Exact tensor completion with sum-of-squares

Aaron Potechin[*]   David Steurer[†]

February 20, 2017

Abstract

We obtain the first polynomial-time algorithm for exact tensor completion that improves over the bound implied by reduction to matrix completion. The algorithm recovers an unknown 3-tensor with r incoherent, orthogonal components in $\mathbb{R}^n$ from $r \cdot \tilde{O}(n^{1.5})$ randomly observed entries of the tensor. This bound improves over the previous best one of $r \cdot \tilde{O}(n^2)$ by reduction to exact matrix completion. Our bound also matches the best known results for the easier problem of approximate tensor completion (Barak & Moitra, 2015).

Our algorithm and analysis extends seminal results for exact matrix completion (Candes & Recht, 2009) to the tensor setting via the sum-of-squares method. The main technical challenge is to show that a small number of randomly chosen monomials are enough to construct a degree-3 polynomial with a precisely planted orthogonal global optima over the sphere and that this fact can be certified within the sum-of-squares proof system.

Keywords: tensor completion, sum-of-squares method, semidefinite programming, exact recovery, matrix polynomials, matrix norm bounds

Contents

1 Introduction

A basic task in machine learning and signal processing is to infer missing data from a small number of observations about the data. An important example is *matrix completiton* which asks to recover an unknown low-rank matrix from a small number of observed entries. This problem has many interesting applications—one of the prominent original motivations was the Netflix Prize that sought improved algorithms for predicting user ratings for movies from a small number of user-provided ratings. After an extensive research effort [CR09, CT10, KMO09, SS05], efficient algorithms with almost optimal, provable recovery guarantees have been obtained: In order to efficiently recover an unknown incoherent n-by-n matrix of rank r it is enough to observe $r \cdot \tilde{O}(n)$ random entries of the matrix [Gro11, Rec11]. One of the remaining challenges is to obtain algorithm for the more general and much less understood *tensor completion problem* where the observations do not just consist of pairwise correlations but also higher-order ones.

Algorithms and analyses for matrix and tensor completion come in three flavors:

1. algorithms analyzed by statistical learning tools like Rademacher complexity [SS05, BM16].
2. iterative algorithms like alternating minimization [JNS13, Har14, HW14].
3. algorithms analyzed by constructing dual certificates for convex programming relaxations [CR09, Gro11, Rec11].

While each of these flavors have different benefits, typically only algorithms of the third flavor achieve exact recovery. (The only exception of this rule we are aware of is a recent fast algorithm for matrix completion [JN15].) For all other algorithms, the analysis exhibits a trade-off between reconstruction error and the required number of observations (even when there is no noise in the input).[1]

In this work, we obtain the first algorithm for exact tensor completion that improves over the bounds implied by reduction to exact matrix completion. The algorithm recovers an unknown 3-tensor with r incoherent, orthogonal components in $\mathbb{R}^n$ from $r \cdot \tilde{O}(n^{1.5})$ randomly observed entries of the tensor. The previous best bound for exact recovery is $r \cdot \tilde{O}(n^2)$, which is implied by reduction to exact matrix completion. (The reduction views 3-tensor on $\mathbb{R}^n$ as an n-by-n^2 matrix. We can recover rank-r matrices of this shape from $r \cdot \tilde{O}(n^2)$ samples, which is best possible.) Our bound also matches the best known results for the easier problem of approximate tensor completion [JO14, BS15, BM16] (the results of the last work also applies to a wider range of tensors and does not require orthogonality).

A problem similar to matrix and tensor completion is matrix and tensor sensing. The goal is to recover an unknown low rank matrix or tensor from a small number of linear measurements. An interesting phenomenon is that for carefully designed measurements

---

[1]We remark that this trade-off is a property of the analysis and not necessarily the algorithm. For example, some algorithms of the first flavor are based on the same convex programming relaxations as exact recovery algorithms. Also for iterative algorithm, the trade-off between reconstruction error and number of sample comes from the requirement of the analysis that each iteration uses fresh samples. For these iterative algorithms, the number of samples depends only logarithmically on the desired accuracy, which means that these analyses imply exact recovery if the bit complexity of the entries is small.

(which actually happen to be rank 1) it is possible to efficiently recover a 3-tensor of rank r with just $O(r \cdot n)$ measurements [FS12].

Sum-of-squares method.   Our algorithm is based on *sum-of-squares* [Sho87, Par00, Las01], a very general and powerful meta-algorithm studied extensively in many scientific communities (see for example the survey [BS14]). In theoretical computer science, the main research focus has been on the capabilities of sum-of-squares for approximation problems [BBH+12], especially in the context of Khot's Unique Games Conjecture [Kho02]. More recently, sum-of-squares emerged as a general approach to inference problems that arise in machine learning and have defied other algorithmic techniques. This approach has lead to improved algorithms for tensor decomposition [BKS15, GM15, HSSS16, MSS16], dictionary learning [BKS15, HM16], tensor principal component analysis [HSS15, RRS16, BGL16], planted sparse vectors [BKS14, HSSS16]. An exciting direction is also to understand limitations of sum-of-squares for inference problems on concrete input distributions [MW15, HSS15, BHK+16].

An appealing feature of the sum-of-squares method is that its capabilities and limitations can be understood through the lens of a simple but surprisingly powerful and intuitive restricted proof system called sum-of-squares or Positivstellensatz system [GV01, Gri01a, Gri01b]. A conceptual contribution of this work is to show that seminal results for inference problem like compressed sensing and matrix completion have natural interpretations as *identifiability proofs* in this system. Furthermore, we show that this interpretation is helpful in order to analyze more challenging inference problems like tensor completion. A promising future direction is to find more examples of inference problems where this lens on inference algorithms and identifiability proofs yields stronger provable guarantees.

A technical contribution of our work is that we develop techniques in order to show that sum-of-squares achieves exact recovery. Most previous works only showed that sum-of-squares gives approximate solutions, which in some cases can be turned to exact solutions by invoking algorithms with local convergence guarantees [GM15, BKS14] or solving successive sum-of-squares relaxations [MSS16].

1.1   Results

We say that a vector $v \in \mathbb{R}^n$ is μ-incoherent with respect to the coordinate basis $e_1, \ldots, e_n$ if for every index $i \in [n]$,

$$\langle e_i, v \rangle^2 \leqslant \tfrac{\mu}{n} \cdot \|v\|^2 . \tag{1.1}$$

We say that a 3-tensor $X \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ is orthogonal of rank r if there are orthogonal vectors $\{u_i\}_{i \in [r]} \subseteq \mathbb{R}^n$, $\{v_i\}_{i \in [r]} \subseteq \mathbb{R}^n$, $\{w_i\}_{i \in [r]} \subseteq \mathbb{R}^n$ such that $X = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$. We say that such a 3-tensor X is μ-incoherent if all of the vectors u_i, v_i, w_i are μ-incoherent.

Theorem 1.1 (main). *There exists a polynomial-time algorithm that given at least $r \cdot \mu^{O(1)} \cdot \tilde{O}(n)^{1.5}$ random entries of an unknown orthogonal μ-incoherent 3-tensor $X \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ of rank r, outputs all entries of X with probability at least $1 - n^{-\omega(1)}$.*

We note that the analysis also shows that the algorithm is robust to inverse polynomial amount of noise in the input (resulting in inverse polynomial amount of error in the output).

We remark that the running time of the algorithm depends polynomially on the bit complexity on X.

2 Techniques

Let $\{u_i\}_{i\in[r]}, \{v_i\}_{i\in[r]}, \{w_i\}_{i\in[r]}$ be three orthonormal sets in $\mathbb{R}^n$. Consider a 3-tensor $X \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ of the form $X = \sum_{i=1}^r \lambda_i \cdot u_i \otimes v_i \otimes w_i$ with $\lambda_1, \ldots, \lambda_n \geqslant 0$. Let $\Omega \subseteq [n]^3$ be a subset of the entries of X.

Our goal is to efficiently reconstruct the unknown tensor X from its restriction X_Ω to the entries in Ω. Ignoring computational efficiency, we first ask if this task is information-theoretically possible. More concretely, for a given set of observations X_Ω, how can we rule out that there exists another rank-r orthogonal 3-tensor $X' \neq X$ that would give rise to the same observations $X'_\Omega = X_\Omega$?[2]

A priori it is not clear how an answer to this information-theoretic question could be related to the goal of obtaining an efficient algorithm. However, it turns out that the sum-of-squares framework allows us to systematically translate a uniqueness proof to an algorithm that efficiently finds the solution. (In addition, this solution also comes with a short certificate for uniqueness.[3])

Uniqueness proof. Let $\Omega \subseteq [n]^3$ be a set of entries and let $X = \sum_{i=1}^r \lambda_i \cdot u_i \otimes v_i \otimes w_i$ be a 3-tensor with $\lambda_1, \ldots, \lambda_r \geqslant 0$.

It turns out that the following two conditions are enough to imply that X_Ω uniquely determines X: The first condition is that the vectors $\{(u_i \otimes v_i \otimes w_i)_\Omega\}$ are linearly independent. The second condition is that exists a 3-linear form T on $\mathbb{R}^n$ with the following properties:

1. in the monomial basis T is supported on Ω so that $T(x, y, z) = \sum_{(i,j,k)\in\Omega} T_{ijk} \cdot x_i y_j x_k$,

2. evaluated over unit vectors, the 3-form T is exactly maximized at the points (u_i, v_i, w_i) so that $T(u_1, v_1, w_1) = \cdots = T(u_r, v_r, w_r) = 1$ and $T(x, y, z) < 1$ for all unit vectors $(x, y, z) \notin \{(u_i, v_i, w_i) \mid i \in [r]\}$.

We show that the two deterministic conditions above are satisfied with high probability if the vectors $\{u_i\}, \{v_i\}, \{w_i\}$ are incoherent and Ω is a random set of entries of size at least $r \cdot \tilde{O}(n^{1.5})$.

Let us sketch the proof that such a 3-linear form T indeed implies uniqueness. Concretely, we claim that if we let X' be a 3-tensor of the form $\sum_{i=1}^{r'} \lambda_i' \cdot u_i' \otimes v_i' \otimes w_i'$ for $\lambda_1', \ldots, \lambda_{r'}' \geqslant 0$

---

[2]We emphasize that we ask here about the uniqueness of X for a fixed set of entries Ω. This questions differs from asking about the uniqueness for a random set of entries, which could be answered by suitably counting the number of low-rank 3-tensors.

[3]This certificate is closely related to certificates in the form of dual solutions for convex programming relaxations that are used in the compressed sensing and matrix completion literature.

and unit vectors $\{u_i'\}, \{v_i'\}, \{w_i'\}$ with $X'_\Omega = X_\Omega$ that minimizes $\sum_{i=1}^{r'} |\lambda_i'|$ then $X' = X$ must hold. We identify T with an element of $\mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ (the coefficient tensor of T in the monomial basis). Let X' be as before. We are to show that $X = X'$. One the one hand, using that $T(x, y, z) \leqslant 1$ for all unit vectors x, y, z,

$$\langle T, X' \rangle = \sum_{i=1}^{r'} \lambda_i' \cdot T(u_i', v_i', w_i') \leqslant \sum_{i=1}^{r'} \lambda_i'.$$

At the same time, using that T is supported on Ω and the fact that $X_\Omega = X'_\Omega$,

$$\langle T, X' \rangle = \langle T, X \rangle = \sum_{i=1}^{r} \lambda_i \cdot T(u_i, v_i, w_i) = \sum_{i=1}^{r} \lambda_i.$$

Since X' minimizes $\sum_{i=1}^{r'} \lambda_i'$, equality has to hold in the previous inequality. It follows that every point (u_i', v_i', w_i') is equal to one of the points (u_j, v_j, w_j), because T is uniquely maximized at the points $\{(u_i, v_i, w_i) \mid i \in [r]\}$. Since we assumed that $\{(u_i \otimes v_i \otimes w_i)_\Omega\}$ is linearly independent, we can conclude that $X = X'$.

When we show that such a 3-linear form T exists, we will actually show something stronger, namely that the second property is not only true but also has a short certificate in form of a "degree-4 sum-of-squares proof", which we describe next. This certificate also enables us to efficiently recover the missing tensor entries.

Uniqueness proof in the sum-of-squares system. A degree-4 sos certificate for the second property of T is an $(n + n^2)$-by-$(n + n^2)$ positive-semidefinite matrix M (acting as a linear operator on $\mathbb{R}^n \oplus (\mathbb{R}^n \otimes \mathbb{R}^n)$) that represents the polynomial $\|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 2T(x, y, z)$, i.e.,

$$\langle (x, y \otimes z), M(x, y \otimes z) \rangle = \|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 2T(x, y, z). \tag{2.1}$$

Furthermore, we require that the kernel of M is precisely the span of the vectors $\{(u_i, v_i \otimes w_i) \mid i \in [r]\}$. Let's see that this matrix M certifies that T has the property that over unit vectors it is exactly maximized at the desired points (u_i, v_i, w_i). Let u, v, w be unit vectors such that (u, v, w) is not a multiple of one of the vectors (u_i, v_i, w_i). Then by orthogonality, both $(u, v \otimes w)$ and $(-u, v \otimes w)$ have non-zero projection on the orthogonal complement of the kernel of M. Therefore, the bounds $0 < \langle (u, v \otimes w), M(u, v \otimes w) \rangle = 2 - 2p(u, v, w)$ and $0 < \langle (-u, v \otimes w), M(-u, v \otimes w) \rangle = 2 + 2p(u, v, w)$ together give the desired conclusion that $|T(u, v, w)| < 1$.

Reconstruction algorithm based on the sum-of-squares system. The existence of a positive semidefinite matrix M as above not only means that reconstruction of X from X_Ω is possible information-theoretically but also efficiently. The sum-of-squares algorithm allows us to efficiently search over low-degree moments of objects called *pseudo-distributions* that generalize probability distributions over real vector spaces. Every pseudo-distribution

μ defines *pseudo-expectation values* $\tilde{\mathbb{E}}_\mu f$ for all low-degree polynomial functions $f(x, y, z)$, which behave in many ways like expectation values under an actual probability distribution. In order to reconstruct X from the observations X_Ω, we use the sum-of-squares algorithm to efficiently find a pseudo-distribution μ that satisfies[4]

$$\tilde{\mathbb{E}}_{\mu(x,y,z)} \|x\|^2 + \|y\|^2 \cdot \|z\|^2 \leqslant 1 \tag{2.2}$$

$$\left(\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z \right)_\Omega = X_\Omega \tag{2.3}$$

Note that the distribution over the vectors (u_i, v_i, w_i) with probabilities λ_i satisfies the above conditions. Our previous discussion about uniqueness shows that the existence of a positive semidefinite matrix M as above implies no other distribution satisfies the above conditions. It turns out that the matrix M implies that this uniqueness holds even among pseudo-distributions in the sense that any pseudo-distribution that satisfies Eqs. (2.2) and (2.3) must satisfy $\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z = X$, which means that the reconstruction is successful.[5]

When do such uniqueness certificates exist? The above discussion shows that in order to achieve reconstruction it is enough to show that uniqueness certificates of the form above exist. We show that these certificates exists with high probability if we choose Ω to be a large enough random subset of entries (under suitable assumptions on X). Our existence proof is based on a randomized procedure to construct such a certificate heavily inspired by similar constructions for matrix completion [Gro11, Rec11]. (We note that this construction uses the unknown tensor X and is therefore not "constructive" in the context of the recovery problem.)

Before describing the construction, we make the requirements on the 3-linear form T more concrete. We identify T with the linear operator from $\mathbb{R}^n \otimes \mathbb{R}^n$ to $\mathbb{R}^n$ such that $T(x, y, z) = \langle x, T(y \otimes z) \rangle$. Furthermore, let T_a be linear operators on $\mathbb{R}^n$ such that $T(x, y, z) = \sum_{a=1}^n x_a \cdot \langle y, T_a z \rangle$. Then, the following conditions on T imply the existence of a uniqueness certificate M (which also means that recover succeeds),

1. every unknown entry $(i, j, k) \notin \Omega$ satisfies $\langle e_i, T(e_j \otimes e_k) \rangle = 0$,

2. every index $i \in [r]$ satisfies $u_i = T(v_i \otimes w_i)$,

3. the matrix $\sum_{a=1}^n T_a \otimes T_a{}^\intercal - \sum_{i=1}^r (v_i \otimes w_i)(v_i \otimes w_i)^\intercal$ has spectral norm at most 0.01.

We note that the uniqueness certificates for matrix completion [Gro11, Rec11] have similar requirements. The key difference is that we need to control the spectral norm of an operator

---

[4]The viewpoint in terms of pseudo-distributions is useful to see how the previous uniqueness proof relates to the algorithm. We can also describe the solutions to the constraints Eqs. (2.2) and (2.3) in terms of linearly constrained positive semidefinite matrices. See alternative description of Algorithm 4.1

[5]The matrix M can also be viewed as a solution to the dual of the convex optimization problem of finding a pseudo-distribution that satisfies conditions Eqs. (2.2) and (2.3).

that depends quadratically on the constructed object T (as opposed to a linear dependence in the matrix completion case). Combined with the fact that the construction of T is iterative (about $\log n$ steps), the spectral norm bound unfortunately requires significant technical work. In particular, we cannot apply general matrix concentration inequalities and instead apply the trace moment method. (See Section 5.)

We also note that the fact that the above requirements allow us to construct the certifcate M is not immediate and requires some new ideas about matrix representations of polynomials, which might be useful elsewhere. (See Appendix A.)

Finally, we note that the transformation applied to T in order to obtain the matrix for the third condition above appears in many works about 3-tensors [HSS15, BM16] with the earliest appearance in a work on refutation algorithms for random 3-SAT instances (see [FO07]).

The iterative construction of the linear operator T exactly follows the recipe from matrix completion [Gro11, Rec11]. Let $\mathcal{R}_\Omega$ be the projection operator into the linear space of operators T that satify the first requirement. Let $\mathcal{P}_T$ be the (affine) projection operator into the affine linear space of operators T that satisfy the second reqirement. We start with $T^{(0)} = X$. At this point we satisfy the second condition. (Also the matrix in the third condition is 0.) In order to enforce the first condition we apply the operator $\mathcal{R}_\Omega$. After this projection, the second condition is most likely no longer satisfied. To enforce the second condition, we apply the affine linear operator $\mathcal{P}_T$ and obtain $T^{(1)} = \mathcal{P}_T(\mathcal{R}_\Omega X)$. The idea is to iterate this construction and show that after a logarithmic number of iterations both the first and second condition are satisfied up to an inverse polynomially small error (which we can correct in a direct way). The main challenge is to show that the iterates obtained in this way satisfy the desired spectral norm bound. (We note that for technical reasons the construction uses fresh randomness Ω for each iteration like in the matrix completion case [Rec11, Gro11]. Since the number of iterations is logarithmic, the total number of required observations remains the same up to a logarithmic factor.)

3 Preliminaries

Unless explicitly stated otherwise, $O(\cdot)$-notation hides absolute multiplicative constants. Concretely, every occurrence of $O(x)$ is a placeholder for some function $f(x)$ that satisfies $\forall x \in \mathbb{R}. |f(x)| \leqslant C|x|$ for some absolute constant $C > 0$. Similarly, $\Omega(x)$ is a placeholder for a function $g(x)$ that satisfies $\forall x \in \mathbb{R}. |g(x)| \geqslant |x|/C$ for some absolute constant $C > 0$.

Our algorithm is based on a generalization of probability distributions over $\mathbb{R}^n$. To define this generalization the following notation for the formal expectation of a function f on $\mathbb{R}^n$ with respect to a finitely-supported function $\mu \colon \mathbb{R}^n \to \mathbb{R}$,

$$\tilde{\mathbb{E}}_\mu f = \sum_{x \in \text{support}(\mu)} \mu(x) \cdot f(x).$$

A *degree-d pseudo-distribution over* $\mathbb{R}^n$ is a finitely-supported function $\mu \colon \mathbb{R}^n \to \mathbb{R}$ such that $\tilde{\mathbb{E}}_\mu 1 = 1$ and $\tilde{\mathbb{E}}_\mu f^2 \geqslant 0$ for every polynomial f of degree at most $d/2$.

A key algorithmic property of pseudo-distributions is that their low-degree moments have an efficient separation oracle. Concretely, the set of degree-d moments $\tilde{\mathbb{E}}_\mu(1, x)^{\otimes d}$ such that μ is a degree-d pseudo-distributions over $\mathbb{R}^n$ has an $n^{O(d)}$-time separation oracle. Therefore, standard convex optimization methods allow us to efficiently optimize linear functions over low-degree moments of pseudo-distributions (even subject to additional convex constraints that have efficient separation oracles) up to arbitrary numerical accuracy.

4 Tensor completion algorithm

In this section, we show that the following algorithm for tensor completion succeeds in recovering the unknown tensor from partial observations assuming the existence of a particular linear operator T. We will state conditions on the unknown tensor that imply that such a linear operator exists with high probability if the observed entries are chosen at random. We use essentially the same convex relaxation as in [BM16] but our analysis differs significantly.

Algorithm 4.1 (Exact tensor completion based on degree-4 sum-of-squares).
Input: locations $\Omega \subseteq [n]^3$ and partial observations X_Ω of an unknown 3-tensor $X \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$.
Operation: Find a degree-4 pseudo-distribution μ on $\mathbb{R}^n \oplus \mathbb{R}^n \oplus \mathbb{R}^n$ such that the third moment matches the observations $(\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z)_\Omega = X_\Omega$ so as to minimize

$$\tilde{\mathbb{E}}_{\mu(x,y,z)} \|x\|^2 + \|y\|^2 \cdot \|z\|^2 .$$

Output the 3-tensor $\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$.
Alternative description: Output a minimum trace, positive semidefinite matrix Y acting on $\mathbb{R}^n \oplus (\mathbb{R}^n \otimes \mathbb{R}^n)$ with blocks $Y_{1,1}$, $Y_{1,2}$ and $Y_{2,2}$ such that $(Y_{1,2})_\Omega = X_\Omega$ matches the observations, and $Y_{2,2}$ satisfies the additional symmetry constraints that each entry $\langle e_j \otimes e_k, Y_{2,2}(e_{j'} \otimes e_{k'}) \rangle$ only depends on the index sets $\{j, j'\}, \{k, k'\}$.

Let $\{u_i\}, \{v_i\}, \{w_i\}$ be three orthonormal sets in $\mathbb{R}^n$, each of cardinality r.

We reason about the recovery guarantees of the algorithm in terms of the following notion of certifcate.

Definition 4.2. We say that a linear operator T from $\mathbb{R}^n \otimes \mathbb{R}^n$ to $\mathbb{R}^n$ is a *degree-4 certificate* for Ω and orthonormal sets $\{u_i\}, \{v_i\}, \{w_i\} \subseteq \mathbb{R}^n$ if the following conditions are satisfies

1. the vectors $\{(u_i \otimes v_j \otimes w_k)_\Omega \mid (i, j, k) \in S\}$ are linearly independent, where $S \subseteq [n]^3$ is the set of triples with at least two identical indices from $[r]$,

2. every entry $(a, b, c) \notin \Omega$ satisfies $\langle e_a, T(e_b \otimes e_c) \rangle = 0$,

3. If we view T as a 3-tensor in $(\mathbb{R}^n)^{\otimes 3}$ whose (a, b, c) entry is $\langle e_a, T(e_b \otimes e_c) \rangle$, every index $i \in [r]$ satisfies $(u_i^\mathsf{T} \otimes v_i^\mathsf{T} \otimes \mathrm{Id})T = w_i$, $(u_i^\mathsf{T} \otimes \mathrm{Id} \otimes w_i^\mathsf{T})T = v_i$, and $(\mathrm{Id} \otimes v_i^\mathsf{T} \otimes w_i^\mathsf{T})T = u_i$.

4. the following matrix has spectral norm at most 0.01,

$$\sum_{a=1}^{n} T_a \otimes T_a^{\mathsf{T}} - \sum_{i=1}^{r} (v_i \otimes w_i)(v_i \otimes w_i)^{\mathsf{T}},$$

where $\{T_a\}$ are matrices such that $\langle x, T(y \otimes x) \rangle = \sum_{a=1}^{n} x_a \cdot \langle y, T_a z \rangle$.

In Section 4.4, we prove that existence of such certifcates implies that the above algorithm successfully recovers the unknown tensor, as formalized by the following theorem.

Theorem 4.3. *Let $X \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ be any 3-tensor of the form $\sum_{i=1}^{r} \lambda_i \cdot u_i \otimes v_i \otimes w_i$ for $\lambda_1, \ldots, \lambda_r \in \mathbb{R}_+$. Let $\Omega \subseteq [n]^3$ be a subset of indices. Suppose there exists degree-4 certificate in the sense of Definition 4.2. Then, given the observations X_Ω the above algorithm recovers the unknown tensor X exactly.*

In Section 4.5, we show that degree-4 scertificates are likely to exist when Ω is a random set of appropriate size.

Theorem 4.4. *Let $\{u_i\}, \{v_i\}, \{w_i\}$ be three orthonormal sets of μ-incoherent vectors in $\mathbb{R}^n$, each of cardinality r. Let $\Omega \subseteq [n]^3$ be a random set of tensor entries of cardinality $m = r \cdot n^{1.5} (\mu \log n)^C$ for an absolute constant $C \geqslant 1$. Then, with probability $1 - n^{-\omega(1)}$, there exists a linear operator T that satisfies the requirements of Definition 4.2.*

Taken together the two theorems above imply our main result Theorem 1.1.

4.1 Simpler proofs via higher-degree sum-of-squares

Unfortunately the proof of Theorem 4.4 requires extremely technical spectral norm bounds for random matrices.

It turns out that less technical norm bounds suffice if we use degree 6 sum-of-squares relaxations. For this more powerful algorithm, weaker certificates are enough to ensure exact recovery and the proof that these weaker certificates exist with high probability is considerably easier than the proof that degree-4 certificates exist with high probability.

In the following we describe this weaker notion of certificates and state their properties. In the subsequent sections we prove properties of these certificates are enough to imply our main result Theorem 1.1.

Algorithm 4.5 (Exact tensor completion based on higher-degree sum-of-squares).
Input: locations $\Omega \subseteq [n]^3$ and partial observations X_Ω of an unknown 3-tensor $X \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$.
Operation: Find a degree-6 pseudo-distribution μ on $\mathbb{R}^n \oplus \mathbb{R}^n \oplus \mathbb{R}^n$ so as to minimize $\tilde{\mathbb{E}}_{\mu(x,y,z)} \|x\|^2 + \|z\|^2$ subject to the following constraints

$$\left(\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z \right)_\Omega = X_\Omega, \tag{4.1}$$

$$\tilde{\mathbb{E}}_{\mu(x,y,z)} (\|y\|^2 - 1) \cdot p(x,y,z) = 0 \text{ for all } p(x,y,z) \in \mathbb{R}[x,y,z]_{\leqslant 14}. \tag{4.2}$$

Output the 3-tensor $\tilde{\mathbb{E}}_{\mu(x,y,z)} \, x \otimes y \otimes z \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$.

Let $\{u_i\}, \{v_i\}, \{w_i\}$ be three orthonormal sets in $\mathbb{R}^n$, each of cardinality r. We reason about the recovery guarantees of the above algorithm in terms of the following notion of certificate. The main difference to degree-4 certificate (Definition 4.2) is that the spectral norm condition is replaced by a condition in terms of sum-of-squares representations.

Definition 4.6. We say that a 3-tensor $T \in (\mathbb{R}^n)^{\otimes 3}$ is a *higher-degree certificate* for Ω and orthonormal sets $\{u_i\}, \{v_i\}, \{w_i\} \subseteq \mathbb{R}^n$ if the following conditions are satisfies

1. the vectors $\{(u_i \otimes v_i \otimes w_i)_\Omega\}_{i \in [r]}$ are linearly independent,

2. every entry $(a, b, c) \notin \Omega$ satisfies $\langle T, (e_a \otimes e_b \otimes e_c) \rangle = 0$,

3. every index $i \in [r]$ satisfies $(u_i^\mathsf{T} \otimes v_i^\mathsf{T} \otimes \mathrm{Id})T = w_i$, $(u_i^\mathsf{T} \otimes \mathrm{Id} \otimes w_i^\mathsf{T})T = v_i$, and $(\mathrm{Id} \otimes v_i^\mathsf{T} \otimes w_i^\mathsf{T})T = u_i$,

4. the following degree-4 polynomials in $\mathbb{R}[x, y, z]$ are sum of squares

$$\|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 1/\varepsilon \cdot \langle T', x \otimes y \otimes z \rangle, \tag{4.3}$$
$$\|y\|^2 + \|x\|^2 \cdot \|z\|^2 - 1/\varepsilon \cdot \langle T', x \otimes y \otimes z \rangle, \tag{4.4}$$
$$\|z\|^2 + \|x\|^2 \cdot \|y\|^2 - 1/\varepsilon \cdot \langle T', x \otimes y \otimes z \rangle. \tag{4.5}$$

where $T' = T - \sum_{i=1}^r u_i \otimes v_i \otimes w_i$ and $\varepsilon > 0$ is an absolute constant (say $\varepsilon = 10^{-6}$).

In the following sections we prove that higher-degree certificates imply that Algorithm 4.5 successfully recovers the desired tensor and that they exist with high probability for random Ω of appropriate size.

4.2 Higher-degree certificates imply exact recovery

Let $\{u_i\}, \{v_i\}, \{w_i\}$ be orthonormal bases in $\mathbb{R}^n$. We say that a degree-ℓ pseudo-distribution $\mu(x, y, z)$ satisfies the constraint $\|y\|^2 = 1$, denoted $\mu \models \{\|y\|^2 = 1\}$, if $\tilde{\mathbb{E}}_{\mu(x,y,z)} \, p(x, y, z) \cdot (1 - \|y\|^2) = 0$ for all polynomials $p \in \mathbb{R}[x, y, z]_{\leqslant \ell - 2}$

We are to show that a higher-degree certificate in the sense of Definition 4.6 implies that Algorithm 4.5 reconstructs the partially observed tensor exactly. A key step of this proof is the following lemma about expectation values of higher degree pseudo-distributions.

Lemma 4.7. *Let $T \in (\mathbb{R}^n)^{\otimes 3}$ be a higher-degree certificate as in Definition 4.6 for the set $\Omega \subseteq [n]^3$ and the vectors $\{u_i\}_{i \in [r]}, \{v_i\}_{i \in [r]}, \{w_i\}_{i \in [r]}$. Then, every degree-16 pseudo-distribution $\mu(x, y, z)$ with $\mu \models \{\|y\|^2 = 1\}$ satisfies*

$$\tilde{\mathbb{E}}_{\mu(x,y,z)} T(x, y, z) \leqslant \tilde{\mathbb{E}}_{\mu(x,y,z)} \frac{\|x\|^2 + \|z\|^2}{2} - \frac{1}{100} \cdot \sum_{i=r+1}^n \left(\langle u_i, x \rangle^2 + \langle w_i, z \rangle^2 \right)$$

$$- \frac{1}{100} \cdot \sum_{i=1}^n \sum_{j \in [n] \setminus \{i\}} \langle v_i, y \rangle^2 \cdot \left(\langle u_j, x \rangle^2 + \langle w_j, z \rangle^2 \right) \quad (4.6)$$

To prove this lemma it will be useful to introduce the sum-of-squares proof system. Before doing that let us observe that the lemma indeed allows us to prove that Algorithm 4.5 works.

Theorem 4.8 (Higher-degree certificates imply exact recovery). *Suppose there exists a higher-degree certificate T in the sense of Definition 4.6 for the set $\Omega \subseteq [n]^3$ and the vectors $\{u_i\}_{i\in[r]}, \{v_i\}_{i\in[r]}, \{w_i\}_{i\in[r]}$. Then, Algorithm 4.5 recovers the partially observed tensor exactly. In other words, if $X = \sum_{i=1}^r \lambda_i \cdot u_i \otimes v_i \otimes w_i$ with $\lambda_1, \ldots, \lambda_r \geqslant 0$ and $\mu(x,y,z)$ is a degree-16 pseudo-distribution with $\mu \models \{\|y\|^2 = 1\}$ that minimizes $\tilde{\mathbb{E}}_{\mu(x,y,z)} \frac{1}{2}(\|x\|^2 + \|z\|^2)$ subject to $(\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z)_\Omega = X_\Omega$, then $\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z = X$.*

Proof. Consider the distribution μ^* over vectors (x, y, z) such that $(\sqrt{\lambda_i n} \cdot u_i, v_i, \sqrt{\lambda_i n} \cdot w_i)$ has probability $1/n$. By construction, $\mathbb{E}_{\mu^*(x,y,z)} x \otimes y \otimes z = X$. We have

$$\tilde{\mathbb{E}}_{\mu(x,y,z)} T(x, y, z) = \mathbb{E}_{\mu^*(x,y,z)} T(x, y, z) = \sum_{i=1}^r \lambda_i = \mathbb{E}_{\mu^*(x,y,z)} \tfrac{1}{2}(\|x\|^2 + \|z\|^2).$$

By Lemma 4.7 and the optimality of μ, it follows that

$$\tilde{\mathbb{E}}_{\mu(x,y,z)} \tfrac{1}{100} \cdot \sum_{i=r+1}^n (\langle u_i, x\rangle^2 + \langle w_i, z\rangle^2) + \tfrac{1}{100} \cdot \sum_{i=1}^n \sum_{j\in[n]\backslash\{i\}} \langle v_i, y\rangle^2 \cdot \left(\langle u_j, x\rangle^2 + \langle w_j, z\rangle^2\right) = 0$$

Since the summands on the left-hand side are squares it follows that each summand has pseudo-expectation 0. It follows that $\tilde{\mathbb{E}}_\mu \langle u_i, x\rangle^2 = \tilde{\mathbb{E}}_\mu \langle v_i, y\rangle^2 = \tilde{\mathbb{E}}_\mu \langle w_i, z\rangle^2 = 0$ for all $i > r$ and $\tilde{\mathbb{E}}_\mu \langle v_i, y\rangle^2 \langle u_j, x\rangle^2 = \tilde{\mathbb{E}}_\mu \langle v_i, y\rangle^2 \langle w_j, x\rangle^2 = 0$ for all $i \neq j$. By the Cauchy–Schwarz inequality for pseudo-expectations, it follows that $\tilde{\mathbb{E}}_{\mu(x,y,z)} \langle x \otimes y \otimes z, u_i \otimes v_j \otimes w_k\rangle = 0$ unless $i = j = k \in [r]$. Consequently, $\tilde{\mathbb{E}}_{\mu(x,y,z)} x \otimes y \otimes z$ is a linear combination of the vectors $\{u_i \otimes v_i \otimes w_i \mid i \in [r]\}$. Finally, the linear independence of the vectors $\{(u_i \otimes v_i \otimes w_i)_\Omega \mid i \in [r]\}$ implies that $\tilde{\mathbb{E}}_\mu x \otimes y \otimes z = X$ as desired. $\qquad\square$

It remains to prove Lemma 4.7. Here it is convenient to use formal notation for sum-of-squares proofs. We will work with polynomials $\mathbb{R}[x, y, z]$ and the polynomial equation $\mathcal{A} = \{\|y\|^2 = 1\}$. For $p \in \mathbb{R}[x, y, z]$, we say that there exists a degree-ℓ SOS proof that $\mathcal{A}$ implies $p \geqslant 0$, denoted $\mathcal{A} \vdash_\ell p \geqslant 0$, if there exists a polynomial $q \in \mathbb{R}[x, y, z]$ of degree at most $\ell - 2$ such that $p + q \cdot (1 - \|y\|^2)$ is a sum of squares of polynomials. This notion proof allows us to reason about pseudo-distributions. In particular, if $\mathcal{A} \vdash_\ell p \geqslant 0$ then every degree-ℓ pseudo-distribution μ with $\mu \models \mathcal{A}$ satisfies $\tilde{\mathbb{E}}_\mu p \geqslant 0$.

We will change coordinates such that $u_i = v_i = w_i = e_i$ is the i-th coordinate vector for every $i \in [n]$. Then, the conditions on T in Definition 4.6 imply that

$$\langle T, (x \otimes y \otimes z)\rangle = \sum_{i=1}^r x_i y_i z_i + T'(x, y, z), \tag{4.7}$$

where T' is a 3-linear form with the property that $T'(x, x, x)$ does not contain squares (i.e. is multilinear). Furthermore, the conditions imply the following SOS proofs for T':

1. $\emptyset \vdash_4 T'(x, y, z) \leqslant \varepsilon \cdot \left(\|x\| + \|y\|^2 \cdot \|z\|^2 \right),$
2. $\emptyset \vdash_4 T'(x, y, z) \leqslant \varepsilon \cdot \left(\|y\| + \|x\|^2 \cdot \|z\|^2 \right),$
3. $\emptyset \vdash_4 T'(x, y, z) \leqslant \varepsilon \cdot \left(\|z\| + \|x\|^2 \cdot \|y\|^2 \right).$

The following lemma gives an upper bound on one of the parts in Eq. (4.7).

Lemma 4.9. *For $\mathcal{A} = \{\|y\|^2 = 1\}$, the following inequality has a degree-6 sum-of-squares proof,*

$$\mathcal{A} \vdash_6 \sum_{i=1}^r x_i y_i z_i \leqslant \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|z\|^2 - \tfrac{1}{4} \sum_{i=r+1}^n (x_i^2 + z_i^2)$$
$$- \tfrac{1}{8} \sum_{i \neq j} y_i^2 \cdot \left(x_j^2 + z_j^2 + y_j^2 \cdot (\|x\|^2 + \|z\|^2) \right). \quad (4.8)$$

Proof. We bound the left-hand side in the lemma as follows,

$$\mathcal{A} \vdash_6 \sum_{i=1}^r x_i y_i z_i \leqslant \sum_{i=1}^r (\tfrac{1}{2}x_i^2 + \tfrac{1}{2}y_i^2 z_i^2) \qquad (4.9)$$

$$\leqslant \tfrac{1}{2}\|x\|^2 - \tfrac{1}{2} \sum_{i>r} x_i^2 + \tfrac{1}{2} \sum_{i=1}^n y_i^2 z_i^2. \qquad (4.10)$$

We can further bound $\sum_i y_i^2 z_i^2$ as follows,

$$\mathcal{A} \vdash_6 \sum_{i=1}^n y_i^2 z_i^2 = \left(\sum_{i=1}^n y_i^2 \right) \cdot \left(\sum_{i=1}^n z_i^2 \right) - \sum_{i \neq j} y_i^2 \cdot z_j^2 \qquad (4.11)$$

$$= \left(\sum_{i=1}^n z_i^2 \right) - \sum_{i \neq j} y_i^2 \cdot z_j^2. \qquad (4.12)$$

We can prove a different bound on $\sum_i y_i^2 z_i^2$ as follows,

$$\mathcal{A} \vdash_6 \sum_{i=1}^n y_i^2 z_i^2 \leqslant \tfrac{1}{2}\|z\|^2 + \tfrac{1}{2} \sum_{i=1}^n y_i^4 z_i^2 \qquad (4.13)$$

$$\leqslant \tfrac{1}{2}\|z\|^2 + \tfrac{1}{2} \sum_{i=1}^n y_i^4 \|z\|^2 \qquad (4.14)$$

$$= \tfrac{1}{2}\|z\|^2 + \tfrac{1}{2} \left(\sum_{i=1}^n y_i^2 \right) \cdot \left(\sum_{i=1}^n y_i^2 \|z\|^2 \right) - \tfrac{1}{2} \sum_{i \neq j} y_i^2 \cdot y_j^2 \|z\|^2 \qquad (4.15)$$

$$= \|z\|^2 - \tfrac{1}{2} \sum_{i \neq j} y_i^2 \cdot y_j^2 \|z\|^2. \qquad (4.16)$$

By combining these three inequalities, we obtain the inequality

$$\mathcal{A} \vdash_6 \sum_{i=1}^r x_i y_i z_i \leqslant \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|z\|^2 - \tfrac{1}{2} \sum_{i>r} x_i^2 - \tfrac{1}{2} \sum_{i \neq j} y_i^2 \cdot z_j^2 - \tfrac{1}{4} \sum_{i \neq j} y_i^2 \cdot y_j^2 \|z\|^2.$$

By symmetry between z and x, the same inequality holds with x and z exchanged. Combining these symmetric inequalities, we obtain the desired inequality

$$\mathcal{A} \vdash_6 \sum_{i=1}^{r} x_i y_i z_i \leqslant \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|z\|^2 - \tfrac{1}{4}\sum_{i>r}(x_i^2 + z_i^2)$$
$$- \tfrac{1}{4}\sum_{i\neq j} y_i^2 \cdot (x_j^2 + z_j^2) - \tfrac{1}{8}\sum_{i\neq j} y_i^2 \cdot y_j^2(\|x\|^2 + \|z\|^2). \quad (4.17)$$

$\square$

It remains to bound the second part in Eq. (4.7), which the following lemma achieves.

Lemma 4.10. $\mathcal{A} \vdash_6 T'(x,y,z) \leqslant \frac{3\varepsilon}{2} \sum_i \sum_{j\neq i} y_i^2 \left(x_j^2 + z_j^2 + \tfrac{1}{2}y_j^2(\|x\|^2 + \|z\|^2) \right)$

Proof. It is enough to show the following inequality for all $i \in [n]$,

$$\mathcal{A} \vdash_6 y_i^2 T'(x,y,z) \leqslant \varepsilon \sum_{j\neq i} \left(\frac{3}{2}y_i^2(x_j^2 + z_j^2) + \frac{1}{2}y_i^2 y_j^2(\|x\|^2 + \|z\|^2) \right)$$

By symmetry it suffices to consider the case $i = 1$. Let $x' = x - x_1 \cdot e_1$, $y' = y - y_1 \cdot e_1$, and $z' = z - z_1 \cdot e_1$. We observe that

$$\mathcal{A} \vdash_4 T'(x,y,z) = T'(x_1 e_1 + x', y_1 e_1 + y', z_1 e_1 + z')$$
$$= T'(x_1 e_1, y', z') + T'(x', y_1 e_1, z')$$
$$+ T'(x', y', z_1 e_1) + T'(x', y', z')$$

We now apply the following inequalities

1. $\mathcal{A} \vdash_4 T'(x_1 e_1, y', z') \leqslant \frac{\varepsilon}{2}\left(x_1^2\|y'\|^2 + \|z'\|^2 \right) \leqslant \frac{\varepsilon}{2}\sum_{j\neq 1}(y_j^2\|x\|^2 + z_j^2)$

2. $\mathcal{A} \vdash_4 T'(x', y_1 e_1, z') \leqslant \frac{\varepsilon}{2}\left(\|x'\|^2 y_1^2 + \|z'\|^2 \right) \leqslant \frac{\varepsilon}{2}\sum_{j\neq 1}(x_j^2 + z_j^2)$

3. $\mathcal{A} \vdash_4 T'(x', y', z_1 e_1) \leqslant \frac{\varepsilon}{2}\left(z_1^2\|y'\|^2 + \|x'\|^2 \right) \leqslant \frac{\varepsilon}{2}\sum_{j\neq 1}(y_j^2\|z\|^2 + x_j^2)$

4. $\mathcal{A} \vdash_4 T'(x', y', z') \leqslant \frac{\varepsilon}{2}\left(\|x'\|^2\|y'\|^2 + \|z'\|^2 \right) \leqslant \frac{\varepsilon}{2}\sum_{j\neq 1}(x_j^2 + z_j^2)$

$\square$

We can now prove Lemma 4.7.

Proof of Lemma 4.7. Taken together, Lemmas 4.9 and 4.10 imply

$$\mathcal{A} \vdash_6 \langle T, x \otimes y \otimes z \rangle \leqslant \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|z\|^2 - (\tfrac{1}{4} - O(\varepsilon))\sum_{i=r+1}^{n}(x_i^2 + z_i^2)$$
$$- (\tfrac{1}{8} - O(\varepsilon))\sum_{i\neq j} y_i^2 \cdot \left(x_j^2 + z_j^2 + y_j^2 \cdot (\|x\|^2 + \|z\|^2) \right), \quad (4.18)$$

where the absolute constant hidden by $O(\cdot)$ notation is at most 10. Therefore for $\varepsilon < 1/100$, as we assumed in Definition 4.6, we get a SOS proof of the inequality,

$$\mathcal{A} \vdash_6 \langle T, x \otimes y \otimes z \rangle \leqslant \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|z\|^2 - \tfrac{1}{8}\sum_{i=r+1}^{n}(x_i^2 + z_i^2)$$
$$- \tfrac{1}{16}\sum_{i \neq j} y_i^2 \cdot \left(x_j^2 + z_j^2 + y_j^2 \cdot (\|x\|^2 + \|z\|^2)\right), \quad (4.19)$$

This SOS proof implies that that every degree-6 pseudo-distribution $\mu(x, y, z)$ with $\mu \models \mathcal{A}$ satisfies the desired inequality,

$$\tilde{\mathbb{E}}_{\mu(x,y,z)} \langle T, x \otimes y \otimes z \rangle \leqslant \tilde{\mathbb{E}}_{\mu(x,y,z)} \tfrac{1}{2}\|x\|^2 + \tfrac{1}{2}\|z\|^2 - \tfrac{1}{8}\sum_{i=r+1}^{n}(x_i^2 + z_i^2)$$
$$- \tfrac{1}{16}\sum_{i \neq j} y_i^2 \cdot \left(x_j^2 + z_j^2 + y_j^2 \cdot (\|x\|^2 + \|z\|^2)\right), \quad (4.20)$$

$\square$

4.3 Constructing the certificate T

In this section we give a procedure for constructing the certificate T. This construction is directly inspired by the construction of the dual certificate in [Gro11, Rec11] (sometimes called quantum golfing). We will then prove that T satisfies all of the conditions for a higher-degree certificate of Ω. In Section 4.5 we will show that T also satisfies the conditions for a degree-4 certificate for Ω.

Let $\{u_i\}, \{v_i\}, \{w_i\} \subseteq \mathbb{R}^n$ be three orthonormal bases, with all vectors μ-incoherent. Let $X = \sum_{i=1}^{r} u_i \otimes v_i \otimes w_i$. Let $\Omega \subseteq [n]^3$ chosen at random such that each element is included independently with probability $m/n^{1.5}$ (so that $|\Omega|$ is tightly concentrated around m).

Let P be the projector on the span of the vectors $u_i \otimes v_j \otimes w_k$ such that an index in $[r]$ appears at least twice in (i, j, k) (i.e., at least one of the conditions $i = j \in [r]$, $i = k \in [r]$, $j = k \in [r]$ is satisfied). Let R_Ω be the linear operator on $\mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ that sets all entries outside of Ω to 0 (so that $(R_\Omega[T])_\Omega = R_\Omega[T]$) and is scaled such that $\mathbb{E}_\Omega R_\Omega = \text{Id}$. Let $\bar{R}_\Omega$ be $\text{Id} - R_\Omega$.

Our goal is to construct $T \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ such that $P[T] = X$, $(T)_\Omega = T$, and the spectral norm condition in Definition 4.2 is satisfied. The idea for constructing T is to start with $T = X$. Then, move to closest point T' that satisfies $R_\Omega[T'] = T'$ Then, move to closest point T'' that satisfies $P[T''] = X$ and repeat. To implement this strategy, we define

$$T^{(k)} = \sum_{j=0}^{k-1} (-1)^j R_{\Omega_{j+1}} (P\bar{R}_{\Omega_j}) \cdots (P\bar{R}_{\Omega_1})[X], \quad (4.21)$$

where $\Omega_1, \ldots, \Omega_k$ are iid samples from the same distribution as Ω.

By induction, we can show the following lemma about linear constraints that the constructed stensors $T^{(k)}$ satisfy.

Lemma 4.11. *For every $k \geqslant 1$, the tensor $T^{(k)}$ satisfies $(T)_\Omega = T$ and*

$$P[T^{(k)}] + (-1)^k P(P\bar{R}_{\Omega_k}) \cdots (P\bar{R}_{\Omega_1})[X] = X.$$

Here, $P(P\bar{R}_{\Omega_k}) \cdots (P\bar{R}_{\Omega_1})[X]$ is an error term that decreases geometrically. In the parameter regime of Theorem 4.4, the norm of this term is $n^{-\omega(1)}$ for some $k = (\log n)^{O(1)}$.

The following lemma shows that it is possible to correct such small errors. This lemma also implies that the linear independence condition in Definition 4.2 is satisfied with high probability. (Therefore, we can ignore this condition in the following.)

Lemma 4.12. *Suppose $m \geqslant rn\mu \cdot (\log n)^{O(1)}$. Then, with probability $1 - n^{-\omega(1)}$ over the choice of Ω, the following holds: For every $E \in \mathbb{R}^n \otimes \mathbb{R}^n \otimes \mathbb{R}^n$ with $P[E] = E$, there exists Y with $(Y)_\Omega = Y$ such that $P[Y] = E$ and $\|Y\|_F \leqslant O(1) \cdot \|E\|_F$.*

Proof. Let $S \subseteq [n]^3$ be such that P is the projector to the vectors $u_i \otimes v_j \otimes w_k$ with $(i, j, k) \in S$. By construction of P we have $|S| \leqslant 3rn$. In order to show the conclusion of the lemma it is enough to show that the vectors $(u_i \otimes v_j \otimes w_k)_\Omega$ with $(i, j, k) \in S$ are well-conditioned in the sense that the ratio of the largest and smallest singular value is $O(1)$. This fact follows from standard matrix concentration inequalities. See Lemma 4.14. $\square$

The main technical challenge is to show that the construction satisfies the condition that the following degree 4 polynomials are sums of squares (where $T' = T - X$).

$$\|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 1/\varepsilon \cdot \langle T', x \otimes y \otimes z \rangle, \tag{4.22}$$

$$\|y\|^2 + \|x\|^2 \cdot \|z\|^2 - 1/\varepsilon \cdot \langle T', x \otimes y \otimes z \rangle, \tag{4.23}$$

$$\|z\|^2 + \|x\|^2 \cdot \|y\|^2 - 1/\varepsilon \cdot \langle T', x \otimes y \otimes z \rangle. \tag{4.24}$$

We show how to prove the first statement, the other statements can be proved with symmetrical arguments. To prove the first statement, we decompose T' into pieces of the form $(\bar{R}_{\Omega_l} P) \cdots (\bar{R}_{\Omega_1} P)(\bar{R}_{\Omega_0} X)$, $P'(\bar{R}_{\Omega_l} P) \cdots (\bar{R}_{\Omega_1} P)(\bar{R}_{\Omega_0} X)$ (where we will split P into parts P'), or E. For each piece A, we prove a norm bound $\|\sum_a A_a \otimes A_a^T\| \leqslant B$. Since $\sum_a A_a \otimes A_a^T$ represents the same polynomial as $A^\intercal A$, this proves that $B\|y\|^2\|z\|^2 - (y \otimes z)^\intercal A^\intercal A(y \otimes z)$ is a degree 4 sum of squares. Now note that $(y \otimes z)^\intercal A^\intercal A(y \otimes z) - \sqrt{B}x^\intercal A(y \otimes z) - \sqrt{B}(y \otimes z)^\intercal A^\intercal x + B\|x\|^2$ is also a sum of squares. Combining these equations and scaling we have that $\|x\|^2 + \|y\|^2\|z\|^2 - \frac{2}{\sqrt{B}}x^\intercal A(y \otimes z)$ is a degree 4 sum of squares.

Thus, it is sufficient to prove norm bounds on $\|\sum_a A_a \otimes A_a^T\|$. We have an appropriate bound in the case when $A = E$ because E has very small Frobenius norm. For the cases when $A = (\bar{R}_{\Omega_l} P) \cdots (\bar{R}_{\Omega_1} P)(\bar{R}_{\Omega_0} X)$ or $A = P(\bar{R}_{\Omega_l} P) \cdots (\bar{R}_{\Omega_1} P)(\bar{R}_{\Omega_0} X)$, we use the following theorem

Theorem 4.13. *Let $A = (\bar{R}_{\Omega_l} P) \cdots (\bar{R}_{\Omega_1} P)(\bar{R}_{\Omega_0} X)$ or $P'(\bar{R}_{\Omega_l} P) \cdots (\bar{R}_{\Omega_1} P)(\bar{R}_{\Omega_0} X)$ where P' is a part of P. There is an absolute constant C such that for any $\alpha > 1$ and $\beta > 0$,*

$$\mathbb{P}\left[\left\|\sum_a A_a \otimes A_a^T\right\| > \alpha^{-(l+1)}\right] < n^{-\beta}$$

as long as $m > C\alpha\beta\mu^{\frac{3}{2}}rn^{1.5} \cdot \log(n)$ and $m > C\alpha\beta\mu^2 rn \log(n)$.

Proof. This theorem follows directly from combining Proposition 5.10, Theorem 6.1, Theorem 7.1, and Theorem 8.1. □

4.3.1 Final correction of error terms

In this section, we prove a spectral norm bound that allows us to correct error terms that are left at the end of the construction. The proof uses the by now standard Matrix Bernstein concentration inequality. Similar proofs appear in the matrix completion literature [Gro11, Rec11].

Let $\{u_i\}, \{v_i\}, \{w_i\} \subseteq \mathbb{R}^n$ be three orthonormal bases, with all vectors μ-incoherent. Let $\Omega \subseteq [n]^3$ be m entires sampled uniformly at random with replacement. (This sampling model is different from what is used in the rest of the proof. However, it is well known that the models are equivalent in terms of the final recovery problem.)

Lemma 4.14. *Let $S \subseteq [n]^3$. Suppose $m = \mu|S|(\log n)^C$ for an absolute constant $C \geqslant 1$. Then with probability $1 - n^{\omega(1)}$ over the choice of Ω, the vectors $(u_i \otimes v_j \otimes w_k)_\Omega$ for $(i, j, k) \in S$ are well-conditioned in the sense that the ratio between the largest and smallest singular value is at most 1.1.*

Proof. For $s = (i, j, k) \in S$, let $y_s = u_i \otimes v_j \otimes w_k$. Let $\Omega = \{\omega_1, \dots, \omega_m\}$, where $\omega_1, \dots, \omega \in [n]^3$ are sampled uniformly at random with replacement. Let A be the S-by-S Gram matrix of the vectors $(y_s)_\Omega$. Then, A is the sum of m identically distributed rank-1 matrices A_i,

$$A = \sum_{i=1}^m A_i \quad \text{with} \quad (A_i)_{s,s'} = (y_s)_{\omega_i} \cdot (y_{s'})_{\omega_i}.$$

Each A_i has expectation $\mathbb{E} A_i = n^{-1.5} \mathrm{Id}$ and spectral norm at most $|S| \cdot \mu/n^{1.5}$. Standard matrix concentration inequalities [Tro12] show that $m \geqslant O(|S|\mu^2 \log n)$ is enough to ensure that the sum is spectral close to its expectation $(m/n^{1.5}) \mathrm{Id}$ in the sense that $0.99A \leq (m/n^{1.5}) \mathrm{Id} \leq 1.1A$. □

4.4 Degree-4 certificates imply exact recovery

In this section we prove Theorem 4.3. We need the following technical lemma, which we prove in Appendix A.

Lemma 4.15. *Let R be self-adjoint linear operator R on $\mathbb{R}^n \otimes \mathbb{R}^n$. Suppose $\langle (v_j \otimes w_k), R(v_i \otimes w_i) \rangle = 0$ for all indices $i, j, k \in [r]$ such that $i \in \{j, k\}$. Then, there exists a self-adjoint linear operator R' on $\mathbb{R}^n \otimes \mathbb{R}^n$ such that $R'(v_i \otimes w_i) = 0$ for all $i \in [r]$, the spectral norm of R' satisfies $\|R'\| \leqslant 10\|R\|$, and R' represents the same polynomial in $\mathbb{R}[y, z]$,*

$$\langle (y \otimes z), R'(y \otimes z) \rangle = \langle (y \otimes z), R(y \otimes z) \rangle.$$

We can now prove that certificates in the sense of Definition 4.2 imply the our algorithm successfully recoves the unknown tensor.

15

Proof of Theorem 4.3. Let T be a certificate in the sense of Definition 4.2.

Our goal is to construct a positive semidefinite matrix M on $\mathbb{R}^n \oplus \mathbb{R}^n \otimes \mathbb{R}^n$ that represents the following polynomial

$$\langle (x, y \otimes z), M(x, y \otimes z) \rangle = \|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 2\langle x, T(y \otimes z) \rangle.$$

Let T_a be matrices such that $\langle x, T(x \otimes y) \rangle = \sum_a x_a \cdot T_a(y, z)$. Since $\|x\|^2 + \sum_{a=1}^n T_a(y, z)^2 - 2\langle x, T(y \otimes z) \rangle = \|x - T(y \otimes z)\|$ is a sum of squares of polynomials, it will be enough to find a positive semidefinite matrix that represents the polynomial $\|y\|^2 \cdot \|z\|^2 - \sum_{a=1}^n T_a(y, z)^2$. (This step is a polynomial version of the Schur complement condition for positive semidefiniteness.) Let R be the following linear operator

$$R = \sum_{a=1}^n T_a \otimes T_a^\intercal - \sum_{i=1}^r (v_i \otimes w_i)(v_i \otimes w_i)^\intercal,$$

Lemma 4.16. *R satisfies the requirement of Lemma 4.15.*

Proof. Consider $\langle (v_j \otimes w_k), R(v_j \otimes w_j) \rangle$. Since v_j is repeated, the value of this expression will be the same if we replace R by an R_2 which represents the same polynomial. Thus, we can replace R by $R_2 = \sum_{a=1}^n T_a^\intercal T_a - \sum_{i=1}^r (v_i \otimes w_i)(v_i \otimes w_i)^\intercal = T^\intercal T - \sum_{i=1}^r (v_i \otimes w_i)(v_i \otimes w_i)^\intercal$

We now observe that $\langle (v_j \otimes w_k), R_2(v_j \otimes w_j) \rangle = \langle (v_j \otimes w_k), T^\intercal(u_j) - (v_j \otimes w_j) \rangle = 0$. By a symmetrical proof, $\langle (v_j \otimes w_k), R(v_k \otimes w_k) \rangle = 0$ as well. $\qquad\square$

By Lemma Lemma 4.15, there exists a self-adjoint linear operator R' that represents the same polynomial as R, has spectral norm $\|R'\| \leq 10\|R\| \leq 0.1$, and sends all vectors $v_i \otimes w_i$ to 0. Since R' sends all vectors $v_i \otimes w_i$ to 0 and $\|R'\| \leq 0.1$, the following matrix

$$R'' = \sum_{i=1}^r (v_i \otimes w_i)(v_i \otimes w_i)^\intercal + R'$$

has r eigenvalues of value 1 (corresponding to the space spanned by $v_i \otimes w_i$) and all other eigenvalues are at most 0.1 (because the non-zero eigenvalues of R' have eigenvectors orthogonal to all $v_i \otimes w_i$). At the same time, R'' represents the following polynomial,

$$\langle (y \otimes z), R''(y \otimes z) \rangle = \sum_{a=1}^n T_a(y, z)^2.$$

Let P be a positive semidefinite matrix that represents the polynomial $\|x\|^2 + \sum_{a=1}^n T_a(y, z)^2 - 2\langle x, T(y \otimes z) \rangle$ (such a matrix exists because the polynomial is a sum of squares). We choose M as follows

$$M = \begin{pmatrix} \mathrm{Id} & -T \\ (T)^\intercal & (T)^\intercal T \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{Id} - R'' \end{pmatrix}$$

Since $R'' \preceq \mathrm{Id}$, this matrix is positive semidefinite. Also, M represents $\|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 2\langle x, T(y \otimes z) \rangle$. Since $u_i = T(v_i \otimes w_i)$ for all $i \in [r]$ and the kernel of $\mathrm{Id} - R''$ only contains span of $v_i \otimes w_i$, the kernel of M is exactly the span of the vectors $(u_i, v_i \otimes w_i)$.

16

Next, we show that the above matrix M implies that Algorithm 4.1 recovers the unknown tensor X. Recall that the algorithm on input X_Ω finds a pseudo-distribution $\mu(x, y, z)$ so as to minimize $\tilde{\mathbb{E}}_\mu \|x\|^2 + \|y\|^2 \cdot \|z\|^2$ such that $(\tilde{\mathbb{E}}_\mu \, x \otimes y \otimes z)_\Omega = X_\Omega$. Since everything is scale invariant, we may assume that $X = \sum_{i=1}^r \lambda_i \cdot u_i \otimes v_i \otimes w_i$ for $\lambda_1, \ldots, \lambda_r \geqslant 0$ and $\sum_i \lambda_i = 1$. Then, a valid pseudo-distribution would be the probability distribution over $(u_1, v_1, w_1), \ldots, (u_r, v_r, w_r)$ with probabilities $\lambda_1, \ldots, \lambda_r$. Let μ be the pseudo-distribution computed by the algorithm. By optimality of μ, we know that the objective value satisfies $\tilde{\mathbb{E}}_\mu \|x\|^2 + \|y\|^2 \cdot \|z\|^2 \leqslant \mathbb{E}_{i \sim \lambda} \|u_i\|^2 + \|v_i\|^2 \cdot \|w_i\|^2 = 2$. Then, if we let $Y = \mathbb{E}_\mu(x, y \otimes z)(x, y \otimes z)^\intercal$,

$$
\begin{aligned}
0 \leqslant \langle M, Y \rangle &= \tilde{\mathbb{E}}_{\mu(x,y,z)} \|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 2\langle x, T(y \otimes z) \rangle \\
&\leqslant 2 - 2 \tilde{\mathbb{E}}_{\mu(x,y,z)} \langle x, T(y \otimes z) \rangle \\
&= 2 - 2 \mathbb{E}_{i \sim \lambda} \langle u_i, T(v_i \otimes w_i) \rangle \\
&= 0
\end{aligned}
$$

The first step uses that M and Y are psd. The second step uses that M represents the polynomial $\|x\|^2 + \|y\|^2 \cdot \|z\|^2 - 2\langle x, T(y \otimes z) \rangle$. The third step uses that μ minimizes the objective function. The fourth step uses that the entries of T are 0 outside of Ω and that μ matches the observations $(\tilde{\mathbb{E}}_\mu \, x \otimes y \otimes z)_\Omega = X_\Omega$. The last step uses that $u_i = T(v_i \otimes w_i)$ for all $i \in [r]$.

We conclude that $\langle M, Y \rangle = 0$, which means that the range of Y is contained in the kernel of M. Therefore, $Y = \sum_{i,j=1}^r \gamma_{i,j} \cdot (u_i, v_i \otimes w_i)(u_j, v_j \otimes w_j)^\intercal$ for scalars $\{\gamma_{i,j}\}$. We claim that the multipliers must satisfy $\gamma_{i,i} = \lambda_i$ and $\gamma_{i,j} = 0$ for all $i \neq j \in [r]$. Indeed since μ matches the observations in Ω,

$$
0 = \sum_{i,j=1}^n (\lambda_i - \gamma_{i,j}\delta_{ij}) \cdot (u_i \otimes v_j \otimes w_j)_\Omega .
$$

Since the vectors $(u_i \otimes v_j \otimes w_j)_\Omega$ are linearly independent, we conclude that $\gamma_{i,j} = \lambda_i \cdot \delta_{ij}$ as desired. (This linear independence was one of the requirements of the certificate in Definition 4.2.) $\qquad\square$

4.5 Degree-4 certificates exist with high probability

In this section we show that our certificate T in fact satisfies the conditions for a degree-4 certificate, proving Theorem 4.4.

We use the same construction as in Section 4.3. The main, remaining technical challenge for Theorem 4.4 is to show that the construction satisfies the spectral norm condition of Definition 4.2. This spectral norm bound follows from the following theorem which we give a proof sketch for in Appendix B.

Theorem 4.17. *Let* $A = (\bar{R}_{\Omega_l}P)\cdots(\bar{R}_{\Omega_1}P)(\bar{R}_{\Omega_0}X)$ *or* $P(\bar{R}_{\Omega_l}P)\cdots(\bar{R}_{\Omega_1}P)(\bar{R}_{\Omega_0}X)$ *and* $B = (\bar{R}_{\Omega_{l'}}P)\cdots(\bar{R}_{\Omega_1}P)(\bar{R}_{\Omega_0}X)$ *or* $P(\bar{R}_{\Omega_{l'}}P)\cdots(\bar{R}_{\Omega_1}P)(\bar{R}_{\Omega_0}X)$. *There is an absolute constant* C *such that for any* $\alpha > 1$ *and* $\beta > 0$,

$$\mathbb{P}\left[\left\|\sum_a A_a \otimes B_a^T\right\| > \alpha^{-(l+l'+2)}\right] < n^{-\beta}$$

as long as $m > C\alpha\beta\mu^{\frac{3}{2}}rn^{1.5} \cdot \log(n)$ *and* $m > C\alpha\beta\mu^2 rn\log(n)$.

Remark 4.18. If it were true in general that $\|\sum_a A_a \otimes B_a^T\| \leqslant \sqrt{\|\sum_a A_a \otimes A_a^T\|}\sqrt{\|\sum_a B_a \otimes B_a^T\|}$ then it would be sufficient to use Theorem 4.13 and we would not need to prove Theorem 4.17. Unfortunately, this is not true in general.

That said, it may be possible to show that even if we do not know directly that $\|\sum_a A_a \otimes B_a^T\|$ is small, since $\|\sum_a A_a \otimes A_a^T\|$ and $\|\sum_a B_a \otimes B_a^T\|$ are both small there must be some alternative matrix representation of $\sum_a A_a \otimes B_a^T$ which has small norm, and this is sufficient. We leave it as an open problem whether this can be done.

We have now all ingredients to prove Theorem 4.4.

Proof of Theorem 4.4. Let $k = (\log n)^C$ for some absolute constant $C \geqslant 1$. Let $E = (-1)^k P(P\bar{R}_{\Omega_k})\cdots(P\bar{R}_{\Omega_1})[X]$. By Lemma 4.12 there exists Y with $(Y)_\Omega = Y$ and $P[Y] = E$ such that $\|Y\|_F \leqslant O(1)\|E\|$. We let $T = T^{(k)} + Y$. This tensor satisfies the desired linear constraints $(T)_\Omega = T$ and $P[T] = X$. Since E has the form of the matrices in Theorem 4.17, the bound in Theorem 4.17 implies $\|E\|_F \leqslant 2^{-k} \cdot n^{10} \leqslant n^{-C+10}$. (Here, we use that the norm in the conclusion of Theorem 4.17 is within a factor of n^{10} of the Frobenius norm.)

We are to prove that the following matrix has spectral norm bounded by 0.01,

$$\sum_{a=1}^n (T)_a \otimes (T)_a^T - \sum_{a=1}^n X_a \otimes X_a^T.$$

We expand the sum according to the definition of $T^{(\ell)}$ in Eq. (4.21). Then, most terms that appear in the expansion have the form as in Theorem 4.17. Since those terms decrease geometrically, we can bound their contribution by 0.001 with probability $1 - n^{-\omega(1)}$. The terms that involve the error correction Y is smaller than 0.001 because Y has polynomially small norm $\|Y\|_F \leqslant n^{-C+10}$. The only remaining terms are cross terms between X and a tensor of the form as in Theorem 4.17. We can bound the total contribution of these terms also bounded by at most 0.001 using Theorem B.27. □

5 Matrix norm bound techniques

In this section, we describe the techniques that we will use to prove probabilistic norm bounds on matrices of the form $Y = \sum_a (\bar{R}_\Omega A)_a \otimes (\bar{R}_\Omega A)_a^T$. We will prove these norm bounds using the trace moment method, which obtains probabilistic bounds on the norm of a matrix Y from bounds on the expected value of $tr((YY^T)^q)$ for sufficiently large q.

This will require analyzing $tr((YY^T)^q)$, which will take the form of a sum of products, where the terms in the product are either entries of A or terms of the form $\bar{R}_\Omega(a, b, c)$ where $\bar{R}_\Omega(a, b, c) = \frac{n^3}{m} - 1$ if $(a, b, c) \in \Omega$ and -1 otherwise. To analyze $tr((YY^T)^q)$, we will group products together which have the same expected behavior on the $\bar{R}_\Omega(a, b, c)$ terms, forming smaller sums of products. For each of these sums, we can then use the same bound on the expected behavior of the $\bar{R}_\Omega(a, b, c)$ terms for each product in the sum. This allows us to move this bound outside of the sum, leaving us with a sum of products of entries of A. We will then bound the value of these sums by carefully choosing the order in which we sum over the indices.

In the reainder of this section and in the next two sections, we allow for our tensors to have asymmetric dimensions. We account for this with the following definitions.

Definition 5.1. We define n_1 to the dimension of the u vectors, n_2 to be the dimension of the v vectors, and n_3 to be the dimension of the w vectors. We define $n_{max} = \max\{n_1, n_2, n_3\}$

5.1 The trace moment method

We use the trace moment method through the following proposition and corollary.

Proposition 5.2. *For any random matrix Y, for any integer $q \geq 1$ and any $\varepsilon > 0$,*

$$Pr\left[||Y|| > \sqrt[2q]{\frac{E\left[tr((YY^T)^q)\right]}{\varepsilon}}\right] < \varepsilon$$

Proof. By Markov's inequality, for all integers $q \geq 1$ and all $\varepsilon > 0$

$$Pr\left[tr((YY^T)^q) > \frac{E\left[tr((YY^T)^q)\right]}{\varepsilon}\right] < \varepsilon$$

The result now follows from the observation that if $||Y|| > \sqrt[2q]{\frac{E\left[tr((YY^T)^q)\right]}{\varepsilon}}$ then $tr((YY^T)^q) > \frac{E\left[tr((YY^T)^q)\right]}{\varepsilon}$. $\square$

Corollary 5.3. *For a given $p \geq 1$, $r \geq 0$, $n > 0$, and $B > 0$, for a random matrix Y, if $E\left[tr\left((YY^T)^q\right)\right] \leq (q^p B)^{2q} n^r$ for all integers $q > 1$ then for all $\beta > 0$,*

$$Pr\left[||Y|| > Be^p\left(\frac{(r+\beta)}{2p}\ln n + 1\right)^p\right] < n^{-\beta}$$

Proof. We take $\varepsilon = n^{-\beta}$ and we choose q to minimize $\sqrt[2q]{\frac{(q^p B)^{2q}n^r}{\varepsilon}} = Bq^p n^{\frac{r+\beta}{2q}}$. Setting the derivative of this expression to 0 we obtain that $(\frac{p}{q} - \frac{r+\beta}{2q^2}\ln n)Bq^p n^{\frac{r+\beta}{2q}} = 0$, so we want $q = \frac{r+\beta}{2p}\ln n$. However, q must be an integer, so we instead take $q = \lceil\frac{r+\beta}{2p}\ln n\rceil$. With this q, we have that

$$Bq^p n^{\frac{r+\beta}{2q}} \leq B\left(\frac{r+\beta}{2p}\ln n + 1\right)^p n^{\frac{p}{\ln n}} = Be^p\left(\frac{(r+\beta)}{2p}\ln n + 1\right)^p$$

19

Applying Proposition 5.2 with q, we obtain that

$$Pr\left[||Y|| > \sqrt[2q]{\frac{E\left[tr((YY^T)^q)\right]}{\varepsilon}}\right] \leqslant Pr\left[||Y|| > Be^p\left(\frac{(r+\beta)}{2p}\ln n + 1\right)^p\right] < n^{-\beta}$$

$\square$

5.2 Partitioning by intersection pattern

As discussed at the beginning of the section, $E\left[tr((YY^T)^q)\right]$ will be a sum of products, where part of these products will be of the form $\prod_{i=1}^{2q'} \bar{R}_\Omega(a_i, b_i, c_i)$. Here, q' may or may not be equal to q, in fact we will often have $q' = 2q$ because each Y will contribute two terms of the form $\bar{R}_\Omega(a, b, c)$ to the product. To handle this part of the product, we partition the terms of our sum based on the intersection pattern of which triples (a_i, b_i, c_i) are equal to each other. Fixing an intersection pattern determines the expected value of $\prod_{i=1}^{2q'} \bar{R}_\Omega(a_i, b_i, c_i)$.

Definition 5.4. We define an intersection pattern to be a set of equalities and inequalities satisfying the following conditions

1. All of the equalities and inequalities are of the form $(a_{i_1}, b_{i_1}, c_{i_1}) = (a_{i_2}, b_{i_2}, c_{i_2})$ or $(a_{i_1}, b_{i_1}, c_{i_1}) \neq (a_{i_2}, b_{i_2}, c_{i_2})$, respectively.

2. For every i_1, i_2, either $(a_{i_1}, b_{i_1}, c_{i_1}) = (a_{i_2}, b_{i_2}, c_{i_2})$ is in the intersections pattern or $(a_{i_1}, b_{i_1}, c_{i_1}) \neq (a_{i_2}, b_{i_2}, c_{i_2})$ is in the intersection pattern

3. All of the equalities and inequalities are consistent with each other, i.e. there exist values of $(a_1, b_1, c_1), \cdots, (a_{2q}, b_{2q}, c_{2q})$ satisfying all of the equalities and inequalities in the intersection pattern.

Proposition 5.5. *For a given* (a, b, c),

1. $E\left[\bar{R}_\Omega(a, b, c)\right] = 0$

2. *For all* $k > 1$, $E\left[\left(\bar{R}_\Omega(a, b, c)\right)^k\right] \leqslant \left(\frac{n_1 n_2 n_3}{m}\right)^{k-1}$

Corollary 5.6. *For a given intersection pattern, if there is any triple* (a, b, c) *which appears exactly once,* $E\left[\prod_{i=1}^{2q'} \bar{R}_\Omega(a_i, b_i, c_i)\right] = 0$. *Otherwise, letting z be the number of distinct triples,* $E\left[\prod_{i=1}^{2q'} \bar{R}_\Omega(a_i, b_i, c_i)\right] \leqslant \left(\frac{n_1 n_2 n_3}{m}\right)^{2q'-z}$

Proof. for a given intersection pattern, let $(a_{i_1}, b_{i_1}, c_{i_1}), \cdots, (a_{i_z}, b_{i_z}, c_{i_z})$ be the distinct triples and let c_j be the number of times the triple $(a_{i_j}, b_{i_j}, c_{i_j})$ appears. We have that

$$E\left[\prod_{i=1}^{2q'} \bar{R}_\Omega(a_i, b_i, c_i)\right] = \prod_{j=1}^{z} E\left[\left(\bar{R}_\Omega(a_{i_j}, b_{i_j}, c_{i_j})\right)^{c_j}\right]$$

20

If $c_j = 1$ for any j then this expression is 0. Otherwise,

$$\prod_{j=1}^{z} E\left[\left(\bar{R}_\Omega(a_{i_j}, b_{i_j}, c_{i_j})\right)^{c_j}\right] \leqslant \prod_{j=1}^{z}\left(\frac{n_1 n_2 n_3}{m}\right)^{c_j - 1} = \left(\frac{n_1 n_2 n_3}{m}\right)^{\left(\sum_{j=1}^{z} c_j\right) - z} = \left(\frac{n_1 n_2 n_3}{m}\right)^{2q' - z}$$

$\square$

5.3 Bounding sums of products of tensor entries

In this subsection, we describe how to bound the sum of products of tensor entries we obtain for a given intersection pattern after moving our bound on the expected value of the $\bar{R}_\Omega(a, b, c)$ terms outside the sum. We represent such a product with a hypergraph as follows.

Definition 5.7. Given a set of distinct indices and a set of tensor entries on those indices, let H be the hypergraph with one vertex for each distinct index and one hyperedge for each tensor entry, where the hyperedge consists of all indices contained in the tensor entry. If the tenor entry appears to the pth power, we take this hyperedge with multiplicity p.

With this definition in mind, we will first preprocess our products.

1. We will preprocess the tensor entries so that every entry appears to an even power using the inequality $|ab| \leqslant \frac{1}{2}(a^2 + b^2)$. This has the effect of taking two hyperedges of our choice in H and replacing them with one doubled hyperedge or the other (we have to consider both possibilities). Note that this step makes all of our terms positive and can only increase their magnitude, so the result will be an upper bound on our actual sum.

2. We will add the missing terms to our sum so that for we sum over every possibility for the distinct indices (even the possibilities which make several of these indices equal and would put us in a different intersection pattern). Note that this can only increase our sum.

Remark 5.8. It is important that we first bound the expected value of the $\bar{R}_\Omega(a, b, c)$ terms and move this bound outside of our sum before adding the missing terms to the sum.

After preprocessing our products, our strategy will be as follows. We will sum over the indices, removing the corresponding vertices from H. As we do this, we will apply appropriate bounds on squared tensor entries, removing the corresponding doubled hyperedge from H. To obtain these bounds, we observe that we can bound the average square of our tensor entries in terms of the number of indices we are averaging over.

Definition 5.9. We say that an order 3 tensor A of dimensions $n_1 \times n_2 \times n_3$ is (B, r, μ)-bounded if the following bounds are true

1. $\max_{a,b,c}\{A_{abc}^2\} \leqslant Br$

2. $\max \{ \max_{b,c} \{ \frac{1}{n_1} \sum_a A^2_{abc} \}, \max_{a,c} \{ \frac{1}{n_2} \sum_b A^2_{abc} \}, \max_{a,b} \{ \frac{1}{n_3} \sum_c A^2_{abc} \} \} \leqslant \frac{B}{\mu}$

3. $\max \{ \max_c \{ \frac{1}{n_1 n_2} \sum_{a,b} A^2_{abc} \}, \max_b \{ \frac{1}{n_1 n_3} \sum_{a,c} A^2_{abc} \}, \max_a \{ \frac{1}{n_2 n_3} \sum_{b,c} A^2_{abc} \} \} \leqslant \frac{B}{\mu^2}$

4. $\frac{1}{n_1 n_2 n_3} \sum_{a,b,c} A^2_{abc} \leqslant \frac{B}{\mu^3}$

More generally, we say that a tensor A is (B, r, μ)-bounded if the following is true

1. The maximum value of an entry of A squared is at most Br

2. Every index which we average over decreases our upper bound by a factor of μ

3. If we are averaging over at least one index then we can delete the factor of r in our bound.

Since r and μ will always be the same, we write B-bounded rather than (B, r, μ)-bounded

To give a sense of why these are the correct type of bounds to use, we now show that X is $\left(\frac{r\mu^3}{n_1 n_2 n_3} \right)$-bounded. In Section 7, we will use an iterative argument to show that with high probability, similar bounds hold for all of the tensors A we will be considering.

Proposition 5.10. X is $\left(\frac{r\mu^3}{n_1 n_2 n_3} \right)$-bounded

Proof. Recall that $X = \sum_{i=1}^r u_i \otimes v_i \otimes w_i$ where the vectors $\{u_i\}$ are orthogonal, the vectors $\{v_i\}$ are orthogonal, and the vectors $\{w_i\}$ are orthogonal. Also recall that for all i, a, b, c, $u_{ia}^2 \leqslant \frac{\mu}{n_1}$, $v_{ib}^2 \leqslant \frac{\mu}{n_2}$, and $w_{ic}^2 \leqslant \frac{\mu}{n_3}$. We now have the following bounds:

1.
$$\max_{a,b,c} \{ X^2_{abc} \} = \max_{a,b,c} \left\{ \sum_{i=1}^r \sum_{i'=1}^r u_{ia} v_{ib} w_{ic} u_{i'a} v_{i'b} w_{i'c} \right\} \leqslant \frac{r^2 \mu^3}{n_1 n_2 n_3}$$

2.
$$\max_{b,c} \left\{ \frac{1}{n_1} \sum_a X^2_{abc} \right\} = \frac{1}{n_1} \max_{b,c} \left\{ \sum_a \sum_{i=1}^r \sum_{i'=1}^r u_{ia} v_{ib} w_{ic} u_{i'a} v_{i'b} w_{i'c} \right\}$$
$$= \frac{1}{n_1} \max_{b,c} \left\{ \sum_{i=1}^r \sum_{i'=1}^r \left(\sum_a u_{ia} u_{i'a} \right) v_{ib} w_{ic} v_{i'b} w_{i'c} \right\}$$
$$= \frac{1}{n_1} \max_{b,c} \left\{ \sum_{i=1}^r v_{ib}^2 w_{ic}^2 \right\}$$
$$\leqslant \frac{r\mu^2}{n_1 n_2 n_3}$$

The other bounds where we sum over one index follow by symmetrical arguments.

22

3.

$$\max_c \left\{ \frac{1}{n_1 n_2} \sum_{a,b} X_{abc}^2 \right\} = \frac{1}{n_1 n_2} \max_c \left\{ \sum_{a,b} \sum_{i=1}^r \sum_{i'=1}^r u_{ia} v_{ib} w_{ic} u_{i'a} v_{i'b} w_{i'c} \right\}$$

$$= \frac{1}{n_1 n_2} \max_c \left\{ \sum_{i=1}^r \sum_{i'=1}^r \left(\sum_a u_{ia} u_{i'a} \right) \left(\sum_b v_{ib} v_{i'b} \right) w_{ic} w_{i'c} \right\}$$

$$= \frac{1}{n_1 n_2} \max_c \left\{ \sum_{i=1}^r w_{ic}^2 \right\}$$

$$\leqslant \frac{r\mu}{n_1 n_2 n_3}$$

The other bounds where we sum over two indices follow by symmetrical arguments.

4.

$$\frac{1}{n_1 n_2 n_3} \sum_{a,b,c} X_{abc}^2 = \frac{1}{n_1 n_2 n_3} \sum_{a,b,c} \sum_{i=1}^r \sum_{i'=1}^r u_{ia} v_{ib} w_{ic} u_{i'a} v_{i'b} w_{i'c}$$

$$= \frac{1}{n_1 n_2 n_3} \sum_{i=1}^r \sum_{i'=1}^r \left(\sum_a u_{ia} u_{i'a} \right) \left(\sum_b v_{ib} v_{i'b} \right) \left(\sum_c w_{ic} w_{i'c} \right)$$

$$= \frac{1}{n_1 n_2 n_3} \sum_{i=1}^r 1$$

$$= \frac{r}{n_1 n_2 n_3}$$

$\square$

With these kinds of bounds in mind, we bound sums of products of tensor entries as follows. We note that we can always apply the entrywise bound for a squared tensor entry. However, to apply any of the other bounds, we must be able to sum over an index or indices where the only term in our product which depends on this index or indices is the squared tensor entry. This can be described in terms of the hypergraph H as follows.

Definition 5.11. Given a hyperedge e in H, define $b(e)$ to the the minimal B such that the tensor entry corresponding to e is B-bounded.

Definition 5.12. We say that a vertex is free in H if it contained in only one hyperedge and this hyperedge appears with multiplicity two.

We can apply our bounds in the following ways.

1. We can always choose a hyperedge e of H, use the entrywise bound of $rb(e)$ on the corresponding squared tensor entry (note the extra factor of r), and reduce the multiplicity of e by two.

23

2. If there is a free vertex incident with a doubled hyperedge e in H, we can sum over all free vertices which are incident with e using the corresponding bound then delete these vertices and the doubled hyperedge e from H. When we do this, we obtain a factor of

$$b(e) \left(\frac{n_1}{\mu}\right)^{\text{\# of deleted a vertices}} \left(\frac{n_2}{\mu}\right)^{\text{\# of deleted b vertices}} \left(\frac{n_3}{\mu}\right)^{\text{\# of deleted c vertices}}$$

The factors of n_1, n_2, n_3 appear because we are summing over these indices and the factors of $\frac{1}{\mu}$ appear because each index we sum over reduces the bound on the average value by a factor of μ.

If we apply these bounds repeatedly until there are no tensor entries/hyperedges left to bound, our final bound on a single sum of products of tensor entries will be

$$\left(\prod_{e \in H} \sqrt{b(e)}\right) \left(\frac{n_1}{\mu}\right)^{\text{\# of } a \text{ indices}} \left(\frac{n_2}{\mu}\right)^{\text{\# of } b \text{ indices}} \left(\frac{n_3}{\mu}\right)^{\text{\# of } c \text{ indices}} r^{\text{\# of entrywise bounds used}}$$

To prove our final upper bound, we will argue that we can always apply these bounds in such a way that the number of times we need to use an entrywise bound is sufficiently small.

5.4 Counting intersection patterns

There will be one more factor in our final bound. This factor will come from the number of possible intersection patterns with a given number z of distinct triples (a, b, c).

Lemma 5.13. *The total number of intersection patterns on $2q'$ triples with z distinct triples (a, b, c) such that every triple (a, b, c) has multiplicity at least two is at most $\binom{2q'}{z} z^{2q'-z} \leqslant 2^{2q'} q'^{2q'-z}$*

Proof. To determine which triples (a, b, c) are equal to each other, it is sufficient to decide which triples are distinct from all previous triples (there are $\binom{2q'}{z}$ choices for this) and for the remaining $2q' - z$ triples, which of the z distinct triples they are equal to (there are $z^{2q'-z}$ choices for this). $\square$

6 Trace Power Calculation for $\bar{R}_\Omega A \otimes (\bar{R}_\Omega A)^T$

In this section, we implement the techniques described in Section 5 to probabilistically bound $\|\bar{R}_\Omega A \otimes (\bar{R}_\Omega A)^T\|$. In particular, we prove the following theorem.

Theorem 6.1. *If A is B-bounded, $C \geqslant 1$, and*

1. $m > 10000C(2+\beta)^2 n_{max} r \mu^2 \ln n_{max}$

2. $m > 10000C(2+\beta)^2 r \sqrt{n_1} \max\{n_2, n_3\} \mu^{\frac{3}{2}} \ln n_{max} \geqslant 10000C(2+\beta)^2 r \sqrt{n_1 n_2 n_3} \mu^{\frac{3}{2}} \ln n_{max}$

24

3. $\mu r \leqslant \min \{n_1, n_2, n_3\}$

then defining $Y = \bar{R}_\Omega A \otimes (\bar{R}_\Omega A)^T$,

$$Pr\left[||Y|| > \frac{Bn_1 n_2 n_3}{Cr\mu^3}\right] < 4n^{-(\beta+1)}$$

Corollary 6.2. *If $C \geqslant 1$ and*

1. $m > 10000C(2 + \beta)^2 n_{max} r \mu^2 \ln n_{max}$

2. $m > 10000C(2 + \beta)^2 r \sqrt{n_1} \max \{n_2, n_3\} \mu^{\frac{3}{2}} \ln n_{max} \geqslant 10000C(2 + \beta)^2 r \sqrt{n_1 n_2 n_3} \mu^{\frac{3}{2}} \ln n_{max}$

3. $\mu r \leqslant \min \{n_1, n_2, n_3\}$

then

$$Pr\left[||\bar{R}_\Omega X \otimes (\bar{R}_\Omega X)^T|| > \frac{1}{C}\right] < 4n^{-(\beta+1)}$$

Proof. This follows immediately from Theorem 6.1 and the fact that X is $\left(\frac{r\mu^3}{n_1 n_2 n_3}\right)$-bounded. □

To prove Theorem 6.1, we break up Y into four parts and then prove probabilistic norm bounds for each part.

Definition 6.3.

1. Define $(Y_1)_{bcb'c'} = Y_{bcb'c'}$ if $b = b'$, $c = c'$ and 0 otherwise.

2. Define $(Y_2)_{bcb'c'} = Y_{bcb'c'}$ if $b = b'$, $c \neq c'$ and 0 otherwise.

3. Define $(Y_3)_{bcb'c'} = Y_{bcb'c'}$ if $b \neq b'$, $c = c'$ and 0 otherwise.

4. Define $(Y_4)_{bcb'c'} = Y_{bcb'c'}$ if $b \neq b'$, $c \neq c'$ and 0 otherwise.

6.1 Structure of $tr((Y_j Y_j^T)^q)$

We have that $Y_{bcb'c'} = \sum_a \bar{R}_\Omega(a, b, c') \bar{R}_\Omega(a, b', c) A_{abc'} A_{ab'c}$. To see the structure of $(YY^T)^q$, we now compute YY^T.

$$(YY^T)_{b_1 c_1 b_2 c_2} =$$
$$\sum_{a_1, a_2, b', c'} \bar{R}_\Omega(a_1, b_1, c') \bar{R}_\Omega(a_1, b', c_1) \bar{R}_\Omega(a_2, b_2, c') \bar{R}_\Omega(a_2, b', c_2) A_{a_1 b_1 c'} A_{a_1 b' c_1} A_{a_2 b_2 c'} A_{a_2 b' c_2}$$

The $\bar{R}_\Omega$ terms will not be part of our hypergraph H (as their expected behavior is determined by the intersection pattern). We can view the first two terms $A_{a_1 b_1 c'}$ and $A_{a_1 b' c_1}$ as an hourglass with upper triangle (b_1, a_1, c') and lower triangle (c_1, a_1, b') (where the vertices in each triangle are listed from left to right). Similarly, we can view the last two terms $A_{a_2 b_2 c'}$

25

and $A_{a_2 b' c_2}$ as an hourglass with upper triangle (c', a_2, b_2) and lower triangle (b', a_2, c_2). Thus, the hypergraph H corresponding to $tr((Y_j Y_j^T)^q)$ will be $2q$ hourglasses glued together where the top vertices of the hourglass alternate between b and c' indices, the bottom vertices of the hourglass alternate between c and b' indices, and the middle vertices of the hourglass are the a indices.

Remark 6.4. While there is no real difference between the b and b' indices and between the c and c' indices, we will keep track of this to make it easier to see the structure of H.

As described in Section 5, we split up $E\left[tr((Y_j Y_j^T)^q)\right]$ based on the intersection pattern of which of the $4q$ triples of the form (a, b, c') or (a, b', c) are equal to each other. We only need to consider patterns where each triple and thus each hyperedge appears at least twice, as otherwise the terms in the sum will have expected value 0. In all cases, letting z be the number of distinct triples in a given intersection pattern, by Corollary 5.6 our bound on the expected value of the $\bar{R}_\Omega$ terms will be $\left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z}$

6.2 Bounds on $\|Y_1\|$

Consider $E\left[tr((Y_1 Y_1^T)^q)\right]$. The constraints that $b' = b$ and $c' = c$ in every Y force all of the b and b' indices to be equal and all of the c and c' indices to be equal, so our hypergraph H consists of a single vertex b, a single vertex c, and two copies of the hyperedge (a_i, b, c) for each $i \in [1, 2q]$. For all intersection patterns, the number of distinct triples z is equal to the number of distinct a indices, which can be anywhere from 1 to $2q$.

We apply our bounds on H as follows.

1. In our preprocessing step, when there are two hyperedges e_1 and e_2 which appear with odd multiplicity, we double one of these hyperedges or the other. Thus, we can assume that all hyperedges appear with even multiplicity.

2. We will apply an entrywise bound $2q - z$ times on hyperedges of multiplicity $\geqslant 4$, reducing the multiplicity by 2 each time.

3. After applying these entrywise bounds, all of the distinct a vertices will be free and we can sum up over these indices one by one.

Recall that the bound from the R_Ω terms is $\left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z}$ and our bound for the other terms is

$$\left(\prod_{e \in H} \sqrt{b(e)}\right) \left(\frac{n_1}{\mu}\right)^{\overset{\text{\# of } a \text{ entries}}{}} \left(\frac{n_2}{\mu}\right)^{\overset{\text{\# of } b \text{ entries}}{}} \left(\frac{n_3}{\mu}\right)^{\overset{\text{\# of } c \text{ entries}}{}} r^{\overset{\text{\# of entrywise bounds used}}{}}$$

where $b(e) = B$ for all our hyperedges. Summing over all $z \in [1, 2q]$ and all intersection patterns using Lemma 5.13, our final bound is

$$2q \cdot 2^{4q} \max_{z \in [1, 2q]} \left\{ (2q)^{4q-z} \left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z} B^{2q} \left(\frac{n_1}{\mu}\right)^z \left(\frac{n_2}{\mu}\right) \left(\frac{n_3}{\mu}\right) r^{2q-z} \right\}$$

26

The inner expression will either be maximized at $z = 2q$ or $z = 1$ and we will always take q to be between $\frac{\ln n_{max}}{2}$ and $\frac{n_{max}}{2}$, so our final bound on $E\left[tr((Y_1 Y_1^T)^q)\right]$ is at most

$$(4q)^{4q} \max \left\{ n_{max} \left(\frac{n_1^2 n_2 n_3 B}{m \mu \ln n_{max}} \right)^{2q} \left(\frac{n_2}{\mu} \right) \left(\frac{n_3}{\mu} \right), \left(\frac{n_1^2 n_2^2 n_3^2 r B}{m^2} \right)^{2q} \frac{m}{r \mu^3} \right\}$$

Since $m > 10000C(2 + \beta)^2 n_{max} r \mu^2 \ln n_{max}$ and $m > 10000C(2 + \beta)^2 r \sqrt{n_1 n_2 n_3} \mu^{\frac{3}{2}} \ln n_{max}$, we have that

$$E\left[tr((Y_1 Y_1^T)^q)\right] < (16q^2)^{2q} \left(\frac{n_1 n_2 n_3 B}{10000C(2 + \beta)^2 r \mu^3 (\ln n_{max})^2} \right)^{2q} n_{max}^3$$

(note that $m < n_{max}^3$ as otherwise the tensor completion problem is trivial). We now recall Corollary 5.3, which says that for a given $p \geqslant 1$, $r \geqslant 0$, $n > 0$, and $B > 0$, for a random matrix Y, if $E\left[tr\left((YY^T)^q\right)\right] \leqslant (q^p B)^{2q} n^r$ for all integers $q > 1$ then for all $\beta > 0$,

$$Pr\left[||Y|| > Be^p \left(\frac{(r + \beta)}{2p} \ln n + 1 \right)^p \right] < n^{-\beta}$$

Using Corollary 5.3 with the appropriate parameters, we can show that for all $\beta > 0$,

$$P\left[||Y_1|| > \frac{16e^2 B n_1 n_2 n_3}{10000 r \mu^3} \right] < n_{max}^{-(\beta+1)}$$

6.3 Bounds on $||Y_2||$ and $||Y_3||$

Consider $E\left[tr((Y_2 Y_2^T)^q)\right]$. The constraint that $b' = b$ in every Y forces all of the b and b' indices to be equal, so our hypergraph H consists of a single vertex b and $4q$ total hyperedges of the form (a, b, c) or (a, b, c'). Ignoring the b vertex (which is part of all the hyperedges), the (a, c) and (a, c') edges form a single connected component. We only need to consider intersection patterns where each triple (a, b, c) or (a, b, c') (and thus each edge (a, c) or (a, c')) appears with multiplicity at least two. For a given intersection pattern, let z be the number of distinct edges.

We apply our bounds on H as follows.

1. In our preprocessing step, when there are two edges e_1 and e_2 which appear with odd multiplicity, we double one of these edges or the other. Thus, we can assume that all edges appear with even multiplicity.

2. We will apply an entrywise bound $2q - z$ times on edges of multiplicity $\geqslant 4$, reducing the multiplicity by 2 each time.

3. After applying these entrywise bounds, all of our edges will have multiplicity 2. We now sum over a free a, c, or c' vertex in H whenever such a vertex exists. Otherwise, there must be a cycle, in which case we use the entrywise bound on one edge of the cycle and delete it.

27

Definition 6.5. Let x be the number of times we delete an edge in a cycle using the entrywise bound.

Lemma 6.6. *The total number of vertices in H (excluding b) is $z + 1 - x$*

Proof. Observe that neither deleting a free vertex nor deleting an edge in a cycle can disconnect H. Also, except for the final edge where both of its vertices will be free, every edge which has a free vertex has exactly one free vertex. Thus, we delete an edge in a cycle x times, removing 0 vertices each time, we delete an edge with one free vertex $z - x - 1$ times, removing 1 vertex each time, and we delete the final edge once, removing the final two vertices. This adds up to $z + 1 - x$ vertices in H. □

Recall that the bound from the R_Ω terms is $\left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z}$ and our bound for the other terms is

$$\left(\prod_{e \in H} \sqrt{b(e)}\right) \left(\frac{n_1}{\mu}\right)^{\text{\# of } a \text{ entries}} \left(\frac{n_2}{\mu}\right)^{\text{\# of } b \text{ entries}} \left(\frac{n_3}{\mu}\right)^{\text{\# of } c \text{ or } c' \text{ entries}} r^{\text{\# of entrywise bounds used}}$$

where $b(e) = B$ for all our hyperedges. Summing over all $z \in [1, 2q]$ and all intersection patterns using Lemma 5.13, our final bound is

$$2q \cdot 2^{4q} \max_{z \in [1,2q], x \in [0,z-1]} \left\{ (2q)^{4q-z} \left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z} B^{2q} \left(\frac{n_{max}}{\mu}\right)^{z-1-x} \left(\frac{n_1 n_2 n_3}{\mu^3}\right) r^{2q-z+x} \right\}$$

Since $\mu r \leqslant n_{max}$, the inner expression will either be maximized when $z = 2q$ and $x = 0$ or when $z = 1$ and $x = 0$. Again, we will always take q to be between $\frac{\ln n_{max}}{2}$ and $\frac{n_{max}}{2}$, so our final bound on $E\left[tr((Y_2 Y_2^T)^q)\right]$ is at most

$$(4q)^{(4q)} \max \left\{ \left(\frac{n_1 n_2 n_3 n_{max} B}{m \mu \ln n_{max}}\right)^{2q} \left(\frac{n_1 n_2 n_3}{\mu^2}\right), \left(\frac{n_1^2 n_2^2 n_3^2 r B}{m^2}\right)^{2q} \frac{m}{r \mu^3} \right\}$$

Since $m > 10000 C(2 + \beta)^2 n_{max} r \mu^2 \ln n_{max}$ and $m > 10000 C(2 + \beta)^2 r \sqrt{n_1 n_2 n_3} \mu^{\frac{3}{2}} \ln n_{max}$, we have that

$$E\left[tr((Y_2 Y_2^T)^q)\right] < (16q^2)^{2q} \left(\frac{n_1 n_2 n_3 B}{10000 C(2 + \beta)^2 r \mu^3 (\ln n_{max})^2}\right)^{2q} n_{max}^3$$

Using Corollary 5.3 with the appropriate parameters (in fact the same ones as before), we can show that for all $\beta > 0$,

$$P\left[\|Y_2\| > \frac{16 e^2 B n_1 n_2 n_3}{10000 r \mu^3}\right] < n_{max}^{-(\beta+1)}$$

By a symmetrical argument, we can obtain the same probabilistic bound on $\|Y_3\|$.

6.4 Bounds on $\|Y_4\|$

Consider $E\left[tr((Y_4 Y_4^T)^q)\right]$. Our hypergraph H consists of $2q$ hyperedges of the form (b, a, c') or (c', a, b) from the top triangles of the hourglasses and $2q$ hyperedges of the form (c, a, b') or (b', a, c) from the bottom triangles of the hourglasses. We only need to consider intersection patterns where each triple (and thus each hyperedge) appears with multiplicity at least two. For a given intersection pattern, let z be the number of distinct hyperedges.

Ignoring the a vertices for now, we can think of H as a graph on the b, b', c, and c' vertices. Note that the (b, c') and (c', b) edges are part of a single connected component and the (c, b') and (b', c) edges are part of a single connected component (these connected components may or may not be the same).

We apply our bounds on H as follows.

1. In our preprocessing step, when there are two hyperedges e_1 and e_2 which appear with odd multiplicity, we double one of these hyperedges or the other. Thus, we can assume that all hyperedges appear with even multiplicity.

2. We will apply an entrywise bound $2q - z$ times on hyperedges of multiplicity $\geqslant 4$, reducing the multiplicity by 2 each time.

3. After applying these entrywise bounds, all of our hyperedges will have multiplicity 2. We now sum over a free b,b',c, or c' vertex in H whenever such a vertex exists. Otherwise, there must be a cycle on the (b, c') and (b', c) parts of the hyperedges, in which case we use the entrywise bound on one hyperedge of the cycle and delete it.

Definition 6.7. Let x be the number of times we delete a hyperedge in a cycle using the entrywise bound.

Lemma 6.8. *Let k be the number of connected components of H. The total number of b,b',c, and c' vertices in H is $z + k - x \leqslant z + 2 - x$*

Proof. The proof is similar to the proof of Lemma 6.6. Observe that neither deleting a free vertex nor deleting an edge in a cycle can disconnect a connected component of H. Also, except for the final edge of a connected component where both of its vertices will be free, every edge which has a free vertex has exactly one free vertex. Thus, we delete an edge in a cycle x times, removing 0 vertices each time, we delete an edge with one free vertex $z - x - k$ times, removing 1 vertex each time, and we delete the final edge of a connected component k times, removing the final $2k$ vertices. This adds up to $z + k - x$ vertices in H. For the inequality, recall that H has at most 2 connected components, one for the (b, c') edges and one for the (c, b') edges. $\square$

Finally, we bound the number of distinct a indices

Proposition 6.9. *The number of distinct a indices is at most $\frac{z}{2}$.*

Proof. Note that by the definition of Y_4, every a index must be part of at least two distinct hyperedges. $\square$

Recall that the bound from the R_Ω terms is $\left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z}$ and our bound for the other terms is

$$\left(\prod_{e \in H} \sqrt{b(e)}\right) \left(\frac{n_1}{\mu}\right)^{\text{\# of } a \text{ entries}} \left(\frac{n_2}{\mu}\right)^{\text{\# of } b \text{ or } b' \text{ entries}} \left(\frac{n_3}{\mu}\right)^{\text{\# of } c \text{ or } c' \text{ entries}} r^{\text{\# of entrywise bounds used}}$$

where $b(e) = B$ for all our hyperedges. Summing over all $z \in [2, 2q]$ and all intersection patterns using Lemma 5.13, our final bound is

$$2q \cdot 2^{4q} \max_{z \in [1, 2q], x \in [0, z-2]} \left\{ (2q)^{4q-z} \left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z} B^{2q} \left(\frac{n_1}{\mu}\right)^{\frac{z}{2}} \left(\frac{\max\{n_2, n_3\}}{\mu}\right)^{z-2-x} \left(\frac{n_2^2 n_3^2}{\mu^4}\right) r^{2q-z+x} \right\}$$

Since $\mu r \leqslant \min\{n_1, n_2, n_3\}$, the inner expression will either be maximized when $z = 2q$ and $x = 0$ or when $z = 2$ and $x = 0$. Again, we will always take q to be between $\frac{\ln n_{max}}{2}$ and $\frac{n_{max}}{2}$, so our final bound on $E\left[tr((Y_2 Y_2^T)^q)\right]$ is at most

$$(4q)^{(4q)} \max \left\{ \left(\frac{n_1 n_2 n_3 \sqrt{n_1} \max\{n_2, n_3\} B}{m \mu^{\frac{3}{2}} \ln n_{max}}\right)^{2q} \left(\frac{n_{max}^3}{\mu^2}\right), \left(\frac{n_1^2 n_2^2 n_3^2 r B}{m^2}\right)^{2q} \frac{m^2}{r^2 \mu^5 n_1} \right\}$$

Since $m > 10000C(2 + \beta)^2 r \sqrt{n_1} \max\{n_2, n_3\} \mu^{\frac{3}{2}} \ln n_{max}$, we have that

$$E\left[tr((Y_4 Y_4^T)^q)\right] < (16q^2)^{2q} \left(\frac{n_1 n_2 n_3 B}{10000C(2 + \beta)^2 r \mu^3 (\ln n_{max})^2}\right)^{2q} n_{max}^6$$

Using Corollary 5.3 with the appropriate parameters, we can show that for all $\beta > 0$,

$$P\left[\|Y_4\| > \frac{16e^2 B n_1 n_2 n_3}{10000 r \mu^3}\right] < n_{max}^{-(\beta+1)}$$

Putting our four bounds together with a union bound, for all $\beta > 0$,

$$P\left[\|Y\| > \frac{B n_1 n_2 n_3}{r \mu^3}\right] \leqslant P\left[\|Y\| > \frac{64e^2 B n_1 n_2 n_3}{10000 r \mu^3}\right] < 4n_{max}^{-(\beta+1)}$$

as needed.

7 Iterative tensor bounds

In this section, we show that with high probability, applying the operator $P\bar{R}_\Omega$ to an order 3 tensor A improves our bounds on it, where we are assuming that Ω is chosen independently of A.

Theorem 7.1. *If A is a B-bounded tensor, $C \geqslant 1$, $\beta > 0$, and*

1. $m > 10000C(2 + \beta)^2 n_{max} r \mu^2 \ln n_{max}$

30

2. $m > 10000C(2 + \beta)^2 r \sqrt{n_1} \max\{n_2, n_3\} \mu^{\frac{3}{2}} \ln n_{max} \geqslant 10000C(2 + \beta)^2 r \sqrt{n_1 n_2 n_3} \mu^{\frac{3}{2}} \ln n_{max}$

3. $\mu r \leqslant \min\{n_1, n_2, n_3\}$

then

$$Pr\left[P\bar{R}_\Omega A \text{ is not } \left(\frac{B}{C}\right)\text{-bounded}\right] < 100 n_{max}^{-(\beta+1)}$$

Proof. We first consider how P acts on a tensor

Definition 7.2.

1. Define P^{UV} to be the projection onto $span\{u_i \otimes v_i \otimes w : i \in [1, r]\}$.

2. Define P^{UW} to be the projection onto $span\{u_i \otimes v \otimes w_i : i \in [1, r]\}$.

3. Define P^{VW} to be the projection onto $span\{u \otimes v_i \otimes w_i : i \in [1, r]\}$.

4. Define P^{UVW} to be the projection onto $span\{u_i \otimes v_i \otimes w_i : i \in [1, r]\}$.

Proposition 7.3. $P = P_{UV} + P_{UW} + P_{VW} - 2P_{UVW}$

With this in mind, we break up the tensor $W = P\bar{R}_\Omega A$ into four parts and then obtain probabilistic bounds for each part. Theorem 7.1 will then follow from the union bound and the inequality $(a + b + c - 2d)^2 \leqslant 5(a^2 + b^2 + c^2 + 2d^2)$.

Definition 7.4.

1. Define $W^{UV} = P_{UV}\bar{R}_\Omega A$.

2. Define $W^{UW} = P_{UW}\bar{R}_\Omega A$.

3. Define $W^{VW} = P_{VW}\bar{R}_\Omega A$.

4. Define $W^{UVW} = P_{UVW}\bar{R}_\Omega A$.

To analyze these parts, we reexpress $P_{UV}, P_{UW}, P_{VW}, P_{UVW}$ in terms of matrices UV, UW, VW, UVW.

Definition 7.5.

1. Define $UV_{aba'b'} = \sum_{i=1}^r u_{ia} v_{ib} u_{ia'} v_{ib'}$

2. Define $UW_{aca'c'} = \sum_{i=1}^r u_{ia} w_{ic} u_{ia'} w_{ic'}$

3. Define $VW_{bcb'c'} = \sum_{i=1}^r v_{ib} w_{ic} v_{ib'} w_{ic'}$

4. Define $UVW_{abca'b'c'} = \sum_{i=1}^r u_{ia} v_{ib} w_{ic} u_{ia'} v_{ib'} w_{ic'}$

Proposition 7.6.

1. UV is $\left(\frac{r\mu^4}{n_1^2 n_2^2}\right)$-bounded.

31

2. UW is $\left(\frac{r\mu^4}{n_1^2 n_2^3}\right)$-bounded.

3. VW is $\left(\frac{r\mu^4}{n_2^2 n_3^3}\right)$-bounded.

4. UVW is $\left(\frac{r\mu^6}{n_1^2 n_2^2 n_3^2}\right)$-bounded.

Proof. These bounds can be proved in the same way as Proposition 5.10. $\square$

Proposition 7.7.

1. $W_{abc}^{UV} = \sum_{a',b'} UV_{aba'b'}\bar{R}_\Omega(a',b',c)A_{a'b'c}$

2. $W_{abc}^{UW} = \sum_{a',c'} UV_{aca'c'}\bar{R}_\Omega(a',b,c')A_{a'bc'}$

3. $W_{abc}^{VW} = \sum_{b',c'} UV_{bcb'c'}\bar{R}_\Omega(a,b',c')A_{ab'c'}$

4. $W_{abc}^{UVW} = \sum_{a',b',c'} UV_{abca'b'c'}\bar{R}_\Omega(a',b',c')A_{a'b'c'}$

Proposition 7.8.

1. $\left(W_{abc}^{UV}\right)^2 = \sum_{a_1',b_1',a_2',b_2'} UV_{aba_1'b_1'}UV_{aba_2'b_2'}\bar{R}_\Omega(a_1',b_1',c)\bar{R}_\Omega(a_2',b_2',c)A_{a_1'b_1'c}A_{a_2'b_2'c}$

2. $\left(W_{abc}^{UW}\right)^2 = \sum_{a_1',c_1',a_2',c_2'} UW_{aca_1'c_1'}UW_{aca_2'c_2'}\bar{R}_\Omega(a_1',b,c_1')\bar{R}_\Omega(a_2',b,c_2')A_{a_1'bc_1'}A_{a_2'bc_2'}$

3. $\left(W_{abc}^{VW}\right)^2 = \sum_{b_1',c_1',b_2',c_2'} VW_{bcb_1'c_1'}VW_{bcb_2'c_2'}\bar{R}_\Omega(a,b_1',c_1')\bar{R}_\Omega(a,b_2',c_2')A_{ab_1'c_1'}A_{ab_2'c_2'}$

4.

$$\left(W_{abc}^{UVW}\right)^2 =$$
$$\sum_{a_1',b_1',c_1',a_2',b_2',c_2'} UVW_{abca_1'b_1'c_1'}UVW_{abca_2'b_2'c_2'}\bar{R}_\Omega(a_1',b_1',c_1')\bar{R}_\Omega(a_2',b_2',c_2')A_{a_1'b_1'c_1'}A_{a_2'b_2'c_2'}$$

We need to probabilistically bound the expressions $\sum_{\text{subset of } \{a,b,c\}} (W_{a,b,c}^{UV,UW,VW, \text{ or } UVW})^2$. For each expression which we need to probabilistically bound, we can obtain this bound by analyzing the expected value of its qth power using the techniques in Section 5 and then using a result similar to Corollary 5.3. We begin by probabilistically bounding $\left(W_{abc}^{UVW}\right)^2$. As the remaining bounds will all be very similar, rather than giving a full proof of the remaining bounds we will only describe the few differences and what effect they have.

Lemma 7.9. *For all a,b,c and all $\beta > 0$, if $m > 10000C(2+\beta)^2 n_{max} r\mu^2 \ln n_{max}$ then*

$$P\left[(W_{abc}^{UVW})^2 > \frac{32e^2 Br\mu}{10000Cn_{max}}\right] < n_{max}^{-(\beta+4)}$$

Proof. Similar to before, we partition our sum based on the intersection pattern of which (a'_i, b'_i, c'_i) are equal. Letting z be the number of distinct triples (a'_i, b'_i, c'_i), the contribution from the $\bar{R}_\Omega(a'_i, b'_i, c'_i)$ terms will be at most a factor of $\left(\frac{n_1 n_2 n_3}{m}\right)^{2q-z}$. Recall that for a given intersection pattern, our bound on the remaining terms is

$$
\left(\prod_{e \in H} \sqrt{b(e)}\right) \left(\frac{n_1}{\mu}\right)^{\text{\# of } a \text{ or } a' \text{ indices}} \left(\frac{n_2}{\mu}\right)^{\text{\# of } b \text{ or } b' \text{ indices}} \left(\frac{n_3}{\mu}\right)^{\text{\# of } c \text{ or } c' \text{ indices}} r^{\text{\# of entrywise bounds used}}
$$

Remark 7.10. Here we will only be summing over a', b', and c', indices, but for other expressions we will be summing over a, b, and c indices as well.

In our hypergraph H, we will have hyperedges (a'_i, b'_i, c'_i) corresponding to the tensor entries $A_{a'_i, b'_i, c'_i}$ and we will have hyperedges $(a, b, c, a'_i, b'_i, c'_i)$ corresponding to the matrix entries $UVW_{a,b,c,a'_i,b'_i,c'_i}$. We have that

$$
\prod_{e \in H} \sqrt{b(e)} = B^q \left(\frac{r \mu^6}{n_1^2 n_2^2 n_3^2}\right)^q
$$

We apply our bounds on H as follows.

1. In our preprocessing step, we ensure that all of the hyperedges (a', b', c') occur with even multiplicity and every distinct (a', b', c') has multiplicity at least two. If all of the hyperedges (a, b, c, a', b', c') have the same a, b, c, which is the case here, then we ensure that all of these hyperedges occur with even multiplicity and (a, b, c, a', b', c') occurs with multiplicity at least 2 for every distinct (a', b', c'). If not, which will happen in other cases, we take the two (a, b, c, a', b', c') hyperedges coming from each $(W_{abc}^{UVW})^2$ and double one or the other. When this happens, all of the (a, b, c) will be distinct because we will be summing over at least one of these indices and our intersection patterns say nothing about the (a, b, c) indices, so this ensures that each (a, b, c) occurs with multiplicity exactly 2.

2. If all of the hyperedges (a, b, c, a', b', c') have the same a, b, c, which happens here, we apply an entrywise bound q times to the (a, b, c, a', b', c') hyperedges. If instead each (a, b, c) appears with multiplicity 2 and always contains at least one free index, which will happen in other cases, we will apply the appropriate bounds on the (a, b, c, a', b', c') hyperedges.

3. We will apply an entrywise bound $q - z$ times on hyperedges (a', b', c') of multiplicity ≥ 4, reducing the multiplicity by 2 each time. After doing this, all our hyperedges will have multiplicity 2. We now ignore the c' vertices and consider the graph on the a', b' vertices. We then sum over a free a' or b' vertex in H whenever such a vertex exists. Otherwise, there must be a cycle (which could be a duplicated edge if we have hyper-edges (a', b', c'_1) and (a', b', c'_2)), in which case we use the entrywise bound on one edge of the cycle and delete it.

Definition 7.11. Let x be the number of times we delete an edge in a cycle using the entrywise bound.

Lemma 7.12. *Let k be the number of connected components of H. The total number of a' and b' vertices in H is $z + k - x \leqslant 2z - 2x$*

Proof. The first part can be proved in exactly the same way as Lemma 6.8. For the inequality, we need to show that $k \leqslant z - x$. To see this, note that there are at most z distinct edges and every time we delete an edge in a cycle, this removes one edge without reducing the number of connected components. After removing all cycles (and no other edges), we must have at least as many edges left as we have connected components, so $z - x \geqslant k$, as needed. $\square$

Summing over all $z \in [1, 2q]$ and all intersection patterns using Lemma 5.13 and noting that there are at most z a',b',c' indices but we must have two fewer a' or b' indices for each time we delete an edge in a cycle using an entrywise bound, our final bound on $E\left[\left(\left(W_{abc}^{UVW}\right)^2\right)^q\right]$ is

$$2q \cdot 2^{2q} \max_{z \in [1, 2q], x \in [0, z-1]} \left\{ q^{2q-z} \left(\frac{n_1 n_2 n_3}{m} \right)^{2q-z} \left(\frac{Br^2 \mu^6}{n_1^2 n_2^2 n_3^2} \right)^q \left(\frac{n_1 n_2 n_3}{\mu^3} \right)^z \left(\frac{\mu}{\min\{n_1, n_2\}} \right)^{2x} r^{q-z+x} \right\}$$

Since $m >> rq$ and $\mu r \leqslant \min\{n_1, n_2, n_3\}$, the inner expression will be maximized when $z = q$ and $x = 0$. Again, we will take q to be between $\frac{\ln n_{max}}{2}$ and $\frac{n_{max}}{2}$ so our final bound on $E\left[\left(\left(W_{abc}^{UVW}\right)^2\right)^q\right]$ is at most

$$(2q)^{(2q)} \left(\frac{2Br^2 \mu^3}{m \ln n_{max}} \right)^q n_{max}$$

Since $m > 10000C(2 + \beta)^2 r n_{max} \mu^2 \ln n_{max}$, we have that for all a, b, c.

$$E\left[\left(\left(W_{abc}^{UVW}\right)^2\right)^q\right] < q^{2q} \left(\frac{8Br\mu}{10000C(2 + \beta)^2 n_{max} (\ln n_{max})^2} \right)^q n_{max}$$

Corollary 7.13. *For a given $p \geqslant 1$, $r \geqslant 0$, $n > 0$, and $B > 0$, for a non-negative expression Z, if $E[Z^q] \leqslant (q^p B)^{2q} n^r$ for all integers $q > 1$ then for all $\beta > 0$,*

$$Pr\left[|Z| > B^2 e^{2p} \left(\frac{(r + \beta)}{2p} \ln n + 1 \right)^{2p} \right] < n^{-\beta}$$

Proof. This can be proved in the same way as Corollary 5.3 except that $|Z|$ takes the place of $\|YY^T\|$ which is why the bound of Corollary 5.3 is squared. $\square$

Using Corollary 7.13 with the appropriate parameters, we can show that for all $\beta > 0$, for all a, b, c

$$P\left[(W_{abc}^{UVW})^2 > \frac{32e^2 Br\mu}{10000C n_{max}} \right] < n_{max}^{-(\beta+4)}$$

$\square$

All of the remaining bounds can be proved in a similar way, we now describe the few differences.

We first consider the remaining bounds involving W^{UVW}. If we average over at least one coordinate, then instead of using entrywise bounds on the (a, b, c, a', b', c') hyperedges, the index or indices we average over will create free vertices, allowing us to bound the (a, b, c, a', b', c') hyperedges without using any entrywise bounds. This reduces our power bound by a factor of r^q and thus reduces our final bound by a factor of r. Moreover, each index that we average over reduces our final bound by a factor of μ. For example, if we are averaging over the a indices, then having these extra indices gives us an additional factor of $\left(\frac{n_1}{\mu}\right)^q$ in our power bound and thus a factor of $\frac{n_1}{\mu}$ in our final bound. However, since we are taking an average rather than a sum, we divide this by n_1, so our final bound is reduced by factor of μ, as needed.

At this point, we just need to consider the bounds involving W^{UV}, as the remaining cases are symmetric. When we analyze W^{UV} rather than W^{UVW}, instead of having $\left(\frac{Br^2\mu^6}{n_1^2 n_2^2 n_3^2}\right)^q$ in our power bound from the (a, b, c, a', b', c') hyperedges, we will have $\left(\frac{Br^2\mu^4}{n_1^2 n_2^2}\right)^q$ from (a, b, a', b') hyperedges, increasing our power bound by a factor of $\left(\frac{n_3^2}{\mu^2}\right)^q$. However, this is partially counteracted by the fact that we are either no longer summing over the c' indices separately from the c indices because we always have that $c_i' = c_i$. This removes a factor of $(\frac{n_3}{\mu})^z$ from our power bound. Thus, our bound on $E\left[\left(\left(W_{abc}^{UV}\right)^2\right)^q\right]$ is now

$$
2q \cdot 2^{2q} \max_{z \in [1, 2q], x \in [0, z-1]} \left\{ q^{2q-z} \left(\frac{n_1 n_2 n_3}{m}\right)^{2q-z} \left(\frac{Br^2\mu^4}{n_1^2 n_2^2}\right)^q \left(\frac{n_1 n_2}{\mu^2}\right)^z \left(\frac{\mu}{\min\{n_1, n_2\}}\right)^{2x} r^{q-z+x} \right\}
$$

We check that it is still optimal to take $z = q$ and $x = 0$. Since $r\mu \leqslant \min\{n_1, n_2, n_3\}$, it is always optimal to take $x = 0$. Now if we reduce z by 1, this gives us a factor of at most $\frac{q n_1 n_2 n_3}{m} \cdot \frac{r\mu^2}{n_1 n_2} = \frac{q r \mu^2 n_3}{m}$. We will take $q \leqslant 10(1 + \beta) \ln n_{max}$ and we have that $m > 10000C(2 + \beta)^2 n_{max} r \mu^2 \ln n_{max}$, so it is indeed still optimal to take $z = q$ and $x = 0$. Thus, the net effect of the differences is a factor of $\left(\frac{n_3}{\mu}\right)^q$ in our power bound and which gives us a factor of $\frac{n_3}{\mu}$ in our final bound. This gives us the following bound.

Lemma 7.14. *For all a, b, c and all $\beta > 0$, if $m > 10000C(2 + \beta)^2 n_{max} r \mu^2 \ln n_{max}$ then*

$$
P\left[(W_{abc}^{UV})^2 > \frac{32 e^2 Br}{10000C}\right] < n_{max}^{-(\beta+4)}
$$

We now consider what happens if we average over one or more of the a, b, and c indices. If we average over the a indices or average over the b indices, then instead of using entrywise bounds on the (a, b, a', b') hyperedges, the index or indices we average over will create free vertices, allowing us to bound the (a, b, a', b') hyperedges without using any entrywise bounds. We can now use the same reasoning as before. The final case is if we only average of the c indices. To handle this, instead of first processing the

(a, b, a', b') hyperedges and then processing the (a', b', c') hyperedges, we will first process the (a', b', c') hyperedges using all of the distinct c' vertices and process the (a, b, a', b') hyperedges in the same way that we processed the (a', b', c') hyperedges before.

Letting $y = z - (\text{\# of distinct } c)$, we have the following bounds on the number of a', b', c indices and the number of times we will use an entrywise bound

1. There are at most z a' indices and there are at most z b' indices.

2. There are at most $2z - 2x$ a' and b' indices, where x is the number of times we use an entrywise bound because of an edge in a cycle.

3. There are $(z - y)$ c indices.

4. The number of times that we will use an entrywise bound is $2q - 2z + x + y$

We check that it is optimal to take $z = q$, $x = 0$, and $y = 0$. Since $r\mu \leqslant \min\{n_1, n_2, n_3\}$, it is always optimal to take $x = y = 0$. Now if we reduce z by 1, this gives us a factor of at most $\frac{qn_1n_2n_3}{m} \cdot \frac{r^2\mu^3}{n_1n_2n_3} = \frac{qr^2\mu^3}{m}$. We will take $q \leqslant 10(1 + \beta)\ln n_{max}$ and we have that $r\mu \leqslant \min\{n_1, n_2, n_3\}$ and $m > 10000C(2 + \beta)^2 n_{max} r\mu^2 \ln n_{max}$, so it is indeed optimal to take $z = q$, $x = 0$, and $y = 0$ and thus the optimal case has no entrywise bounds. This implies that the same bounds hold when we average over the c indices as when we average over the a or b indices and we have all of our needed probabilistic bounds. Theorem 7.1 now follows from the inequality $W_{abc}^2 \leqslant 5\left((W_{abc}^{UV})^2 + (W_{abc}^{UW})^2 + (W_{abc}^{VW})^2 + 2(W_{abc}^{UVW})^2\right)$ and union bounds. $\qquad\square$

8 Trace Power Calculation for $P\bar{R}_\Omega A \otimes (P\bar{R}_\Omega A)^T$

In this section, we give a sketch of how to prove the following theorem.

Theorem 8.1. *If A is B-bounded, $C \geqslant 1$, and*

1. $m > 10000C(2 + \beta)^2 n_{max} r\mu^2 \ln n_{max}$

2. $m > 10000C(2 + \beta)^2 r\sqrt{n_1}\max\{n_2, n_3\}\mu^{\frac{3}{2}}\ln n_{max} \geqslant 10000C(2 + \beta)^2 r\sqrt{n_1 n_2 n_3}\mu^{\frac{3}{2}}\ln n_{max}$

3. $\mu r \leqslant \min\{n_1, n_2, n_3\}$

then

$$Pr\left[||Y|| > \frac{Bn_1 n_2 n_3}{Cr\mu^3}\right] < 4n^{-(\beta+1)}$$

whenever Y is any of the following:

1. $Y = P^{UV}\bar{R}_\Omega A \otimes (P^{UV}\bar{R}_\Omega A)^T$

2. $Y = P^{UW}\bar{R}_\Omega A \otimes (P^{UW}\bar{R}_\Omega A)^T$

3. $Y = P^{VW}\bar{R}_\Omega A \otimes (P^{VW}\bar{R}_\Omega A)^T$

4. $Y = P^{UVW}\bar{R}_\Omega A \otimes (P^{UVW}\bar{R}_\Omega A)^T$

Proof Sketch. This can be proved using the techniques of Sections 5 and 6 with one additional trick. We first consider the P^{UVW} case and then describe the small difference with the other cases.

The structure of $tr\left((YY^T)^q\right)$ is as follows. We have (a', b', c') hyperedges and we have hyperedges (a, b, c, a', b', c') which we can view as an outer triangle (a, b, c) and an inner triangle (a', b', c'). The outer triangles form hourglasses as before while the inner triangles sit inside the outer triangles.

The $\bar{R}_\Omega$ terms only involve the (a', b', c') triples so our intersection patterns only describe these indices. Thus, we sum over all of the a, b, c indices freely. We now use the following additional trick. We decompose each $UVW_{abca'b'c'}$ as $\sum_{i=1}^r u_{ia}v_{ib}w_{ic}u_{ia'}v_{ib'}w_{ic'}$. Now every vertex in the outer triangles appears in two hyperedges. When we sum over that vertex, we get something like $\sum_a u_{i_1 a}u_{i_2 a}$. This is 0 unless $i_1 = i_2$ and is 1 if $i_1 = i_2$. This in fact forces a global choice for i among the UVW terms, giving a global factor of r which is negligible. This also means that the vertices in the outer triangles give a factor of exactly 1, so they can be ignored! For the remaining terms of UVW, we use the bounds $u_{ia}^2 \leq \frac{\mu}{n_1}$, $v_{ib}^2 \leq \frac{\mu}{n_2}$, and $w_{ic}^2 \leq \frac{\mu}{n_3}$.

We now consider the contribution from summing over the a', b', c' vertices, the contribution from the $\bar{R}_\Omega$ terms, and the contribution from the entries of A. Letting z be the number of distinct triples (a', b', c') in the given intersection pattern, the contribution from the $\bar{R}_\Omega$ terms will be $\left(\frac{n_1 n_2 n_3}{m}\right)^{4q-z}$. The contribution from the entries of A from the $b(e)$ is B^{2q}. Letting x be the number of times that we have to use an entrywise bound on a doubled edge because it is in a cycle, we have the following bounds on the number of indices and the number of times we use an entrywise bound.

1. The number of distinct a indices, the number of distinct b indices, and the number of distinct c indices are all at most z

2. The total number of distinct indices is at most $3z - x$.

3. The number of times we use an entrywise bound is $2q - z + x$

It is easily checked that it is optimal to take $z = 2q$ and $x = 0$. In this case, the factors from summing over the indices and the remaining factors from the UVW terms cancel each other out! Thus, ignoring the terms from the number of intersection patterns (which gives a function of q), we are left with roughly $B^{2q}\left(\frac{n_1 n_2 n_3}{m}\right)^{2q}$.

This is much less than our previous bound of roughly $\left(\frac{n_1 n_2 n_3 \sqrt{n_1}\max\{n_2,n_3\}B}{m\mu^{\frac{3}{2}}}\right)^{2q}$ which we had for Y_4, so we get a correspondingly smaller norm bound, as needed.

The analysis us exactly the same for the P^{UV}, P^{UW}, and P^{VW} cases except for the following difference (which increases the bound, but still makes it much less than we had for Y_4). One of the outer indices is merged with one of the indices and we no longer have

UVW, UV, UW, or VW terms for that index. This means that instead of having perfect cancellation between the factors from summing over the indices and the remaining factors from the UVW terms, we have factors of $\frac{n_1}{\mu}$, $\frac{n_2}{\mu}$, or $\frac{n_3}{\mu}$ from summing over those indices.

Our bound will now be roughly at most $B^{2q} \left(\frac{n_1 n_2 n_3 n_{max}}{m \mu} \right)^{2q}$ which is still much less than the bound we had for Y_4 □

References

[BBH⁺12] Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou, *Hypercontractivity, sum-of-squares proofs, and their applications*, STOC, ACM, 2012, pp. 307–326. 2

[BGL16] Vijay V. S. P. Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee, *Certifying random polynomials over the unit sphere via sum of squares hierarchy*, CoRR **abs/1605.00903** (2016). 2

[BHK⁺16] Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin, *A nearly tight sum-of-squares lower bound for the planted clique problem*, CoRR **abs/1604.03084** (2016). 2

[BKS14] Boaz Barak, Jonathan A. Kelner, and David Steurer, *Rounding sum-of-squares relaxations*, STOC, ACM, 2014, pp. 31–40. 2

[BKS15] _____ , *Dictionary learning and tensor decomposition via the sum-of-squares method*, STOC, ACM, 2015, pp. 143–151. 2

[BM16] Boaz Barak and Ankur Moitra, *Noisy tensor completion via the sum-of-squares hierarchy*, COLT, JMLR Workshop and Conference Proceedings, vol. 49, JMLR.org, 2016, pp. 417–445. 1, 6, 7

[BS14] Boaz Barak and David Steurer, *Sum-of-squares proofs and the quest toward optimal algorithms*, Electronic Colloquium on Computational Complexity (ECCC) **21** (2014), 59. 2

[BS15] Srinadh Bhojanapalli and Sujay Sanghavi, *A new sampling technique for tensors*, CoRR **abs/1502.05023** (2015). 1

[CR09] Emmanuel J. Candès and Benjamin Recht, *Exact matrix completion via convex optimization*, Foundations of Computational Mathematics **9** (2009), no. 6, 717–772. 1

[CT10] Emmanuel J. Candès and Terence Tao, *The power of convex relaxation: near-optimal matrix completion*, IEEE Trans. Information Theory **56** (2010), no. 5, 2053–2080. 1

[FO07]   Uriel Feige and Eran Ofek, *Easily refutable subformulas of large random 3cnf formulas*, Theory of Computing **3** (2007), no. 1, 25–43. 6

[FS12]   Michael A. Forbes and Amir Shpilka, *On identity testing of tensors, low-rank recovery and compressed sensing*, STOC, ACM, 2012, pp. 163–172. 2

[GM15]   Rong Ge and Tengyu Ma, *Decomposing overcomplete 3rd order tensors using sum-of-squares algorithms*, APPROX-RANDOM, LIPIcs, vol. 40, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015, pp. 829–849. 2

[Gri01a]   Dima Grigoriev, *Complexity of positivstellensatz proofs for the knapsack*, Computational Complexity **10** (2001), no. 2, 139–154. 2

[Gri01b]   _____, *Linear lower bound on degrees of positivstellensatz calculus proofs for the parity*, Theor. Comput. Sci. **259** (2001), no. 1-2, 613–622. 2

[Gro11]   David Gross, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Trans. Information Theory **57** (2011), no. 3, 1548–1566. 1, 5, 6, 13, 15

[GV01]   Dima Grigoriev and Nicolai Vorobjov, *Complexity of null-and positivstellensatz proofs*, Ann. Pure Appl. Logic **113** (2001), no. 1-3, 153–160. 2

[Har14]   Moritz Hardt, *Understanding alternating minimization for matrix completion*, FOCS, IEEE Computer Society, 2014, pp. 651–660. 1

[HM16]   Elad Hazan and Tengyu Ma, *A non-generative framework and convex relaxations for unsupervised learning*, CoRR **abs/1610.01132** (2016). 2

[HSS15]   Samuel B. Hopkins, Jonathan Shi, and David Steurer, *Tensor principal component analysis via sum-of-square proofs*, COLT, JMLR Workshop and Conference Proceedings, vol. 40, JMLR.org, 2015, pp. 956–1006. 2, 6

[HSSS16]   Samuel B. Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer, *Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors*, STOC, ACM, 2016, pp. 178–191. 2

[HW14]   Moritz Hardt and Mary Wootters, *Fast matrix completion without the condition number*, COLT, JMLR Workshop and Conference Proceedings, vol. 35, JMLR.org, 2014, pp. 638–678. 1

[JN15]   Prateek Jain and Praneeth Netrapalli, *Fast exact matrix completion with finite samples*, COLT, JMLR Workshop and Conference Proceedings, vol. 40, JMLR.org, 2015, pp. 1007–1034. 1

[JNS13]   Prateek Jain, Praneeth Netrapalli, and Sujay Sanghavi, *Low-rank matrix completion using alternating minimization*, STOC, ACM, 2013, pp. 665–674. 1

[JO14]   Prateek Jain and Sewoong Oh, *Provable tensor factorization with missing data*, NIPS, 2014, pp. 1431–1439. 1

[Kho02]   Subhash Khot, *On the power of unique 2-prover 1-round games*, STOC, ACM, 2002, pp. 767–775. 2

[KMO09] Raghunandan H. Keshavan, Andrea Montanari, and Sewoong Oh, *Matrix completion from noisy entries*, NIPS, Curran Associates, Inc., 2009, pp. 952–960. 1

[Las01]   Jean B. Lasserre, *Global optimization with polynomials and the problem of moments*, SIAM J. Optim. **11** (2000/01), no. 3, 796–817. MR 1814045 2

[MSS16]   Tengyu Ma, Jonathan Shi, and David Steurer, *Polynomial-time tensor decompositions with sum-of-squares*, CoRR **abs/1610.01980** (2016). 2

[MW15]   Tengyu Ma and Avi Wigderson, *Sum-of-squares lower bounds for sparse PCA*, NIPS, 2015, pp. 1612–1620. 2

[Par00]   Pablo A Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, Ph.D. thesis, California Institute of Technology, 2000. 2

[Rec11]   Benjamin Recht, *A simpler approach to matrix completion*, Journal of Machine Learning Research **12** (2011), 3413–3430. 1, 5, 6, 13, 15

[RRS16]   Prasad Raghavendra, Satish Rao, and Tselil Schramm, *Strongly refuting random csps below the spectral threshold*, CoRR **abs/1605.00058** (2016). 2

[Sho87]   N. Z. Shor, *An approach to obtaining global extrema in polynomial problems of mathematical programming*, Kibernetika (Kiev) (1987), no. 5, 102–106, 136. MR 931698 2

[SS05]   Nathan Srebro and Adi Shraibman, *Rank, trace-norm and max-norm*, COLT, Lecture Notes in Computer Science, vol. 3559, Springer, 2005, pp. 545–560. 1

[Tro12]   Joel A. Tropp, *User-friendly tail bounds for sums of random matrices*, Foundations of Computational Mathematics **12** (2012), no. 4, 389–434. 15

A Controlling the kernel of matrix representations

We prove the following lemma in this section which was an ingredient of the proof of Theorem 4.3. Let $\{u_i\}, \{v_i\}, \{w_i\}$ be three orthonormal bases of $\mathbb{R}^n$.

Lemma (Restatement Lemma 4.15). *Let R be self-adjoint linear operator R on $\mathbb{R}^n \otimes \mathbb{R}^n$. Suppose $\langle (v_j \otimes w_k), R(v_i \otimes w_i) \rangle = 0$ for all indices $i, j, k \in [r]$ such that $i \in \{j, k\}$. Then, there exists a*

self-adjoint linear operator R' on $\mathbb{R}^n \otimes \mathbb{R}^n$ such that $R'(v_i \otimes w_i) = 0$ for all $i \in [r]$, the spectral norm of R' satisfies $\|R'\| \leqslant 10\|R\|$, and R' represents the same polynomial in $\mathbb{R}[y, z]$,

$$\langle (y \otimes z), R'(y \otimes z) \rangle = \langle (y \otimes z), R(y \otimes z) \rangle.$$

Proof. We write

$$R(v_i \otimes w_i) = \sum_{jk} c_{iijk} v_j \otimes w_k$$

Then the condition on the bilinear form of R implies that for all i, j, k, $c_{iiik} = 0$ and $c_{iiji} = 0$.

We now take Z to be the following matrix

$$Z = \sum_{i,j,k} c_{iijk} \left((v_j \otimes w_k)(v_i \otimes w_i)^T - (v_j \otimes w_i)(v_i \otimes w_k)^T - (v_i \otimes w_k)(v_j \otimes w_i)^T + (v_i \otimes w_i)(v_j \otimes w_k)^T \right)$$

$$+ \sum_{ij} \frac{c_{iijj}}{2} \left((v_j \otimes w_j)(v_i \otimes w_i)^T - (v_j \otimes w_i)(v_i \otimes w_j)^T - (v_i \otimes w_j)(v_j \otimes w_i)^T + (v_i \otimes w_i)(v_j \otimes w_j)^T \right)$$

It can be verified directly that Z represents the 0 polynomial and has the same behavior on each of the $(v_i \otimes w_i)$ as R. The factor of $\frac{1}{2}$ in the second sum comes from the fact that $c_{jjii} = c_{iijj}$ and the fourth term for c_{jjii} matches the first term for c_{iijj}

We choose $R' = R - Z$. In order to show the bound $\|R'\| \leqslant 10\|R\|$ it is enough to show that $\|Z\| \leqslant 9\|R\|$

We analyze the norm of Z as follows. We break Z into parts according to each type of term and analyze each part separately. Define X to be the subspace spanned by the $(v_i \otimes w_i)$, define P_X to be the projection onto X and define $P_X^\perp$ to be the projection onto the subspace orthogonal to X.

For the part $\sum_{ijk} c_{iijk}(v_j \otimes w_k)(v_i \otimes w_i)^T$, note that $\sum_{ijk} c_{iijk}(v_j \otimes w_k)(v_i \otimes w_i)^T = P_X^\perp R' P_X$ so it has norm at most $\|R\|$.

For the part $\sum_{ijk} c_{iijk}(v_j \otimes w_i)(v_i \otimes w_k)^T$, note that under a change of basis this is equivalent to a block-diagonal matrix with blocks $\sum_{jk} c_{iijk} v_j w_k^T$. The norm of each such block is at most its Frobenius norm, which is the norm of $\sum_{jk} c_{iijk}(v_j \otimes w_k) = R'(v_i \otimes w_i)$. Thus, this part also has norm at most $\|R\|$. Using similar arguments, we can bound the norm of the other parts by $\|R\|$ as well, obtaining that $\|Z\| \leqslant 8\|R\|$. $\qquad\square$

B   Full Trace Power Calculation

In this section, we analyze $\|\sum_a A_a \otimes B_a^T\|$ where $A = (\bar{R}_{\Omega_l} P_l) \cdots (\bar{R}_{\Omega_1} P_1)(\bar{R}_{\Omega_0} X)$ or $A = P_{l+1}(\bar{R}_{\Omega_l} P_l) \cdots (\bar{R}_{\Omega_1} P_1)(\bar{R}_{\Omega_0} X)$ for some projection operators $P_1, \cdots, P_l, P_{l+1}$ and $B = (\bar{R}_{\Omega_{l'}} P_{l'}) \cdots (\bar{R}_{\Omega_1} P_1')(\bar{R}_{\Omega_0} X)$ or $B = P_{l'+1}(\bar{R}_{\Omega_{l'}} P_{l'}) \cdots (\bar{R}_{\Omega_1} P_1')(\bar{R}_{\Omega_0} X)$ for some projection operators $P_1', \cdots, P_{l'}', P_{l'+1}'$. In particular, we prove the following theorem using the trace power method.

Theorem B.1. *There is an absolute constant C such that for any $\alpha > 1$ and $\beta > 0$,*

$$Pr\left[\|\sum_a A_a \otimes B_a^T\| > \alpha^{-(l+l'+2)} \right] < n^{-\beta}$$

41

as long as

1. $r\mu \leqslant \min\{n_1, n_2, n_3\}$

2. $m > C\alpha\beta\mu^{\frac{3}{2}}r\sqrt{n_1}\max\{n_2, n_3\}log(\max\{n_1, n_2, n_3\})$

3. $m > C\alpha\beta\mu^2 r\max\{n_1, n_2, n_3\}log(\max\{n_1, n_2, n_3\})$

Remark B.2. In this draft, we only sketch the case where we do not have projection operators in front. To handle the cases where there are projection operators in front, we can use the same ideas that are sketched out in Section 8,

B.1 Term Structure

When we expand out the sums in $tr\left(((\sum_a A_a \otimes B_a^T)(\sum_a A_a \otimes B_a^T)^T)^q\right)$, our terms will have the following structure. We label the indices so that each $\bar{R}_{\Omega_i}$ operator has its own indices (a_{ij}, b_{ij}, c_{ij}) or $(a'_{ij}, b'_{ij}, c'_{ij})$. Many of these indices will be equal.

1. For all $i \in [0, l]$ and all $j \in [1, 2q]$ we have indices (a_{ij}, b_{ij}, c_{ij}) and a corresponding term $\bar{R}_{\Omega_i}(a_{ij}, b_{ij}, c_{ij})$ in the product.

2. For all $i \in [0, l']$ and all $j \in [1, 2q]$ we have indices $(a'_{ij}, b'_{ij}, c'_{ij})$ and a corresponding term $\bar{R}_{\Omega_i}(a'_{ij}, b'_{ij}, c'_{ij})$ in the product.

3. For all $j \in [1, 2q]$ we have a term $X_{a_0 b_0 c_0}$ and a term $X_{a'_0 b'_0 c'_0}$ in the product.

4. For all $i \in [0, l]$ and all $j \in [1, 2q]$ we have a term $P_i(a_{ij}, b_{ij}, c_{ij}, a_{(i-1)j}, b_{(i-1)j}, c_{(i-1)j})$ in the product.

5. For all $i \in [0, l']$ and all $j \in [1, 2q]$ we have a term $P'_i(a'_{ij}, b'_{ij}, c'_{ij}, a'_{(i-1)j}, b'_{(i-1)j}, c'_{(i-1)j})$ in the product.

We represent the terms in the product graphically as follows.

Definition B.3.

1. For all i, we represent the terms $\bar{R}_{\Omega_i}(a_{ij}, b_{ij}, c_{ij})$ and $\bar{R}_{\Omega_i}(a'_{ij}, b'_{ij}, c'_{ij})$ by triangles. We call these triangles R_i-triangles and R'_i-triangles respectively.

2. For all i and j,

   (a) If $P_i = P_{UV}$ then we represent $P_i(a_{ij}, b_{ij}, c_{ij}, a_{(i-1)j}, b_{(i-1)j}, b_{(i-1)j})$ by a hyperedge $(a_{ij}, b_{ij}, a_{(i-1)j}, b_{(i-1)j})$. We call this hyperedge a UV-hyperedge.

   (b) If $P_i = P_{UW}$ then we represent $P_i(a_{ij}, b_{ij}, c_{ij}, a_{(i-1)j}, b_{(i-1)j}, b_{(i-1)j})$ by a hyperedge $(a_{ij}, c_{ij}, a_{(i-1)j}, c_{(i-1)j})$. We call this hyperedge a UW-hyperedge.

   (c) If $P_i = P_{VW}$ then we represent $P_i(a_{ij}, b_{ij}, c_{ij}, a_{(i-1)j}, b_{(i-1)j}, b_{(i-1)j})$ by a hyperedge $(b_{ij}, c_{ij}, b_{(i-1)j}, c_{(i-1)j})$. We call this hyperedge a VW-hyperedge.

42

(d) If $P_i = P_{UVW}$ then we represent $P_i(a_{ij}, b_{ij}, c_{ij}, a_{(i-1)j}, b_{(i-1)j}, b_{(i-1)j})$ by a hyperedge $(a_{ij}, b_{ij}, c_{ij}, a_{(i-1)j}, b_{(i-1)j}, c_{(i-1)j})$. We call this hyperedge a UVW-hyperedge.

We represent the P_i' terms by hyperedges in a similar manner.

3. For all $j \in [1, 2q]$, we represent the term $X_{a_{0j}b_{0j}c_{0j}}$ with a hyperedge (a_{0j}, b_{0j}, c_{0j}) and we represent the term $X_{a_{0j}'b_{0j}'c_{0j}'}$ with a hyperedge $(a_{0j}', b_{0j}', c_{0j}')$. We call these hyperedges X-hyperedges.

We have the following equalities among the indices:

1. For all $j \in [1, 2q]$, $a_{lj} = a_{l'j}'$

2. For all $j \in [1, 2q]$, if j is even then $b_{lj} = b_{l(j+1)}$ and $c_{l'j}' = c_{l'(j+1)}'$

3. For all $j \in [1, 2q]$, if j is odd then $c_{lj} = c_{l(j+1)}$ and $b_{l'j}' = b_{l'(j+1)}'$

4. For all $i \in [1, l]$ and all $j \in [1, 2q]$, if $P_i = P_{UV}$ then $c_{ij} = c_{(i-1)j}$, if $P_i = P_{UW}$ then $b_{ij} = b_{(i-1)j}$, and if $P_i = P_{VW}$ then $a_{ij} = a_{(i-1)j}$

5. For all $i \in [1, l]$ and all $j \in [1, 2q]$, if $P_i' = P_{UV}$ then $c_{ij}' = c_{(i-1)j}'$, if $P_i' = P_{UW}$ then $b_{ij}' = b_{(i-1)j}'$, and if $P_i' = P_{VW}$ then $a_{ij}' = a_{(i-1)j}'$

B.2 Techniques

In this section, we describe how to bound the expected value of

$$tr\left(\left(\left(\sum_a A_a \otimes B_a\right)\left(\sum_a A_a \otimes B_a\right)^T\right)^q\right)$$

We first consider the $\bar{R}_{\Omega_i}$ terms, which for a given choice of the indices are as follows:

$$\left(\prod_{i=0}^{l}\prod_{j=1}^{2q} \bar{R}_{\Omega_i}(a_{ij}, b_{ij}, c_{ij})\right)\left(\prod_{i=0}^{l'}\prod_{j=1}^{2q} \bar{R}_{\Omega_i}(a_{ij}', b_{ij}', c_{ij}')\right)$$

For a given choice of the indices, the expected value of this part can be bounded as follows

Definition B.4. For all i, let z_i be the number of distinct R_i-triangles and let z_i' be the number of distinct R_i'-triangles. If a triangle appears as both an R_i-triangle and as an R_i'-triangle then it contributes $\frac{1}{2}$ to both z_i and z_i' (so the total number of distinct triangles at level i is $z_i + z_i'$)

Lemma B.5. *For a given choice of the indices $\{a_{ij}, b_{ij}, c_{ij}\}$ and $\{a_{ij}', b_{ij}', c_{ij}'\}$*

1. *If any triangle appears exactly once at some level i then*

$$E\left[\left(\prod_{i=0}^{l}\prod_{j=1}^{2q} \bar{R}_{\Omega_i}(a_{ij}, b_{ij}, c_{ij})\right)\left(\prod_{i=0}^{l'}\prod_{j=1}^{2q} \bar{R}_{\Omega_i}(a_{ij}', b_{ij}', c_{ij}')\right)\right] = 0$$

43

2. *If for all i, all of the triangles which appear at level i appear at least twice then*

$$0 < E\left[\left(\prod_{i=0}^{l}\prod_{j=1}^{2q}\bar{R}_{\Omega_i}(a_{ij}, b_{ij}, c_{ij})\right)\left(\prod_{i=0}^{l'}\prod_{j=1}^{2q}\bar{R}_{\Omega_i}(a'_{ij}, b'_{ij}, c'_{ij})\right)\right]$$

$$\leq \left(\frac{n_1 n_2 n_3}{m}\right)^{\sum_{i=0}^{l}(2q-z_i)+\sum_{i=0}^{l'}(2q-z'_i)}$$

Proof. If there is any triangle (a, b, c) which appears exactly once in level i then $\bar{R}_{\Omega_i}(a_i, b_i, c_i)$ has expectation 0 and is independent of every other term in the product so the entire product has value 0. Otherwise, note that for $k > 1$, $0 < E\left[(R_{\Omega_i}(a, b, c))^k\right] \leq \left(\frac{n_1 n_2 n_3}{m}\right)^{k-1}$. Further note that $R_{\Omega_i}(a, b, c)$ terms with either different i or different a, b, c are independent of each other. Thus, using this bound, each copy of a triangle beyond the first gives us a factor of $\left(\frac{n_1 n_2 n_3}{m}\right)$. The total number of factors which we obtain is the total number of triangles minus the number of distinct triangles (where triangles at different levels are automatically distinct) and the result follows. $\square$

We now note that this bound holds for all sets of indices that follow the same intersection pattern of which R_i-triangles and R'_i-triangles are equal to each other. Thus, we can group all terms which have the same intersection pattern together, using this bound on all of them.

Each such intersection pattern forces additional equalities between the indices. After taking these equalities into account, we must sum over the remaining distinct indices. We now analyze what happens with the remaining terms of the product as we sum over these indices. We begin by considering how well we can bound the sum of entries of X squared if we sum over 0,1,2, or all 3 indices.

Lemma B.6.

1. $\max_{abc}\{X^2_{abc}\} \leq \frac{r^2\mu^3}{n_1 n_2 n_3}$

2. $\max_{bc}\{\sum_a X^2_{abc}\} \leq \frac{r\mu^2}{n_2 n_3}$

3. $\max_c\{\sum_{a,b} X^2_{abc}\} \leq \frac{r\mu}{n_3}$

4. $\sum_{a,b,c} X^2_{abc} = r$

Proof. For the first statement,

$$X^2_{abc} = \sum_{i,j} u_{ia}v_{ib}w_{ic}u_{ja}v_{jb}w_{jc} \leq r^2 \max_i\{u^2_{ia}v^2_{ib}w^2_{ic}\} \leq \frac{r^2\mu^3}{n_1 n_2 n_3}$$

For the second statement,

$$\sum_a X^2_{abc} = \sum_{i,j}\left(\sum_a u_{ia}u_{ja}\right)v_{ib}w_{ic}v_{jb}w_{jc} = \sum_{i,a} u^2_{ia}v^2_{ib}w^2_{ic} \leq \frac{\mu^2}{n_2 n_3}\sum_{a,i} u^2_{ia} = \frac{r\mu^2}{n_2 n_3}$$

For the third statement,

$$\sum_{a,b} X_{abc}^2 = \sum_{i,j,b} \left(\sum_a u_{ia} u_{ja} \right) v_{ib} w_{ic} v_{jb} w_{jc} = \sum_{i,a,b} u_{ia}^2 v_{ib}^2 w_{ic}^2 \le \frac{\mu}{n_3} \sum_i \left(\sum_a u_{ia}^2 \sum_b v_{ib}^2 \right) = \frac{r\mu}{n_3}$$

The final statement can be proved in a similar way. $\qquad\square$

Note that every index we sum over reduces the average value by μ. Further note that if we do not sum over any indices, there is an extra factor of r in our bound. Following similar logic, similar statements hold for the P_i and P_i' terms.

We utilize this as follows. We start with a hypergraph H which represents the current terms in our product. We first preprocess our product using the inequality $|ab| \le \frac{x}{2}a^2 + \frac{b^2}{2x}$ (carefully choosing each a, b, and x) to make all of our hyperedges have even multiplicity. Note that when doing this, we cannot fully control which doubled hyperedges we will have; if we apply this on hyperedges e_1 and e_2 we could end up with two copies of e_1 or two copies of e_2.

Now if we have a hyperedge with multiplicity 4 or more, we use the entrywise bound to reduce its multiplicity by 2. For example, if our sum was $\sum_a X_{abc}^4$ then we would use the inequality

$$\sum_a X_{abc}^4 \le \max_{abc} \{X_{abc}^2\} \sum_a X_{abc}^2$$

to bound this sum.

Once every hyperedge appears with power 2, we choose an ordering for how we will bound the hyperedges. For each hyperedge, we sum over all indices which are currently only incident with that hyperedge, take the appropriate bound, and then delete the hyperedge and these indices from our current hypergraph H. We account for all of this with the following definitions:

Definition B.7.

1. Define the base value of an X-hyperedge e to be $v(e) = \sqrt{\frac{r\mu^3}{n_1 n_2 n_3}}$

2. Define the base value of a UV-hyperedge e to be $v(e) = \sqrt{\frac{r\mu^4}{n_1^2 n_2^2}}$

3. Define the base value of a UW-hyperedge e to be $v(e) = \sqrt{\frac{r\mu^4}{n_1^2 n_3^2}}$

4. Define the base value of a VW-hyperedge e to be $v(e) = \sqrt{\frac{r\mu^4}{n_2^2 n_3^2}}$

5. Define the base value of an UVW-hyperedge e to be $v(e) = \sqrt{\frac{r\mu^6}{n_1^2 n_2^2 n_3^2}}$

Definition B.8. We say that a index in our hypergraph H is free if it is incident with at most one hyperedge.

45

For a given intersection pattern, assuming that every vertex is incident with at least one hyperedge after the preprocessing, our final bound will be

$$
\left(\frac{n_1 n_2 n_3}{m}\right)^{\sum_{i=0}^{l}(2q-z_i)+\sum_{i=0}^{l'}(2q-z_i')} \left(\prod_e v(e)\right) \left(\frac{n_1}{\mu}\right)^{\text{\# of a indices}}
$$

$$
\left(\frac{n_2}{\mu}\right)^{\text{\# of b indices}} \left(\frac{n_3}{\mu}\right)^{\text{\# of c indices}} r^{\text{\# of doubled hyperedges we bound with no free index}}
$$

To see this, note that from the discussion above, when an index a,b, or c is free and we sum over it, we obtain n_1, n_2, or n_3 terms respectively but this also reduces the current bound we are using by a factor of μ. This will happen precisely one time for every index which is incident to at least one edge. Thus, for this part the ordering doesn't really matter. However, there is an extra factor of r whenever we bound a doubled hyperedge with no free index (including when this hyperedge has multiplicity 4 or higher and we reduce its multiplicity by 2). We want to avoid this extra factor of r as much as possible. We describe how to do this in subsection B.4.

Remark B.9. When summing over an index, we may not acutally sum over all possiblities because this could create equalities between triangles which should not be equal according to the intersection pattern. However, adding in these missing terms can only increase the sum, so it is still an upper bound.

B.3 Bounding the number of indices

In this subsection, we describe bounds on the number of each type of index for a given intersection pattern. We then define a coefficient Δ which is the discrepency between the our bounds and the actual number of indices and reexpress our boun in terms of Δ.

We make the following simplifying assumption about our sums.

1. For all $i \in [0, l']$, we either have that $a_{ij}' = a_{ij}$ for all $j \in [1, 2q]$ or $a_{ij}' \neq a_{ij}$ for all $j \in [1, 2q]$.

2. For all $i \in [0, l']$, we either have that $b_{ij}' = b_{ij}$ for all $j \in [1, 2q]$ or $b_{ij}' \neq b_{ij}$ for all $j \in [1, 2q]$.

3. For all $i \in [0, l']$, we either have that $c_{ij}' = c_{ij}$ for all $j \in [1, 2q]$ or $c_{ij}' \neq c_{ij}$ for all $j \in [1, 2q]$.

Moreover, all of these choices are fixed beforehand. We justify this assumption with a random partitioning argument in subsection B.5.

With this setup, we first bound the number of each type of index which appears.

Definition B.10. For all i,

1. We define x_{ia} to be the number of distinct indices a_{ij} which do not appear at a higher level, we define x_{ib} to be the number of distinct indices b_{ij} which do not appear at a higher level, and we define x_{ic} to be the number of distinct indices c_{ij} which do not appear at a higher level.

46

2. We define x'_{ia} to be the number of distinct indices a'_{ij} which do not appear at a higher level, we define x'_{ib} to be the number of distinct indices b'_{ij} which do not appear at a higher level, and we define x'_{ic} to be the number of distinct indices c'_{ij} which do not appear at a higher level.

In the case where we have an equality $a'_{ij} = a_{ij}$ and this index does not appear at a higher level, we instead count it as $\frac{1}{2}$ for x_{ia} and $\frac{1}{2}$ for x'_{ia} (and similarly for b and c).

Recall that we defined z_i to be the number of distinct R_i-triangles and we defined z'_i to be the nmber of distinct R'_i-triangles. z_i and z'_i give the following bounds on the coefficients

Proposition B.11.

1. For all $i < l$, $x_{ia} \leqslant z_i$, $x_{ib} \leqslant z_i$, and $x_{ic} \leqslant z_i$.

2. For all $i < l'$, $x'_{ia} \leqslant z'_i$, $x'_{ib} \leqslant z'_i$, and $x'_{ic} \leqslant z'_i$.

3. If $l' \neq l$, $b'_{lj} \neq b_{lj}$, or $c'_{lj} \neq c_{lj}$,

   (a) $x_{la} + x'_{l'a} \leqslant \min\{z_l, z'_{l'}\}$

   (b) $x_{lb} + x_{lc} \leqslant z_l + 1$

   (c) $x'_{l'b} + x'_{l'c} \leqslant z'_{l'} + 1$

4. In the special case that $l' = l$, $b'_{lj} = b_{lj}$, and $c'_{lj} = c_{lj}$,

   (a) $x_{la} + x'_{l'a} = z_l + z'_{l'}$

   (b) $x_{lb} + x_{lc} = x'_{l'b} + x'_{l'c} = 1$

Proof. The first two statements and 3(a) follow from the observation that distinct vertices must be in distinct triangles. For 3(b), note that if we take the b, c edges from each R_i-triangle, the resulting graph is connected. Thus, each distinct such edge (which must come from a distinct triangle) after the first edge can only add one new vertex and the result follows. 3(c) can be proved analogously.

For the fourth statement, note that in this case all of the b_{lj} and b'_{lj} indices are equal to a single index b and all of the c_{lj} and c'_{lj} indices are equal to a single index c. Thus, the number of distinct a_{lj} is equal to the number of distinct R_l and R'_l triangles. $\square$

With these bounds in mind, we define x^{max} coefficients which represent the maximum number of distinct indices we can expect (given the structure of A and B and the values z_i, z'_i) and Δ coefficients which describe the discrepency between this maximum and the number of distinct indices which we actually have.

Definition B.12.

1. For all $i < l$,

47

(a) If $P_{i+1} = P_{UV}$ then we define $x_{ia}^{max} = x_{ib}^{max} = z_i$. We define $\Delta_{ia} = x_{ia}^{max} - x_{ia}$, $\Delta_{ib} = x_{ib}^{max} - x_{ib}$, and $\Delta_{ic} = 0$.

(b) If $P_{i+1} = P_{UW}$ then we define $x_{ia}^{max} = x_{ic}^{max} = z_i$. We define $\Delta_{ia} = x_{ia}^{max} - x_{ia}$, $\Delta_{ic} = x_{ic}^{max} - x_{ic}$, and $\Delta_{ib} = 0$.

(c) If $P_{i+1} = P_{VW}$ then we define $x_{ia}^{max} = x_{ib}^{max} = z_i$. We define $\Delta_{ib} = x_{ib}^{max} - x_{ib}$, $\Delta_{ic} = x_{ic}^{max} - x_{ic}$, and $\Delta_{ia} = 0$.

(d) If $P_{i+1} = P_{UVW}$ then we define $x_{ia}^{max} = x_{ib}^{max} = x_{ic}^{max} = z_i$. We define $\Delta_{ia} = x_{ia}^{max} - x_{ia}$, $\Delta_{ib} = x_{ib}^{max} - x_{ib}$, and $\Delta_{ic} = x_{ic}^{max} - x_{ic}$.

2. For all $i < l'$,

(a) If $P'_{i+1} = P_{UV}$ then we define $x'^{max}_{ia} = x'^{max}_{ib} = z'_i$. We define $\Delta'_{ia} = x'^{max}_{ia} - x'_{ia}$, $\Delta'_{ib} = x'^{max}_{ib} - x'_{ib}$, and $\Delta'_{ic} = 0$.

(b) If $P'_{i+1} = P_{UW}$ then we define $x'^{max}_{ia} = x'^{max}_{ic} = z'_i$. We define $\Delta'_{ia} = x'^{max}_{ia} - x'_{ia}$, $\Delta'_{ic} = x'^{max}_{ic} - x'_{ic}$, and $\Delta'_{ib} = 0$.

(c) If $P'_{i+1} = P_{VW}$ then we define $x'^{max}_{ia} = x'^{max}_{ib} = z'_i$. We define $\Delta'_{ib} = x'^{max}_{ib} - x'_{ib}$, $\Delta'_{ic} = x'^{max}_{ic} - x'_{ic}$, and $\Delta'_{ia} = 0$.

(d) If $P'_{i+1} = P_{UVW}$ then we define $x'^{max}_{ia} = x'^{max}_{ib} = x'^{max}_{ic} = z'_i$. We define $\Delta_{ia} = x'^{max}_{ia} - x'_{ia}$, $\Delta'_{ib} = x'^{max}_{ib} - x'_{ib}$, and $\Delta'_{ic} = x'^{max}_{ic} - x'_{ic}$.

3. If $l' \neq l$, $b'_{l'j} \neq b_{lj}$, or $c'_{l'j} \neq c_{lj}$ then we define $x_{ll'a}^{max} = \min\{z_l, z'_{l'}\}$, we define $x_{lbc}^{max} = z_l + 1$, and we define $x'^{max}_{lbc} = z'_{l'} + 1$. In the special case where $l' = l$, $b'_{l'j} = b_{lj}$, and $c'_{l'j} = c_{lj}$, we define $x_{ll'a}^{max} = z_l + z'_l$ and $x_{lbc}^{max} = x'^{max}_{l'bc} = 1$. In both of these cases, we define $\Delta_{ll'a} = x_{ll'a}^{max} - x_{la} - x'_{l'a}$, $\Delta_{lbc} = x_{lbc}^{max} - x_{lb} - x_{lc}$, and $\Delta'_{l'bc} = x'^{max}_{l'bc} - x'_{l'b} - x'_{l'c}$.

Definition B.13. We define

$$\Delta = \Delta_{ll'a} + \Delta_{lbc} + \Delta'_{l'bc} + \sum_{i=0}^{l-1}(\Delta_{ia} + \Delta_{ib} + \Delta_{ic}) + \sum_{i=0}^{l'-1}(\Delta'_{ia} + \Delta'_{ib} + \Delta'_{ic})$$

We now reexpress our bound in terms of Δ.

Lemma B.14. *For a given intersection pattern and choices for the equalities or inequalities between the a_{ij}, b_{ij}, c_{ij} and $a'_{ij}, b'_{ij}, c'_{ij}$ indices, we can obtain a bound which is a product of*

$$\frac{n_2 n_3}{\mu^2} \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta} r^{\text{\# of doubled hyperedges we bound with no free index}-\left(\sum_{i=0}^{l}(2q-z_i)+\sum_{i=0}^{l'}(2q-z'_i)\right)}$$

and terms of the form $\frac{r\mu^{\frac{3}{2}}\sqrt{n_1}\max\{n_2,n_3\}}{m}$, $\frac{r\mu^2\max\{n_1,n_2,n_3\}}{m}$, or $\frac{r\mu^3}{m}$

Proof. Recall that our bound was

$$\left(\frac{n_1 n_2 n_3}{m}\right)^{\sum_{i=0}^{l}(2q-z_i)+\sum_{i=0}^{l'}(2q-z'_i)} \left(\prod_e v(e)\right)\left(\frac{n_1}{\mu}\right)^{\text{\# of } a \text{ indices}}$$

48

$$\left(\frac{n_2}{\mu}\right)^{\text{\# of b indices}} \quad \left(\frac{n_3}{\mu}\right)^{\text{\# of c indices}} \quad r^{\text{\# of doubled hyperedges we bound with no free index}}$$

For all $i < l$, we consider the part of this bound which comes from P_{i+1} and the indices a_i, b_i, c_i which do not appear at a higher level. Similary, for all $i < l'$, we consider the part of this bound which comes from P'_{i+1} and the indices a'_i, b'_i, c'_i which do not appear at a higher level. Finally, we consider the part of this bound that comes from the X hyperedges, the R_l-triangles, the $R'_{l'}$-triangles, and their indices.

1. If $P_{i+1} = P_{UV}$ then we can decompose the corresponding terms into the following parts:

    (a) $(\frac{r\mu^4}{n_1^2 n_2^2})^q$ from the hyperedges.

    (b) $\left(\frac{n_1 n_2}{\mu^2}\right)^q$ from the q potential new a, b, and c indices.

    (c) $(\frac{n_1 n_2 n_3}{m})^q$ from the q potential distinct triangles.

    (d) $\left(r \cdot \frac{\mu^2}{n_1 n_2} \cdot \frac{n_1 n_2 n_3}{m}\right)^{q-z_i} = \left(\frac{r\mu^2 n_3}{m}\right)^{q-z_i}$ from the actual number of distinct triangles, the corresponding reduced maximum number of potential new indices, and the factors of r which we take from $r^{\text{\# of doubled hyperedges we bound with no free index}}$

    (e) $\left(\frac{\mu}{n_1}\right)^{\Delta_{ia}} \left(\frac{\mu}{n_2}\right)^{\Delta_{ib}} \left(\frac{\mu}{n_3}\right)^{\Delta_{ic}} \leqslant \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta_{ia}+\Delta_{ib}+\Delta_{ic}}$ from the actual number of new indices which we have

    Putting everything together we obtain

    $$\left(\frac{r\mu^2 n_3}{m}\right)^{2q-z_i} \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta_{ia}+\Delta_{ib}+\Delta_{ic}}$$

    Similar arguments apply if $P_{i+1} = P_{VW}$ or P_{VW}

2. If $P_{i+1} = P_{UVW}$ then we can decompose the corresponding terms into the following parts:

    (a) $(\frac{r\mu^6}{n_1^2 n_2^2 n_3^2})^q$ from the hyperedges.

    (b) $\left(\frac{n_1 n_2 n_3}{\mu^2}\right)^q$ from the q potential new a, b, and c indices.

    (c) $(\frac{n_1 n_2 n_3}{m})^q$ from the q potential distinct triangles.

    (d) $\left(r \cdot \frac{\mu^3}{n_1 n_2 n_3} \cdot \frac{n_1 n_2 n_3}{m}\right)^{q-z_i} = \left(\frac{r\mu^3}{m}\right)^{q-z_i}$ from the actual number of distinct triangles, the corresponding reduced maximum number of potential new indices, and the factors of r which we take from $r^{\text{\# of doubled hyperedges we bound with no free index}}$

    (e) $\left(\frac{\mu}{n_1}\right)^{\Delta_{ia}} \left(\frac{\mu}{n_2}\right)^{\Delta_{ib}} \left(\frac{\mu}{n_3}\right)^{\Delta_{ic}} \leqslant \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta_{ia}+\Delta_{ib}+\Delta_{ic}}$ from the actual number of new indices which we have

Putting everything together we obtain

$$\left(\frac{r\mu^3}{m}\right)^{2q-z_i} \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta_{ia}+\Delta_{ib}+\Delta_{ic}}$$

Similar arguements holds for the P' terms.

3. If $l' \neq l$, $b'_{lj} \neq b_{lj}$, or $c'_{lj} \neq c_{lj}$ then our remaining terms are as follows

   (a) $(\frac{r\mu^3}{n_1 n_2 n_3})^{2q}$ from the hyperedges.

   (b) $\frac{n_2 n_3}{\mu^2} \left(\frac{n_1(\max\{n_2,n_3\})^2}{\mu^3}\right)^q$ from the q potential a indices and $2q + 2$ potential b or c indices (which must have at least one b index and at least one c index).

   (c) $(\frac{n_1 n_2 n_3}{m})^{2q}$ from the $2q$ potential distinct triangles.

   (d) $\left(r \cdot \frac{\mu^{\frac{3}{2}}}{\sqrt{n_1}\max\{n_2,n_3\}} \cdot \frac{n_1 n_2 n_3}{m}\right)^{2q-z_l-z'_l} \leqslant \left(\frac{r\mu^{3/2}\sqrt{n_1}\max\{n_2,n_3\}}{m}\right)^{2q-z_l-z'_l}$ from the actual number of distinct triangles, the corresponding reduced maximum number of potential new indices, and the factors of r which we take from $r^{\#\text{ of doubled hyperedges we bound with no free index}}$

   (e) $\left(\frac{\mu}{n_1}\right)^{\Delta_{ll'a}} \left(\frac{\mu}{\max\{n_2,n_3\}}\right)^{\Delta_{lbc}+\Delta'_{l'bc}} \leqslant \left(\frac{\mu}{\min\{n_1,n_2,n_3\}}\right)^{\Delta_{ll'a}+\Delta_{lbc}+\Delta'_{l'bc}}$ from the actual number of new indices which we have

   Putting everything together we obtain

$$\frac{n_2 n_3}{\mu^2} \left(\frac{r\mu^{3/2}\sqrt{n_1}\max\{n_2, n_3\}}{m}\right)^{4q-z_l-z'_l} \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta_{ll'a}+\Delta_{lbc}+\Delta'_{l'bc}}$$

4. In the special case that $l' = l$, $b'_{lj} = b_{lj}$, and $c'_{lj} = c_{lj}$, we have the same terms except that now there is only one b and c index and there are $2q$ potential a indices. Following similar logic we obtain a bound of

$$\frac{n_2 n_3}{\mu^2} \left(\frac{r\mu^2 n_1}{m}\right)^{4q-z_l-z'_l} \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta_{ll'a}+\Delta_{lbc}+\Delta'_{l'bc}}$$

$\square$

With this lemma in hand, to show our bound it is sufficient to show that we can choose an ordering on the hyperedges such that the number of times we bound a doubled hyperedge without a free index is at most $\Delta + \sum_{i=0}^{l} (2q - z_i) + \sum_{i=0}^{l'} (2q - z'_i)$

B.4   Choosing an ordering

In this section, we describe how to choose a good ordering for bounding the hyperedges.

Lemma B.15. *For any structure for A and B (including equalities or inequalities between a_{ij}, b_{ij}, c_{ij} and $a'_{ij}, b'_{ij}, c'_{ij}$) and any intersection pattern, there is a way to double the hyperedges using the inequality $|ab| \leqslant \frac{x}{2}a^2 + \frac{1}{2x}b^2$ and then bound the doubled hyperedges one by one so that*

1. *After doubling the hyperedges, every index is part of at least one hyperedge.*

2. *The number of times that we bound a doubled hyperedge without a free index is at most*
   $$\Delta + \sum_{i=0}^{l}(2q - z_i) + \sum_{i=0}^{l'}(2q - z'_i)$$

Proof. To double the X-hyperedges, we choose pairs of X-hyperedges corresponding to the same triangle. This guarantees us at least one doubled hyperedge for every triangle at level 0. We double any remaining X-hyperedges arbitrarily.

We show by induction on i that we cover all indices with these hyperedges. The base case $i = 0$ is already done. If we have already covered all indices at level $i - 1$ then consider the hyperedges corresponding to the projection operators P_i and P'_i. All of these hyperedges go between a triangle at level $i - 1$ and a triangle at level i. We double pairs of these hyperedges which correspond to the same triangle at level i. This guarantees that for every triangle at level i, there is at least one doubled hyperedge corresponding to it. This hyperedge may not cover all three of the vertices of the triangle, but if it misses one, this one must be equal to a vertex at the level below which was already covered by assumption. We double the remaining hyperedges corresponding to the projection operators P_i and P'_i arbitrarily.

When performing this doubling, whenever the two hyperedges e_1 and e_2 have the same base value, we use the inequality $|e_1 e_2| \leqslant \frac{e_1^2 + e_2^2}{2}$. In the rare case when they have different base values, we use the inequality $|e_1 e_2| \leqslant \frac{v(e_2)}{2v(e_1)}e_1^2 + \frac{v(e_1)}{2v(e_2)}e_2^2$ to preserve the product of the base values.

Note that by this construction, for every triangle at level $i \geqslant 1$, there is a doubled hyperedge corresponding to some P_i or P'_i which goes between this triangle and a lower triangle, but we don't know which one.

We now describe our ordering on the hyperedges. To find this ordering, we consider the following multi-graph.

Definition B.16. We define the multi-graph G to have vertex set $V(G) = \cup_{ij}\{a_{ij}, b_{ij}, c_{ij}, a'_{ij}, b'_{ij}, c'_{ij}\}$ (with all equalities implied by the intersection pattern, the structure of the matrices A and B, and the choices for equalities or inequalities between the primed indices and unprimed indices.). We take the edges of G as follows. For all $i < l$ and for each distinct triangle (a_{ij}, b_{ij}, c_{ij}) or $(a'_{ij}, b'_{ij}, c'_{ij})$, we take the elements which do not appear in a higher level. If this is true for two of the three elements (which will be the case most of the time) we take the corresponding edge. If this is true for all three elements, we choose two of them to take as an edge, making this choice so that we take the same type of edge for all triangles at that level. If this is only true for one element, we take a loop on that element.

There are two cases for what happens with $i = l$

1. If $l' \neq l$, $b'_{lj} \neq b_{lj}$, or $c'_{lj} \neq c_{lj}$ then for every triangle (a_{lj}, b_{lj}, c_{lj}) we take the edge (b_{lj}, c_{lj}). If $l' = l$ then for every triangle $(a'_{lj}, b'_{lj}, c'_{lj})$ we take the edge (b'_{lj}, c'_{lj})

2. If $l' = l$, $b'_{lj} = b_{lj}$, and $c'_{lj} = c_{lj}$ then we take loops on every distinct element a_{lj}.

We analyze Δ in terms of this G. If we have a fixed budget of edges and want to maximize the number of vertices which we have, we want to have as many connected components as possible and we want each connected component to have the minimal number of edges. We define weights on the connected components of G measuring how far they are from satisfying these ideals.

Definition B.17. Given a connected component C of G, we define $w_{edge}(C)$ to be the number of non-loop edges it contains plus 1 minus the number of vertices it contains.

Definition B.18. Given a connected component C of G, we define $w_{triangle}(C)$ as follows

1. If C does not contain any b_{lj}, c_{lj}, $b'_{l'j}$, or $c'_{l'j}$ then we define $w_{triangle}(C)$ to be the number of distinct triangles whose corresponding edge in G is in C minus 1.

2. If C is the connected component containing b_{lj} and c_{lj} for all j then we set $w_{triangle}(C) = 0$

3. If C is the connected component containing $b'_{l'j}$ and $c'_{l'j}$ for all j then we set $w_{triangle}(C)$ to be the number of distinct $R_{l'}$ triangles $(a_{l'j}, b_{l'j}, c_{l'j})$ whose corresponding edge in G is in C.

4. If C is a connected component containing some $c'_{l'j}$ but no $b'_{l'j}$ (because all of the $b'_{l'j}$ appeared at a higher level) or vice versa, then we define $w_{triangle}(C)$ to be the number of distinct triangles whose corresponding edge in G is in C minus 1.

Definition B.19. We say that a vertex a_{ij} is bad if

1. The projector P_{i+1} involves the vertex a_{ij} (i.e. we do not have the constraint $a_{(i+1)j} = a_{ij}$ directly)

2. a_{ij} appears at a higher level.

We define badness similarly for the a', b, b', c, c' indices. Note that we could have a_i be bad while a'_i is not bad even if $a'_i = a_i$ (in fact this equality must be true in this case).

Lemma B.20.

$$\Delta \geqslant \sum_C (w_{edge}(C) + w_{triangle}(C)) + \sum_{i<l:a_{ij},b_{ij},\text{ or } c_{ij} \text{ is bad}} z_i + \sum_{i<l:a'_{ij},b'_{ij},\text{ or } c'_{ij} \text{ is bad}} z'_i$$

Proof. As discussed above, every time a connected component contains an extra edge above what it needs to be connected, this reduces the number of indices we can have by 1. Similarly, in the optimal case we have one connected component per triangle (with the exception of the R_l-triangles and perhaps the $R'_{l'}$-triangles), so every time a connected

component contains an extra triangle (or rather the edge corresponding to that triangle), this reduces the number of connected components by 1. For the remaining terms, note that if there are bad vertices, our previous bounds assumed that we would have new indices of that type but we do not. The resulting difference in the bounds is the corresponding z_i or z_i'. Note that this also works out in the special case that $a_i' = a_i$, $b_i' = b_i$, $c_i' = c_i$. Here we can view each a index as being half a_i and half a_i' and similarly for the b and c indices. □

With this lemma in hand, our strategy is as follows. We choose an ordering on the hyperedges so that each time we fail to have a free index, we can attribute it to one of the terms described above. We first preprocess our doubled hyperedges so that each hyperedge appears with multiplicity exactly 2. This requires bounding $\sum_{i=0}^{l} (2q - z_i) + \sum_{i=0}^{l'} (2q - z_i')$ doubled hyperedges with no free index. At this point, there is a one to one correspondence between our doubled hyperedges and edges of G. Note that this correspondence is somewhat strange, we only know that each edge in G is part of the upper level triangle for its corresponding hyperedge.

We now describe our procedures for ordering the hyperedges

Definition B.21. We say that a vertex v is an anchor for an edge e of G if either

1. $v, e \subseteq \{a_{ij}, b_{ij}, c_{ij}\}$ for some i and j and v appears at a higher level.

2. $v, e \subseteq \{a_{ij}', b_{ij}', c_{ij}'\}$ for some i and j and v appears at a higher level.

Definition B.22. For an anchor vertex v_{anchor}, define $E_i(v_{anchor})$ to be the set of all edges at level i which have v_{anchor} as an anchor vertex.

Definition B.23. We say that a vertex v or edge e is uncovered if it is not incident with any hyperedges between its level and the level above and covered otherwise. For a vertex v which is not part of G at level i, we say that v is uncovered at level i if there is no $j \geq 0$ such that v incident with a hyperedge between level $i + j$ and $i + j + 1$.

Definition B.24. We say that a vertex v is released at level i if there are no hyperedges remaining between level i and $i - 1$ whose upper and lower triangles both contain v.

Our main recursive procedure is as follows. We are considering a collection of connected component of the graph at level i where everything is uncovered except possibly for one edge e_r. If an edge e_r is covered and has anchor vertex v_{anchor} then we assume that this collection contains all of $E_i(v_{anchor})$ and that v_{anchor} is uncovered at level i.

We first consider the case when there are no bad vertices (we will consider the cases where we have bad vertices afterwards). If G contains a cycle, we can delete an edge and its corresponding hyperedge to break the cycle, accounting for this by decreasing $w_{edge}(C)$. Otherwise, unless C is just the single edge e_r, there must be a vertex v and edge e in C such that $e \neq e_r$ and e is the only edge incident with v.

We now consider the hyperedge corresponding to e. If v is part of this hyperedge then we can delete e and this hyperedge and continue. Otherwise, v must be an anchor vertex for many edges at the level below. Moreover, v is uncovered at level $i - 1$. We now consider $E_{i-1}(v)$. If $E_{i-1}(v)$ and everything connected to it is uncovered except for the edge

e'_r which is the bottom edge of the hyperedge corresponding to e, then we can apply our procedure recursively on $E_{i-1}(v)$ and everything connected to it. Otherwise, $E_{i-1}(v)$ must be connected to $E_{i-1}(v'_{anchor})$ for some other anchor vertex v'_{anchor} which has not yet been released at level i. Note that since there are no bad vertices, $E_{i-1}(v) \cap E_{i-1}(v'_{anchor}) = \emptyset$. Thus, there is a contribution of at least 1 to $w_{triangle}$ of one of these connected components from the connection between $E_{i-1}(v)$ and $E_{i-1}(v'_{anchor})$. Using this contribution, we can delete e and continue. After doing this, v is released at level i.

Remark B.25. Whenever we have a connection between E_{i-1} for two anchor vertices, we relase one of them at level i immediately after taking this connection into account. This ensures that we do not double count contributions to $w_{triangle}$.

If we are left with the single edge e_r then there are several cases. Letting v be the anchor vertex for e_r, if v goes down to the level below then consider the hyperedge corresponding to e_r and let e'_r be its bottom edge. Since we have deleted all edges in $E_i(v)$ except for e_r, either all of $E_{i-1}(v)$ except for e'_r is uncovered or $E_{i-1}(v)$ is connected to $E_{i-1}(v'_{anchor})$ for a different anchor vertex v'_{anchor} which has not yet been released at level i. In the first case, we can apply our recursive procedure on $E_{i-1}(v)$ and all edges connected to it. In the second case, we instead delete e_r as before and go back to the level above. Again, after doing this, v is now released at level i.

If v does not go down to the level below (or we are already at the bottom) then the hyperedge coresponding to e_r contains v. Moreover, by our assumption v is uncovered at level i. Thus, v is a free index for e_r so we can delete e_r and go back to the level above.

This procedure will succeed in the case that there are no bad vertices. We now handle bad vertices by reducing to the case where there are no bad vertices.

We consider the case where are below level l' and we do not haave that $a'_{ij} = a_{ij}$, $b'_{ij} = b_{ij}$, and $c'_{ij} = c_{ij}$. We will handle these cases separately.

If the a'_{ij} are bad vertices, this must be because of equalities $a'_{ij} = a_{ij}$. We handle this by replacing each a'_{ij} with a new vertex and running our procudure on this altered graph. This will cause failures when we try to use a'_{ij} or a_{ij} as a free index. That said, once we've tried to use all but one of a set of equal vertices, the final one will succeed, so the number of additional failures is at most z'_i. We can account for this using the term $\sum_{i<l:a'_{ij},b'_{ij}, \text{ or } c'_{ij} \text{ is bad}} z'_i$. We handle bad $a_{ij}, b_{ij}, b'_{ij}, c_{ij}, c'_{ij}$ vertices in a similar manner.

In the case that $a'_{ij} = a_{ij}$, $b'_{ij} = b_{ij}$, and $c'_{ij} = c_{ij}$, if the a'_{ij} and b_{ij} are bad vertices, this must be because of the equalities $a'_{ij} = a_{ij}$ and $b'_{ij} = b_{ij}$. We handle this by creating a new vertex for each a'_{ij}, having the hyperedges between levels i and $i+1$ use the old vertices, and having the hyperedges at lower levels use the new vertices. We modify G so that instead of loops at level i, the edges involve these new vertices. This makes it so that the only anchor vertices for edges at level i are the vertices $b'_{(i+1)j}$. Since all edges of G now have a unique anchor, the recursive procedure succeeds. We can accomplish this with the terms $\sum_{i<l:a_{ij},b_{ij}, \text{ or } c_{ij} \text{ is bad}} z_i + \sum_{i<l:a'_{ij},b'_{ij}, \text{ or } c'_{ij} \text{ is bad}} z'_i$.

We consider level l' separately. If the bottom of level l' contains bad vertices, we cannot make these vertices distinct. However, if this happens then we have loops in G for the bottom triangles at level l'. These triangles are distinct from the triangles on top at level l'.

We handle this by using $w_{triangle}$ to delete edges from G at this level so that each component contains at most one loop. When we run the procedure, we can use $w_{triangle}$ when $E_{i-1}(v)$ is connected to a loop as well as when it is connected to $E_{i-1}(v'_{anchor})$ for some other anchor vertex v'_{anchor} which has not been released at level i. This allows us to process each component of G at level i until we are left with either a covered edge or a single loop, both of which can be handled by our procedure. □

B.5   Counting intersection patterns and random partitioning

There are two pieces left to add. First, all of our analysis so far was for a given intersection pattern. We must sum over all intersection patterns.

Lemma B.26. *For all i, there are at most $\binom{2q}{z_i}(z_i)^{2q-z_i} \leqslant 2^{2q}(2q)^{2q-z_i}$ choices for which R_i-triangles are equal to each other.*

Proof. To specify a partition of the 2q R_i-triangles into z_i parts, we specify which triangles are distinct from all previous triangles. There are $\binom{2q}{z_i}$ choices for which triangles these are. For the remaining triangles, we specify which previous triangle they are equal to. There are at most $(z_i)^{2q-z_i}$ choices for this. □

In our bound, we can group this with the other factors corresponding to the R_i-triangles. Since we take q to be $O(log n)$, this is fine as m has a $log(n)$ factor.

Second, we justify our assumption that

1. For all $i \in [0, l']$, we either have that $a'_{ij} = a_{ij}$ for all $j \in [1, 2q]$ or $a'_{ij} \neq a_{ij}$ for all $j \in [1, 2q]$.

2. For all $i \in [0, l']$, we either have that $b'_{ij} = b_{ij}$ for all $j \in [1, 2q]$ or $b'_{ij} \neq b_{ij}$ for all $j \in [1, 2q]$.

3. For all $i \in [0, l']$, we either have that $c'_{ij} = c_{ij}$ for all $j \in [1, 2q]$ or $c'_{ij} \neq c_{ij}$ for all $j \in [1, 2q]$.

To achieve this, instead of looking at the entire matrix $\sum_a A_a \otimes B_a^T$, we split it into parts based on the equalities/inequalities we're looking at. To obtain the case where indices a and a' are always equal, we just restrict ourselves in $\sum_a A_a \otimes B_a^T$ to the terms where this is the case. To obtain the case where indices a and a' are never equal, we choose a random partition V, V^c of the indices and restrict ourselves in $\sum_a A_a \otimes B_a^T$ to the terms where $a \in V$ and $a' \in V^c$. If there are multiple indices that we wish to fork over, we apply this argument to each one (choosing the vertex partitions independently).

This construction has the property that if we take the expectation over all the possible vertex partitions, we obtain a constant times the part of $\sum_a A_a \otimes B_a^T$ we are interested in. Using this, it can be shown that probabilistic nrom bounds on these restricted matrices imply probabilistic norm bounds on the original matrix. For details, see Lemma 27 of "Bounds on the Norms of Uniform Low Degree Graph Matrices". From the above subsections, we have probabilistic norm bounds on the restricted matrices and the result follows.

B.6 Other Cross Terms

In this subsection, we sketch how the argument differs when $B = X$ rather than $B = \bar{R}_{\Omega_0} X$ or $B = P'_0 \bar{R}_{\Omega_0} X$.

Theorem B.27. *There is an absolute constant C such that for any $\alpha > 1$ and $\beta > 0$,*

$$Pr\left[|| \sum_a A_a \otimes X^T || > \alpha^{-(l+1)}\right] < n^{-\beta}$$

as long as

1. $r\mu \leqslant \min\{n_1, n_2, n_3\}$

2. $m > C\alpha\beta\mu^{\frac{3}{2}} r \sqrt{n_1} \max\{n_2, n_3\} log(\max\{n_1, n_2, n_3\})$

3. $m > C\alpha\beta\mu^2 r \max\{n_1, n_2, n_3\} log(\max\{n_1, n_2, n_3\})$

Proof sketch: The terms from X directly are

$$\prod_{j=1}^{2q} X_{a'_{0j} b'_{0j} c'_{0j}} = \prod_{j=1}^{2q} \left(\sum_{i_j} u_{i_j a'_{0j}} v_{i_j b'_{0j}} w_{i_j c'_{0j}} \right)$$

Note that the R_Ω factors are completely independent of the b'_{0j} and c'_{0j} indices. Thus, we can sum over the b'_{0j} and c'_{0j} indices first. When we do, this zeros out all terms except the ones where all of the i_j are equal. Moreover, all of the v and w terms sum to 1. The u_{i_j} terms can be bounded by $\left(\frac{\mu}{n_1}\right)^q$. We now compare the bound we had before with the bound we have here.

For $\bar{R}_{\Omega_0} X$ we had factors

1. $\left(\frac{r\mu^3}{n_1 n_2 n_3}\right)^{2q}$ from the X-hyperedges.

2. $\frac{n_2 n_3}{\mu^2} \left(\frac{n_1 (\max\{n_2, n_3\})^2}{\mu^3}\right)^q$ from the q potential a indices and $2q + 2$ potential b or c indices.

3. $\left(\frac{n_1 n_2 n_3}{m}\right)^{2q}$ from the $2q$ potential distinct triangles.

4. $\left(r \cdot \frac{\mu^{\frac{3}{2}}}{\sqrt{n_1} \max\{n_2, n_3\}} \cdot \frac{n_1 n_2 n_3}{m}\right)^{2q-z_l-z'_l} \leqslant \left(\frac{r\mu^{3/2} \sqrt{n_1} \max\{n_2, n_3\}}{m}\right)^{2q-z_l-z'_l}$ from the actual number of distinct triangles, the corresponding reduced maximum number of potential new indices, and the factors of r which we take from $r^{\#}$ of doubled hyperedges we bound with no free index

5. $\left(\frac{\mu}{n_1}\right)^{\Delta_{ll'a}} \left(\frac{\mu}{\max\{n_2, n_3\}}\right)^{\Delta_{lbc}+\Delta'_{l'bc}} \leqslant \left(\frac{\mu}{\min\{n_1, n_2, n_3\}}\right)^{\Delta_{ll'a}+\Delta_{lbc}+\Delta'_{l'bc}}$ from the actual number of new indices which we have

We now have the following factors instead:

1. $(\frac{r\mu^3}{n_1 n_2 n_3})^q r(\frac{\mu}{n_1})^q$ from the X-hyperedges.

2. $\frac{\max\{n_2,n_3\}}{\mu} \left(\frac{n_1 \max\{n_2,n_3\}}{\mu^2}\right)^q$ from the q potential a indices and $q+1$ potential b or c indices.

3. $\left(\frac{n_1 n_2 n_3}{m}\right)^q$ from the q potential distinct triangles.

4. $\left(r \cdot \frac{\mu^2}{n_1 \max\{n_2,n_3\}} \cdot \frac{n_1 n_2 n_3}{m}\right)^{q-z_l} \leqslant \left(\frac{r\mu^2 \max\{n_2,n_3\}}{m}\right)^{q-z_l}$ from the actual number of distinct triangles, the corresponding reduced maximum number of potential new indices, and the factors of r which we take from $r^{\#\text{ of doubled hyperedges we bound with no free index}}$

5. $\left(\frac{\mu}{n_1}\right)^{\Delta_{ll'a}} \left(\frac{\mu}{\max\{n_2,n_3\}}\right)^{\Delta_{lbc}+\Delta'_{l'bc}} \leqslant \left(\frac{\mu}{\min\{n_1,n_2,n_3\}}\right)^{\Delta_{ll'a}+\Delta_{lbc}+\Delta'_{l'bc}}$ from the actual number of new indices which we have

The difference is in the first three terms, grouping these terms together gives

$$\frac{r\max\{n_2,n_3\}}{\mu}\left(\frac{r\mu^2\max\{n_2,n_3\}}{m}\right)^q$$

By our assumption, $m \geqslant Cr\mu^2\max\{n_2 n_3\}log(n)^2$ so we are fine. $\qquad\square$