SIAM J. MATH. DATA SCI.
Vol. 3, No. 4, pp. 1117–1140

Deterministic Tensor Completion with Hypergraph Expanders*

Kameron Decker Harris† and Yizhe Zhu‡

Abstract. We provide a novel analysis of low-rank tensor completion based on hypergraph expanders. As a proxy for rank, we minimize the max-quasinorm of the tensor, which generalizes the max-norm for matrices. Our analysis is deterministic and shows that the number of samples required to approximately recover an order-t tensor with at most n entries per dimension is linear in n, under the assumption that the rank and order of the tensor are $O(1)$. As steps in our proof, we find a new expander mixing lemma for a t-partite, t-uniform regular hypergraph model and prove several new properties about the tensor max-quasinorm. To the best of our knowledge, this is the first deterministic analysis of tensor completion. We develop a practical algorithm that solves a relaxed version of the max-quasinorm minimization problem, and we demonstrate its efficacy with numerical experiments.

Key words. hypergraph, expander, tensor completion, linear algebra

AMS subject classifications. 15A69, 05C90, 05C65, 05C48

DOI. 10.1137/20M1379745

1. Introduction.

1.1. Matrix and tensor completion.

Classical compressed sensing considers the recovery of high-dimensional structured signals from a small number of samples. These signals are typically represented by sparse vectors or low-rank matrices. A natural generalization is to study recovery of higher-order tensors, i.e., a multidimensional array of real numbers with more than two indices, using similar low-rank assumptions. However, much less is understood about compressed sensing of tensors.

Matrix completion is the problem of reconstructing a matrix from a subset of entries, leveraging prior knowledge such as its rank. The sparsity pattern of observed entries can be thought of as the adjacency or biadjacency matrix of a graph, where each edge corresponds to an observed entry in the matrix. There are two general sampling approaches studied for matrix completion. During probabilistic sampling, the entries in the matrix are observed at random according to Erdős–Rényi [34, 54], random regular bipartite [24, 9], or more general graph models [35]. Deterministic sampling, on the other hand, studies precisely what kinds of graphs are good for matrix completion and offers some advantages: One does not have to sample different entries for new matrices, and any recovery guarantees are deterministic

*Received by the editors December 9, 2020; accepted for publication (in revised form) July 28, 2021; published electronically October 6, 2021.
https://doi.org/10.1137/20M1379745
Funding: The first author was supported by a Washington Research Foundation Postdoctoral Fellowship. The second author was partially supported by NSF DMS-1949617.

†Department of Computer Science, Western Washington University, Bellingham, WA 98225 USA (kameron.harris@wwu.edu).
‡Department of Mathematics, University of California, San Diego, La Jolla, CA 92093 USA (yiz084@ucsd.edu).

without failure probability. It has been shown in [29, 6, 10] that expander graphs, which have pseudo-random properties, are a good way to sample deterministically for matrix completion. A deterministic theory of matrix completion based on graph limits, a different approach, appeared very recently [13].

Tensor completion, in which we observe a subset of the entries in a tensor and attempt to fill in the unobserved values, is a useful problem with a number of data science applications [43, 57]. But fewer numerical and theoretical linear algebra tools exist for working with tensors than for matrices. For example, computing the spectral norm of a tensor, computing its low rank decomposition, and computing the eigenvectors all turn out to be NP-hard [30].

Let the tensor of interest T be order-t, each dimension of size n, and have $\text{rank}(T) = r$, i.e., $T \in \bigotimes_{i=1}^{t} \mathbb{R}^n$ (we introduce our notation more fully in section 1.7). In this paper, tensor rank will always be using the canonical polyadic (CP) decomposition [36]. A fundamental question in tensor completion is how many observations via uniform sampling are required to guarantee recovery of an unknown tensor with high probability. A naive lower bound for the sample complexity is $\Omega(nrt)$, the number of unknown parameters in a CP decomposition.

Compared to the classical matrix completion problem, an important phenomenon in tensor completion is the trade-off between computational and statistical complexity. One way to reduce tensors to matrices is by flattening the order t-tensor into an $n^{\lceil t/2 \rceil} \times n^{\lceil t/2 \rceil}$ matrix, which can be solved in polynomial time. Yet current results using flattening have sample complexity at best $O\left(rn^{\lceil t/2 \rceil}\right)$ [25, 49]. A different approach using nuclear norm minimization was studied in [62] with sample complexity $\tilde{O}(n^{t/2})$, where $\tilde{O}(\cdot)$ hides polylog factors. However, it is shown in [22] that computing the nuclear norm of a given tensor is NP-hard. The best known polynomial time algorithms require $O(n^{t/2})$ sample complexity for an order-t tensor, including spectral algorithms [48, 60], gradient descent [11, 60], alternating minimization [32, 41], and convex relaxation via sum of squares [4, 53] or iterative collaborative filtering [56]. There is still a huge gap between the sample complexity of the existing polynomial algorithms and the statistical lower bound. In [4], Barak and Moitra conjectured that for an order-3 tensor, $\Omega(n^{3/2})$ many samples are needed for any polynomial algorithms by connecting it to the literature on refuting random 3-SAT. All of the above results are concerned with uniform sampling. With adaptive sampling, $O(n)$ sample complexity was obtained in [37, 63]. Very recently, Yu provided an algorithm that estimates a subclass of low-rank tensor with nearly linear samples [61].

In [26], Ghadermarzy, Plan, and Yilmaz studied tensor completion without reducing it to a matrix case by minimizing a max-quasinorm (satisfying all properties of the norm except a modified triangle inequality, which we call the "max-qnorm") 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 (see Definition 4.3 for further details). This is a generalization of the max-norm for matrices that many have shown yields good matrix completion results [58, 55, 21, 29, 12, 20]. Assuming that the observed entries are sampled from some probability distribution, it was shown that solving a max-qnorm constrained least-squares problem results in $O\left(\frac{nt}{\varepsilon^2}\right)$ sample complexity when $r = O(1)$ and even faster rates

in ε for the case of zero noise [26]. However, it is not clear if minimizing the max-qnorm is NP-hard.

We study the deterministic analogue of the tensor completion approach proposed in [26]. The deterministic analysis leads to a sample complexity which is also linear in n, albeit with weaker dependence on other parameters (see section 1.4). We assume that the observed entries correspond to the edges in an expander hypergraph. It has been known that revealing entries of a low-rank matrix according to the edges of an expander graph allows matrix completion with small errors [29, 6]. To the best of our knowledge, our work is the first generalization of such a connection to hypergraph expanders and tensor completion. A deterministic algorithm on low-rank tensor approximation was studied recently in [50] based on multiparty communication complexity. However, as pointed out by the authors of [50], their analysis cannot be applied to tensor completion problems directly.

1.2. Expanders and mixing. The expander mixing lemma for d-regular graphs (e.g., [14]) states the following: Let G be a d-regular graph on n vertices with $\lambda = \max\{\lambda_2, |\lambda_n|\} < d$. For any two sets $V_1, V_2 \subseteq V(G)$, let $e(V_1, V_2) = |\{(x, y) \in V_1 \times V_2 : xy \in E(G)\}|$ be the number of edges between V_1 and V_2. Then we have that

$$\left| e(V_1, V_2) - \frac{d|V_1||V_2|}{n} \right| \leq \lambda \sqrt{|V_1||V_2| \left(1 - \frac{|V_1|}{n}\right)\left(1 - \frac{|V_2|}{n}\right)}. \tag{1.1}$$

Inequality (1.1) tells us regular graphs with small λ have the *expansion property*, where the number of edges between any two sets is well approximated by the number of edges we would expect if they were drawn at random. The quality of such an approximation is controlled by λ. It is known from the Alon–Boppana bound that $\lambda \geq 2\sqrt{d-1} - o(1)$, and regular graphs that achieve this bound are called Ramanujan. Deterministic and random constructions of Ramanujan (or nearly so) graphs have been extensively studied [5, 45, 8, 10, 47].

Higher-order, i.e., hypergraph, expanders have received significant attention in combinatorics and theoretical computer science [44]. There are several expander mixing lemmas in the literature based on the spectral norm of tensors [23, 52, 15, 39]. An obstacle to applying such results to tensor completion is that in most cases the second eigenvalues of tensors are unknown, even approximately. In [19], an expander mixing lemma similar to (1.1) based on the second eigenvalue of the adjacency matrix of regular hypergraphs was derived. However, for our application we need an expander mixing lemma that estimates the number of hyperedges among t different vertex subsets.

One exception is the work of Friedman and Wigderson [23], who studied a t-uniform hypergraph model on n vertices with dn^{t-1} hyperedges chosen randomly with $d \geq C \log n$. They proved that the second eigenvalue of the associated tensor $\lambda = O((\log n)^{t/2}\sqrt{d})$. However, this is a relatively dense random hypergraph model, since the number of edges grows superlinearly with n for $t > 2$. Thus Friedman and Widgerson's model only applies when one has the ability to make many measurements, as opposed to the more realistic "big data" scenario constrained to $\tilde{O}(n)$ observations. If we sample the original tensor according to the hyperedge set of a hypergraph, we would like the number of hyperedges to be small, in order to be able to represent a small number of samples. From previous results on matrix completion [29, 6],

we expect that the reconstruction error should be controlled by a parameter that is related to the expansion property of the hypergraph.

Our tensor completion analysis can also be applied to observation models where entries are revealed according to a general t-uniform hypergraph. Define the *spectral norm* of a tensor T as

$$\|T\| = \sup_{v_1,\dots,v_k \in S^{n-1}} \left| \sum_{i_1,\dots,i_k=1}^{n} T_{i_1,\dots,i_k} v_1(i_1) \cdots v_k(i_k) \right|, \tag{1.2}$$

where S^{n-1} is the unit sphere in $\mathbb{R}^n$. The mixing lemma for any t-uniform hypergraph was first considered in [23] and was generalized in [15]. We state the expander mixing lemma for t-partite t-uniform hypergraphs as follows. A quick proof of Lemma 1.1 is provided in section SM1.1 of the supplement.

Lemma 1.1. *Let $H = (V, E)$ be a t-uniform t-partite hypergraph with vertex set $V = V_1 \cup \cdots \cup V_t$ and adjacency tensor T_H. For any subsets $W_1 \subset V_1, \dots, W_t \subset V_t$, define*

$$e(W_1, \dots, W_t) = |\{e = (v_1, \dots, v_t) \in E, v_i \in W_i, 1 \le i \le t\}|.$$

Then the following holds:

$$\left| e(W_1, \dots, W_t) - \frac{|E|}{n^t} |W_1| \cdots |W_t| \right| \le \lambda_2(H) \sqrt{|W_1| \cdots |W_t|}, \tag{1.3}$$

where

$$\lambda_2(H) = \left\| T_H - \frac{|E|}{n^t} J \right\|, \tag{1.4}$$

and J is the all-ones tensor.

1.3. Main results. In this paper we seek the connection between two topics, hypergraph expanders and tensor completion, using the tensor max-quasinorm introduced in [26].

We revisit and generalize the sparse, deterministic hypergraph construction introduced in [7]. We construct a t-uniform t-partite hypergraph by taking a d-regular "base" graph and forming a hypergraph from its walks of length t. In this model, each node is of degree d^{t-1}, corresponding to nd^{t-1} samples. An advantage of our expander mixing result is that the expansion property of the hypergraph is controlled by the expansion in a d-regular graph (Theorem 3.1). This is easy to compute and optimize using known constructions of d-regular expanders. Based on such hypergraphs, we perform a deterministic analysis of an optimization problem similar to that analyzed by Ghadermarzy et al. [26].

Our main contributions can be summarized as follows: First, we obtain a variant of the expander mixing inequalities from [7, 3] and generalize the t-partite t-uniform regular hypergraph construction in [7] to t-partite t-uniform quasi-regular hypergraphs; see sections 2 and 3. The new expander mixing result provides a better error control for tensor completion. This improvement might also be useful for the application of hypergraph codes studied in [7]. Next, we perform a deterministic analysis of an optimization problem similar to that analyzed

by Ghadermarzy, Plan, and Yilmaz [26]. Our proof is based on the techniques used to study matrix completion in [29] (see also [9]). We prove several useful linear algebra facts about the max-quasinorm for tensors in order to prove the main results on the tensor completion error; these facts may also be of interest on their own. Finally, we show proof-of-concept numerical results on minimal max-quasinorm completion.

For a deterministic hypergraph H, if we have a good estimate of $\lambda_2(H)$ defined in (1.4), we can obtain the following tensor completion error bound. Here we state it for t-uniform t-partite hypergraphs to avoid symmetric sampling, which might be wasteful and will increase the sample size by a factor of $t!$, but the analysis can be extended to general t-uniform hypergraphs.

Theorem 1.2. *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 $\lambda_2(H)$. Then solving*

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

will result in the following error bound:

$$\frac{1}{n^t}\|\hat{T} - T\|_F^2 \le \frac{2^{2t} n^{t/2} K_G^{t-1} \lambda_2(H)}{|E|} \|T\|_{\max}^2, \tag{1.5}$$

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

Although it is NP hard to compute $\lambda_2(H)$ for a given hypergraph H, in [23, 32, 15, 64], the upper bound on $\lambda_2(H)$ is provided for some random hypergraph models. Let T_H be the adjacency tensor of a given H. From the definition of $\lambda_2(H)$ and (1.2), suppose H has $|E| = o(n^t)$ many hyperedges; we have

$$\lambda_2(H) \ge \max_{i_1,\ldots,i_t} \left| (T_H)_{i_1,\ldots,i_t} - \frac{|E|}{n^t} \right| = \Omega(1).$$

Therefore, if we want to use (1.5) to control the mean square error by ε, the number of samples must be $\Omega(n^{t/2})$. In [32], an estimate of $\lambda_2(H)$ for 3-uniform Erdős–Rényi random hypergraphs with $p = \frac{c \log n}{n^{3/2}}$ was obtained and our Theorem 1.2 can be applied.

However, using a hypergraph expander model based on regular graphs, we can get a better error bound without using $\lambda_2(H)$. We state our main results on the deterministic bounds for tensor completion based on hypergraph expander models. A formal definition of this hypergraph model is given in section 2. When $t = 2$, it reduces to the result in [29].

Theorem 1.3. *Given a hypercubic tensor T of order t, reveal its entries according to a t-partite, t-uniform, d^{t-1}-regular hypergraph $H = (V, E)$ constructed from a d-regular graph G of size n with second eigenvalue (in absolute value) $\lambda \in (0, d)$ (see section 2.1). Then solving*

$$\hat{T} = \arg\min_{T'} \|T'\|_{\max} \quad \text{such that} \quad T'_e = T_e \quad \text{for all} \quad e \in E \tag{1.6}$$

will result in the following mean squared error bound:

$$\frac{1}{n^t}\|\hat{T} - T\|_F^2 \le C_t \|T\|_{\max}^2 \frac{\lambda}{d}, \tag{1.7}$$

where $C_t = 2^t(2t-3)K_G^{t-1}$, *and* $K_G \leq 1.783$ *is Grothendieck's constant over* $\mathbb{R}$.

Remark 1.4. From Theorem 4.6, $\|T\|_{\max}^2 \leq r^{t^2-t-1}|T|_\infty$, where r is the rank of the tensor T. The number of revealed entries in T is nd^{t-1}. The analysis of sample complexity for the algorithm is given in section 1.4.

If the tensor T is not hypercubic, say $T \in \bigotimes_{i=1}^t \mathbb{R}^{n_i}$, we can still apply Theorem 1.3 by taking $n = \max_{1\leq i\leq t} n_i$ and embed T in $\bigotimes_{i=1}^t \mathbb{R}^n$ by filling extra entries with zeros. But when $n_1, \ldots, n_t$ are heterogeneous, this might be wasteful. In order to handle heterogeneous dimensions, we generalize the construction in [3, 7] to construct t-uniform t-partite quasi-regular hypergraphs with good expansion properties, based on bipartite biregular expanders (see section 3.2), which yields the following theorem. The proof of Theorem 1.5 is similar to that of Theorem 1.3, and we include it in section SM1.2.

Theorem 1.5. *Given a tensor* $T \in \bigotimes_{i=1}^t \mathbb{R}^{n_i}$, *reveal its entries according to a* t*-partite,* t*-uniform, quasi-regular hypergraph* $H = (V, E)$ *constructed from a collection of* (d_{2i-1}, d_{2i})*-biregular bipartite graphs with second eigenvalues* $\lambda^{(i)}$ *for* $1 \leq i \leq t-1$ *(see section 2.2). Then solving*

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

will result in the following mean squared error bound:

$$\frac{1}{\prod_{i=1}^t n_i}\|\hat{T} - T\|_F^2 \leq 2^t K_G^{t-1}\left(\frac{\lambda^{(1)}}{\sqrt{d_1 d_2}} + \sum_{k=2}^{t-1}\frac{2\lambda^{(k)}}{\sqrt{d_{2k-1}d_{2k}}}\right)\|T\|_{\max}^2.$$

When $t = 2$, Theorem 1.5 reduces to Theorem 24 in [9] for deterministic matrix completion with bipartite biregular graphs.

The main motivation to study the mixing properties of hypergraph expanders in sections 2 and 3 is that the key parameter to control the mixing properties is the spectral gap, which is well understood in spectral graph theory and easy to compute. This gives us explicit error control in Theorems 1.3 and 1.5.

1.4. Sample complexity. We focus on the sample complexity analysis for the hypergraph model we used in Theorem 1.3, and other models can be analyzed in a similar way. Recall that the number of edges in H is nd^{t-1}, equal to the number of samples. Suppose we have an expander graph G, where $\lambda = O(\sqrt{d})$. In order to guarantee the right-hand side in (1.7) is bounded by ε, Theorems 1.3 and 4.6 say that, assuming $t = O(1)$, we require

$$|E| = O(\|T\|_{\max}^{4t-4}\varepsilon^{-2(t-1)}n) = O\left(\frac{nr^{2(t-1)(t^2-t-1)}}{\varepsilon^{2(t-1)}}\right) \tag{1.8}$$

samples, which is linear in n. The computations are shown in section SM1.12.

In Theorem 1.3, the dependence on rank is exponential in t, since the best known dependence of the max-qnorm on rank is $\|T\|_{\max} = O(\sqrt{r^{t^2-t-1}})$. For matrices, our results have $|E| = O(nr^2/\varepsilon^2)$, which is the same sample complexity derived in [29]. A better understanding of the max-qnorm for tensors may lead to better dependence on the rank.

The sample complexity in Theorem 1.2 can be similarly analyzed. To guarantee an ε mean squared error, we have $|E| = O(\lambda_2(H)\|T\|_{\max}^2 \varepsilon^{-1} n^{t/2})$. The dependence on $\|T\|_{\max}$ and ε in this bound is better, but the dependence on n is much weaker compared to (1.8).

1.5. Computational complexity. The computational complexity for solving the optimization problem (1.6) is not clear. Solving it might be NP-hard, but we provide a practical algorithm to approximately solve it.

1.6. Organization of the paper. In section 2, we construct the hypergraph expanders with good mixing properties. In section 3, we prove an expander mixing lemma for such hypergraphs. In section 4, we prove several useful properties of the max-quasinorm for tensors. In section 5, we leverage these properties to analyze the above tensor completion algorithm and prove the main results. We extend our result for tensor completion with errors in the observed entries, which can model noise or adversarial corruptions. In section 6 we provide a numerical algorithm for finding tensors with the minimum complexity. We conclude with a discussion of limitations and future directions in section 7. Omitted proofs are provided in section SM1 of the supplement.

1.7. Notation. The notation we use throughout the paper comes from the review by Kolda and Bader [36]. We use lowercase symbols u for vectors and 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, which will be defined in section 4. We use $\bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ for the space of all order-t tensors with n_i entries in the ith 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 entrywise ℓ_p norms of matrices and tensors and always include the subscript to avoid confusion with set cardinality.

2. Construction of hypergraph expanders.

We start with the definition of a hypergraph and some basic properties.

Definition 2.1 (hypergraph). *A* hypergraph H *consists of a set of vertices V and a set of hyperedges E, where each hyperedge is a nonempty set of V, the vertices that participate in that hyperedge. The hypergraph H is* t-uniform *for an integer $t \geq 2$ if every hyperedge $e \in E$ contains exactly t vertices. The* degree *of vertex i is the number of all hyperedges containing i. A hypergraph is* d-regular *if all of its vertices have degree d.*

A t-uniform hypergraph is t-partite *if its vertex set V can be decomposed as $V_1 \cup V_2 \cup \cdots \cup V_t$ such that each hyperedge $e \in E$ consists of t vertices $v_1, \dots, v_t$ such that $v_i \in V_i$ for $1 \leq i \leq t$. For a t-uniform, t-partite hypergraph $H = (V, E)$, we denote each hyperedge e as an ordered tuple $(v_1, \dots, v_t)$ where $v_k \in V_k, 1 \leq k \leq t$.*

Definition 2.2 (adjacency tensor for t-uniform t-partite hypergraphs). *Let $H = (V, E)$ be a*

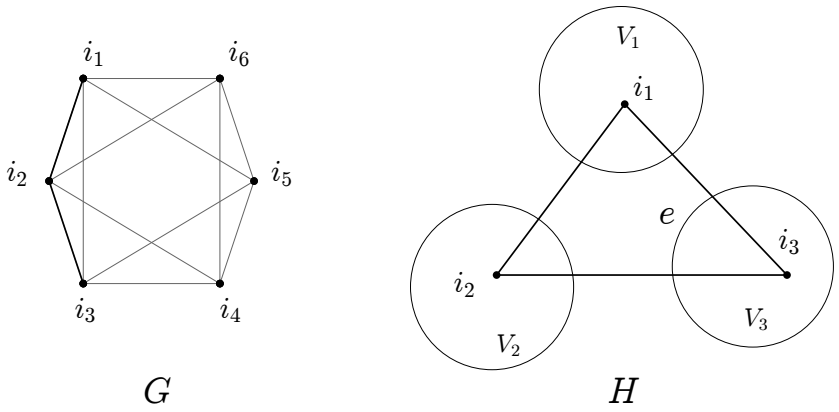

Figure 2.1. *An example hyperedge in the $t = 3$ case: We depict the base graph G on the left and a single edge in the hypergraph H on the right. (i_1, i_2, i_3) forms an hyperedge e in H if and only if (i_1, i_2, i_3) is a walk in G.*

t-uniform t-partite hypergraph with vertex set $V = V_1 \cup \cdots \cup V_t$ such that $|V_k| = n_k, 1 \le k \le t$. We define the adjacency tensor $T \in \otimes_{k=1}^{t} \{0,1\}^{n_k}$ as

$$T_{i_1,\ldots,i_t} = \begin{cases} 1 & \text{if } (i_1, \ldots, i_t) \in E, i_k \in V_k, 1 \le k \le t, \\ 0 & \text{otherwise.} \end{cases}$$

2.1. Construction of regular hypergraph expanders. Let $G = (V(G), E(G))$ be a connected d-regular graph on n vertices with second largest eigenvalue (in absolute value) $\lambda \in (0, d)$. We construct a t-partite, t-uniform, d^{t-1}-regular hypergraph $H = (V, E)$ from G as follows.

Definition 2.3 (regular hypergraph expander). *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.*

Given the description above, we have $|V| = nt$ and $|E| = nd^{t-1}$, since E contains all possible walks of length $t-1$ in G. Moreover, every vertex is contained in exactly d^{t-1} many hyperedges, so H is regular. From our definition of the hyperedges in H, the order of the walk in G matters. For example, two walks $i_1 \to i_2 \to i_3$ and $i_3 \to i_2 \to i_1$ correspond to different hyperedges (i_1, i_2, i_3) and (i_3, i_2, i_1) in H when $i_1 \neq i_3$. When $t = 2$, H is a bipartite d-regular graph with $2n$ vertices. See Figure 2.1 for an example of the construction with $t = 3$.

This construction was used by [3] to derandomize graph products and by [7] to construct error correcting codes. Both groups' results depended on analyzing the expansion properties of this hypergraph model.

2.2. Generalization to quasi-regular hypergraphs.

Definition 2.4 (bipartite biregular graph). *A graph $G = (V, E)$ is* bipartite *if the vertex set V can be partitioned into disjoint vertex sets V_1, V_2 such that every edge connects a vertex in*

V_1 to a vertex in V_2. A bipartite graph G is (d_1, d_2)-biregular *if each vertex in V_1 has degree d_1 and each vertex in V_2 has degree d_2.*

We can also generalize our construction in section 2.1 to t-partite quasi-regular hypergraph H with $|V_1| = n_1, \dots, |V_t| = n_t$. The idea is to combine $t-1$ many bipartite biregular graphs in the following way.

Definition 2.5 (quasi-regular hypergraph expander). *Take $\{d_i\}_{1\le i\le 2t-2}$ such that $n_i d_{2i-1} = n_{i+1} d_{2i}$ for $1 \le i \le t-1$. Let $G_1, \dots, G_{t-1}$ be bipartite biregular graphs such that G_i is (d_{2i-1}, d_{2i})-biregular with two vertex sets of size n_i and n_{i+1}, respectively. We construct H such that a hyperedge $e = (v_1, \dots, v_t)$ is in H if and only if $(v_i, v_{i+1}) \in E(G_i)$ for all $1 \le i \le t-1$. Now we have $|E(H)| = n_1 \prod_{i=1}^{t-1} d_{2i-1}$. Denote the second largest eigenvalue of G_i by $\lambda^{(i)}$.*

Note that the hypergraph H is not regular, but each vertex in V_i has the same degree for $1 \le i \le t$, as shown in the following lemma. The proof of Lemma 2.6 is given in section SM1.3.

Lemma 2.6. *Let H be the t-uniform t-partite quasi-regular hypergraph defined above; then the degree of each vertex in V_i is $(\prod_{k=1}^{i-1} d_{2k})(\prod_{k=i}^{t-1} d_{2k-1})$.*

Deterministic or random construction of a bipartite biregular graph with a small second eigenvalue was considered in [10, 9], which can be used to construct the quasi-regular hypergraph in Definition 2.5.

3. Expander mixing.

In this section, we prove novel and tighter expansion properties in the hypergraph models we constructed in section 2. Later, we will apply them to tensor completion.

3.1. Regular hypergraph expanders.

Let G be a d-regular graph on n vertices with $\lambda \in (0, d)$, and let H be the corresponding t-partite, t-uniform hypergraph constructed as in section 2.1. We get the following mixing lemma for H. The mixing rate is essentially controlled by the second eigenvalue of the d-regular graph G. This is an advantage over other expander mixing lemmas for hypergraphs [23, 15, 52], since in our model the parameters that control the mixing rate are explicit and easy to compute.

Theorem 3.1. *Given a base graph G with $\lambda \in (0, d)$, form the hypergraph H following the construction in section 2. Let $W_i \subseteq V_i$, $1 \le i \le t$, be any nonempty subsets. Denote $\alpha_i := \frac{|W_i|}{n} \in [0, 1], 1 \le i \le t$, and let*

$$e(W_1, \dots, W_t) := |\{(v_1, \dots, v_t) \in W_1 \times \cdots \times W_t : (v_1, \dots, v_t) \in E(H)\}|.$$

Then the following expander mixing property holds:

$$\left| \frac{e(W_1, \dots, W_t)}{n d^{t-1}} - \prod_{i=1}^{t} \alpha_i \right|$$

$$\le \frac{\lambda}{d} \left(\sqrt{\alpha_1(1-\alpha_1)\alpha_2(1-\alpha_2)} \prod_{k=3}^{t} \alpha_k + \sum_{i=3}^{t} \sqrt{\alpha_1 \alpha_i (1-\alpha_i)} \prod_{k=i+1}^{t} \alpha_k \right) \le \frac{(2t-3)\lambda}{4d}. \tag{3.1}$$

Remark 3.2. When $t = 2$, Theorem 3.1 reduces to the expander mixing lemma (1.1) for regular graphs. The results in [7, 3] estimate the ratio between $\frac{e(W_1,\dots,W_t)}{nd^{t-1}}$ and $\prod_{i=1}^{t}\alpha_i$. In particular, Lemma 3 in [7] only provides an upper bound, and Theorem 4 of [2] gives a two-sided bound with an extra assumption that $\alpha_i > 6\lambda/d$ for all $1 \le i \le t$. Our new expander mixing result provides a two-sided control of the difference between the two quantities without any restriction on α_i, which is important for our tensor completion analysis, and the results in [3, 7] are not directly applicable. The ratio $\frac{e(W_1,\dots,W_t)}{nd^{t-1}}$ describes the hitting property of expander walks, which has many applications in theoretical computer science [1, 33, 31, 16].

Let A be the adjacency matrix of G and $M = \frac{1}{d}A$ be the transition probability matrix for the simple random walk on G. Define $u = \frac{1}{n}(1,\dots,1)^\top$. Then u is the first eigenvector of M with eigenvalue 1 and the second eigenvalue of M is λ/d. Define

$$P_k = \sum_{i\in W_k} e_i e_i^\top, \quad \mathbf{1}_{W_k} = \sum_{i\in W_k} e_i, \tag{3.2}$$

where e_i is the basis vector with 1 in the ith coordinate and 0 elsewhere. Then P_k is a projection matrix such that $P_k e_i = \mathbf{1}\{i \in W_k\}e_i$. Let P be the projection onto the orthogonal space of u. As shown in [3], considering the simple random walk of t steps in G with a uniformly chosen initial vertex, the probability that the simple random walk stays in W_i at step i for all $1 \le i \le t$ can be written as

$$\frac{e(W_1,\dots,W_t)}{nd^{t-1}} = \|P_t M P_{t-1} M \cdots P_2 M P_1 u\|_1. \tag{3.3}$$

Let $v_1 = P_1 u = \frac{1}{n}\mathbf{1}_{W_1}, v_{i+1} = P_{i+1}Mv_i$. Then

$$\frac{e(W_1,\dots,W_t)}{nd^{t-1}} = \|v_t\|_1, \quad \|v_1\|_1 = \alpha_1, \quad \|v_1\|_2 = \frac{\sqrt{\alpha_1}}{\sqrt{n}}. \tag{3.4}$$

Decompose $v_i = x_i + y_i$, where x_i is the part of v that is a scalar multiple of u, and $y_i = Pv_i$ is orthogonal to u. We first prove the following lemma. The argument is based on the proof of the expander mixing lemma (1.1) for d-regular graphs (see, e.g., [14]).

Lemma 3.3. *For* $1 \le i \le t-1$,

$$|\|v_{i+1}\|_1 - \alpha_i\|v_i\|_1| \le \frac{\lambda}{d}\sqrt{n\alpha_{i+1}(1-\alpha_{i+1})}\|y_i\|_2.$$

Proof. Since the entries of v_i represent probabilities, each v_i is entrywise nonnegative. Denote $\mathbf{1} = (1,\dots,1)^\top$. We have

$$\begin{aligned}
\|v_{i+1}\|_1 &= \|P_{i+1}Mv_i\|_1 = \mathbf{1}_{W_{i+1}}^\top Mv_i \\
&= \alpha_{i+1}\mathbf{1}^\top M\left(\frac{\|v_i\|_1}{n}\mathbf{1}\right) + \left(\mathbf{1}_{W_{i+1}} - \alpha_{i+1}\mathbf{1}\right)^\top M\left(v_i - \frac{\|v_i\|_1}{n}\mathbf{1}\right) \\
&\quad + \left(\mathbf{1}_{W_{i+1}} - \alpha_{i+1}\mathbf{1}\right)^\top M\frac{\|v_i\|_1}{n}\mathbf{1} + \alpha_{i+1}\mathbf{1}^\top M\left(v_i - \frac{\|v_i\|_1}{n}\mathbf{1}\right).
\end{aligned}$$

Note that, since $\mathbf{1}$ is the eigenvector of M and M^T with eigenvalue 1,

$$(\mathbf{1}_{W_{i+1}} - \alpha_{i+1}\mathbf{1})^\top M \frac{\|v_i\|_1}{n}\mathbf{1} = 0, \quad \alpha_{i+1}\mathbf{1}^\top M\left(v_i - \frac{\|v_i\|_1}{n}\mathbf{1}\right) = 0.$$

We obtain

$$\|v_{i+1}\|_1 = \alpha_{i+1}\|v_i\|_1 + \left(\mathbf{1}_{W_{i+1}} - \alpha_{i+1}\mathbf{1}\right)^\top M\left(v_i - \frac{\|v_i\|_1}{n}\mathbf{1}\right).$$

Therefore, by the Cauchy inequality, $|\|v_{i+1}\|_1 - \alpha_{i+1}\|v_i\|_1|$ can be bounded by

$$\begin{aligned}\left|\left(\mathbf{1}_{W_{i+1}} - \alpha_{i+1}\mathbf{1}^\top\right) M\left(v_i - \frac{\|v_i\|_1}{n}\mathbf{1}\right)\right| &\le \frac{\lambda}{d}\|\mathbf{1}_{W_{i+1}} - \alpha_{i+1}\mathbf{1}^\top\|_2 \left\|v_i - \frac{\|v_i\|_1}{n}\mathbf{1}\right\|_2 \\ &= \frac{\lambda}{d}\sqrt{n\alpha_{i+1}(1-\alpha_{i+1})}\|y_i\|_2.\end{aligned}$$

■

With Lemma 3.3, we finish the proof of Theorem 3.1.

Proof of Theorem 3.1. We would like to control $\|y_i\|_2$. Note that

$$\|y_1\|_2 = \left\|v_1 - \frac{\|v_1\|_1}{n}\mathbf{1}\right\|_2 = \frac{1}{n}\left\|\mathbf{1}_{W_1} - \frac{|W_1|}{n}\mathbf{1}\right\|_2 = \sqrt{\frac{\alpha_1(1-\alpha_1)}{n}}.$$

For $i \ge 2$, $\|y_i\|_2 \le \|v_i\|_2 = \|P_i M v_{i-1}\|_2 \le \|v_{i-1}\|_2$. We obtain $\|y_i\|_2 \le \|v_1\|_2 = \sqrt{\alpha_1/n}$. From Lemma 3.3, we have

$$\begin{aligned}|\|v_t\|_1 - \alpha_t\|v_{t-1}\|_1| &\le \frac{\lambda}{d}\sqrt{\alpha_1\alpha_t(1-\alpha_t)}, \\ |\alpha_t\|v_{t-1}\|_1 - \alpha_{t-1}\alpha_t\|v_{t-2}\|_1| &\le \frac{\lambda}{d}\alpha_t\sqrt{\alpha_1\alpha_{t-1}(1-\alpha_{t-1})}, \\ &\vdots \\ |\alpha_4\cdots\alpha_t\|v_3\|_1 - \alpha_3\cdots\alpha_t\|v_2\|_1| &\le \frac{\lambda}{d}(\alpha_4\cdots\alpha_t)\sqrt{\alpha_1\alpha_3(1-\alpha_3)}, \\ |\alpha_3\cdots\alpha_t\|v_2\|_1 - \alpha_2\cdots\alpha_t\|v_1\|_1| &\le \frac{\lambda}{d}(\alpha_3\cdots\alpha_t)\sqrt{\alpha_1(1-\alpha_1)\alpha_2(1-\alpha_2)}. \end{aligned} \tag{3.5}$$

Since $\|v_1\|_1 = \alpha_1$, for $t \ge 2$, combining the inequalities above and applying the triangle inequality leads to

$$\begin{aligned}&|\|v_t\|_1 - \alpha_1\cdots\alpha_t| \\ &\le \frac{\lambda}{d}\left(\alpha_3\cdots\alpha_t\sqrt{\alpha_1(1-\alpha_1)\alpha_2(1-\alpha_2)} + \sum_{i=3}^{t}\sqrt{\alpha_1\alpha_i(1-\alpha_i)}\prod_{k=i+1}^{t}\alpha_k\right).\end{aligned} \tag{3.6}$$

In the $i = t$ term, the empty product is defined to equal 1. Using the inequality $\sqrt{x(1-x)} \le \frac{1}{2}$ for any $x \in [0,1]$ and that $\alpha_i \in [0,1]$, (3.6) implies

$$\left|\frac{e(W_1,\dots,W_t)}{nd^{t-1}} - \alpha_1\cdots\alpha_t\right| \le \frac{\lambda}{d}\left(\frac{1}{4} + \frac{t-2}{2}\right). \tag{3.7}$$

This completes the proof. ■

3.2. Quasi-regular hypergraph expanders. The quasi-regular hypergraph constructed in section 2.2 has a similar expansion property as follows.

Theorem 3.4. *Let H be the hypergraph constructed in Definition 2.5. Let $W_i \subseteq V_i$, $1 \le i \le t$, be any nonempty subsets. Denote $\alpha_i := \frac{|W_i|}{n_i} \in [0,1], 1 \le i \le t$, and let $e(W_1, \dots, W_t)$ be the number of hyperedges between $W_1, \dots, W_t$. Then the following expansion property holds:*

$$\begin{aligned}
&\left| \frac{e(W_1, \dots, W_t)}{n_1 d_1 \cdots d_{2t-3}} - \prod_{i=1}^{t} \alpha_i \right| \\
&\le \frac{\lambda^{(1)}}{\sqrt{d_1 d_2}} \sqrt{\alpha_1(1-\alpha_1)\alpha_2(1-\alpha_2)} \prod_{k=3}^{t} \alpha_k + \sum_{k=2}^{t-1} \frac{\lambda^{(k)}}{\sqrt{d_{2k-1} d_{2k}}} \sqrt{\alpha_1 \alpha_{k+1}(1-\alpha_{k+1})} \prod_{i=k+2}^{t} \alpha_i \\
&\le \frac{\lambda^{(1)}}{4\sqrt{d_1 d_2}} + \sum_{k=2}^{t-1} \frac{\lambda^{(k)}}{2\sqrt{d_{2k-1} d_{2k}}}.
\end{aligned}$$

When $t = 2$, Theorem 3.4 reduces to the expander mixing lemma for bipartite biregular graphs [28, 17]. We present the proof of Theorem 3.4 in section SM1.4 of the supplement.

4. Tensor complexity. In order to complete a partially observed matrix or tensor, some kind of prior knowledge of its structure is required. The tensor that is output by the learning algorithm will then be the least complex one that is consistent with the observations. Consistency may be defined as either exactly matching the observed entries—in the case of zero noise—or being close to them under some loss—in the case where the observations are corrupted. We now argue for the use of the tensor max-quasinorm (see Definition 4.3 below) as a measure of complexity. Towards this aim, we also show a number of previously unknown properties about the max-quasinorm.

For matrices, the most common measure of complexity is the rank. In the tensor setting, there are various definitions of rank [36]. However, in this paper we will work with the rank defined via the CP decomposition as

$$\operatorname{rank}(T) = \min \left\{ r \;\middle|\; T = \sum_{i=1}^{r} u_i^{(1)} \circ \cdots \circ u_i^{(t)} \right\}, \tag{4.1}$$

where each vector $u_i^{(j)} \in \mathbb{R}^n$. Note that the decomposition above is atomic and equivalent to the decomposition used to define matrix rank when $t = 2$. The sum is composed of r rank-1 tensors expressed as the outer products $u_i^{(1)} \circ \cdots \circ u_i^{(t)}$. Our analysis uses Kronecker and Hadamard products of tensors. These are generalizations of the usual definitions for matrices in the obvious way.

Definition 4.1. *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*

$$(T \otimes S)_{k_1,\dots,k_t} = T_{i_1,\dots,i_t} S_{j_1,\dots,j_t} \quad \text{for} \quad k_1 = j_1 + m_1(i_1 - 1), \dots, k_t = j_t + m_t(i_t - 1).$$

Definition 4.2. *Let $T \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ and $S \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$. We define the* Hadamard product *of two tensors $(T * S) \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$ as the tensor with indices $(T * S)_{i_1,\dots,i_t} = T_{i_1,\dots,i_t} S_{i_1,\dots,i_t}$.*

4.1. Max-qnorm. The max-norm of a matrix (also called γ_2 norm) is a common relaxation of rank. It was originally proposed in the theory of Banach spaces [59] but has found applications in communication complexity [40, 38, 46] and matrix completion [58, 55, 21, 29, 12, 20]. For a matrix A, the max-norm of A is defined as

$$\|A\|_{\max} := \min_{U,V:A=UV^\top} \|U\|_{2,\infty}\|V\|_{2,\infty}.$$

We can generalize its definition to tensors, following [26], with the caveat that it then becomes a quasinorm since the triangle inequality is not satisfied.

Definition 4.3. *Define the* max-quasinorm *(or* max-qnorm*) of an order-t tensor $T \in \bigotimes_{i=1}^t \mathbb{R}^{n_i}$ as*

$$\|T\|_{\max} = \min_{T=U^{(1)}\circ\cdots\circ U^{(t)}} \prod_{i=1}^t \|U^{(i)}\|_{2,\infty}, \quad \textit{where} \quad \|U\|_{2,\infty} = \max_{\|x\|_2=1} \|Ux\|_\infty,$$

i.e., the maximum ℓ_2 norm of any row of U, and each of the $U^{(i)} \in \mathbb{R}^{n_i \times r}$ for some r.

The following lemma provides some basic properties of the max-quasinorm for tensors.

Lemma 4.4 ([26], Theorem 4). *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* $c \in \mathbb{R}$.
3. $\|T + S\|_{\max} \leq (\|T\|_{\max}^{2/t} + \|S\|_{\max}^{2/t})^{t/2} \leq 2^{t/2-1}(\|T\|_{\max} + \|S\|_{\max})$.

Note that property 3 in Lemma 4.4 implies that $\|\cdot\|_{\max}$ is a so-called p norm with $p = 2/t$ and also a quasinorm with constant $2^{t/2-1}$ [18]. Finally, in the matrix case it is a norm. As a matrix norm, many properties and equivalent definitions of the max-norm are known, and it can be computed via semidefinite programming [40, 38, 46]. In the tensor case, much less is known about the max-qnorm. We now prove generalizations of some of these properties that hold for tensors (proof in section SM1.5).

Theorem 4.5. *Let $T \in \bigotimes_{i=1}^t \mathbb{R}^{n_i}$ and $S \in \bigotimes_{i=1}^t \mathbb{R}^{m_i}$. The following max-qnorm properties hold:*

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 $T, S \in \bigotimes_{i=1}^t \mathbb{R}^{n_i}$.*
4. $\|T * T\|_{\max} \leq \|T\|_{\max}^2$.

For any matrix A, there is a surprising relationship between max-norm and rank:

$$|A|_\infty \leq \|A\|_{\max} \leq \sqrt{\operatorname{rank}(A)} \cdot |A|_\infty, \tag{4.2}$$

which does not depend on the size of A. We denote the entrywise infinity norm $|A|_\infty = \max_{i,j} |A_{ij}|$, and similarly $|T|_\infty = \max_{i_1,\ldots,i_t} |T_{i_1,\ldots,i_t}|$ for tensors. The proof of (4.2) is a result of John's theorem and is given in [40]. For the tensor generalization, we prove the following theorem.

Theorem 4.6. *Let* $T \in \bigotimes_{i=1}^{t} \mathbb{R}^n$ *with* $\operatorname{rank}(T) = r$. *Then we have that*

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

The proof of Theorem 4.6 is included in section SM1.9. This improves Theorem 7(ii) in [26] by a factor of $\sqrt{r}$. If we take $t = 2$, we get that $\|A\|_{\max} \le \sqrt{\operatorname{rank}(A)} \cdot |A|_\infty$, which is the same as (4.2). It remains an open question whether a better bound exists for all $t \ge 2$ in terms of the dependence on t. The numerical experiments of [26], which used a bisection method to estimate the max-qnorm of tensors of known rank, suggest that an improvement is possible. They study tensors formed from random factors, finding that increasing t by one leads to an approximately $\sqrt{r}$ increase. This suggests the conjecture that perhaps $\|T\|_{\max} \le \sqrt{r^{t-1}} \cdot |T|_\infty$ is the optimal bound. Further support for this scaling with r and t comes from incoherent tensors. Our definition is inspired by but slightly different from that used by Barak and Moitra [4] but reduces to the usual matrix incoherence condition.

Definition 4.7. *A rank* r *tensor* T *is said to be* C-incoherent *if there exists a rank-*r *factorization* $T = U^{(1)} \circ \cdots \circ U^{(t)}$ *such that* $|U^{(i)}|_\infty \le C$ *for* $i \in [t]$.

Proposition 4.8. *Let* T *be a* C*-incoherent tensor. Then* $\|T\|_{\max} \le C^t \sqrt{r^t}$, *and* $|T|_\infty \le C^t$.

The proof of Proposition 4.8 is included in section SM1.6. In any case, Theorem 4.6 is still useful for low-rank tensor completion, as it implies that an upper bound on the generalization error in terms of the max-qnorm can be translated into a bound that depends on the rank. That upper bound does not depend on n, which is crucial for attaining sample complexity linear in n.

We have found an improved lower bound on $\|T\|_{\max}$ via tensor matricization, sometimes called tensor unfolding or tensor flattening. The proof is in section SM1.7.

Definition 4.9 (Kolda and Bader [36]). *Let* $T \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$. *For* $1 \le i \le t$, *the mode-*i matricization *of* T *is a matrix denoted by* $T_{[i]} \in \mathbb{R}^{n_i} \times \mathbb{R}^{\prod_{j \ne i} n_j}$ *such that for any index* $(j_1, \ldots, j_t)$,

$$\left(T_{[i]}\right)_{j_i,k} = T_{j_1,\ldots,j_t},$$

with $k = 1 + \sum_{s=1, s\ne i}^{t} (j_s - 1) N_s$ *and* $N_s = \prod_{m=1, m\ne i}^{s-1} n_m$.

Proposition 4.10. *Let* $T \in \bigotimes_{i=1}^{t} \mathbb{R}^{n_i}$. *Then* $\|T\|_{\max} \ge \max_{1\le i\le t} \|T_{[i]}\|_{\max} \ge \max_{i_1,\ldots,i_t} |T_{i_1,\ldots,i_t}|$.

4.2. Sign tensors. In order to connect expansion properties of the hypergraph H to the error of our proposed tensor completion algorithm, 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

$$\operatorname{rank}_\pm(S) = \inf \left\{ r \;\middle|\; S = \sum_{i=1}^{r} s_i^{(1)} \circ \cdots \circ s_i^{(t)},\ s_i^{(j)} \in \{\pm 1\}^{n_i} \text{ for } i \in [r] \text{ and } j \in [t] \right\}. \tag{4.3}$$

Using rank-1 sign tensors as our atoms, we can construct a nuclear norm [26].

Definition 4.11. *We define the* sign nuclear norm *for a tensor* T *as*

$$\|T\|_\pm = \inf \left\{ \sum_{i=1}^{r} |\alpha_i| \;\middle|\; T = \sum_{i=1}^{r} \alpha_i S_i \text{ where } \alpha_i \in \mathbb{R},\ \operatorname{rank}_\pm(S_i) = 1 \right\}. \tag{4.4}$$

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 [26, 29]. The sign nuclear norm is called the "atomic M-norm" by [26] and the "atomic norm" by [29].

The next lemma which relates $\|\cdot\|_{\pm}$ and $\|\cdot\|_{\max}$ follows from a multilinear generalization of Grothendieck's inequality (the proof is given in section SM1.8). We use K_G to denote Grothendieck's constant over the reals. For detailed background, see [59].

Lemma 4.12. *The sign nuclear norm and max-qnorm satisfy* $\|T\|_{\pm} \leq K_G^{t-1}\|T\|_{\max}$, *where* K_G *is Grothendieck's constant over* $\mathbb{R}$.

5. Tensor completion.

5.1. Proof of Theorem 1.2.

Consider a rank-1 sign tensor $S = s_1 \circ \cdots \circ s_t$ with $s_j \in \{\pm 1\}^n$. Let J be the tensor of all ones and $S' = \frac{1}{2}(S + J)$, so that S' is shifted to be a tensor of zeros and ones. Then

$$S'_{i_1,\ldots,i_t} = \begin{cases} 1 & \text{if } (s_1)_{i_1} \cdots (s_t)_{i_t} = 1, \\ 0 & \text{if } (s_1)_{i_1} \cdots (s_t)_{i_t} = -1. \end{cases} \tag{5.1}$$

Define the sets

$$W_j := \{i \in [n] : (s_j)_i = -1\}. \tag{5.2}$$

Let $\mathcal{S}_t$ be the set of even t-strings in $\{0,1\}^t$. An even string has an even number of 1's in it; e.g., for $t = 3$ we have 000, 110, 101, 011 as even strings. The number of these strings is $|\mathcal{S}_t| = 2^{t-1}$, so we can enumerate all possible even t-strings from 1 to 2^{t-1}, denoted by $w_1, \ldots, w_{2^{t-1}} \in \{0,1\}^t$. Now for all $1 \leq i \leq t$ and $1 \leq j \leq 2^{t-1}$, we define the sets $W_{i,j}$ by

$$W_{i,j} = \begin{cases} W_i & \text{if } (w_j)_i = 1, \\ [n] \setminus W_i & \text{if } (w_j)_i = 0. \end{cases} \tag{5.3}$$

By considering the sign of entries in the components of S, we have the following decomposition for S' as a sum of rank-1 tensors (a derivation of (5.4) is provided in section SM1.10 of the supplement):

$$S' = \sum_{j=1}^{2^{t-1}} 1_{W_{1,j}} \circ \cdots \circ 1_{W_{t,j}}. \tag{5.4}$$

Now we consider the deviation in the sample mean from the mean over all entries in S:

$$\begin{aligned}\left|\frac{1}{n^t}\sum_{e\in[n]^t} S_e - \frac{1}{|E|}\sum_{e\in E} S_e\right| &= \left|\frac{1}{n^t}\sum_{e\in[n]^t}(2S'_e-1) - \frac{1}{|E|}\sum_{e\in E}(2S'_e-1)\right| \\ &= 2\left|\frac{1}{n^t}\sum_{e\in[n]^t} S'_e - \frac{1}{|E|}\sum_{e\in E} S'_e\right| \\ &= 2\left|\frac{\sum_{j=1}^{2^{t-1}}\prod_{i=1}^{t}|W_{i,j}|}{n^t} - \frac{\sum_{j=1}^{2^{t-1}} e(W_{1,j},\dots,W_{t,j})}{|E|}\right| \\ &\le 2\sum_{j=1}^{2^{t-1}}\left|\frac{|W_{1,j}|\cdots|W_{t,j}|}{n^t} - \frac{e(W_{1,j},\dots,W_{t,j})}{|E|}\right|.\end{aligned}$$

Applying (1.3) in Lemma 1.1 to the sets $W_{1,j}\subseteq V_1,\dots,W_{t,j}\subseteq V_t$ for each $1\le j\le 2^{t-1}$, we get that

$$\begin{aligned}\left|\frac{1}{n^t}\sum_{e\in[n]^t} S_e - \frac{1}{|E|}\sum_{e\in E} S_e\right| &\le 2\sum_{j=1}^{2^{t-1}}\frac{\lambda_2(H)}{|E|}\sqrt{|W_{1,j}|\cdots|W_{t,j}|} \\ &= 2\sum_{j=1}^{2^{t-1}}\frac{\lambda_2(H)n^{t/2}}{|E|}\sqrt{\frac{|W_{1,j}|\cdots|W_{t,j}|}{n^t}} \le \frac{2^t n^{t/2}\lambda_2(H)}{|E|}. \end{aligned} \tag{5.5}$$

We now write the tensor $T=\sum_i \alpha_i S_i$ as a sum of rank-1 sign tensors S_i, with coefficients $\alpha_i\in\mathbb{R}$. Let $\|\cdot\|_\pm$ be the tensor sign nuclear norm, i.e., $\|T\|_\pm=\sum_i|\alpha_i|$. Then for a general tensor T, we can apply (5.5) to each S_i, and by the triangle inequality we get

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

This 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,

$$\begin{aligned}&\frac{1}{n^t}\|\hat{T}-T\|_F^2 = \left|\frac{1}{n^t}\sum_{e\in[n]^t} R_e\right| \\ &\le \frac{2^t n^{t/2}\lambda_2(H)}{|E|}\|R\|_\pm \le \frac{2^t n^{t/2}\lambda_2(H)}{|E|}K_G^{t-1}\,\|R\|_{\max} && \text{(Lemma 4.12)} \\ &\le \frac{2^t n^{t/2}\lambda_2(H)}{|E|}K_G^{t-1}\,\|\hat{T}-T\|_{\max}^2 && \text{(Theorem 4.5, part 4)} \\ &\le \frac{2^{2t-2} n^{t/2}\lambda_2(H)}{|E|}K_G^{t-1}\left(\|\hat{T}\|_{\max}+\|T\|_{\max}\right)^2. && \text{(Lemma 4.4)} \end{aligned} \tag{5.6}$$

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

5.2. Proof of Theorem 1.3. Consider a hypercubic tensor T of order t, i.e., $T \in \mathbb{R}^{n\times\cdots\times n}$. We sample the entry T_e whenever $e = (i_1, \dots, i_t)$ is a hyperedge in H defined in section 2.1. Then the sample size is $|E| = nd^{t-1}$. Consider a rank-1 sign tensor $S = s_1 \circ \cdots \circ s_t$ with $s_j \in \{\pm 1\}^n$. Following the same steps in the proof of Theorem 1.2, we obtain

$$\left| \frac{1}{n^t} \sum_{e\in[n]^t} S_e - \frac{1}{|E|} \sum_{e\in E} S_e \right| \le 2 \sum_{j=1}^{2^{t-1}} \left| \frac{|W_{1,j}| \cdots |W_{t,j}|}{n^t} - \frac{e(W_{1,j}, \dots, W_{t,j})}{nd^{t-1}} \right|.$$

Applying Theorem 3.1 to the sets $W_{1,j} \subseteq V_1, \dots, W_{t,j} \subseteq V_t$ for each $1 \le j \le 2^{t-1}$, we get that

$$\left| \frac{1}{n^t} \sum_{e\in[n]^t} S_e - \frac{1}{nd^{t-1}} \sum_{e\in E} S_e \right| \le 2 \sum_{j=1}^{2^{t-1}} \frac{(2t-3)\lambda}{4d} = 2^{t-2}(2t-3)\frac{\lambda}{d}. \tag{5.7}$$

We now write the tensor $T = \sum_i \alpha_i S_i$ as a sum of rank-1 sign tensors S_i, with coefficients $\alpha_i \in \mathbb{R}$. Define $R = (\hat{T} - T) * (\hat{T} - T)$. Following the same steps in the proof of Theorem 1.2, we have

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

Since $\|\hat{T}\|_{\max} \le \|T\|_{\max}$, this leads to the final result with a constant $C_t = 2^t(2t-3)K_G^{t-1}$.

5.3. Tensor completion with erroneous observations. Now we turn to the case when our observations Z of the original tensor T are corrupted by errors ν. We will call this noise, but it can be anything, even chosen adversarially, so long as it is bounded. Let $Z \in \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ be the tensor we observe with $Z_e = 0$ if $e \notin E$ and $Z_e = T_e + \nu_e$ for $e \in E$. In this case, we study the solution to the following optimization problem:

$$\begin{aligned} \min_X \quad & \|X\|_{\max}, \\ \text{subject to} \quad & \frac{1}{|E|} \sum_{e\in E} (X_e - Z_e)^2 \le \delta^2, \end{aligned} \tag{5.8}$$

for some $\delta > 0$. The parameter δ is a bound on the root mean squared error of the observations. In a probabilistic setting, we may pick this parameter so that the constraint holds with sufficiently high probability. We obtain the following corollary of Theorem 1.3, with the proof in section SM1.11 of the supplement.

Corollary 5.1. *Let E be the hyperedge set of H defined in section 2. Suppose we observe $Z_e = T_e + \nu_e$ for all $e \in E$ with bounded mean square error satisfying*

$$\frac{1}{|E|} \sum_{e\in E} \nu_e^2 \le \delta^2. \tag{5.9}$$

Then solving the optimization problem (5.8) *will give us a solution* $\hat{T}$ *that satisfies*

$$\frac{1}{n^t}\|\hat{T} - T\|_F^2 \leq 2^t(2t-3)K_G^{t-1}\|T\|_{\max}^2\frac{\lambda}{d} + 4\delta^2. \tag{5.10}$$

Remark 5.2. Inequality (5.10) has an error with an irreducible term $O(\delta^2)$, in contrast to results such as [26], where the error goes to zero even in the presence of noise. In our case, this term cannot be overcome because the errors do not have any of the nice properties of noise. For example, all observations Z could be shifted by an amount δ, which would bias the estimation of T to $T + \delta J$.

6. Practical algorithm for finding minimum complexity tensors.

In such tensor factorization problems, one usually picks a rank r and alternately minimizes the objective function over the factors $U^{(i)}$, making the overall approach coordinate descent. Note that the optimization over a single factor is in fact a convex problem for which there exist polynomial time algorithms. However, care must be taken in designing these convex subproblems for efficiency. We will relax the problem so as to make it amenable to proximal gradient descent methods. This will require us to compute the proximal operator of $\|\cdot\|_{2,\infty}$ and the gradient of the smooth part of the modified objective.

6.1. Relaxed algorithm.

We propose the following more practical relaxation of (5.8). Rather than deal with the constrained problem, we instead optimize

$$\begin{aligned} \min_{X,R} \quad & \|X\|_{\max} + \frac{\kappa}{2}\|M * (X - Z - R)\|_F^2 + \frac{\beta}{2}\|M * R\|_F^2 \\ \text{subject to} \quad & \|M * R\|_F \leq \delta. \end{aligned} \tag{6.1}$$

The mask tensor M has $M_e = 1$ if $e \in E$ and $M_e = 0$ otherwise. This makes M equivalent to the adjacency operator of the observation hypergraph H, so the constraint is the same as in the original noisy problem (5.8). However, we have relaxed the problem by introducing the auxiliary variable R. Here we absorb $\sqrt{|E|}$ into the noise magnitude δ to avoid clutter. The relaxation parameter κ is taken to be large, which keeps $R \approx X - Z$ within the observation mask; outside the mask, we may assume $R = 0$. The parameter β is a small smoothing of the hard constraint taken for technical reasons laid out in section SM2 of the supplement. For all experiments, we take $\kappa = 100$ and $\beta = 1$.

The abilities to compute gradients and evaluate the prox of the $\ell_{2,\infty}$ norm are all the ingredients needed to implement coordinate descent for (6.1). The details of how these are computed and a variable projection step to minimize out R are provided in section SM2 of the supplement. At each coordinate descent step, we apply the accelerated proximal gradient method to the cost. The cost in terms of the coordinate blocks $\{U^{(i)}\}_{i=1}^t$ is written as

$$C = \frac{1}{2}\left(\kappa(1-\mu)^2 + \mu^2\beta\right)\|M * (U^{(1)} \circ \cdots \circ U^{(t)} - Z)\|_F^2 + \prod_{i=1}^t \|U^{(i)}\|_{2,\infty}, \tag{6.2}$$

with $\mu = \mu(U^{(1)} \circ \cdots \circ U^{(t)})$ given by (SM2.1). After the first iteration of coordinate descent, we find that it is important to rescale the factors to have equal norms across all columns.

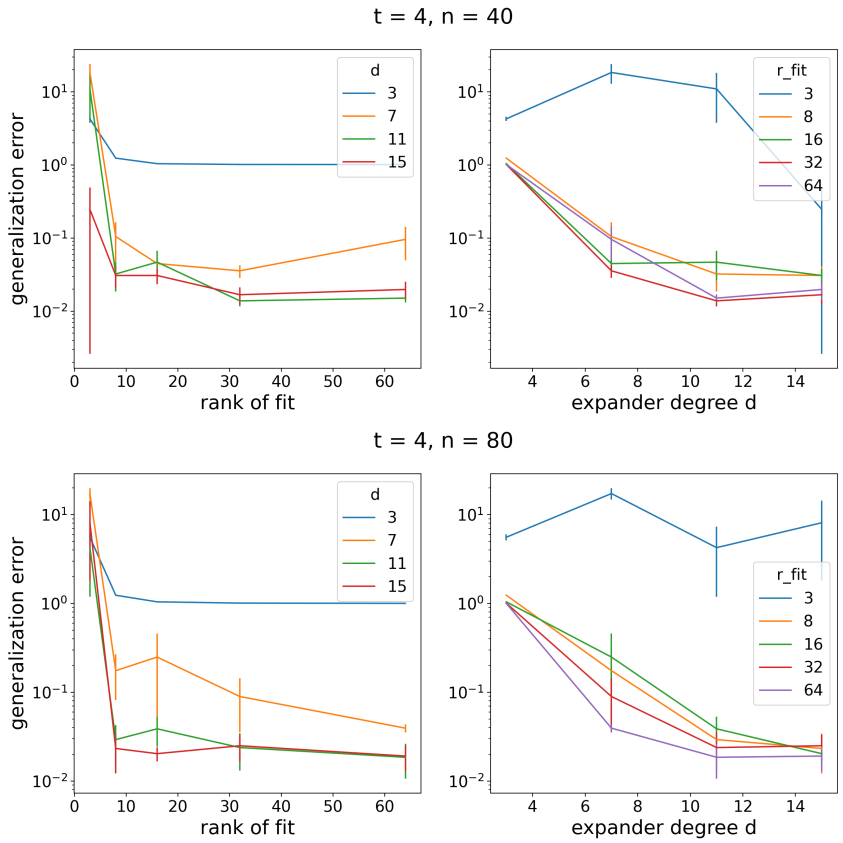

Figure 6.1. *Relative error of reconstruction for $t = 4$ and $n = 40, 80$. The results are the average of 6 tensors with standard errors shown.*

This step does not affect the goodness of fit, but it can cause the max-qnorm penalty term to increase. However, we have found that it makes the algorithm more stable. Not shown in (6.2), we also include a small squared Frobenius penalty $0.01 \cdot \|U\|_F^2$ on each factor for numerical stability. Python code to implement this method and reproduce our experiments is available from https://github.com/kharris/max-qnorm-tensor-completion.

6.2. Numerical experiments. We generate random rank $r = 3$, order t tensors by drawing the entries of $U^{(i)}$ independently from the uniform distribution Unif$[-1, 1]$ for $i = 1, \ldots, t$. The resulting tensor T is rescaled to have $\|T\|_F = \sqrt{nd^t}$, so that the root-mean-square of its entries is 1. Entries are sampled using a random d-regular expander graph as described in section 2.1, and no noise was added. In principle, we could have used one of any number of deterministic d-regular expander constructions, but these are more difficult to implement for arbitrary d. Simulations were run for parameter ranges $n \in \{20, 40, 80\}, t \in \{3, 4\}, d \in \{3, 7, 11, 15\}$. We fit with tensors of rank $r_{\text{fit}} \in \{3, 8, 16, 32, 64\}$ and residual parameter $\delta = 0.05\sqrt{|E|}$, with $\kappa = 100$, $\beta = 1$. We report generalization error, defined as

$$\text{generalization error} := \frac{\|\hat{T} - T\|_F}{\|T\|_F}.$$

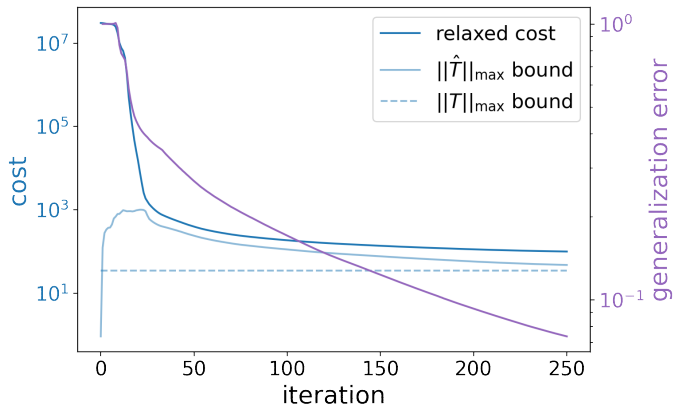

Figure 6.2. *The behavior of the optimization algorithm by iteration for a test case with* $t = 4, n = 400, d = 12, r_{\text{fit}} = 16$. *Initially, the algorithm focuses on fitting the data, causing the max-qnorm to increase until the residuals are comparable to* δ, *at which point the cost is mostly due to the max-qnorm. In the final stage, the max-qnorm is decreased, leading to a decrease in the generalization error. Note that the max-qnorm lines are upper bounds, and the algorithm has not fully converged in* 250 *iterations.*

The results for and $t = 4$ are shown in Figure 6.1, with a similar Figure SM1 for $t = 3$ in the supplement. In either dimension, errors $< 10\%$ are achieved whenever d and r_{fit} are sufficiently high. The fraction of observed entries in this setting is $(d/n)^{t-1}$. For sufficiently large r_{fit}, we see error $< 10\%$ once $d \geq 11$, or observing 2% $(n = 40, t = 3)$, 0.6% $(n = 40, t = 4)$, 0.3% $(n = 80, t = 3)$, and 0.04% $(n = 80, t = 4)$ of the total entries.

It is interesting to note that having $r_{\text{fit}} \gg r$ helps the optimization algorithm find a better solution. Furthermore, having a highly overparameterized model does not appreciably hurt its generalization ability. Once the rank of the fit is above a threshold around 10, the error stays essentially flat. Overparameterizing the rank appears to help the optimization find a good solution, while the max-qnorm penalty controls the complexity of this solution to prevent overfitting.

When we compare $\prod_{i=1}^{t} \|U^{(i)}\|_{2,\infty}$ and $\prod_{i=1}^{t} \|\hat{U}^{(i)}\|_{2,\infty}$, which are upper bounds on the max-qnorm of the truth T and estimate $\hat{T}$, we find that the estimate generalizes well when its upper bound is less than or equal to that of the truth. This is evident in Figure 6.2, where we see how the cost and generalization error change with iteration for $t = 4, n = 400, d = 12, r_{\text{fit}} = 16$, and otherwise the same parameters as before. This corresponds to a fraction 2.7×10^{-5} of the total entries observed. The overall relaxed cost (6.2) is shown along with the max-qnorm upper bounds. Initially, the cost is dominated by the residual term. The residuals decrease while the bound on $\|\hat{T}\|_{\max}$ actually increases. Once the mean-square residuals are as small as δ^2, the cost is dominated by the max-qnorm complexity term, and the $\|\hat{T}\|_{\max}$ bound decreases. These stages can be understood by the variable projection parameter varying between $\mu \approx 0$ and $\mu \approx 1$; see section SM2.1. Throughout this process the generalization error decreases, although it has not fully converged in 250 iterations.

Finally, we also have performed an experiment with $n = 1000, t = 4, d = 12, r_{\text{fit}} = 16$, and otherwise the same parameters as before. After running for 271 iterations, we stopped the

optimization and found a generalization error of 3.4%. This is a very small fraction 1.7×10^{-6} of the total entries in T. This and the $n = 400$ experiment show that good recovery is possible even when $d \ll n$.

7. Discussion.

We have deterministically analyzed tensor completion using the max-qnorm as a measure of complexity and hypergraph sampling. Our main results show that, by finding the tensor with smallest max-qnorm that is consistent with the observations, one may obtain a good estimate of the true tensor. The error of this estimate depends on the expansion properties of the hypergraph model. Auxiliary to this main result are a number of newly proven facts about the max-qnorm which may be of interest to specialists in communication complexity.

We show that proximal algorithms based on a relaxation of constrained max-qnorm minimization are practical to implement and provide code to reproduce our results. Although our numerical study is mainly proof-of-concept, the method does work in the problem sizes that were tested. These were mainly small-scale ($n < 100$), but promising results were found in medium-scale tests ($n = 400, 1000$) with very small fractions of the total entries observed. Sampling based on d-regular expander graphs, as opposed to sampling tensor entries uniformly at random, was successful in these experiments.

A number of theoretical and practical considerations still remain. Theoretically, it would be nice to have other constructions of sparse hypergraph expanders. Completely deterministic constructions could be useful for applications, e.g., for compressive sensing in hardware where a single sensor reuses its "observation mask" over and over.

Finding the right complexity measures for tensors is still an open problem. Here we study max-qnorm because it is amenable to deterministic analysis; however, there are likely other measures that work better in some settings. For example, if the factors are known to be smooth, e.g., due to spatial autocorrelation in images, then other regularization will be beneficial [51].

The optimization problem is nonconvex, but our simulations show that good solutions can be found when the fit uses an overparameterized rank. It may be possible to explain this using the techniques of [27, 42].

A thorough numerical comparison of the performance of max-qnorm minimization versus simple Frobenius norm minimization would be informative but outside the scope of this paper. Simulations of larger size will be needed to see whether the numerical results support the $O(n)$ sample complexity that the theory predicts. Furthermore, our experiments were with incoherent, random tensors, which could make the problem easier. It would also be interesting to try expander sampling with real datasets. It will be important to reconcile various conjectured hardness results [4] with the practical success of these algorithms in many settings.

Acknowledgments. We thank Ioana Dumitriu for support and discussions. We are grateful to Paul Beame for discussions of communication complexity and to Adi Shraibman for sharing details of the proof of the Hadamard product bound on the max-norm for matrices.

REFERENCES

[1] M. Ajtai, J. Komlós, and E. Szemerédi, *Deterministic simulation in logspace*, in Proceedings of the 19th Annual ACM Symposium on Theory of Computing, ACM, New York, 1987, pp. 132–140.

[2] N. Alon, *Eigenvalues and expanders*, Combinatorica, 6 (1986), pp. 83–96.

[3] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman, *Derandomized graph products*, Comput. Complexity, 5 (1995), pp. 60–75.

[4] B. Barak and A. Moitra, *Noisy tensor completion via the sum-of-squares hierarchy*, in Proceedings of the Conference on Learning Theory, New York, NY, 2016, pp. 417–445.

[5] A. Ben-Aroya and A. Ta-Shma, *A combinatorial construction of almost-Ramanujan graphs using the zig-zag product*, SIAM J. Comput., 40 (2011), pp. 267–290, https://doi.org/10.1137/080732651.

[6] S. Bhojanapalli and P. Jain, *Universal matrix completion*, in Proceedings of the International Conference on Machine Learning, Beijing, China, 2014, pp. 1881–1889.

[7] Y. Bilu and S. Hoory, *On codes from hypergraphs*, European J. Combin., 25 (2004), pp. 339–354.

[8] Y. Bilu and N. Linial, *Lifts, discrepancy and nearly optimal spectral gap*, Combinatorica, 26 (2006), pp. 495–519.

[9] G. Brito, I. Dumitriu, and K. D. Harris, *Spectral Gap in Random Bipartite Biregular Graphs and Its Applications*, preprint, https://arxiv.org/abs/1804.07808v1, 2018.

[10] S. P. Burnwal and M. Vidyasagar, *Deterministic completion of rectangular matrices using asymmetric Ramanujan graphs: Exact and stable recovery*, IEEE Trans. Signal Process., 68 (2020), pp. 3834–3848.

[11] C. Cai, G. Li, H. V. Poor, and Y. Chen, *Nonconvex low-rank symmetric tensor completion from noisy data*, in Advances in Neural Information Processing Systems 2019, NeurIPS, San Diego, CA, 2019.

[12] T. Cai and W.-X. Zhou, *Matrix completion via max-norm constrained optimization*, Electron. J. Stat., 10 (2016), pp. 1493–1525.

[13] S. Chatterjee, *A deterministic theory of low rank matrix completion*, IEEE Trans. Inform. Theory, 66 (2020), pp. 8046–8055.

[14] F. Chung, *Spectral Graph Theory*, CBMS Regional Conf. Ser. in Math 92, AMS, Providence, RI, 1997.

[15] E. Cohen, D. Mubayi, P. Ralli, and P. Tetali, *Inverse expander mixing for hypergraphs*, Electron. J. Combin., 23 (2016), pp. 2–20.

[16] G. Cohen, N. Peri, and A. Ta-Shma, *Expander random walks: A Fourier-analytic approach*, Electron. Colloquium Comput. Complex., 27 (2020), 6.

[17] S. De Winter, J. Schillewaert, and J. Verstraete, *Large incidence-free sets in geometries*, Electron. J. Combin., 19 (2012), 24.

[18] S. J. Dilworth, *The dimension of Euclidean subspaces of quasinormed spaces*, Math. Proc. Cambridge Philos. Soc., 97 (1985), pp. 311–320.

[19] I. Dumitriu and Y. Zhu, *Spectra of Random Regular Hypergraphs*, preprint, https://arxiv.org/abs/1905.06487, 2019.

[20] S. Foucart, D. Needell, R. Pathak, Y. Plan, and M. Wootters, *Weighted matrix completion from non-random, non-uniform sampling patterns*, IEEE Trans. Inform. Theory, 67 (2021), pp. 1264–1290.

[21] R. Foygel and N. Srebro, *Concentration based guarantees for low-rank matrix reconstruction*, in Proceedings of the 24th Annual Conference on Learning Theory, Budapest, Hungary, 2011, pp. 315–340.

[22] S. Friedland and L.-H. Lim, *Nuclear norm of higher-order tensors*, Math. Comp., 87 (2018), pp. 1255–1281.

[23] J. Friedman and A. Wigderson, *On the second eigenvalue of hypergraphs*, Combinatorica, 15 (1995), pp. 43–65.

[24] D. Gamarnik, Q. Li, and H. Zhang, *Matrix completion from $O(n)$ samples in linear time*, in Proceedings of the 30th Conference on Learning Theory, Amsterdam, The Netherlands, 2017, pp. 940–947.

[25] S. Gandy, B. Recht, and I. Yamada, *Tensor completion and low-n-rank tensor recovery via convex optimization*, Inverse Problems, 27 (2011), 025010.

[26] N. Ghadermarzy, Y. Plan, and Ö. Yilmaz, *Near-optimal sample complexity for convex tensor completion*, Inf. Inference, 8 (2019), pp. 577–619.

[27] B. D. Haeffele and R. Vidal, *Global optimality in neural network training*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Washington, DC, 2017, pp. 7331–7339.

[28] W. H. Haemers, *Interlacing eigenvalues and graphs*, Linear Algebra Appl., 226 (1995), pp. 593–616.
[29] E. Heiman, G. Schechtman, and A. Shraibman, *Deterministic algorithms for matrix completion*, Random Structures Algorithms, 45 (2014), pp. 306–317.
[30] C. J. Hillar and L.-H. Lim, *Most tensor problems are NP-hard*, J. ACM, 60 (2013), 45.
[31] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc., 43 (2006), pp. 439–561.
[32] P. Jain and S. Oh, *Provable tensor factorization with missing data*, in Advances in Neural Information Processing Systems 2014, NeurIPS, San Diego, CA, 2014, pp. 1431–1439.
[33] N. Kahale, *Eigenvalues and expansion of regular graphs*, J. ACM, 42 (1995), pp. 1091–1106.
[34] R. H. Keshavan, A. Montanari, and S. Oh, *Matrix completion from a few entries*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2980–2998.
[35] O. Klopp, *Noisy low-rank matrix completion with general sampling distribution*, Bernoulli, 20 (2014), pp. 282–303.
[36] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500, https://doi.org/10.1137/07070111X.
[37] A. Krishnamurthy and A. Singh, *Low-rank matrix and tensor completion via adaptive sampling*, in Advances in Neural Information Processing Systems 2013, NeurIPS, San Diego, CA, 2013, pp. 836–844.
[38] T. Lee, A. Shraibman, and R. Špalek, *A direct product theorem for discrepancy*, in Proceedings of the 23rd Annual IEEE Conference on Computational Complexity, IEEE, Washington, DC, 2008, pp. 71–80.
[39] J. Lenz and D. Mubayi, *Eigenvalues and linear quasirandom hypergraphs*, Forum Math. Sigma, 3 (2015), e2.
[40] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman, *Complexity measures of sign matrices*, Combinatorica, 27 (2007), pp. 439–463.
[41] A. Liu and A. Moitra, *Tensor completion made practical*, in Advances in Neural Information Processing Systems 2020, NeurIPS, San Diego, CA, 2020.
[42] C. Liu, L. Zhu, and M. Belkin, *Loss Landscapes and Optimization in Over-parameterized Non-linear Systems and Neural Networks*, preprint, https://arxiv.org/abs/2003.00307, 2020.
[43] J. Liu, P. Musialski, P. Wonka, and J. Ye, *Tensor completion for estimating missing values in visual data*, IEEE Trans. Pattern Anal. Mach. Intell., 35 (2012), pp. 208–220.
[44] A. Lubotzky, *High dimensional expanders*, in Proceedings of the International Congress of Mathematicians, ICM 2018, World Scientific, Hackensack, NJ, 2018, pp. 705–730.
[45] A. W. Marcus, D. A. Spielman, and N. Srivastava, *Interlacing families* I: *Bipartite Ramanujan graphs of all degrees*, Ann. of Math. (2), 182 (2015), pp. 307–325.
[46] J. Matoušek, A. Nikolov, and K. Talwar, *Factorization norms and hereditary discrepancy*, Int. Math. Res. Not. IMRN, 2020 (2020), pp. 751–780.
[47] S. Mohanty, R. O'Donnell, and P. Paredes, *Explicit near-Ramanujan graphs of every degree*, in Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, 2020, pp. 510–523.
[48] A. Montanari and N. Sun, *Spectral algorithms for tensor completion*, Comm. Pure Appl. Math., 71 (2018), pp. 2381–2425.
[49] C. Mu, B. Huang, J. Wright, and D. Goldfarb, *Square deal: Lower bounds and improved relaxations for tensor recovery*, in Proceedings of the 31st International Conference on Machine Learning, Beijing, China, 2014, pp. 73–81.
[50] C. Musco, C. Musco, and D. P. Woodruff, *Simple heuristics yield provable algorithms for masked low-rank approximation*, in 12th Innovations in Theoretical Computer Science Conference (ITCS 2021), LIPIcs. Leibniz Int. Proc. Inform. 185, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, Germany, 2021, 6.
[51] J. Y. Park, K. Carr, S. Zheng, Y. Yue, and R. Yu, *Multiresolution tensor learning for efficient and interpretable spatial analysis*, in Proceedings of the International Conference on Machine Learning, PMLR, 2020, pp. 7499–7509.
[52] O. Parzanchevski, *Mixing in high-dimensional expanders*, Combin. Probab. Comput., 26 (2017), pp. 746–761.

[53] A. Potechin and D. Steurer, *Exact tensor completion with sum-of-squares*, in Proceedings of the 2017 Conference on Learning Theory, volume 65, PMLR, 2017, pp. 1619–1673.
[54] B. Recht, *A simpler approach to matrix completion*, J. Mach. Learn. Res., 12 (2011), pp. 3413–3430.
[55] J. D. M. Rennie and N. Srebro, *Fast maximum margin matrix factorization for collaborative prediction*, in Proceedings of the 22nd International Conference on Machine learning, ACM, New York, 2005, pp. 713–719.
[56] D. Shah and C. L. Yu, *Iterative collaborative filtering for sparse noisy tensor estimation*, in Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), IEEE, Washington, DC, 2019, pp. 41–45.
[57] Q. Song, H. Ge, J. Caverlee, and X. Hu, *Tensor completion algorithms in big data analytics*, ACM Trans. Knowledge Discovery from Data (TKDD), 13 (2019), 6.
[58] N. Srebro and A. Shraibman, *Rank, trace-norm and max-norm*, in Proceedings of the International Conference on Computational Learning Theory, Springer, New York, 2005, pp. 545–560.
[59] N. Tomczak-Jaegermann, *Banach-Mazur Distances and Finite-Dimensional Operator Ideals*, Longman Scientific & Technical, Harlow, UK, 1989.
[60] D. Xia and M. Yuan, *On polynomial time methods for exact low-rank tensor completion*, Found. Comput. Math., 19 (2019), pp. 1265–1313.
[61] C. L. Yu, *Tensor Estimation with Nearly Linear Samples*, preprint, https://arxiv.org/abs/2007.00736, 2020.
[62] M. Yuan and C.-H. Zhang, *On tensor completion via nuclear norm minimization*, Found. Comput. Math., 16 (2016), pp. 1031–1068.
[63] A. Zhang, *Cross: Efficient low-rank tensor completion*, Ann. Statist., 47 (2019), pp. 936–964.
[64] Z. Zhou and Y. Zhu, *Sparse random tensors: Concentration, regularization and applications*, Electron. J. Stat., 15 (2021), pp. 2483–2516.