Non-backtracking tensor and long matrix completion

Ludovic Stephan

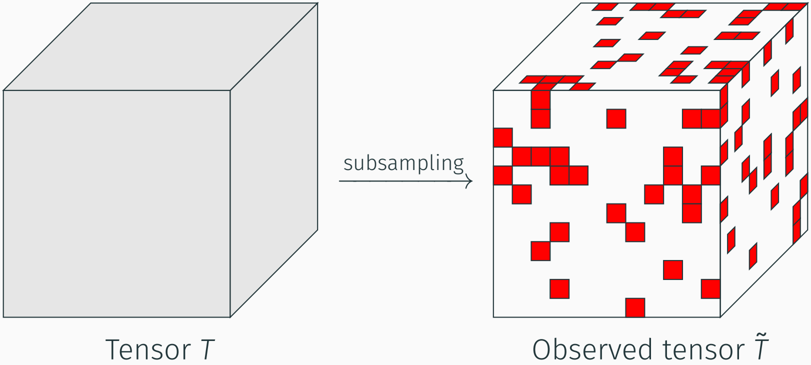
École Polytechnique Fédérale de Lausanne

Yizhe Zhu

University of California, Irvine

Problem setting

What is tensor completion ?



What is tensor completion ?

Formally:

- T is a tensor of size $n_1 \times \dots n_k$
- the observed tensor $\tilde{T}$ is defined as

$$\tilde{T}_{i_1,\dots,i_k} = \begin{cases} T_{i_1,\dots,i_k} & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

What is tensor completion ?

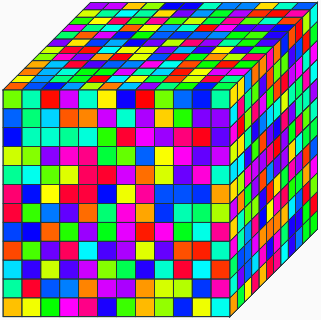
Formally:

- T is a tensor of size $n_1 \times \ldots n_k$
- the observed tensor $\tilde{T}$ is defined as

$$\tilde{T}_{i_1,\ldots,i_k} = \begin{cases} T_{i_1,\ldots,i_k} & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

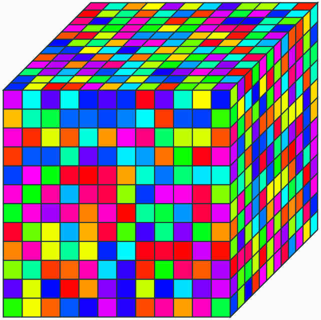
Goal: recover T from $\tilde{T}$

The problems begin

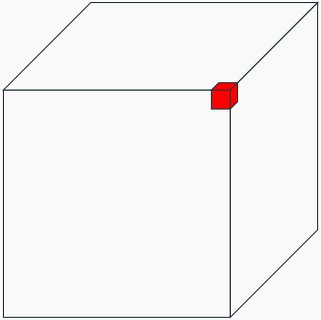


Too many degrees of freedom!

The problems begin

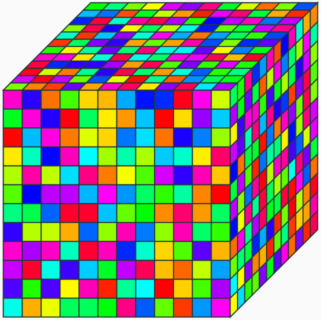


Too many degrees of freedom!

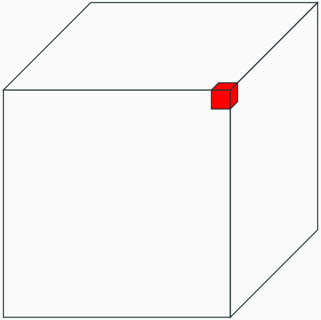


Too localized!

The problems begin



Too many degrees of freedom!



Too localized!

Both need $p = 1$

New assumptions

- T has low CP-rank:

$$T = \sum_{i=1}^{r} \lambda_i \left(w_i^{(1)} \otimes \cdots \otimes w_i^{(k)} \right)$$

$\Rightarrow r \times (n_1 + \cdots + n_k)$ degrees of freedom

New assumptions

- T has low CP-rank:

$$T = \sum_{i=1}^{r} \lambda_i \left(w_i^{(1)} \otimes \cdots \otimes w_i^{(k)} \right)$$

$\Rightarrow r \times (n_1 + \cdots + n_k)$ degrees of freedom

- T is delocalized:

$$\|T\|_\infty \simeq \left(\prod n_i\right)^{-1/2} \|T\|_F$$

or (a little stronger)

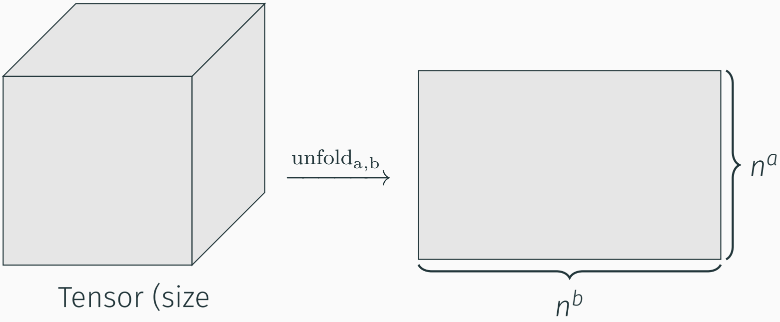
$$\|w_i^{(j)}\|_\infty \simeq n_i^{-1/2}$$

Unfoldings and hardness

Computational complexity problem: everything with tensors is hard [Hillar, Lim ’09]

- spectral norm
- eigenvalues/singular values
- low-rank approximations
- etc.

Unfoldings and hardness



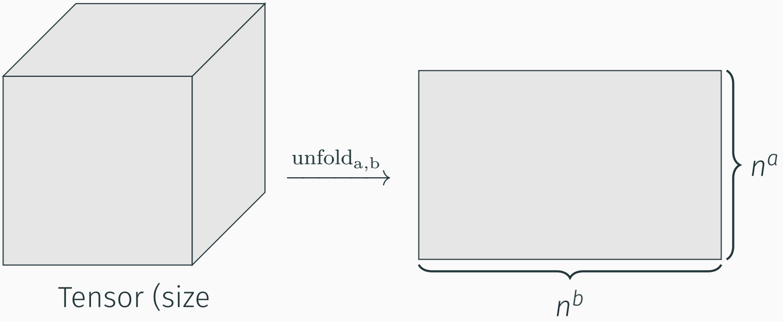
Tensor (size $n \times \cdots \times n$)

Unfolding matrix (size $n^a \times n^b$)

"Grouping" indices:

$$T_{i_1,\ldots,i_k} = M_{(i_1,\ldots,i_a),(i_{a+1},\ldots,i_k)}$$

Unfoldings and hardness



Tensor (size $n \times \cdots \times n$)

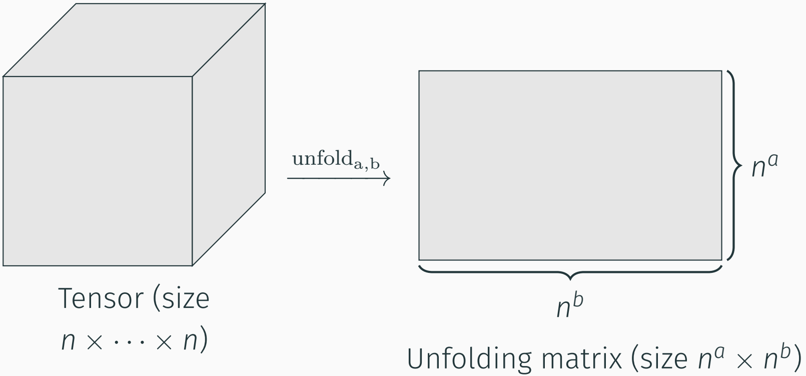
Unfolding matrix (size $n^a \times n^b$)

"Grouping" indices:

$$T_{i_1,\dots,i_k} = M_{(i_1,\dots,i_a),(i_{a+1},\dots,i_k)}$$

Tensor completion $\Leftarrow$ Long matrix completion

Unfoldings and hardness



Tensor (size $n \times \cdots \times n$)

Unfolding matrix (size $n^a \times n^b$)

"Grouping" indices:

$$T_{i_1,\ldots,i_k} = M_{(i_1,\ldots,i_a),(i_{a+1},\ldots,i_k)}$$

Tensor completion $\not\Rightarrow$ Long matrix completion

Unfoldings and hardness

Statistical-computational gap:

- NP-hard algorithms [Ghadermarzy et al '19]: tensor-based minimization methods
  $\rightarrow$ works until $p = O(n^{-(k-1)})$

Unfoldings and hardness

Statistical-computational gap:

- NP-hard algorithms [Ghadermarzy et al '19]: tensor-based minimization methods
  $\rightarrow$ works until $p = O(n^{-(k-1)})$

- Unfolding-based algorithms [Montanari and Sun '16, Liu and Moitra '20, Cai et al. '21...]
  $\rightarrow$ works until $p = O(n^{-k/2})$

Unfoldings and hardness

Statistical-computational gap:

- NP-hard algorithms [Ghadermarzy et al ’19]: tensor-based minimization methods
  $\rightarrow$ works until $p = O(n^{-(k-1)})$

- Unfolding-based algorithms [Montanari and Sun ’16, Liu and Moitra ’20, Cai et al. ’21...]
  $\rightarrow$ works until $p = O(n^{-k/2} \times \text{polylog}(n))$

What happens if $p \propto n^{-k/2}$?

Not a trivial improvement

Running example: rank-one tensor $T = x \otimes x \otimes x$, $x \in \{-1, 1\}^n$

Constant "degree": $p = dn^{-3/2}$

Typical algorithm:

- unfold $\tilde{T}$ into $A = \mathrm{unfold}_{1,2}(\tilde{T})$
- take the SVD of A (+ postprocessing)

Not a trivial improvement

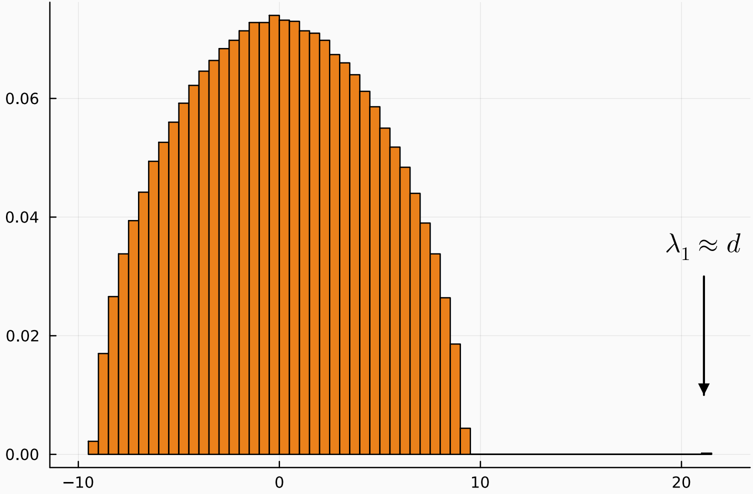


Figure: $AA^\top, d = 20$

Not a trivial improvement

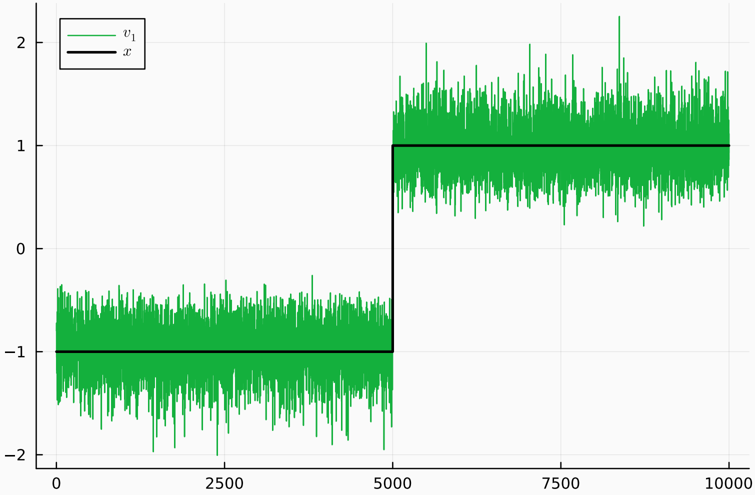


Figure: $AA^\top, d = 20$

Not a trivial improvement

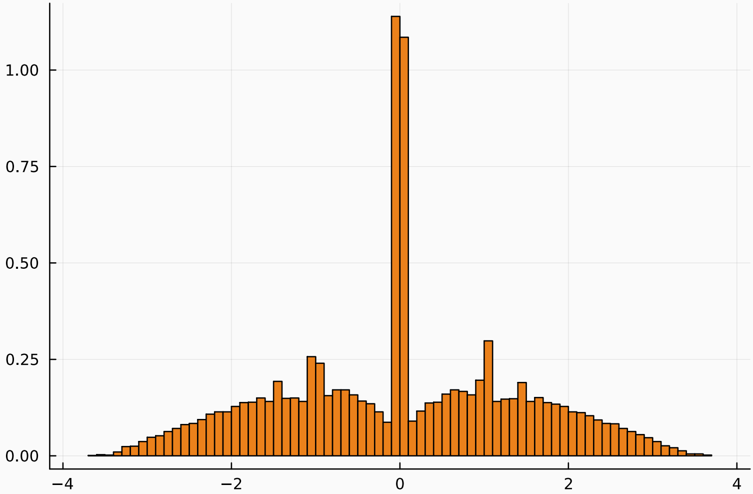


Figure: $AA^\top, d = 2$

Not a trivial improvement

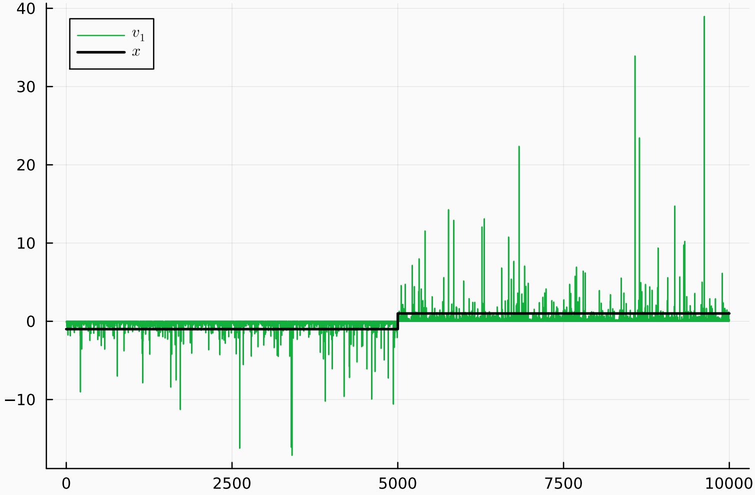


Figure: $AA^\top, d = 2$

Where are we so far ?

Recap:

- existing methods to not reach the exact conjectured threshold for tensor completion
- it is not a technical but a conceptual issue
- it suffices to solve long matrix completion

Our solution: non-backtracking wedge matrix

Setting

Long matrix reconstruction:

- Rectangular matrix M of size $n \times m$ ($m \gg n$), with SVD

$$M = \sum_{i=1}^{r} \nu_i \phi_i \psi_i^\top$$

- Masking matrix X with $X_{ij} \sim \mathrm{Ber}(p)$
- Observed matrix:

$$A = \frac{X \circ M}{p} \quad \text{so that} \quad \mathbb{E}[A] = M$$

Setting

Long matrix reconstruction:

- Rectangular matrix M of size $n \times m$ ($m \gg n$), with SVD

$$M = \sum_{i=1}^{r} \nu_i \phi_i \psi_i^\top$$

- Masking matrix X with $X_{ij} \sim \mathrm{Ber}(p)$
- Observed matrix:

$$A = \frac{X \circ M}{p} \quad \text{so that} \quad \mathbb{E}[A] = M$$

Assumptions:

$$r, \sqrt{n}\|\phi_i\|_\infty = O(\mathrm{polylog}(n))$$

Non-backtracking wedge matrix

We can view A as a weighted bipartite graph G with vertex sets $V_1 = [n]$ and $V_2 = [m]$.

The *non-backtracking wedge matrix* B is defined on *oriented wedges* in G

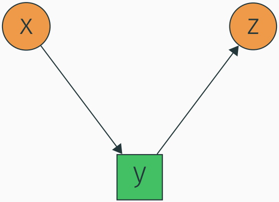
$$\vec{E}_2 = \{(x, y, z) \in V_1 \times V_2 \times V_1, z \neq x\}$$

Non-backtracking wedge matrix

We can view A as a weighted bipartite graph G with vertex sets $V_1 = [n]$ and $V_2 = [m]$.

The *non-backtracking wedge matrix* B is defined on *oriented wedges* in G

$$\vec{E}_2 = \{(x, y, z) \in V_1 \times V_2 \times V_1, z \neq x\}$$

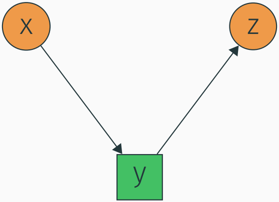


Non-backtracking wedge matrix

We can view A as a weighted bipartite graph G with vertex sets $V_1 = [n]$ and $V_2 = [m]$.

The *non-backtracking wedge matrix* B is defined on *oriented wedges* in G

$$\vec{E}_2 = \{(x, y, z) \in V_1 \times V_2 \times V_1, z \neq x\}$$



$\Rightarrow$ B has size $\sim d^2 n$: independent from m

Non-backtracking wedge matrix

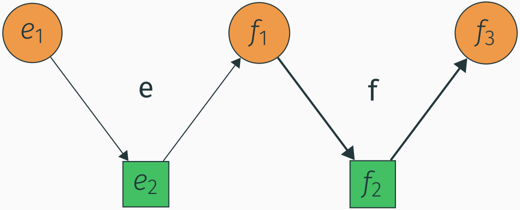
Defined as

$$B_{ef} = \begin{cases} A_{f_1 f_2} A_{f_3 f_2} & \text{if } e_3 = f_1 \text{ and } e_2 \neq f_2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Non-backtracking wedge matrix

Defined as

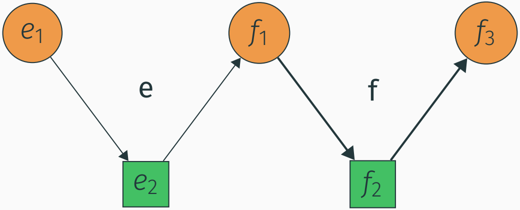
$$B_{ef} = \begin{cases} A_{f_1 f_2} A_{f_3 f_2} & \text{if } e_3 = f_1 \text{ and } e_2 \neq f_2 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$



Non-backtracking wedge matrix

Defined as

$$B_{ef} = \begin{cases} A_{f_1 f_2} A_{f_3 f_2} & \text{if } e_3 = f_1 \text{ and } e_2 \neq f_2 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$



Weight assignment is a convention !

Thresholds

Two important thresholds:

$$\vartheta_1 = \sqrt{\|\mathrm{Var}(A)\|}$$

- naturally occuring
- likely sharp
- decreases as $d^{-1/2}$

Thresholds

Two important thresholds:

$$\vartheta_1 = \sqrt{\|\mathrm{Var}(A)\|}$$

- naturally occuring
- likely sharp
- decreases as $d^{-1/2}$

$$\vartheta_2 = \|A\|_\infty$$

- less natural
- mostly spurious
- decreases as d^{-1}

Thresholds

Two important thresholds:

$$\vartheta_1 = \sqrt{\|\mathrm{Var}(A)\|}$$

- naturally occuring
- likely sharp
- decreases as $d^{-1/2}$

$$\vartheta_2 = \|A\|_\infty$$

- less natural
- mostly spurious
- decreases as d^{-1}

Total threshold:

$$\vartheta = \max(\vartheta_1, \vartheta_2)$$

Our results: eigenvalues

Theorem (S. and Zhu '23)

- *(Outliers) For any ν_i satistfying $\nu_i > \vartheta$, there exists an eigenvalue λ_i of B with*

$$|\lambda_i - \nu_i^2| = O(n^{-c})$$

Our results: eigenvalues

Theorem (S. and Zhu ’23)

- (**Outliers**) *For any ν_i satistfying $\nu_i > \vartheta$, there exists an eigenvalue λ_i of B with*

$$|\lambda_i - \nu_i^2| = O(n^{-c})$$

- (**Bulk**) *All other eigenvalues are asymptotically confined in a circle of radius ϑ^2*

Results: eigenvalues

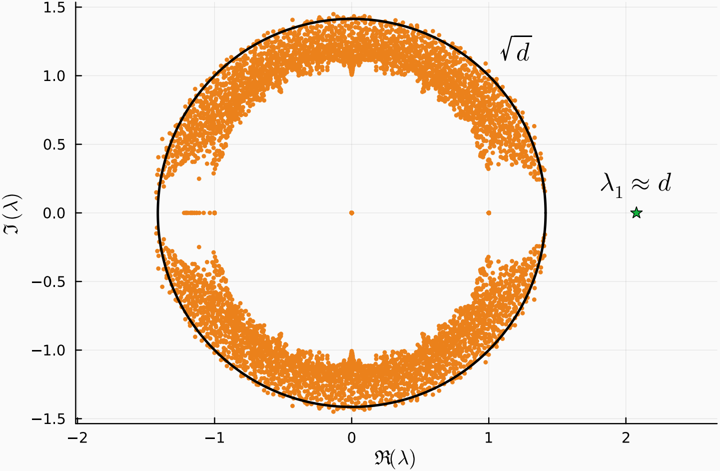


Figure: $B, d = 2$

Embedding eigenvectors

Need an embedding procedure from $\mathbb{R}^{\vec{E}_2}$ to $\mathbb{R}^n$

Embedding eigenvectors

Need an embedding procedure from $\mathbb{R}^{\vec{E}_2}$ to $\mathbb{R}^n$

- For a right eigenvector ξ^R:

$$\zeta^R(x) = \sum_{e:e_1=x} A_{e_1e_2}A_{e_3e_2}\xi^R(e)$$

- For a left eigenvector ξ^L:

$$\zeta^L(x) = \sum_{e:e_1=x} \xi^L(e)$$

Embedding eigenvectors

Need an embedding procedure from $\mathbb{R}^{\vec{E}_2}$ to $\mathbb{R}^n$

- For a right eigenvector ξ^R:

$$\zeta^R(x) = \sum_{e:e_1=x} A_{e_1e_2}A_{e_3e_2}\xi^R(e)$$

- For a left eigenvector ξ^L:

$$\zeta^L(x) = \sum_{e:e_1=x} \xi^L(e)$$

Depends on the weight convention !

Our results: eigenvectors

Theorem (S. and Zhu '23)

Assume that $\nu_i > \vartheta$, and let $\xi_i^{L/R}$ the left/right eigenvectors associated to λ_i. Then, there exists a γ_i such that

$$\gamma_i = 1 - O(d^{-1})$$

and

$$\left| \langle \zeta_i^{L/R}, \phi_i \rangle - \sqrt{\gamma_i} \right| = O(n^{-c})$$

Our results: eigenvectors

Theorem (S. and Zhu '23)

Assume that $\nu_i > \vartheta$, and let $\xi_i^{L/R}$ the left/right eigenvectors associated to λ_i. Then, there exists a γ_i such that

$$\gamma_i = 1 - O(d^{-1})$$

and

$$\left| \langle \zeta_i^{L/R}, \phi_i \rangle - \sqrt{\gamma_i} \right| = O(n^{-c})$$

Weak recovery when $d \to \infty$

Our results: eigenvectors

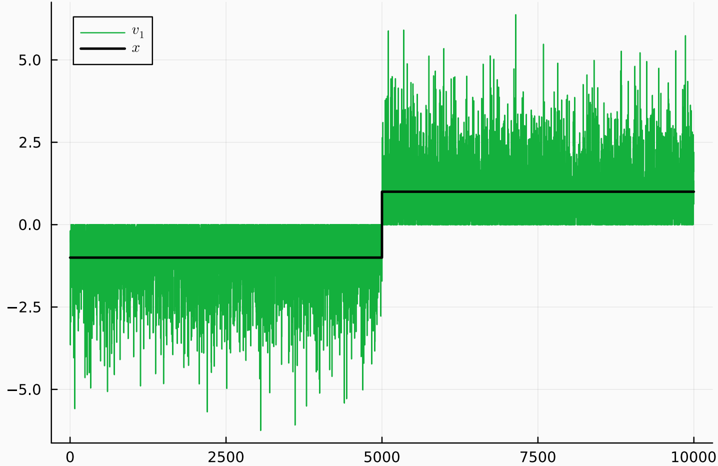
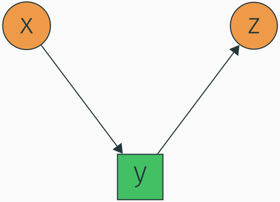


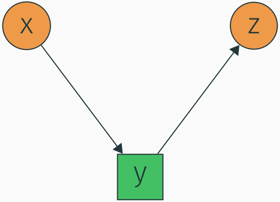
Figure: $B, d = 2$

Proof idea

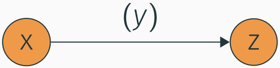


- $\deg(y) = 2$ w.h.p
- “independent” wedges
- Associated weight:
  $W_e = A_{xy}A_{zy}$

Proof idea



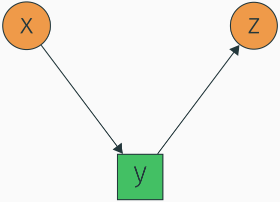
“$\Leftrightarrow$”



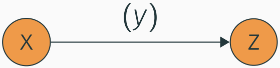
- $\mathbf{deg}(y) = 2$ w.h.p
- “independent” wedges
- Associated weight: $W_e = A_{xy}A_{zy}$

- Erdős-Rényi graph $G \sim \mathcal{G}(n, d^2/n)$
- Associated weight: $W_e = A_{xY}A_{zY}, Y \sim \mathbf{Unif}([m])$

Proof idea



“⇔”



- $\mathbf{deg}(y) = 2$ w.h.p
- “independent” wedges
- Associated weight: $W_e = A_{xy}A_{zy}$

- Erdős-Rényi graph $G \sim \mathcal{G}(n, d^2/n)$
- Associated weight: $W_e = A_{xY}A_{zY}, Y \sim \mathbf{Unif}([m])$

ER with random weights [Stephan and Massoulié ’20]

Takeaways and directions

- Does not work for finite aspect ratio ($m = O(n)$): is there a unified spectral algorithm with [Bordenave et al. ’21] ?

Takeaways and directions

- Does not work for finite aspect ratio ($m = O(n)$): is there a unified spectral algorithm with [Bordenave et al. '21] ?
- Our algorithm does not recover the right singular vectors of M (consistent with [Montanari and Wu '22])

Takeaways and directions

- Does not work for finite aspect ratio ($m = O(n)$): is there a unified spectral algorithm with [Bordenave et al. '21] ?
- Our algorithm does not recover the right singular vectors of M (consistent with [Montanari and Wu '22])
- We only handle the case $r = \mathrm{polylog}(n)$; what happens when $r = n^{\kappa}$? Can we reach the optimal sample complexity $rd^{3/2}$?

Thank you !

arxiv:2304.02077