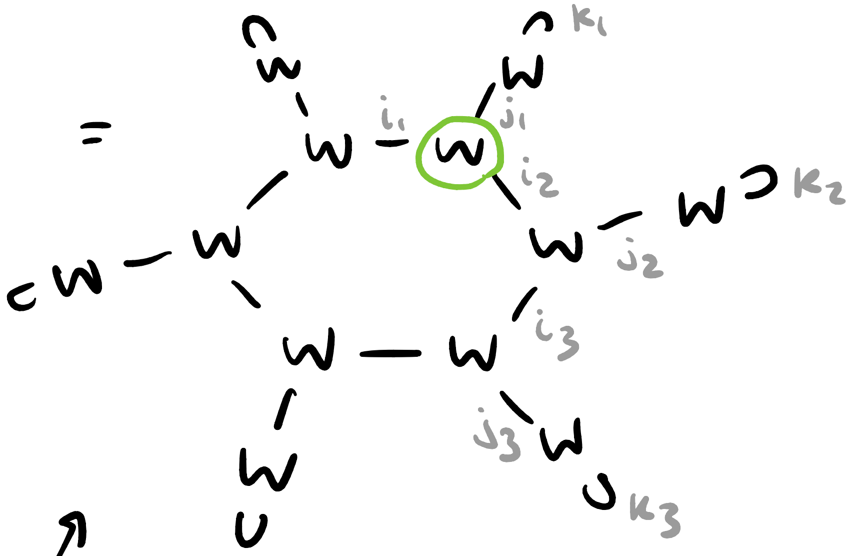
$$\mathbb{E}[\mathrm{Tr}(M^{2k})] = \mathbb{E} \sum_i \lambda_i^{2k} \ge \mathbb{E}\, \lambda_{max}^{2k} = \mathbb{E}\left[ \|M\|^{2k} \right]$$

Markov: $$\Pr\left( \|M\|^{2k} \ge t \right) \le \frac{\mathbb{E}\left[ \|M\|^{2k} \right]}{t} \le \frac{\mathbb{E}\, \mathrm{Tr}(M^{2k})}{t}$$

$\mathrm{Tr}(M^{2k}) =$ 

$k=3$

In general: 2k copies of M

$=$ 


$M =$ 

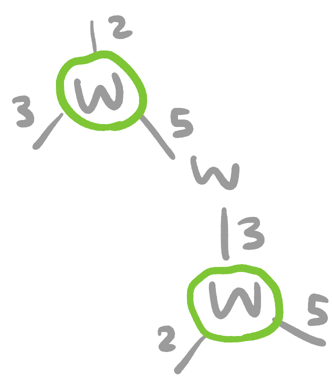
Want $\mathbb{E}[\,\cdot\,]$

$W_{1,3,6} = W_{3,1,6}$

$$= \sum_{\substack{i_1, i_2, \dots \\ j_1, j_2, \dots \\ k_1, k_2, \dots}} \underbrace{\mathbb{E}\, W_{i_1, j_1, i_2} W_{\dots} W_{\dots} \cdots}_{0 \text{ unless labeling } (i_\ell, j_\ell, k_\ell) \text{ is "valid"}}$$

$\phi : E(G) \to [n]$

each $W_{abc}$ appear an even # of times

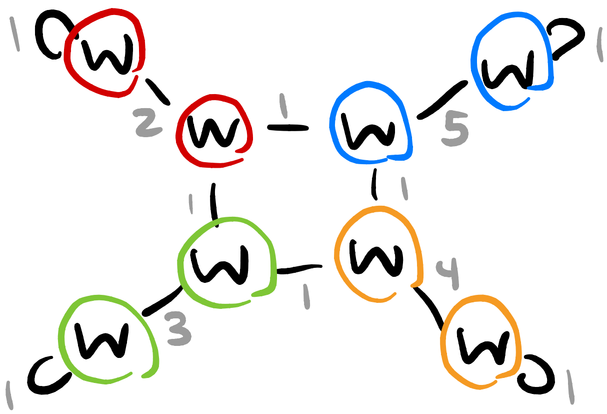



$\mathbb{E}[W_{2,3,5}] = 0$

$\mathbb{E}[W^2_{2,3,5}] = 1$

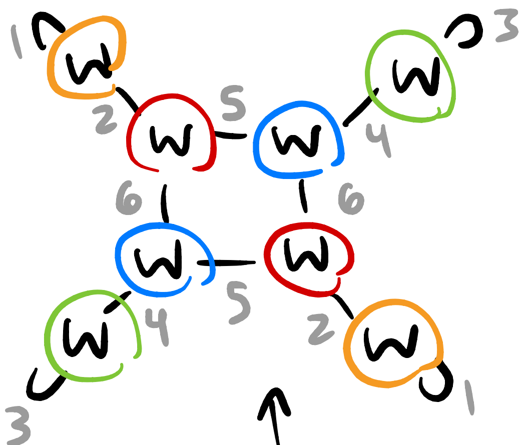
$$\Rightarrow \mathbb{E}\,\mathrm{Tr}(M^{2k}) = \Theta(\#\text{ valid labelings})$$

Count valid labelings $(k=2)$



5 "free" labels

$\# : \quad \Theta(n^5)$



$\# : \quad \Theta(n^6)$

Claim: This is the optimal configuration

$$\Rightarrow \mathbb{E}\,\mathrm{Tr}(M^{2k}) = \Theta(n^{3k})$$

$\#$ total edges: $3 \cdot 2k = 6k$

$\#$ free labels: $3k$

$$\Rightarrow \|M\|^{2k} \lesssim \Theta(n^{3k})$$

$$\|M\| \lesssim \Theta(n^{3/2})$$

Signal $\|\cdot\| = \lambda^2 n^3$ vs noise $\|\cdot\| \approx n^{3/2}$

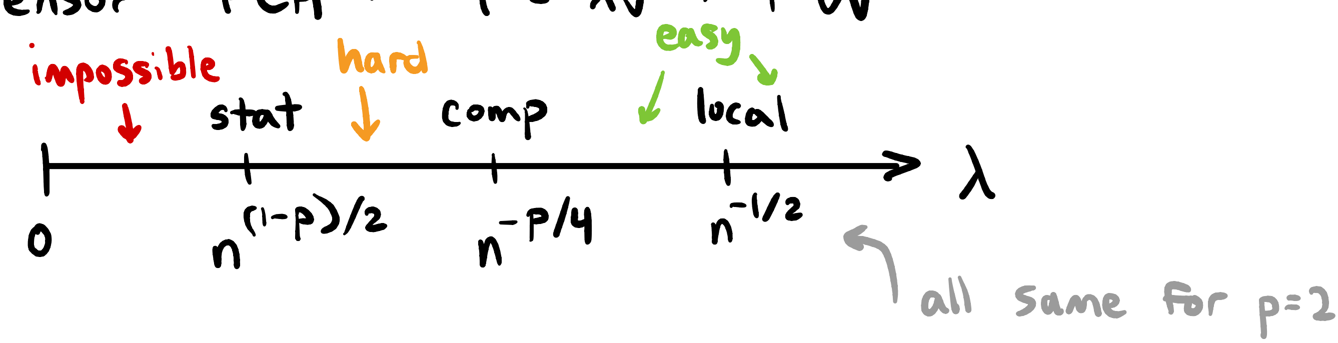
$\Rightarrow$ Succeed when $\lambda^2 n^3 \gg n^{3/2} \Rightarrow \lambda \gg n^{-3/4}$

This approach:

- Works for general networks [MW'18]
- But proof is difficult...
- I would like to see better tools for this style of analysis...

To summarize...

Tensor PCA: $T = \lambda v^{\otimes p} + W$




Statistically impossible | possible by exhaustive search

"low-degree" algs fail | known poly-time algs

"local algs"

[HKPRSS'17]

[KWB'19]

[KMW'24]

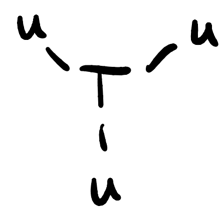
stat-comp gap

comp-local gap

"Local" algs:

- Tensor power method [MR '14]


- Gradient descent [Ben Arous, Gheissari, Jagannath '18]

  ↳ objective $\max_{\substack{u \\ \|u\| = \sqrt{n}}}$ 


- AMP [MR '14]

* Potential lesson for ML... Gradient descent is not always optimal!

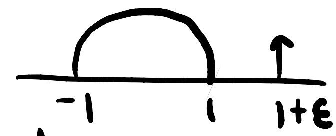
"Low-degree lower bound":

$$\inf_{\substack{f : \mathbb{R}^{n \times n \times n} \to \mathbb{R}^n \\ \deg(f) \leq D}} \mathbb{E} \|f(T) - v\|^2 \geq \dots$$

see [Schramm, W '22]

$D \gg \log n$

Why $\log(n)$?  $M$ w/ spectrum




# iterations of power method needed to compute leading eigenvec

Want $\lambda_1^t \gtrsim \sum_{i=2}^{n} \lambda_i^t$  $\lambda_1 \geq \lambda_2 \geq \cdots$

$$(1+\varepsilon)^t \gtrsim n \cdot 1^t$$

$$t \log(1+\varepsilon) \gtrsim \log n$$

$$t \gtrsim \frac{\log n}{\varepsilon}$$