Spectral gap-based deterministic tensor completion

Kameron Decker Harris
Department of Computer Science
Western Washington University
Bellingham, WA, USA
harri267@wwu.edu

Oscar López
Harbor Branch Oceanographic Institute
Florida Atlantic University
Fort Pierce, FL, USA
lopezo@fau.edu

Angus Read
Department of Computer Science
Western Washington University
Bellingham, WA, USA
reada2@wwu.edu

Yizhe Zhu
Department of Mathematics
University of California, Irvine
Irvine, CA, USA
yizhe.zhu@uci.edu

June 13, 2023

Abstract

Tensor completion is a core machine learning algorithm used in recommender systems and other domains with missing data. While the matrix case is well-understood, theoretical results for tensor problems are limited, particularly when the sampling patterns are deterministic. Here we bound the generalization error of the solutions of two tensor completion methods, Poisson loss and atomic norm minimization, providing tighter bounds in terms of the target tensor rank. If the ground-truth tensor is order t with CP-rank r, the dependence on r is improved from $r^{2(t-1)(t^2-t-1)}$ in [16] to $r^{2(t-1)(3t-5)}$. The error in our bounds is deterministically controlled by the spectral gap of the sampling sparsity pattern. We also prove several new properties for the atomic tensor norm, reducing the rank dependence from r^{3t-3} in [14] to r^{3t-5} under random sampling schemes. A limitation is that atomic norm minimization, while theoretically interesting, leads to inefficient algorithms. However, numerical experiments illustrate the dependence of the reconstruction error on the spectral gap for the practical max-quasinorm, ridge penalty, and Poisson loss minimization algorithms. This view through the spectral gap is a promising window for further study of tensor algorithms.

1 Introduction

In many situations, incomplete data measurements are the rule due to data corruption, impracticality, or impossibility of filling in all information that is desired. Tensor completion is a general technique used to fill in missing multivariate data using assumptions of prior structure. It is the natural generalization of matrix completion, made famous as a winning solution to the Netflix prize [2]: A system to recommend films to users was built by filling in a low-rank matrix of scores for each pair of movies by users.

In tensor completion, we seek to fully or partially recover the entries of an unknown order t tensor (i.e., a multidimensional $n \times \cdots \times n$ array with $t \geq 2$ indices) T given some limited set of observations $\{T_e\}_{e \in E}$. The ground truth is assumed to have CP-rank r [20]. Mathematical analysis of such methods often assumes that the observed entries in the tensor $E \subseteq [n]^t$ are random, and the best known polynomial-time algorithms require $\tilde{O}(n^{t/2})$ sample complexity for an order t tensor [19, 22, 24, 6]. On the other hand, a number of papers have studied deterministic sampling for matrix completion [17, 4, 3, 5, 10, 12]. For tensors, only a few recent results have studied error bounds in the deterministic setting. [9, Theorem 3.1] gave a general bound for non-uniform Frobenius norm error for any deterministic sampling pattern with an NP-hard algorithm.

However, the dependence on rank in the bound is not specified. An HOSVD method was also analyzed in [9] for low Tucker rank tensor completion.

Closest to the work we present here, [16] found bounds for minimizing a quasinorm penalized problem [14] in terms a *spectral gap*. For regular graphs where each vertex has the same degree, *smaller second eigenvalue corresponds to larger spectral gap*. Viewing the sampling pattern as the adjacency tensor of a t-uniform hypergraph H constructed from a regular graph G, they showed that the second eigenvalue of G could be used to control the reconstruction error. Expander graphs, which have the smallest possible second eigenvalue/largest spectral gap, are thus optimal for that theory. In addition, using the definition of the second eigenvalue for the adjacency tensor of a hypergraph [13], they showed that the for any t-uniform hypergraph (not necessarily regular) as a sampling pattern, its second eigenvalue could also be used to control the reconstruction error, but with a larger sample complexity bound.

Our main contribution is a collection of theoretical results suggesting that the spectral gap influences the quality of reconstruction. We first show how the rank dependence of previous generalization bounds [16] can be tightened if the optimization is done with a different atomic norm penalty. Several properties of this atomic norm are proven, improving upon the results in [14] where the atomic norm was introduced. We also extend the analysis of tensor count data with a Poisson likelihood model [21] to deterministic observations. Finally, we present numerical experiments to showcase how a larger spectral gap leads to less error. These results together provide strong evidence that the spectral gap of the observation mask is important for controlling the error of tensor completion. Our notation and some basic definitions are given in App. A.1.

2 Theoretical results

2.1 Tensor atomic norm properties

In order to connect the expansion properties of the sampling pattern hypergraph to the error of the algorithms, we will work with sign tensors. A sign tensor S has all entries equal to $+1$ or -1, i.e. $S \in \bigotimes_{i=1}^{t} \{\pm 1\}^{n_i}$. The **sign rank** of a sign tensor S is defined as

$$\operatorname{rank}_{\pm}(S) = \inf \left\{ r \;\middle|\; S = \sum_{i=1}^{r} s_i^{(1)} \circ \cdots \circ s_i^{(t)} \right\}, \tag{2.1}$$

where $s_i^{(j)} \in \{\pm 1\}^{n_i}, i \in [r], j \in [t]$. We define the **atomic norm** for a tensor T as

$$\|T\|_{\pm} = \inf \left\{ \sum_{i=1}^{r} |\alpha_i| \;\middle|\; T = \sum_{i=1}^{r} \alpha_i S_i \right\}, \tag{2.2}$$

where $\alpha_i \in \mathbb{R}$, $S_i \in \bigotimes_{i=1}^{t} \{\pm 1\}^{n_i}, i \in [r]$. This is called *tensor atomic-M norm* in [14]. In (2.2), the "atoms" are simple sign tensors.

Note that the set of all rank-1 sign tensors forms a basis for $\bigotimes_{i=1}^{t} \mathbb{R}^n$, so this decomposition into rank-1 sign tensors is always possible; furthermore, this is a norm for tensors and matrices [14, 17], and it is commonly used in compressed sensing and matrix completion [8, 11]. The following results give several useful properties of the tensor atomic norm, which will be used in our tensor completion analysis (proofs in Apps. A.2 and A.3, respectively):

Theorem 2.1. *Let $T \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ and $S \in \bigotimes_{i=1}^{t} \mathbb{R}^{m_i}$. The following atomic norm properties hold:*

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

Lemma 2.2. *Let $u_i \in \mathbb{R}^{n_i}$ with $\|u_i\|_\infty \le 1$ for $i \in [t]$ and rank-1 $T = u_1 \circ u_2 \circ \cdots \circ u_t$. Then $\|T\|_\pm \le 1$.*

We also need to compare the atomic norm of a tensor T and its CP-rank to obtain rank-dependent bounds. The following Theorem improves the rank dependence in [14, Theorem 7] by a factor of r (proof in App. A.4):

Theorem 2.3. *Let $T \in \bigotimes_{i=1}^t \mathbb{R}^{n_i}$ and $\operatorname{rank}(T) = r$. Then*

$$|T|_\infty \le \|T\|_\pm \le K_G \sqrt{r^{3t-5}} |T|_\infty, \tag{2.3}$$

where $K_G \le 1.783$ is Grothendieck's constant over $\mathbb{R}$.

In contrast, for the max-quasinorm (see App. A.1), it was shown in [16] that

$$|T|_\infty \le \|T\|_{\max} \le \sqrt{r^{t^2-t-1}}\, |T|_\infty.$$

Since $3t - 5 \le t^2 - t - 1$ for all integers $t \ge 2$, atomic norm-based analysis and optimization will yield a better rank dependence in the generalization error bound.

While the main focus of this paper is on deterministic sampling patterns, we note that Theorem 2.3 can be used to improve upon results in the literature that consider random sampling schemes. In [14], the authors introduce the atomic norm to show that $O\left(nr^{3t-3}\right)$ entries chosen randomly are sufficient to provide an accurate approximation of a rank r tensor in $\bigotimes_{i=1}^t \mathbb{R}^n$. Using our main result, Theorem 8 in [14] can be applied with $R = K_G\sqrt{r^{3t-5}}|T|_\infty$ to reduce the sampling complexity to $O\left(nr^{3t-5}\right)$. To the best of our knowledge, the results here provide the best sampling complexity to date for tensors of general order under random and deterministic sampling patterns.

2.2 Deterministic tensor completion

Equipped with the properties of the atomic norm proved in Theorem 2.1, we give an improved generalization error bound for deterministic tensor completion (proof in App. A.5):

Theorem 2.4. *Given a hypercubic tensor T of order t, reveal its entries according to a t-uniform t-partite hypergraph $H = (V, E)$ with $V = V_1 \cup \cdots \cup V_t$, $|V_1| = \cdots = |V_t| = n$, and second eigenvalue $\lambda_2(H)$. Let $\hat{T}$ satisfy*

$$\hat{T} = \arg\min_{T'} \|T'\|_\pm \quad \text{such that} \quad T'_e = T_e \quad \text{for all} \quad e \in E.$$

Then the following bound holds:

$$\frac{1}{n^t}\|\hat{T} - T\|_F^2 \le \frac{2^{t+2} n^{t/2} \lambda_2(H)}{|E|} \|T\|_\pm^2, \tag{2.4}$$

where $\lambda_2(H) = \left\| T_H - \frac{|E|}{n^t} J \right\|$, J is the all-ones tensor, and T_H is the adjacency tensor of H such that $T_{i_1,\ldots,i_t} = \mathbf{1}\{(i_1, \ldots, i_t) \in E\}$.

The above result is for any t-uniform, t-partite hypergraphs, and the error bound depends on the second eigenvalue of their adjacency tensors. Concentration for the second eigenvalue of adjacency tensors for random hypergraphs was considered in [13, 25].

Computing $\lambda_2(H)$ is costly since the tensor spectral norm is NP-hard [18]. However, we can "lift" a graph into a hypergraph in a way that gives a bound in terms of graph eigenvalues. Let $G = (V(G), E(G))$ be a connected d-regular graph on n vertices with the second largest eigenvalue (in absolute value) $\lambda \in (0, d)$. The following construction of a t-partite, t-uniform, d^{t-1}-regular hypergraph $H = (V, E)$ from G was given in [16]. See Fig 1 for an illustration of the "edge lifting" operation from a regular graph to a regular hypergraph graph when $t = 3$ and $d = 4$.

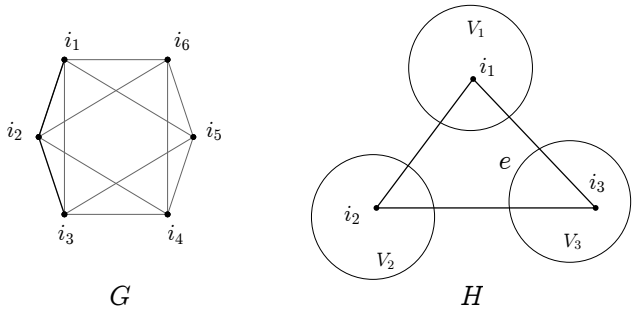

Figure 1: Lifting a graph G into a hypergraph H when $t = 3$: We depict a 4-connected ring base graph G on the left and a single edge in the hypergraph H on the right. (i_1, i_2, i_3) forms an hyperedge e in H if and only if (i_1, i_2, i_3) is a walk in G.

Definition 2.5 (Edge lifting for regular hypergraphs)**.** Let $V = V_1 \cup V_2 \cup \cdots \cup V_t$ be the disjoint union of t vertex sets such that $|V_1| = \cdots = |V_t| = n$. The hyperedges of H correspond to all walks of length $t-1$ in G: $(v_1, \ldots, v_t)$ is a hyperedge in H if and only if $(i_1, \ldots, i_t)$ is a walk of length $t-1$ in G.

Theorem 2.6. *Given a hypercubic tensor T of order t, reveal its entries according to a t-partite, t-uniform, d^{t-1}-regular hypergraph $H = (V, E)$ lifted from a d-regular graph G of size n with second eigenvalue (in absolute value) $\lambda \in (0, d)$. Then solving*

$$\begin{aligned} \hat{T} &= \arg\min_{T'} \|T'\|_{\pm} \\ &\quad \textit{such that} \quad T'_e = T_e \quad \textit{for all} \quad e \in E \end{aligned} \tag{2.5}$$

will result in the following mean squared error bound:

$$\begin{aligned} \frac{1}{n^t}\|\hat{T} - T\|_F^2 &\le 2^t(2t-3)\frac{\lambda}{d}\|T\|_{\pm}^2 \\ &\le 2^t(2t-3)\frac{\lambda}{d}K_G^2 r^{3t-5}|T|_\infty^2. \end{aligned} \tag{2.6}$$

Theorem 2.6 is proven in App. A.6. Suppose we lift an expander graph G with $\lambda = O(\sqrt{d})$ and $t, |T|_\infty = O(1)$. In order to have the right hand side of (2.6) bounded by ε, we need to take

$$|E| = O\left(\frac{nr^{2(t-1)(3t-5)}}{\varepsilon^{2(t-1)}}\right) \tag{2.7}$$

many samples. The sample complexity in [16] is $O\left(\frac{nr^{2(t-1)(t^2-t-1)}}{\varepsilon^{2(t-1)}}\right)$. Optimizing the atomic norm instead of the max-quasinorm yields a better bound, although it requires costly integer programming.

3 Poisson Tensor Regression

We may also use the atomic norm properties to obtain error bounds in the case of noisy tensor completion. In this scenario, we seek to estimate a parametric tensor T from noisy, incomplete observations so that T is never observed directly. We specifically consider the Poisson tensor completion problem [21], where we observe count data $X \in \bigotimes_{i=1}^t \mathbb{N}_+^{n_i}$ obeying

$$X_e \sim \text{Poisson}(T_e) \quad \text{for all} \quad e \in E, \tag{3.1}$$

where $T \in \bigotimes_{i=1}^{t}[\beta,\alpha]^{n_i}$ specifies the range of possible Poisson parameters (with $0 < \beta \leq \alpha$). Given X_e for $e \in E$, we approximate T via a Poisson maximum likelihood estimator

$$\hat{T} = \underset{Z \in S_r(\beta,\alpha)}{\arg\max} \sum_{e \in E} X_e \log(Z_e) - Z_e \tag{3.2}$$

where our parametric tensor search space is

$$S_r(\beta,\alpha) = \left\{ Z \in \bigotimes_{i=1}^{t}[\beta,\alpha]^{n_i} \mid \operatorname{rank}(Z) \leq r \right\}.$$

This leads to our main result for the Poisson regression (proof in App. A.7):

Theorem 3.1. *Let the hypercubical parameter tensor T and observations X be generated as above, with entries revealed according to a t-uniform, t-partite hypergraph $H = (V,E)$ with $V = V_1 \cup \cdots \cup V_t$, $|V_1| = \cdots = |V_t| = n$, and second eigenvalue $\lambda_2(H)$. Then with probability exceeding $1 - \frac{2}{|E|}$, there exists an absolute constant $C > 0$ such that $\hat{T}$ satisfies:*

$$\begin{aligned} \frac{1}{n^t}\|\hat{T} - T\|_F^2 \leq & \frac{C\alpha^3 t^{3/2}\sqrt{nr^{3t-5}}\log_2(n)}{\beta\sqrt{|E|}} \\ & + \frac{2^{t+2}n^{t/2}\lambda_2(H)K_G^2\alpha^2 r^{3t-5}}{|E|}. \end{aligned} \tag{3.3}$$

Furthermore, if the entries are revealed according to a t-partite, t-uniform, d^{t-1}-regular hypergraph $H = (V,E)$ constructed from a d-regular graph G of size n with second eigenvalue (in absolute value) $\lambda \in (0,d)$, then with probability exceeding $1 - \frac{2}{nd^{t-1}}$,

$$\begin{aligned} \frac{1}{n^t}\|\hat{T} - T\|_F^2 \leq & \frac{C\alpha^3 t^{3/2}\sqrt{r^{3t-5}}\log_2(n)}{\beta\sqrt{d^{t-1}}} \\ & + \alpha^2 2^t (2t-3)\frac{\lambda}{d}K_G^2 r^{3t-5}. \end{aligned} \tag{3.4}$$

If all r, t, α, β are independent of n, and we take an expander graph with $\lambda = O(\sqrt{d})$, (3.4) says if nd^{t-1} is $\omega(n\log^2 n)$, then the generalization error goes to 0 as $n \to \infty$. Fixing t, α, β, the sample size is $O\left(\frac{nr^{2(t-1)(3t-5)}}{\varepsilon^{2(t-1)}} + \frac{nr^{3t-5}\log^2(n)}{\varepsilon}\right)$ for an ε-approximation. The result is no longer deterministic due to the random nature of the observed counts X, but not the observation mask. We roughly obtain the same sampling complexity as in Theorem 2.6 with an extra $O\left(\frac{nr^{3t-5}\log^2(n)}{\varepsilon}\right)$ term.

We end this section by noting that the derived noisy tensor completion error bounds (3.3) and (3.4) are dependent on the Poisson parameter bounds β, α. Such dependence is typical in the literature under random sampling schemes [7, 21]. Theoretical results therein exhibit analogous dependence on the distributional parameter space, with numerical experiments that further validate this dependence as β and α are varied.

4 Numerical Experiments

We now present some numerical experiments that explore our derived error bounds in practice. The goal of this section is to showcase how the accuracy of estimators $\hat{T}$ depends on the spectral gap of the associated hypergraph or "lifted" regular graph that specifies the sampled tensor entries. To do so, we will generate subsets of revealed tensor entries E in different manners that vary the second eigenvalue of the associated adjacency tensor or lifted graph. The following experiments keep the cardinality $|E|$ fixed so that only the distribution of the sampled entries contributes to the behavior of the reconstruction errors.

We tested the max-quasinorm minimization algorithm of [16] with graph lifting for a graph with $d = 15$ and random target tensor with $n = 100$, $t = 3$, and $r = 3$. To vary the eigenvalue $\lambda_2(G)$, we start by creating a d-connected ring graph (see Figure 1). We then perform a number of random edge swaps where we randomly select two edges and switch their endpoints, which preserves the regularity of the graph. This is the classical switch Markov chain for generating random regular graphs [15]. As the number of swaps increases, λ_2 decreases until it stabilizes at approximately $2\sqrt{d-1}$. We vary the number of swaps between 0 and 600 to give as wide of a λ_2 range as possible. This graph is then lifted into a hypergraph/tensor mask of observed entries.

Figure 2 shows the results from these experiments. We observe a significant positive correlation between generalization error and λ_2, increasing nonlinearly for the larger values of the eigenvalue. To check if this pattern persisted when limiting to lower λ_2 values, we removed the points with large λ_2 values that were dominating the plot. We can see there still is a positive correlation when the points are removed, providing further evidence that λ_2 is a determinant of generalization error. Further details and similar plots for two additional square loss algorithms, as well as regression reports, are included in App. A.8 and the supplemental files.

We also experimented with the Poisson regression algorithm [21]. In this case, we did not use graph lifting but came up with a procedure to generate mask tensors with varying $\lambda_2(H)$ by starting from a grid-like mask and randomly shuffling entries. The tensor eigenvalue $\lambda_2(H)$ was estimated by a rank-1 fit to $(T_H - \frac{|E|}{n^t} J)$. We again saw a strong correlation between λ_2 and error. Further details and plots of results are shown in App. A.9.

5 Conclusion

We provided an improved analysis of deterministic tensor completion based on the spectral gap of expander graphs in [16] and applied the results for Poisson tensor regression. Our new numerical experiments support the dependence of generalization error on the spectral gap. Our main contribution also improves upon previous results that consider random sampling schemes, providing the best sampling complexity to date for general order tensor completion problems in terms of the CP-rank.

It would be interesting to see if our analysis can be extended for deterministic non-uniform low CP-rank tensor completion following the line of work [12, 9]. However, more properties of the tensor atomic or max-quasinorm are needed. In particular, in the matrix case, we have the following relation between the max-norm and operator norm: $\|A * B\| \leq \|A\|_{\max} \|B\|$ for any two $n \times n$ matrices A, B, which is crucial in the proof of [12, Theorem 15] and [9, Theorem B.1]. Generalizing this inequality for these tensor factorization norms is an interesting question for future work.

Our work also contains some limitations. First, we did not study the computational complexity of our optimization algorithms, although we expect them to be NP-hard [1]. Directly minimizing the atomic norm as in Theorem 2.4 requires integer programming and is not efficient in practice. Finally, while many numerical results show a good correlation with λ_2, there is significant unexplained variance at a given gap and across algorithms. Like many results, these theoretical bounds are not tight enough to quantitatively predict performance, and they are far from the parameter counting lower-bound of $O(nrt)$ for the CP decomposition.

Acknowledgments

The authors are listed in alphabetical order. We would like to thank Daniel Dunlavy for his guidance in setting up the Poisson tensor regression numerical experiments.

O.L. is supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

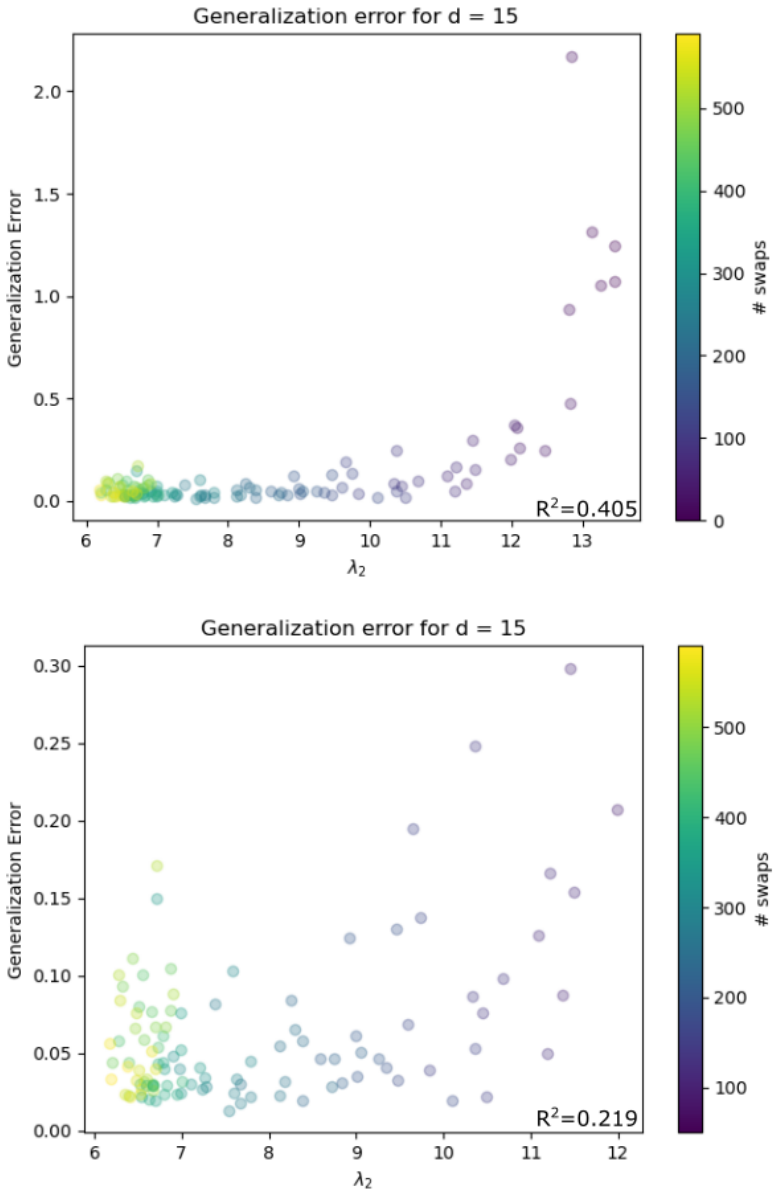

Figure 2: Numerical experiments show that reconstruction error correlates with $\lambda_2(G)$ using the max-quasinorm minimization method and graph lifting. This graph has $d = 15$ and $n = 100$, and the tensor order $t = 3$, for 2.25% of the entries sampled. Above: The full range of graph eigenvalues by varying edge swaps from 0 to 600. Below: Linear fits show significant positive correlation even when points with $\lambda_2 > 12$ are removed. Coefficients of determination R^2 for linear fits of error versus λ_2 are included.

Y.Z. is partially supported by NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning.

References

[1] Boaz Barak and Ankur Moitra. Noisy tensor completion via the sum-of-squares hierarchy. In *Conference on Learning Theory*, pages 417–445, 2016.

[2] J. Bennett and S. Lanning. The netflix prize. In *Proceedings of the KDD Cup Workshop 2007*, pages 3–6. ACM, August 2007.

[3] Srinadh Bhojanapalli and Prateek Jain. Universal matrix completion. In *International Conference on Machine Learning*, pages 1881–1889, 2014.

[4] Gerandy Brito, Ioana Dumitriu, and Kameron Decker Harris. Spectral gap in random bipartite biregular graphs and applications. *Combinatorics, Probability and Computing*, 31(2):229–267, 2022.

[5] Shantanu Prasad Burnwal and Mathukumalli Vidyasagar. Deterministic completion of rectangular matrices using asymmetric Ramanujan graphs: Exact and stable recovery. *IEEE Transactions on Signal Processing*, 68:3834–3848, 2020.

[6] Changxiao Cai, Gen Li, H Vincent Poor, and Yuxin Chen. Nonconvex low-rank symmetric tensor completion from noisy data. *Advances in neural information processing systems*, 2019.

[7] Yang Cao and Yao Xie. Poisson matrix recovery and completion. *IEEE Transactions on Signal Processing*, 64(6):1609–1620, 2016.

[8] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.

[9] Zehan Chao, Longxiu Huang, and Deanna Needell. HOSVD-based algorithm for weighted tensor completion. *Journal of Imaging*, 7(7):110, 2021.

[10] Sourav Chatterjee. A deterministic theory of low rank matrix completion. *IEEE Transactions on Information Theory*, 66(12):8046–8055, 2020.

[11] Yuejie Chi and Maxime Ferreira Da Costa. Harnessing sparsity over the continuum: Atomic norm minimization for superresolution. *IEEE Signal Processing Magazine*, 37(2):39–57, 2020.

[12] Simon Foucart, Deanna Needell, Reese Pathak, Yaniv Plan, and Mary Wootters. Weighted matrix completion from non-random, non-uniform sampling patterns. *IEEE Transactions on Information Theory*, 67(2):1264–1290, 2020.

[13] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995.

[14] Navid Ghadermarzy, Yaniv Plan, and Özgür Yilmaz. Near-optimal sample complexity for convex tensor completion. *Information and Inference: A Journal of the IMA*, 8(3):577–619, 2019.

[15] Catherine Greenhill. Generating graphs randomly. *arXiv preprint arXiv:2201.04888*, 2022.

[16] Kameron Decker Harris and Yizhe Zhu. Deterministic tensor completion with hypergraph expanders. *SIAM Journal on Mathematics of Data Science*, 3(4):1117–1140, 2021.

[17] Eyal Heiman, Gideon Schechtman, and Adi Shraibman. Deterministic algorithms for matrix completion. *Random Structures & Algorithms*, 45(2):306–317, 2014.

[18] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

[19] Prateek Jain and Sewoong Oh. Provable tensor factorization with missing data. In *Advances in Neural Information Processing Systems*, pages 1431–1439, 2014.

[20] Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

[21] Oscar López, Daniel M Dunlavy, and Richard B Lehoucq. Zero-truncated Poisson regression for sparse multiway count data corrupted by false zeros. *arXiv preprint arXiv:2201.10014*, 2022.

[22] Andrea Montanari and Nike Sun. Spectral algorithms for tensor completion. *Communications on Pure and Applied Mathematics*, 71(11):2381–2425, 2018.

[23] Cyrus Rashtchian. Bounded matrix rigidity and John's theorem. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, page 93, 2016.

[24] Dong Xia and Ming Yuan. On polynomial time methods for exact low-rank tensor completion. *Foundations of Computational Mathematics*, 19(6):1265–1313, 2019.

[25] Zhixin Zhou and Yizhe Zhu. Sparse random tensors: Concentration, regularization and applications. *Electronic Journal of Statistics*, 15(1):2483–2516, 2021.

A Appendix

A.1 Notation and definitions

We use lowercase symbols u for vectors, uppercase U for matrices and tensors. The symbol "$\circ$" denotes the outer product of vectors, i.e. $T = u \circ v \circ w$ denotes the order 3, rank-1 tensor with entry $T_{i,j,k} = u_i v_j w_k$. We also use this symbol for the outer product of matrices as appears in the rank-r decomposition of a tensor $T = U^{(1)} \circ U^{(2)} \circ U^{(3)}$, where each matrix $U^{(i)}$ has r columns, so that $T_{i,j,k} = \sum_{l=1}^{r} U^{(1)}_{i,l} U^{(2)}_{j,l} U^{(3)}_{k,l}$, and $T = \bigcirc_{i=1}^{t} U^{(i)}$ is shorthand for the same order t, rank-r tensor. The symbols $\otimes$ and $*$ denote Kronecker and Hadamard products, respectively. We use $\bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ for the space of all order t tensors with n_i entries in the i-th dimension. We use $1_A \in \mathbb{R}^n$ as the indicator vector of a set $A \subseteq [n]$, i.e. $(1_A)_i = 1$ if $i \in A$ and 0 otherwise. For any order t tensor $T \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ and subsets $I_i \subseteq [n_i]$, denote $T_{I_1,\dots,I_t}$ to be the subtensor restricted on the index set $I_1 \times \cdots \times I_t$. Norms $\|\cdot\|$ are by default the ℓ_2 norm for vectors and operator norm for matrices and tensors. We use the notation $|\cdot|_p$ for entry-wise ℓ_p norms of matrices and tensors.

Let $T \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ and $S \in \bigotimes_{i=1}^{t} \mathbb{R}^{m_i}$. We define the **Kronecker product** of two tensors $(T \otimes S) \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i m_i}$ as the tensor with entries

$$(T \otimes S)_{k_1,\dots,k_t} = T_{i_1,\dots,i_t} S_{j_1,\dots,j_t}$$

for $k_1 = j_1 + m_1(i_1 - 1), \dots, k_t = j_t + m_t(i_t - 1)$.

Let $T, S \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$. We define the **Hadamard product** of two tensors $(T * S) \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ as the tensor with indices $(T * S)_{i_1,\dots,i_t} = T_{i_1,\dots,i_t} S_{i_1,\dots,i_t}$.

For matrices, the most common measure of complexity is the rank. In the tensor setting, there are various definitions of rank [20]. However, in this paper, we will work with the **CP-rank** defined as

$$\operatorname{rank}(T) = \min \left\{ r \;\middle|\; T = \sum_{i=1}^{r} u_i^{(1)} \circ \cdots \circ u_i^{(t)} \right\}, \tag{A.1}$$

where each vectors $u_i^{(j)} \in \mathbb{R}^n$.

In [14], Ghadermarzy, Plan, and Yilmaz studied tensor completion without reducing it to a matrix case by minimizing a **max-quasinorm** as a proxy for rank. This is defined as

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

where the factorization is a CP decomposition of T.

Define the **spectral norm** of a tensor T as

$$\|T\| = \sup_{v_1,\dots,v_t \in S^{n-1}} \left| \sum_{i_1,\dots,i_t=1}^{n} T_{i_1,\dots,i_t} v_1(i_1) \cdots v_t(i_t) \right|, \tag{A.2}$$

where S^{n-1} is the unit sphere in $\mathbb{R}^n$.

A **d-regular graph** on n vertices is a graph where each vertex has the same degree d. The **adjacency matrix** of a graph $G = (V, E)$ is a $|V| \times |V|$ symmetric matrix such that $A_{ij} = \mathbf{1}\{\{i, j\} \in E\}$ for all $i, j \in V$. The **second eigenvalue** (in absolute value) of G, denoted by $\lambda(G)$, is defined as $\lambda = \max\{|\lambda_2(A)|, |\lambda_n(A)|\}$.

A **hypergraph** $H = (V, E)$ consists of a set V of vertices and a set E of hyperedges such that each hyperedge is a nonempty set of V. A hypergraph H is **k-uniform** for an integer $k \geq 2$ if every hyperedge $e \in E$ contains exactly k vertices. The *degree* of i, is the number of all hyperedges incident to i. A hypergraph is **d-regular** if all of its vertices have degree d. A k-uniform hypergraph is **k-partite** if we can decompose the vertex set as a disjoint union $V = V_1 \cup \cdots \cup V_k$ such that each hyperedge in E contains exactly one vertex in $V_i, 1 \leq i \leq k$. The **adjacency tensor** T of a k-uniform hypergraph $H = (V, E)$ on n vertices is a k-th order symmetric tensor of size n such that $T_{i_1,\dots,i_k} = \mathbf{1}\{\{i_1, \dots, i_k\} \in E\}$. The second eigenvalue of H, denoted by $\lambda_2(H)$, is defined as $\lambda_2(H) = \left\| T - \frac{|E|}{n^t} J \right\|$, where J is the all-ones tensor.

A.2 Proof of Theorem 2.1

We prove Claims (1–4) in order. For Claim (1), write T using the decomposition which attains the atomic norm, $T = \bigcirc_{i=1}^{t} U^{(i)}$ for some $U^{(i)} \in \mathbb{R}^{n_i \times r}, 1 \le i \le t$. Then a single entry of T can be written as

$$T_{i_1,\dots,i_t} = \sum_{i=1}^{r} \alpha_i u_{i_1,i}^{(1)} \cdots u_{i_t,i}^{(t)},$$

where $u_{i_k,i}^{k} \in \{-1,+1\}$ for $1 \le i_k \le n_k, 1 \le i \le r, 1 \le k \le t$,

$$U^{(1)} = [\alpha_1 u_1^{(1)}, \cdots, \alpha_r u_r^{(1)}],$$

and $U^{(k)}, 2 \le k \le t$ are matrices with column vectors given by $u_1^{(k)}, \dots, u_r^{(k)} \in \mathbb{R}^{n_1}$. By the definition of atomic norm, we have

$$\|T\|_{\pm} = \sum_{i=1}^{r} |\alpha_i|.$$

We know that $T_{I_1,\dots,I_t} = \bigcirc_{i=1}^{t} U_{I_i,:}^{(i)}$, where $U_{I_i,:}^{(i)}$ denotes the submatrix of $U^{(i)}$ with the column restricted on I_i. Therefore $T_{I_1,\dots,I_t}$ can be written as a linear combination of rank-1 sign tensors with the sum of absolute value of weights given by $\sum_{i=1}^{r} |\alpha_i|$. By the definition of the atomic norm, this is an upper bound, so $\|T_{I_1,\dots,I_t}\|_{\pm} \le \|T\|_{\pm}$. This proves Claim (1). For Claim (2), let

$$T = \bigcirc_{i=1}^{t} T^{(i)} \quad \text{and} \quad S = \bigcirc_{i=1}^{t} S^{(i)}$$

be the rank r_1 and r_2 decompositions of T and S that attain their atomic norms. We can write

$$T_{i_1,\dots,i_t} = \sum_{i=1}^{r_1} \alpha_i u_{i_1,i}^{(1)} \cdots u_{i_t,i}^{(t)}, \quad S_{j_1,\dots,j_t} = \sum_{j=1}^{r_2} \beta_j v_{j_1,j}^{(1)} \cdots v_{j_t,j}^{(t)}, \tag{A.3}$$

where $v_{j_k,j}^{k} \in \{-1,+1\}$ for $1 \le j_k \le n_k, 1 \le j \le r, 1 \le k \le t$,

$$V^{(1)} = [\beta_1 v_1^{(1)}, \cdots, \beta_{r_2} v_{r_2}^{(1)}],$$

and $V^{(k)}, 2 \le k \le t$ are matrices with column vectors given by $v_1^{(k)}, \dots, v_r^{(k)}$. And by definition, $\|S\|_{\pm} = \sum_{j=1}^{r_2} |\beta_j|$. Then since

$$\begin{aligned}
(T \otimes S)_{k_1,\dots,k_t} = T_{i_1,\dots,i_t} S_{j_1,\dots,j_t} &= \left(\sum_{l=1}^{r_1} U_{i_1,l}^{(1)} \cdots U_{i_t,l}^{(t)} \right) \left(\sum_{l'=1}^{r_2} V_{j_1,l'}^{(1)} \cdots V_{j_t,l'}^{(t)} \right) \\
&= \sum_{l=1}^{r_1} \sum_{l'=1}^{r_2} \left(U_{i_1,l}^{(1)} V_{j_1,l'}^{(1)} \right) \cdots \left(U_{i_t,l}^{(t)} V_{j_t,l'}^{(t)} \right) \\
&= \sum_{p=1}^{r_1 r_2} \left(U^{(1)} \otimes V^{(1)} \right)_{k_1,p} \cdots \left(T^{(t)} \otimes V^{(t)} \right)_{k_t,p}
\end{aligned}$$

for $k_s = j_s + m_s(i_s - 1)$ for all $s = 1, \dots, t$ and $p = l' + r_2(l-1)$, we have that $T \otimes S = \bigcirc_{i=1}^{t} (U^{(i)} \otimes V^{(i)})$. This gives a way to write $T \otimes S$ as a weighted sum of rank-1 sign tensors with the sum of absolute value of weights given by

$$\sum_{i=1}^{r_1} \sum_{j=1}^{r_2} |\alpha_i \beta_j| = \|T\|_{\pm} \|S\|_{\pm}.$$

Therefore $\|T \otimes S\|_{\pm} \le \|T\|_{\pm} \|S\|_{\pm}$. This completes the proof of Claim (2). For Claim (3), note that every entry in $T * S$ appears in $T \otimes S$, since

$$(T * S)_{i_1,\dots,i_t} = (T \otimes S)_{i_1 + n_1(i_1-1),\dots,i_t + n_t(i_t-1)}.$$

So we have that $T * S = (T \otimes S)_{I_1,\dots,I_t}$ for some subsets of indices $I_1, \dots, I_t$, and by Claim (1), the result follows. Finally, from Claims (2) and (3), $\|T * T\|_{\pm} \le \|T \otimes T\|_{\pm} \le \|T\|_{\pm}^2$, and Claim (4) follows.

A.3 Proof of Lemma 2.2

Since $\|u_i\|_\infty \le 1$, $u_i \in [-1,1]^{n_i}$, where $[-1,1]^{n_i}$ is a convex hull of the set $\{-1,1\}^{n_i}$, u_i can be written as a convex combination of $\{-1,1\}^{n_i}$ such that

$$u_i = \sum_{k\in[2^{n_i}]} \lambda_k^{(i)} v_k^i,$$

where $v_k^i, k \in [2^{n_i}]$ are all possible vectors in $\{-1,1\}^{n_i}$ and $\sum_k |\lambda_k^{(i)}| = 1$. Therefore we have

$$T = u_1 \circ u_2 \cdots \circ u_t = \sum_{k_1,\ldots,k_t} \lambda_{k_1}^{(1)} \cdots \lambda_{k_t}^{(t)} v_{k_1}^{(1)} \circ \cdots \circ v_{k_t}^{(t)},$$

which is a decomposition of T as a linear combination of sign rank-1 tensors. So

$$\|T\|_\pm \le \sum_{k_1,\ldots,k_t} |\lambda_{k_1}^{(1)} \cdots \lambda_{k_t}^{(t)}| \le 1.$$

A.4 Proof of Theorem 2.3

The lower bound was shown in [14, Theorem 7]. We now focus on the upper bound. When $t = 2$, using Grothendieck's inequality, it was shown in [17, Theorem 7] that $\|T\|_\pm \le K_G \|T\|_{\max}$. From John's Theorem (see, for example [23, Corollary 2.2]), $\|T\|_{\max} \le \sqrt{r}|T|_\infty$. This proves (2.3) when $t = 2$.

For $t \ge 3$, we will use induction. Let T be an order t tensor with rank r and $|T|_\infty \le 1$. Then T has the rank-r decomposition as

$$T = \sum_{j=1}^r v_j^1 \circ v_j^2 \circ \cdots \circ v_j^t.$$

Matricizing along mode-1 we obtain $T_{[1]} \in \mathbb{R}^{n_1 \times (n_2 \cdots n_t)}$ such that

$$T_{[1]} = \sum_{i=1}^r v_i^1 \circ (v_i^2 \otimes \cdots \otimes v_i^t).$$

Let W be a $\mathbb{R}^{(n_2\cdots n_t)\times r}$ matrix such that $W(:,i) = v_i^2 \otimes \cdots \otimes v_i^t$. By John's Theorem ([23, Corollary 2.2]), there exists an $S \in \mathbb{R}^{r\times r}$ such that

$$T_{[1]} = X \circ Y, \tag{A.4}$$

where $X = V^{(1)}S \in \mathbb{R}^{n_1\times r}, Y = WS^{-1} \in \mathbb{R}^{n_2\cdots n_t \times r}$, and $\|X\|_{2,\infty} \le r^{1/2}, \|Y\|_{2,\infty} \le 1$. This also implies $\|Y\|_\infty \le \|Y\|_{2,\infty} \le 1$. Since each column of Y is a linear combination of columns of W, for some constants $\{\gamma_{ij}\}$,

$$Y(:,i) = \sum_{j=1}^r \gamma_{ij}(v_j^2 \otimes \cdots \otimes v_j^t).$$

Then

$$E_i := \sum_{j=1}^r \gamma_{ij}(v_j^2 \circ \cdots \circ v_j^t)$$

is a order $(t-1)$ tensor of rank at most r with $\|E_i\|_\infty \le 1$. By induction, we have $\|E_i\|_\pm \le K_G\sqrt{r^{3t-7}}$. Then by the definition of the atomic norm in (2.2), there exists a decomposition of E_i such that for some integer r_i,

$$E_i = \sum_{j=1}^{r_i} \lambda_j^{(i)} u_{i,j}^2 \circ \cdots \circ u_{i,j}^t,$$

where

$$\sum_{j=1}^{r_i} |\lambda_j^{(i)}| \le K_G \sqrt{r^{3t-8}},$$

and $\|u_{i,j}^d\|_\infty \le 1$ for $2 \le d \le t$. This gives a decomposition of T such that

$$T = \sum_{i=1}^{r} x_i \circ \left(\sum_{j=1}^{r_i} \lambda_j^{(i)} u_{i,j}^2 \circ \cdots \circ u_{i,j}^t \right) = \sum_{i=1}^{r} \sum_{j=1}^{r_i} \lambda_j^{(i)} \|x_i\|_\infty \left(\frac{x_i}{\|x_i\|_\infty} \right) \circ u_{i,j}^2 \circ \cdots \circ u_{i,j}^t, \tag{A.5}$$

where x_i is the i-th column vector of X and

$$\|x_i\|_\infty \le \|x_i\|_2 \le \|X\|_{2,\infty} \le r^{1/2}.$$

Here we assume all $x_i \neq 0$ for $i \in [r]$. Otherwise, we can ignore the terms involving a zero vector. (A.5) gives a decomposition of T into a linear combination of $\sum_{i=1}^{r} r_i$ many rank-1 tensors where each component vector has ℓ_∞-norm at most 1. Therefore by Lemma 2.2 and the triangle inequality,

$$\|T\|_\pm \le \sum_{i=1}^{r} \sum_{j=1}^{r_i} |\lambda_j^{(i)}| \|x_i\|_\infty \le r^{3/2} K_G \sqrt{r^{3t-8}} = K_G \sqrt{r^{3t-5}}.$$

A.5 Proof of Theorem 2.4

Following every step in the proof of [16, Theorem 1.2], we have

$$\left| \frac{1}{n^t} \sum_{e \in [n]^t} T_e - \frac{1}{|E|} \sum_{e \in E} T_e \right| \le \frac{2^t n^{t/2} \lambda_2(H)}{|E|} \|T\|_\pm$$

holds for any tensor T. Now we apply this inequality to the tensor of squared residuals $R := (\hat{T} - T) * (\hat{T} - T)$. Since we solve for $\hat{T}$ with equality constraints, we have that $R_e = 0$ for all $e \in E$. Thus, using Claim (4) in Theorem 2.1,

$$\begin{aligned} \frac{1}{n^t} \|\hat{T} - T\|_F^2 &= \left| \frac{1}{n^t} \sum_{e \in [n]^t} R_e \right| \\ &\le \frac{2^t n^{t/2} \lambda_2(H)}{|E|} \|R\|_\pm \le \frac{2^t n^{t/2} \lambda_2(H)}{|E|} \|\hat{T} - T\|_\pm^2 \le \frac{2^t n^{t/2} \lambda_2(H)}{|E|} \left(\|\hat{T}\|_\pm + \|T\|_\pm \right)^2. \end{aligned} \tag{A.6}$$

Since $\hat{T}$ is the output of our optimization routine and T is feasible, $\|\hat{T}\|_\pm \le \|T\|_\pm$. This leads to the final result.

A.6 Proof of Theorem 2.6

Following the same steps in the proof of [16, Theorem 1.3],

$$\left| \frac{1}{n^t} \sum_{e \in [n]^t} T_e - \frac{1}{n d^{t-1}} \sum_{e \in E} T_e \right| \le 2 \sum_{j=1}^{2^{t-1}} \frac{(2t-3)\lambda}{4d} = 2^{t-2}(2t-3)\frac{\lambda}{d}. \tag{A.7}$$

Define $R = (\hat{T} - T) * (\hat{T} - T)$. Then we have

$$\frac{1}{n^t} \|\hat{T} - T\|_F^2 = \left| \frac{1}{n^t} \sum_{e \in [n]^t} R_e \right| \le 2^{t-2}(2t-3)\frac{\lambda}{d} \left(\|\hat{T}\|_\pm + \|T\|_\pm \right)^2 \le 2^t (2t-3)\frac{\lambda}{d} \|T\|_\pm^2. \tag{A.8}$$

A.7 Proof of Theorem 3.1

We closely follow the proof of Theorem 3 in [21]. The proof here is modified by using Lemma 2.3 to bound the atomic norm (which improves the rank dependence by a factor of r) and applying bounds (A.6) and (A.8) to replace the uniform random sampling assumption on E.

Let $r_\pm = \sup_{Z \in S_r(\alpha,\beta)} \|Z\|_\pm$. As in the proof of Theorem 3 in [21], up to equation 3.14, for any E we obtain that with probability greater than $1 - 2|E|^{-1}$

$$\frac{1}{|E|}\sum_{e\in E} R_e \le \frac{128\alpha\sqrt{tn}(2r_\pm + 2)}{\beta\sqrt{|E|}}\left(\alpha(e^2-2) + 3\log_2(|E|)\right) \le \frac{C\alpha^2 t^{3/2}\sqrt{n} r_\pm \log_2(n)}{\beta\sqrt{|E|}},$$

where $R = (\hat{T} - T) * (\hat{T} - T)$ and $C > 0$ is an absolute constant. Adding and subtracting $\frac{1}{n^t}\sum_{e\in[n]^t} R_e$ to the left-hand side and rearranging, we have shown

$$\left|\frac{1}{n^t}\sum_{e\in[n]^t} R_e\right| \le \frac{C\alpha^2 t^{3/2}\sqrt{n} r_\pm \log_2(n)}{\beta\sqrt{|E|}} + \left|\frac{1}{n^t}\sum_{e\in[n]^t} R_e - \frac{1}{|E|}\sum_{e\in E} R_e\right|. \tag{A.9}$$

Bounding the last term as in (A.6) gives

$$\left|\frac{1}{n^t}\sum_{e\in[n]^t} R_e\right| \le \frac{C\alpha^2 t^{3/2}\sqrt{n} r_\pm \log_2(n)}{\beta\sqrt{|E|}} + \frac{2^{t+2} n^{t/2}\lambda_2(H) r_\pm^2}{|E|},$$

which will establish (3.3). To prove (3.4), the additional assumptions imposed on H can be used to instead bound the last term of (A.9) as in (A.8) and obtain

$$\left|\frac{1}{n^t}\sum_{e\in[n]^t} R_e\right| \le \frac{C\alpha^2 t^{3/2}\sqrt{n} r_\pm \log_2(n)}{\beta\sqrt{|E|}} + 2^t(2t-3)\frac{\lambda}{d} r_\pm^2.$$

The proof ends by using $|E| = nd^{t-1}$ and applying Lemma 2.3 to see that $r_\pm \le K_G \alpha\sqrt{r^{3t-5}}$.

A.8 Square loss experiments

Theorems 2.4 and 2.6 both concern constrained regression where the observations are fit exactly. However, for noisy data, one can use a square loss and get similar bounds as in [16]. We tested the performance of three least-squares tensor completion algorithms for varying spectral gaps. These differed in their regularization and optimization routines, and we refer to the algorithms as "ridge," "ridge projected," and "max-quasinorm." All of these algorithms are implemented in the code provided at https://github.com/kamdh/max-qnorm-tensor-completion

The standard ridge and projected version of it consider a sum of squares penalty on the factor matrices $\sum_{i=1}^t \|U^{(i)}\|_F^2$ as in ridge regression. The ridge algorithm attempts to solve for

$$\text{Ridge:} \quad \min_{T' = U^{(1)}\circ\cdots\circ U^{(t)}} \|P(T'-T)\|_F^2 + \varepsilon\sum_{i=1}^t \|U^{(i)}\|_F^2, \tag{A.10}$$

where $T' = U^{(1)} \circ \cdots \circ U^{(t)}$ is the CP decomposition into factor matrices, and P is a projection operator that zeroes out unobserved entries in the data tensor T. The optimization routine performs alternating minimization over the factor matrices–coordinate descent–using the conjugate gradient method. This is one of the simpler tensor completion algorithms that one could imagine.

The projected ridge method attempts to solve a constrained version of the ridge problem

$$\min_{T' = U^{(1)}\circ\cdots\circ U^{(t)}} \sum_{i=1}^t \|U^{(i)}\|_F^2 \quad \text{s.t.} \quad \|P(T'-T)\|_F \le \delta. \tag{A.11}$$

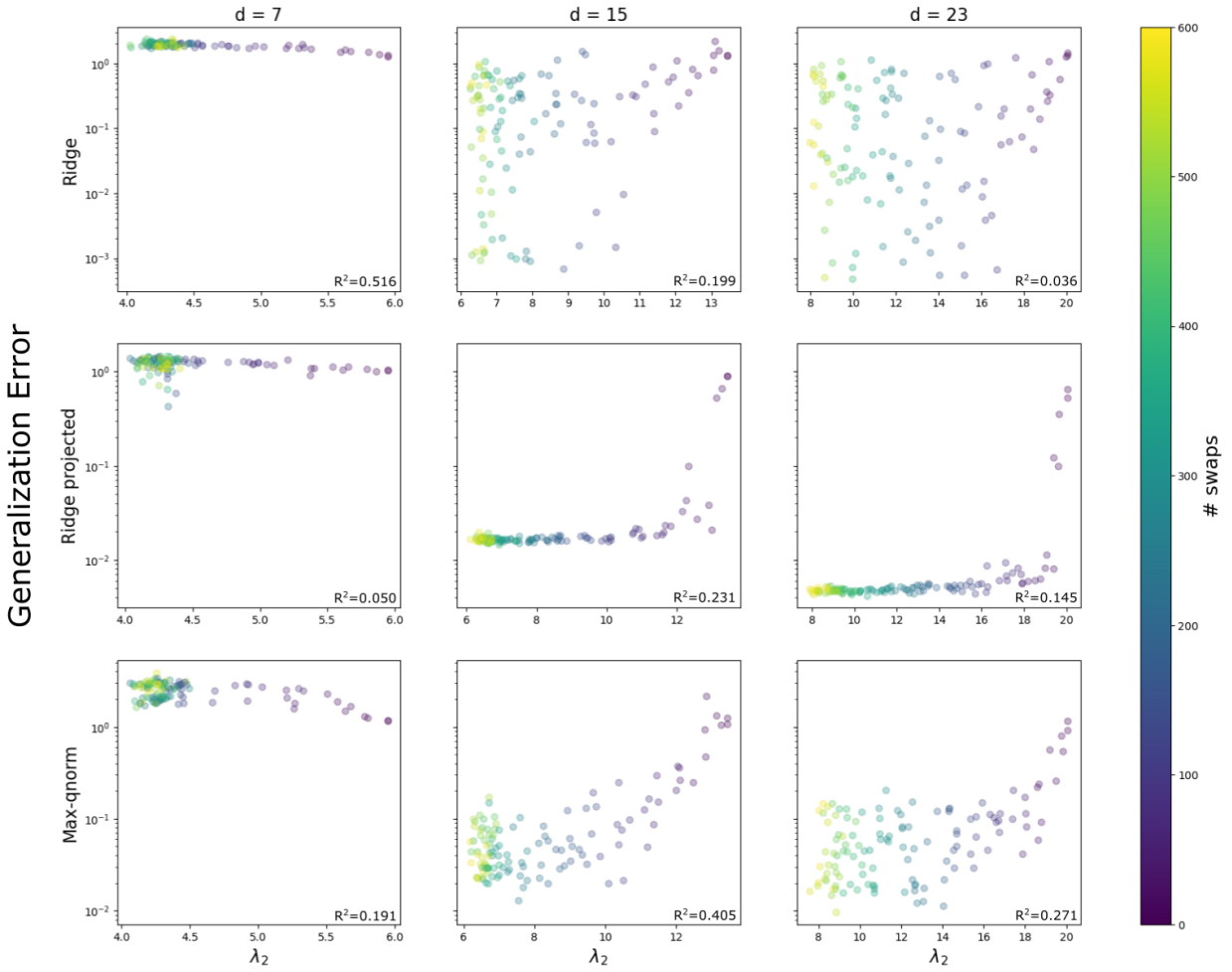

Figure 3: Numerical experiments show that reconstruction error (y-axes, log scale for visualization) correlates with $\lambda_2(G)$ for three different algorithms and varying graph degrees d =7, 15, 23. Coefficients of determination (R^2, linear fit of error $\sim \lambda_2$ *without* log scaling to be consistent with Fig. 2) are inlaid, and full regression reports are provided in the supplemental materials. In the sparsest sampling regime ($d = 7$), no method performs well, and the correlation is less consistent. In the denser regimes, the algorithms perform better, although ridge exhibits very high variance in error, while max-quasinorm and projected ridge are more consistent.

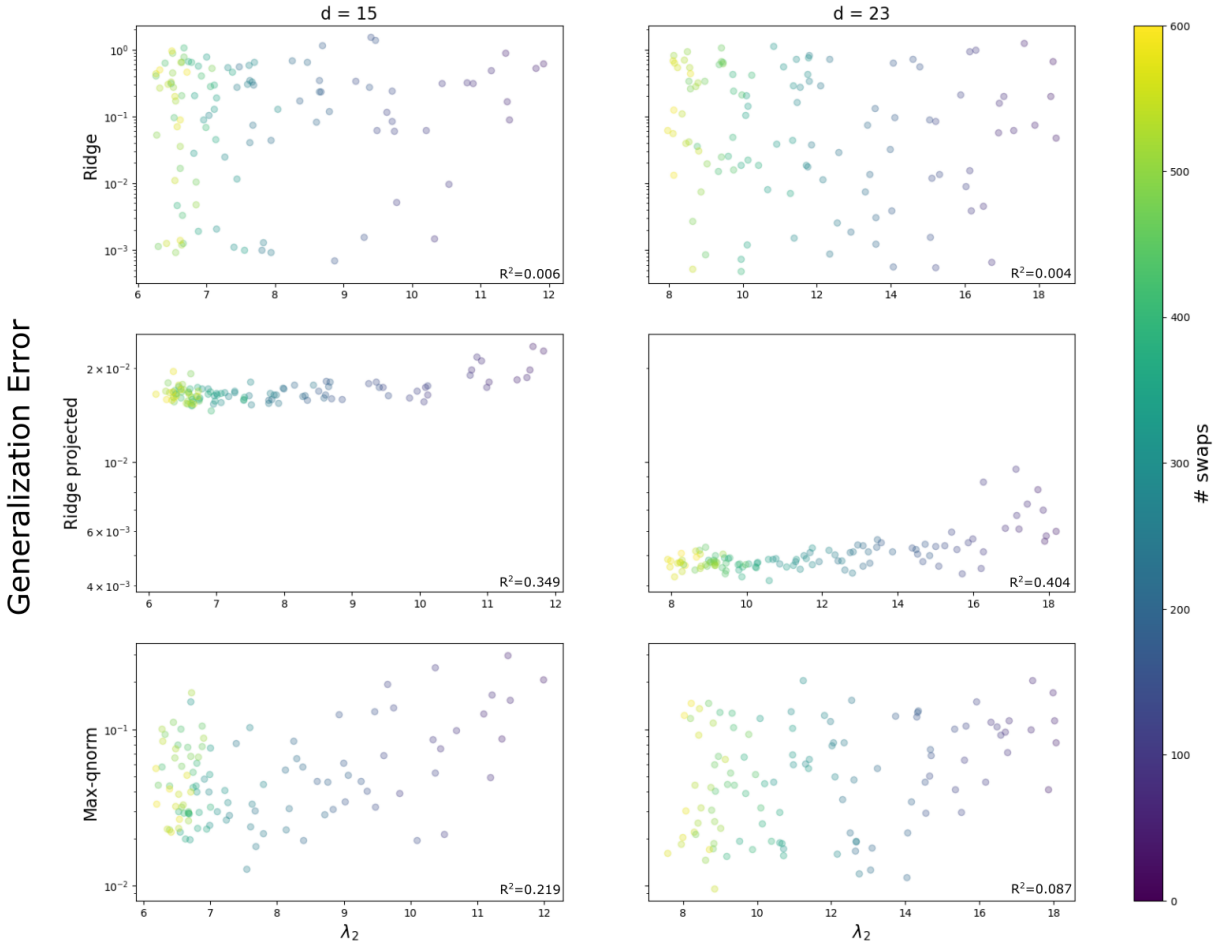

Figure 4: Same as for Fig. 3 except cutting off the largest λ_2 values and just depicting $d = 15, 23$. Numerical experiments show that reconstruction error (y-axes, log scale for visualization) correlates with $\lambda_2(G)$ for three different algorithms and varying graph degrees. Coefficients of determination (R^2, linear fit of error $\sim \lambda_2$ *without* log scaling to be consistent with Fig. 2) are inlaid, and full regression reports are provided in the supplemental materials. Here, the correlations are generally weaker except for the ridge projected algorithm, which has the best performance.

To deal with the hard constraint on the square residuals, we use an analogous relaxation and variable projection technique as the max-quasinorm algorithm from [16]. This leads to the relaxed optimization problem

$$\text{Ridge projected:} \quad \min_{T',R} \sum_{i=1}^{t} \|U^{(i)}\|_F^2 + \frac{\kappa}{2}\|P(T'-T-R)\|_F^2 + \beta\|R\|_F^2 \quad \text{s.t.} \quad \|R\|_F \leq \delta, \tag{A.12}$$

where again $T' = U^{(1)} \circ \cdots \circ U^{(t)}$. The parameters we used are $\delta = 0.05\sqrt{|E|}$, $\varepsilon = 0.01$, $r_{\text{fit}} = 10r$, $\kappa = 100$, $\beta = 1$, `rebalance = True`, `init = 'svdrand'`. Due to its intriguingly good performance (see below), we plan to study this algorithm in more detail in a future work.

The final algorithm is the max-quasinorm algorithm studied in detail in [16]. This algorithm is the closest to atomic norm minimization that we know of that's also practical. Atomic norm minimization is challenging since it would require integer optimization. For completeness, the optimization problem is

$$\text{Max-quasinorm:} \quad \min_{T',R} \|T\|_{\max} + \frac{\kappa}{2}\|P(T'-T-R)\|_F^2 + \beta\|R\|_F^2 \quad \text{s.t.} \quad \|P(R)\|_F \leq \delta. \tag{A.13}$$

The parameters for the max-quasinorm algorithm were $\delta = 0.05\sqrt{|E|}$, $\varepsilon = 0.01$, $r_{\text{fit}} = 10r$, $\kappa = 100$, $\beta = 1$, `rebalance = True`, `init = 'svdrand'`.

The target tensors were size $n = 100$, order $t = 3$, rank $r = 3$, and their factors were generated from a uniform distribution $U[0, 1]$ and rescaled to have Hilbert-Schmidt norm $\sqrt{n^t}$. After fitting $\hat{T}$, we measure generalization error $\|\hat{T} - T\|_F$ using the tensor Frobenius norm (root mean square error). Due to the normalization of the target tensor, an error of 1 corresponds to 100% relative error.

We sampled the tensor entries using graph lifting, where a d-regular graph G is lifted into a hypergraph by the t-path traversal method described in the main text. We start by taking the d-connected ring on n nodes, where each node is connected to its d nearest neighbors with periodic boundary conditions. This is a deterministic graph with eigenvalue $\lambda_2(G) \approx d-1$ in experiments. In order to vary $\lambda_2(G)$, we pick edge pairs at random and swap their endpoints. These edge swaps preserve the degree distribution, but after many swaps, the graph distribution approaches that of the d-regular random graph, which has $\lambda_2(G) \approx 2\sqrt{d-1}$, approximately as small as possible. We take $d \in \{7, 15, 23\}$ to generate hypergraphs with varying proportions 0.5%, 2.3%, 5.3% of observed entries.

Besides the results shown in Figure 2 for $d = 15$ and just the max-quasinorm algorithm, we show supporting results for other parameters and algorithms in Figure 3. These also show a significant correlation between $\lambda_2(G)$ and the reconstruction error of the tensor for all cases except $d = 7$. In that case, there are too few observations to correctly learn the tensor no matter the gap. We performed linear regression of error versus λ_2. Coefficients of determination are given in the plot, and the full regression reports are given in supplemental tables. Finally, Figure 4 shows the same data over a smaller range of λ_2.

A.9 Poisson loss experiments

We used the Poisson max-likelihood algorithm from [21] with code provided by those authors. The algorithm was run with its default parameters. The target tensors were size $n = 100$, order $t = 3$, and rank $r = 3$. Their factors were generated randomly from the uniform distribution and rescaled so the resulting tensor had entries in $[1, 6]$.

The sampling was performed without graph lifting. We found that a regularly spaced "grid" tensor had a large $\lambda_2(H)$, and that swapping grid entries with random entries caused the eigenvalue to decrease. To generate the grid for a particular sampling fraction, we uniformly spaced ones in a length n^3 linear array that gets reshaped into an $n \times n \times n$ array. To vary $\lambda_2(H)$, we then shuffle some fraction of those ones into random locations. In all experiments, we observe 5% of the entries and shuffle between 10% and 100% of those grid points with random locations.

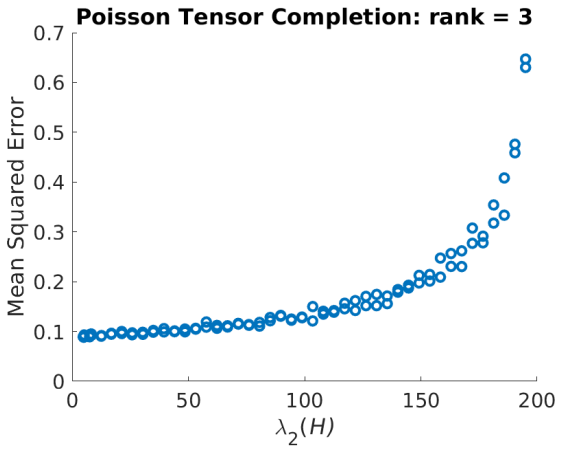

Figure 5: Results of Poisson tensor regression mean squared error for masks with varying $\lambda_2(H)$. A strong correlation is exhibited for target tensor with rank $r = 3$.

To estimate $\lambda_2(H)$, we fit a rank-1 tensor R to $(T_H - \frac{|E|}{n^t}J)$ using alternating least squares, so that $R \approx (T_H - \frac{|E|}{n^t}J)$. The Frobenius norm $\|R\|_F$ gives an estimate of the second eigenvalue. Figure 5 shows how the mean squared error $\frac{1}{n^3}\|\hat{T} - T\|_F^2$ varies with $\lambda_2(H)$. Again, we see a strong positive correlation between error and the eigenvalue.