Tensor complexity and completion using hypergraph expanders

Kameron Decker Harris

Paul G. Allen School of Computer Science & Engineering, Biology, UW
WWU Computer Science

Joint work with Yizhe Zhu, UCSD Math

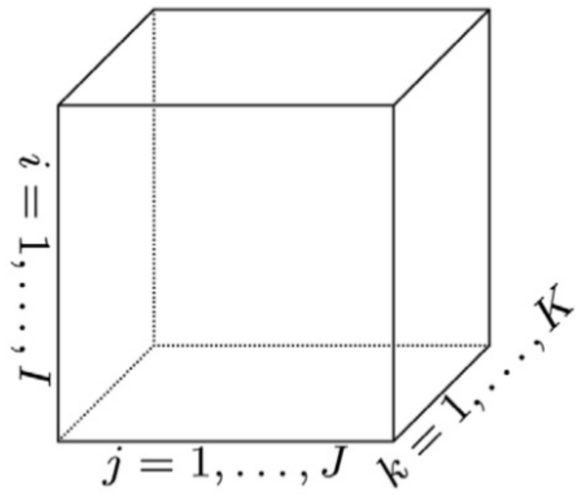
Overview

1) Introduction
2) Tensor complexity: norms & quasinorms
3) Hypergraph observation model
4) Tensor completion bound

Overview

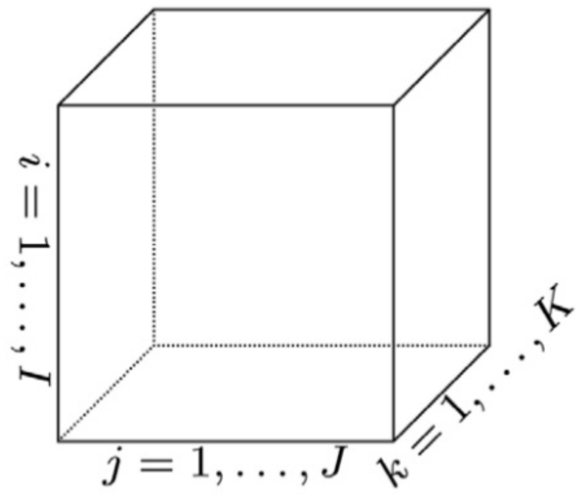
1) **Introduction**
2) Tensor complexity: norms & quasinorms
3) Hypergraph observation model
4) Tensor completion bound

What is a tensor?



A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

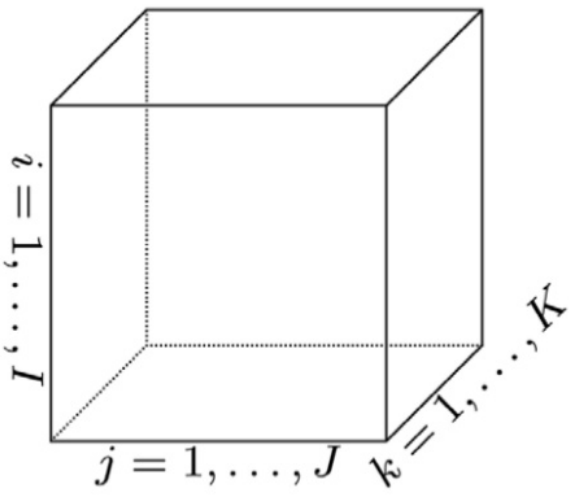
What is a tensor?



A third-order tensor: $\mathfrak{X} \in \mathbb{R}^{I \times J \times K}$

For us, a multi-dimensional array

What is a tensor?

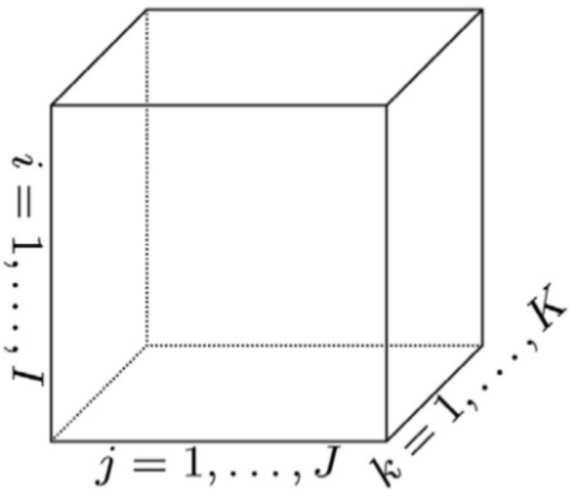


A third-order tensor: $\mathfrak{X} \in \mathbb{R}^{I \times J \times K}$

For us, a multi-dimensional array

Also:

What is a tensor?



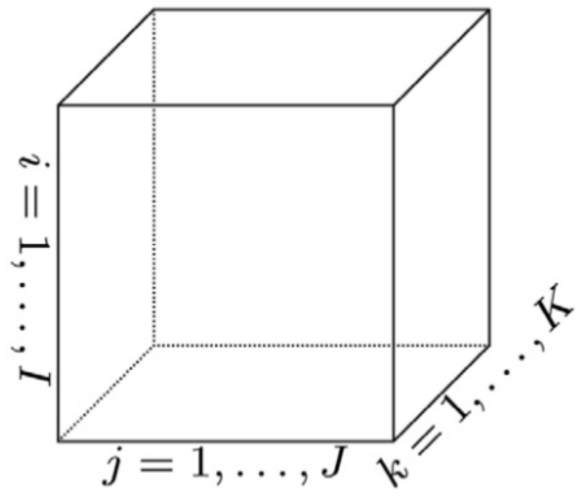
A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

For us, a multi-dimensional array

Also:

- multi-linear operator

What is a tensor?

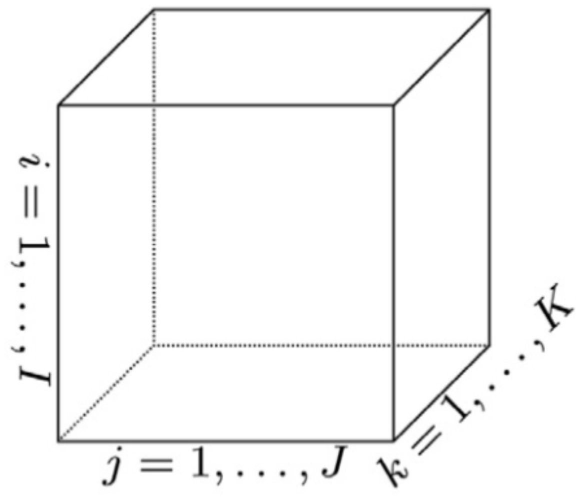


A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

For us, a multi-dimensional array

Also:
- multi-linear operator
- homogeneous polynomial

What is a tensor?



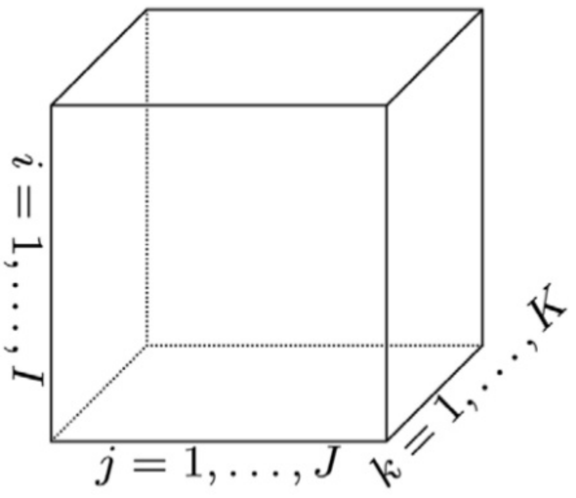
A third-order tensor: $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times K}$

For us, a multi-dimensional array

Also:

- multi-linear operator
- homogeneous polynomial
- **different** than a matrix

What is a tensor?



A third-order tensor: $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

For us, a multi-dimensional array

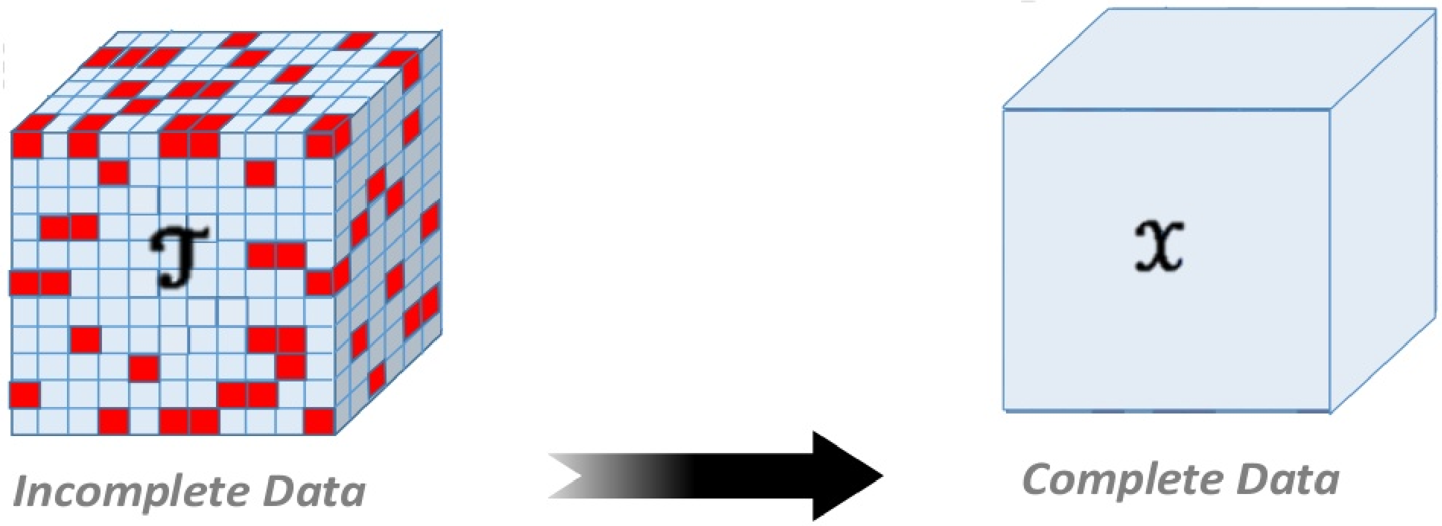
Also:

- multi-linear operator
- homogeneous polynomial
- **different** than a matrix

Excellent introduction: Kolda & Bader. SIAM Rev (2009)

Many problems cast as tensor completion

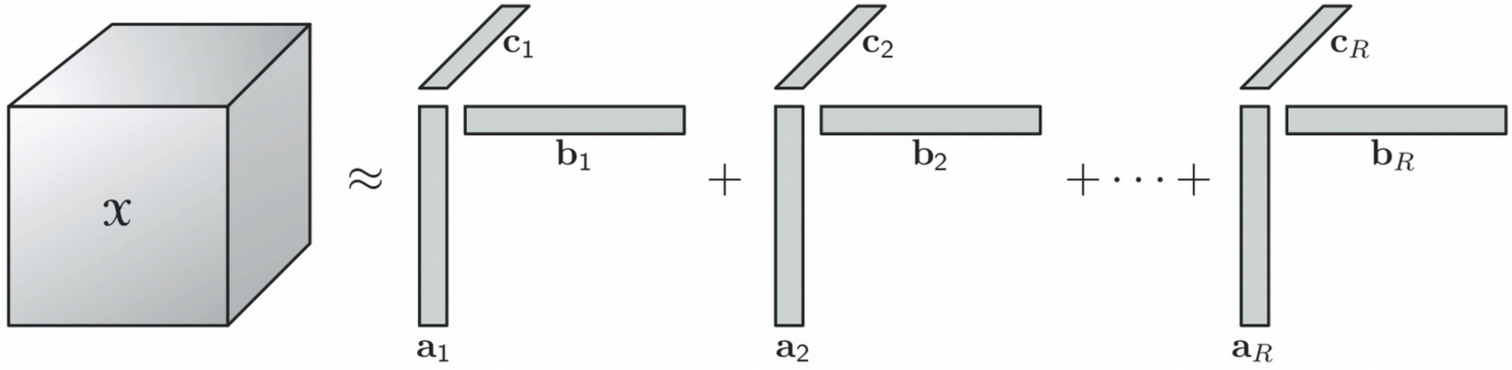
- Use low-rank structure to infer missing data



Song, Ge, Caverlee, Hu. KDD (2019)

But tensor rank is not like in matrices

- More than one version (CP/Kruskal, Tucker, unfolding rank)
- "Most tensor problems are **NP-hard**", Hillar & Lim. ACM (2013)



CP decomposition of a three-way array.

Kolda & Bader (2009)

Some of what’s known in completion problems

Some of what's known in completion problems

- Matrix results:
  - O(*nr* log(*n*)) samples suffice, fast spectral algorithms
  - Keshavan, Montanari, Oh. JMLR (2010); early work by Fazel, Recht, Candes, *inter alia*

Some of what's known in completion problems

- Matrix results:
  - O(nr log(n)) samples suffice, fast spectral algorithms
  - Keshavan, Montanari, Oh. JMLR (2010); early work by Fazel, Recht, Candes, *inter alia*
- Tensor results:
  - Flatten, O($n^{t/2}$ r polylog(n)) samples, matrix algorithms
  - Gandy, Recht, Yamada (2011); Mu et al. (2014); Montanari & Sun (2018)

Some of what's known in completion problems

- Matrix results:
  - O(nr log(n)) samples suffice, fast spectral algorithms
  - Keshavan, Montanari, Oh. JMLR (2010); early work by Fazel, Recht, Candes, *inter alia*
- Tensor results:
  - Flatten, O($n^{t/2}$ r polylog(n)) samples, matrix algorithms
  - Gandy, Recht, Yamada (2011); Mu et al. (2014); Montanari & Sun (2018)
  - SOS relaxation $t = 3$, O($n^{3/2}$ r polylog(n)) samples, poly algorithm
  - Barak & Moitra. JMLR (2016) … connection to random 3-SAT & hardness result

Some of what's known in completion problems

- Matrix results:
  - $O(nr \log(n))$ samples suffice, fast spectral algorithms
  - Keshavan, Montanari, Oh. JMLR (2010); early work by Fazel, Recht, Candes, *inter alia*
- Tensor results:
  - Flatten, $O(n^{t/2}\, r \,\text{polylog}(n))$ samples, matrix algorithms
  - Gandy, Recht, Yamada (2011); Mu et al. (2014); Montanari & Sun (2018)
  - SOS relaxation $t = 3$, $O(n^{3/2}\, r \,\text{polylog}(n))$ samples, poly algorithm
  - Barak & Moitra. JMLR (2016) … connection to random 3-SAT & hardness result
  - "Max-quasinorm" constraint, random observations, $O(nt\, ||T||^2_{\max})$ samples
  - Ghadermarzy, Plan, Yilmaz. Information & Inference (2018)

Some of what's known in completion problems

- Matrix results:
  - O(nr log(n)) samples suffice, fast spectral algorithms
  - Keshavan, Montanari, Oh. JMLR (2010); early work by Fazel, Recht, Candes, *inter alia*
- Tensor results:
  - Flatten, O($n^{t/2}$ r polylog(n)) samples, matrix algorithms
  - Gandy, Recht, Yamada (2011); Mu et al. (2014); Montanari & Sun (2018)
  - SOS relaxation t = 3, O($n^{3/2}$ r polylog(n)) samples, poly algorithm
  - Barak & Moitra. JMLR (2016) … connection to random 3-SAT & hardness result
  - "Max-quasinorm" constraint, random observations, O(nt $||T||^2_{max}$) samples
  - Ghadermarzy, Plan, Yilmaz. Information & Inference (2018)
  - Deterministic observations? Communication complexity & rank
  - Musco, Musco, Woodruff. Arxiv (2020) … polynomial algorithms **if** rank is relaxed, very relevant for us, but different objective

Outline of our results

Outline of our results

- Improved understanding of the *max-quasinorm*
  - New inequalities, rank bounds, relationship to an atomic norm

Outline of our results

- Improved understanding of the *max-quasinorm*
  - New inequalities, rank bounds, relationship to an atomic norm
- Hypergraph sampling model
  - Construction from expander graphs, new mixing inequality

Outline of our results

- Improved understanding of the *max-quasinorm*
  - New inequalities, rank bounds, relationship to an atomic norm
- Hypergraph sampling model
  - Construction from expander graphs, new mixing inequality
- Deterministic bound of generalization error for completion

$$\min_T \|T\|_{\max} \text{ s.t. } \|\Omega * (T - Z)\| \leq \delta$$

Tensor notation setup

$$T \in \bigotimes_{i=1}^{t} \mathbb{R}^n$$

order t

$$(T)_{i_1,\ldots,i_t} = \sum_{j=1}^{r} \underbrace{U^{(1)}_{i_1,j} \cdots U^{(t)}_{i_t,j}}_{\text{factor matrices}}$$

rank-r tensor (CP),
$r\ n\ t$ = # parameters

$$T = U^{(1)} \circ \cdots \circ U^{(t)}$$

shorthand for rank decomposition

$$T \otimes S, \qquad T * S$$

Kronecker, Hadamard products

Overview

1) Introduction
2) **Tensor complexity: norms & quasinorms**
3) Hypergraph observation model
4) Tensor completion bound

Matrix complexity

Matrix complexity

- Rank $$\mathrm{rank}(A) = \sum_i 1_{\{\sigma_i \neq 0\}}$$

Matrix complexity

- Rank

$$\mathrm{rank}(A) = \sum_i 1_{\{\sigma_i \neq 0\}}$$

- Nuclear/trace norm

$$\|A\|_* = \sum_i |\sigma_i| = \min_{A=UV^\intercal} \|U\|_F \|V\|_F = \min_{A=UV^\intercal} \frac{1}{2}(\|U\|_F^2 + \|V\|_F^2)$$

Srebro & Shraibman. COLT (2005)

Max-norm for matrices

$$\|A\|_{\max} = \min_{UV^\intercal = A} \|U\|_{2,\infty} \|V\|_{2,\infty} \text{ where } \|U\|_{2,\infty} = \max_i \|U_{i,:}\|_2$$

Max-norm for matrices

$$\|A\|_{\max} = \min_{UV^\intercal = A} \|U\|_{2,\infty}\|V\|_{2,\infty} \text{ where } \|U\|_{2,\infty} = \max_i \|U_{i,:}\|_2$$

Functional analysts call it $\gamma_2(A)$ ... factorization through L_2

Tomczak-Jaegermann 1989

Max-norm for matrices

$$\|A\|_{\max} = \min_{UV^\intercal = A} \|U\|_{2,\infty}\|V\|_{2,\infty} \text{ where } \|U\|_{2,\infty} = \max_i \|U_{i,:}\|_2$$

Functional analysts call it $\gamma_2(A)$ ... factorization through L_2

Tomczak-Jaegermann 1989

Communication complexity & discrepancy

Linial et al. 2007; Lee et al. 2008; Matoušek et al. 2014

Max-norm for matrices

$$\|A\|_{\max} = \min_{UV^\intercal = A} \|U\|_{2,\infty} \|V\|_{2,\infty} \text{ where } \|U\|_{2,\infty} = \max_i \|U_{i,:}\|_2$$

Functional analysts call it γ_2(A) ... factorization through L_2

Tomczak-Jaegermann 1989

Communication complexity & discrepancy

Linial et al. 2007; Lee et al. 2008; Matoušek et al. 2014

Matrix completion, incoherence, leverage

Srebro & Shraibman 2005; Heiman et al. 2014; Cai & Zhou 2016; Foucart et al. 2017

Max-quasinorm of a tensor

$$\|T\|_{\max} = \min_{T=U^{(1)}\circ\cdots\circ U^{(t)}} \prod_{i=1}^{t} \|U^{(i)}\|_{2,\infty}$$

Ghadermarzy, Plan, Yilmaz. Information & Inference (2018)

Max-quasinorm of a tensor

$$\|T\|_{\max} = \min_{T=U^{(1)}\circ\cdots\circ U^{(t)}} \prod_{i=1}^{t} \|U^{(i)}\|_{2,\infty}$$

Ghadermarzy, Plan, Yilmaz. Information & Inference (2018)

Lemma 1. Let $t \geq 2$, then any two order-t tensors T and S of the same shape satisfy the following properties:

1. $\|T\|_{\max} = 0$ if and only if $T = 0$.
2. $\|cT\|_{\max} = |c|\|T\|_{\max}$, where $c \in \mathcal{R}$.
3. $\|T + S\|_{\max} \leq \left(\|T\|_{\max}^{2/t} + \|S\|_{\max}^{2/t}\right)^{t/2} \leq 2^{t/2-1}\left(\|T\|_{\max} + \|S\|_{\max}\right)$.

p-norm, $p = 2/t$
Dilworth (1985)

quasi-triangle inequality

New results for max-qnorm

Theorem 1. *Let* $T \in \bigotimes_{i=1}^{t} {}^{n_i}$ *and* $S \in \bigotimes_{i=1}^{t} {}^{m_i}$, *then:*

1. $\|T_{I_1,\ldots,I_t}\|_{\max} \leq \|T\|_{\max}$ *for any subsets of indices* $I_i \subseteq [n_i]$
2. $\|T \otimes S\|_{\max} \leq \|T\|_{\max}\|S\|_{\max}$
3. $\|T * S\|_{\max} \leq \|T \otimes S\|_{\max}$, *where* $T, S \in \bigotimes_{i=1}^{t} {}^{n_i}$
4. $\|T * T\|_{\max} \leq \|T\|_{\max}^2$

Generalizes a number of results in
"A Direct Product Theorem for Discrepancy". Lee, Shraibman, Špalek (2008)

Comparison to $\ell_\infty^{n^t}$

Lower bound: $\|T\|_{\max} \geq \max_{1 \leq i \leq t} \|T_{[i]}\|_{\max} \geq |T|_\infty$

Comparison to $\ell_\infty^{n^t}$

Lower bound: $\|T\|_{\max} \geq \max_{1 \leq i \leq t} \|T_{[i]}\|_{\max} \geq |T|_\infty$

Ghadermarzy et al. showed $\|T\|_{\max} \leq \sqrt{r^{t(t-1)}} \cdot |T|_\infty$

Comparison to $\ell_\infty^{n^t}$

Lower bound: $\|T\|_{\max} \geq \max_{1 \leq i \leq t} \|T_{[i]}\|_{\max} \geq |T|_\infty$

Ghadermarzy et al. showed $\|T\|_{\max} \leq \sqrt{r^{t(t-1)}} \cdot |T|_\infty$

But for a matrix, $\|A\|_{\max} \leq \sqrt{r} \cdot |A|_\infty$

Comparison to $\ell_\infty^{n^t}$

Lower bound: $\|T\|_{\max} \geq \max_{1 \leq i \leq t} \|T_{[i]}\|_{\max} \geq |T|_\infty$

Ghadermarzy et al. showed $\|T\|_{\max} \leq \sqrt{r^{t(t-1)}} \cdot |T|_\infty$

But for a matrix, $\|A\|_{\max} \leq \sqrt{r} \cdot |A|_\infty$

A guess: $\|T\|_{\max} \leq \sqrt{r^{t-1}} \cdot |T|_\infty$

Sign tensors & another norm

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix} = \begin{bmatrix} + \\ - \\ + \end{bmatrix} \begin{bmatrix} + & - & + \end{bmatrix}$$

rank-1 sign tensor

Sign nuclear norm:

$$\|T\|_{\pm} = \inf \left\{ \sum_{i=1}^{r} |\alpha_i| \;\middle|\; T = \sum_{i=1}^{r} \alpha_i S_i \text{ where } \alpha_i \in \mathbb{R}, \, \mathrm{rank}_{\pm}(S_i) = 1 \right\}$$

Relation between sign and max-qnorm

Lemma 2. *The sign nuclear norm and max-qnorm satisfy*

$$\|T\|_{\pm} \leq K_G^{t-1} \|T\|_{\max},$$

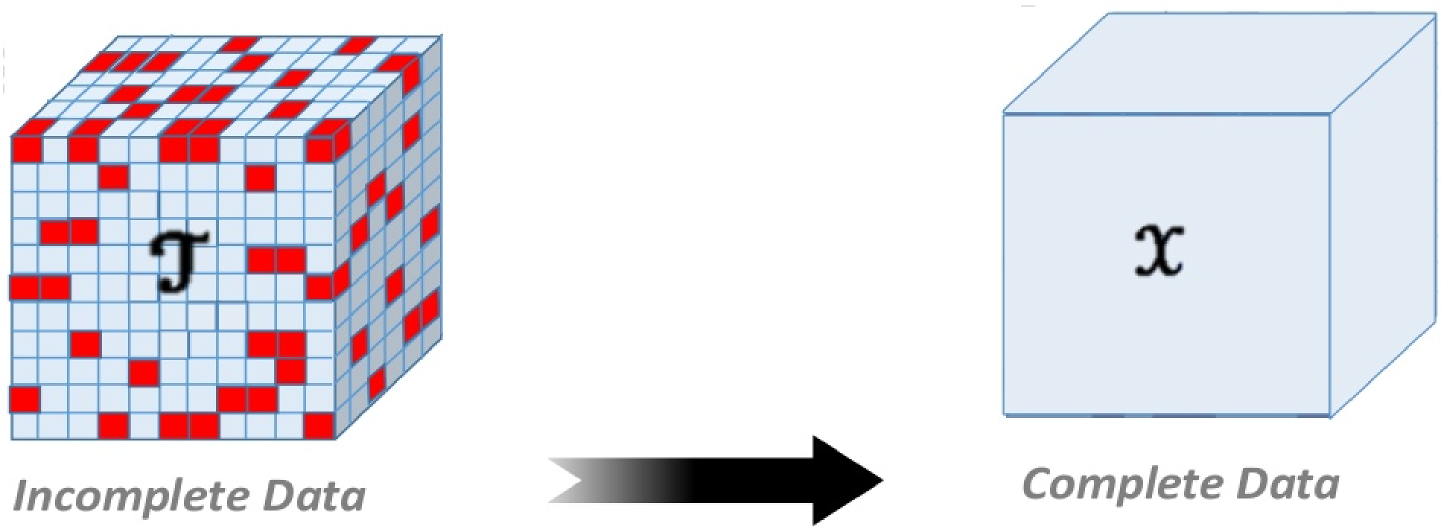
where K_G is the Grothendieck constant over $\mathbb{R}$.

(Tightens a result by Ghadermarzy et al.)

Overview

1) Introduction
2) Tensor complexity: norms & quasinorms
3) **Hypergraph observation model**
4) Tensor completion bound

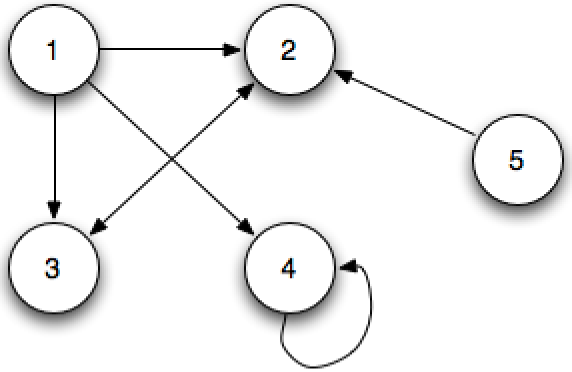
Modeling observations



sparse binary mask

Image: Song, Ge, Caverlee, Hu. KDD (2019)

Adjacency matrix of a graph



	1	2	3	4	5
1	0	1	1	1	0
2	0	0	1	0	0
3	0	1	0	0	0
4	0	0	0	1	0
5	0	1	0	0	0

Image: David Babcock, http://ycpcs.github.io/cs360-spring2015/lectures/lecture15.html

Hypergraph → Adjacency tensor

Properties we require

- *t*-uniform: all hyperedges contain *t* vertices
- *t*-partite: non-symmetric tensors

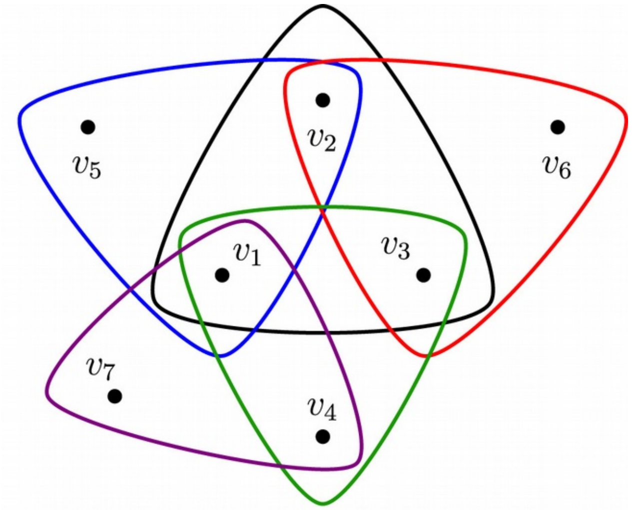


Image: Akhtar & Maity. Disc Appl Math (2017)

Lofty goal: *H* has "good" mixing

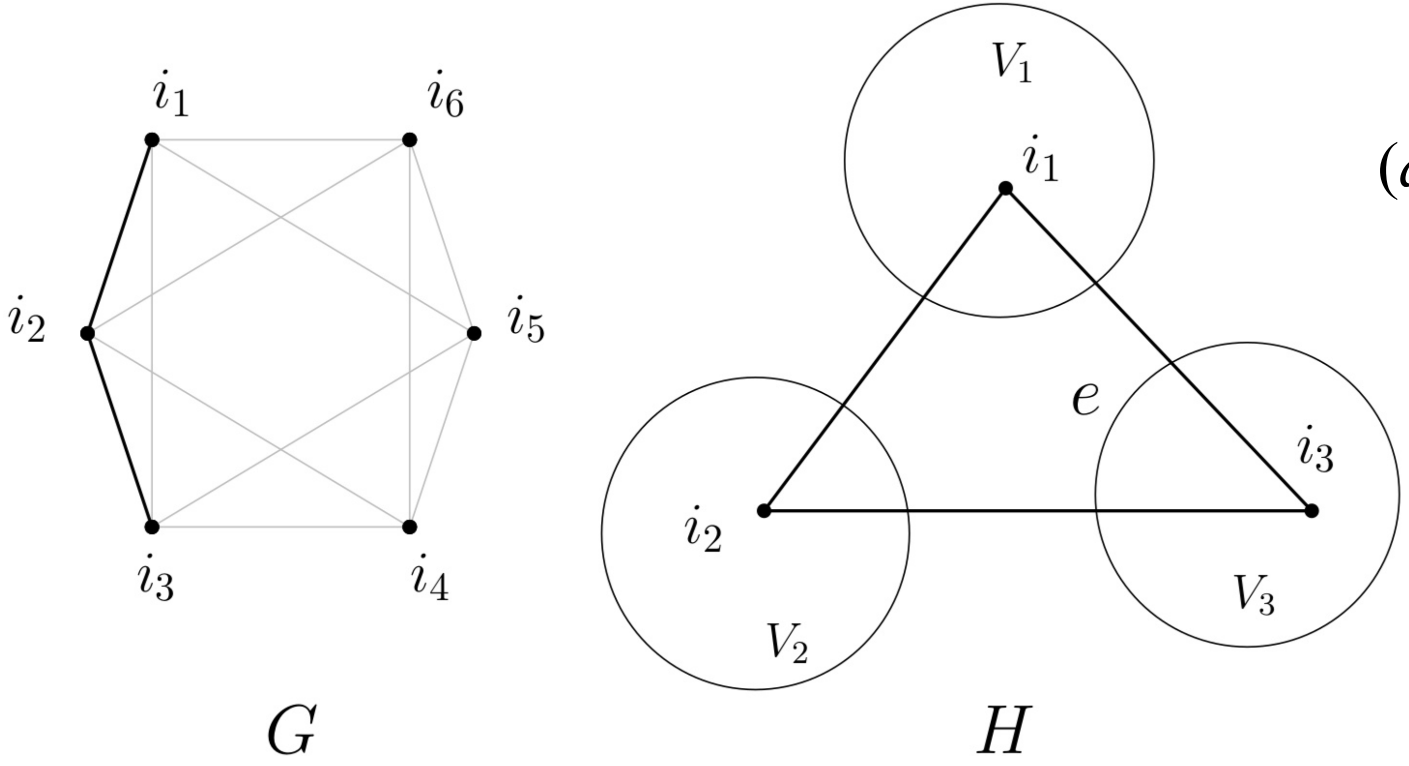
$$\left|\, e(V_1, \ldots, V_t) - \alpha |V_1| \cdots |V_t| \,\right| \leq \lambda \sqrt{|V_1| \cdots |V_t|}$$

$$\lambda = \|T - \alpha J\|_\sigma$$

"Second eigenvalue" of *T*(*H*)

$$\|T\|_\sigma = \sup_{v_1, \ldots, v_t \in S^{n-1}} \left| \sum_{i_1, \ldots, i_t = 1}^{n} T_{i_1, \ldots, i_t} v_1(i_1) \cdots v_t(i_t) \right|$$

Expander construction



(d = 4, t = 3)

has $nd^{t\text{-}1}$ many edges

Alon et al. Computational Complexity (1995)
Bilu & Hoory. EJ Combinatorics (2004)

(Weak) mixing lemma

Lemma 3. *Construct H from the d-regular base graph G. Then:*

$$\left|\frac{e(W_1,\ldots,W_t)}{nd^{t-1}} - \prod_{i=1}^{t}\alpha_i\right| \le \left(\left(1+\frac{\lambda}{d}\right)^{t-1} - 1\right)\min\left\{\frac{1}{4}, \prod_{i=1}^{t}\max\left\{\sqrt{\alpha_i}, \sqrt{1-\alpha_i}\right\}\right\}$$

where

$$\alpha_i = \frac{|W_i|}{n} \qquad \text{and} \qquad \lambda = \lambda_2(G).$$

Tightens results from Alon et al. (1995); Bilu & Hoory (2004)

Overview

1) Introduction
2) Tensor complexity: norms & quasinorms
3) Hypergraph observation model
4) **Tensor completion bound**

Max-qnorm penalized guarantee

- Suppose we can *solve*

$$\hat{T} = \text{argmin}_{T'} \|T'\|_{\max} \text{ s.t. } \|\Omega * (T' - Z)\|_F \leq \delta$$

Ghadermarzy, Plan, Yilmaz (2018)
Heiman et al. (2014)
Brito, Dumitriu, Harris (2018)

Max-qnorm penalized guarantee

- Suppose we can *solve*

$$\hat{T} = \text{argmin}_{T'} \|T'\|_{\max} \text{ s.t. } \|\Omega * (T' - Z)\|_F \leq \delta$$

observation mask
i.e. adjacency tensor

Ghadermarzy, Plan, Yilmaz (2018)
Heiman et al. (2014)
Brito, Dumitriu, Harris (2018)

Max-qnorm penalized guarantee

- Suppose we can *solve*

$$\hat{T} = \text{argmin}_{T'} \|T'\|_{\max} \text{ s.t. } \|\Omega * (T' - Z)\|_F \leq \delta$$

observation mask
i.e. adjacency tensor

data

Ghadermarzy, Plan, Yilmaz (2018)
Heiman et al. (2014)
Brito, Dumitriu, Harris (2018)

Max-qnorm penalized guarantee

- Suppose we can *solve*

noise level

$$\hat{T} = \text{argmin}_{T'} \|T'\|_{\max} \text{ s.t. } \|\Omega * (T' - Z)\|_F \leq \delta$$

observation mask
i.e. adjacency tensor

data

Ghadermarzy, Plan, Yilmaz (2018)
Heiman et al. (2014)
Brito, Dumitriu, Harris (2018)

Max-qnorm penalized guarantee

- Suppose we can *solve*

noise level

$$\hat{T} = \mathrm{argmin}_{T'} \|T'\|_{\max} \text{ s.t. } \|\Omega * (T' - Z)\|_F \leq \delta$$

observation mask
i.e. adjacency tensor

data

Theorem 2:

$$\frac{1}{n^t}\|\hat{T} - T\|_F^2 \leq (4K_G)^{t-1}\|T\|_{\max}^2 \left(\left(1 + \frac{\lambda}{d}\right)^{t-1} - 1\right) + 4\delta^2$$

Ghadermarzy, Plan, Yilmaz (2018)
Heiman et al. (2014)
Brito, Dumitriu, Harris (2018)

Sample complexity

- # observations $|E(H)| = nd^{t-1}$
- use an expander G $\implies \frac{\lambda}{d} = \frac{1}{\sqrt{d}}$

Required # samples:

$$|E| = n\left(\frac{(t-1)(4K_G)^{t-1}\|T\|_{\max}^2}{\varepsilon}\right)^{2(t-1)}$$

Proof sketch (part 1)

Proof sketch (part 1)

1) Write a tensor Q decomposed into sign tensors

Proof sketch (part 1)

1) Write a tensor Q decomposed into sign tensors
2) Interpret rank-1 sign tensor as indicator of sets

Proof sketch (part 1)

1) Write a tensor Q decomposed into sign tensors
2) Interpret rank-1 sign tensor as indicator of sets
3) Use mixing lemma to bound variation from sample

$$\left| \frac{1}{n^t} \sum_{e \in [n]^t} Q_e - \frac{1}{nd^{t-1}} \sum_{e \in E} Q_e \right| \leq 2^{t-2} \left(\left(1 + \frac{\lambda}{d}\right)^{t-1} - 1 \right) \|Q\|_{\pm}$$

Proof sketch (part 2)

1) Take $Q = (\hat{T} - T) * (\hat{T} - T)$ squared residuals

2) Then we have:

$$\|Q\|_{\pm} \leq 2^{t-2} K_G^{t-1} \left(\|\hat{T}\|_{\max} + \|T\|_{\max} \right)^2$$
$$\leq 2^t K_G^{t-1} \|T\|_{\max}^2$$

Conclusions

- New inequalities for tensor complexity measures
- Better weak mixing for expander hypergraph
- New analysis of tensor completion
  - Linear dependence on n
  - Better bound possible with 2^{nd} eigenvalue of H

Computational challenges

Computational challenges

- Difficulty of optimization?

Computational challenges

- Difficulty of optimization?
- Barak & Moitra. JMLR (2016): SOS + $\|\cdot\|_*$
  - random observations ↔ random 3-SAT, *conjectured* hard
  - sign tensor = solution, observations = clauses

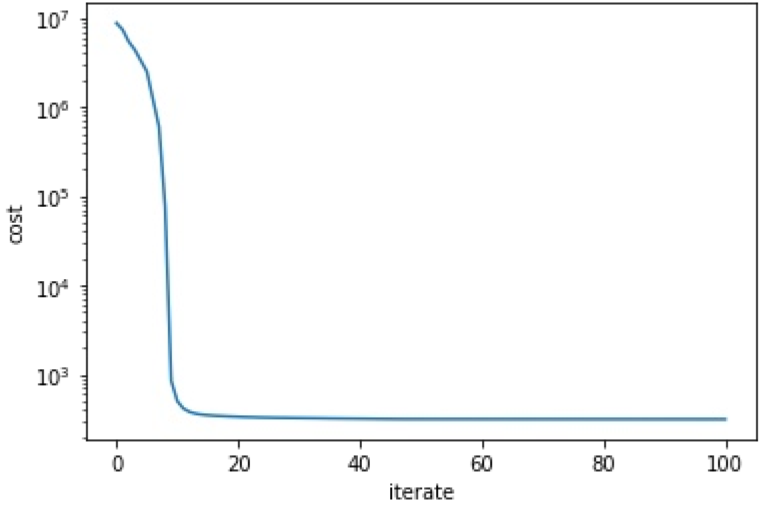
Computational challenges

- Difficulty of optimization?
- Barak & Moitra. JMLR (2016): SOS + $\|\cdot\|_*$
  - random observations ↔ random 3-SAT, *conjectured* hard
  - sign tensor = solution, observations = clauses
- Approx solution, get bound w/ $\max(\|\hat{T}\|_{\max}, \|T\|_{\max})$

Computational challenges

- Difficulty of optimization?
- Barak & Moitra. JMLR (2016): SOS + $\|\cdot\|_*$
  - random observations ↔ random 3-SAT, *conjectured* hard
  - sign tensor = solution, observations = clauses
- Approx solution, get bound w/ $\max(\|\hat{T}\|_{\max}, \|T\|_{\max})$
- Suggests even approx solutions hard!

Early numerical experiments



$n = 200,\ t = 3,\ r = 4,\ |E| = 55224$

$\sim 0.7\%$

Coordinate descent on factors $U^{(i)}$

- $r > \text{rank}(T)$ helps
- overfit residuals okay
- stagnation for low rank, too few observations

Thank you!

- Acknowledgements:
  - **coauthor: Yizhe Zhu**
  - Ioana Dumitriu
  - Paul Beame
  - Adi Shraibman

- Funding:
  - NSF DMS-1712630 (YZ)

Washington Research
FOUNDATION

"Deterministic tensor completion with hypergraph expanders"

https://arxiv.org/abs/1910.10692

Ideas for $\|T\|_{\max} \leq \sqrt{r^t} \cdot |T|_\infty$

Ideas for $\|T\|_{\max} \leq \sqrt{rt} \cdot |T|_\infty$

- Banach-Mazur distance:

$$d(X,Y) = \inf\{\|L\|_{X\to Y}\|L^{-1}\|_{Y\to X} \ : \ L \text{ isomorphism}\}$$

Ideas for $\|T\|_{\max} \leq \sqrt{r^t} \cdot |T|_\infty$

- Banach-Mazur distance:

$$d(X,Y) = \inf\{\|L\|_{X\to Y}\|L^{-1}\|_{Y\to X} \;:\; L \text{ isomorphism}\}$$

- By Dilworth (1985), *r*-dim *p*-norm space $d(X, \ell_2) \leq r^{\frac{2}{p}-\frac{1}{2}}$

Ideas for $\|T\|_{\max} \leq \sqrt{r^t} \cdot |T|_\infty$

- Banach-Mazur distance:

$$d(X,Y) = \inf\{\|L\|_{X\to Y}\|L^{-1}\|_{Y\to X} \ : \ L \text{ isomorphism}\}$$

- By Dilworth (1985), *r*-dim *p*-norm space $d(X,\ell_2) \leq r^{\frac{2}{p}-\frac{1}{2}}$
- Let $Z = \operatorname{span}\left\{U^{(1)}_{:,i} \circ \cdots \circ U^{(t)}_{:,i}\right\}_{i=1}^{r}, \quad X = \left(\mathbb{R}^{n^t}, \|\cdot\|_{\max}\right) \cap Z, \quad Y = \left(\mathbb{R}^{n^t}, |\cdot|_\infty\right) \cap Z$

Ideas for $\|T\|_{\max} \le \sqrt{r^t} \cdot |T|_\infty$

- Banach-Mazur distance:

$$d(X,Y) = \inf\{\|L\|_{X\to Y}\|L^{-1}\|_{Y\to X} \ : \ L \text{ isomorphism}\}$$

- By Dilworth (1985), *r*-dim *p*-norm space $\quad d(X,\ell_2) \le r^{\frac{2}{p}-\frac{1}{2}}$
- Let $Z = \operatorname{span}\left\{U^{(1)}_{:,i} \circ \cdots \circ U^{(t)}_{:,i}\right\}_{i=1}^{r}, \quad X = \left(\mathbb{R}^{n^t}, \|\cdot\|_{\max}\right) \cap Z, \quad Y = \left(\mathbb{R}^{n^t}, |\cdot|_\infty\right) \cap Z$
- Then $\ d(X,Y) \le d(X,\ell_2)d(Y,\ell_2) \le \sqrt{r^t}$

Ideas for $\|T\|_{\max} \leq \sqrt{r^t} \cdot |T|_\infty$

- Banach-Mazur distance:

$$d(X,Y) = \inf\{\|L\|_{X\to Y}\|L^{-1}\|_{Y\to X} \ : \ L \text{ isomorphism}\}$$

- By Dilworth (1985), *r*-dim *p*-norm space $d(X,\ell_2) \leq r^{\frac{2}{p}-\frac{1}{2}}$
- Let $Z = \mathrm{span}\left\{U^{(1)}_{:,i} \circ \cdots \circ U^{(t)}_{:,i}\right\}_{i=1}^{r}, \quad X = \left(\mathbb{R}^{n^t}, \|\cdot\|_{\max}\right) \cap Z, \quad Y = \left(\mathbb{R}^{n^t}, |\cdot|_\infty\right) \cap Z$
- Then $d(X,Y) \leq d(X,\ell_2)d(Y,\ell_2) \leq \sqrt{r^t}$
- Consider norm inequality $c|T|_\infty \leq \|T\|_{\max} \leq C|T|_\infty$