CS 59300 – Algorithms for Data Science

Classical and Quantum approaches

Lecture 5 (09/16)

Tensor Methods (V)

https://ruizhezhang.com/course_fall_2025.html

Recap

We introduced the tensor network diagram

We discussed some applications in quantum computing (simulating quantum circuits) and quantum physics (MPS, DMRG, etc)

Today: we'll talk about a classical application of tensor network: orbit recovery

Today's plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

Orbit recovery

Setup:

- Let $x \in \mathbb{R}^n$ be an unknown signal
- Let G be a group with group action $\mathcal{P}: G \to \mathbb{R}^{n \times n}$
- We get measurements of the form:

$$y_i = \mathcal{P}(g_i)x + \eta_i$$

  - g_i is an independent, uniformly random element from G (under the Haar measure)
  - η_i is an independent Gaussian noise $\mathcal{N}(0, \sigma^2 I)$

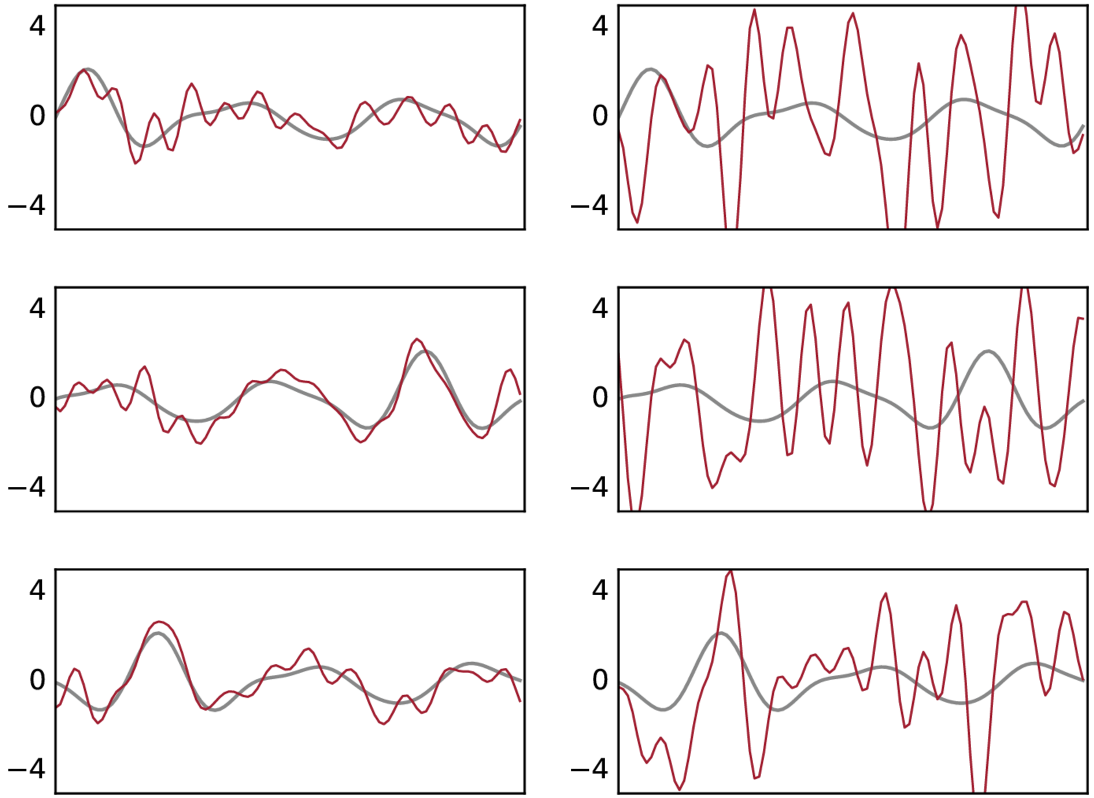
Goal: recover $\hat{x}$ close to some element in the orbit

$$\{\mathcal{P}(g)x \mid g \in G\}$$

For simplicity, we'll use $g \cdot x$ to denote the group action

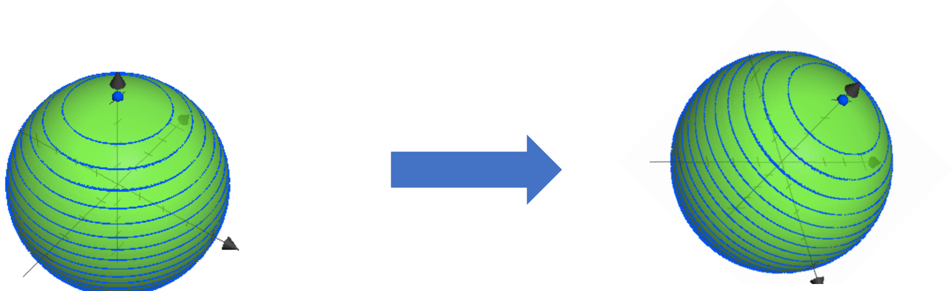
Example 1: multi-reference alignment (MRA)

- Discrete MRA: $G = \mathbb{Z}_n$ (random shift)
- Continuous MRA: $G = SO(2)$ (2D random rotation)

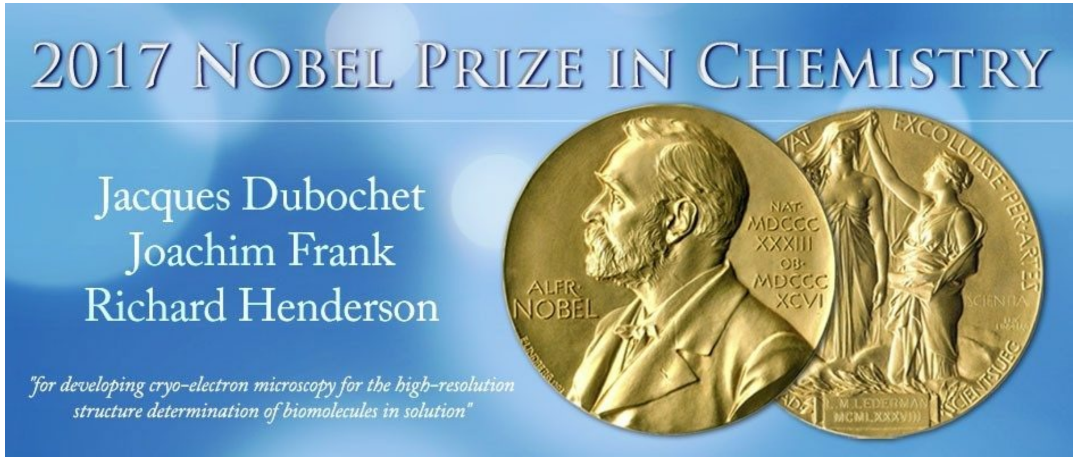


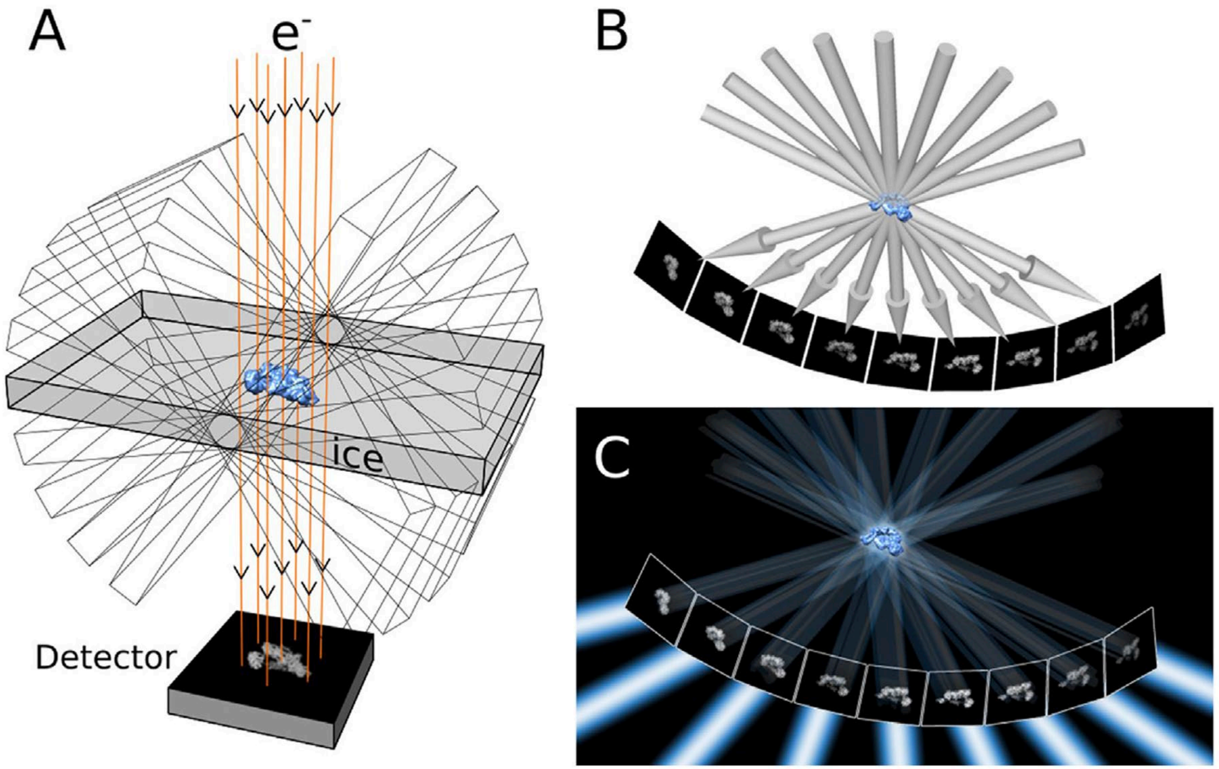
Example 2: cryo-electron tomography (cryo-ET)

- $G = SO(3)$



+ noise

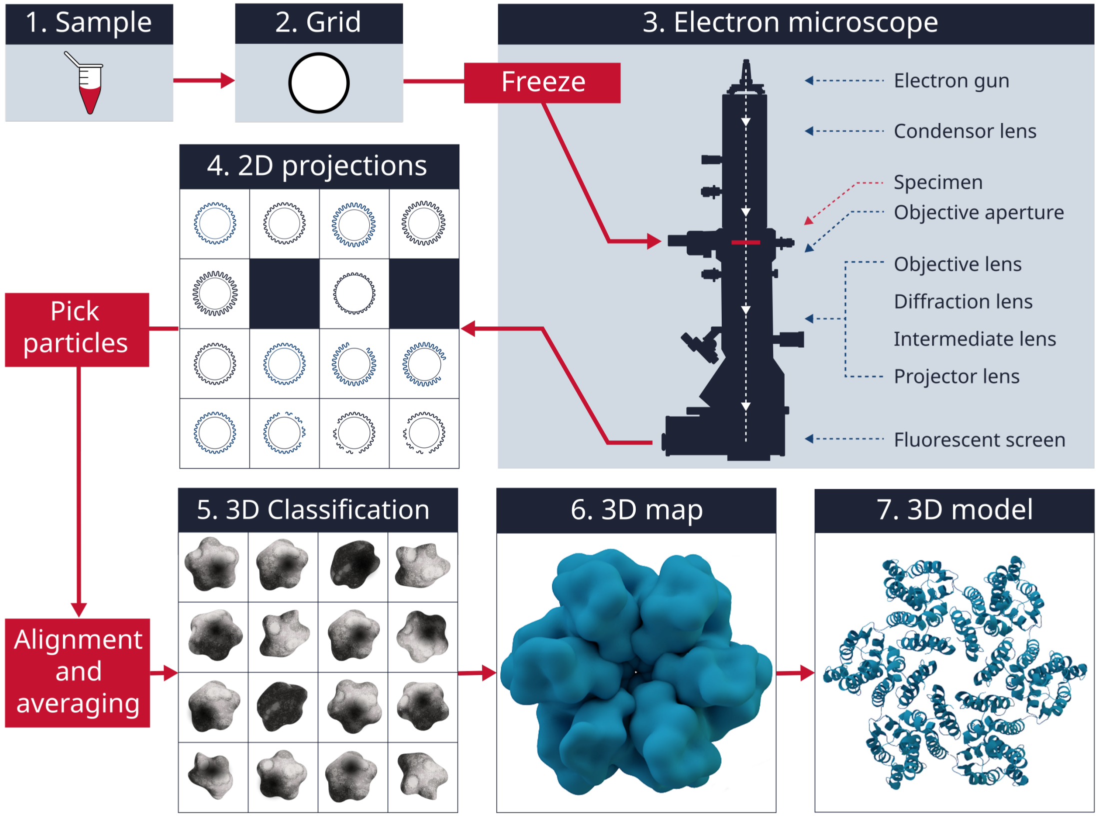




Cryo-electron microscopy (cryo-EM)

- Cryo-ET + 2D projection

$$y_i = \Pi(g_i \cdot x) + \eta_i$$



Today’s plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

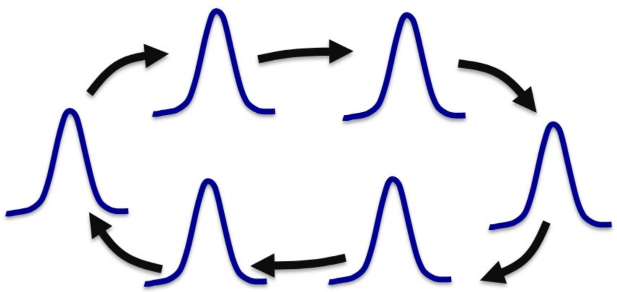
Discrete MRA

- For $g \in G = \mathbb{Z}_n$, we have

$$(g \cdot x)_i = x_{i-g \pmod n}$$

- **Perry-Weed-Bandeira-Rigollet-Singer, 2017:** algorithm for discrete MRA with optimal sample complexity ($m \sim \sigma^6$)
- We have seen this algorithm: learning mixtures of Gaussians
- Each sample $y_i = g_i \cdot x + \eta_i$. If g_i is fixed, then

$$y_i \sim \mathcal{N}(g_i \cdot x, \sigma^2 I)$$



- Thus, the sample distribution can be expressed as a mixture of Gaussian with n components:

$$\mu_i = g_i \cdot x \quad \forall i \in [n]$$

- To apply Jennrich's algorithm, we compute the 3rd-moment, which needs $\sim \sigma^6$ samples

Continuous MRA

For $g \in G = SO(2)$, what is $g \cdot x$ for $x \in \mathbb{R}^n$?

- $$x = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & \cdots & x_{\frac{n}{2}} \\ x_{-1} & x_{-2} & x_{-3} & x_4 & \cdots & x_{-\frac{n}{2}} \end{bmatrix}$$

$$\begin{pmatrix} x_j \\ x_{-j} \end{pmatrix} \overset{g}{\longmapsto} \begin{bmatrix} \cos(jg) & -\sin(jg) \\ \sin(jg) & \cos(jg) \end{bmatrix} \begin{pmatrix} x_j \\ x_{-j} \end{pmatrix}$$

The linear map $g \in \mathbb{R}^{n \times n}$ is block-diagonal

- G can be parameterized by $g \in (0, 2\pi]$
- It is convenient to work in the Fourier basis:

$$\hat{x}_j := \frac{1}{\sqrt{2}}\left(x_j + \mathbf{i}x_{-j}\right), \qquad \hat{x}_{-j} := \frac{1}{\sqrt{2}}\left(x_j - \mathbf{i}x_{-j}\right) \quad \forall\, j > 0$$

- You can check that in the Fourier basis, g is a diagonal matrix ($\hat{x}_j \mapsto e^{\mathbf{i}jg}\hat{x}_j$)

Method of moments

- Define the p-th moment in the Fourier basis:

$$\hat{T}_p(\hat{x}) \coloneqq \mathbb{E}_g\left[(g \cdot \hat{x})^{\otimes p}\right] \in \mathbb{R}^{n^{\times p}}$$

- For any coordinates $j_1, \ldots, j_p \in [n]$, we have

$$\begin{aligned}
\hat{T}_p(\hat{x})_{j_1,\ldots,j_p} &= \mathbb{E}_g\left[(g \cdot \hat{x})_{j_1}(g \cdot \hat{x})_{j_2} \cdots (g \cdot \hat{x})_{j_p}\right] \\
&= \mathbb{E}_g\left[e^{\mathbf{i}gj_1}\hat{x}_{j_1} e^{\mathbf{i}gj_2}\hat{x}_{j_2} \cdots e^{\mathbf{i}gj_p}\hat{x}_{j_p}\right] \\
&= \mathbb{E}_g\left[e^{\mathbf{i}g(j_1+\cdots+j_p)}\right]\hat{x}_{j_1} \cdots \hat{x}_{j_p} \\
&= \mathbf{1}_{j_1+\cdots+j_p=0} \cdot \hat{x}_{j_1} \cdots \hat{x}_{j_p}
\end{aligned}$$

- Given access to $\hat{T}_1, \ldots, \hat{T}_p$, can we reverse-engineer $\hat{x}$?

Frequency marching

$$\hat{T}_p(\hat{x})_{j_1,\ldots,j_p} = \mathbf{1}_{j_1+\cdots+j_p=0} \cdot \hat{x}_{j_1} \cdots \hat{x}_{j_p}$$

- Consider the second moment:

$$\hat{T}_2(\hat{x})_{j,-j} = \hat{x}_j \hat{x}_{-j} = \left|\hat{x}_j\right|^2 \quad \forall\, j > 0$$

- Thus, we can learn $\left|\hat{x}_j\right|$ for every j from $\hat{T}_2$
- To learn the phases, consider the third moment:

$$\hat{T}_3(\hat{x})_{j_1,j_2,-(j_1+j_2)} = \hat{x}_{j_1}\hat{x}_{j_2}\hat{x}_{-(j_1+j_2)}$$

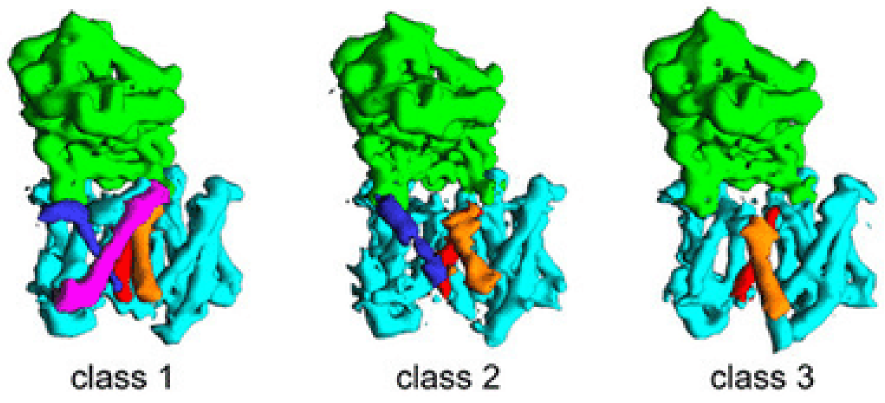
- Since $(g \cdot \hat{x})_1 = e^{\mathbf{i}g}\hat{x}_1$, there exists an orbit such that $\hat{x}_1$'s phase $\phi_1 = 0$
- Then, using $\hat{T}_3(\hat{x})_{-1,-1,2} = |\hat{x}_1|^2\hat{x}_2$ and $|\hat{x}_1|$, we learn ϕ_2
- Next, using $\hat{T}_3(\hat{x})_{-1,-2,3} = \hat{x}_1\hat{x}_{-2}\hat{x}_3$ and $\hat{x}_{-2} = \hat{x}_2^*$, we learn ϕ_3
- We can repeat this procedure until we have learned all the phases

Today’s plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

Heterogeneous MRA

What if there are multiple molecules (or multiple conformations of the same molecule)?



- Suppose there are K unknown vectors $x^1, x^2, \ldots, x^K$
- We observe $y_i = g_i \cdot x^{k_i} + \eta_i$, where $k_i \sim_u [K]$, $g_i \sim_u G$, and $\eta_i \sim \mathcal{N}(0, \sigma^2 I)$
- Can we still use frequency marching to recover $\{x^k\}$?

Frequency marching does not work

- Consider the Fourier transform of the p-th moment:

$$\hat{T}_p(\{\hat{x}^k\}) \coloneqq \mathbb{E}_{k,g}\left[\left(g \cdot \hat{x}^k\right)^{\otimes p}\right] = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_g[(g \cdot \hat{x})^{\otimes p}]$$

- Thus, we have

$$\hat{T}_p(\{\hat{x}^k\})_{j_1,\dots,j_p} = \mathbf{1}_{j_1+\dots+j_p=0} \cdot \frac{1}{K}\sum_{k=1}^{K} \hat{x}^k_{j_1} \cdots \hat{x}^k_{j_p}$$

- Frequency marching breaks down from the first step:

Signals are entangled!

$$\hat{T}_2(\{\hat{x}^k\})_{j,-j} = \frac{1}{K}\sum_{k=1}^{K} \left|x^k\right|^2$$

Does tensor decomposition help?

- Consider the third moment:

$$T_3(\{x^k\}) \coloneqq \frac{1}{K|G|}\sum_{k=1}^{K}\sum_{g\in G}(g\cdot x^k)^{\otimes 3}$$

- If G is a finite group, then this is a rank-$K|G|$ tensor
- If undercomplete, then we can just run Jennrich's algorithm, and we're done!
- If overcomplete, we may either use higher moments (e.g., T_5) or assume x^k are random vectors

- Unfortunately, $SO(2)$ is a continuous group (or Lie group) → this tensor has ∞ rank
- Moreover, the decomposition is not unique: if $T_3 = \sum_{i=1}^{r} a_i^{\otimes 3}$ is a solution, then there are ∞-many solutions $T_3 = \sum_{i=1}^{r}(g\cdot a_i)^{\otimes 3}$ for any $g \in [G]$

Today’s plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

List recovery for average-case heterogeneous MRA

Theorem (Moitra-Wein, 2018).

Let $x^1, \ldots, x^K \in \mathbb{R}^n$ be drawn independently from $\mathcal{N}\left(0, \frac{1}{n} I\right)$.

We are given the tensor $\mathcal{T} = T + E \in \mathbb{R}^{n \times n \times n}$ where $\|E\| \leq 1/\mathrm{poly}(n)$ and

$$T = \sum_{k=1}^{K} \mathbb{E}_g\left[(g \cdot x^K)^{\otimes 3}\right]$$

There is an algorithm that runs in time $\mathrm{poly}(n)$ and outputs a list of unit vectors $\tau_1, \ldots, \tau_L \in \mathbb{R}^n$ with $L = \mathrm{poly}(n)$ that has the following guarantee. With high probability over both $x^1, \ldots, x^K$ and the algorithm's randomness,

$$\forall k \in [K], \exists\, i \in [L] \ \text{ s.t. } \ \langle \tau_i, x^k \rangle^2 \geq 0.99$$

General recipe for spectral methods

Given input tensor T

- **Step 1:** Construct a new tensor B by contracting multiple copies of T according to a tensor network
- **Step 2:** Flatten B to form a symmetric matrix M
- **Step 3:** Compute the leading eigenvector of M

We use the trace method to show that the top eigenvector is close to the orbit of x^k

Today's plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

Interlude: the trace method

Let M be a random matrix, and our goal is to bound its spectral norm

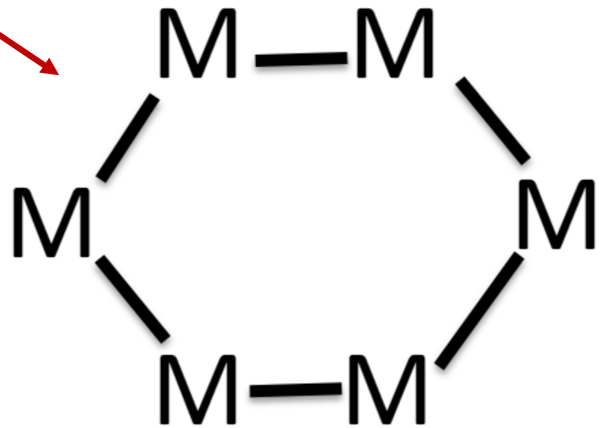
Basic idea:

$$\mathrm{tr}[M^{2k}] = \sum_i \lambda_i^{2k} \geq \|M\|^{2k}$$

Applying Markov's inequality, we get the bound

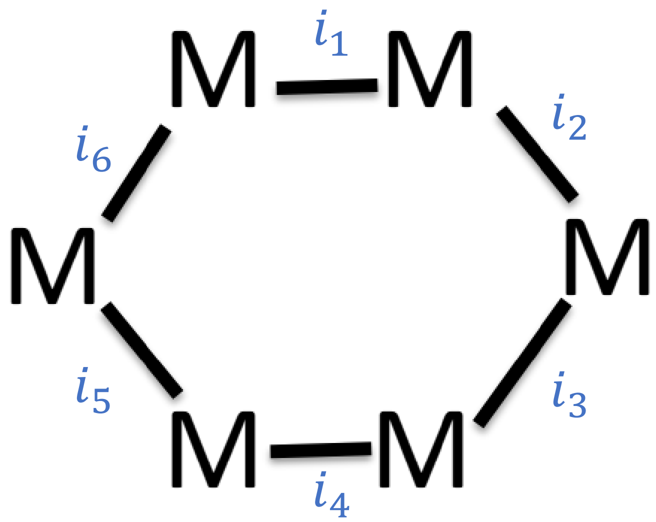
$$\Pr[\|M\| \geq t] = \Pr[\|M\|^{2k} \geq t^{2k}] \leq \frac{\mathbb{E}\left[\mathrm{tr}[M^{2k}]\right]}{t^{2k}}$$

Tensor network!



Example

Suppose M is an $n \times n$ symmetric matrix with *i.i.d.* Rademacher entries and zeros along the diagonal



$$\mathbb{E}\, M_{i_1,i_2} M_{i_2,i_3} M_{i_3,i_4} M_{i_4,i_5} M_{i_5,i_6} M_{i_6,i_1} = 1$$

iff every $\{i_j, i_{j+1}\}$ occurs even number of times

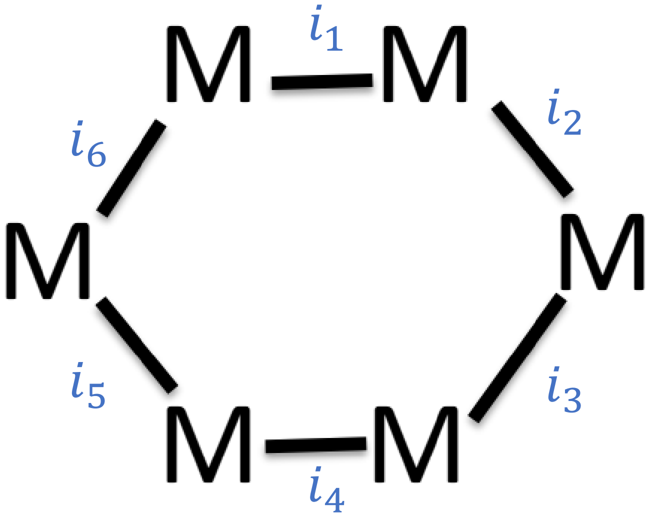
Combinatorial problem: $\mathrm{tr}[M^{2k}]$ equals to the number of sequences $(i_1, \ldots, i_{2k}) \in [n]^{2k}$ such that every $\{i_j, i_{j+1}\}$ occurs in the sequence an even number of times

- At most $k + 1$ distinct labels in the sequences

$k = 2$: $(i_1, \ldots, i_4) = (a, b, c, b)$ $\quad a \quad b \quad c \quad$ $(i_1, \ldots, i_4) = (a, b, a, c)$ $\quad a \quad b \quad c$

Example

Suppose M is an $n \times n$ symmetric matrix with *i.i.d.* Rademacher entries and zeros along the diagonal



$$\mathbb{E}\, M_{i_1,i_2} M_{i_2,i_3} M_{i_3,i_4} M_{i_4,i_5} M_{i_5,i_6} M_{i_6,i_1} = 1$$

iff every $\{i_j, i_{j+1}\}$ occurs even number of times

Combinatorial problem: $\mathrm{tr}[M^{2k}]$ equals to the number of sequences $(i_1, \dots, i_{2k}) \in [n]^{2k}$ such that every $\{i_j, i_{j+1}\}$ occurs in the sequence an even number of times

- At most $k+1$ distinct labels in the sequences
- $\mathrm{tr}[M^{2k}] \le n^{k+1} \cdot (k+1)^{2k}$
- $\|M\| \le \sqrt{n} \log n$ (by taking $k \sim \log n$)

Furedi-Komlos: $\|M\| \simeq \sqrt{n}$

Today's plan

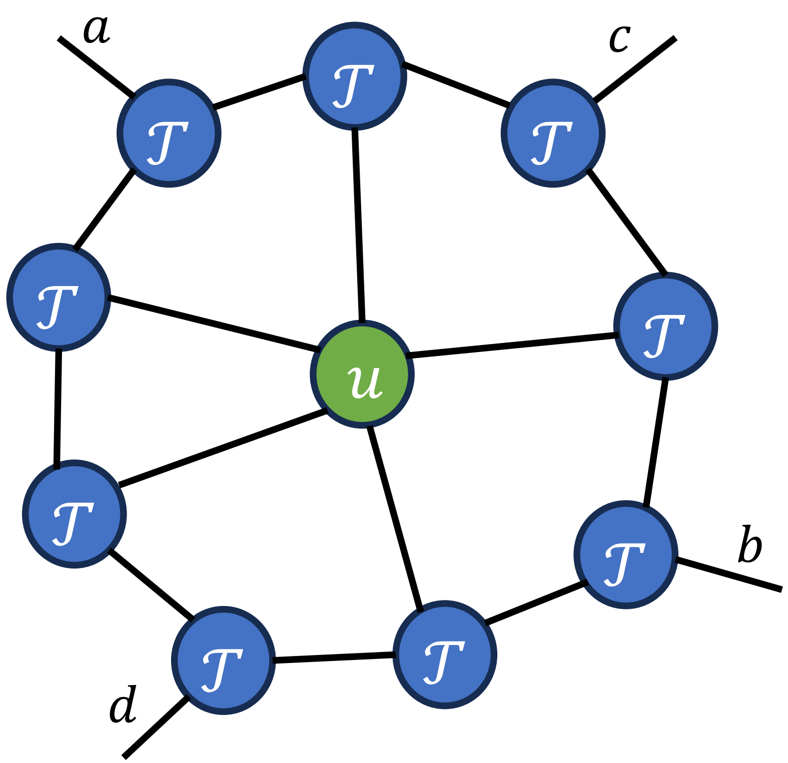
- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

The blueprint

- The algorithm takes a random order-5 tensor u (with i.i.d. $\mathcal{N}(0,1)$ entries)
- The hope is that u has non-trivial correlation with some x in the orbit of one of $x^1, \ldots, x^K$
- Compute the following tensor network:

Q: Why do we need a random tensor u?

A: Symmetry-breaking



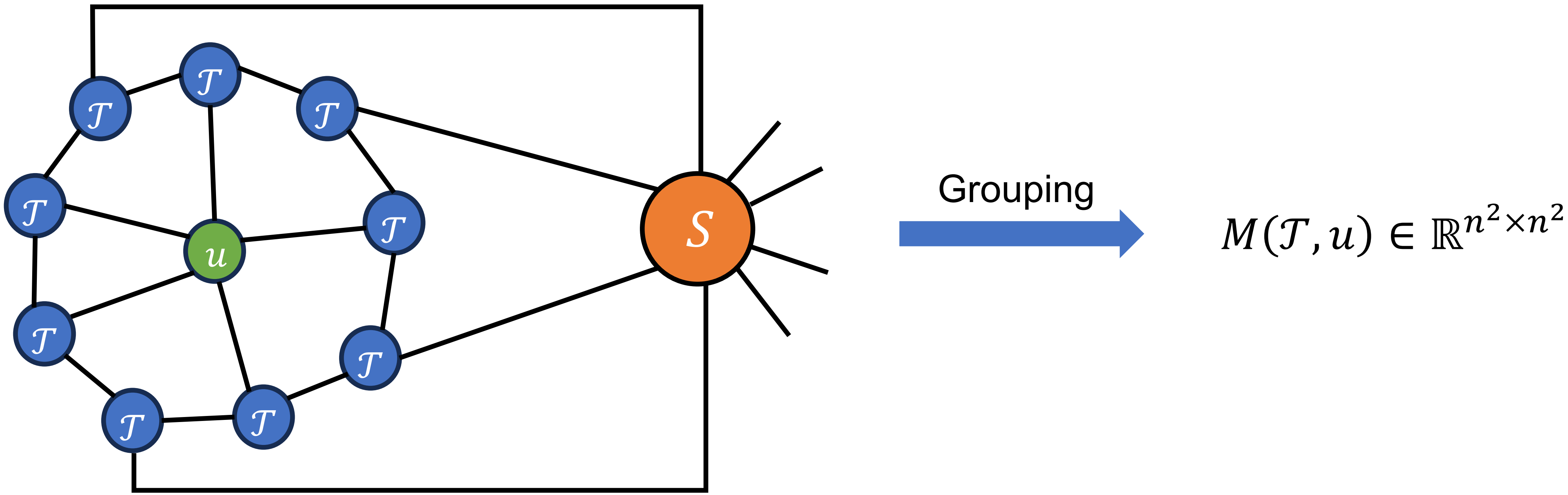
Grouping $\{(a,b),(c,d)\}$ $\longrightarrow$ $\widetilde{M}(\mathcal{T},u) \in \mathbb{R}^{n^2 \times n^2}$

We want to show that if $u = x^{\otimes 5}$, then

$\widetilde{M}(\mathcal{T},u) \approx x^{\otimes 2}\left(x^{\otimes 2}\right)^{\top}$

The blueprint

- Use a simple tensor $S \in \mathbb{R}^{n^{\times 8}}$ to correct the tensor network:



Main technical step

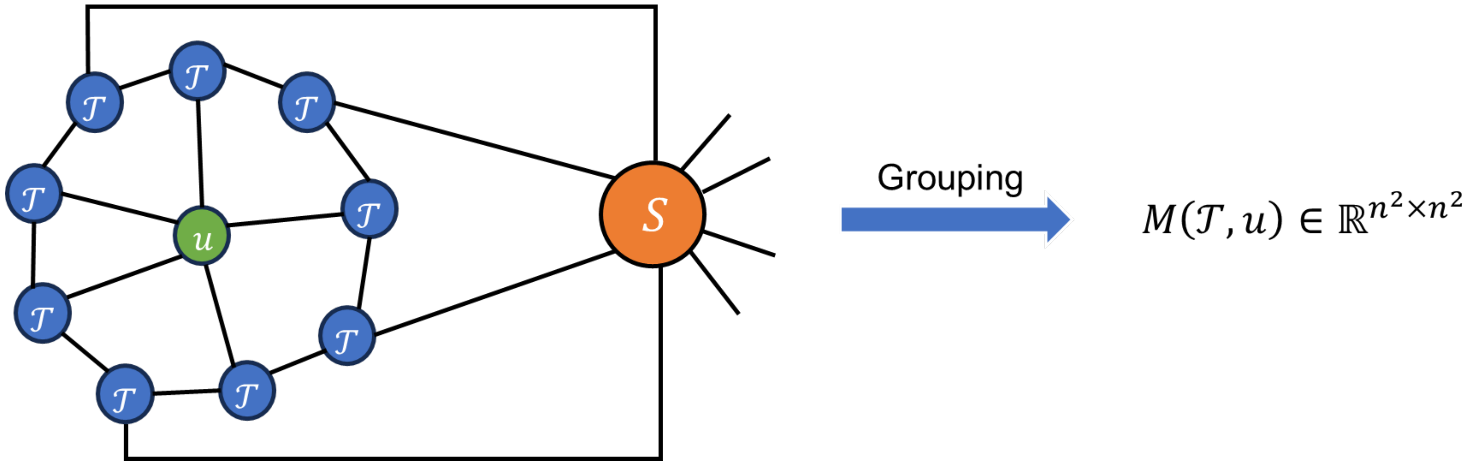
Technical theorem.

There is a matrix $M(\mathcal{T}, u) \in \mathbb{R}^{n^2 \times n^2}$ (computable in poly-time from $\mathcal{T}$ and u) with the following guarantee.

Let $v \in \mathbb{R}^{n^2}$ be the leading eigenvector of $(M(\mathcal{T}, u) + M(\mathcal{T}, u)^\top)/2$.

Let $V = \mathrm{mat}(v) \in \mathbb{R}^{n \times n}$ and let $\tau \in \mathbb{R}^n$ be the top eigenvector of $(V + V^\top)/2$.

With high probability over $\{x^k\}$, for any $k \in [K]$, we have $\langle \tau, x^k \rangle^2 \geq 0.99$ with probability $1/\mathrm{poly}(n)$.



Main technical step

Technical theorem.

There is a matrix $M(\mathcal{T}, u) \in \mathbb{R}^{n^2 \times n^2}$ (computable in poly-time from $\mathcal{T}$ and u) with the following guarantee. Let $v \in \mathbb{R}^{n^2}$ be the leading eigenvector of $(M(\mathcal{T}, u) + M(\mathcal{T}, u)^\top)/2$. Let $V = \mathrm{mat}(v) \in \mathbb{R}^{n \times n}$ and let $\tau \in \mathbb{R}^n$ be the top eigenvector of $(V + V^\top)/2$. With high probability over $\{x^k\}$, for any $k \in [K]$, we have $\langle \tau, x^k \rangle^2 \geq 0.99$ with probability $1/\mathrm{poly}(n)$.

Proof of the main theorem:

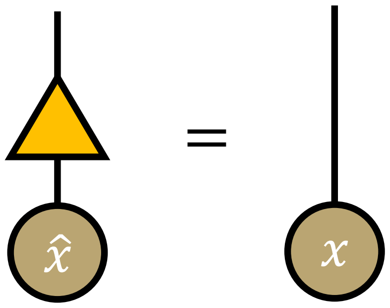
- Sample $u_1, \ldots, u_L$ and use the technical theorem to obtain $\tau_1, \ldots, \tau_L$
- For any k, the overall failure probability is $\leq (1 - 1/\mathrm{poly}(n))^L = \exp(-L/\mathrm{poly}(n))$
- By union bound over all $k \in [K]$, the total failure probability is $\leq K \exp(-L/\mathrm{poly}(n)) = o(1)$

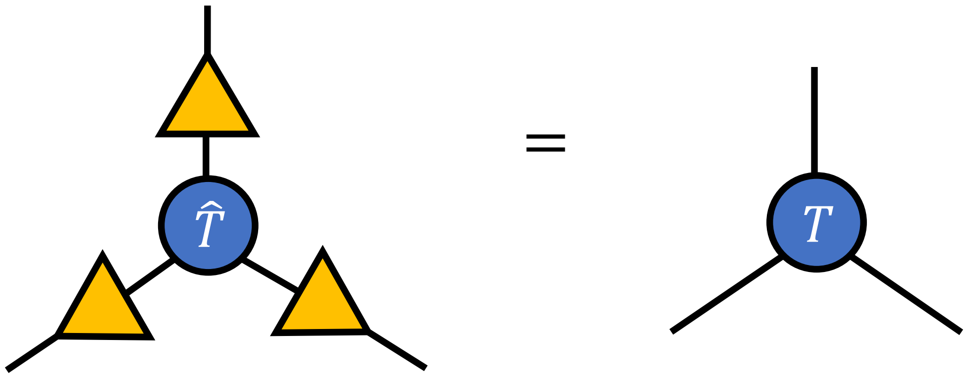
∎

Fourier transform in tensor network

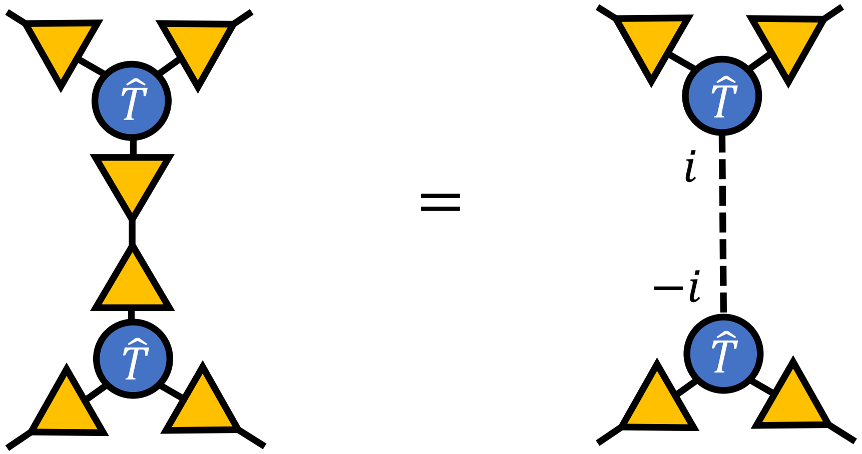
$$\hat{x}_j := \frac{1}{\sqrt{2}}\left(x_j + \mathbf{i}x_{-j}\right), \quad \hat{x}_{-j} := \frac{1}{\sqrt{2}}\left(x_j - \mathbf{i}x_{-j}\right)$$

We first define Δ to be the Fourier transform unitary matrix that transforms $\hat{x}$ to x





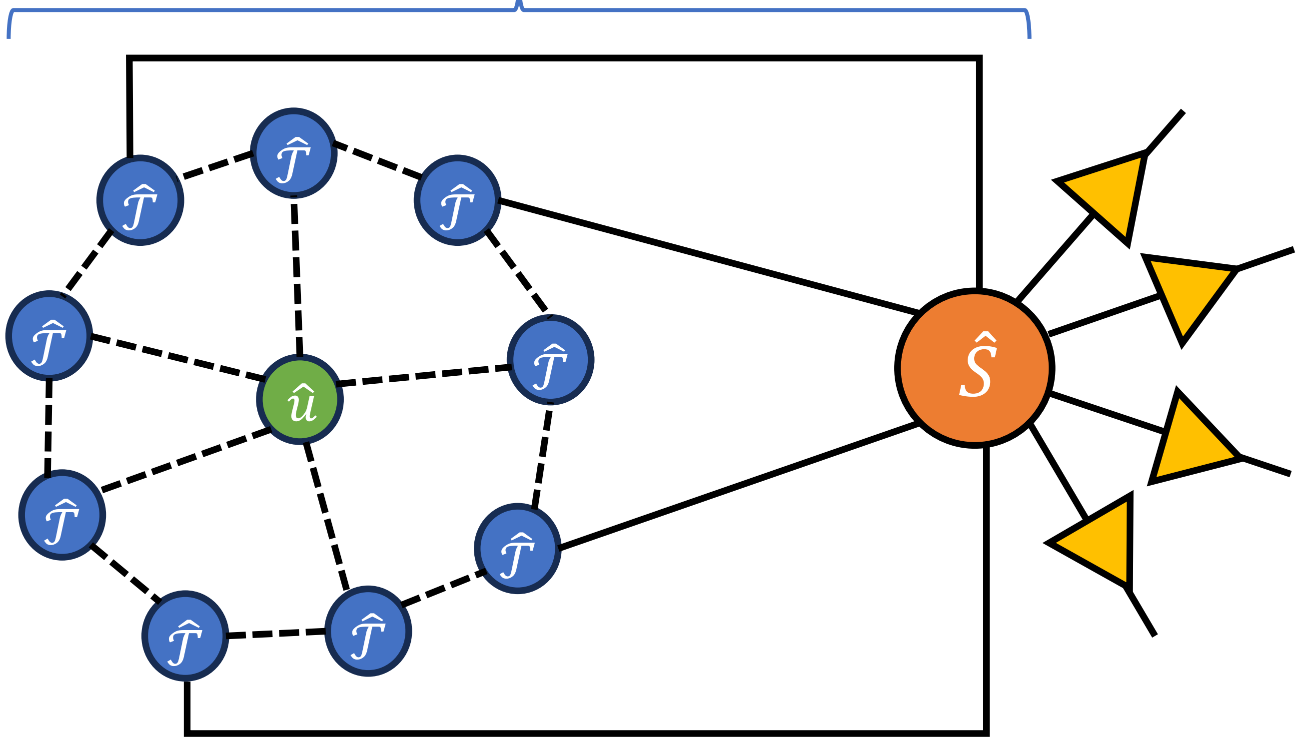
You can check that $(\Delta^{\top}\Delta)_{ij} = \mathbf{1}_{i=-j}$



Corrected tensor network

$$M(\mathcal{T}, u) = (\Delta \otimes \Delta)\widehat{M}(\mathcal{T}, u)(\Delta \otimes \Delta)^{\mathsf{T}}$$

$\widehat{\mathcal{M}}(\mathcal{T}, u)$



$$(\hat{S}T)_{abcd} = S_{abcd}T_{abcd}$$

Today’s plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

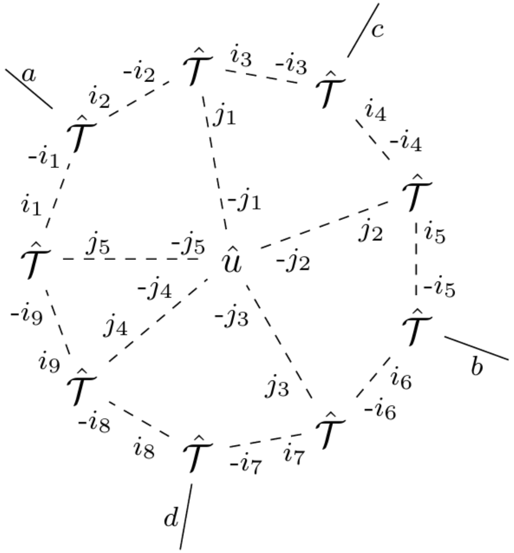
Correcting the signal

Goal: If we correctly guess $u = \left(x^k\right)^{\otimes 5}$, then $M(\mathcal{T}, u) \approx \left(x^k \otimes x^k\right)\left(x^k \otimes x^k\right)^\top \equiv X^k$

- Wlog, we can write $u = \alpha\left(x^1\right)^{\otimes 5} + \tilde{u}$, where $\tilde{u} \perp \left(x^1\right)^{\otimes 5}$ (noise)
- Let $T^1 \coloneqq \mathbb{E}_g\left[\left(g \cdot x^1\right)^{\otimes 3}\right]$ denote the 3rd moment of x^1. Then

Pure signal

$$\left(\hat{T}^1\right)_{j_1 j_2 j_3} = \mathbf{1}_{j_1+j_2+j_3=0}\, \hat{x}^1_{j_1} \hat{x}^1_{j_2} \hat{x}^1_{j_3}$$



We want to match $M\left(T^1, \left(x^1\right)^{\otimes 5}\right)$ to X^1:

- $\widehat{M}\left(T^1, \left(x^1\right)^{\otimes 5}\right)_{ab,cd} = S_{abcd} \cdot s_{abcd} \cdot \hat{x}^1_a \hat{x}^1_b \hat{x}^1_c \hat{x}^1_d$, where

$$s_{abcd} \coloneqq \sum_{i_1,\ldots,i_9} \sum_{j_1,\ldots,j_5} \left(\mathbf{1}_{a+i_2=i_1} \cdots \mathbf{1}_{i_1+j_5=i_9}\right) \left|\hat{x}^1_{i_1}\right|^2 \cdots \left|\hat{x}^1_{i_9}\right|^2 \left|\hat{x}^1_{j_1}\right|^2 \cdots \left|\hat{x}^1_{j_5}\right|^2$$

$$s_{abcd} \coloneqq \sum_{i_1,\dots,i_9} \sum_{j_1,\dots,j_5} \left(\mathbf{1}_{a+i_2=i_1} \cdots \mathbf{1}_{i_1+j_5=i_9}\right) \left|\hat{x}^1_{i_1}\right|^2 \cdots \left|\hat{x}^1_{i_9}\right|^2 \left|\hat{x}^1_{j_1}\right|^2 \cdots \left|\hat{x}^1_{j_5}\right|^2$$

$$\left(\widehat{M}\right)_{ab,cd} \coloneqq S_{abcd} s_{abcd} \cdot \hat{x}^1_a \hat{x}^1_b \hat{x}^1_c \hat{x}^1_d$$

- Define

$$S_{abcd} \coloneqq \begin{cases} 0 & \text{if } a = -b \text{ or } c = -d \\ \dfrac{1}{\mathbb{E}_{x^1}[s_{abcd}]} & \text{otherwise} \end{cases}$$

- We will show that s_{abcd} is concentrated around its mean (which is independent of x^1)

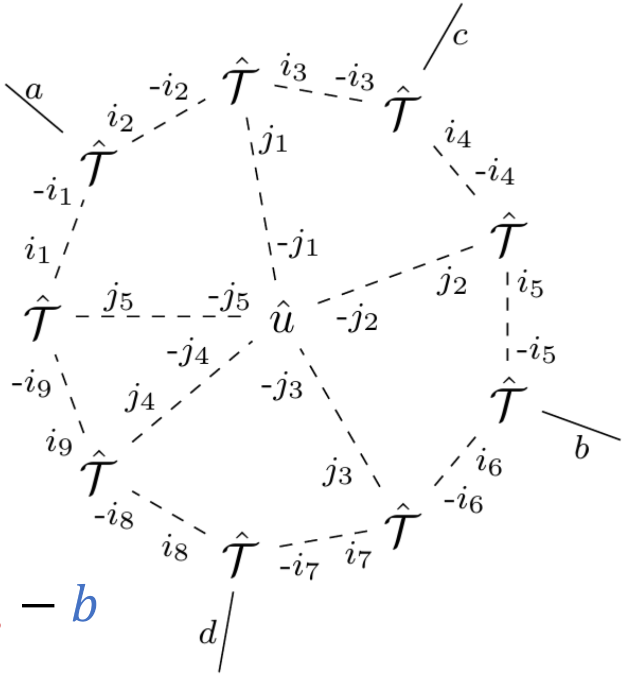
Proposition.

For $a \neq -b$ and $c \neq -d$, we have $|S_{abcd} s_{abcd} - 1| = \mathcal{O}(n^{-0.1})$ with overwhelming probability over x^1. Therefore, $\left\|M\left(T^1, (x^1)^{\otimes 5}\right) - X^1\right\| = o(1)$.

Linear constraints on the indices

- $i_1 - i_2 = a$
- $i_1 + j_5 - i_9 = 0$
- $i_9 + j_4 - i_8 = 0$
- $i_7 - i_8 = d$
- $i_7 + j_3 - i_6 = 0$
- $i_5 - i_6 = b$
- $i_5 + j_2 - i_4 = 0$
- $i_3 - i_4 = c$
- $i_3 + j_1 - i_2 = 0$

- $i_1 = i_9 - j_5$
- $i_2 = i_9 - j_5 - a$
- $i_3 = i_9 - j_5 - a - j_1$
- $i_4 = i_9 - j_5 - a - j_1 - c$
- $i_5 = i_9 + j_4 + d + j_3 + b$
- $i_6 = i_9 + j_4 + d + j_3$
- $i_7 = i_9 + j_4 + d$
- $i_8 = i_9 + j_4$
- $j_2 = -j_5 - a - j_1 - c - j_4 - d - j_3 - b$



For any fixed (a, b, c, d), there are 5 “free” indices $(i_9, j_1, j_3, j_4, j_5)$

- Thus, the number of solutions is upper-bounded by n^5
- We can also prove that the number of valid solution is lower-bounded by cn^5 for some small constant c, by a careful counting argument

Moments of random vector

Lemma. Let $x^1 \sim \mathcal{N}(0, I/n)$. Then, we have

- $\mathbb{E}_{x^1}\left[\left|\hat{x}_i^1\right|^{2k}\right] = k!\, n^{-k}$
- If $k_1 \neq k_2$, then $\mathbb{E}_{x^1}\left[\left(\hat{x}_i^1\right)^{k_1}\left(\hat{x}_{-i}^1\right)^{k_2}\right] = 0$
- If $i \neq \pm j$, then $\hat{x}_i^1$ and $\hat{x}_j^1$ are independent

$\Rightarrow$ $\mathbb{E}_{x^1}\left[\left|\hat{x}_{i_1}^1\right|^2 \cdots \left|\hat{x}_{i_9}^1\right|^2 \left|\hat{x}_{j_1}^1\right|^2 \cdots \left|\hat{x}_{j_5}^1\right|^2\right]$
$= \Theta(n^{-14})$

Proof.

- $\hat{x}_i^1 \sim \mathcal{N}(0, I/(2n)) + \mathbf{i}\mathcal{N}(0, I/(2n))$ and $\hat{x}_{-i}^1 = \left(\hat{x}_i^1\right)^*$
- $\left|\hat{x}_i^1\right|^2 \sim \frac{1}{2n}\chi^2$

- Expectation:

$$\mathbb{E}_{x^1}[s_{abcd}] = \sum_{i_1,\dots,i_9} \sum_{j_1,\dots,j_5} \left(\mathbf{1}_{a+i_2=i_1} \cdots \mathbf{1}_{i_1+j_5=i_9}\right) \mathbb{E}_{x^1}\left[\left|\hat{x}^1_{i_1}\right|^2 \cdots \left|\hat{x}^1_{i_9}\right|^2 \left|\hat{x}^1_{j_1}\right|^2 \cdots \left|\hat{x}^1_{j_5}\right|^2\right]$$

$$= \sum_{i_1,\dots,i_9} \sum_{j_1,\dots,j_5} \left(\mathbf{1}_{a+i_2=i_1} \cdots \mathbf{1}_{i_1+j_5=i_9}\right) \Theta(n^{-14})$$

S_{abcd} only depends on n

$$= \Theta\left(n^5 \cdot n^{-14}\right) = \Theta(n^{-9})$$

- Variance:

$$\mathrm{Var}[s_{abcd}] = \mathrm{Var}\left[\sum_t Z_t\right] = \sum_t \mathrm{Var}[Z_t] + \sum_{t \neq t'} \mathrm{Cov}(Z_t, Z_{t'})$$

$$\leq \Theta(n^5) \cdot \mathcal{O}(n^{-28}) + \mathcal{O}(n^9) \cdot \mathcal{O}(n^{-28}) = \mathcal{O}(n^{-19})$$

- By Chebyshev's inequality, they imply that $|s_{abcd} - \mathbb{E}[s_{abcd}]| \leq n^{-9.1}$ with probability $1 - 1/\mathrm{poly}(n)$
- Note that s_{abcd} is a degree-14 polynomial of Gaussian variables. Gaussian hypercontractivity can boost the probability to $1 - \exp\left(-\mathrm{poly}(n)\right)$

Proposition.

For $a \neq -b$ and $c \neq -d$, we have $|S_{abcd}s_{abcd} - 1| = \mathcal{O}(n^{-0.1})$ with overwhelming probability over x^1. Therefore, $\left\|M\left(T^1, (x^1)^{\otimes 5}\right) - X^1\right\| = o(1)$.

- We have proved that for $a \neq -b$ and $c \neq -d$,

$$\left(\widehat{M}\right)_{ab,cd} = S_{abcd}s_{abcd} \cdot \hat{x}_a^1\hat{x}_b^1\hat{x}_c^1\hat{x}_d^1 = (1 \pm n^{-0.1})\hat{x}_a^1\hat{x}_b^1\hat{x}_c^1\hat{x}_d^1$$

- Thus, we have

$$\begin{aligned}\left\|M\left(T^1, (x^1)^{\otimes 5}\right) - X^1\right\| &= \left\|M\left(T^1, (x^1)^{\otimes 5}\right) - (x^1 \otimes x^1)(x^1 \otimes x^1)^\top\right\| \\ &= \left\|\widehat{M}\left(T^1, (x^1)^{\otimes 5}\right) - (\hat{x}^1 \otimes \hat{x}^1)(\hat{x}^1 \otimes \hat{x}^1)^\top\right\| \\ &\leq \left\|\widehat{M}\left(T^1, (x^1)^{\otimes 5}\right) - (\hat{x}^1 \otimes \hat{x}^1)(\hat{x}^1 \otimes \hat{x}^1)^\top\right\|_F \\ &\leq \sqrt{n^4 \cdot (n^{-0.1} \cdot n^{-2})^2 + 2n^3 \cdot (n^{-2})^2} \\ &= o(1)\end{aligned}$$

- The proposition is then proved

$\blacksquare$

If correctly guess $u = x^k$, then

$$M \approx \left(x^k \otimes x^k\right)\left(x^k \otimes x^k\right)^\top$$

Road map

$u = \alpha x^1 + \tilde{u}, \qquad \tilde{u} \perp x^1$

$$M(\mathcal{T}, u) = \alpha M\left(T^1, (x^1)^{\otimes 5}\right) + \alpha\left(M\left(T, (x^1)^{\otimes 5}\right) - M\left(T^1, (x^1)^{\otimes 5}\right)\right) + \left(M(\mathcal{T}, u) - M(T, u)\right) + M(T, \tilde{u})$$

If correctly guess $u = x^k$, then $M \approx \left(x^k \otimes x^k\right)\left(x^k \otimes x^k\right)^\top$

Road map

$u = \alpha x^1 + \tilde{u}, \qquad \tilde{u} \perp x^1$

$$M(\mathcal{T}, u) = \boxed{\alpha M\left(T^1, (x^1)^{\otimes 5}\right)} + \alpha\left(M\left(T, (x^1)^{\otimes 5}\right) - M\left(T^1, (x^1)^{\otimes 5}\right)\right) + \left(M(\mathcal{T}, u) - M(T, u)\right) + M(T, \tilde{u})$$

$$= \left(1 \pm o(1)\right)\left(x^k \otimes x^k\right)\left(x^k \otimes x^k\right)^\top$$

If correctly guess $u = x^k$, then $M \approx (x^k \otimes x^k)(x^k \otimes x^k)^\top$

Road map

$$u = \alpha x^1 + \tilde{u}, \qquad \tilde{u} \perp x^1$$

$$M(\mathcal{T}, u) = \alpha M\left(T^1, (x^1)^{\otimes 5}\right) + \boxed{\alpha\left(M\left(T, (x^1)^{\otimes 5}\right) - M\left(T^1, (x^1)^{\otimes 5}\right)\right) + \left(M(\mathcal{T}, u) - M(T, u)\right)} + M(T, \tilde{u})$$

$$\left\| M\left(T, (x^1)^{\otimes 5}\right) - M\left(T^1, (x^1)^{\otimes 5}\right) \right\| = o(1)$$

$$\| M(\mathcal{T}, u) - M(T, u) \| = o(1)$$

If correctly guess $u = x^k$, then
$M \approx (x^k \otimes x^k)(x^k \otimes x^k)^\top$

Road map

$u = \alpha x^1 + \tilde{u}, \qquad \tilde{u} \perp x^1$

$$M(\mathcal{T}, u) = \alpha M\left(T^1, (x^1)^{\otimes 5}\right) + \alpha\left(M\left(T, (x^1)^{\otimes 5}\right) - M\left(T^1, (x^1)^{\otimes 5}\right)\right) + \left(M(\mathcal{T}, u) - M(T, u)\right) + \boxed{M(T, \tilde{u})}$$

Non-negligible!

- Random tensor contraction ($\tilde{u} \sim \mathcal{N}(0, \Sigma)$)
- The trace method to upper bound $\|\widetilde{W}\|$
- Combinatorial problem of counting labels

$\Rightarrow$ $\|M(T, \tilde{u})\| = \sqrt{\log n}$

$\widetilde{W} =$

Road map

If correctly guess $u = x^k$, then
$M \approx (x^k \otimes x^k)(x^k \otimes x^k)^\top$

$u = \alpha x^1 + \tilde{u}, \qquad \tilde{u} \perp x^1$
$$M(\mathcal{T}, u) = \alpha M\left(T^1, (x^1)^{\otimes 5}\right) + \alpha \left(M\left(T, (x^1)^{\otimes 5}\right) - M\left(T^1, (x^1)^{\otimes 5}\right)\right) + \left(M(\mathcal{T}, u) - M(T, u)\right) + M(T, \tilde{u})$$

✅ ✅ ✅ ✅

x^1 can be recovered from the top eigenvector of $M(\mathcal{T}, u)$

As long as we sample sufficiently many u's, we can "hit" all x^k w.h.p.

Today’s plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
    - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

Towards proving the technical theorem

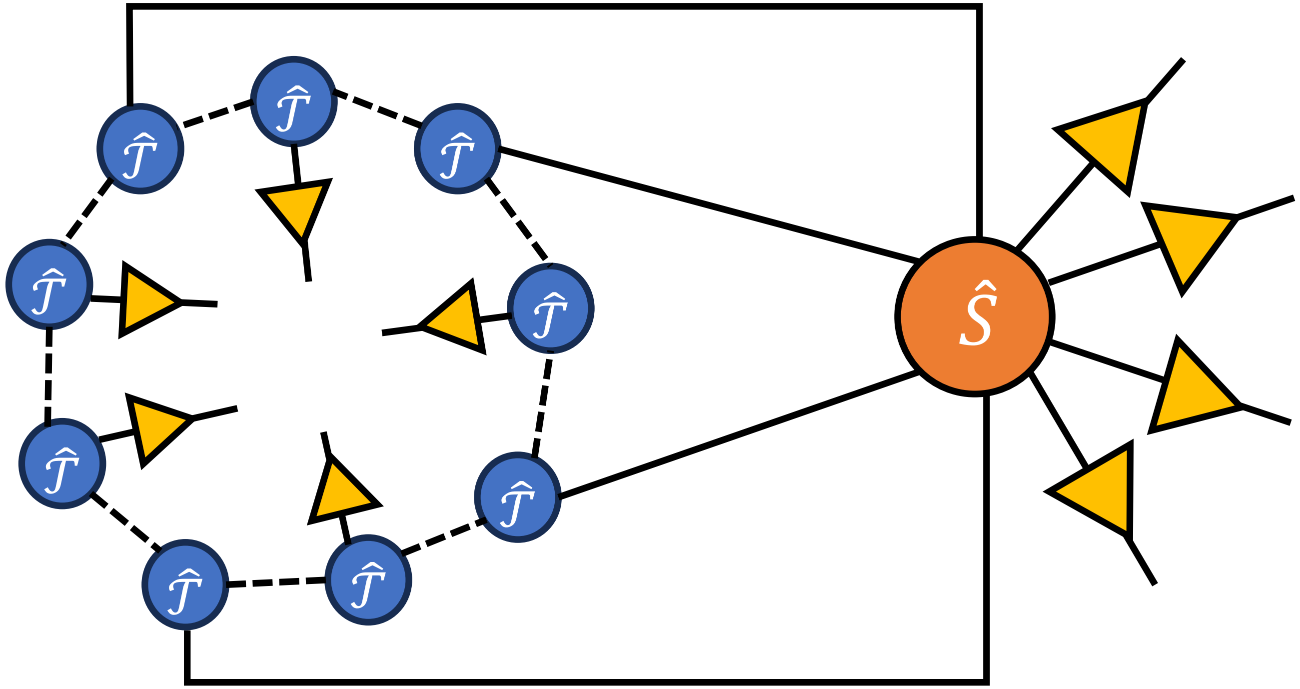
Recall that $u = \alpha(x^1)^{\otimes 5} + \tilde{u}$, and our final goal is to analyze

$$M(\mathcal{T}, u) = \underbrace{\alpha M\left(T^1, (x^1)^{\otimes 5}\right)}_{\approx X^1} + \underbrace{\alpha\left(M\left(T, (x^1)^{\otimes 5}\right) - M\left(T^1, (x^1)^{\otimes 5}\right)\right)}_{\text{heterogeneous signal term}} + \underbrace{M(T, \tilde{u})}_{\text{noise term}} + \underbrace{\left(M(\mathcal{T}, u) - M(T, u)\right)}_{\text{error term}}$$

Key Proposition.

There is an event $\mathcal{E}$ depending only on $\{x^k\}$ that happens with high probability over the randomness of $\{x^k\}$. Conditioned on $\mathcal{E}$, we have $\|M(T, \tilde{u})\| = \mathcal{O}\left(\sqrt{\log n}\right)$ with high probability over the randomness of $\tilde{u}$.

$$M(T, \tilde{u}) = (I_{d^2} \otimes I_{d^2} \otimes \tilde{u})W$$

$W \in \mathbb{R}^{n^2 \times n^2 \times n^5} :=$ 

Interlude: random tensor contraction

Theorem (Ma-Shi-Steurer, 2016).

Let $W \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{R}^r$ be an order-3 tensor. Let $\tilde{u} \sim \mathcal{N}(0, \Sigma)$ with $r \times r$ covariance matrix satisfying $0 \preccurlyeq \Sigma \preccurlyeq I$. Define

$$L \coloneqq \max\left\{\left\|W_{\{1\},\{23\}}\right\|, \left\|W_{\{13\},\{2\}}\right\|\right\} .$$

Then for any $t \geq 0$,

$$\Pr_{\tilde{u}}\left[\left\|\left(I_p \otimes I_q \otimes \tilde{u}\right)W\right\| \geq t\right] \leq 4(p+q)e^{-\frac{t^2}{2L^2}}$$

- L serves as the Lipschitz parameter

Lipschitz property

- Define $A(u) \coloneqq (I_p \otimes I_q \otimes u)W = \sum_{k\in[r]} u_k W_k$
- For any $u, v \in \mathbb{R}^r$, we have

$$\|A(u) - A(v)\| = \left\| \sum_{k\in[r]} (u_k - v_k) W_k \right\|$$

$$= \sup_{\substack{x\in\mathbb{R}^p : \|x\|=1, \\ y\in\mathbb{R}^q : \|y\|=1}} \left| \sum_{k\in[r]} (u_k - v_k) \langle W_k, xy^\top \rangle \right|$$

$$\leq \sup_{\substack{x\in\mathbb{R}^p : \|x\|=1, \\ y\in\mathbb{R}^q : \|y\|=1}} \|u - v\| \cdot \left(\sum_{k\in[r]} \langle W_k, xy^\top \rangle^2 \right)^{1/2}$$

$$\sup_{\substack{x\in\mathbb{R}^p:\|x\|=1,\\ y\in\mathbb{R}^q:\|y\|=1}} \left(\sum_{k\in[r]} \langle W_k, xy^\top\rangle^2\right)^{1/2} = \sup_{\|x\|=1}\sup_{\|y\|=1}\left(\sum_{k\in[r]} (x^\top W_k y)^2\right)^{1/2}$$

$$\leq \sup_{\|x\|=1}\left(\sum_{k\in[r]} \sup_{\|y\|=1} (x^\top W_k y)^2\right)^{1/2}$$

$$= \sup_{\|x\|=1}\left(\sum_{k\in[r]} \left\|W_k^\top x\right\|^2\right)^{1/2} = \left\|W_{\{1\},\{23\}}\right\|$$

$$\sup_{\substack{x\in\mathbb{R}^p:\|x\|=1,\\ y\in\mathbb{R}^q:\|y\|=1}} \left(\sum_{k\in[r]} \langle W_k, xy^\top\rangle^2\right)^{1/2} = \sup_{\|y\|=1}\sup_{\|x\|=1}\left(\sum_{k\in[r]} (x^\top W_k y)^2\right)^{1/2}$$

$$\leq \sup_{\|y\|=1}\left(\sum_{k\in[r]} \sup_{\|x\|=1} (x^\top W_k y)^2\right)^{1/2}$$

$$= \sup_{\|y\|=1}\left(\sum_{k\in[r]} \|W_k y\|^2\right)^{1/2} = \left\|W_{\{13\},\{2\}}\right\|$$

$$\Longrightarrow \quad \|A(u) - A(v)\| \leq L \cdot \|u - v\|$$

$u \mapsto \|A(u)\|$ is L-Lipschitz

Back to the proof of Key Proposition

- Applying that theorem to $M(T, \tilde{u}) = (I_{n^2} \otimes I_{n^2} \otimes \tilde{u})W$, we have
$$\Pr_{\tilde{u}}[\|M(T, \tilde{u})\| \geq tL] \leq n^2 e^{-t^2/2} \ ,$$
where
$$L \coloneqq \max\left\{\left\|W_{ab,cd\,j_1 \ldots j_5}\right\|, \left\|W_{ab\,j_1 \ldots j_5,cd}\right\|\right\}$$
- By symmetry, we just consider the first one, and define $\widetilde{W} \in \mathbb{R}^{n^2 \times n^7}$:
$$\widetilde{W}_{ab,cd\,j_1 \ldots j_5} \coloneqq W_{ab,cd,j_1 \ldots j_5}$$
- We need to upper bound $\left\|\widetilde{W}\right\|$

Recap: the trace method

Theorem.

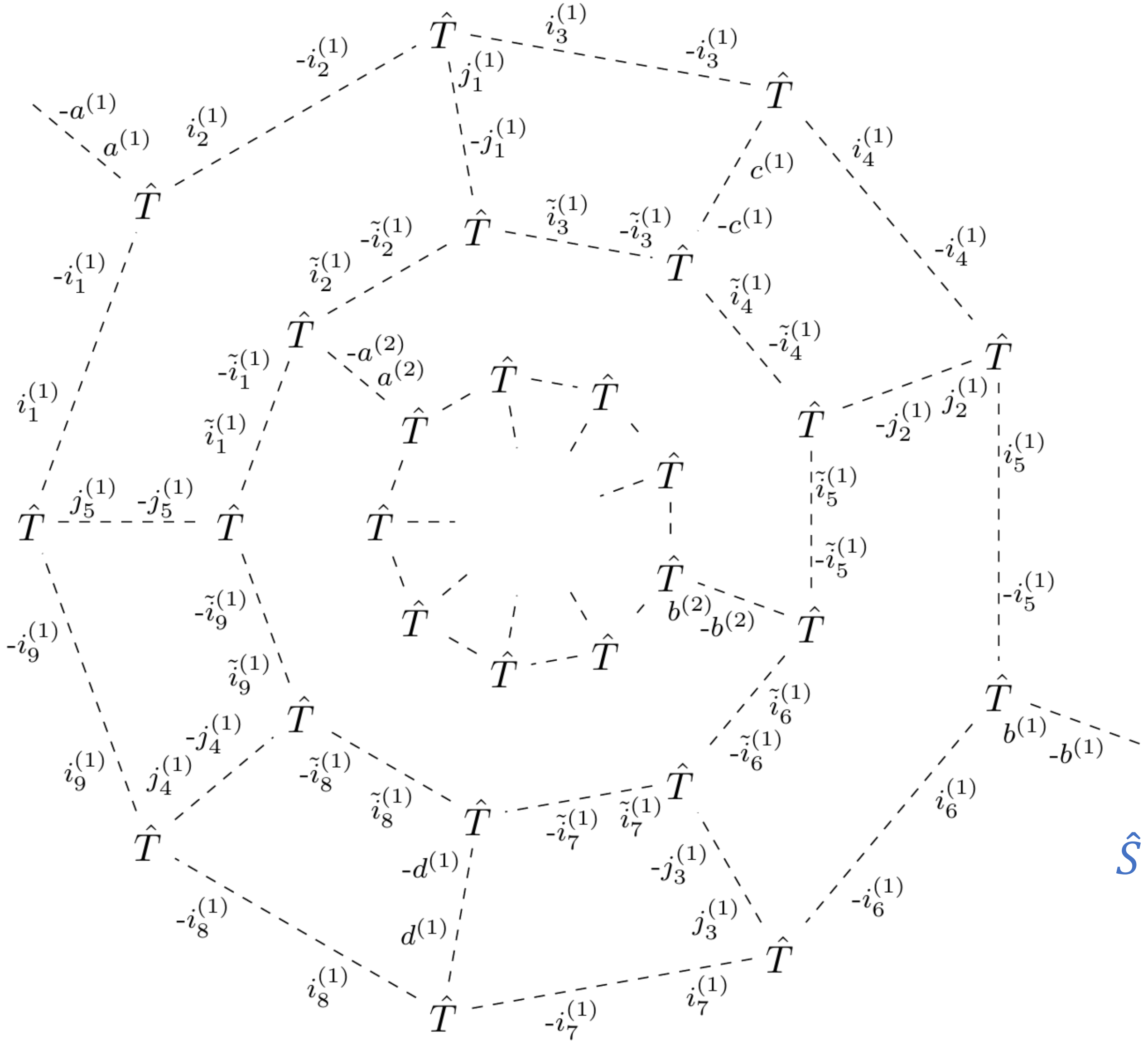
For any real-valued random matrix Y, for any integer $q \geq 1$ and any $\epsilon > 0$,

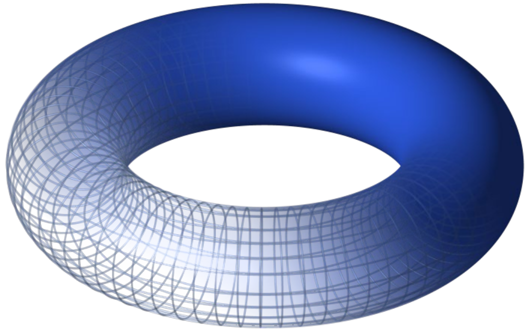
$$\Pr\left[\|Y\| > \left(\frac{\mathbb{E}\left[\mathrm{tr}[(YY^\top)^q]\right]}{\epsilon}\right)^{\frac{1}{2q}}\right] \leq \epsilon$$

- $\mathrm{tr}\left[\left(\widetilde{W}\widetilde{W}^\top\right)^q\right]$ can be represented as a huge TN $\mathcal{G}_q$ by connecting $2q$ copies of the TN for $\widetilde{W}$ in a ring
- Computing $\mathbb{E}\left[\mathrm{tr}\left[\left(\widetilde{W}\widetilde{W}^\top\right)^q\right]\right]$ is reduced to a combinatorial problem of labeling $\mathcal{G}_q$

Tensor network for the trace method

$\mathrm{tr}\left[\left(\widetilde{W}\widetilde{W}^{\top}\right)^{q}\right]$

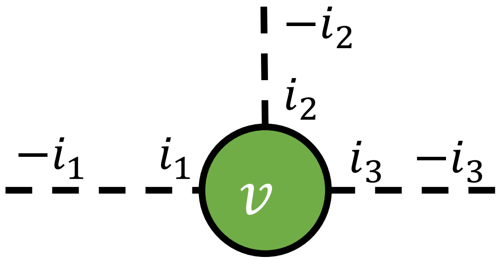




$\hat{S}$ is not shown

Labeling the TN graph

A labeling $\mathcal{L}$ of $\mathcal{G}_q$ is to assign every edge a pair of indices $(i_e, -i_e)$ for any $i_e \in [n/2]$, and assign every vertex v (i.e., $\hat{T}$ in the graph) an index $k_v \in [K]$.



$$\mathcal{L}(v) := \mathbf{1}_{i_1+i_2+i_3=0} x_{i_1}^{k_v} x_{i_2}^{k_v} x_{i_3}^{k_v}$$

$$\mathbb{E}\left[\mathrm{tr}\left[\left(\widetilde{W}\widetilde{W}^\top\right)^q\right]\right] = \mathbb{E}\left[\sum_{\mathcal{L}} S_{\mathcal{L}} \prod_{v} \mathcal{L}(v)\right]$$

where

$$S_{\mathcal{L}} := \prod_{l=1}^{q} S_{a^{(l)}b^{(l)}c^{(l)}d^{(l)}} \cdot S_{(-a^{(l+1)})(-b^{(l+1)})(-c^{(l)})(-d^{(l)})}$$

Labeling the TN graph

$$\mathbb{E}\left[\sum_{\mathcal{L}} S_{\mathcal{L}} \prod_{v} \mathcal{L}(v)\right] = \sum_{\mathcal{L}} S_{\mathcal{L}} \mathbb{E}\left[\prod_{v} \mathcal{L}(v)\right]$$

- Let $c(\mathcal{L})$ be the number of repeated labels:

$$c(\mathcal{L}) \coloneqq \sum_{i \in [p/2]} \max\{0, (\#\text{edges with label} \pm i) - 1\}$$

- The vertices with the same label k_v form a region. Let $r(\mathcal{L})$ be the number of regions:

$$r(\mathcal{L}) \coloneqq \#\text{distinct } k_v \text{ values}$$

- A labeling $\mathcal{L}$ is valid if:

1) $i_1 + i_2 + i_3 = 0$ at every vertex $\Rightarrow$ $\mathcal{L}(v) \neq 0$
2) $a^{(l)} \neq -b^{(l)}$ and $c^{(l)} \neq -d^{(l)}$ for every layer l $\Rightarrow$ $S_{\mathcal{L}} \neq 0$
3) Each region has as many incident i labels as incident $-i$ labels $\Rightarrow$ $\mathbb{E}[\ldots] \neq 0$

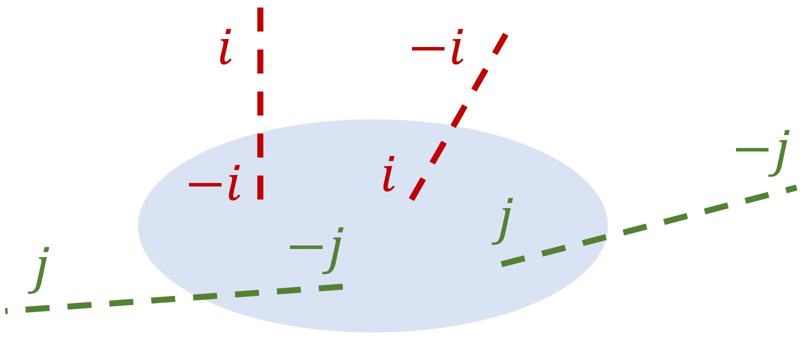
Labeling the TN graph

Claim. For any valid labeling $\mathcal{L}$ with $r(\mathcal{L}) > 1$, we have $c(\mathcal{L}) \geq r(\mathcal{L})/2$.

Proof.

- Since the number of regions > 1, and each region has an even number of cut edges (otherwise condition (3) is violated)
- So, the total number of cut edges is $\geq r(\mathcal{L})$
- Since every cut edge with label $(i, -i)$ can be matched to another edge with label $(-i, i)$, the number of repeated edges is at least $r(\mathcal{L})/2$

$\blacksquare$



Counting the valid labelings

Lemma. The number of valid edge-labelings with exactly $c(\mathcal{L})$ repeated labels is at most

$$q^{\mathcal{O}(c(\mathcal{L}))} n^{1+9q-c(\mathcal{L})/25}$$

Lemma. For a valid edge-labeling with c repeated labels, the number of valid vertex-labelings is at most

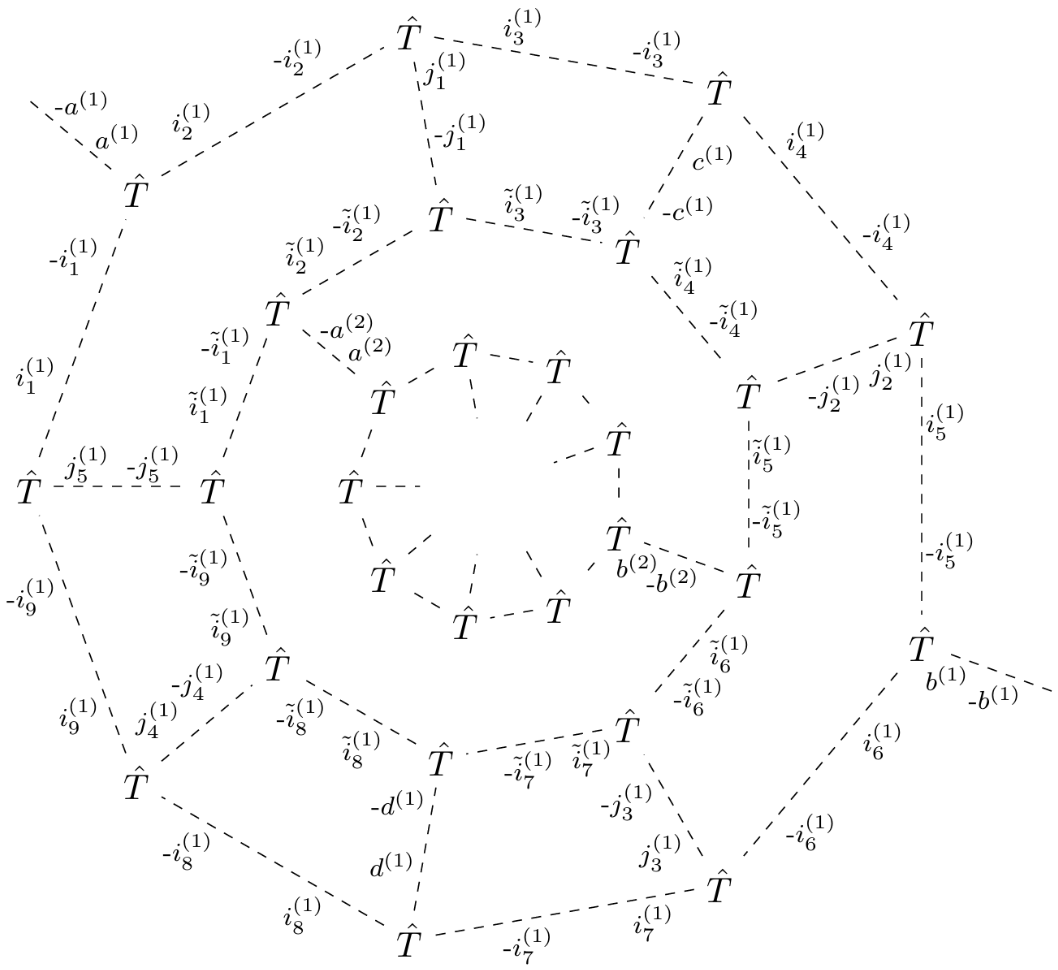
$$c(\mathcal{L})^{\mathcal{O}(q)} \exp\big(c(\mathcal{L})\big)$$

Proof.

- $r(\mathcal{L}) \leq 2c(\mathcal{L}) + 1$
- #vertices $= 18q$, and the number of ways to assign regions is $\leq (2c+1)^{18q}$
- #labelings for regions is $\leq K^{2c+1}$

$\blacksquare$

Linear constraints in the TN graph

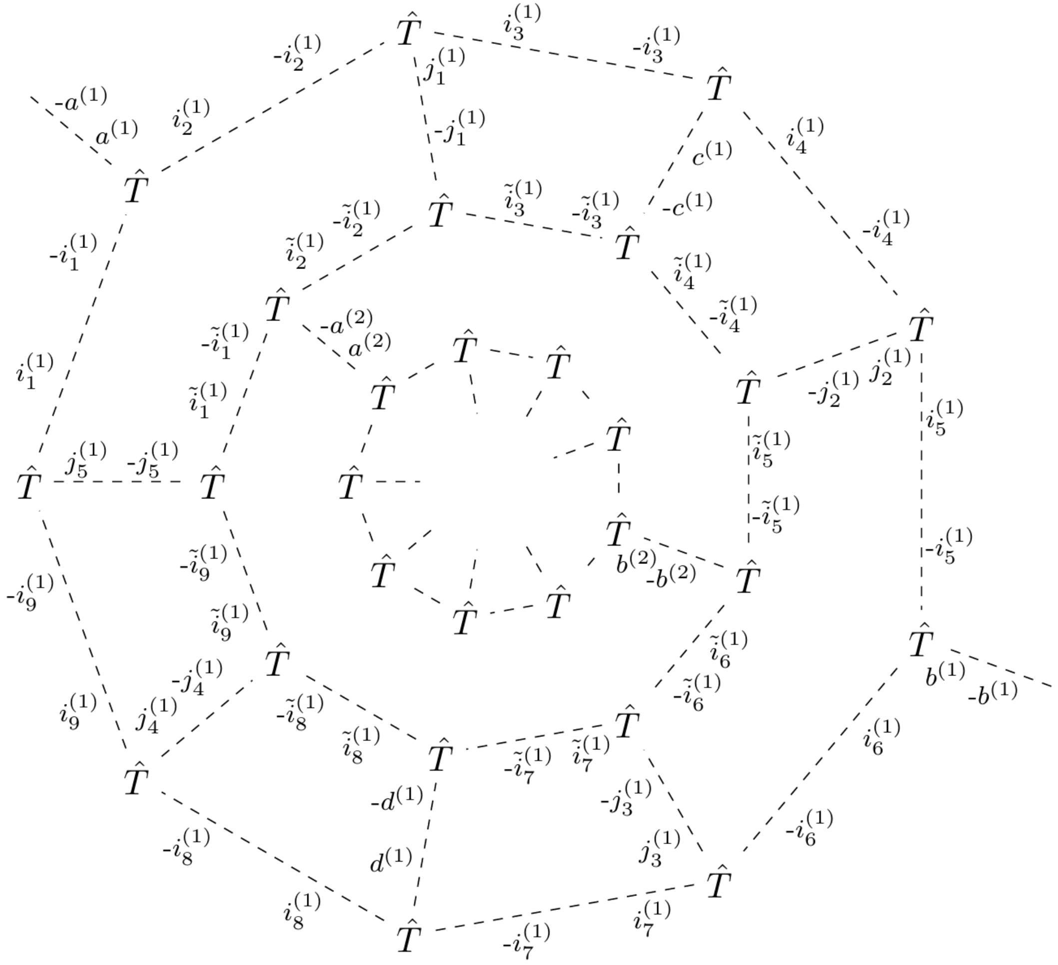


- $a^{(1)} + j_1^{(1)} + c^{(1)} + j_2^{(1)} + b^{(1)} + j_3^{(1)} + d^{(1)} + j_4^{(1)} + j_5^{(1)} = 0$

- $-a^{(2)} - j_1^{(1)} - c^{(1)} - j_2^{(1)} - b^{(2)} - j_3^{(1)} - d^{(1)} - j_4^{(1)} - j_5^{(1)} = 0$

$$\begin{aligned} w &:= a^{(l)} + b^{(l)} \\ &= -j_1^{(l)} - c^{(l)} - j_2^{(l)} - \boldsymbol{j_3}^{(l)} - \boldsymbol{d}^{(l)} - j_4^{(l)} - j_5^{(l)} \\ &= a^{(l+1)} + b^{(l+1)} \end{aligned}$$

Number of free labels



$$
\begin{aligned}
{\color{red}w} &:= a^{(l)} + b^{(l)} \\
&= -j_1^{(l)} - c^{(l)} - j_2^{(l)} - j_3^{(l)} - d^{(l)} - j_4^{(l)} - j_5^{(l)} \\
&= a^{(l+1)} + b^{(l+1)}
\end{aligned}
$$

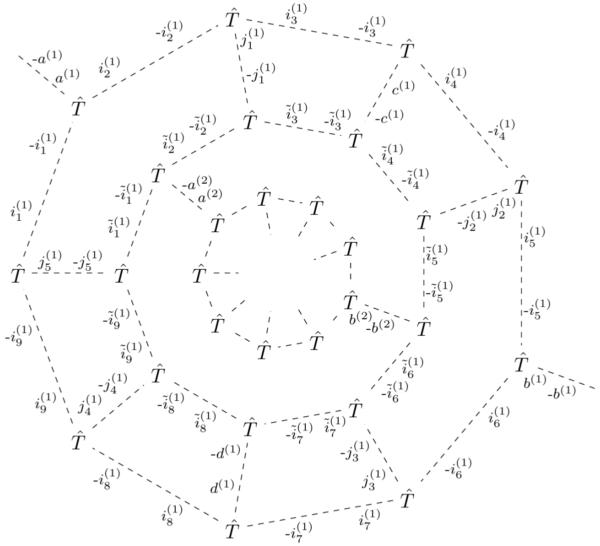
- w has $(2d + 1)$ choices
- $a^{(l)}$ is free
- $c^{(l)}, d^{(l)}, j_1^{(l)}, j_2^{(l)}, j_3^{(l)}, j_4^{(l)}$ are free
- $i_1^{(l)}$ is free
- $\tilde{\imath}_1^{(l)}$ is free
- So, each l has ${\color{red}9}$ free labels

Number of repetition patterns

- $\mathcal{L}$ is required to have $c(\mathcal{L})$ repeated edge labels
- Suppose we label the edges layer-by-layer, and within a layer, there is an ordering for the edges
- We say an edge is repeated if its label equal to a previous label up to sign in the ordering

repetition pattern:

- The total number of edges is $27q$
- At most $(27q)^{c(\mathcal{L})}$ ways to choose which edges are repeated
- At most $(2 \times 27q)^{c(\mathcal{L})}$ ways to choose which previous edges to repeat
- So, there are $q^{\mathcal{O}(c(\mathcal{L}))}$ repetition patterns



Repetition pattern reduces free labels

Fix w and fix a repetition pattern. We need to bound the number of labelings

- For each l, the labels are partitioned into two classes: $\{ab\}$ and $\{cdji\tilde{\imath}\}$
- Each class has ≤ 25 edges, while there are $c(\mathcal{L})$ repeated edges in total. So, at least $c(\mathcal{L})/25$ classes has repeated edges
- By a case study, you can check that each class with a repeated edge will has one fewer free label

If $\{a^{(l)}, b^{(l)}\}$ has a repeated edge, there are three cases:

- If $a^{(l)}$ or $b^{(l)}$ repeats a previous edge outside of this class, then the other label is determined, since $a^{(l)} + b^{(l)} = w$
- If $a^{(l)} = b^{(l)}$, then the labels are also determined $(= w/2)$
- If $a^{(l)} = -b^{(l)}$, then it violates the condition (2) and the labeling is not valid

Thus, $a^{(l)}$ is no longer a free label

Total number of valid edge-labelings

$$\underbrace{(2n+1)}_{w} \cdot \underbrace{q^{\mathcal{O}(c(\mathcal{L}))}}_{\text{\#repetition patterns}} \cdot \underbrace{n^{9q-c(\mathcal{L})/25}}_{\text{\#free labelings}} = q^{\mathcal{O}(c(\mathcal{L}))} n^{1+9q-c(\mathcal{L})/25}$$

Now, we are ready to compute the q-th moment

$$\mathbb{E}\left[\mathrm{tr}\left[\left(\widetilde{W}\widetilde{W}^{\top}\right)^{q}\right]\right] = \sum_{\text{valid }\mathcal{L}} S_{\mathcal{L}}\mathbb{E}\left[\prod_{v}\mathcal{L}(v)\right]$$

$$\leq (n^9)^{2q}\sum_{\text{valid }\mathcal{L}}\mathbb{E}\left[\prod_{v}\mathcal{L}(v)\right]$$

$$\leq n^{18q}\sum_{c=0}^{27q}\underbrace{\left(n^{-27q}\cdot(c+1)!\right)}_{\text{Gaussian moments}}\cdot\underbrace{\left(c^{\mathcal{O}(q)}\exp(c)\right)}_{\text{\#vertex-labelings}}\cdot\underbrace{\left(q^{\mathcal{O}(c)}n^{1+9q-c/25}\right)}_{\text{\#edge-labelings}}$$

$$= \exp(q)\cdot n\cdot\sum_{c=0}^{27q}\left(n^{-1/25}q^{\mathcal{O}(1)}\right)^{c}c^{q}$$

- If we set $q = \log n$, then we have

$$\mathbb{E}\left[\mathrm{tr}\left[\left(\widetilde{W}\widetilde{W}^{\top}\right)^{q}\right]\right] = n^{\mathcal{O}(1)}$$

- So,

$$\mathbb{E}\left[\mathrm{tr}\left[\left(\widetilde{W}\widetilde{W}^{\top}\right)^{q}\right]\right]^{\frac{1}{2q}} = n^{\frac{\mathcal{O}(1)}{\log n}} = \mathcal{O}(1)$$

Bound the noise term

Key Proposition.

There is an event $\mathcal{E}$ depending only on $\{x^k\}$ that happens with high probability over the randomness of $\{x^k\}$. Conditioned on $\mathcal{E}$, we have $\|M(T,\tilde{u})\| = \mathcal{O}\left(\sqrt{\log n}\right)$ with high probability over the randomness of $\tilde{u}$.

Proof.

- $\Pr\left[\|\widetilde{W}\| > \left(\frac{\mathbb{E}\left[\operatorname{tr}\left[\left(\widetilde{W}\widetilde{W}^\top\right)^{\log n}\right]\right]}{1/n}\right)^{1/(2\log n)}\right] \le \frac{1}{n}$
- We have with probability $1 - 1/n$, $\|\widetilde{W}\| = \mathcal{O}(1)$. Let $\mathcal{E}$ be this event
- Conditioned on $\mathcal{E}$,
$$\Pr_{\tilde{u}}[\|M(T,\tilde{u})\| \ge t\mathcal{O}(1)] \le n^2 e^{-t^2/2}$$
- Thus, with probability $1 - 1/\mathrm{poly}(n)$,
$$\|M(T,\tilde{u})\| = \mathcal{O}\left(\sqrt{\log n}\right)$$

■

Technical theorem.

There is a matrix $M(\mathcal{T}, u) \in \mathbb{R}^{n^2 \times n^2}$ (computable in poly-time from $\mathcal{T}$ and u) with the following guarantee. Let $v \in \mathbb{R}^{n^2}$ be the leading eigenvector of $(M(\mathcal{T}, u) + M(\mathcal{T}, u)^\top)/2$. Let $V = \mathrm{mat}(v) \in \mathbb{R}^{n \times n}$ and let $\tau \in \mathbb{R}^n$ be the top eigenvector of $(V + V^\top)/2$. With high probability over $\{x^k\}$, for any $k \in [K]$, we have $\langle \tau, x^k \rangle^2 \geq 0.99$ with probability $1/\mathrm{poly}(n)$.

Proof.

We have proved that, for any $k \in [K]$,

$$M(\mathcal{T}, u) = \alpha \underbrace{M\left(T^k, \left(x^k\right)^{\otimes 5}\right)}_{(1 \pm o(1))(x^k \otimes x^k)(x^k \otimes x^k)^\top} + \alpha \underbrace{\left(M\left(T, \left(x^k\right)^{\otimes 5}\right) - M\left(T^1, \left(x^k\right)^{\otimes 5}\right)\right)}_{\leq o(1)} + \underbrace{M(T, \tilde{u})}_{\mathcal{O}(\sqrt{\log n})} + \underbrace{(M(\mathcal{T}, u) - M(T, u))}_{\leq o(1)}$$

To conclude the proof,

1. Show that the top eigenvector of $M_{\mathrm{sym}} \coloneqq (M(\mathcal{T}, u) + M(\mathcal{T}, u)^\top)/2$ is close to $\left(x^k\right)^{\otimes 2}$
2. Show that the top eigenvector of $V_{\mathrm{sym}} \coloneqq (V + V^\top)/2$ is close to x^k

(Possibly in your problem set)

Today's plan

- Problem setup and examples
- Some easy algorithms
- Heterogeneous setting
  - Main result: a spectral algorithm for heterogeneous MRA
- The trace method
- The blueprint of the algorithm
- Proof: the signal part
- Proof: the noise part
- Generalizations

Generalizations

- Orbit recovery with other groups?
  - → **Bandeira-Blum-Smith-Kileel-Perry-Wein-Weed**: the sample complexity of list recovery for a compact group G is $\Theta(\sigma^{2d^\star})$ generically, where σ is the noise-level, and $d^\star$ is the degree of the invariant ring for G
  - → **Liu-Moitra**: Smoothed analysis for SO(3) orbit recovery (cryo-ET)

Invariant theory and orbit recovery

- The ring of invariant polynomials consists of all polynomials q that satisfy

$$q(g \cdot x) = q(x) \quad \forall g \in G$$

- The invariant ring determines x up to its orbit, i.e., $y \in \mathrm{orbit}(x) \Longleftrightarrow q(y) = q(x)\ \forall$ invariant q
- Method of moment with group symmetry: How many moments suffices $\Longleftrightarrow$ At what degree do invariant polynomials generate the full invariant ring