New Tools for Smoothed Analysis: Least Singular Value Bounds for Random Matrices with Dependent Entries

Aditya Bhaskara
University of Utah
Salt Lake City, USA
bhaskaraaditya@gmail.com

Eric Evert
Northwestern University
Evanston, USA
eric.evert@northwestern.edu

Vaidehi Srinivas
Northwestern University
Evanston, USA
vaidehi@u.northwestern.edu

Aravindan Vijayaraghavan
Northwestern University
Evanston, USA
aravindv@northwestern.edu

ABSTRACT

We develop new techniques for proving lower bounds on the least singular value of random matrices with limited randomness. The matrices we consider have entries that are given by polynomials of a few underlying base random variables. This setting captures a core technical challenge for obtaining smoothed analysis guarantees in many algorithmic settings. Least singular value bounds often involve showing strong anti-concentration inequalities that are intricate and much less understood compared to concentration (or large deviation) bounds.

First, we introduce a general technique for proving anti-concentration that uses well-conditionedness properties of the Jacobian of a polynomial map, and show how to combine this with a hierarchical ε-net argument to prove least singular value bounds. Our second tool is a new statement about least singular values to reason about higher-order lifts of smoothed matrices and the action of linear operators on them.

Apart from getting simpler proofs of existing smoothed analysis results, we use these tools to now handle more general families of random matrices. This allows us to produce smoothed analysis guarantees in several previously open settings. These new settings include smoothed analysis guarantees for power sum decompositions and certifying robust entanglement of subspaces, where prior work could only establish least singular value bounds for fully random instances or only show non-robust genericity guarantees.

CCS CONCEPTS

• **Theory of computation → Design and analysis of algorithms**; **Randomness, geometry and discrete structures**.

KEYWORDS

smoothed analysis, tensors, random matrices, least singular values, matrix anticoncentration, unsupervised learning, quantum entanglement

ACM Reference Format:

Aditya Bhaskara, Eric Evert, Vaidehi Srinivas, and Aravindan Vijayaraghavan. 2024. New Tools for Smoothed Analysis: Least Singular Value Bounds for Random Matrices with Dependent Entries. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC '24), June 24–28, 2024, Vancouver, BC, Canada.* ACM, New York, NY, USA, 31 pages. https://doi.org/10.1145/3618260.3649765

1 INTRODUCTION

Over the past two decades, there has been significant progress in using algebraic methods for high-dimensional statistical estimation (e.g., [2]). Techniques like tensor decomposition have been used for parameter estimation in mixture models [3, 10, 14], shallow neural networks [5, 25], stochastic block models [2], and more [26]. Recently, more sophisticated decomposition methods based on tensor networks [21], circuit complexity [12] and algebraic geometry [12, 19] have given to rise to new algorithms for many problems in high-dimensional geometry and parameter estimation. These algorithms start by building appropriate algebraic structures that "encode" the hidden parameters of interest. Then, they use the algebraic techniques described above for recovering the solution.

Unfortunately, in most of these applications, the recovery problem turns out to be NP hard in general. So the algorithms have provable recovery guarantees only under certain *algebraic* conditions. Typically, these conditions can be formulated in terms of appropriately defined matrices being well-conditioned, i.e., having a non-negligible least singular value. Furthermore, the least singular value determines the sample complexity and running time, and so it is important to obtain inverse polynomial bounds.

Now it is natural to ask: *do the algebraic conditions typically hold?* Due to NP hardness, we know there exist parameters for which the conditions do not hold. But how common or rare are such parameter

STOC '24, June 24–28, 2024, Vancouver, BC, Canada
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0383-6/24/06
https://doi.org/10.1145/3618260.3649765

settings/instances? A strong way to address this question is via the framework of smoothed analysis, developed in the seminal work of Spielman and Teng [23, 27, 28]. A condition is said to hold in a smoothed analysis setting if for *any* instance, a small random perturbation of magnitude, say $\rho = 1/n^2$, where n is the input size, results in an instance that satisfies the condition with high probability. Smoothed analysis guarantees show that any potential bad instance is isolated or degenerate: most other instances in a small ball around it have good guarantees. On the one hand, smoothed analysis gives a much stronger guarantee than *average case analysis*, where one shows that the condition holds w.h.p. for a random choice of parameters from some distribution. On the other hand, it provides quantitative, robust analogs of *genericity* results in algebraic settings, which are needed in most algorithmic applications.

Considering the flavor of the algebraic non-degeneracy conditions, the problem of smoothed analysis boils down to the following: *given a matrix $\mathcal{M}$ whose entries are functions (typically polynomials) of some base variables, does randomly perturbing the variables result in $\mathcal{M}$ having a non-negligible least singular value with high probability?*

This question is non-trivial even in very specialized settings, as it is a statement about anti-concentration — a topic that is less understood in probability theory than concentration or large deviation bounds. For example when the underlying variables form a matrix $U \in \mathbb{R}^{n\times m}$, the structured matrix $\mathcal{M} = U \odot U = \big(u_i \otimes u_i\big)_{i\in[m]}$,[1] represents the Khatri-Rao product, and has been the subject of much past work [4, 9, 11] that developed intricate arguments specialized for this setting. Least singular value bounds of $\mathcal{M} = \widetilde{U} \odot \widetilde{U}$ for randomly perturbed $\widetilde{U}$ have lead to smoothed analysis guarantees for several problems including tensor decomposition [9], recovering assemblies of neurons [4], parameter estimation of latent variable models like mixtures of Gaussians [13], hidden Markov models [11], independent component analysis [15] and even learning shallow neural networks [5]. Another approach is to use concentration bounds to prove lower bounds on the least singular value [7, 22, 29?] for analyzing random instances; these techniques based on concentration bounds cannot handle smoothed instances. We lack a broader toolkit that allows us to analyze more general classes of random matrices that arise in many other smoothed analysis settings of interest.

Consider, for example the symmetric lift of the matrix $\widetilde{U}$ represented by

$$\widetilde{U}^{\circledast 2} := ((\tilde{u}_i \otimes \tilde{u}_j + \tilde{u}_j \otimes \tilde{u}_i) : 1 \le i \le j \le m),$$

where the columns (up to reshaping) give a basis for the space of all the symmetric matrices that are supported on the subspace $\tilde{U}$. Here $\circledast$ denotes the symmetrized Kronecker product.

Question 1.1. For a linear operator Φ acting on the space of symmetric $n \times n$ matrices (e.g., a projection matrix), can we obtain an inverse polynomial lower bound with high probability on the least singular value of the matrix

$$\mathcal{M} = \Phi(\widetilde{U}^{\circledast 2}) = \Big(\Phi(\tilde{u}_i \otimes \tilde{u}_j + \tilde{u}_j \otimes \tilde{u}_i) : 1 \le i \le j \le m\Big),$$

when $m \le cn$ for a sufficiently small $c \in (0, 1)$?

The new techniques developed in this paper, to our knowledge, give the first inverse polynomial lower bound on the least singular value of $\mathcal{M}$, and its higher order generalizations; see Theorem 1.4. As it turns out, this already captures the Khatri-Rao product $\widetilde{U} \odot \widetilde{U}$ setting as a special case by setting $m = 1$ and Φ appropriately. One interpretation of the statement is that $\widetilde{U} \circledast \widetilde{U}$ acts like "truly random" subspace in the lifted space $\mathsf{Sym}(\mathbb{R}^n \otimes \mathbb{R}^n)$ with the same dimension. With high probability, a random subspace of $\mathsf{Sym}(\mathbb{R}^n \otimes \mathbb{R}^n)$[2] with dimension $o(n^2)$ will not contain any vector near the kernel of Φ. The affirmative answer to the above question shows that the lifted space that corresponds to column space of $(\widetilde{U})^{\circledast 2}$ behaves similarly and is far from the kernel of Φ! In other words, it is rotationally well-spread; it is not too aligned with any specific subspace. Note that $\widetilde{U}$ only has about nm truly independent coordinates or "bits", whereas a random subspace of the same dimension has $c \cdot n^2 m^2$ independent coordinates. Hence the lift $\mathcal{U}^{\circledast 2}$ of a smoothed subspace $\mathcal{U}$ acts "pseudorandom" – it acts like a random subspace in the lifted space with respect to all linear operators of reasonable rank.

Matrices of this flavor arise in open questions about the smoothed analysis of various algebraic algorithms for problems like robust certification of quantum entanglement in subspaces, certifying distance from varieties [19], and decomposition into sums of powers of polynomials [7, 12]. Specifically, rank-1 matrices (of unit norm) correspond to separable or non-entangled states in bipartite quantum systems. For a certain specific choice of Φ, the positive resolution of Question 1.1 certifies that a smoothed subspace of $n_1 \times n_2$ matrices of dimension cn_1n_2 (for some $c > 0$) is far from any rank-1 matrix of unit norm. Moreover, in the recent algebraic algorithms of [7, 12], they consider generic or random subspaces $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_t \subset \mathbb{R}^n$ and they need to argue that the corresponding dth order lifts $\mathcal{U}_1^{\circledast d}, \mathcal{U}_2^{\circledast d}, \dots, \mathcal{U}_t^{\circledast d}$ are far from each other.

Our results give a novel and modular way to analyze such matrices. Our contributions are two fold:

- We give new tools for proving least singular value lower bounds via ε-nets. This involves identifying a key property that is sufficient for carrying forth net based arguments, and giving a new tool for proving such a property.
- We consider matrices that have the structure of a linear operator applied to higher-order lifts corresponding to the Kronecker product, and give new techniques to reason about the least singular value. This resolves open questions raised in [7, 12, 19].

1.1 Our Results

1.1.1 Hierarchical Nets. Our first set of results focus on ε-net based arguments for proving bounds for least singular values. Suppose we have a random matrix $\mathcal{M}$, the idea is to consider a fixed "test" vector α, prove that $\|\mathcal{M}\alpha\|$ is large enough with high probability, and then take a union bound over "all possible vectors α". As the set of candidate α is infinite, the idea is to take a fine enough net over possible vectors α. The challenge when dealing with structured matrices (of the kind discussed above) is that for a single test vector

[1] Here, $\otimes$ represents the standard tensor product or Kronecker product.

[2] $\mathsf{Sym}(\mathbb{R}^n \otimes \mathbb{R}^n)$ is the space of all symmetric $n \times n$ matrices.

α, we do not obtain a sufficiently strong probability guarantee. This is because the individual columns of $\mathcal{M}$ may not have "sufficient randomness", and since we do not know how α spreads its mass across columns, the bound will be weak. Our main observation is that in the matrices we consider for our application, as long as α is *well spread*, we can obtain a much stronger bound. We refer to this as a "combination amplifies anticoncentration" (CAA) property of $\mathcal{M}$.

CAA Property (Informal Definition). We say that $\mathcal{M}$ has the CAA property if for every $k \geq 1$, for any test vector α that has k entries of magnitude $\geq \delta$, we have that $\|\mathcal{M}\alpha\| \geq \Omega(\delta)$, with probability $1 - \exp(-\omega(k))$.

Formally, to capture the $\omega(k)$ term, we have a parameter β. See Definition 4.1 for details. Our first result is that for any matrix with this property, we have a bound on $\sigma_{\min}(\mathcal{M})$.

Informal Theorem 1.2. *Suppose $\mathcal{M}$ is a random matrix with m columns and that $\mathcal{M}$ satisfies the CAA property with parameter $\beta > 0$. Then with high probability (indeed, exponentially small probability of failure), we have $\sigma_{\min}(\mathcal{M}) > \mathrm{poly}(1/m)$. (See Theorem 4.2 for the formal statement.)*

The proof uses a novel ε-net construction. Nets that use structural properties of the test vector α have been used in prior works in the context of proving least singular value bounds, notably in the celebrated work of Rudelson and Vershynin [24]. In proving our result, the natural approach of constructing a hierarchy of nets based on increasing k (and using some threshold δ) does not work. Informally, this is because the error from ignoring terms that are slightly smaller than δ can add up significantly, causing the argument to fail. We introduce a new hierarchical construction that overcomes this problem.

The next question we consider is how to prove that the CAA property holds in a particular context. This can be shown via a direct argument when $\mathcal{M}$ is simple, e.g., a random matrix with independent entries. However, for matrices with more structured entries, it can need a careful analysis. To handle this, we develop a new tool for proving anticoncentration that we believe is of independent interest.

1.1.2 Anti-concentration of a Vector of Polynomials.

Consider $P(x) := (p_1(x), p_2(x), \dots, p_N(x))$, where each p_i is a polynomial of n "base" random variables. Suppose we wish to show anti-concentration bounds for $P(\tilde{x})$, where $\tilde{x}$ is a perturbation of some x (i.e., we wish to bound the probability that $P(\tilde{x})$ is within a small ball of a point y is small, for all y). One hope is to use a coordinate-wise bound (e.g., using known results like [30]) and take the product over $1, 2, \dots, N$. It is easy to see that this is too good to be true: consider an example where p_i are all equal; here having N coordinates is the same as having just one. So we need a good metric for "how different" the polynomials p_i are for a *typical* x. We capture this notion using the Jacobian of the polynomial map P. Recall that in this case, the Jacobian $J(x)$ is a matrix with one column per p_i, containing the vector of partial derivatives, $\nabla p_i(x)$.

Jacobian rank property (Informal Definition). We say that $P(x)$ has the Jacobian rank property if for every x, at a slightly perturbed point $\tilde{x}$, $J(\tilde{x})$ has at least k singular values that are *large enough* (where k is a parameter).

We refer to Definition B.1 for the formal statement. Our result here is that this property implies anticoncentration:

Informal Theorem 1.3. *Suppose $P(x)$ defined as above satisfies the Jacobian rank property with parameter k. Then for a perturbation of any point x, we have that $\forall y, \mathbb{P}[\|P(\tilde{x}) - y\| < \varepsilon] < \exp(-\Omega(k))$. (Here, ε is a quantity that depends on the dimensions, k, the perturbation, and the singular value guarantee; see Theorem 4.7 for the formal statement.)*

Intuitively, the Jacobian having several large singular values must result in anticoncentration (because $P(x)$ locally behaves linearly). However, the challenging aspect is that the Jacobian need not *always* have many large singular values. Our assumption (Jacobian rank property) is itself made for a perturbed vector, i.e., we assume that $J(\tilde{x})$ has many high singular values with high probability. Further, the magnitude of these singular values will depend on the perturbation: if a "bad" x was perturbed by ρ, $J(\tilde{x})$ will have most of the large singular values being $\approx \rho$. Dealing with this issue turns out to be the main challenge in proving the theorem (see Theorem 4.7 for a formal statement).

As an application of the Jacobian rank method, we re-prove the main result of [9] and [4]. They consider random matrices $\mathcal{M}$ where the ith column is $\tilde{u}_i \otimes \tilde{v}_i$, and $\tilde{u}_i, \tilde{v}_i$ are perturbed vectors in $\mathbb{R}^n$. We show that this $\mathcal{M}$ satisfies the CAA property, and thus our first result (above) implies a condition number lower bound. In order to prove the CAA property, we consider a combination of the columns $\sum_i \alpha_i(\tilde{u}_i \otimes \tilde{v}_i)$ and prove that if α has k entries $\geq \delta$, then the Jacobian has $nk/2$ large singular values. Using our second result, we obtain a strong anticoncentration bound, thus completing the proof. This technique also lets us tackle Question 1.1 described above, but in what follows, we describe a different technique that also generalizes to higher orders.

1.1.3 Structured Matrices from Kronecker Products.

Next, we consider a general class of structured matrices that are obtained by taking the symmetrized Kronecker product of some ρ-perturbation $\tilde{U}$ of an underlying matrix U and applying a linear operator Φ. Here, $\tilde{U}$ is a ρ-perturbation of U means $\tilde{U} = U + \mathcal{N}(0, \rho^2)$. In other words, the matrix of interest is $\mathcal{M} = \Phi\widetilde{U}^{\circledast d}$, where d is a constant. For such a matrix, we can ask the question: are there conditions on Φ under which we can prove that $\sigma_{\min}(\mathcal{M})$ is large, with high probability over the perturbation? We provide an affirmative answer to this question in terms of the rank of Φ.

This question captures a variety of settings studied previously. For example, [11] studies matrices $\mathcal{M}$ whose columns are tensor products of some underlying vectors (i.e., the columns have the form $u_{i_1} \otimes u_{i_2} \otimes \cdots \otimes u_{i_d}$). This turns out to be a special case of our setting above. Likewise, in the work of [7], one of the matrices they consider is an $\mathcal{M}$ formed by concatenating the Kronecker products of a collection of underlying matrices, and the analysis of their algorithm relies on $\sigma_{\min}(\mathcal{M})$ being non-negligible. This also falls into our setting by choosing Φ appropriately (as we show in Corollary 5.3). Finally, as we discuss in our applications, the setting $\mathcal{M} = \Phi\widetilde{U}^{\circledast d}$ also directly appears in the work of [19].

The following is an informal statement of our result. $\mathsf{Sym}_d(\mathbb{R}^n)$ will refer to a symmetrization of $(\mathbb{R}^n)^{\otimes d}$.[3] Also, as before, $\sigma_{\min}$ corresponds to right singular vectors.

Informal Theorem 1.4. *Suppose Φ is a matrix of rank $\delta\binom{n+d-1}{d}$ for some constant $\delta > 0$, and let U be any $n \times m$ matrix. Let $\widetilde{U}$ be a ρ-perturbation of U. Then as long as $m \leq cn$ for some constant c, we have $\geq 1 - \exp(-\Omega(n))$,*

$$\sigma_{\min}(\Phi\widetilde{U}^{\circledast d}) \geq \mathrm{poly}\left(\rho, \frac{1}{n}\right).$$

(See Theorem 5.1 for a formal statement.)

Note that the above Theorem 1.4 with $d = 2$ answers Question 1.1 affirmatively. It also proves a similar statement about how the column space of a dth order lift $\widetilde{U}^{\circledast d}$ behaves like a random subspace of the lifted space of the same dimension with respect to linear operators in the lifted space of reasonable rank, even though we have only dnm random "bits" as opposed to $\Omega_d((mn)^d)$. As we describe in Section 2, the proof relies on first moving to non-symmetric products via a new decoupling argument. In the case of non-symmetric products, we end up having to analyze the least singular value of a matrix of the form $\Phi(\widetilde{U}^{(1)} \otimes \widetilde{U}^{(2)} \otimes \cdots \otimes \widetilde{U}^{(d)})$. This can be interpreted as a "modal contraction" (or dimension reduction of the mode) defined by $\{\widetilde{U}^{(i)}\}$ applied to the tensor Φ. We then show how to analyze such *smoothed modal contractions*, which ends up being one of our technical contributions (see Section 2.3 and Theorem 5.2).

1.1.4 Applications.

Certifying distance from variety and quantum entanglement. Our first application is to the problem of certifying that a variety is "far" from a *generic* linear subspace. As a simple motivation, suppose we have a linear subspace $\mathcal{X}$ of dimension δn in $\mathbb{R}^n$ (assume $\delta < 1/2$). Then for a randomly ρ-perturbed subspace $\widetilde{\mathcal{U}}$ of dimension $< n/2$, we can show that the two spaces have no overlap in a strong sense: every unit vector $u \in \mathcal{X}$ is at a distance $\Omega(\rho)$ from $\widetilde{\mathcal{U}}$. It is natural to ask if a similar statement holds when $\mathcal{X}$ is an algebraic variety (as opposed to a subspace). This problem also has applications to quantum information (see [19] and references therein). Furthermore, we can ask if there is an efficient algorithm that can *certify* that every unit vector in $\mathcal{X}$ is far from $\widetilde{\mathcal{U}}$.

We answer both these questions in the affirmative.

Informal Theorem 1.5. *Suppose $\mathcal{X} \subset \mathbb{R}^n$ is an irreducible variety cut out by $\delta\binom{n+d-1}{d}$ homogeneous degree d polynomials. There exists a $c > 0$ such that for any ρ-perturbed subspace $\widetilde{\mathcal{U}}$ of dimension at most cn, with probability $1 - \exp(-\Omega(n))$, every unit vector in $\mathcal{X}$ has distance $\geq$ poly $\left(\rho, \frac{1}{n}\right)$ to $\widetilde{\mathcal{U}}$. Further, this can be certified by an efficient algorithm. (See Theorem D.1 for the formal statement.)*

The recent work of [19] gave an algorithm that we also use, but our new least singular value bounds imply the quantitative distance lower bound stated above. Applying this theorem with the variety of rank-1 matrices gives the following direct corollary.

Corollary 1.6. *There is a polynomial time algorithm that given a random ρ-perturbed subspace $\widetilde{\mathcal{U}}$ of $n_1 \times n_2$ matrices of dimension $m \leq cn_1n_2$ (for some universal constant $c > 0$) certifies w.h.p. that $\widetilde{\mathcal{U}}$ is at least $\mathrm{poly}(\rho, 1/n)$ far from every rank-1 matrix of unit norm.*

The above theorem also has a direct implication to robustly certifying entanglement of different kinds, which we describe in Section D.

Decomposing sums of powers of polynomials. Our second application is to the problem of "decomposing power sums" of polynomials, a question that has applications to learning mixtures of distributions. In the simplest setting, [12] and [7] consider the following problem: given a polynomial $p(\mathbf{x})$ that can be expressed as

$$p(\mathbf{x}) = \sum_{t \in [m]} a_t(\mathbf{x})^3 + e(\mathbf{x})$$

where a_t are quadratic polynomials and $e(\mathbf{x})$ is a small enough error term, the goal is to recover $\{a_t(\mathbf{x})\}_{t\in[m]}$.[4] The work of [7] gave an algorithm for this problem, but their analysis relies on certain *non-degeneracy* conditions, which can be formulated as a lower bound on the least singular value of appropriate matrices. They prove that these conditions hold if the instances (i.e., the polynomials a_t) are *random*, using the machinery of graph matrices [1]. However, the question of obtaining a smoothed analysis guarantee is left open. As discussed earlier, a smoothed analysis guarantee is much stronger than a guarantee for random instances, as it shows that even in the neighborhood of hard instances, most instances are easy.

Their analysis requires least singular value bounds for various matrices that arise from higher order lifts and polynomials of some underlying random variables. For example, they require least singular value bounds on matrices of the form $\Phi(\tilde{U}^{\circledast 3})$, for a specific symmetrization operator Φ that acts on the lifted space. Another type of matrix that they analyze are *block Kronecker products*, of the form $V = [\tilde{U}_1^{\circledast 2} \dots \tilde{U}_m^{\circledast 2}]$ that arise from different partial derivatives.[5] These kinds of matrices are ideal candidates for our techniques.

Informal Theorem 1.7. *For the matrices $\mathcal{M}$ arising in the analysis of [7], a ρ-perturbation of the parameters of a_t results in $\sigma_{\min}(\mathcal{M}) \geq \mathrm{poly}(\rho, 1/n)$, with probability $1 - \exp(-\mathrm{poly}(m, n))$. (This corresponds the formal statements of propositions E.1, E.2, and E.3.)*

These least singular bounds allow us to conclude that the algorithm of [7] indeed has a smoothed analysis guarantee. In Section E, we outline the algorithm of [7], identify the different non-degeneracy conditions required and show that each of these conditions holds for *smoothed*/perturbed polynomials a_t. Interestingly, we can avoid

[3] The latter can be viewed as having a coordinate for all "ordered" monomials of degree d in n variables (e.g., x_ix_j and x_jx_i correspond to different coordinates), while the former collects the terms with the same product. See Section 3 for a formal description.

[4] This corresponds to the setting $K = 2, D = 1$ in their framework. We focus only on this setting, as it turns out to be representative of their techniques.

[5] The actual matrix is slightly different, and is described in detail in Section E.

the technically heavy machinery of graph matrices, while obtaining stronger (smoothed) results. We hope our new techniques can also help obtain smoothed analysis guarantees for other algebraic methods like the framework of [12].

2 PROOF OVERVIEW AND TECHNIQUES

2.1 Improved Net Analyses

ε-Nets and limitations. The classic approach to proving least singular value bounds is an ε-net argument. The argument proceeds by trying to prove that $\|\mathcal{M}\alpha\|$ is large for all α in the unit sphere. It does so by constructing a fine "net" over points in the sphere with the properties that (a) the net has a small number of points, and hence a union bound can establish the desired bound for points in the net, and (b) for every other point α in the sphere, there is a point α' in the net that is close enough, and hence the bound for α' "translates" to a bound for α. However, in settings where the columns $\widetilde{X}_i$ of $\mathcal{M}$ have "limited randomness", this approach cannot be applied in many parameter regimes of interest. The simplest example is one where each $\widetilde{X}_i$ is of the form $\tilde{u}_i \otimes \tilde{u}_i$, where $\tilde{u}_i \in \mathbb{R}^n$ and we have around $m = n^2/4$ such vectors. In this case, (a) above causes a problem: the size of a net for unit vectors in a sphere in $\mathbb{R}^m$ is $\exp(m) = \exp(n^2/4)$. This is much too big for applying a union bound, since each column only has "n bits" of randomness, so the failure probability we can obtain for a general α is $\exp(-n)$. For this specific example, the works [4, 9] overcome this limitation by considering more ad-hoc methods for showing least singular value bounds, not based on ε-nets.

Main idea from Section 4.1. As described above, the limited randomness in each column $\widetilde{X}_i$ limits the probability with which we can show that $\mathbb{P}[\|\mathcal{M}\alpha\|]$ is large. However, we observe that in many settings, as long as we consider an α that is *spread out*, we can show that $\mathbb{P}[\|\mathcal{M}\alpha\|]$ is large with a significantly better probability. Informally, in this case, the randomness across many different columns gets "accumulated", thus amplifying the resulting bound. We refer to this phenomenon as *combination amplifies anticoncentration* (CAA) (described informally in Section 1.1; see Definition 4.1). Our first theorem states that the CAA property automatically implies a lower bound on $\sigma_{\min}(\mathcal{M})$ with high probability.

To outline the proof of the theorem, let us consider some unit vector $\alpha \in \mathbb{R}^m$. If α has say $m/2$ "large enough" entries, then the CAA property implies that $\|\mathcal{M}\alpha\|$ is non-negligible with probability $1 - \exp(-m)$ (roughly), and so we can take a union bound over a (standard) ε-net, and we would be done. However, suppose α had only k entries that are large enough (defined as $> \delta$ for some threshold), and $k \ll m$. In this case, the CAA property implies that $\|\mathcal{M}\alpha\| \geq c\delta$ with probability roughly $1 - \exp(-k)$. While this is large enough to allow a union bound over just the large entries of α (placing a zero in the other entries), the problem is that there can be *many* entries in α that are just slightly smaller than δ. In this case, having $\|\mathcal{M}\alpha_{\geq\delta}\| \geq c\delta$ (where $\alpha_{\geq\delta}$ is the vector α restricted to the entries $\geq \delta$ in magnitude, and zeros everywhere else) does not let us conclude that $\|\mathcal{M}\alpha\| > 0$, unless c is very large. Since we cannot ensure that c is large, we need a different argument. The idea will be to use the fact that our definition of the CAA comes with a slack parameter β. In particular, for α as above with k values of magnitude $\geq \delta$, it allows us to take a union bound over $k \cdot m^{\beta}$ parameters. Thus, if we knew that there are at most $k \cdot m^{\beta}$ entries that are "slightly smaller" (by a factor roughly θ) than δ, we can include them in the ε-net. Defining θ appropriately, we can ensure that the problem described above (where the slightly smaller entries cancel out the $\mathcal{M}\alpha_{\geq\delta}$) does not occur. The problem now is when α has $> k \cdot m^{\beta}$ entries of magnitude between $\theta\delta$ and δ. While this is indeed a problem for this value of δ, it turns out that we can try to work with $\theta\delta$ instead. Now the problem can recur, but it cannot recur more than $(1/\beta)$ times (because each time, k grows by an m^{β} factor). This allows to define a hierarchical net, which helps us identify the threshold δ for which the ratio of the number of entries $\geq \theta\delta$ and $\geq \delta$ is smaller than m^{β}.

By carefully bounding the sizes of all the nets and setting θ appropriately, Theorem 4.2 follows.

2.2 Jacobian Based Anticoncentration

As described in Section 1.1, proving smoothed analysis bounds often requires dealing with a vector of polynomials

$$P(x) = (p_1(x), \ldots, p_N(x))$$

in some underlying variables x. The goal is to show that for every x, evaluating P at a ρ-perturbed point $\tilde{x}$ gives a vector that is not too small in magnitude. (A slight generalization is to show that $P(\tilde{x})$ is not too close to any fixed y.)

We first observe that such a statement is not hard to prove if we know that the Jacobian $J(x)$ of $P(x)$ has many large singular values at *every* x, and if the perturbation ρ is small enough. This is because around the given point x, we can consider the linear approximation of $P(\tilde{x})$ given by the Jacobian. Now as long as the perturbation has a high enough projection onto the span of the corresponding singular vectors of $J(x)$, $P(\tilde{x})$ can be shown to have desired anticoncentration properties (by using the standard anticoncentration result for Gaussians). Finally, if $J(x)$ has k large singular values, a random ρ-perturbation will have a large enough projection to the span of the singular vectors with probability $1 - \exp(-k)$.

Now, in the applications we are interested in, the polynomials P tend to have the Jacobian property above for "typical" points x, but not all x. Our main result here is to show that this property suffices. Specifically, suppose we know that for every x, the Jacobian at a ρ perturbed point has k singular values of magnitude $\geq c\rho$ with high probability. Then, in order to show anticoncentration, we view the ρ perturbation of x as occurring in two independent steps: first perturb by $\rho\sqrt{1 - z^2}$ for some parameter z, and then perturb by ρz. The key observation is that for Gaussian perturbations, this is identical to a ρ perturbation!

This gives an approach for proving anticoncentration. We use the fact that the first perturbation yields a point with sufficiently many large Jacobian singular values with high probability, and combine this with our earlier result (discussed above) to show that if z is small enough, the linear approximation can indeed be used for the second perturbation, and this yields the desired anticoncentration bound.

Applications. The simplest application for our framework is the setting where $\mathcal{M}$ has columns being $\tilde{u}_i \otimes \tilde{v}_i$, for some ρ-perturbations of underlying vectors u_i, v_i. (This setting was studied in [4, 9] and already had applications to parameter recovery in statistical models.) Here, we can show that $\mathcal{M}$ has the CAA property. To show this, we consider some combination $\sum_i \alpha_i(\tilde{u}_i \otimes \tilde{v}_i)$ with k "large" coefficients in α, and show that in this case, the Jacobian property holds. Specifically, we show that the Jacobian has $\Omega(kn)$ large singular values. This establishes the CAA property, which in turn implies a lower bound on $\sigma_{\min}(\mathcal{M})$. This gives an alternative proof of the results of the works above.

2.3 Structured Matrices from Kronecker Products and Higher-order lifts

Our second set of techniques allow us to handle structured matrices that arise from the action of a linear operator on Kronecker products, as described in Question 1.1. For simplicity let us focus on the setting when $d = 2$, and let $\Phi : \mathrm{Sym}(\mathbb{R}^n \otimes \mathbb{R}^n) \to \mathbb{R}^k$ be an (orthogonal) projection matrix of rank $R \geq 0.01n^2$ acting on the space of symmetric matrices $\mathrm{Sym}(\mathbb{R}^n \otimes \mathbb{R}^n)$ (in general Φ can also be any linear operator of large rank). Let $m = o(n)$ and $\widetilde{U} \in \mathbb{R}^{n\times m}$ be a small random ρ-perturbation of arbitrary matrix $U \in \mathbb{R}^{n\times m}$. The columns of the matrix $\widetilde{U}^{\circledast 2}$ are linearly independent with high probability, and span the symmetric lift of the column space of $\widetilde{U}$. An arbitrary subspace of $\mathrm{Sym}(\mathbb{R}^n \otimes \mathbb{R}^n)$ of the same dimension may intersect non-trivially, or lie close to the kernel of Φ. Theorem 1.4 shows that the column space of $\widetilde{U}^{\circledast 2}$ for a smoothed $\widetilde{U}$ is in fact far from the kernel of Φ with high probability. Note that $\widetilde{U}$ only has about nm truly independent coordinates or "bits", whereas a random subspace (matrix) of the same dimension has $c \cdot n^2 m^2$ independent coordinates.

Challenge with existing approaches. This setting captures many kinds of random matrices that have been studied earlier including [4, 9, 11]. For example, [11] studies the setting when a fixed polynomial map $f : \mathbb{R}^n \to \mathbb{R}^k$ applied to a randomly perturbed vector $\tilde{u}_i$ to produce the ith column $f(\tilde{u}_i)$. It turns out to be a special case of our setting above when $m = 1$. These works use the *leave-one-out approach* to lower bound the least singular value, where they establish that every column has a non-negligible component orthogonal to the span of the rest of the columns (see Lemma 3.1). However this approach crucially relies on the columns bringing in independent randomness.[6] This does not hold in our setting, since every column share randomness with $\Omega(m)$ other columns.

In the recent algebraic algorithms of [7, 12] for decomposing sum of powers of polynomials, the analysis of the algorithm involves analyzing the least singular value of different random matrices. One such matrix $\mathcal{M}$ is formed by concatenating the Kronecker products of a collection of underlying matrices. This allows us to reason about that the non-overlap or distance between the lifts of a collection of subspaces. The work of [7] analyzed the *fully random* setting and proves least singular value bounds with intricate arguments involving graph matrices, matrix concentration, and other ideas. Specifically, like in [29], they show that $\mathbb{E}[\mathcal{M}]$ has good least singular value, and then prove deviation bounds on the largest singular value of $\mathcal{M} - \mathbb{E}[\mathcal{M}]$ to get a bound of $\sigma_{\min}(\mathbb{E}[\mathcal{M}]) - \|\mathcal{M} - \mathbb{E}[\mathcal{M}]\|$. But this approach does not extend to the smoothed setting, since the underlying arbitrary matrix U makes it challenging to get good bounds for $\|\mathcal{M} - \mathbb{E}[\mathcal{M}]\|$.

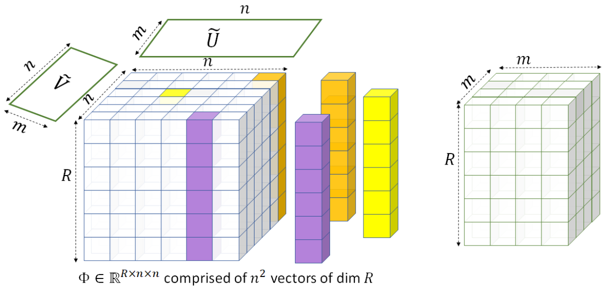
n
m
$\widetilde{U}$
n
n
$\widetilde{V}$
n
m
R
m
m
R
$\Phi \in \mathbb{R}^{R\times n\times n}$ comprised of n^2 vectors of dim R

Figure 1: The figure shows the setting of Theorem 2.1 with $d = 2$. *Left*: The linear operator $\Phi : \mathbb{R}^{n\times n} \to \mathbb{R}^R$ interpreted as a tensor consisting of a $n \times n$ array of R-dimensional vectors. There are *smoothed* or random contractions applied using matrices $\widetilde{U}, \widetilde{V} \in \mathbb{R}^{n\times m}$. *Right*: The operator $\Phi(\widetilde{U} \otimes \widetilde{V}) : \mathbb{R}^{m\times m} \to \mathbb{R}^R$ interpreted as an m^2 array of R-dimensional vectors. Theorem 2.1 shows that under the conditions of the theorem, with high probability the robust rank is m^2.

For the smoothed case, when $d = 2$, it turns out that we can use ideas similar to those described in Sections 2.1 and 2.2 to show Theorem 1.4. However, the approach runs into technical issues for larger d. Thus, we develop an alternate technique to analyze higher-order lifts that proves Theorem 1.4 for all constant d. In order to prove Theorem 1.4 we first move to a decoupled setting where we are analyzing the action of a linear operator on decoupled products of the form

$$\Phi(\widetilde{U} \otimes \widetilde{V}),$$

where $\widetilde{V}$ has a random component that is independent of $\widetilde{U}$. This new decoupling step leverages symmetry and the Taylor expansion and carefully groups together terms in a way that decouples the randomness. The main technical statement we prove is the following non-symmetric version of Theorem 1.4 which analyzes a linear operator acting on a Kronecker product of different smoothed matrices.

Informal Theorem 2.1 (Non-symmetric version for $d = 2$ and modal contractions)**.** *Suppose $\Phi \in \mathbb{R}^{R\times n^d}$ is a matrix with at least $\Omega(n^2)$ singular values larger than 1, and let $\widetilde{U}, \widetilde{V}$ be random ρ-perturbations of arbitrary matrices U, V. Then if $m \leq cn$ for an appropriate small constant $c > 0$, we have with probability $\geq 1 - \exp(-\Omega(n))$ that*

$$\sigma_{\min}\Big(\Phi(\widetilde{U} \otimes \widetilde{V})\Big) \geq \mathrm{poly}\left(\rho, \frac{1}{n}\right).$$

(See Theorem 5.2 for the formal statement for general d.)

[6] The work of [11] also handles some specific settings with a small overlap across columns, but these specialized ideas do not extend more generally to our setting.

Smoothed modal contractions. While Φ is specified as a linear operator or a matrix of dimension $R \times n^2$ in Theorem 2.1, one can alternately view Φ as a order-3 tensor of dimensions $R \times n \times n$ as shown in Figure 1. Theorem 2.1 then gives a lower bound for the multilinear rank[7] (or its robust analog) under smoothed modal contractions (dimension reduction) along the modes of dimension n each. The proof of this theorem is by induction on the order d. We perform each modal contraction one at a time. As shown in Figure 2, we first do modal contraction by $\widetilde{V}$ to obtain a $R \times n \times m$ tensor W and then by $\widetilde{U}$ to form the final $R \times m \times m$ tensor. We need to argue about the (robust) ranks of the matrix slices (we also call them blocks) and tensors obtained in intermediate steps. For any matrix M (potentially a matrix slice of the tensor Φ) of large (robust) rank $k > 1.1m$, a smoothed contraction $M\tilde{U}$ has full rank m (i.e., non-negligible least singular value) with probability $1 - \exp(-\Omega(k))$. To argue that the final tensor (when flattened) has full rank m^2, we need to argue that for the tensor in the intermediate step W, each of the m slices (along the contracted mode) has rank at least $\Omega(n)$. The original rank of Φ was large, so we know that a constant fraction of the slices $\Phi_1, \ldots, \Phi_n$ must have rank $\Omega(n)$. But this alone may not be enough since many of the slices can be identical, in which case the m slices are not sufficiently different from each other.

We can use the large rank of Φ to argue that a constant fraction of the matrix slices should have large "marginal rank" i.e., they have large rank even if we project out the column spaces of the slices that were chosen before it. While this strategy may work in the non-robust setting, this incurs an exponential blowup in the least singular value. Instead we use the following *randomized* strategy of finding a collection of blocks or slices $S_1 \subset [n]$, each of which has a *large "relative rank"*, even after we project out the column spaces of all the other blocks in S_1 (we show these statements in a robust sense, formalized using appropriate least singular values).

Finding many blocks with large relative rank. We note that while the idea is quite intuitive, the proof of the corresponding claim (Lemma 5.4) is non-trivial because we require that in any selected block, there must be many vectors with a large component orthogonal to the *entire span* of the other selected blocks. As a simple example, consider setting $n_2 = 2t$ and

$$\Phi_1 = \{e_1, e_2, \ldots, e_t, \varepsilon e_{t+1}, \varepsilon e_{t+2}, \ldots, \varepsilon e_{2t}\},$$
$$\text{and } \Phi_2 = \{\varepsilon e_1, \varepsilon e_2, \ldots, \varepsilon e_t, e_{t+1}, e_{t+2}, \ldots, e_{2t}\}.$$

In this case, even if ε is tiny, we cannot choose both the blocks, because the span of the vectors in Φ_2 contains all the vectors in Φ_1.

The proof will proceed by first identifying a set of roughly $R = \Omega(n^2)$ vectors (spread across the blocks) that form a well conditioned matrix, followed by randomly restricting to a subset of the blocks. We start with the following claim, which gives us the first step.

Claim 2.2 (Same as Lemma C.2)**.** *Suppose A is an $m \times n$ matrix such that $\sigma_k(A) \geq \theta$. Then there exists a submatrix A_S with $|S| = k$ columns, such that $\sigma_k(A_S) \geq \theta/\sqrt{nk}$.*

The lemma is a robust version of the simple statement that if $\sigma_k(A) > 0$, then there exist k linearly independent columns. The proof of the claim is elegant and uses the choice of a so-called Auerbach basis or a well-conditioned basis for the column span.

The outline of the main argument is as follows:

(1) First find a submatrix M of $R = \delta n^2$ columns of Φ such that $\sigma_R(M)$ is large

(2) Randomly sample a subset $T \subseteq [n]$ of the blocks.

(3) Discard any block $j \in T$ that has fewer than $\delta n/6$ vectors with a non-negligible component orthogonal to the span of $\cup_{r \in (T \setminus \{j\})} \Phi_r$; argue that there are $\Omega(\delta n)$ blocks remaining.

We remark that the above idea of a random restriction to obtain many blocks with large relative rank (in a robust sense) seems of independent interest and also comes in handy in the application to power sum decompositions (Claim E.5).

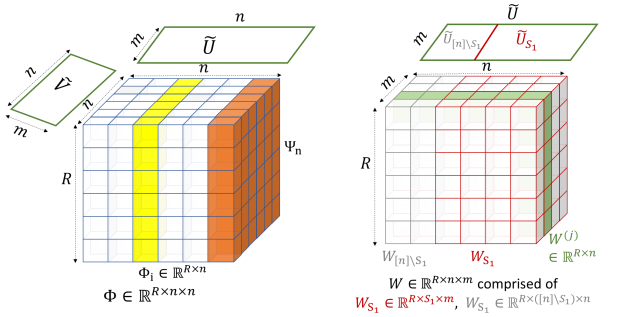
n
$\widetilde{U}$
m
n
$\widetilde{V}$
n
m
n
R
Ψ_n
$\Phi_i \in \mathbb{R}^{R\times n}$
$\Phi \in \mathbb{R}^{R\times n\times n}$
$\widetilde{U}$
m
$\widetilde{U}_{[n]\setminus S_1}$
$\widetilde{U}_{S_1}$
n
m
R
$W^{(j)}$
$\in \mathbb{R}^{R\times n}$
$W_{[n]\setminus S_1}$
W_{S_1}
$W \in \mathbb{R}^{R\times n\times m}$ comprised of
$W_{S_1} \in \mathbb{R}^{R\times S_1\times m}$, $W_{S_1} \in \mathbb{R}^{R\times([n]\setminus S_1)\times n}$

Figure 2: *Left*: The setting of $d = 2$ with linear operator $\Phi : \mathbb{R}^{n\times n} \to \mathbb{R}^R$ having slices $\Phi_1, \ldots, \Phi_n \in \mathbb{R}^{R\times n}$. The modal contractions $\widetilde{U}, \widetilde{V} \in \mathbb{R}^{n\times m}$ have not yet been applied. *Right*: After modal contraction along $\widetilde{V} \in \mathbb{R}^{n\times m}$, we get $W \in \mathbb{R}^{R\times n\times m}$ with slices $W_1, \ldots, W_n$. $W_{S_1} \in \mathbb{R}^{R\times S_1\times m}$ represents the slices obtained from the "good" blocks $S_1 \subset [n]$, and $W_{[n]\setminus S_1} \in \mathbb{R}^{R\times([n]\setminus S_1)\times m}$ represents the remaining slices. The random modal contraction $\widetilde{U}$ is also split into $\widetilde{U}_{S_1} \in \mathbb{R}^{S_1\times m}, \widetilde{U}_{[n]\setminus S_1} \in \mathbb{R}^{[n]\setminus S_1\times m}$.

Finishing the inductive argument. As shown in Figure 2, after modal contraction along $\tilde{V} \in \mathbb{R}^{n\times m}$, we get $W \in \mathbb{R}^{R\times n\times m}$ with slices $W_1, \ldots, W_n$.

Now we would like to argue that when we perform a smoothed contraction with $\widetilde{U}$, the contracted slices have large rank, while simultaneously preserving the relative rank across the slices. Let $W_{S_1} \in \mathbb{R}^{R\times S_1\times m}$ represent the subtensor corresponding to the slices obtained from the "good" blocks $S_1 \subset [n]$ (which have large relative rank), and let $W_{[n]\setminus S_1} \in \mathbb{R}^{R\times([n]\setminus S_1)\times m}$ represent the remaining slices. Also let $W^{(j)} \in \mathbb{R}^{R\times n}$ denote the matrix slices along the alternate mode for each $j \in [m]$. We can show that the randomly contracted matrices $W_{S_1}^{(j)}$ have large relative rank with respect to each other. The random modal contraction $\widetilde{U}$ can also now be

[7] The multilinear rank(s) of a tensor is the rank of the matrix after flattening all but one mode of the tensor.

split into $\widetilde{U}_{S_1} \in \mathbb{R}^{S_1 \times m}, \widetilde{U}_{[n]\setminus S_1} \in \mathbb{R}^{[n]\setminus S_1 \times m}$. The final matrix slice obtained for each $j \in [m^{d-1}]$ can be written as

$$M^{(j)} = W^{(j)}_{S_1} \widetilde{U}_{S_1} + W^{(j)}_{[n]\setminus S_1} \widetilde{U}_{[n]\setminus S_1},$$

where the randomness in the two summands is independent. Arguing that the high relative rank across the slices is preserved involves some work, and this is achieved in Lemma 5.5. The lemma proves that with high probability, every test unit vector $\alpha \in \mathbb{R}^{m\cdot m}$ has non-negligible value of $\|M\alpha\|_2$. A standard argument would consider a net over all potential unit vectors $\alpha \in \mathbb{R}^{m\cdot m}$. However this approach fails here, since we cannot get high enough concentration (of the form $e^{-\Omega(m^2)}$) that is required for this argument. Instead, we argue that if there were such a test vector $\alpha \in \mathbb{R}^{m\cdot m}$, there exists a block $j^* \in [m]$ where we get a highly unlikely event. This allows us to conclude the inductive proof that establishes Theorem 2.1.

3 PRELIMINARIES

We now introduce our basic definitions and notation. For a matrix $U \in \mathbb{R}^{n\times m}$, let $\|U\|$ and $\|U\|_F$ denote the operator and Frobenius norms of U, respectively. Central to the paper are ρ-smoothed matrices. In particular, given a matrix $U \in \mathbb{R}^{n\times m}$, we let $\tilde{U} = U + E$ where $E \in \mathcal{N}(0, \rho^2)$. We commonly call $\tilde{U}$ a ρ-smoothing of U or a ρ-perturbation of U. Similar notation is used for vector inputs $x = (x_1, \ldots, x_n)$ to a polynomial $p : \mathbb{R}^n \to \mathbb{R}^m$. I.e., $\tilde{x} = x + \eta$ where $\eta \in \mathcal{N}(0, \rho^2)$. Thus, for example, $p(\tilde{x})$ is the evaluation of p on a ρ-smoothed x.

Products. We also frequently use the Kronecker product, denoted $\otimes$, and the Khatri-Rao product, denoted $\odot$. Given matrices, $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{k\times \ell}$, the Kronecker product $A \otimes B$ is the block matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \ldots & a_{1m_1}B \\ \vdots & \ddots & \vdots \\ a_{n_1 1}B & \ldots & a_{n_1 m_1}B \end{bmatrix} \in R^{nk\times m\ell}.$$

We let $A^{\otimes d} \in R^{n^d \times m^d}$ denote the Kroncker product of a total of d copies of A. In the case that $m = \ell$, the Khatri-Rao product $A \odot B$ is defined by

$$A \odot B = \begin{bmatrix} \uparrow & & \uparrow \\ a_1 \otimes b_1 & \ldots & a_m \otimes b_m \\ \downarrow & & \downarrow \end{bmatrix} \in \mathbb{R}^{nk\times m}.$$

Here a_j and b_j denote the jth column of A and B, respectively, and $a_j \otimes b_j$ is the Kronecker product (or simply the tensor product) of these columns.

For vector spaces $\mathcal{U}, \mathcal{V}$, the tensor product space $\mathcal{U} \otimes \mathcal{V} = \{u \otimes v : u \in \mathcal{U}, v \in \mathcal{V}\}$. When $\mathcal{U} = \mathcal{V}$, we also call $\mathcal{U}^{\otimes 2} = \mathcal{U} \otimes \mathcal{U}$ a lift of the space $\mathcal{U}$ (of degree/order 2). This can also be generalized to d-wise products and lifts. When $\mathcal{U} = \mathbb{R}^n$, the space $(\mathbb{R}^n)^{\otimes d}$ corresponds to the space of all d-th order tensors of dimensions $n \times n \cdots \times n$. This is isomorphic to the space $\mathbb{R}^{n^d}$; each tensor can be flattened to form a vector in n^d dimensions i.e., $(\mathbb{R}^n)^{\otimes d} \cong \mathbb{R}^{n^d}$.

Symmetrized products. We are often concerned with symmetrized versions of matrix products. To handle these, we introduce a (partially) symmetrized Kronecker product $\circledast$ which is defined for tuples of matrices $(U^{(1)}, \ldots, U^{(d)})$ where $U^{(j)} \in \mathbb{R}^{n_j \times m}$. We define $U^{(1)} \circledast U^{(2)} \circledast \ldots \circledast U^{(d)} \in \mathbb{R}^{\Pi_{i=1}^d n_j \times \binom{m+d-1}{d}}$ to be the matrix with columns indexed by tuples $(i_1, i_2, \ldots, i_d)$ with $1 \le i_1 \le i_2 \le \cdots \le i_d \le m$ where the column corresponding to $(i_1, i_2, \ldots, i_d)$ is

$$\frac{1}{|S_d|} \sum_{\pi \in S_d} u^{(1)}_{i_{\pi(1)}} \otimes u^{(2)}_{i_{\pi(2)}} \otimes \cdots \otimes u^{(d)}_{i_{\pi(d)}}.$$

Here S_d denotes the symmetric group on $[d]$ and $u^{(j)}_{i_{\pi(j)}}$ denotes the $i_{\pi(j)}$th column of $U^{(j)}$. For example, for matrices $U, V \in \mathbb{R}^{n\times m}$, the column of $U \circledast V$ corresponding to a tuple (i, j) with $i \le j$ is

$$\frac{1}{2}(u_i \otimes v_j + u_j \otimes v_j).$$

In the case that $i = j$, this reduces to $u_i \otimes v_i$. For a matrix $U \in \mathbb{R}^{n\times m}$, we let $U^{\circledast d} \in \mathbb{R}^{n^d \times \binom{m+d-1}{d}}$ denote the $\circledast$ product of a total of d copies of U. The product $\circledast$ can be viewed as a partially symmetrized version of the Kronecker product since all columns of $U^{\circledast d}$ are symmetric with respect to the natural symmetrization of $\mathbb{R}^{n^d} \cong (\mathbb{R}^n)^{\otimes d}$.

Along these lines, we introduce the operator $\mathsf{Sym}_d : \mathbb{R}^{n^d} \to \mathbb{R}^{n^d}$ which symmetrizes elements of $\mathbb{R}^{n^d}$ with respect to the identification $\mathbb{R}^{n^d} \cong (\mathbb{R}^n)^{\otimes d}$. With this notation, we have that

$$\mathsf{Sym}_d(U^{\circledast d}) = U^{\circledast d}.$$

Moreover, the columns of the matrix $U^{\circledast d}$ are precisely the *unique* columns of the matrix $\mathsf{Sym}_d(U^{\otimes d})$.

Finally, for a vector space $\mathcal{U}$, we have that $\mathcal{U}^{\circledast d} = \mathsf{Sym}_d(\mathcal{U}^{\otimes d})$ is the space of symmetric dth tensors over the spacd $\mathcal{U}$. We also call this the symmetric dth order left of the space $\mathcal{U}$.

Leave-one-out distance. The leave-one-out distance of a matrix U is a useful tool for analyzing least singular values. Given $U \in \mathbb{R}^{n\times m}$, define the leave-one-out distance $\ell(U)$ by

$$\ell(U) = \min_i \operatorname{dist}\left(u_i, \operatorname{Span}\{u_j : j \neq i\}\right).$$

The least singular value of U is related to the leave-one-out distance of U through the following lemma [24].

Lemma 3.1 (Leave one out distance). *Let $U \in \mathbb{R}^{n\times m}$. Then*

$$\frac{\ell(U)}{\sqrt{m}} \le \sigma_{\min}(U) \le \ell(U).$$

See also Lemma A.2 for a block-version of leave-one-out singular value bounds.

In our work we also encounter the Jacobian of a polynomial map. Given a vector valued function $P(x) = (p_1(x), p_2(x), \ldots, p_N(x))$ over underlying variables $x = (x_1, x_2, \ldots, x_n)$, the Jacobian is defined as the $(n \times N)$ matrix of partial derivatives where the (i, j)th entry is $\frac{\partial p_j}{\partial x_i}$. Thus, the linear approximation of $P(x)$ around a point x is simply $P(x + \eta) = P(x) + J(x)^T \eta$.

4 HIERARCHICAL NETS AND ANTI-CONCENTRATION FROM JACOBIAN CONDITIONING

A complete version of this section, including all deferred proofs, can be found in Appendix B. In this section, we will primarily deal with a matrix $\mathcal{M}$ of dimensions $N \times m$ where $m < N$. The columns will be denoted by $\widetilde{X}_i$, and we wish to show a lower bound on $\sigma_m(\mathcal{M})$.

In this section, we describe the finer ε-net argument outlined in Section 2. We begin with a formal definition of the CAA property.

Definition 4.1 (CAA property). We say that a random matrix $\mathcal{M}$ with m columns has the CAA property with parameter $\beta > 0$, if for all $k \geq 1$, for all test vectors $\alpha \in \mathbb{R}^m$ with at least k coordinates of magnitude δ, there exist $\lambda > 0$ and $c \geq \frac{8}{\beta}$ (dependent only on $\mathcal{M}$) such that

$$\forall h \in (0,1), \quad \mathbb{P}[\|\mathcal{M}\alpha\| < \delta h/\lambda] \leq \exp\left(-c \min(m, km^{\beta}) \log(1/h)\right).$$

Remark. We note that the condition $c \geq 8/\beta$ may seem strong; however, as we will see in applications, it is satisfied as long as m is small enough compared to N, the number of rows of the matrix.

4.1 Hierarchical nets

The following shows that the CAA property implies a least singular value guarantee.

Theorem 4.2. *Suppose $\mathcal{M}$ is a random matrix with m columns and that $\mathcal{M}$ satisfies the CAA property with some parameter $\beta > 0$. Suppose additionally that we have the spectral norm bound $\|\mathcal{M}\| \leq L$ with probability $1-\eta$. Then with probability at least $1-\exp(-m^{\beta})-\eta$, we have*

$$\sigma_m(\mathcal{M}) \geq \frac{1}{(Lm\lambda)^{2\lceil \frac{1}{\beta} \rceil}},$$

where λ comes from the CAA property.

As discussed in Section 2, the natural approach to proving such a result would be to take nets based on the sparsity of the test vector α. In other words, if there are k nonzero values of magnitude $\delta > 0$, the CAA property yields a least singular value lower bound of δ/λ (choosing h to be a small constant), and we can take a union bound over a net of size $\exp(k)$. The issue with this argument is that α might have many other non-zero values that are *slightly* smaller than δ, and these might lead to a zero singular value (unless it so happened that $\lambda < 1/m$, which we do not have a control of). Of course, in this case, we should have worked with a slightly smaller value of δ, but this issue may recur, so we need a more careful argument.

The rest of this subsection will focus on proving Theorem 4.2. For defining the nets, we will use threshold values $\tau_1 = 1/m$, $\tau_2 = \theta/m$, and so on (more generally, $\tau_j = \theta^{j-1}/m$). θ is a parameter that will be chosen appropriately; for now we simply use $\theta \in (0, 1/m)$.

We construct a sequence of nets $\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_{s-1}$ as follows. The net $\mathcal{N}_1$ is a set of vectors parametrized by pairs $(r_1, r_2) \in \mathbb{N}^2$, where: (a) $1 \leq r_1 \leq m^{1-\beta}$, (b) $r_2 \leq m^{\beta} r_1$. For each pair (r_1, r_2), we include all the vectors whose entries are integer multiples of $\frac{\theta}{m}$ with have exactly $(r_1 + r_2)$ non-zero entries, of which r_1 entries are in $(\tau_1, 1]$ and r_2 entries are in $[\tau_2, \tau_1]$. Thus, the number of vectors in $\mathcal{N}_1$ for a single pair (r_1, r_2) is bounded by:

$$\binom{m}{r_1}\binom{m}{r_2}\left(\frac{m}{\theta}\right)^{r_1}\left(\frac{m}{\theta}\right)^{r_2} < \left(\frac{m}{\theta}\right)^{2(r_1+r_2)}.$$

The next net $\mathcal{N}_2$ has vectors parametrized by $(r_1, r_2, r_3) \in \mathbb{N}^3$, where (a) $r_2 \leq m^{1-\beta}$, (b) $r_3 \leq m^{\beta} r_2$, and additionally, (c) $r_2 \geq m^{\beta} r_1$. For each such tuple, we include vectors that have exactly $(r_1 + r_2 + r_3)$ non-zero entries (in the corresponding τ ranges as above), and have values that are all integer multiples of θ^2/m.

More generally, the vectors of $\mathcal{N}_j$ will be parametrized by $(r_1, r_2, \dots, r_{j+1}) \in \mathbb{N}^{j+1}$, where (a) $r_j \leq m^{1-\beta}$, (b) $r_{j+1} \leq m^{\beta} r_j$, and additionally, (c) for $1 \leq i < j$, we have $r_{i+1} > m^{\beta} r_i$. In other words, r_{j+1} is the first value that does not grow by a factor m^{β}. For every such tuple, $\mathcal{N}_j$ includes all vectors that have exactly $(r_1 + \cdots + r_{j+1})$ non-zero entries, each of which is an integer multiple of $\frac{\theta^j}{m}$, and exactly r_i of them in the range $(\tau_i, \tau_{i-1}]$ for all $i \leq j+1$.

We have nets of this form for $j = 1, 2, \dots, s-1$, where $s = \lceil \frac{1}{\beta} \rceil$. We now have the following claim.

Claim 4.3. *Fix any $1 \leq j < s$. We have*

$$\mathbb{P}\left[\exists \alpha \in \mathcal{N}_j, \|\mathcal{M}\alpha\| < \frac{\theta^{j-\frac{1}{2}}}{m\lambda}\right] < \exp\left(-\tfrac{1}{2} c m^{j\beta}\right).$$

Finally, we have a bigger net for all "dense" vectors α, that have at least $m^{1-\beta}$ coordinates of magnitude $\geq \frac{\theta^{s-1}}{m}$. This net consists of vectors $\in \mathbb{R}^m$ for which (a) every coordinate is an integer multiple of θ^s/m (between 0 and 1), and (b) at least $m^{1-\beta}$ coordinates are $\geq \frac{\theta^{s-1}}{m}$. Call this net $\mathcal{N}_0$. An easy upper bound for the size is

$$|\mathcal{N}_0| \leq \left(\frac{m}{\theta^s}\right)^m.$$

Using this, we have the following:

Claim 4.4.

$$\mathbb{P}\left[\exists \alpha \in \mathcal{N}_0 : \|\mathcal{M}\alpha\| < \frac{\theta^{s-\frac{1}{2}}}{m\lambda}\right] < \exp\left(-\frac{c}{2} m\right).$$

One of the advantages of our ε-net argument is that if we only care about "well spread" vectors, we can obtain a much stronger concentration bound (Eq (10)).

Observation 4.5. *Suppose $\mathcal{M}$ is a random matrix that satisfies the CAA property with parameter β. Let us call a test vector α (of length ≤ 1) "dense" if it has at least $m^{1-\beta}$ coordinates of magnitude $> \delta$. Then*

$$\mathbb{P}\left[\exists \text{ dense } \alpha : \|\mathcal{M}\alpha\| < \frac{1}{(Lm\lambda)^{2\lceil \frac{1}{\beta} \rceil}}\right] < \exp\left(-\tfrac{1}{2} c m\right).$$

Note that in the above claim, m could be quite large compared to n. The observation follows immediately from (10), but we will use it later in Section 4.3.

4.2 Anticoncentration of a vector of homogeneous polynomials

We consider the following setting: let $p_1, p_2, \dots, p_N$ be a collection of homogeneous polynomials over n variables $(x_1, x_2, \dots, x_n)$, and define

$$P(x) = \begin{bmatrix} p_1(x) \\ p_2(x) \\ \vdots \\ p_N(x) \end{bmatrix} \qquad (1)$$

Our goal will be to show anticoncentration results for P. Specifically, we want to prove that $\mathbb{P}[\|P(\tilde{x}) - y\| < \varepsilon]$ is small for all y, where $\tilde{x}$ is a perturbation of some (arbitrary) vector $x \in \mathbb{R}^n$. We give a sufficient condition for proving such a result, in terms of the Jacobian of P. (See Section 3 for background.)

Definition 4.6 (Jacobian rank property)**.** We say that P has the Jacobian rank property with parameters (k, c, γ) if for all $\rho > 0$ and for all x, the matrix $J(\tilde{x})$ has at least k singular values of magnitude $\geq c\rho$, with probability at least $1 - \gamma$. Here, $\tilde{x} = x + \eta$, where $\eta \sim \mathcal{N}(0, \rho^2)$ is a perturbation of the vector x.

Comment. Indeed, all of our results will hold if we only have the required condition for *small enough* perturbations ρ. To keep the results simple, we work with the stronger definition.

For many interesting settings of P, the Jacobian rank property turns out to be quite simple to prove. Our main result now is that the property above implies an anticoncentration bound for P.

Theorem 4.7. *Suppose $P(x)$ defined as above satisfies the Jacobian rank property with parameters (k, c, γ), and suppose further that the Jacobian P' is M-Lipschitz in our domain of interest. Let x be any point and let $\tilde{x}$ be a ρ-perturbation. Then for any $h > 0$, we have*

$$\forall y \in \mathbb{R}^N,\ \mathbb{P}\left[\|P(\tilde{x}) - y\| < \frac{c\rho^2 h}{64Mnk}\right] \leq \gamma + \exp(-\frac{1}{4} \cdot k \log(1/h)).$$

A key ingredient in the proof is the following "linearization" based lemma.

Lemma 4.8. *Suppose x is a point at which the Jacobian $J(x)$ of a polynomial P has at least k singular values of magnitude $\geq \tau$. Also suppose that the norm of the Hessian of P is bounded by M in the domain of interest. Then, for "small" perturbations, $0 < \rho < \frac{\tau}{4Mnk}$, we have that for any $\varepsilon > 0$,*

$$\forall y,\ \mathbb{P}[\|P(\tilde{x}) - y\| < \varepsilon] < \left(\frac{2\varepsilon}{\tau\rho}\right)^k + \left(\frac{2M\rho nk}{\tau}\right)^{k/2}.$$

We remark that the lemma does not imply Theorem 4.7 directly because it only applies to the case where the perturbation ρ is much smaller than the singular value threshold τ.

4.3 Jacobian rank property for Khatri Rao products

As the first application, let us use the machinery from the previous sections to prove the following.

Theorem 4.9. *Suppose $U, V \in \mathbb{R}^{n \times m}$ and suppose their entries are independently perturbed (by Gaussians $\mathcal{N}(0, \rho^2)$) to obtain $\tilde{U}$ and $\tilde{V}$. Then whenever $m \leq n^2/C$ for some absolute constant C, we have*

$$\sigma_{\min}(\tilde{U} \odot \tilde{V}) \geq \operatorname{poly}\left(\rho, \frac{1}{n}\right),$$

with probability $1 - \exp(-\Omega(n))$.

Note that the result is stronger in terms of the success probability than the main result of [9] and matches the result of [4]. The following lemma is the main ingredient of the proof, as it proves the CAA property for $\tilde{U} \odot \tilde{V}$. Theorem 4.9 then follows immediately from Theorem 4.2.

Lemma 4.10. *Suppose $\alpha \in \mathbb{R}^m$ be a unit vector at least k of whose coordinates have magnitude $\geq \delta$. Let U, V be arbitrary (as above), and let $\tilde{U}$ and $\tilde{V}$ be ρ perturbations. Define $P(\tilde{U}, \tilde{V}) = \sum_i \alpha_i \tilde{u}_i \otimes \tilde{v}_i$. Then for $M = (m + n)^2$ and all $h > 0$, we have*

$$\mathbb{P}\left[\|P(\tilde{U}, \tilde{V})\| < \delta h \cdot \frac{\rho^2}{64Mnk}\right] < \exp\left(-\frac{1}{16} kn \log(1/h)\right).$$

Remark. To see why this satisfies the CAA property (hypothesis of Theorem 4.2), note that as long as $m < n^2/C$ for a sufficiently large (absolute) constant C, the term $\frac{kn}{16} \geq 16 \min(m, km^{1/2})$, thus it satisfies the condition with $\beta = 1/2$.

The Jacobian property used to show Lemma 4.10 can be extended to higher order Khatri-Rao products. We give details in Section B.3.

5 HIGHER ORDER LIFTS AND STRUCTURED MATRICES FROM KRONECKER PRODUCTS

A complete version of this section, including all deferred proofs can be found in Appendix C. We provide the following theorem.

Theorem 5.1. *Suppose $d \in \mathbb{N}$, and let $\Phi : \mathsf{Sym}^d(\mathbb{R}^n) \to \mathbb{R}^D$ be an orthogonal projection of rank $R = \delta\binom{n+d-1}{d}$ for some constant $\delta > 0$, and let $\mathsf{Sym}_d : (\mathbb{R}^n)^{\otimes d} \to \mathsf{Sym}^d(\mathbb{R}^n)$ be the orthogonal projection on to the symmetric subspace of $(\mathbb{R}^n)^{\otimes d}$. Let $U = (u_i : i \in [m]) \in \mathbb{R}^{n \times m}$ be an arbitrary matrix, and let $\tilde{U}$ be a random ρ-perturbation of U. Then there exists a constant $c_d > 0$ such that for $m \leq c_d \delta n$, with probability at least $1 - \exp\big(-\Omega_{d,\delta}(n)\big)$, we have the least singular value*

$$\sigma_{\binom{m+d-1}{d}}(\Phi \tilde{U}^{\circledast d}) \geq \frac{\rho^d}{n^{O(d)}}, \text{ where}$$

$$\tilde{U}^{\circledast d} := \Big(\mathsf{Sym}_d(\tilde{u}_{i_1} \otimes \tilde{u}_{i_2} \cdots \otimes \tilde{u}_{i_d}) : 1 \leq i_1 \leq i_2 \leq \cdots \leq i_d \leq n\Big). \qquad (2)$$

In the above statement, one can also consider an arbitrary linear operator Φ and suffer an extra factor of $\sigma_R(\Phi)$ in the least singular value bound (by considering the projector onto the span of the top

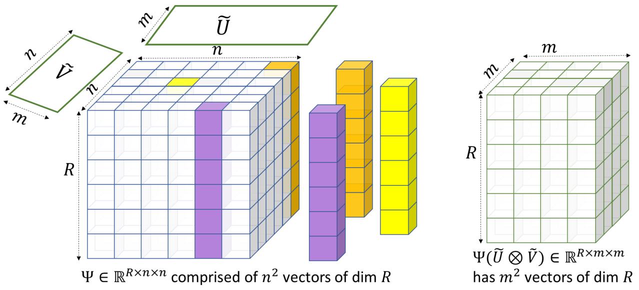
m
$\tilde{U}$
n
n
$\tilde{V}$
n
m
R
$\Psi \in \mathbb{R}^{R\times n\times n}$ comprised of n^2 vectors of dim R
m
m
R
$\Psi(\tilde{U}\otimes\tilde{V}) \in \mathbb{R}^{R\times m\times m}$
has m^2 vectors of dim R

Figure 3: *Left*: The linear operator $\Psi : \mathbb{R}^{n\times n} \to \mathbb{R}^R$ interpreted as a tensor consisting of a $n \times n$ array of R-dimensional vectors. There are *smoothed* or random contractions applied using matrices $\tilde{U}, \tilde{V} \in \mathbb{R}^{n\times m}$. *Right*: The operator $\Psi(\tilde{U} \otimes \tilde{V}) : \mathbb{R}^{m\times m} \to \mathbb{R}^R$ interpreted as an m^2 array of R-dimensional vectors. Theorem 5.2 shows that under the conditions of the theorem, with high probability the robust rank of this operator is m^2 i.e, the least singular value of $R \times m^2$ matrix is inverse polynomial.

R singular vectors). In the rest of the section, we assume that Φ is an orthogonal projector of rank R without loss of generality.

Theorem 5.1 follows from the following theorem (Theorem 5.2) which gives a non-symmetric analog of the same statement. The proof of Theorem 5.1 follows from a reduction to Theorem 5.2 that is given by Lemma C.4. In what follows, $\Psi \in \mathbb{R}^{R\times n^d}$ denotes the natural matrix representation of Φ such that $\Psi x^{\otimes d} = \Phi(x^{\otimes d})$ for all $x \in \mathbb{R}^n$.

Theorem 5.2. *Suppose $\ell \in \mathbb{N}$, $R = \delta\binom{n+d-1}{d}$ for some constant $\delta > 0$ and let $\Psi : (\mathbb{R}^n)^{\otimes \ell} \to \mathbb{R}^D$ be a linear operator with $\sigma_R(\Psi) \geq 1$. Suppose random matrices $\tilde{U}^{(1)}, \dots, \tilde{U}^{(d)} \in \mathbb{R}^{n\times m}$ are generated as follows:*

$$\forall j \in [d],\ \tilde{U}^{(j)} = U^{(j)} + Z^{(j)},\ \text{where } Z^{(j)} \sim_{i.i.d} \mathcal{N}(0,\rho^2)^{n\times m} \text{ and is independent of } U^{(j)}, \tag{3}$$

while $U^{(j)} \in \mathbb{R}^{n\times m}$ is arbitrary and can also depend on $\tilde{U}^{(j+1)}, \dots, \tilde{U}^{(d)}$. Then there exists constants $c_d, c'_d > 0$ and an absolute constant $c_0 \geq 1$ such that for $m \leq c_d\delta n$, with probability at least $1 - \exp\left(-\Omega_{d,\delta}(n)\right)$, we have

$$\sigma_{m^d}\left(\Psi(\tilde{U}^{(1)} \otimes \cdots \otimes \tilde{U}^{(d)})\right) \geq \frac{c'_d \rho^d}{n^{c_0 d}}. \tag{4}$$

While Ψ is specified as a matrix of dimension $R \times n^d$ in Theorem 5.2, one can alternately view Ψ as a $(d+1)$-order tensor of dimensions $R\times n\times n\times \cdots \times n$ as shown in Figure 4. Theorem 5.2 then gives a lower bound for the multilinear rank (in fact, for its robust analog) under smoothed modal contractions along the d modes of dimension n each.

Applying Theorem 5.1 along with the block leave-one-out approach (see Lemma A.2) we arrive at the following corollary.

Corollary 5.3. *Suppose $d, t \in \mathbb{N}$ and let $1 \geq \delta_1 > \delta_2 > 0$ be given. Also let $\Phi : \mathrm{Sym}^d(\mathbb{R}^n) \to \mathbb{R}^D$ be an orthogonal projection of rank $R \geq \delta_1\binom{n+d-1}{d}$. Let $\{U_j\}_{j=1}^t \subset \mathbb{R}^{n\times m}$ be an arbitrary collection of $n \times m$ matrices, and for each j, let $\tilde{U}_j$ be a random ρ-perturbation of U_j. Then there exists a constant $c_d > 0$ such that if $t\binom{m+d-1}{d} \leq \delta_2\binom{n+d-1}{d}$ and $m \leq c_d(\delta_1 - \delta_2)n$, then with probability at least $1 - \exp\left(-\Omega_{d,\delta_1,\delta_2}(n)\right)$, we have the least singular value*

$$\sigma_{t\binom{m+d-1}{d}}\left(\Phi\left[\tilde{U}_1^{\otimes d} \quad \tilde{U}_2^{\otimes d} \quad \dots \quad \tilde{U}_t^{\otimes d}\right]\right) \geq \frac{\rho^d}{\sqrt{t}\,n^{O(d)}}. \tag{5}$$

5.1 Proof of Theorem 5.2

We will prove Theorem 5.2 for general d by induction on d. The following crucial lemma considers a linear operator Ψ acting on the space $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$, and shows that if Ψ has large rank $\Omega(n_1 n_2)$, then it has many "blocks" of large relative rank as described in Section 2.3.

Lemma 5.4. *Let $\Psi \in \mathbb{R}^{R\times(n_1 n_2)}$ be a projection matrix of rank $R = \delta n_1 n_2$ for some constant $\delta > 0$, and let $\Psi = [\Psi_1\ \Psi_2\ \dots\ \Psi_{n_1}]$ where the blocks $\Psi_i \in \mathbb{R}^{R\times n_2}$ $\forall i \in [n_1]$. Then there exists constants $c_1, c_2, c_3 > 0$ and a subset $S_1 \subset [n_1]$ with $|S_1| \geq c_1\delta n_1$ such that*

$$\forall i \in S_1,\ \sigma_{c_2\delta n_2}\left(\Pi^{\perp}_{S_1\setminus\{i\}}\Psi_i\right) \geq \frac{1}{(nk)^{c_3}}, \tag{6}$$

where $\Pi^{\perp}_S$ is the projection orthogonal to $\mathrm{span}\big(\cup_{i\in S} \mathrm{colspan}(\Psi_i)\big)$.

We note that while the statement of Lemma 5.4 is quite intuitive, the proof is non-trivial because we require that in any selected block, there must be many vectors with a large component orthogonal to the *entire span* of the other selected blocks. We prove this lemma in Section C.2 by restricting to randomly chosen columns as described in the overview (Section 2.3).

The following lemma will be important in the inductive proof of the theorem. It reasons about the robust rank (also called multi-linear rank) after the modal contraction by a smoothed matrix along a specific mode. The lemma is proved in slightly more generality; we will use it for the theorem with $\varepsilon = 1$.

Lemma 5.5 (Robust rank under random contractions)**.** *Suppose $\varepsilon \in (0,1]$ is a constant. For every constant $\gamma, C > 0$, there is a constant $c \in (0,1)$ such that the following holds for all $s = 2^{o(k)}$. Consider matrices $A_1, A_2, \dots, A_s \in \mathbb{R}^{R\times k}$, $C_1, \dots, C_s \in \mathbb{R}^{R\times m}$ and $\forall j \in [s]$ let $\Pi^{\perp}_{-j}$ denote the projector orthogonal to the span of the column spaces of $\{A_{j'} : j' \neq j, j' \in [s]\}$. Suppose the following conditions are satisfied:*

$$\forall j \in [s],\ \sigma_{\varepsilon k}(\Pi^{\perp}_{-j}A_j) \geq k^{-\gamma} \tag{7}$$

and $\sigma_1(A_j), \sigma_1(C_j) \leq k^C$. For a random ρ-perturbed matrix $\tilde{U} \in \mathbb{R}^{k\times m}$ with $m \leq c\varepsilon k$, we have with probability at least $1 - \exp(-\Omega(\varepsilon k))$ that

$$\text{if } \forall j \in [s],\ M_j = C_j + A_j\tilde{U},\ \text{then } \sigma_{sm}\Big(M_1 \mid \cdots \mid M_s\Big) \geq \frac{\rho}{2k^{\gamma+1}\sqrt{s}}.$$

Finally, we reduce the the setting of symmetric products to that of non-symmetric products. We provide details in Section C.3.

6 ACKNOWLEDGMENTS

Vaidehi Srinivas and Aravindan Vijayaraghavan were supported by the National Science Foundation under Grant Nos. CCF-1652491, ECCS-2216970. Vaidehi Srinivas was also supported by the Northwestern Presidential Fellowship. Aditya Bhaskara was supported by the National Science Foundation under Grant Nos. CCF-2008688 and CCF-2047288. We thank the (anonymous) reviewers for their detailed feedback which helped improve the presentation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. 2016. Graph Matrices: Norm Bounds and Applications. *arXiv: Combinatorics* (2016). https://api.semanticscholar.org/CorpusID:211252816

[2] Anima Anandkumar, Rong Ge, Daniel J. Hsu, Sham M. Kakade, and Matus Telgarsky. 2015. Tensor Decompositions for Learning Latent Variable Models (A Survey for ALT). In *Algorithmic Learning Theory - 26th International Conference, ALT 2015, Banff, AB, Canada, October 4-6, 2015, Proceedings (Lecture Notes in Computer Science, Vol. 9355)*, Kamalika Chaudhuri, Claudio Gentile, and Sandra Zilles (Eds.). Springer, 19–38. https://doi.org/10.1007/978-3-319-24486-0_2

[3] Animashree Anandkumar, Daniel J. Hsu, and Sham M. Kakade. 2012. A Method of Moments for Mixture Models and Hidden Markov Models. In *COLT 2012 - The 25th Annual Conference on Learning Theory, June 25-27, 2012, Edinburgh, Scotland (JMLR Proceedings, Vol. 23)*, Shie Mannor, Nathan Srebro, and Robert C. Williamson (Eds.). JMLR.org, 33.1–33.34. http://proceedings.mlr.press/v23/anandkumar12/anandkumar12.pdf

[4] Nima Anari, Constantinos Daskalakis, Wolfgang Maass, Christos Papadimitriou, Amin Saberi, and Santosh Vempala. 2018. Smoothed Analysis of Discrete Tensor Decomposition and Assemblies of Neurons. In *Advances in Neural Information Processing Systems*.

[5] Pranjal Awasthi, Alex Tang, and Aravindan Vijayaraghavan. 2021. Efficient Algorithms for Learning Depth-2 Neural Networks with General ReLU Activations. In *Proceedings of the Neural Information Processing Systems (NeurIPS)*.

[6] Baruch Awerbuch and Robert Kleinberg. 2008. Online linear optimization and adaptive routing. *J. Comput. System Sci.* 74, 1 (2008), 97–114.

[7] Mitali Bafna, Jun-Ting Hsieh, Pravesh K. Kothari, and Jeff Xu. 2022. Polynomial-Time Power-Sum Decomposition of Polynomials. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. 956–967. https://doi.org/10.1109/FOCS54457.2022.00094

[8] Boaz Barak, Pravesh K. Kothari, and David Steurer. 2017. Quantum entanglement, sum of squares, and the log rank conjecture. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*.

[9] Aditya Bhaskara, Moses Charikar, Ankur Moitra, and Aravindan Vijayaraghavan. 2014. Smoothed Analysis of Tensor Decompositions. In *Symposium on the Theory of Computing (STOC)*.

[10] Aditya Bhaskara, Moses Charikar, and Aravindan Vijayaraghavan. 2014. Uniqueness of Tensor Decompositions with Applications to Polynomial Identifiability. *Conference on Learning Theory* (2014).

[11] Aditya Bhaskara, Aidao Chen, Aidan Perreault, and Aravindan Vijayaraghavan. 2019. Smoothed Analysis in Unsupervised Learning via Decoupling. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*.

[12] Ankit Garg, Neeraj Kayal, and Chandan Saha. 2020. Learning sums of powers of low-degree polynomials in the non-degenerate case. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. 889–899. https://doi.org/10.1109/FOCS46700.2020.00087

[13] Rong Ge, Qingqing Huang, and Sham M. Kakade. 2015. Learning Mixtures of Gaussians in High Dimensions. In *Symposium on Theory of Computing*.

[14] Navin Goyal, Santosh Vempala, and Ying Xiao. 2014. Fourier PCA and Robust Tensor Decomposition. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing* (New York, New York) *(STOC '14)*. Association for Computing Machinery, New York, NY, USA, 584–593. https://doi.org/10.1145/2591796.2591875

[15] Navin Goyal, Santosh Vempala, and Ying Xiao. 2014. Fourier PCA and Robust Tensor Decomposition. In *Symposium on the Theory of Computing (STOC)* (New York, New York) *(STOC '14)*. Association for Computing Machinery, New York, NY, USA, 584–593. https://doi.org/10.1145/2591796.2591875

[16] Venkatesan Guruswami and Ali Kemal Sinop. 2012. Optimal column-based low-rank matrix reconstruction. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, Yuval Rabani (Ed.). SIAM, 1207–1214. https://doi.org/10.1137/1.9781611973099.95

[17] Aram W. Harrow and Ashley Montanaro. 2010. An Efficient Test for Product States with Applications to Quantum Merlin-Arthur Games. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 633–642. https://doi.org/10.1109/FOCS.2010.66

[18] Elad Hazan and Zohar Karnin. 2016. Volumetric spanners: an efficient exploration basis for learning. *Journal of Machine Learning Research* (2016).

[19] Nathaniel Johnston, Benjamin Lovitz, and Aravindan Vijayaraghavan. 2023. Computing linear sections of varieties: quantum entanglement, tensor decompositions and beyond. In *Proceedings of the IEEE conference on the Foundations of Computer Science (FOCS)*.

[20] J. Lindenstrauss and L. Tzafriri. 2013. *Classical Banach Spaces I: Sequence Spaces*. Springer Berlin Heidelberg.

[21] Ankur Moitra and Alexander S. Wein. 2019. Spectral methods from tensor networks. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, Moses Charikar and Edith Cohen (Eds.). ACM, 926–937. https://doi.org/10.1145/3313276.3316357

[22] Goutham Rajendran and Madhur Tulsiani. [n. d.]. *Concentration of polynomial random matrices via Efron-Stein inequalities*. 3614–3653. https://doi.org/10.1137/1.9781611977554.ch138 arXiv:https://epubs.siam.org/doi/pdf/10.1137/1.9781611977554.ch138

[23] Tim Roughgarden. 2020. *Beyond the Worst-Case Analysis of Algorithms*. Cambridge University Press.

[24] Mark Rudelson and Roman Vershynin. 2008. The Littlewood–Offord problem and invertibility of random matrices. *Advances in Mathematics* 218, 2 (2008), 600 – 633. https://doi.org/10.1016/j.aim.2008.01.010

[25] Hanie Sedghi and Anima Anandkumar. 2016. Training Input-Output Recurrent Neural Networks through Spectral Methods. *CoRR* abs/1603.00954 (2016). arXiv:1603.00954 http://arxiv.org/abs/1603.00954

[26] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos. 2017. Tensor Decomposition for Signal Processing and Machine Learning. 65, 13 (July 2017), 3551–3582.

[27] Daniel A. Spielman and Shang-Hua Teng. 2004. Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time. *J. ACM* 51, 3 (may 2004), 385–463. https://doi.org/10.1145/990308.990310

[28] Shang-Hua Teng. 2023. "Intelligent Heuristics Are the Future of Computing". *ACM Trans. Intell. Syst. Technol.* (oct 2023). https://doi.org/10.1145/3627708 Just Accepted.

[29] Roman Vershynin. 2020. Concentration inequalities for random tensors. *Bernoulli* 26, 4 (2020), 3139 – 3162. https://doi.org/10.3150/20-BEJ1218

[30] Van Vu. 2017. *Anti-concentration Inequalities for Polynomials*. 801–810. https://doi.org/10.1007/978-3-319-44479-6_32

A AUXILIARY LEMMAS

Lemma A.1. *Let $\tilde{u} \in \mathbb{R}^n$ be a ρ-smoothed vector and fix $\delta > 0$. There exists a universal constant $c > 0$ such that*

$$Pr[\|\tilde{u}\| < \delta] \leq \left(\frac{c\delta}{\rho}\right)^n$$

Proof. The PDF of $\tilde{u}$ is bounded above by $\left(\rho\sqrt{2\pi})^{-n}\right)$ from which we obtain that

$$Pr[\|\tilde{u}\| < \delta] \leq \int_{B(0,\delta)} \left(\rho\sqrt{2\pi}\right)^{-n} dx = \left(\rho\sqrt{2\pi}\right)^{-n} \mathrm{vol}\big(B(0,\delta)\big)$$

$$= \left(\rho\sqrt{2\pi}\right)^{-n} \frac{\left(\delta\sqrt{\pi}\right)^n}{\Gamma\left(\frac{n}{2}+1\right)} = \left(\frac{\delta}{\rho\sqrt{2}}\right)^n \frac{1}{\Gamma(\frac{n}{2}+1)}.$$

Now, using Stirling's approximation we have

$$\Gamma\left(\frac{n}{2}+1\right) \geq \sqrt{\pi n}\left(\sqrt{\frac{n}{2e}}\right)^n.$$

Substituting this probability in above gives the desired probability upper bound. □

A.1 Block leave-one-out bounds

Lemma A.2. *Let $U_1, \ldots, U_t \in \mathbb{R}^{n\times m}$ and for each $j = 1, \ldots, t$ let $\Pi^{\perp}_{-j}$ be the projection on to the orthogonal complement of*

$$\mathrm{Ran}\left(\begin{bmatrix} U_1 & \ldots & U_{j-1} & U_{j+1} & \ldots U_t. \end{bmatrix}\right)$$

Define $\ell_B(\{U_j\}) = \min_j \sigma_{\min}(\Pi^{\perp}_{-j}U_j)$. Then

$$\frac{\ell_B(\{U_j\})}{\sqrt{t}} \leq \sigma_{\min}\left(\begin{bmatrix} U_1 & \ldots & U_t \end{bmatrix}\right) \leq \ell_B(\{U_j\}).$$

Proof. Let $\alpha \in \mathbb{R}^{tm}$ be a unit vector and write $\alpha = \alpha_1 \oplus \cdots \oplus \alpha_t$. Intuitively, α_j records the entries of alpha that are coefficients of the columns of U_j in the product $\begin{bmatrix} U_1 & \ldots & U_t \end{bmatrix}\alpha$. Since α is a unit vector, there must exist some index j such that $\|\alpha_j\| \geq \frac{1}{\sqrt{t}}$. From this we obtain

$$\left\|\Pi^{\perp}_{-j}U_j\alpha_j\right\| = \left\|\Pi^{\perp}_{-j}\begin{bmatrix} U_1 & \ldots & U_t \end{bmatrix}\alpha\right\| \leq \left\|\begin{bmatrix} U_1 & \ldots & U_t \end{bmatrix}\alpha\right\|.$$

Using the fact our lower bound on the norm of α_j and the fact that $\ell_B(\{U_j\})$ lower bounds the least singular value of $\Pi^{\perp}_{-j}U_j$, we obtain

$$\frac{\ell_B(\{U_j\})}{\sqrt{t}} \leq \left\|\Pi^{\perp}_{-j}U_j\alpha_j\right\|$$

from which the desired lower bound follows. The proof of the upper bound is straightforward. □

A.2 Bounds on singular values of smoothed matrices

Lemma A.3. *Let $V \in \mathbb{R}^{n\times k}$ be an arbitrary matrix with $k \leq n$, and let $\tilde{V}$ be a ρ perturbation. Suppose α_i are some scalars such that $\alpha_i \geq \delta$ for all $i \leq k$. Then for any $h \in (0, 1/2)$,*

$$\mathbb{P}\left[\sigma_{k/2}\left(\tilde{V}diag(\alpha)\right) < h\sigma\delta\right] \leq \exp(-\frac{1}{8}kn\log(1/h)).$$

Proof. Let us denote $W = \tilde{V}\mathrm{diag}(\alpha)$, for convenience. Suppose that $\sigma_{k/2}(W) < h\sigma\delta$. This implies that there exists a set J of $k/2$ columns of W with the property that the rest of the columns together have a squared projection at most $k^2(h\sigma\delta)^2$ orthogonal to the span of the columns in J.[8]

Now take any subset of columns J with $|J| = k/2$; the probability that any column $i \notin J$ has a projection of length $< h\delta\rho$ orthogonal to the span of J is at most $h^{n-\frac{k}{2}}$. (This is because for each of the $n - \frac{k}{2}$ directions orthogonal to the span of J, we must have a component $< h\delta\rho$, and we can use the standard Gaussian anticoncentration for each direction.) Since there are $k/2$ columns $i \notin J$, the probability that all of them satisfy the condition is $\leq h^{\frac{k}{2}(n-\frac{k}{2})}$.

The total number of choices for J is clearly at most 2^k, thus taking a union bound, we have that the probability is at most

$$2^k h^{\frac{k}{2}(n-\frac{k}{2})} \leq \exp(-\frac{1}{8}kn\log(1/h)).$$

□

B DEFERRED PROOFS FROM SECTION 4

In this section, we will primarily deal with a matrix $\mathcal{M}$ of dimensions $N \times m$ where $m < N$. The columns will be denoted by $\widetilde{X}_i$, and we wish to show a lower bound on $\sigma_m(\mathcal{M})$.

In this section, we describe the finer ε-net argument outlined in Section 2. We begin with a formal definition of the CAA property.

Definition 4.1 (CAA property)**.** We say that a random matrix $\mathcal{M}$ with m columns has the CAA property with parameter $\beta > 0$, if for all $k \geq 1$, for all test vectors $\alpha \in \mathbb{R}^m$ with at least k coordinates of magnitude δ, there exist $\lambda > 0$ and $c \geq \frac{8}{\beta}$ (dependent only on $\mathcal{M}$) such that

$$\forall h \in (0,1), \quad \mathbb{P}[\|\mathcal{M}\alpha\| < \delta h/\lambda] \leq \exp\left(-c\min(m, km^{\beta})\log(1/h)\right).$$

Remark. We note that the condition $c \geq 8/\beta$ may seem strong; however, as we will see in applications, it is satisfied as long as m is small enough compared to N, the number of rows of the matrix.

B.1 Hierarchical nets

The following shows that the CAA property implies a least singular value guarantee.

Theorem 4.2. *Suppose $\mathcal{M}$ is a random matrix with m columns and that $\mathcal{M}$ satisfies the CAA property with some parameter $\beta > 0$. Suppose additionally that we have the spectral norm bound $\|\mathcal{M}\| \leq L$ with probability $1-\eta$. Then with probability at least $1-\exp(-m^{\beta})-\eta$, we have*

$$\sigma_m(\mathcal{M}) \geq \frac{1}{(Lm\lambda)^{2^{\lceil\frac{1}{\beta}\rceil}}},$$

where λ comes from the CAA property.

[8] Here, we are using the well known connection between the low rank error and an approximation via columns [16].

As discussed in Section 2, the natural approach to proving such a result would be to take nets based on the sparsity of the test vector α. In other words, if there are k nonzero values of magnitude $\delta > 0$, the CAA property yields a least singular value lower bound of δ/λ (choosing h to be a small constant), and we can take a union bound over a net of size $\exp(k)$. The issue with this argument is that α might have many other non-zero values that are *slightly* smaller than δ, and these might lead to a zero singular value (unless it so happened that $\lambda < 1/m$, which we do not have a control of). Of course, in this case, we should have worked with a slightly smaller value of δ, but this issue may recur, so we need a more careful argument.

The rest of this subsection will focus on proving Theorem 4.2. For defining the nets, we will use threshold values $\tau_1 = 1/m$, $\tau_2 = \theta/m$, and so on (more generally, $\tau_j = \theta^{j-1}/m$). θ is a parameter that will be chosen appropriately; for now we simply use $\theta \in (0, 1/m)$.

We construct a sequence of nets $\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_{s-1}$ as follows. The net $\mathcal{N}_1$ is a set of vectors parametrized by pairs $(r_1, r_2) \in \mathbb{N}^2$, where: (a) $1 \le r_1 \le m^{1-\beta}$, (b) $r_2 \le m^\beta r_1$. For each pair (r_1, r_2), we include all the vectors whose entries are integer multiples of $\frac{\theta}{m}$ with have exactly $(r_1 + r_2)$ non-zero entries, of which r_1 entries are in $(\tau_1, 1]$ and r_2 entries are in $[\tau_2, \tau_1]$.

Thus, the number of vectors in $\mathcal{N}_1$ for a single pair (r_1, r_2) is bounded by:

$$\binom{m}{r_1}\binom{m}{r_2}\left(\frac{m}{\theta}\right)^{r_1}\left(\frac{m}{\theta}\right)^{r_2} < \left(\frac{m}{\theta}\right)^{2(r_1+r_2)}.$$

The next net $\mathcal{N}_2$ has vectors parametrized by $(r_1, r_2, r_3) \in \mathbb{N}^3$, where (a) $r_2 \le m^{1-\beta}$, (b) $r_3 \le m^\beta r_2$, and additionally, (c) $r_2 \ge m^\beta r_1$. For each such tuple, we include vectors that have exactly $(r_1 + r_2 + r_3)$ non-zero entries (in the corresponding τ ranges as above), and have values that are all integer multiples of θ^2/m.

More generally, the vectors of $\mathcal{N}_j$ will be parametrized by $(r_1, r_2, \dots, r_{j+1}) \in \mathbb{N}^{j+1}$, where (a) $r_j \le m^{1-\beta}$, (b) $r_{j+1} \le m^\beta r_j$, and additionally, (c) for $1 \le i < j$, we have $r_{i+1} > m^\beta r_i$. In other words, r_{j+1} is the first value that does not grow by a factor m^β. For every such tuple, $\mathcal{N}_j$ includes all vectors that have exactly $(r_1 + \cdots + r_{j+1})$ non-zero entries, each of which is an integer multiple of $\frac{\theta^j}{m}$, and exactly r_i of them in the range $(\tau_i, \tau_{i-1}]$ for all $i \le j + 1$.

We have nets of this form for $j = 1, 2, \dots, s - 1$, where $s = \lceil \frac{1}{\beta} \rceil$. We now have the following claim.

Claim 4.3. *Fix any* $1 \le j < s$. *We have*

$$\mathbb{P}\left[\exists \alpha \in \mathcal{N}_j, \|\mathcal{M}\alpha\| < \frac{\theta^{j-\frac{1}{2}}}{m\lambda}\right] < \exp\left(-\tfrac{1}{2} c m^{j\beta}\right).$$

Proof. First consider any single $\alpha \in \mathcal{N}_j$. By assumption, it has r_j coordinates with magnitude $\ge \frac{\theta^{j-1}}{m}$. Thus, from the CAA property,

$$\mathbb{P}\left[\|\mathcal{M}\alpha\| < \frac{\theta^{j-1}}{m}\frac{h}{\lambda}\right] \le \exp\left(-c r_j m^\beta \log(1/h)\right). \tag{8}$$

The number of vectors in $\mathcal{N}_j$ for a given tuple of r_j values is clearly bounded by $\left(\frac{m}{\theta^j}\right)^{r_1+r_2+\cdots+r_{j+1}}$. We choose $h = \theta^{1/2}$, and $\theta < 1/m$, and argue that as long as these are true,

$$(r_1 + \cdots + r_{j+1}) \log \frac{m}{\theta^j} \le \frac{c}{2} r_j m^\beta \log(1/h). \tag{9}$$

We can simplify this by noting that from our assumptions on r_j, $r_j m^\beta \ge \frac{1}{2}(r_1 + r_2 + \cdots + r_{j+1})$. Thus to show (9), it suffices to have

$$\frac{c}{4} \log \frac{1}{h} \ge \log \frac{m}{\theta^j}.$$

Since $\theta < 1/m$, we have $\theta^j/m > \theta^{j+1} > \theta^s$. Since $c \ge 8s$ and $h = \theta^{1/2}$, the above inequality holds, and so inequality (9) also holds.

Next, we observe that for any tuple of $\{r_i\}$ values in $\mathcal{N}_j$, since $r_1 \ge 1$, we have $r_j \ge m^{(j-1)\beta}$, which implies that $r_j m^\beta \ge m^{j\beta}$. Using this fact, together with (9), we first take a union bound over all $\alpha \in \mathcal{N}_j$ corresponding to a given tuple $(r_i)_{1 \le i \le j+1}$ (call this set $\mathcal{N}_j'$ for now), and obtain

$$\mathbb{P}\left[\exists \alpha \in \mathcal{N}_j' : \|\mathcal{M}\alpha\| < \frac{\theta^{j-\frac{1}{2}}}{m\lambda}\right] < \exp\left(-\frac{c}{2} m^{j\beta} \log(1/h)\right).$$

Now since the total number of tuples is easily bounded by $m^{j+1} \le m^s$, taking a further union bound over these choices and simplifying, we obtain the claim. □

Finally, we have a bigger net for all "dense" vectors α, that have at least $m^{1-\beta}$ coordinates of magnitude $\ge \frac{\theta^{s-1}}{m}$. This net consists of vectors $\in \mathbb{R}^m$ for which (a) every coordinate is an integer multiple of θ^s/m (between 0 and 1), and (b) at least $m^{1-\beta}$ coordinates are $\ge \frac{\theta^{s-1}}{m}$. Call this net $\mathcal{N}_0$.

An easy upper bound for the size is

$$|\mathcal{N}_0| \le \left(\frac{m}{\theta^s}\right)^m.$$

Using this, we have the following:

Claim 4.4.

$$\mathbb{P}\left[\exists \alpha \in \mathcal{N}_0 : \|\mathcal{M}\alpha\| < \frac{\theta^{s-\frac{1}{2}}}{m\lambda}\right] < \exp\left(-\frac{c}{2} m\right).$$

Proof. First, consider any fixed $\alpha \in \mathcal{N}_0$. Using part (b) of the definition of $\mathcal{N}_0$, we can use the CAA property to obtain

$$\mathbb{P}\left[\|\mathcal{M}\alpha\| < \frac{\theta^{s-1}h}{m\lambda}\right] < \exp\left(-cm\log(1/h)\right),$$

for any parameter h. As before, we show that this is small enough to take a union bound over $|\mathcal{N}_0|$ terms. Specifically, we argue that setting $h = \theta^{1/2}$,

$$\log |\mathcal{N}_0| \le m \log \frac{m}{\theta^s} \le \frac{1}{4} cm \log \frac{1}{\theta}.$$

The latter holds because $\theta < 1/m$, and $c \ge 8s \ge 4(s + 1)$. This completes the proof. □

We can now complete the proof of Theorem 4.2 as follows.

Proof of Theorem 4.2. Consider any $\alpha \in \mathbb{R}^m$ with $\|\alpha\| = 1$. Also, suppose we condition on the event that $\|\mathcal{M}\| \le L$ (which happens with probability $1-\eta$). Let $s = \lceil \frac{1}{\beta} \rceil$ as before. Now define $r_1, r_2, \dots r_s$ to be the number of entries of α in the intervals $(\frac{1}{m}, 1], (\frac{\theta}{m}, \frac{1}{m}]$, and so on. We consider two cases:

Case 1. $(r_1 + r_2 + \cdots + r_s) \ge m^{1-\beta}$. I.e., there are sufficiently many "large" coordinates in α.

In this case, we observe that there exists a vector $\alpha' \in \mathcal{N}_0$ such that $\|\alpha - \alpha'\| \le \theta^s$. Now, we have from Claim 4.4 that with high probability, $\|\mathcal{M}\alpha'\| \ge \frac{\theta^{s-\frac{1}{2}}}{m\lambda}$. If this holds, then $\|\mathcal{M}\alpha\| \ge \frac{\theta^{s-\frac{1}{2}}}{m\lambda} - L\theta^s$, where L is the spectral norm bound on $\mathcal{M}$. We choose θ such that

$$L\theta^{1/2} < \frac{1}{m\lambda} \iff \theta < \frac{1}{m^2 L^2 \lambda^2}.$$

Thus for this value of θ,

$$\mathbb{P}\left[\exists \alpha, \|\alpha\| = 1, \text{ satisfying (Case 1)}: \ \|\mathcal{M}\alpha\| < \frac{\theta^{s-\frac{1}{2}}}{2m\lambda}\right] < \exp\left(-\frac{c}{2}m\right). \tag{10}$$

Case 2. $(r_1 + r_2 + \cdots + r_s) < m^{1-\beta}$. In this case, we claim that we must have some index $j < s$ such that $r_{j+1}/r_j \le m^{\beta}$. This is because $r_1 \ge 1$ (a unit vector must have some entry $> 1/m$), and if the above does not hold, then $r_j > m^{(j-1)\beta}$. Plugging $j = s$ gives a contradiction to our assumption that $(r_1 + r_2 + \dots r_s) < m^{1-\beta}$.

Thus, let j be the smallest index for which we have $r_{j+1}/r_j \le m^{\beta}$. We can now consider the net $\mathcal{N}_j$ and find some $\alpha' \in \mathcal{N}_j$ such that $\|\alpha - \alpha'\| < \theta^j$.

We can use an argument identical to the one in (Case 1), this time leveraging Claim 4.3, to conclude that for all j,

$$\mathbb{P}\left[\exists \alpha, \|\alpha\| = 1, \text{satisfying (Case 2)}: \|\mathcal{M}\alpha\| < \frac{\theta^{s-\frac{1}{2}}}{2m\lambda}\right] < \sum_{1 \le j < s} \exp\left(-\frac{c}{2} m^{j\beta}\right).$$

The first term on the RHS dominates, and plugging in the chosen values of θ, s, this completes the proof. □

One of the advantages of our ε-net argument is that if we only care about "well spread" vectors, we can obtain a much stronger concentration bound (Eq (10)).

Observation 4.5. *Suppose $\mathcal{M}$ is a random matrix that satisfies the CAA property with parameter β. Let us call a test vector α (of length ≤ 1) "dense" if it has at least $m^{1-\beta}$ coordinates of magnitude $> \delta$. Then*

$$\mathbb{P}\left[\exists \text{ dense } \alpha : \|\mathcal{M}\alpha\| < \frac{1}{(Lm\lambda)^{2\lceil \frac{1}{\beta} \rceil}}\right] < \exp\left(-\tfrac{1}{2}cm\right).$$

Note that in the above claim, m could be quite large compared to n. The observation follows immediately from (10), but we will use it later in Section B.3.

B.2 Anticoncentration of a vector of homogeneous polynomials

We consider the following setting: let $p_1, p_2, \dots, p_N$ be a collection of homogeneous polynomials over n variables $(x_1, x_2, \dots, x_n)$, and define

$$P(x) = \begin{bmatrix} p_1(x) \\ p_2(x) \\ \vdots \\ p_N(x) \end{bmatrix} \tag{11}$$

Our goal will be to show anticoncentration results for P. Specifically, we want to prove that $\mathbb{P}[\|P(\tilde{x}) - y\| < \varepsilon]$ is small for all y, where $\tilde{x}$ is a perturbation of some (arbitrary) vector $x \in \mathbb{R}^n$. We give a sufficient condition for proving such a result, in terms of the Jacobian of P. (See Section 3 for background.)

Definition B.1 (Jacobian rank property)**.** We say that P has the Jacobian rank property with parameters (k, c, γ) if for all $\rho > 0$ and for all x, the matrix $J(\tilde{x})$ has at least k singular values of magnitude $\ge c\rho$, with probability at least $1 - \gamma$. Here, $\tilde{x} = x + \eta$, where $\eta \sim \mathcal{N}(0, \rho^2)$ is a perturbation of the vector x.

Comment. Indeed, all of our results will hold if we only have the required condition for *small enough* perturbations ρ. To keep the results simple, we work with the stronger definition.

For many interesting settings of P, the Jacobian rank property turns out to be quite simple to prove. Our main result now is that the property above implies an anticoncentration bound for P.

Theorem 4.7. *Suppose $P(x)$ defined as above satisfies the Jacobian rank property with parameters (k, c, γ), and suppose further that the Jacobian P' is M-Lipschitz in our domain of interest. Let x be any point and let $\tilde{x}$ be a ρ-perturbation. Then for any $h > 0$, we have*

$$\forall y \in \mathbb{R}^N, \ \mathbb{P}\left[\|P(\tilde{x}) - y\| < \frac{c\rho^2 h}{64Mnk}\right] \le \gamma + \exp(-\frac{1}{4} \cdot k \log(1/h)).$$

A key ingredient in the proof is the following "linearization" based lemma.

Lemma 4.8. *Suppose x is a point at which the Jacobian $J(x)$ of a polynomial P has at least k singular values of magnitude $\ge \tau$. Also suppose that the norm of the Hessian of P is bounded by M in the domain of interest. Then, for "small" perturbations, $0 < \rho < \frac{\tau}{4Mnk}$, we have that for any $\varepsilon > 0$,*

$$\forall y, \ \mathbb{P}[\|P(\tilde{x}) - y\| < \varepsilon] < \left(\frac{2\varepsilon}{\tau\rho}\right)^k + \left(\frac{2M\rho nk}{\tau}\right)^{k/2}.$$

We remark that the lemma does not imply Theorem 4.7 directly because it only applies to the case where the perturbation ρ is much smaller than the singular value threshold τ.

Proof. Let η be the random perturbation of x as in the lemma statement. We have

$$P(x+\eta) = P(x) + J(x)^T \eta + E(\eta),$$

where $E(\eta)$ is an error term, bounded in magnitude by $M\|\eta\|^2$ because of our assumption on the Hessian. Now, the desired probability is equivalent to

$$\mathbb{P}\left[J(x)^T \eta + E(\eta) \in \text{Ball}(y - P(x), \varepsilon)\right].$$

From the bound on η, the above probability can be upper bounded by

$$\mathbb{P}\left[J(x)^T \eta \in \text{Ball}\left(y - P(x), \varepsilon + M\|\eta\|^2\right)\right].$$

Let us denote the event in the parentheses above by $\mathcal{E}$. Now, consider the top k singular directions of $J(x)$; suppose the eigenvalues are $\sigma_1, \sigma_2, \dots, \sigma_k$, and suppose $\eta_1, \eta_2, \dots, \eta_k$ are the components of η along these directions. By hypothesis, $\sigma_i \ge \tau$ for all $i \le k$. Thus if $\mathcal{E}$ occurs, we also have,

$$\forall i \le k,\ \sigma_i \eta_i \in \text{Ball}\left((y - P(x))_i, \varepsilon + M\|\eta\|^2\right). \tag{12}$$

Let $\theta > 1$ be a parameter that we set later. We note that by Gaussian tail bounds,

$$\mathbb{P}[\|\eta\|^2 > n\rho^2 k\theta] \le \exp(-k\theta).$$

In what follows, let us condition on the event $\|\eta\|^2 \le n\rho^2 k\theta$. Then, the probability in (12) is upper bounded by

$$\left(\frac{\varepsilon + M\rho^2 nk\theta}{\tau\rho}\right)^k \le \left(\frac{2\varepsilon}{\tau\rho}\right)^k + \left(\frac{2M\rho nk\theta}{\tau}\right)^k.$$

We will choose the parameter $\theta = \log\left(\frac{\tau}{M\rho nk}\right)$. This ensures that the term $\exp(-k\theta)$ is the same order of magnitude as the last term on the RHS above. Simplifying, we obtain the desired claim. □

Proof of Theorem 4.7. The main idea in the proof is to view the perturbation $x \to \tilde{x}$ as occurring in two independent steps $x \to x' \to \tilde{x}$, where the first perturbation has norm $\rho\sqrt{1-z^2}$ and the second perturbation has norm ρz. By standard properties of Gaussian perturbations, this is equivalent to a ρ perturbation of x. We pick the parameter $z < 1/2$ carefully (later).

Using the Jacobian rank property of P on the first perturbation, we have that with probability $\ge 1 - \gamma$, $J(x')$ has at least k singular values of magnitude $\ge c(\rho/2)$ (we are using the fact that $z < 1/2$). Let us call this value τ, which we will use to apply Lemma 4.8. As long as we choose z such that

$$z\rho < \frac{\tau}{4Mnk} = \frac{c\rho}{8Mnk},$$

we can apply the Lemma to conclude that for any $\varepsilon > 0$,

$$\forall y,\ \mathbb{P}[\|P(\tilde{x}) - y\| < \varepsilon] < \left(\frac{2\varepsilon}{\tau z\rho}\right)^k + \left(\frac{2Mz\rho nk}{\tau}\right)^{k/2}$$
$$= \left(\frac{4\varepsilon}{z\rho^2}\right)^k + \left(\frac{4Mznk}{c}\right)^{k/2}.$$

Let $0 < f < 1/2$ be a parameter that we will fix shortly. We first choose $z = \frac{cf}{8Mnk}$, so that the latter term above becomes $(f/2)^{k/2}$. Then, we pick $\varepsilon = \frac{fz\rho^2}{8}$, so that the former term becomes $(f/2)^k$. Putting these together, we have that for all $f \in (0, 1/2)$,

$$\forall y,\ \mathbb{P}\left[\|P(\tilde{x}) - y\| < \frac{f^2 c\rho^2}{64Mnk}\right] < f^{k/2} = \exp\left(-\frac{1}{2}k\log(1/f)\right).$$

Writing $h = f^2$ and incorporating the failure probability of the Jacobian rank guarantee, the theorem follows. □

B.3 Jacobian rank property for Khatri Rao products

As the first application, let us use the machinery from the previous sections to prove the following.

Theorem 4.9. *Suppose $U, V \in \mathbb{R}^{n\times m}$ and suppose their entries are independently perturbed (by Gaussians $\mathcal{N}(0, \rho^2)$) to obtain $\tilde{U}$ and $\tilde{V}$. Then whenever $m \le n^2/C$ for some absolute constant C, we have*

$$\sigma_{\min}(\tilde{U} \odot \tilde{V}) \ge \text{poly}\left(\rho, \frac{1}{n}\right),$$

with probability $1 - \exp(-\Omega(n))$.

Note that the result is stronger in terms of the success probability than the main result of [9] and matches the result of [4]. The following lemma is the main ingredient of the proof, as it proves the CAA property for $\tilde{U} \odot \tilde{V}$. Theorem 4.9 then follows immediately from Theorem 4.2.

Lemma 4.10. *Suppose $\alpha \in \mathbb{R}^m$ be a unit vector at least k of whose coordinates have magnitude $\ge \delta$. Let U, V be arbitrary (as above), and let $\tilde{U}$ and $\tilde{V}$ be ρ perturbations. Define $P(\tilde{U}, \tilde{V}) = \sum_i \alpha_i \tilde{u}_i \otimes \tilde{v}_i$. Then for $M = (m+n)^2$ and all $h > 0$, we have*

$$\mathbb{P}\left[\|P(\tilde{U}, \tilde{V})\| < \delta h \cdot \frac{\rho^2}{64Mnk}\right] < \exp\left(-\frac{1}{16}kn\log(1/h)\right).$$

Remark. To see why this satisfies the CAA property (hypothesis of Theorem 4.2), note that as long as $m < n^2/C$ for a sufficiently large (absolute) constant C, the term $\frac{kn}{16} \ge 16\min(m, km^{1/2})$, thus it satisfies the condition with $\beta = 1/2$.

Proof. Recall that P is a map that has mn variables whose output is an n^2 dimensional vector. We will argue (using the k large coordinates of α) that its Jacobian has sufficiently many nontrivial eigenvalues with high probability. To see this, observe that for a single term $\alpha_i(\tilde{u}_i \otimes \tilde{v}_i)$, the Jacobian (with respect to only u_i variables) is simply an $n^2 \times n$ matrix, structured as follows: in the jth column, the jth "block" of size n is $\alpha_i \tilde{v}_i$, and the rest of the entries are 0. This holds for all j. Thus if $|\alpha_i| \ge \delta$, this matrix has n singular values $\ge \delta\|\tilde{v}_i\|$.

Next, if we take $\sum_i \alpha_i(\tilde{u}_i \otimes \tilde{v}_i)$, as the set of variables is different for every i, the overall Jacobian is the concatenation of the matrices described above (which is an $(n^2 \times nm)$ matrix). Thus, suppose we consider indices $I = \{i : |\alpha_i| > \delta\}$ and form the matrix (call it W) with columns $\{\alpha_i \tilde{v}_i\}_{i\in I}$. If we argue that W has k' large singular values, then the structure above will imply that the Jacobian has nk' large singular values.

Thus, let us focus on W. Lemma A.3 now shows that for any h, W has at least $k/2$ singular values of magnitude $\geq h\delta\rho$, with probability at least $1-\exp(-\frac{1}{8}kn\log(1/h))$. Thus, the Jacobian has $nk/2$ singular values of magnitude $\geq h\delta\rho$ (with the same probability). Thus, we can apply Theorem 4.7, with parameters $c = h\delta$ and rank $nk/2$. We obtain, for any $h > 0$,

$$\mathbb{P}\left[\|P(\tilde{x})\| < \frac{\rho^2 h^2 \delta}{64Mnk}\right] \leq 2\exp\left(-\frac{1}{8}kn\log(1/h)\right).$$

Replacing h^2 with h, the lemma follows. □

Higher order Khatri-Rao products. The Jacobian property used to show Lemma 4.10 can be extended to higher order Khatri-Rao products. We outline the argument for third order tensors: suppose $\{\tilde{u}_i, \tilde{v}_i, \tilde{w}_i\}$ are ρ-perturbed vectors, and define $P = \sum_i \alpha_i(\tilde{u}_i \otimes \tilde{v}_i \otimes \tilde{w}_i)$, for a coefficient vector α that has k coordinates of magnitude $\geq \delta$. Now, the Jacobian (with respect to the $\tilde{u}$ variables) will have as a sub-matrix, the matrix W whose columns are $\{\alpha_i(\tilde{v}_i \otimes \tilde{w}_i)\}$. Now, we need to show that W has at least $k/2$ large singular values, for k up to $\Omega(n^2)$. Instead of a direct argument, we can now use our result for Khatri-Rao products of two matrices!

The natural idea is to use our result for Khatri-Rao products (i.e., Theorem 4.9) directly. While this shows that all k singular values of W are large enough, the success probability we obtain is not high enough. We would ideally want a success probability close to $1-\exp(-kn)$ (and not "merely" $1-\exp(-n)$ as the Theorem gives us). The key observation is that such an improved bound is possible if we start with the weaker goal of obtaining $k/2$ singular values of large magnitude. Indeed, the following simple lemma shows that for a matrix W with k columns to have $k/2$ large singular values, it suffices to show that $\|W\alpha\|$ is large for all "well spread" vectors α. Specifically, it shows that if W had fewer than $k/2$ large singular values, the space spanned by the small singular values must have a well-spread vector.

Lemma B.2. *Suppose $S \subset \mathbb{R}^n$ is a subspace of dimension k. Then there exists a unit vector $u \in S$ that has at least k entries of magnitude $\geq \frac{1}{k\sqrt{n}}$.*

Proof. Let $U \in \mathbb{R}^{n\times k}$ be a matrix whose columns form an orthonormal basis for S. We can now apply Lemma C.2 to U^T to conclude that there exists a subset J of the *rows* of U such that $\sigma_k(U_{|J}) \leq 1/\sqrt{nk}$. [Here, $U_{|J}$ refers to the matrix U restricted to the rows J.]

Now this implies that *every* vector in the column span of $U_{|J}$ can be expressed as $U_{|J}\alpha$, where $\alpha \in \mathbb{R}^k$ and $\|\alpha\| \leq \sqrt{nk}$. In particular, we can conclude that the vector with $1/\sqrt{k}$ in all k coordinates can be so expressed. By considering $\alpha' = \frac{\alpha}{\|\alpha\|}$, we have that $U\alpha'$ has entries $\geq \frac{1}{\sqrt{k}\|\alpha\|} \geq \frac{1}{k\sqrt{n}}$ in all the entries corresponding to J.

Noting that $|J| = k$ completes the proof. □

Next, for well-spread vectors, we can use Observation 4.5 to conclude that $\|W\alpha\|$ is large with very high probability (around $1-\exp(-kn)$, as desired). Thus, unless W has $k/2$ large singular values, we have a contradiction. Then, we can complete the proof as before, except now we apply Theorem 4.2 with $\beta = 1/3$. We omit the details as they are identical to Theorem 4.9.

Other applications. In the subsequent section, we will see other natural candidates for $\mathcal{M}$ including, most significantly, matrices obtained by applying a linear operator to a Kronecker product of matrices on some base variables. It is natural to ask if we can prove these results using the Jacobian based techniques we saw above. It turns out that this is possible for second order Kronecker products (here the CAA property corresponds to amplification for test "matrices" α that have large rank). But the method is not strong enough to handle higher order Kronecker products. We omit the details, since we can handle the general case via our new technique of smoothed contractions.

C DEFERRED PROOFS FROM SECTION 5

We provide the following theorem.

Theorem 5.1. *Suppose $d \in \mathbb{N}$, and let $\Phi : \mathsf{Sym}^d(\mathbb{R}^n) \to \mathbb{R}^D$ be an orthogonal projection of rank $R = \delta\binom{n+d-1}{d}$ for some constant $\delta > 0$, and let $\mathsf{Sym}_d : (\mathbb{R}^n)^{\otimes d} \to \mathsf{Sym}^d(\mathbb{R}^n)$ be the orthogonal projection on to the symmetric subspace of $(\mathbb{R}^n)^{\otimes d}$. Let $U = (u_i : i \in [m]) \in \mathbb{R}^{n\times m}$ be an arbitrary matrix, and let $\tilde{U}$ be a random ρ-perturbation of U. Then there exists a constant $c_d > 0$ such that for $m \leq c_d\delta n$, with probability at least $1-\exp\big(-\Omega_{d,\delta}(n)\big)$, we have the least singular value*

$$\sigma_{\binom{m+d-1}{d}}(\Phi\tilde{U}^{\circledast d}) \geq \frac{\rho^d}{n^{O(d)}}, \text{ where}$$

$$\tilde{U}^{\circledast d} := \Big(\mathsf{Sym}_d(\tilde{u}_{i_1} \otimes \tilde{u}_{i_2} \cdots \otimes \tilde{u}_{i_d}) : 1 \leq i_1 \leq i_2 \leq \cdots \leq i_d \leq n\Big). \tag{2}$$

In the above statement, one can also consider an arbitrary linear operator Φ and suffer an extra factor of $\sigma_R(\Phi)$ in the least singular value bound (by considering the projector onto the span of the top R singular vectors). In the rest of the section, we assume that Φ is an orthogonal projector of rank R without loss of generality.

Theorem 5.1 follows from the following theorem (Theorem 5.2) which gives a non-symmetric analog of the same statement. The proof of Theorem 5.1 follows from a reduction to Theorem 5.2 that is given by Lemma C.4. In what follows, $\Psi \in \mathbb{R}^{R\times n^d}$ denotes the natural matrix representation of Φ such that $\Psi x^{\otimes d} = \Phi(x^{\otimes d})$ for all $x \in \mathbb{R}^n$.

Theorem 5.2. *Suppose $\ell \in \mathbb{N}$, $R = \delta\binom{n+d-1}{d}$ for some constant $\delta > 0$ and let $\Psi : (\mathbb{R}^n)^{\otimes \ell} \to \mathbb{R}^D$ be a linear operator with $\sigma_R(\Psi) \geq 1$. Suppose random matrices $\tilde{U}^{(1)}, \ldots, \tilde{U}^{(d)} \in \mathbb{R}^{n\times m}$ are generated as follows:*

$$\forall j \in [d],\ \tilde{U}^{(j)} = U^{(j)} + Z^{(j)},\ \text{where } Z^{(j)} \sim_{i.i.d} \mathcal{N}(0,\rho^2)^{n\times m} \text{ and is independent of } U^{(j)}, \tag{3}$$

while $U^{(j)} \in \mathbb{R}^{n\times m}$ is arbitrary and can also depend on $\tilde{U}^{(j+1)}, \ldots, \tilde{U}^{(d)}$. Then there exists constants $c_d, c'_d > 0$ and an absolute constant $c_0 \geq 1$

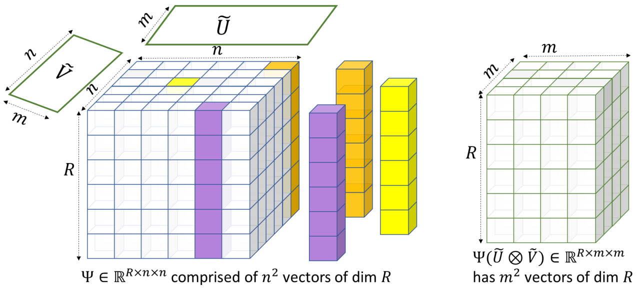
m
$\tilde{U}$
n
$\tilde{V}$
n
n
m
R
$\Psi \in \mathbb{R}^{R\times n\times n}$ comprised of n^2 vectors of dim R
m
m
R
$\Psi(\tilde{U} \otimes \tilde{V}) \in \mathbb{R}^{R\times m\times m}$
has m^2 vectors of dim R

Figure 4: *Left*: The linear operator $\Psi : \mathbb{R}^{n\times n} \to \mathbb{R}^R$ interpreted as a tensor consisting of a $n \times n$ array of R-dimensional vectors. There are *smoothed* or random contractions applied using matrices $\tilde{U}, \tilde{V} \in \mathbb{R}^{n\times m}$. *Right*: The operator $\Psi(\tilde{U} \otimes \tilde{V}) : \mathbb{R}^{m\times m} \to \mathbb{R}^R$ interpreted as an m^2 array of R-dimensional vectors. Theorem 5.2 shows that under the conditions of the theorem, with high probability the robust rank of this operator is m^2 i.e, the least singular value of $R \times m^2$ matrix is inverse polynomial.

such that for $m \le c_d \delta n$, with probability at least $1 - \exp\big(-\Omega_{d,\delta}(n)\big)$, we have

$$\sigma_{m^d}\Big(\Psi\big(\tilde{U}^{(1)} \otimes \cdots \otimes \tilde{U}^{(d)}\big)\Big) \ge \frac{c'_d \rho^d}{n^{c_0 d}}. \tag{4}$$

While Ψ is specified as a matrix of dimension $R \times n^d$ in Theorem 5.2, one can alternately view Ψ as a $(d+1)$-order tensor of dimensions $R \times n \times n \times \cdots \times n$ as shown in Figure 4. Theorem 5.2 then gives a lower bound for the multilinear rank (in fact, for its robust analog) under smoothed modal contractions along the d modes of dimension n each.

In the above statements, one can also consider an arbitrary linear operator Φ with $\sigma_R(\Phi) \ge 1$ and obtain the same consequence. Applying Theorem 5.1 along with the block leave-one-out approach (see Lemma A.2) we arrive at the following corollary.

Corollary 5.3. *Suppose $d, t \in \mathbb{N}$ and let $1 \ge \delta_1 > \delta_2 > 0$ be given. Also let $\Phi : \mathsf{Sym}^d(\mathbb{R}^n) \to \mathbb{R}^D$ be an orthogonal projection of rank $R \ge \delta_1\binom{n+d-1}{d}$. Let $\{U_j\}_{j=1}^t \subset \mathbb{R}^{n\times m}$ be an arbitrary collection of $n \times m$ matrices, and for each j, let $\tilde{U}_j$ be a random ρ-perturbation of U_j. Then there exists a constant $c_d > 0$ such that if $t\binom{m+d-1}{d} \le \delta_2\binom{n+d-1}{d}$ and $m \le c_d(\delta_1 - \delta_2)n$, then with probability at least $1 - \exp\big(-\Omega_{d,\delta_1,\delta_2}(n)\big)$, we have the least singular value*

$$\sigma_{t\binom{m+d-1}{d}}\Big(\Phi\big[\tilde{U}_1^{\circledast d} \quad \tilde{U}_2^{\circledast d} \quad \ldots \quad \tilde{U}_t^{\circledast d}\big]\Big) \ge \frac{\rho^d}{\sqrt{t}\,n^{O(d)}}. \tag{5}$$

C.1 Proof of Theorem 5.2

We will prove Theorem 5.2 for general d by induction on d. The following crucial lemma considers a linear operator Ψ acting on the space $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$, and shows that if Ψ has large rank $\Omega(n_1 n_2)$, then it has many "blocks" of large relative rank as described in Section 2.3.

Lemma 5.4. *Let $\Psi \in \mathbb{R}^{R\times(n_1 n_2)}$ be a projection matrix of rank $R = \delta n_1 n_2$ for some constant $\delta > 0$, and let $\Psi = [\Psi_1 \ \Psi_2 \ \ldots \ \Psi_{n_1}]$ where the blocks $\Psi_i \in \mathbb{R}^{R\times n_2}$ $\forall i \in [n_1]$. Then there exists constants $c_1, c_2, c_3 > 0$ and a subset $S_1 \subset [n_1]$ with $|S_1| \ge c_1 \delta n_1$ such that*

$$\forall i \in S_1,\ \sigma_{c_2 \delta n_2}\Big(\Pi^{\perp}_{S_1 \setminus \{i\}} \Psi_i\Big) \ge \frac{1}{(nk)^{c_3}}, \tag{6}$$

where $\Pi^{\perp}_S$ is the projection orthogonal to $\operatorname{span}\big(\cup_{i\in S} \operatorname{colspan}(\Psi_i)\big)$.

We prove this lemma in Section C.2 by restricting to randomly chosen columns as described in the overview (Section 2.3). We now proceed with the proof of Theorem 5.2 assuming the above lemma.

The following lemma will be important in the inductive proof of the theorem. It reasons about the robust rank (also called multi-linear rank) after the modal contraction by a smoothed matrix along a specific mode. The lemma is proved in slightly more generality; we will use it for the theorem with $\varepsilon = 1$.

Lemma 5.5 (Robust rank under random contractions). *Suppose $\varepsilon \in (0, 1]$ is a constant. For every constant $\gamma, C > 0$, there is a constant $c \in (0, 1)$ such that the following holds for all $s = 2^{o(k)}$. Consider matrices $A_1, A_2, \ldots, A_s \in \mathbb{R}^{R\times k}$, $C_1, \ldots, C_s \in \mathbb{R}^{R\times m}$ and $\forall j \in [s]$ let $\Pi^{\perp}_{-j}$ denote the projector orthogonal to the span of the column spaces of $\{A_{j'} : j' \ne j, j' \in [s]\}$. Suppose the following conditions are satisfied:*

$$\forall j \in [s],\ \sigma_{\varepsilon k}(\Pi^{\perp}_{-j} A_j) \ge k^{-\gamma} \tag{7}$$

and $\sigma_1(A_j), \sigma_1(C_j) \le k^C$. For a random ρ-perturbed matrix $\tilde{U} \in \mathbb{R}^{k\times m}$ with $m \le c\varepsilon k$, we have with probability at least $1 - \exp(-\Omega(\varepsilon k))$ that

$$\text{if } \forall j \in [s],\ M_j = C_j + A_j \tilde{U},\ \text{then } \sigma_{sm}\Big(M_1 \mid \cdots \mid M_s\Big) \ge \frac{\rho}{2k^{\gamma+1}\sqrt{s}}.$$

Proof. Let $\tilde{U} = U + Z$ where $Z \sim N(0, \rho^2)^{k\times m}$ is the random perturbation. Denote $M = (M_1 \mid \cdots \mid M_s)$. Recall $\Pi^{\perp}_{-j^*}$ is the projector orthogonal to the span of the column spaces of $\{A_j : j \ne j^*, j \in [s]\}$. We prove that with high probability, for any (test) unit vector $\alpha \in \mathbb{R}^{s\cdot m}$, we have $\|M\alpha\|_2$ is non-negligible. A standard argument would consider a net over all potential unit vectors $\alpha \in \mathbb{R}^{sm}$. However this approach fails here, since we cannot get high enough concentration (of the form $e^{-\Omega(sm)}$) that is required for this argument. Instead, we argue that if there were such a test vector $\alpha \in \mathbb{R}^{s\cdot m}$, there exists a block $j^* \in [s]$ where we observe a highly unlikely event.

We will use the following simple claim that is proven using a standard net argument.

Claim C.1. *In the above notation, given a (fixed) vector $w \in \mathbb{R}^R$, and a random matrix $Z \sim N(0, \rho^2)^{R\times k}$ with i.i.d entries, we have that with probability at least $1 - \exp(-\Omega(k))$ that*

$$\forall v \in \mathbb{R}^m \text{ with } \|v\|_2 = 1 \text{ and } \forall j \in [s],\ \|w + \Pi^{\perp}_{-j}(A_j Z)v\|_2 \ge \frac{\rho}{2k^{\gamma+1}}$$

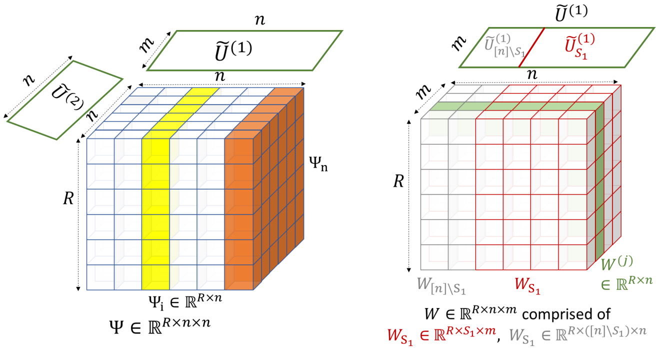
n
m
$\tilde{U}^{(1)}$
n
$\tilde{U}^{(2)}$
n
n
R
Ψ_n
$\Psi_i \in \mathbb{R}^{R\times n}$
$\Psi \in \mathbb{R}^{R\times n\times n}$
$\tilde{U}^{(1)}$
m
$\tilde{U}^{(1)}_{[n]\setminus S_1}$
$\tilde{U}^{(1)}_{S_1}$
n
m
R
$W_{[n]\setminus S_1}$
W_{S_1}
$W^{(j)} \in \mathbb{R}^{R\times n}$
$W \in \mathbb{R}^{R\times n\times m}$ comprised of
$W_{S_1} \in \mathbb{R}^{R\times S_1\times m}$, $W_{S_1} \in \mathbb{R}^{R\times([n]\setminus S_1)\times n}$

Figure 5: *Left*: The setting of $d = 2$ with linear operator $\Psi : \mathbb{R}^{n\times n} \to \mathbb{R}^R$ having slices $\Psi_1, \dots, \Psi_n \in \mathbb{R}^{R\times n}$. The modal contractions $\tilde{U}^{(1)}, \tilde{U}^{(2)} \in \mathbb{R}^{n\times m}$ have not yet been applied. *Right*: After modal contraction along $U^{(2)} \in \mathbb{R}^{n\times m}$, we get $W \in \mathbb{R}^{R\times n\times m}$ with lateral slices $W_1, \dots, W_n$. The subtensor $W_{S_1} \in \mathbb{R}^{R\times|S_1|\times m}$ represents the slices obtained from the "good" blocks $S_1 \subset [n]$, and $W_{[n]\setminus S_1} \in \mathbb{R}^{R\times|[n]\setminus S_1|\times m}$ represents the remaining slices. The random modal contraction $\tilde{U}^{(1)}$ can also now be split into $\tilde{U}^{(1)}_{S_1} \in \mathbb{R}^{S_1\times m}, \tilde{U}^{(1)}_{[n]\setminus S_1} \in \mathbb{R}^{[n]\setminus S_1\times m}$. Let $W^{(j)} \in \mathbb{R}^{R\times n}$ denote the jth frontal slice for each $j \in [m^{d-1}]$. Then the final matrix slice obtained for each $j \in [m^{d-1}]$ can be written as $M^{(j)} = W^{(j)}_{S_1}\tilde{U}^{(1)}_{S_1} + W^{(j)}_{[n]\setminus S_1}\tilde{U}^{(1)}_{[n]\setminus S_1}$, where the randomness in the two summands is independent.

Proof of Claim. We first prove the claim using a net argument over test vectors $v \in \mathbb{R}^m$. Consider a fixed $j \in [s]$; we will do a union bound over all $j \in [s]$.

Let $v \in \mathbb{R}^m$ be a fixed unit vector. Let $A' = \Pi^{\perp}_{-j}A_j$. Observe that $Zv \sim N(0, \rho^2 I)$ is a random Gaussian vector with i.i.d entries each with mean 0 and variance 1. By assumption $\sigma_{\varepsilon k}(A') = \sigma_{\varepsilon k}(\Pi^{\perp}_{-j}A_j) \geq k^{-\gamma}$. Let the SVD of $A' = E\mathrm{diag}(\mu)F^{\top} = \sum_{i=1}^{k}\mu_i e_i f_i^{\top}$ where $\mu_i \geq 0\ \forall i \in [k]$ and $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{\varepsilon k} \geq k^{-\gamma}$.

$$A'(Zv) = E\mathrm{diag}(\mu)(FZv) = E\mathrm{diag}(\mu)\zeta = \sum_{i=1}^{k}\zeta_i\mu_i e_i,$$

where $\zeta \sim N(0, I)$ is a random Gaussian vector with i.i.d entries. Moreover $\mu_i \geq k^{-\gamma}$ for all $i \in [\varepsilon k]$. Hence $\forall \delta \in (0, 1/2)$,

$$\begin{aligned}
&\mathbb{P}\left[\|w + \Pi^{\perp}_{-j}A_jZv\| < \frac{\delta\rho}{k^{\gamma}}\right] \\
&\leq \mathbb{P}\left[\forall i \in [\varepsilon k], |\langle w, e_i\rangle + \langle e_i, \Pi^{\perp}_{-j}A_jZv\rangle| < \frac{\delta\rho}{k^{\gamma}}\right] \\
&= \prod_{i=1}^{\varepsilon k}\mathbb{P}\left[|\langle w, e_i\rangle + \mu_i\zeta_i| < \frac{\delta\rho}{k^{\gamma}}\right] \leq (c_0\delta)^{\varepsilon k},
\end{aligned}$$

for some absolute constant $c_0 > 0$. The equality used the independence of $(\zeta_i : i \in [k])$, while the last inequality used that $\mu_i \geq k^{-\gamma}$ for $i \leq \varepsilon k$ along with standard anti-concentration of a Gaussian r.v.

Set $\varepsilon' := \delta/(4k^{C+1/2})$. Consider an ε'-net $\mathcal{N}_{\varepsilon'}$ over unit vectors in $\mathbb{R}^m$. By a union bound over $\mathcal{N}_{\varepsilon'}$, we get

$$\begin{aligned}
&\mathbb{P}\left[\forall j \in [s], \forall \hat{v} \in \mathcal{N}_{\varepsilon}, \|w + \Pi^{\perp}_{-j}A_jZ\hat{v}\| \geq \frac{\delta\rho}{k^{\gamma}}\right] \\
&\geq 1 - s(c\delta)^{\varepsilon k}|\mathcal{N}_{\varepsilon'}| \\
&\geq 1 - \exp\Big(-\varepsilon k\log\left(\tfrac{1}{c\delta}\right) + \log s + m\log(2/\varepsilon')\Big) \\
&\geq 1 - \exp(-\Omega(k)),
\end{aligned}$$

by picking $\delta = 1/k$, and $m \leq c\varepsilon k$ for an appropriately small constant $c > 0$ (depending on C). Finally conditioned on the above event, for any unit vector $v \in \mathbb{R}^k$, we can consider the closest point $\hat{v}$ in $\mathcal{N}_{\varepsilon'}$ and conclude that $\forall j \in [s], \|\Pi^{\perp}_{-j}A_jZv\| \geq \|\Pi^{\perp}_{-j}A_jZv\| - \|A_j\|\|Z\|\|v - \hat{v}\| \leq O(k^C \cdot \sqrt{k}\rho)\varepsilon' \geq \frac{\delta\rho}{2k^{\gamma}}$.

Finishing the proof of Lemma 5.5 Let us condition on the event that the conclusion of Claim C.1 holds; note that this holds with probability at least $1 - \exp(-\Omega(\varepsilon k))$.

Suppose for contradiction there exists a vector $\alpha \in \mathbb{R}^{s\cdot k}$ such that $\|M\alpha\|_2 < \frac{\rho}{k^{\gamma+1}\sqrt{s}}$. The vector $M\alpha \in \mathbb{R}^R$ is

$$\begin{aligned}
M\alpha = \sum_{j\in[s]}M_j\alpha^{(j)} &= \sum_{j\in[s]}(C_j + A_j(U + Z))\alpha^{(j)} &(13)\\
&= \sum_{j\in[s]}(C_j + A_jU)\alpha^{(j)} + \sum_{j\in[s]}A_jZ\alpha^{(j)} \\
&= b_\alpha + \sum_{j\in[s]}A_jZ\alpha^{(j)}, &(14)
\end{aligned}$$

where b_α is a fixed vector in $\mathbb{R}^R$. We have for all $j^* \in [s]$

$$\begin{aligned}
\Pi^{\perp}_{-j^*}M\alpha &= \Pi^{\perp}_{-j^*}b_\alpha + \sum_{j\in[s]}\Pi^{\perp}_{-j^*}A_jZ\alpha^{(j)} \\
&= \Pi^{\perp}_{-j^*}b_\alpha + \Pi^{\perp}_{-j^*}A_{j^*}(Z\alpha^{(j^*)}), &(15)
\end{aligned}$$

In the above, (15) holds since $\Pi^{\perp}_{-j^*}$ is orthogonal to the column spaces of all $j \neq j^*$ and $A' = \Pi^{\perp}_{-j^*}A_{j^*}$.

Now consider any index $j^* \in [s]$ such that $\|\alpha^{(j^*)}\|_2 \geq 1/\sqrt{s}$ (note $\sum_{j=1}^{s}\|\alpha^{(j)}\|_2^2 = 1$). Now applying Claim C.1 with $j = j^*, v = \alpha^{(j^*)}/\|\alpha^{(j^*)}\|$ and $w = \Pi^{\perp}_{-j^*}b_\alpha/\|\alpha^{(j^*)}\|$, we get that

$$\|M\alpha\|_2 \geq \|\Pi^{\perp}_{-j^*}M\alpha\| \geq \frac{1}{\sqrt{s}} \cdot \frac{\rho}{2k^{\gamma+1}},$$

which contradicts the assumption. This concludes the proof.

□

Proof of Theorem 5.2. We now proceed by induction on d. For the proof it will be useful to think of ρ as a sufficiently small inverse polynomial (this is without loss of generality and suffers only a poly(n) extra factor in the bound).

The base case $d = 1$ follows by simple random matrix arguments; specifically, Lemma 5.5 applied with $s = 1, d = 1$ implies it.

For higher d, we will apply the induction hypothesis for modal contractions along the last $d - 1$ modes using matrices $\tilde{U}^{(2)}, \dots, \tilde{U}^{(d)}$, and then finally apply modal contraction along $\tilde{U}^{(1)}$.

Set $n_1 = n, n_2 = n^{d-1}$. First applying Lemma 5.4 with Ψ, we get a set of blocks $\{\Psi_i : i \in S_1\}$ with $|S_1| = \Omega(\delta n_1)$, satisfying (6). Define for each $i \in S_1$, $\mathcal{V}_i := \text{colspan}(\Psi_i)$, and $\mathcal{V}_{-i} := \text{span}(\cup_{j \in S_1, j \neq i}\text{colspan}(\Psi_i))$ and let $\Pi^{\perp}_{-i}$ be the projection matrix for the subspace orthogonal to $\mathcal{V}_{-i}$.

In other words, suppose for each $i \in [n]$ that $W_i := \Psi_i\big(\tilde{U}^{(2)} \otimes \tilde{U}^{(3)} \otimes \cdots \otimes \tilde{U}^{(d)}\big) \in \mathbb{R}^{R \times m^{d-1}}$. Then for absolute constants $c, c' > 0, c'' > 0$,

$$\sigma_{c\delta n^{d-1}}\Big(\Pi^{\perp}_{-i}\Psi_i\Big) \geq \frac{c''}{(nk)^{c'}} \text{ for all } i \in S_1.$$

By using the induction hypothesis with order $(d-1)$ with the matrices $\{\Pi^{\perp}_{-i}\Psi_i\}$ along with a union bound over the n blocks, for appropriate constants $c_0 > 0$ and $c'_{d-1} > 0$,

$$\begin{aligned}\forall i \in S_1,\ \sigma_{m^{d-1}}\Big(\Pi^{\perp}_{-i}W_i\Big) &= \sigma_{m^{d-1}}\Big(\Psi_i\big(\tilde{U}^{(2)} \otimes \cdots \otimes \tilde{U}^{(d)}\big)\Big)\\ &\geq \frac{c'_{d-1}\rho^{d-1}}{n^{c_0(d-1)}}.\end{aligned}$$

Let $W \in \mathbb{R}^{R \times n \times m^{d-1}}$ be the tensor obtained by stacking the matrices $W_i \in \mathbb{R}^{R \times m^{d-1}}$ as shown in the Figure 5. Let $W_{S_1} \in \mathbb{R}^{R \times S_1 \times m^{d-1}}$ denote the subtensor comprising just the slices $i \in S_1$, and let $W_{[n]\setminus S_1}$ be the remaining portion. For each $j \in [m^{d-1}]$, let $W^{(j)} \in \mathbb{R}^{R \times n}$ be obtained from the slices along the third mode. We will use $W^{(j)}_{S_1}, W^{(j)}_{[n]\setminus S_1}$ to denote the portions of the slices $W^{(j)}$ formed by the columns S_1 and $[n] \setminus S_1$ respectively. If $W^{\text{flat}}_{S_1} \in \mathbb{R}^{R \times (|S_1|m^{d-1})}$ is the matrix obtained by flattening W_{S_1} appropriately, then by Lemma A.2 on the block leave-one-out distance,

$$\begin{aligned}\sigma_{|S_1|m^{d-1}}\Big(W^{\text{flat}}_{S_1}\Big) &= \sigma_{|S_1|m^{d-1}}\Big(W^{(1)}_{S_1} \mid \cdots \mid W^{(m^{d-1})}_{S_1}\Big)\\ &\geq \frac{c'_{d-1}\rho^{d-1}}{n^{c_0(d-1)+\frac{1}{2}}}.\end{aligned} \tag{16}$$

The final matrix is obtained by concatenating the matrices $M^{(j)} \in \mathbb{R}^{R \times m}$ for each $j \in [m^{d-1}]$, where

$$\begin{aligned}M^{(j)} &= W^{(j)}\tilde{U}^{(1)} = W^{(j)}_{S_1}\tilde{U}^{(1)}_{S_1} + W^{(j)}_{[n]\setminus S_1}\tilde{U}^{(1)}_{[n]\setminus S_1}\\ &= W^{(j)}_{S_1}\tilde{U}^{(1)}_{S_1} + C^{(j)}, \text{ where } C^{(j)} := W^{(j)}_{[n]\setminus S_1}\tilde{U}^{(1)}_{[n]\setminus S_1}.\end{aligned}$$

Consider any fixed $j \in [m^{d-1}]$. We will treat $C^{(j)}$ as fixed matrices. Note that the randomness in $\tilde{U}^{(1)}_{S_1}$ is independent of the randomness in $\tilde{U}^{(1)}_{[n]\setminus S_1}$. Now we can apply Lemma 5.5 with $s = m^{d-1}, A_j = W^{(j)}_{S_1}$ and $\tilde{U} = \tilde{U}^{(1)}_{S_1}$ and $C_j = C^{(j)}$ to conclude the inductive proof of Theorem 5.2. □

C.2 Finding many blocks with large relative rank

We note that while the statement of the lemma is quite intuitive, the proof is non-trivial because we require that in any selected block, there must be many vectors with a large component orthogonal to the *entire span* of the other selected blocks. As a simple example, consider setting $n_2 = 2t$ and $\Psi_1 = \{e_1, e_2, \ldots, e_t, \varepsilon e_{t+1}, \varepsilon e_{t+2}, \ldots, \varepsilon e_{2t}\}$, and $\Psi_2 = \{\varepsilon e_1, \varepsilon e_2, \ldots, \varepsilon e_t, e_{t+1}, e_{t+2}, \ldots, e_{2t}\}$. In this case, even if ε is tiny, we cannot choose both the blocks, because the span of the vectors in Ψ_2 contains all the vectors in Ψ_1.

The proof will proceed by first identifying a set of roughly R vectors (spread across the blocks) that form a well conditioned matrix, followed by randomly restricting to a subset of the blocks.

We start with the following lemma, which gives us the first step.

Lemma C.2. *Suppose A is an $m \times n$ matrix such that $\sigma_k(A) \geq \theta$. Then there exists a submatrix A_S with $|S| = k$ columns, such that $\sigma_k(A_S) \geq \theta/\sqrt{nk}$.*

Remark. The lemma is a robust version of the simple statement that if $\sigma_k(A) > 0$, then there exist k linearly independent columns.

Proof. We start by noting that we can restrict to the case $m = k$. This is because we can project the columns of A onto the span of the top k singular vectors of A and pick the S using the resulting matrix. Formally, if Π is the $(k \times m)$ matrix that defines the projection, then we work with ΠA. (By definition, $\sigma_k(\Pi A) = \sigma_k(A) \geq \theta$, so the hypothesis of the lemma holds.) For the obtained set S, it is easy to see that the vectors before projection will satisfy, for any test vector α,

$$\|\sum_i \alpha_i v_i\| \geq \|\sum_i \alpha_i \Pi v_i\|.$$

Thus if we show a lower bound for $\sigma_k(\Pi A_S)$, the same bound holds for $\sigma_k(A_S)$. So in what follows, assume that $m = k$.

Next, we find an Auerbach basis [20] (also referred to as a Barycentric spanner or a well-conditioned basis [6, 18]) for the columns of A. Recall that this is a subset of the columns of A defined by a subset S of indices such that $|S| = k$, and for all $i \in [n]$, A_i can be expressed as $\sum_{j \in S}\alpha_j A_j$ with $|\alpha_j| \leq 1$.

We claim that for this choice of S, we have a lower bound on $\sigma_k(A_S)$. Suppose not; suppose $\|\sum_{i \in S}\alpha_i A_i\| < \theta/\sqrt{nk}$ for some unit vector α (whose non-zero entries are indexed by S). Since α is a unit vector with at most k non-zeros, one of its coefficients, say α_j, must be $\geq 1/\sqrt{k}$. Thus we have $A_j = x + w$, for some $x \in \text{span}(A_{S\setminus\{j\}})$ and $\|w\| \leq \theta/\sqrt{n}$.

Next, consider any column A_ℓ for $\ell \notin S$. From the above, we have that A_ℓ's projection orthogonal to the span of $A_{S\setminus\{j\}}$ is at most $\theta/\sqrt{n}$ (because of the Auerbach basis property, and the fact that A_j is almost in the span of $A_{S\setminus\{j\}}$). This implies that the squared rank-$(k-1)$ approximation error (in the Frobenius norm) of the matrix A is $\leq (n-k)\theta^2/n < \theta^2$, which contradicts the fact that $\sigma_k(A) \geq \theta$. □

We can now complete the proof of Lemma 5.4.

Proof of Lemma 5.4. The outline of the argument is as follows:

(1) First find a subset M of $R = \delta n_1 n_2$ columns of Ψ such that $\sigma_R(M)$ is large (using Lemma C.2).

(2) Randomly sample a subset $T \subseteq [n_1]$ of the blocks.

(3) Discard any block $j \in T$ that has fewer than $\delta n_2/6$ vectors with a non-negligible component orthogonal to the span of $\cup_{r \in (T \setminus \{j\})} \Psi_r$; argue that there are $\Omega(\delta n_1)$ blocks remaining.

The first step is a direct application of Lemma C.2; we thus obtain M with $R = \delta n_1 n_2$ columns such that

$$\sigma_R(M) \geq \frac{1}{n_1 n_2 \sqrt{\delta}}. \tag{17}$$

For convenience, we will denote the columns of M by $v_1, v_2, \ldots, v_R$.

Now for the second step of the outline: $T \subseteq [n_1]$ is selected by including each block j in T with probability equal to $|M \cap \Psi_j|/6n_2$. I.e., the probability is proportional to the fraction of the "M" columns contained in a block. For convenience, we will write $\alpha_j = |M \cap \Psi_j|/n_2$.

Step (3) of the outline is thus the bulk of the argument. We start by introducing two random variables. First, for $j \in [n_1]$, define X_j to be the indicator that is 1 if block j is chosen in T and 0 otherwise. Thus by definition, $\mathbb{P}[X_j = 1] = \alpha_j/6$, and the X_j are independent for different j. Second, for $i \in [R]$, if j is the index of the block that contains v_i, we define Y_i to be 1 if the vector v_i has a projection of length $\geq \frac{1}{R n_1 n_2 \sqrt{\delta}}$ orthogonal to the span of all the columns in $\cup_{r \in T \setminus \{j\}} \Psi_r$ and 0 otherwise.

Now, note that a block j "survives" step (3) of the outline above if (a) $j \in T$ to start with, and (b) $\sum_{v_i \in \Psi_j} Y_i \geq \delta n_2/6$. [This is a sufficient condition for survival, not an equivalence.] Thus, if Q_j is a random variable indicating if block j survives, we can write

$$Q_j \geq \frac{X_j \left(\sum_{v_i \in \Psi_j} Y_i - \frac{\delta n_2}{6} \right)_+}{n_2 \alpha_j}. \tag{18}$$

Here, for a random variable Z, the notation $(Z)_+$ denotes $\max\{Z, 0\}$. We will use the RHS expression to give a positive lower bound on $\mathbb{E}[\sum_{j \in [n_1]} Q_j]$. Note that this will complete the proof of the lemma, because we are only interested in an existential statement.

To this end, the key observation is that for any $v_i \in \Psi_j$, the random variable Y_i is *independent* of X_j. This is because by definition, Y_i indicates if v_i had a large enough component orthogonal to the span of the *other* chosen blocks (irrespective of whether block j is chosen or not). Thus, since $\mathbb{E}[X_j] = \alpha_j/6$, we have that

$$\mathbb{E}\left[\frac{X_j \left(\sum_{v_i \in \Psi_j} Y_i - \frac{\delta n_2}{6} \right)_+}{n_2 \alpha_j} \right]$$

$$= \frac{1}{6n_2} \mathbb{E}\left[\left(\sum_{v_i \in \Psi_j} Y_i - \frac{\delta n_2}{6} \right)_+ \right] \geq \frac{1}{6n_2} \left(\mathbb{E}[\sum_{v_i \in \Psi_j} Y_i] - \frac{\delta n_2}{6} \right).$$

Thus, we have

$$\sum_{j \in [n_1]} Q_j \geq \frac{1}{6n_2} \left(\mathbb{E}\left[\sum_{i \in [R]} Y_i \right] - \frac{\delta n_1 n_2}{6} \right). \tag{19}$$

So it complete the proof, it suffices to prove that $\mathbb{E}[\sum_i Y_i]$ is sufficiently large. We do this by introducing an auxiliary random variable Z_i. For any $i \in [R]$, define Z_i to be the random variable that is 1 if v_i has a projection of length $\geq \frac{1}{R n_1 n_2 \sqrt{\delta}}$ orthogonal to the span of the vectors in *all* the chosen blocks, $\cup_{r \in T} \Psi_r$.

Thus by definition, the inequality $Z_i \leq Y_i$ always holds, and Z_i will be zero if $X_j = 1$ (where Ψ_j is the block that contains v_i). We will prove that in fact, $\mathbb{E}[\sum_i Z_i]$ is large. Observe that by the law of conditional expectation,

$$\mathbb{E}[\sum_i Z_i] = \sum_T \mathbb{P}[T] \cdot \mathbb{E}[\sum_i Z_i | T].$$

Indeed, the last term is deterministic conditioned on T (so it is simply the number of i for which Z_i is 1 for the chosen T). We split the sum into two, depending on $|T|$.

$$\mathbb{E}\left[\sum_i Z_i \right] = \sum_{T : |T| > 2n_1 \delta/3} \mathbb{P}[T] \cdot \mathbb{E}\left[\sum_i Z_i | T \right]$$
$$+ \sum_{T : |T| \leq 2n_1 \delta/3} \mathbb{P}[T] \cdot \mathbb{E}\left[\sum_i Z_i | T \right].$$

We will simply ignore the first sum, as our goal is to obtain a lower bound. To show that this is good enough, we first observe that

$$\mathbb{E}[|T|] = \sum_{j \in [n_1]} X_j = \sum_j \frac{\alpha_j}{6} \leq \frac{\delta n_1}{6}.$$

Thus by Markov's inequality, $\mathbb{P}[|T| \leq 2n_1 \delta/3] \geq 3/4$. Let us thus condition on one such T.

Claim. For any T with $|T| \leq 2\delta n_1/3$, we have $\sum_{i \in [R]} Z_i \geq \frac{\delta n_1 n_2}{3}$.

Informally, v_i are vectors that are all "well conditioned", and thus many of them must have a component orthogonal to any subspace of dimension $< R/2$.

This can be made formal as follows: let $\mathcal{S}$ be the subspace span $(\cup_{r \in T} \Psi_r)$. Clearly, its dimension is $\leq |T| n_2 \leq \delta 2 n_1 n_2/3 = 2R/3$. Now from our definition of $\{v_i\}$, the matrix M whose columns are the v_i has $\sigma_R(M)$ bounded as in (17). Thus, if $\Pi_{\mathcal{S}}^{\perp}$ is the matrix that projects every vector to the space $\mathcal{S}^{\perp}$, we have, by the Min-Max characterization of eigenvalues,

$$\sigma_{R - \dim(\mathcal{S})}(\Pi_{\mathcal{S}}^{\perp} M) \geq \sigma_R(M) \geq \frac{1}{n_1 n_2 \sqrt{\delta}}.$$

Thus, at least $R - \dim(\mathcal{S}) \geq \delta n_1 n_2/2$ columns of $\Pi_{\mathcal{S}}^{\perp} M$ must have *length* $\geq \frac{1}{R n_1 n_2 \sqrt{\delta}}$.[9]

This will let us conclude that $\mathbf{E}[\sum_i Z_i] \geq \delta n_1 n_2/3$, thus completing the proof of the claim.

Next, we use the claim together with our observations above to conclude that

$$\mathbb{E}\left[\sum_i Z_i \right] \geq \frac{\delta n_1 n_2}{3} \cdot \mathbb{P}\left[|T| \leq 2n_1 \delta/3 \right] \geq \frac{\delta n_1 n_2}{3} \cdot \frac{3}{4} = \frac{\delta n_1 n_2}{4}.$$

Plugging this into (19), we obtain $\sum_{j \in [n_1]} Q_j \geq \Omega(n_1)$, thus completing the proof. □

[9] Here we are using the simple observation that if $\sigma_k(X) \geq \delta$ for a matrix X with C columns, then at least k of the columns must be $\geq \delta/C$. This holds because if not, we can project to the space orthogonal to at most $(k-1)$ columns and have every column being of length $< \delta/C$, which means the max singular value of the matrix with these projected columns is $< \delta$); this contradicts the assumption on $\sigma_k(X) \geq \delta$.

C.3 From Symmetric to Non-Symmetric Products

Recall $U \in \mathbb{R}^{n\times m}$ and $\tilde{U} = U + Z$ where $U = (u_i : i \in [m])$ is an arbitrary matrix and $Z \in \mathbb{R}^{n\times m}$ is a random matrix with i.i.d. entries drawn from $\mathcal{N}(0,\rho^2)$. In what follows, $\Phi : \mathsf{Sym}(\mathbb{R}^{n^d}) \to \mathbb{R}^r$ denotes an operator acting on the symmetric space, and let $\Psi \in \mathbb{R}^{r\times n^d}$ denote the natural matrix representation of Φ such that $\Phi(x^{\otimes d}) = \Psi x^{\otimes d}$, and where where every row of Ψ corresponds to a symmetric matrix.

Additionally, we define $\mathrm{Perm}_{\mathrm{avg}} \in \mathbb{R}^{m^d \times \binom{m+d-1}{d}}$ to be the unique matrix with the property that for any collection of matrices $\{V^{(i)}\}_{i=1}^d \subset \mathbb{R}^{n\times m}$, we have that $\left(\otimes_{i=1}^d V^{(i)}\right)\mathrm{Perm}_{\mathrm{avg}} \in \mathbb{R}^{n^d\times\binom{m+d-1}{d}}$ is the matrix with columns indexed by tuples $(i_1, i_2, \dots, i_d)$ with $1 \le i_1 \le i_2 \le \cdots \le i_d \le m$, where the column corresponding to $(i_1, i_2, \dots, i_d)$ is given by $\frac{1}{|S_d|}\left(\sum_{\pi\in S_d} \otimes_{j=1}^d V^{(j)}_{i_{\pi(j)}}\right)$. Here S_d denotes the permutation group on d indices and $V^{(j)}_{i_{\pi(j)}}$ is the $i_{\pi(j)}$th column of $V^{(j)}$.

The matrix $\mathrm{Perm}_{\mathrm{avg}}$ has two important properties that we note. First, for any matrix U, we have that $(U^{\otimes d})\mathrm{Perm}_{\mathrm{avg}}$ is the matrix with columns $\mathsf{Sym}_d(\tilde{u}_{i_1} \otimes \tilde{u}_{i_2} \otimes \cdots \otimes u_{i_d})$ where $1 \le i_1 \le i_2 \cdots \le i_d \le m$. That is $(U^{\otimes d})\mathrm{Perm}_{\mathrm{avg}} = U^{\circledast d}$. It follows that

$$\begin{aligned}&\Psi\left(U^{\otimes d}\right)\mathrm{Perm}_{\mathrm{avg}}\\ =&\Big(\Phi\big(\mathsf{Sym}_d\ \tilde{u}_{i_1} \otimes \cdots \otimes \tilde{u}_{i_d}\big) : 1 \le i_1 \le \cdots \le i_d \le m\Big).\end{aligned}$$

Second, since Ψ is determined by ts action on $\mathsf{Sym}(\mathbb{R}^{\otimes d})$, we obtain that for any collection of matrices $\{V^{(i)}\}_{i=1}^d \subset \mathbb{R}^{n\times m}$ and any permutation $\pi \in S_d$, one has

$$\Psi\left(\otimes_{i=1}^d V^{(i)}\right)\mathrm{Perm}_{\mathrm{avg}} = \Psi\left(\otimes_{i=1}^d V^{(\pi(i))}\right)\mathrm{Perm}_{\mathrm{avg}}. \qquad (20)$$

The following lemma is useful in our reduction from nonsymmetric products to symmetric products.

Lemma C.3. *Given $d \in \mathbb{N}$, for each $i = 1, \dots, d$, let $Z_j \sim \mathcal{N}(0,\rho_j^2)^{n\times m}$ and set $\rho^2 = \rho_1^2 + \cdots + \rho_d^2$ so that $Z := Z_1 + \cdots + Z_d \sim \mathcal{N}(0,\rho^2)^{n\times m}$. Also let $\tilde{U} = U + Z \in \mathbb{R}^{n\times m}$ be a ρ-smoothed matrix. For each $\ell \in [d]$, set $\tilde{V}^{(\ell)} = \tilde{U} - Z_1 - \cdots - Z_\ell$. Then one has*

$$\Psi\left(\tilde{U}^{\otimes d}\right)\mathrm{Perm}_{avg} = \Psi\left(\otimes_{j=1}^d(\tilde{V}^{(j)} + (d-j+1)Z_j)\right)\mathrm{Perm}_{avg} + \Psi E$$

where the error matrix E is a random matrix that satisfies $\|E\|_F \le c_d(1+\|U\|^{d-2})\rho^2(nm)^{\frac{d}{2}})$ with probability at least $1-\exp(-\Omega(nm))$, for some constant $c_d > 0$ that only depends on d.

Proof. The proof follows an induction argument that carefully leverages symmetry (equation (20)), and groups together terms in a way that decouples the randomness.

We give the proof by induction on k. In particular, we will show that for $k \le d$, we have

$$\begin{aligned}&\Psi\left(\tilde{U}^{\otimes d}\right)\mathrm{Perm}_{\mathrm{avg}}\\ =&\Psi\left(\otimes_{j=1}^k(\tilde{V}^{(j)} + (d-j+1)Z_j) \otimes (\tilde{V}^{(k)})^{\otimes d-k}\right)\mathrm{Perm}_{\mathrm{avg}} + \Psi\left(\sum_{j=1}^k E_j\right)\end{aligned}$$

where for each j we have $\|E_j\|_F \le c'_d O((1+\|U\|^{d-2})\rho_j^2(nm)^{\frac{d}{2}})$ with probability at least $1-\exp(-\Omega(nm))$ for some constant c'_d that depends only on d. Setting $\tilde{V}^{(0)} := \tilde{U}$, the statement trivially holds in the base case of $k = 0$. We now suppose the result is true for $k = \ell$ and prove it true for $\ell + 1$. Set

$$W_\ell := \otimes_{j=1}^\ell(\tilde{V}^{(j)} + (d-j+1)Z_j).$$

Then using our induction hypothesis with the identity $\tilde{V}^{(\ell)} = \tilde{V}^{(\ell+1)} + Z_{\ell+1}$, we obtain

$$\begin{aligned}\Psi\left(\tilde{U}^{\otimes d}\right)\mathrm{Perm}_{\mathrm{avg}} &= \Psi\left(W_\ell \otimes (\tilde{V}^{(\ell)})^{\otimes d-\ell}\right)\mathrm{Perm}_{\mathrm{avg}} + \Psi\left(\sum_{j=1}^\ell E_j\right)\\ &= \Psi\left(W_\ell \otimes (\tilde{V}^{(\ell+1)} + Z_{\ell+1})^{\otimes d-\ell}\right)\mathrm{Perm}_{\mathrm{avg}}\\ &\quad + \Psi\left(\sum_{j=1}^\ell E_j\right)\end{aligned}$$

Applying equation (20) and expanding with the binomial theorem gives

$$\begin{aligned}&\Psi\left(W_\ell \otimes \left(\tilde{V}^{(\ell+1)} + Z_{\ell+1})\right)^{\otimes d-\ell}\right)\mathrm{Perm}_{\mathrm{avg}}\\ =&\Psi\bigg(W_\ell \otimes \Big(\sum_{j=0}^{d-\ell}\binom{d-\ell}{j}\left(Z_{\ell+1}^{\otimes j}\right) \otimes \left((\tilde{V}^{(\ell+1)})^{\otimes d-\ell-j}\right)\Big)\bigg)\mathrm{Perm}_{\mathrm{avg}}\\ =&\Psi\bigg(W_\ell \otimes \bigg(\left(\tilde{V}^{(\ell+1)}\right)^{\otimes d-\ell} + (d-\ell)\left(Z_{\ell+1}\right) \otimes \left(\tilde{V}^{(\ell)}\right)^{\otimes d-\ell+1}\\ &\qquad + E'_{\ell+1}\bigg)\bigg)\mathrm{Perm}_{\mathrm{avg}}\\ =&\Psi\left(W_\ell \otimes \left(\tilde{V}^{(\ell+1)} + (d-\ell)Z_{\ell+1}\right)\right) \otimes \left((\tilde{V}^{(\ell+1)})^{\otimes d-\ell-1}\right)\mathrm{Perm}_{\mathrm{avg}}\\ &+ \Psi(W_\ell \otimes E'_{\ell+1})\mathrm{Perm}_{\mathrm{avg}}\end{aligned}$$

Here

$$E'_{\ell+1} := \sum_{j=2}^{d-\ell}\binom{d-\ell}{j} Z_{\ell+1}^{\otimes j} \otimes (V^{(\ell+1)})^{\otimes d-j}.$$

Setting $E_{\ell+1} := (W_\ell \otimes E'_{\ell+1})\mathrm{Perm}_{\mathrm{avg}}$, we obtain that $\|E_{\ell+1}\|_F \le c'_d O((1+\|U\|^{d-2})\rho_j^2(nm)^{\frac{d}{2}})$ with probability at least $1-\exp(-\Omega(nm))$. From here, observing that

$$W_{\ell+1} = W_\ell \otimes \left(\tilde{V}^{(\ell+1)} + (d-\ell)Z_{\ell+1}\right)$$

completes the proof by induction.

□

Lemma C.4 (Symmetric to Non-symmetric Products). *Suppose $d \in \mathbb{Z}_+$ be a positive integer. Suppose $\Phi : \mathsf{Sym}(\mathbb{R}^{n^d}) \to \mathbb{R}^r$ with $\|\Phi\| \le 1$. For every $\rho_1, \dots, \rho_d > 0$ with $\sum_{j=1}^d \rho_j^2 = \rho^2$ and $\delta \in (0,1)$*

the following holds when $\tilde{U} = U + Z$ is drawn as described above with the entries of Z being drawn i.i.d from $\mathcal{N}(0, \rho^2)$:

$\forall t \geq 0,$

$$\begin{aligned}&\mathbb{P}\left[\sigma_{\binom{m+d-1}{d}}\Big(\Phi\big(\mathsf{Sym}_d \tilde{u}_{i_1} \otimes \cdots \otimes \tilde{u}_{i_d}\big) : 1 \leq i_1 \leq \cdots \leq i_d \leq m\Big) \leq t\right] \\ &\leq \mathbb{P}\left[\sigma_{m^d}\left(\Psi\left(\otimes_{j=1}^{d}(\tilde{V}^{(j)} + (d-j+1)Z_j)\right)\right)\right. \\ &\qquad\qquad \left. \leq \sqrt{d!}\cdot t + \|E\|\log(1/\delta)\right] + \delta. \qquad (21)\end{aligned}$$

Here $V^{(\ell)} = \tilde{U} - Z_1 - \cdots - Z_\ell$ and for each $\ell \in [d]$ and Z_ℓ is a random matrix with i.i.d entries drawn from $\mathcal{N}(0, \rho_\ell^2)$ and E is the error matrix appearing in Lemma C.3 which has norm $\|E\| = O((1 + \|U\|^{d-2})\rho^2(n+m)^{\frac{d}{2}})$.

Proof. First observe that

$$\Phi\big(\mathsf{Sym}_d \tilde{u}_{i_1} \otimes \cdots \otimes \tilde{u}_{i_d}\big) : 1 \leq i_1 \leq \cdots \leq i_d \leq m\Big) = \Psi\left(\tilde{U}^{\otimes d}\right)\mathrm{Perm}_{\mathrm{avg}}$$

Using Lemma C.3 shows then shows that

$$\Psi\left(\tilde{U}^{\otimes d}\right)\mathrm{Perm}_{\mathrm{avg}} = \Psi\left(\otimes_{j=1}^{d}(\tilde{V}^{(j)} + (d-j+1)Z_j)\right)\mathrm{Perm}_{\mathrm{avg}} + \Psi E.$$

Using the fact that $\mathrm{Perm}_{\mathrm{avg}}$ has columns with disjoint support each with ℓ_2 norm at least $\frac{1}{\sqrt{d!}}$ shows that the matrix $\mathrm{Perm}_{\mathrm{avg}}$ has full column rank with all singular values in $[\frac{1}{\sqrt{d!}}, 1]$. Combining this with the above equality gives

$$\begin{aligned}&\sigma_{\binom{m+d-1}{d}}\left(\Psi\left(\tilde{U}^{\otimes d}\right)\mathrm{Perm}_{\mathrm{avg}}\right) \\ &\geq \frac{1}{\sqrt{d!}}\sigma_{m^d}\left(\Psi\left(\otimes_{j=1}^{d}(\tilde{V}^{(j-1)} + jZ_j)\right)\right) - \|\Psi\|\|E\|,\end{aligned}$$

from which the desired singular value lower bound follows. □

C.4 Proof of Corollary 5.3

Corollary 5.3. *Suppose $d, t \in \mathbb{N}$ and let $1 \geq \delta_1 > \delta_2 > 0$ be given. Also let $\Phi : \mathsf{Sym}^d(\mathbb{R}^n) \to \mathbb{R}^D$ be an orthogonal projection of rank $R \geq \delta_1\binom{n+d-1}{d}$. Let $\{U_j\}_{j=1}^{t} \subset \mathbb{R}^{n\times m}$ be an arbitrary collection of $n \times m$ matrices, and for each j, let $\tilde{U}_j$ be a random ρ-perturbation of U_j. Then there exists a constant $c_d > 0$ such that if $t\binom{m+d-1}{d} \leq \delta_2\binom{n+d-1}{d}$ and $m \leq c_d(\delta_1 - \delta_2)n$, then with probability at least $1 - \exp\big(-\Omega_{d,\delta_1,\delta_2}(n)\big)$, we have the least singular value*

$$\sigma_{t\binom{m+d-1}{d}}\left(\Phi\begin{bmatrix}\tilde{U}_1^{\circledast d} & \tilde{U}_2^{\circledast d} & \ldots & \tilde{U}_t^{\circledast d}\end{bmatrix}\right) \geq \frac{\rho^d}{\sqrt{t}n^{O(d)}}. \qquad (5)$$

Proof. For each $j = 1, \ldots, t$, let $\Pi_{-j}^{\perp}$ denote the orthogonal projection onto

$$\mathrm{Ran}(\Phi) \cap \mathrm{Ran}\left(\begin{bmatrix}\tilde{U}_1^{\circledast d} & \ldots & \tilde{U}_{j-1}^{\circledast d} & \tilde{U}_{j+1}^{\circledast d} & \ldots & \tilde{U}_t^{\circledast d}\end{bmatrix}\right)^{\perp}$$

We first lower bound the least singular value of $\Pi_{-j}(\tilde{U}_j^{\circledast d})$. Observe that from our assumptions, the rank of Π_{-j} is at least

$$\delta_1\binom{n+d-1}{d} + (1-\delta_2)\binom{n+d-1}{d} - \binom{n+d-1}{d} = (\delta_1 - \delta_2)\binom{n+d-1}{d}.$$

Taking c_d to be the constant from Theorem 5.1, we can then apply Theorem 5.1 to conclude that with probability at least $1 - \exp(-\Omega_{d,\delta_1,\delta_2}(n))$ we have

$$\sigma_{\binom{m+d-1}{d}}(\Pi_{-j}\tilde{U}_j^{\otimes d}) \geq \frac{\rho^d}{n^{O(d)}}.$$

The result now follows by applying the block leave one out bound of Lemma A.2. □

D APPLICATION: CERTIFYING QUANTUM ENTANGLEMENT AND LINEAR SECTIONS OF VARIETIES

We consider the setting in the work of Johnston, Lovitz and Vijayaraghavan [19], where we are given a conic algebraic variety $\mathcal{X}$ and a linear subspace $\mathcal{U}$. The subspace $\mathcal{U}$ is specified by a basis while the variety $\mathcal{X}$ is specified by a set of polynomials that cut it out. Conic varieties (or equivalently, projective varieties) are closed under scalar multiplication and are cut out by homogeneous polynomials, which can further be chosen to all have the same degree d. In [19], they give a polynomial time algorithm that certifies that the intersection $\mathcal{U} \cap \mathcal{X}$ is trivial i.e., $\mathcal{U} \cap \mathcal{X} = \{0\}$ for a *generic* subspace $\mathcal{U}$ up to a certain dimension m. In this section, we prove a robust analogue of this statement in the smoothed analysis setting.

In the smoothed setting, the subspace $\widetilde{\mathcal{U}}$ is spanned by ρ-perturbed vectors $\tilde{u}_1, \ldots, \tilde{u}_m$, with $\forall i \in [m],\ \tilde{u}_i = u_i + z_i$ where $\|u_i\| \leq 1$ and $z_i \sim_{i.i.d} N(0, \rho^2)^n$. The subspace $\widetilde{\mathcal{U}}$ is specified in terms of any orthonormal basis $\hat{u}_1, \ldots, \hat{u}_m$. The variety is specified by a set of degree-d homogenous polynomials that cut out the variety $\mathcal{X}$. Our goal is to certify that every element of $v \in \mathcal{X}$ (with $\|v\| = 1$) is far from the subspace $\widetilde{\mathcal{U}}$.

Theorem D.1. *Let $\mathcal{X} \subseteq \mathbb{R}^n$ be an irreducible variety cut out by $p = \delta\binom{n+d-1}{d}$ linearly independent homogeneous degree-d polynomials $f_1, \ldots, f_p \in \mathbb{R}[x_1, \ldots, x_n]_d$, for constants $d \geq 2$ and $\delta \in (0, 1)$. There exists a constant $c_d > 0$ (that only depends on d) such that for a randomly ρ-perturbed subspace $\widetilde{\mathcal{U}} \subseteq \mathbb{R}^n$ of dimension $m \leq c_d \cdot \delta n$ as described above, we have that with probability $1 - \exp(-\Omega(n))$, the algorithm in Figure 6 on input $f_1, \ldots, f_p$ and a basis $u_1, \ldots, u_m$ for $\widetilde{\mathcal{U}}$ certifies in polynomial time (i.e., $(n/\rho)^{O(d)}$) that*

$$\forall v \in \mathcal{X} \text{ with } \|v\| = 1,\ \mathit{dist}(v, \widetilde{\mathcal{U}}) := \inf_{u \in \widetilde{\mathcal{U}}}\|u - v\|_2 \geq \frac{\rho^d}{n^{O(d)}}. \qquad (22)$$

This theorem gives a robust analog of the genericity statement in [19] that certifies trivial intersection with a generic subspace (Theorem 2 with $s = 0$). We remark that [19] also considers a version of the problem where there are also some generic elements of $\mathcal{X}$ planted in the subspace $\widetilde{\mathcal{U}}$. While it is natural to think of generic elements of $\mathcal{X}$ (using the induced Zariski topology on $\mathcal{X}$), it is not clear how to define an appropriate smoothed analysis model that captures the planted setting. This is an interesting research direction that is beyond the scope of this work.

We use the algorithm of Johnston, Lovitz and Vijayaraghavan [19], which is based on Hilbert's projective Nullstellensatz certificates. Recall that $\mathcal{X}$ is a conic variety that is cut out by a finite set of homogeneous degree-d polynomials $f_1, \ldots, f_p \in \mathbb{R}[x_1, \ldots, x_n]_d$. Viewing

$f_1, \dots, f_p \in (\mathbb{R}^n)^{\otimes d}$ as vectors in the dual space, we consider the map

$$\Phi^d_{\mathcal{X}} : (\mathbb{R}^n)^{\otimes d} \to \mathbb{R}^p \tag{23}$$

$$v \mapsto (f_1(\mathsf{Sym}_d v), \dots, f_p(\mathsf{Sym}_d v))^\top. \tag{24}$$

Note that for rank-1 v, i.e. $v = x^{\otimes d}$ for some $x \in \mathbb{R}^n$, $\Phi^d_{\mathcal{X}} = \mathbf{0}$ exactly when x lies in the variety $\mathcal{X}$. By picking an orthonormal basis for the vector space spanned by the dual vectors $f_1, \dots, f_p$, we can assume without loss of generality that the operator $\Phi^d_{\mathcal{X}}$ is an orthogonal projection matrix with rank p. Moreover we can assume that the given basis $u_1, \dots, u_m$ for $\widetilde{\mathcal{U}}$ is an orthonormal basis.

Input: A basis $\{\hat{u}_1, \dots, \hat{u}_m\}$ for a linear subspace $\widetilde{\mathcal{U}} \subseteq \mathbb{R}^n$, and a collection of homogeneous degree-d polynomials $f_1, \dots, f_p$ that cut out a conic variety $\mathcal{X} \subseteq \mathbb{R}^n$.

(1) Compute the least singular value of the following matrix

$$\eta = \sigma_{\binom{m+d-1}{d}}\Big(\Phi^d_{\mathcal{X}}(\hat{u}_{i_1} \otimes \cdots \otimes \hat{u}_{i_d}) : 1 \le i_1 \le \cdots \le i_d \le m\Big). \tag{25}$$

(2) If $\eta > 0$, output: "$\widetilde{\mathcal{U}}$ is at least η-far from $\mathcal{X}$."

(3) Otherwise, output: "Don't know."

Figure 6: Algorithm certifying that $\widetilde{\mathcal{U}}$ is far from $\mathcal{X}$.

Proof. From standard random matrix theory (or see e.g., Claim C.1) with probability $1 - \exp(-\Omega(n))$ we have for some constant $c_1 > 0$, $\sigma_m(\tilde{U}) \ge c\rho/(n^{c_1})$ and $\sigma_1(\tilde{U}) \le O(\sqrt{n}(1+\rho))$. Conditioned on the above event, every unit vector $u \in \widetilde{\mathcal{U}}$ can be expressed as

$$u = \sum_{i=1}^{m} \alpha_i \tilde{u}_i, \text{ for some } \alpha \in \mathbb{R}^m$$

$$\text{where } \frac{\Omega(1+\rho)}{\sqrt{n}} \le \|\alpha\|_2 \le O\big(n^{c_1}/\rho\big). \tag{26}$$

Moreover, from Theorem 5.1, with probability at least $1 - \exp(-\Omega(n))$ we have for some $c_d > 0$ (that is constant for constant d) that

$$\sigma_{\binom{m+d-1}{d}}\Big(\Phi^d_{\mathcal{X}}\ \widetilde{U}^{\otimes d}\Big) \ge \frac{c_d \rho^d}{n^{O(d)}}. \tag{27}$$

We condition on all the above high probability events (about the least singular values) that hold with probability $1 - \exp(-\Omega(n))$.

Consider any vector $v \in \mathcal{X} \subset \mathbb{R}^n$ such that $\|v\| = 1$. Let $u^* \in \widetilde{\mathcal{U}}$ be the closest vector to v in the subspace $\widetilde{\mathcal{U}}$. On the one hand, from (26) applied with u^*

$$\Phi^d_{\mathcal{X}}((u^*)^{\otimes d}) = \sum_{i_1, i_2 \dots, i_d = 1}^{m} (\alpha_{i_1} \dots \alpha_{i_d}) \cdot \Phi^d_{\mathcal{X}}\big(\tilde{u}_{i_1} \otimes \cdots \otimes \tilde{u}_{i_d}\big)$$

$$= \sum_{1 \le i_1 \le i_2 \le \dots, i_d \le m} \beta_{i_1, i_2, \dots, i_d} \cdot \Phi^d_{\mathcal{X}}\big(\tilde{u}_{i_1} \otimes \cdots \otimes \tilde{u}_{i_d}\big),$$

where $\beta_{i_1, \dots, i_d} = c_{i_1, \dots, i_d} \alpha_{i_1} \dots \alpha_{i_d}$ and $c_{i_1, \dots, i_d}$ is a constant coefficient in $[1, d!]$.[10] Also $1 \le \|\beta\| \le \sqrt{d!} \cdot \|\alpha\|_2$. Hence, from (27)

$$\Big\|\Phi^d_{\mathcal{X}}((u^*)^{\otimes d})\Big\|_2 = \Big\| \sum_{1 \le i_1 \le i_2 \le \dots, i_d \le m} \beta_{i_1, i_2, \dots, i_d} \Phi^d_{\mathcal{X}}\big(\tilde{u}_{i_1} \otimes \cdots \otimes \tilde{u}_{i_d}\big)\Big\|_2$$

$$\ge \frac{c_d \rho^d}{n^{O(d)}} \cdot \|\beta\|_2 \ge \frac{c_d \rho^d}{n^{O(d)}}.$$

On the other hand, we have $\Phi^d_{\mathcal{X}}(v^{\otimes d}) = 0$ since $v \in \mathcal{X}$. As $\|\Phi^d_{\mathcal{X}}\| \le 1$ by assumption, we get the following sequence of inequalities

$$\|v - u^*\| \ge \frac{1}{d}\Big\|v^{\otimes d} - (u^*)^{\otimes d}\Big\| \ge \frac{1}{d}\Big\|\Phi^d_{\mathcal{X}} v^{\otimes d} - \Phi^d_{\mathcal{X}}(u^*)^{\otimes d}\Big\| \ge \frac{c_d \rho^d}{n^{O(d)}}.$$

This proves the theorem. □

D.1 Certifying quantum entanglement

We can instantiate the above theorem with specific choices of the variety $\mathcal{X}$ to get algorithms that certify that a smoothed subspace is robustly entangled i.e., it is far from any non-entangled state, for different notions of entanglement. In what follows, we restrict to the real domain for all of our statements and proofs. However, quantum states are defined over the complex domain, and we do not handle the complex domain in this version of the paper.

We start with the bipartite setting of dimension $n_1 \times n_2$, which is captured by matrices $\mathbb{R}^{n_1 \times n_2}$. The set of *separable* states is captured by the variety of rank-1 matrices:

$$\mathcal{X}_1 = \{M \in \mathbb{R}^{n_1 \times n_2} : \mathsf{rank}(M) \le 1\}. \tag{28}$$

A pure state v that is non-separable i.e., $v \notin \mathcal{X}_1$ is said to be *entangled*. For any $\varepsilon > 0$ we say that a subspace $\mathcal{U} \subset \mathbb{R}^{n_1 \times n_2}$ is said to be ε-robustly entangled if every unit vector in $\mathcal{X}_1$ is at least ε far from the subspace $\mathcal{U}$ i.e.,

$$(\varepsilon\text{-robust entanglement})\ \ \mathrm{dist}(\mathcal{U}, \mathcal{X}_1) = \inf_{\substack{v \in \mathcal{X}_1 : \|v\|_2 = 1, \\ u \in \mathcal{U}}} \|v - u\|_2 \ge \varepsilon. \tag{29}$$

More generally, the determinantal variety $\mathcal{X}_r$ corresponds to states that have *Schmidt rank* at most r where

$$\mathcal{X}_r = \{M \in \mathbb{R}^{n_1 \times n_2} : \mathsf{rank}(M) \le r\}.$$

$$(\varepsilon\text{-robust } r\text{-entanglement})\ \ \mathrm{dist}(\mathcal{U}, \mathcal{X}_r) = \inf_{\substack{v \in \mathcal{X}_r : \|v\|_2 = 1, \\ u \in \mathcal{U}}} \|v - u\|_2 \ge \varepsilon. \tag{30}$$

captures the ε-robustly r-entanglement of a subspace, for an $\varepsilon > 0$. Such subspaces are only close to highly entangled states. Entangled subspaces corresponds to the setting when $r = 1$. The variety $\mathcal{X}_r$ is cut out by $p = \binom{n_1}{r+1}\binom{n_2}{r+1}$ linearly independent homogenous polynomials of degree $r + 1$;[11] see [19]. The following corollary then follows immediately from Theorem D.1.

Corollary D.2. *Let n_1, n_2 be positive integers, let $r < \min\{n_1, n_2\}$ be a positive integer. There exists constants $c_r, c'_r > 0$ (that only depends on r), an absolute constant $c > 0$ and an algorithm that given a randomly ρ-perturbed subspace $\widetilde{\mathcal{U}} \subseteq \mathbb{R}^{n_1 n_2}$ of dimension $m \le c_r \cdot n_1 n_2$, runs in $(n_1 n_2/\rho)^{O(r)}$ time and certifies with probability*

[10] More precisely, $c_{i_1, \dots, i_d}$ is the number of entries in the tensor/dual form that correspond to the relevant monomial in the polynomial. Thus, $c_{i_1, \dots, i_d} = t_1! t_2! \dots t_\ell!$ where ℓ is the number of distinct indices among $i_1, \dots, i_d$, and $t_1, \dots, t_\ell$ are the frequencies of these ℓ distinct indices.

[11] corresponding to all the determinants of the $(r+1) \times (r+1)$ submatrices being 0

at least $1 - \exp(-\Omega(n_1 n_2))$ *that* $\widetilde{\mathcal{U}}$ *is* η*-robustly* r*-entangled for* $\eta = c'_r \rho^r / (n_1 n_2)^{cr}$ *i.e.,*

$$dist(\widetilde{\mathcal{U}}, \mathcal{X}_r) \geq \frac{c'_r \rho^{r+1}}{(n_1 n_2)^{cr}}. \tag{31}$$

Note that when $r = 1$, we get Corollary 1.6 about robustly certifying entanglement of smoothed subspaces. This is the robust generalization of Corollary 28 in [19] which certifies that $\widetilde{\mathcal{U}} \cap \mathcal{X}_r = \{0\}$ for a generic subspace $\widetilde{\mathcal{U}}$ (this can be seen as a special case when $\eta \to 0$). Our result gives a way of certifying a lower bound on the distance of a subspace from the set of rank-1 matrices for smoothed subspaces. This certification problem is closely related to the best separable state problem [17] which also relates this question to several other important questions in quantum information theory and polynomial optimization.

Comparison to Barak, Kothari and Steurer [8]. Barak, Kothari and Steurer [8] give an algorithm for an ε-promise version of the entanglement certification problem, where given an arbitrary subspace $\mathcal{U} \in \mathbb{R}^{n \times n}$, the goal is distinguish between

- **YES:** There is a unit vector $v \in \mathcal{X}_r$ that also lies in $\mathcal{U}$.
- **NO:** For all unit vectors $v \in \mathcal{X}_r$, $\|v - u\| \geq \varepsilon$ for all $u \in \mathcal{U}$.

The algorithm in [8] gives a $2^{O(\sqrt{n})/\varepsilon}$ time algorithm to solve the above promise problem. Harrow and Montanaro [17] presented evidence through conditional hardness results that polynomial time algorithms may not exist even for constant $\varepsilon > 0$ in the worst-case, i.e., for arbitrary subspaces. In contrast, our algorithm gives polynomial time algorithms for smoothed subspaces up to dimension $c \cdot n^2$ (for some constant $c > 0$) even for ε that is inverse polynomially small.

We can also use Theorem D.1 with other choices of $\mathcal{X}$ to get robust analogs of the other certification results in [19]. These include multi-partite entanglement notions like complete entanglement and genuine entanglement which have been studied in quantum information. They follow by applying Theorem D.1 along with the corresponding claims in [19]. We state one such result for complete entanglement. Denote the set of separable order d-tensors

$$\mathcal{X}_{sep} = \{v_1 \otimes v_2 \otimes \cdots \otimes v_d : v_1 \in \mathbb{R}^{n_1}, \ldots v_d \in \mathbb{R}^{n_d}\}. \tag{32}$$

A ε-robust completely entangled subspace $\mathcal{U}$ is one which is ε-far from every unit vector in $\mathcal{X}_{sep}$. Again by using the fact that $\mathcal{X}_{sep}$ is cut out by $p = \binom{n_1 n_2 \ldots n_d + 1}{2} - \binom{n_1+1}{2} \cdot \cdots \cdot \binom{n_d+1}{2}$ linearly independent homogenous polynomials of degree 2 (see Section 2.3 of [19]) we get the following corollary.

Corollary D.3. *Let* $n_1, n_2, \ldots, n_d$ *be positive integers. There exists constants* $c_d, c'_d > 0$ *(that only depends on d), an absolute constant* $c > 0$ *and an algorithm that given a randomly* ρ*-perturbed subspace* $\widetilde{\mathcal{U}} \subseteq \mathbb{R}^{n_1 n_2 \ldots n_d}$ *of dimension* $m \leq c_d \cdot n_1 n_2 \ldots n_d$*, runs in* $\mathrm{poly}(n_1 n_2 \ldots n_d / \rho)$ *time and certifies with probability at least* $1 - \exp(-\Omega(n_1 \ldots n_d))$ *that* $\widetilde{\mathcal{U}}$ *is* η*-robustly completely entangled for* $\eta = c'_d \rho^2 / (n_1 n_2 \cdot n_d)^c$ *i.e.,*

$$dist(\widetilde{\mathcal{U}}, \mathcal{X}_r) \geq \frac{c'_d \rho^2}{(n_1 n_2 \cdot n_d)^c}. \tag{33}$$

E APPLICATION: DECOMPOSING OF SUMS OF POWERS OF POLYNOMIALS

E.1 Overview of Bafna, Hsieh, Kothari and Xu [7]

One application of our techniques is to extend the results of Bafna, Hsieh, Kothari, and Xu [7] (which in turn build on [12]) to the smoothed analysis setting. In [7] they consider the problem of *decomposing power-sum polynomials.* In the most fundamental setting they consider n-variate polynomials of the form

$$\widehat{p}(\mathbf{x}) = \sum_{t \leq m} a_t(\mathbf{x})^3 + e(\mathbf{x}),$$

where $\mathbf{x} = [x_1, \ldots, x_n]$, each $a_t(\mathbf{x})$ is a homogeneous quadratic polynomial, and $e(\mathbf{x})$ is a polynomial of small norm.[12] The goal is to recover the underlying $a_i(\mathbf{x})$s from $\widehat{p}(\mathbf{x})$. [7] give an algorithm that is able to recover these components when there are up to $m \sim \widetilde{O}(n)$ components, while withstanding noise of inverse polynomial magnitude, when each of the components $a_i(\mathbf{x})$ is drawn *randomly* from a mean-0 distribution.

A natural setting for this problem is one in which each $a_i(\mathbf{x})$ is *perturbed*, rather than fully random. In this setting, [7] are only able to show that their algorithm can withstand noise of inverse exponential magnitude. Our contribution is to provide a new analysis of the random matrices that arise in their algorithm, and thus give a smoothed analysis guarantee for the algorithm of [7] that is robust to noise of inverse polynomial magnitude.

In this section, we provide an overview of their algorithm and highlight the three main matrices that arise. First, we rephrase the question about polynomials as a question about tensors. That is, we identify each homogeneous quadratic polynomial $a_t(\mathbf{x})$ with a symmetric coefficient matrix A_t such that $a_t(\mathbf{x}) = \mathbf{x}^\top A_t \mathbf{x}$. Then, one way to represent $p(\mathbf{x})$ is as a symmetric order-6 coefficient tensor P_{sym} defined by

$$\mathrm{vec}(P_{\text{sym}}) = \mathrm{Sym}_6\Big(\mathrm{vec}\Big(\sum_{t \leq m} A_t^{\otimes 3}\Big)\Big),$$

where Sym_6 is a linear operator that performs a 6-way symmetrization operation.[13] Since we are given the input in the form of a polynomial, we only have access to this symmetrized form of the tensor.

[12] [7] also consider a more general setting where $\widehat{p}(\mathbf{x}) = \sum_{t \leq m} a_t(\mathbf{x})^{3D} + e(\mathbf{x})$, where each $a_t(\mathbf{x})$ is a homogeneous polynomial of degree K, that is then taken to the $3D$th power for some integer $D \geq 1$. That is, the basic case we present above is their result for $K = 2, D = 1$. We focus on this case because the techniques for this setting already form the basis of their algorithms for the other settings, and our goal is to give a proof-of-concept for the kind of applications that we expect for our techniques.

[13] In terms of polynomials, Sym_6 essentially combines terms that are the same due to commutativity, i.e. $x_1x_2x_3x_4x_5x_6 = x_2x_3x_4x_5x_6x_1$. Note that, even though the underlying A_ts are (2-way) symmetric matrices, and each $A_t^{\otimes 3}$ is taking a 3-way symmetric product, $A_t^{\otimes 3}$ is *not* necessarily 6-way symmetric. One way to think of this is that the 2-way symmetry of the A_ts allows the variables of $x_1x_2 \cdot x_3x_4 \cdot x_5x_6$ to commute with their pairwise partners, and the 3-way symmetry of the tensor product allows the pairs to commute with each other. However, this does not allow x_2 and x_3 to individually commute with each other, for example. Thus, the Sym_6 operator can change the structure of P significantly.

The main observation that the [7] algorithm is built on is that, if we could recover the asymmetric version of the tensor

$$P_{\text{asym}} = \sum_{t \le m} A_t^{\otimes 3},$$

then we could use standard tensor decomposition techniques to recover the underlying A_ts. In general, recovering an asymmetric tensor from a symmetric tensor is impossible, as there is a whole subspace of asymmetric tensors that could be mapped to the same symmetric tensor. However, if the $A_t^{\otimes 3}$ all belonged to a known generic low-dimensional subspace, then the symmetrization operator would actually be invertible over this subspace.

In [7] they focus on recovering a basis for the subspace spanned by the (vectorized) A_ts. If we recover such a basis, represented by the columns of a matrix $C = [C_1, \dots, C_m]$, then we have that (the vectorized form of) P_{asym} lives in the column span of $C^{\circledast 3}$, where $\circledast$ denotes the symmetrized Kronecker product. We can then invert the symmetrization operator and retrieve P_{asym}, as long as the following claim holds.

Proposition E.1 (Symmetrization invertible over Kronecker basis)**.** *Given $C \in \mathbb{R}^{n^2 \times m}$ representing a basis for a subspace of ρ-perturbed symmetric matrices, with $m \le cn^2$ for some absolute constant $c \in (0,1)$, $\mathrm{Sym}_6\, C^{\circledast 3}$ is* robustly *invertible w.h.p. That is, with probability at least $1 - \exp(-\Omega(n))$,*

$$\sigma_{\min}\left(\mathrm{Sym}_6\, C^{\circledast 3}\right) \ge \mathrm{poly}\left(\rho, \frac{1}{mn}\right).$$

Proposition E.1 follows from Theorem 5.1. First, we can consider $C = QA$ where Q is a basis change matrix (which is well-conditioned w.h.p), and A is the matrix of vectorized A_ts. Then, we are interested in $\mathrm{Sym}_6\, C^{\circledast 3} = \mathrm{Sym}_6\, Q^{\circledast 3} A^{\circledast 3}$. The rank of Sym_6 is $\binom{n+5}{6}$ is only a constant factor less than the dimension of the total dimension of the tensored space $(\binom{n+1}{2})^3$. Thus this statement follows from Theorem 5.1, considering the linear operator $\mathrm{Sym}_6\, R^{\otimes 3}$ applied to symmetric lifts of the ρ-perturbed A_ts. .

To recover the subspace of the A_ts, [7] use the partial derivatives of $p(\mathbf{x})$. Specifically, the form of the second partial derivatives are

$$\frac{\partial^2}{\partial x_i \partial x_j} p(\mathbf{x}) = \frac{\partial^2}{\partial x_i \partial x_j} \sum_{t \le m} a_t(\mathbf{x})^3 = \sum_{t \le m} a_t(\mathbf{x}) q_{t,i,j}(\mathbf{x}),$$

for some homogeneous quadratic polynomials $q_{t,i,j}(\mathbf{x})$. Now, we have a space $\mathcal{U}$ spanned by these polynomial combinations of the $a_t(\mathbf{x})$, and we would like to recover the underlying quadratic $a_t(\mathbf{x})$s from this space.

$\mathcal{U}$ is related to the space $\mathcal{V}$, which spans *all* quartic multiples of the $a_t(\mathbf{x})$. That is,

$$\mathcal{V} = \{a_t(\mathbf{x}) q(\mathbf{x}) \ : \ t \le m \text{ and } q(\mathbf{x}) \text{ is a quadratic polynomial}\}.$$

Given $\mathcal{V}$, [7] show a way to recover the span of the underlying $a_t(\mathbf{x})$. So first, we would like to find $\mathcal{V}$ starting from $\mathcal{U}$. By definition, we have that $\mathcal{U} \subseteq \mathcal{V}$. However, it is *not* true that $\mathcal{U} = \mathcal{V}$. This can be observed by dimension-counting: there are only $\binom{n+1}{2}$ partial derivatives, which cannot span the space of multiples which has dimension $m\binom{n+1}{2}$ for generic $a_t(\mathbf{x})$.

The [7] algorithm gets around this by projecting $\mathcal{U}$ to a smaller dimensional space. In particular, they show that by projecting $\mathcal{U}$ onto $\ell \in O(\sqrt{n})$ variables, they recover the span of all quartic multiples of the $a_t(\mathbf{x})$ restricted to these ℓ variables, i.e. the projected $\mathcal{V}$. We already know that the projected $\mathcal{U}$ must be contained in the projected $\mathcal{V}$. Thus to show equality, we only need to show that the rank of the projected $\mathcal{U}$ matches the rank of the projected $\mathcal{V}$. [7] do this by showing that a certain system of equations that recovers an element of the projected $\mathcal{V}$ from a corresponding element of the projected $\mathcal{U}$ is solvable. This ends up boiling down to the following proposition.

Proposition E.2 (Projected $\mathcal{U}$ same dimension as projected $\mathcal{V}$)**.** *Fix parameters $m, n, \ell \in \mathbb{N}$ where and let $\tilde{U}_1, \dots, \tilde{U}_m$ be a collection of ρ-smoothed $n \times n$ matrices which satisfy $\max_t \|U_t\|_F \le L$. Let $M \in \mathbb{R}^{n \times \ell}$ be a column selection matrix and set $\tilde{S}_t = \tilde{U}_t M$ for each $t = 1, \dots, m$. Also let $V \in \mathbb{R}^{n^2 \times m\binom{\ell+1}{2}+m}$ be the block matrix*

$$V := \left[\tilde{S}_1 \circledast \tilde{S}_1 \quad \dots \quad \tilde{S}_m \circledast \tilde{S}_m \quad \mathrm{vec}(\tilde{U}_1) \quad \dots \quad \mathrm{vec}(\tilde{U}_m)\right].$$

Finally, assume the parameter $r := n^2 - n\ell - m\binom{\ell+1}{2} - m + 1$ satisfies $r \ge \delta n^2$ for some $\delta \in (0,1)$. Then there exists constants $c, c' > 0$ (potentially depending on δ and L) such that with probability at least $1 - \exp(-\Omega_\delta(n))$ we have

$$\sigma_{m\binom{\ell+1}{2}+m}(V) \ge \frac{c' \rho^4}{n^c}.$$

Now, we shift our view to the case where we have $\mathcal{V}$, the space of all quartic multiples of $a_t(\mathbf{x})$.[14] [7] consider an equation of the form

$$\sum_{t \le m} a_t(\mathbf{x}) q_t(\mathbf{x}) = 0, \tag{34}$$

where the $a_t(\mathbf{x})$ are generic, and the $q_t(\mathbf{x})$ are variables. They note that since generic polynomials are *irreducible*, the only solutions to this system have a certain structure. In particular, the space of solutions are spanned by solutions of the form

$$\begin{aligned} \text{for } i \ne j: \quad & q_i(\mathbf{x}) = a_j(\mathbf{x}), \\ & q_j(\mathbf{x}) = -a_i(\mathbf{x}), \\ & q_k(\mathbf{x}) = 0 \qquad\qquad \forall k \ne i, j. \end{aligned} \tag{35}$$

and the dimension of the solution space is $\binom{m}{2}$. Note that solutions of this form always exist, even when the $a_t(\mathbf{x})$ are not irreducible. A key observation of [7] is that when $a_t(\mathbf{x})$ are irreducible, these are the *only* solutions. Now consider

$$\sum_{t \le m} a_t(\mathbf{x}) q_t(\mathbf{x}) = a_0(\mathbf{x}) q_0(\mathbf{x}), \tag{36}$$

where $a_0(\mathbf{x})$ is a fresh generic polynomial chosen by our algorithm. The space of polynomials that have the form of the LHS of (36) is precisely $\mathcal{V}$. Once we choose an $a_0(\mathbf{x})$, the algorithm can also construct the space of all quartic multiples of $a_0(\mathbf{x})$, which we denote $\mathcal{V}_0$. Now, consider all solutions to (36) where $q_0(\mathbf{x})$ is nonzero.

[14] In reality, we have access to the space of quartic multiples of the projected $a_t(\mathbf{x})$. However, we can repeat this step for various choices of the projection, and recover the $a_t(\mathbf{x})$ restricted to different coordinates. For simplicity of notation, we will still refer to the dimension of the (projected) polynomial as n.

By the characterization of solutions in (35), we know that if it is nonzero, $q_0(\mathbf{x})$ must live in the span of the $a_t(\mathbf{x})$. Thus, we have that

$$(\mathcal{V}\cap\mathcal{V}_0)/a_0(\mathbf{x}) = \mathrm{span}\{a_t(\mathbf{x}) : t \le m\}.$$

To go from a generic guarantee to a smoothed guarantee, we must make the characterization (35) of the solution space of this equation robust. Thus, rather than only requiring the $a_t(\mathbf{x})$ to be irreducible, we require the stronger condition that they are perturbed (smoothed). To formalize this, once again we represent our polynomials as symmetric tensors. Denote the coefficient matrix of $a_t(\mathbf{x})$ be A_t, and the coefficient matrix of $q_t(\mathbf{x})$ be Q_t. Then, we can write (34) as

$$\mathrm{Sym}_4\ \mathrm{vec}\left(\sum_{t\le m} A_t \otimes Q_t\right) = 0.$$

If we combine our Q_ts into one large vector, with the appropriate rearranging of indices[15], we can write this as

$$\mathrm{Sym}_4\left(I_{\binom{n+1}{2}} \otimes A\right)\mathbf{q} = 0,$$

where A is the matrix such that column i is (the vectorized) A_t, and $\mathbf{q}$ is the appropriate vectorization of the Q_ts. Recall that we want to show that the dimension of the space of solutions $\mathbf{q}$ is exactly $\binom{m}{2}$. Thus we can write the robust condition as following. In what follows $N_2 := \binom{n+1}{2}$ is the dimension of the space of homogenous quadratic polynomials over n variables.

Proposition E.3 (Not too many solutions to polynomial equation system)**.** *The space of solutions to the above polynomial equation system has rank at most $\binom{m}{2}$ in a robust sense. That is, let $A \in \mathbb{R}^{N_2\times m}$ be made up of perturbed symmetric (when viewed as matrices) columns. Then there exists constants $c, c', c'' > 0$ such that when $m \le cN_2$, with probability at least* $1-\exp(-\Omega(n))$

$$\sigma_{mN_2-\binom{m}{2}}\left(\mathrm{Sym}_4\ \left(I_{N_2\times N_2}\otimes A\right)\right) \ge \frac{c''\min\{\rho,1\}^2}{n^{c'}},$$

where $\otimes$ denotes the Kronecker product and $N_2 = \binom{n+1}{2}$.

In [7] they prove analogues of propositions E.1, E.2, and E.3, for the case where the underlying matrices are fully random (mean-0), and they are able to show that these propositions hold with inverse-polynomial failure probability. Our framework allows us to prove these propositions for perturbed matrices, and show that the failure probability is *exponentially* small. These improvements allow us to conclude that the algorithm of [7] provides a smoothed analysis guarantee. We omit the analysis for general values of K, D in this version of the paper.

E.2 Least singular value bounds for Proposition E.2

Proposition E.2 (Projected $\mathcal{U}$ same dimension as projected $\mathcal{V}$)**.** *Fix parameters $m, n, \ell \in \mathbb{N}$ where and let $\tilde{U}_1, \dots, \tilde{U}_m$ be a collection of ρ-smoothed $n \times n$ matrices which satisfy* $\max_t \|U_t\|_F \le L$*. Let $M \in \mathbb{R}^{n\times\ell}$ be a column selection matrix and set $\tilde{S}_t = \tilde{U}_t M$ for each $t = 1, \dots, m$. Also let $V \in \mathbb{R}^{n^2\times m\binom{\ell+1}{2}+m}$ be the block matrix*

$$V := \begin{bmatrix}\tilde{S}_1 \circledast \tilde{S}_1 & \dots & \tilde{S}_m \circledast \tilde{S}_m & \mathrm{vec}(\tilde{U}_1) & \dots & \mathrm{vec}(\tilde{U}_m)\end{bmatrix}.$$

Finally, assume the parameter $r := n^2 - n\ell - m\binom{\ell+1}{2} - m + 1$ satisfies $r \ge \delta n^2$ for some $\delta \in (0,1)$. Then there exists constants $c, c' > 0$ (potentially depending on δ and L) such that with probability at least $1 - \exp(-\Omega_\delta(n))$ we have

$$\sigma_{m\binom{\ell+1}{2}+m}(V) \ge \frac{c'\rho^4}{n^c}.$$

Proof. Without loss of generality assume that $\tilde{S}_t$ is the first ℓ columns of $\tilde{U}_t$. Also, we have by standard concentration bounds, $\|\tilde{U}_t\|_F \le L + \rho\cdot\mathrm{poly}(n)$ with probability at least $1-\exp(-\Omega(n))$. We condition on the success event for the rest of the proof, and assume without loss of generality that the upper bound L already includes the additive term $\rho\mathrm{poly}(n)$.

We first argue that each column of the form $\mathrm{vec}(\tilde{U}_t)$ has a large component that is orthogonal to the remaining columns of V. To this end, set

$$V_t := \begin{bmatrix} S & \tilde{U}_{(-t)}\end{bmatrix}$$

where

$$S = \begin{bmatrix}\tilde{S}_1 \circledast \tilde{S}_1 & \dots & \tilde{S}_m \circledast \tilde{S}_m\end{bmatrix}$$

and

$$\tilde{U}_{(-t)} = \begin{bmatrix}\mathrm{vec}(\tilde{U}_1) & \dots & \mathrm{vec}(\tilde{U}_{t-1}) & \mathrm{vec}(\tilde{U}_{t+1}) & \dots & \mathrm{vec}(\tilde{U}_m)\end{bmatrix}.$$

We can restrict to considering only the last $n^2 - n\ell$ entries of each column of this matrix. By assumption, each $\tilde{S}_t$ is made up the first ℓ columns of $\tilde{U}_t$, so the smoothing in the last $n^2-n\ell$ entries of $\mathrm{vec}(\tilde{U}_t)$ is independent from the smoothing in the last $n^2 - n\ell$ entries of the columns of V_t.

Now, let $Q : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2-n\ell}$ be the matrix that restricts onto the last $n^2 - n\ell$ entries of a vector and let $\Pi^{\perp}_{RV_t}$ be the projection onto the orthogonal complement of the range of QV_t. Note that the matrix $QV_t \in \mathbb{R}^{n^2-n\ell\times m\binom{\ell+1}{2}+m-1}$ has rank at most $m\binom{\ell+1}{2}+m-1$, so using Lemma A.1 we have

$$\mathbb{P}[\|\Pi^{\perp}_{QV_t} Q\mathrm{vec}(\tilde{U}_t)\| \ge \varepsilon] \ge 1 - \left(\frac{c''\varepsilon}{\rho}\right)^r,$$

where $r = n^2 - n\ell - m\binom{\ell+1}{2} - m + 1$ and $c'' > 0$ is an absolute constant. The probability that the above holds for all $t = 1, \dots, m$ is then at least

$$1 - m\left(\frac{c''\varepsilon}{\rho}\right)^r.$$

Now suppose that $\sigma_{\min}(V) < \varepsilon$. Then there must exist some test unit vector $\alpha \in \mathbb{R}^{m\binom{\ell+1}{2}+m}$ such that $\|V\alpha\| < \varepsilon$. Write $\alpha = \alpha^{(1)} \oplus \alpha^{(2)}$ where $\alpha^1 \in \mathbb{R}^{m\binom{\ell+1}{2}}$ and $\alpha^2 \in \mathbb{R}^m$. Intuitively, $\alpha^{(1)}$ contains the entries of α that are coefficients of columns of V which are from the matrices $\tilde{S}_t \circledast \tilde{S}_t$ while $\alpha^{(2)}$ contains the coefficients of the columns of V of the form $\mathrm{vec}(\tilde{U}_t)$.

[15] Think of $Q = [Q_1\ Q_2\ \dots\ Q_m]$, where the Q_is are vectorized into columns. Then we vectorize Q in row-major order, rather than column major order.

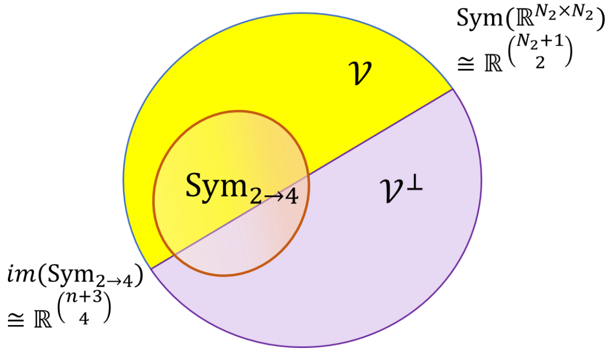
$\mathrm{Sym}(\mathbb{R}^{N_2 \times N_2})$
$\cong \mathbb{R}^{\binom{N_2+1}{2}}$
$\mathcal{V}$
$\mathrm{Sym}_{2\to 4}$
$\mathcal{V}^{\perp}$
$im(\mathrm{Sym}_{2\to 4})$
$\cong \mathbb{R}^{\binom{n+3}{4}}$

Figure 7: In this illustration, $N_2 = \binom{n+1}{2}$ is the dimension of the space of homogeneous quadratic polynomials. The picture shows the space corresponding to the symmetric lift $(\mathbb{R}^{N_2})^{\circledast 2} \cong \mathrm{Sym}(\mathbb{R}^{N_2 \times N_2})$ being expressed as $\mathcal{V} \oplus \mathcal{V}^{\perp}$, where $\mathcal{V} = (\mathcal{A} \circledast \mathcal{A}) \oplus (\mathcal{A} \circledast \mathcal{A}^{\perp})$ as defined in (37). $\mathrm{Sym}_{2\to 4}$ denotes the orthogonal projector onto the space of fully symmetric tensors in $(\mathbb{R}^n)^{\otimes 4}$. Lemma E.4 shows that $\ker(\mathrm{Sym}_{2\to 4}) \cap \mathcal{V} = \{0\}$ w.h.p. Note that $\mathrm{Sym}_{2\to 4}$ may not be a projection matrix when restricted to $\mathcal{V}$.

We consider two cases. In the first case suppose that $\|\alpha^{(2)}\| \le \frac{\sqrt{\varepsilon}}{2\sqrt{mL}}$. In this case we upper bound the probability that $\|V\alpha\| \le \sqrt{\varepsilon}$, which in turn upper bounds the probability that $\|V\alpha\| \le \varepsilon$. We have

$$\begin{aligned}\|V\alpha\| = \Big\| \big[\tilde{S}_1 \circledast \tilde{S}_1 \quad \ldots \quad \tilde{S}_m \circledast \tilde{S}_m\big]\, \alpha^{(1)} \\ + \big[\mathrm{vec}(\tilde{U}_1) \quad \ldots \quad \mathrm{vec}(\tilde{U}_m)\big]\, \alpha^{(2)} \Big\| \\ \ge \Big\| \big[\tilde{S}_1 \circledast \tilde{S}_1 \quad \ldots \quad \tilde{S}_m \circledast \tilde{S}_m\big]\, \alpha^{(1)} \Big\| - \sqrt{mL}\|\alpha^{(2)}\| \ge \frac{\sqrt{\varepsilon}}{2}.\end{aligned}$$

This would imply that

$$\sigma_{\min}\left(\big[\tilde{S}_1 \circledast \tilde{S}_1 \quad \ldots \quad \tilde{S}_m \circledast \tilde{S}_m\big] \right) < \varepsilon/4.$$

Taking $\varepsilon = \frac{c' \min\{1,\rho^4\}}{n^c}$ as in the statement of the proposition, we can use Corollary 5.3 to conclude that the probability that such an α exists is at most $\exp\big(-\Omega_m(n)\big)$.

In the second case, we have that $\|\alpha^{(2)}\| > \frac{\varepsilon}{2\sqrt{mL}}$. Therefore, $\alpha^{(2)}$ must have some component $\alpha_j^{(2)}$ with magnitude at least $|\alpha_j^{(2)}| \ge \frac{\sqrt{\varepsilon}}{2mL}$. It follows that

$$\|\Pi_{QV_j}^{\perp} Q\mathrm{vec}(\tilde{U}_j)\| \|\alpha_j^{(2)}\| = \|\Pi_{QV_j}^{\perp} QV\alpha\| \le \|V\alpha\| < \varepsilon.$$

Here the first equality follows from the fact that the only nonzero column of $\Pi_{QV_j}^{\perp} QV$ is $\Pi_{QV_j}^{\perp} Q\mathrm{vec}(\tilde{U}_j)$ which implies $\Pi_{QV_j}^{\perp} QV\alpha = \Pi_{QV_j}^{\perp} Q\mathrm{vec}(\tilde{U}_j)\alpha_j^{(2)}$. From this we obtain

$$\|\Pi_{QV_j}^{\perp} Q\mathrm{vec}(\tilde{U}_j)\| < 2mL\sqrt{\varepsilon}.$$

Using the discussion above, the probability that there exists some $j \in [m]$ for which this inequality holds is at most $m\left(\frac{c'' mL\sqrt{\varepsilon}}{\rho}\right)^r$.

Taking $\varepsilon = \frac{c' \min\{1,\rho^4\}}{n^c}$ as in the statement of the proposition and recalling that $r \ge \delta n^2$, this can be written as $m\left(\frac{c'' mL\sqrt{\varepsilon}}{\rho}\right)^r = \exp\big(-\Omega_{m,\delta}(n)\big)$

Combining these two cases, we see that the probability that there exists a unit vector α such that $\|V\alpha\| \le \frac{c' \min\{1,\rho^4\}}{n^c}$ is bounded above by

$$\exp\big(-\Omega_{\delta,m}(n)\big) + \exp\big(-\Omega_m(n)\big) = \exp\big(-\Omega_{\delta,m}(n)\big).$$

which completes the proof.

□

E.3 Bounding the solutions for system of equations: Proof of Proposition E.3

Recall that $N_2 = \binom{n+1}{2}$ is the dimension of the space of all symmetric homogeneous polynomials of degree 2 in n variables. Let $\mathrm{Sym}_{2\to 4} : (\mathbb{R}^n)^{\circledast 2} \otimes (\mathbb{R}^n)^{\circledast 2} \to (\mathbb{R}^n)^{\circledast 4}$ be the orthogonal projector onto the fully symmetric space over 4th-order tensors. Note that if $v_1, v_2 \in (\mathbb{R}^n)^{\circledast 2} \cong \mathbb{R}^{N_2}$, we have that $\mathrm{Sym}_{2\to 4}(v_1 \otimes v_2) = \mathrm{Sym}_{2\to 4}(v_2 \otimes v_1) = \frac{1}{2}\mathrm{Sym}_{2\to 4}(v_1 \otimes v_2 + v_2 \otimes v_1)$. (Note that there are other symmetries that are also captured by $\mathrm{Sym}_{2\to 4}$.)

In the smoothed setting $A_1, \ldots, A_m \in \mathbb{R}^{N_2} \cong (\mathbb{R}^n)^{\circledast 2}$ are randomly ρ-perturbed and represent the polynomials $a_1(x), \ldots, a_m(x)$, and let $\mathcal{A} := \mathrm{span}(\{A_i : i \in [m]\})$. With high probability, the A_i are linearly independent, in which case we can let $F_{m+1}, \ldots, F_{N_2} \in \mathbb{R}^{N_2}$ be a (random) orthonormal basis for $\mathcal{A}^{\perp}$. Together $A_1, \ldots, A_m, F_{m+1}, \ldots, F_{N_2}$ form a basis for $\mathbb{R}^{N_2}$. Consider the space $\mathcal{V} \subset (\mathbb{R}^{N_2})^{\circledast 2}$ given by

$$\begin{aligned}\mathcal{V} := \mathrm{span}\Big(\big\{A_i \otimes A_j + A_j \otimes A_i : 1 \le i \le j \le m\big\} \\ \bigcup \big\{A_i \otimes F_j + F_j \otimes A_i : i \in [m], j \in [N_2] \setminus [m]\big\}\Big) \\ = (\mathcal{A} \circledast \mathcal{A}) \oplus (\mathcal{A} \circledast \mathcal{A}^{\perp}). \qquad (37)\end{aligned}$$

Observe that $\mathcal{V}^{\perp} = \mathcal{A}^{\perp} \circledast \mathcal{A}^{\perp} \subset (\mathbb{R}^{N_2})^{\circledast 2}$. Proposition E.3 is challenging to show because it is not easy to reason about the nullspace of the system of equations directly. Instead we argue about the larger vector space $\mathbb{R}^{N_2} \circledast \mathbb{R}^{N_2}$ to help identify a basis for the range space of the polynomial system. Claim E.3 is implied by the following crucial lemma.

Lemma E.4 ($\mathrm{Sym}_{2\to 4}$ does not annihilate any vector in $\mathcal{V}$ for smoothed instances). *Consider the following matrix $M \in \mathbb{R}^{\binom{n+3}{4} \times (mN_2 - \binom{m}{2})}$ formed by columns*

$$columns(M) = \big\{\mathrm{Sym}_{2\to 4}(A_i \otimes A_j + A_j \otimes A_i) : 1 \le i \le j \le m\big\} \qquad (38)$$

$$\bigcup \big\{\mathrm{Sym}_{2\to 4}(A_i \otimes F_j + F_j \otimes A_i) : 1 \le i \le m, m+1 \le j \le N_2\big\}.$$

There exists a constant $c > 0$, such that when $m < cN_2$, with probability at least $1 - \exp(-\Omega(n))$, we have that M has full column rank in a robust sense i.e.,

$$\sigma_{\min}(M) \ge \frac{\rho^2}{n^{O(1)}}. \qquad (39)$$

Note that $\binom{m+1}{2}+m(N_2-m) = mN_2-\binom{m}{2}$ is the number of columns of the above matrix. . The above lemma shows that while $\mathsf{Sym}_{2\to 4}$ is an orthogonal projection matrix of rank $\binom{n+3}{4}$ acting on a space of dimension $\binom{N_2+1}{2}$ (which is larger by a constant factor ≈ 3), it *does not annihilate* any vector in the vector space $\mathcal{V}$. In other words we show that $\ker(\mathsf{Sym}_{2\to 4}) \cap \mathcal{V} = \{0\}$. See Figure 7 for an illustration. Moreover this is true in a robust sense. We now show why the above lemma suffices for the Proposition E.3.

Proof of Proposition E.3 from Lemma E.4. We know that $\sigma_m(A) \geq \rho/n^{O(1)}$ with probability $1-\exp(-\Omega(N_2))$, since $m < cN_2$. We condition on this event for the rest of the argument. The proof just follows from identifying the null space and a dimension counting argument. From the properties of $\mathsf{Sym}_{2\to 4}$ stated earlier, for any $v_1, v_2 \in \mathbb{R}^{N_2} \cong (\mathbb{R}^n)^{\circledast 2} \cong \mathbb{R}^{N_2}$ we have

$$\mathsf{Sym}_4(v_1 \otimes v_2) = \mathsf{Sym}_{2\to 4}\Big(\frac{(v_1\otimes v_2 + v_2 \otimes v_1)}{2}\Big) = \mathsf{Sym}_{2\to 4}(v_1\otimes v_2).$$

Hence, we can restrict to the subspace $\mathbb{R}^{N_2} \circledast \mathbb{R}^{N_2}$ as considered in Figure 7. Now we observe that

$$\forall w \in (\mathcal{A}^\perp) \circledast (\mathcal{A}^\perp), v \in \mathcal{A} \circledast \mathbb{R}^{N_2}, \text{ we have } \langle w, v\rangle = 0, \tag{40}$$
$$\text{i.e., } \forall j_1, j_2 \in \{m+1, \dots, N_2\},$$
$$\forall i \in [m], v \in \mathbb{R}^{N_2},$$
$$\langle (F_{j_1}\otimes F_{j_2} + F_{j_2} \otimes F_{j_1}), (A_i \otimes v + v \otimes A_i)\rangle = 0.$$

The subspace $\mathcal{V}^\perp = \mathcal{A}^\perp \circledast \mathcal{A}^\perp$ has dimension $\binom{N_2-m+1}{2}$. The space orthogonal to this is exactly $\mathcal{V}$ and has dimension

$$\begin{aligned}\dim(\mathcal{V}) &= \binom{N_2+1}{2} - \binom{N_2-m+1}{2}\\ &= \frac{N_2(N_2+1)}{2} - \frac{(N_2-m)(N_2-m+1)}{2}\\ &= mN_2 - \binom{m}{2}.\end{aligned}$$

Moreover, each of the $mN_2 - \binom{m}{2}$ columns of the matrix M from (38) belong to the column space of the matrix $\mathsf{Sym}_4\ (I_{N_2\times N_2} \otimes A)$ of Proposition E.3. Hence Lemma E.4 implies Proposition E.3. □

Proof of Lemma E.4 using random restrictions and contraction. We now proceed to the proof of Lemma E.4. We use the ideas we have developed in the previous sections to tackle the random matrix in Lemma E.4. As in Section 5, we will view $\mathsf{Sym}_{2\to 4}$ as a tensor $W_{2\to 4} \in \mathbb{R}^{\binom{n+3}{4}\times N_2 \times N_2}$; this tensor has rank $\binom{n+3}{4}$ when viewed as a flattened $\mathbb{R}^{\binom{n+3}{4}\times N_2^2}$ matrix. The random matrix in Lemma E.4 is very similar to the random matrices analyzed in Section 5 through random contractions. However in this case, we are doing *lopsided* "random contractions" corresponding to $A^{\circledast 2} + A \otimes F$ where[16] $A \in \mathbb{R}^{N_2 \times m}, F \in \mathbb{R}^{N_2\times(N-m)}$, and we want the entire set of $m \times N_2$ array of vectors in $R = \binom{n+3}{4}$ dimensions to be linearly independent in a robust sense. While there is a clear contraction along one mode (N_2 to $m \ll N_2$), there is no effective reduction in the dimension along the other mode (it remains N_2 since the concatenated matrix $[A, F]$ still has N_2 columns).

To tackle this, we will use a stronger property of the tensor $W_{2\to 4}$ (the tensor corresponding to the operator $\mathsf{Sym}_{2\to 4}$) that ensures that we can keep all of the dimensions corresponding to one of the modes. Consider the slices $W_1, \dots, W_{N_2} \in \mathbb{R}^{R\times N_2}$ (it does not matter which of the last two modes we consider since it is symmetric), and let $\Pi^\perp_{-i}$ denote the projector perpendicular to the span of $\cup_{i'\in[N_2]\setminus\{i\}}\mathrm{cols}(W_{i'})$. We would like to argue that for all $i \in [N_2]$, $\Pi^\perp_{-i}W_i$ has a large rank of $\Omega(N_2)$ in a robust sense i.e., $\forall i \in [N_2]$, $\sigma_{\Omega(N_2)}(\Pi^\perp_{-i}W_i)$ is inverse polynomial with high probability. However, this is unfortunately too good to be true.

The key idea here is to use the *random restriction* idea from Section **??**. Set $\delta := 1/16$. We define a random set $T \subseteq [N_2]$ such that each $i \in [N_2]$ belongs to T independently with probability δ each. For each $i \in [N_2]$, let $W_{i,T} \in \mathbb{R}^{R\times T}$ denote the restriction of W_i to the columns given by T. Similarly, let $\Pi^\perp_{-i,T}$ denote the (orthogonal) projection matrix that is perpendicular to the span of $\cup_{i'\in[N_2]\setminus\{i\}}\mathrm{cols}(W_{i',T})$. We show the following claim.

Claim E.5. *(Large relative rank when random restricted to $T \subseteq [N_2]$) In the above notation, there is a constant $c \geq 1/64$ such that with probability $1-\exp(-\Omega(N_2))$ over the random choice of $T \subseteq [N_2]$*

$$\forall i \in [N_2],\ \sigma_{cN_2}(\Pi^\perp_{-i,T}W_{i,T}) \geq 1, \tag{41}$$

where $\Pi_{-i,T}$ is the projector onto the span of the union of the columns of $W_{i',T}$ for all $i' \neq i$. In particular, there exists a $T \subseteq [N_2]$ such that (41) holds (deterministically).

Proof. Set $\delta := 1/16$. In this proof, we will have unordered 4-tuples represented by multiset of the form $\{i_1,i_2,i_3,i_4\}$ where $i_1,\dots,i_4 \in [n]$; similarly we will use multisets of the form $\{i_1,i_2\}$ for unordered pairs. For example $\{1,3,3,5\}$ is the same as $\{3,5,1,3\}$. We will use the multiset $\{i_1,i_2,i_3,i_4\}$ given by the canonical ordering $1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq n$ to uniquely represent the unordered 4-tuple. Note that some indices may be repeated.

Consider any $i = \{i_1,i_2\} \in [N_2]$ where $1 \leq i_1 \leq i_2 \leq n$. We will now argue about $\Pi_{-i,T}W_i$ Consider a fixed $j = \{j_1,j_2\} \in T$ with $j_1 \leq j_2$. Now the vector corresponding to $W(:,\{i_1,i_2\},\{j_1,j_2\})$ has exactly one entry that is 1 corresponding to the index given by the unordered tuple $\{i_1,i_2,j_1,j_2\}$, and 0 otherwise. In other words, if e_i refers to the ith standard basis vector (in appropriate dimensions), then $\mathsf{Sym}_{2\to 4}(e_{\{i_1,i_2\}}, e_{\{j_1,j_2\}}) = e_{\{i_1,i_2,j_1,j_2\}}$. But there are at most $\binom{4}{2}-1$ other pairs $i' \in [N_2], j' \in [N_2]$ whose vectors $W(:,i',j') \in \mathbb{R}^{\binom{n+3}{4}}$ have a non-zero entry corresponding at the $\{i_1,i_2,j_1,j_2\}$th index. Each of these other $\binom{4}{2}-1$ choices for $i' = \{i'_1, i'_2\}$ has its corresponding $j' = \{j'_1, j'_2\}$ present in T with probability at most δ (note that $j' \neq j$ since $i' \neq i$, and hence conditioning on $j \in T$ does not affect whether $j' \in T$). Hence, by a union bound, we have

$$\mathbb{P}\left[(\Pi^\perp_{-i,T}W_i)e_{\{i_1,i_2,j_1,j_2\}} = e_{\{i_1,i_2,j_1,j_2\}}\right] \geq 1 - \left(\binom{4}{2}-1\right)\delta \geq 1-5\delta.$$

Moreover the above event is independent for each $j = \{j_1,j_2\} \in [N_2]$. The expected number of indices $j = \{j_1,j_2\}$ which belong to T and the above event $(\Pi^\perp_{-i,T}W_{i,T})e_{\{i_1,i_2,j_1,j_2\}} = e_{\{i_1,i_2,j_1,j_2\}}$ holds is at least

$$\delta(1-5\delta)|N_2| \geq \delta/2|N_2|.$$

[16] Note that this does not quite correspond to a Kronecker product as in Section 5, but we will see that similar arguments can be applied here.

These unordered pairs $\{j_1, j_2\}$ that satisfy $(\Pi^{\perp}_{-i,T} W_{i,T}) e_{\{i_1,i_2,j_1,j_2\}} = e_{\{i_1,i_2,j_1,j_2\}}$ define an entire subspace of vectors v such that $\|\Pi^{\perp}_{-i,T} W_{i,T} v\| \geq \|v\|$. Hence by the variational characterization of singular values and standard large deviation bounds we have that

$$\mathbb{P}\left[\sigma_{\delta N_2/4}(\Pi^{\perp}_{-i,T} W_{i,T}) \geq 1\right] \geq 1 - \exp(-\Omega(N_2)).$$

By a union bound over $i \in [N_2]$, we get the statement of our claim.

□

We now use Claim E.5 along with contraction along the smoothed direction A to get our final lemma. Recall $R = \binom{n+3}{4}$. Let $\Phi \in \mathbb{R}^{R \times N_2^2}$ denotes the natural matrix representation of $\mathrm{Sym}_{2\to 4}$ acting on $\mathbb{R}^{N_2^2}$ obtained by flattening $\mathrm{Sym}_{2\to 4}$ appropriately. Define the matrix $M \in \mathbb{R}^{R \times (\binom{m+1}{2} + m(N_2 - m))}$ to be the block matrix

$$M = \Big[\Phi(A \circledast A)\ ,\ \Phi(A \otimes F)\Big] = \Phi\Big[A \circledast A\ ,\ A \otimes F\Big],$$

where $F \in \mathbb{R}^{N_2 \times (N_2 - m)}$ matrix with columns $F_{m+1}, \ldots, F_{N_2}$. To analyze the least singular value of M, we first need to apply the decoupling technique in used Lemma C.4.

Let $A = B + Z_1 + Z_2$ where $Z_1, Z_2 \in \mathbb{R}^{N_2 \times m}$ are random Gaussian matrixes with independent entries $N(0, \rho_1^2)$ and $N(0, \rho_2^2)$ respectively with $\rho_1^2 + \rho_2^2 = \rho^2$. Then by the arguments in Lemma C.4, it suffices to consider the matrix $M' \in \mathbb{R}^{R \times m N_2}$ given below and prove an inverse polynomial lower bound on its least singular value. That is, we need to show that with probability at least $1 - \exp(-\Omega(n))$

$$\sigma_{mN_2}(M') \geq \frac{\Omega(\rho)}{n^{O(1)}}, \text{ where} \tag{42}$$

$$M' := \Phi\Big[(B + Z_1) \otimes (B + Z_1 + 2Z_2)\ ,\ (B + Z_1 + Z_2) \otimes F\Big]. \tag{43}$$

The above bound (42) will follow from the following two simpler claims.

Claim E.6. *In the above notation, the matrix $Q = [B + Z_1 + 2Z_2\ ,\ F]$ is full rank in a robust sense i.e., with probability $1 - \exp(-\Omega(N_2))$*

$$\sigma_{N_2}(Q) \geq \frac{c\rho}{n^{O(1)}},$$

for some absolute constant $c > 0$.

Note that the matrix $[B + Z_1 + Z_2\ ,\ F] = [A\ ,\ F]$ has inverse polynomial least singular value by design, since F was chosen to complete the basis and $\sigma_m(A)$ is lower bounded w.h.p. We just need to show that adding the random matrix $[Z_2\ ,\ 0]$ does not affect its least singular value. This is shown by rewriting the random matrices and a standard net argument in Appendix **??**. The second claim shows that after applying a random modal contraction with the random matrix $U = [B + Z_1\ ,\ Z_2] \in \mathbb{R}^{N_2 \times (2m)}$, the matrix is well-conditioned.

Proof. Let $Q_1 = \big(B + Z_1 + Z_2\ ,\ F\big) \in \mathbb{R}^{N_2 \times N_2}$ and $Q_2 = (Z_2\ ,\ 0) \in \mathbb{R}^{N_2 \times N_2}$. First we note that with probability $1 - \exp(-\Omega(N_2))$, there exists constants $c_1, c_2 > 0$ such that

$$\sigma_{N_2}(Q_1) = \sigma_{N_2}\big(B + Z_1 + Z_2\ ,\ F\big) = \sigma_{N_2}(A, F) \geq \frac{c_2 \rho}{n^{c_1}} =: \tau\rho.$$

This is because $\sigma_m(A) \geq \frac{\rho}{n^{\Omega(1)}}$ with high probability by standard random matrix theory, and since F was chosen to complete the basis. We just need to show that adding the random matrix Q_2 does not affect decrease its least singular value by a lot.

Recall $Z_1 \sim N(0, \rho_1^2)^{N_2 \times m}, Z_2 \sim N(0, \rho_2^2)^{N_2 \times m}$. Next we rewrite the random matrices $Z_1 + 2Z_2$ and Z_2 as follows after Gram-Schmidt orthogonalization. Suppose $Y_1, Y_2 \in \mathbb{R}^{N_2 \times m}$ are random matrices with independent entries given by

$$Y_1 \sim_{iid} N(0, \lambda_1^2)^{N_2 \times m},\ Y_2 \sim_{iid} N(0, \lambda_2^2)^{N_2 \times m},$$

$$\text{where } \lambda_1^2 = \rho_1^2 + \rho_2^2, \lambda_2^2 = \frac{\rho_1^2 \rho_2^2}{\rho_1^2 + \rho_2^2},$$

$$\begin{aligned} Z_1 + Z_2 &= Y_1, \\ \text{and } Z_2 &= \tfrac{\rho_2^2}{\rho_1^2 + \rho_2^2} Y_1 + Y_2, \\ \text{i.e., } Y_2 &= \tfrac{1}{\rho_1^2 + \rho_2^2}(\rho_2^2 Z_1 - \rho_1^2 Z_2) \end{aligned} \tag{44}$$

Note that Y_1 and Y_2 are mutually independent (these are Gaussian r.v.s, and one can verify entrywise that their correlation is 0). Let $\gamma = \rho_2^2/(\rho_1^2 + \rho_2^2)$. Note that $\gamma \in [c_5 N_2^{-c_4}, c_6 N_2^{c_4}]$ for some constants $c_4, c_5, c_6 > 0$. We know that $Q_1 = (B + Y_1 \mid F)$ and $Q_2 = (\gamma Y_1 + Y_2 \mid 0)$.

In the rest of this proof we condition on $\sigma_{N_2}(Q_1) \geq$ Consider any test unit vector $\alpha \in \mathbb{R}^{N_2}$, and let $\alpha_{[m]}$ be the restriction to the first m coordinates. We will now show that for any constant $C > 0$, $\|Q\alpha\| \geq \Omega(\rho^2/n^{O(C)})$ with probability $1 - \exp(-CN_2 \log(N_2))$. We have

$$Q\alpha = Q_1\alpha + \gamma Y_1 \alpha_{[m]} + Y_2 \alpha_{[m]}, \text{ where } \|Q_1\alpha\|_2 \geq \tau\rho.$$

We know further that with high probability $\|Y_1\|, \|Y_2\| \leq c'\rho\sqrt{N_2}$ for some constant $c' > 0$. We split into two cases depending on whether **(a)** $\|\alpha_{[m]}\|_2 \leq \tau/(2c'(1+\gamma)\sqrt{N_2})$, or **(b)** otherwise.

In case **(a)**, we have

$$\begin{aligned} \|Q\alpha\|_2 &\geq \|Q_1\alpha\|_2 - \|\gamma Y_1 \alpha_{[m]}\|_2 - \|Y_2 \alpha_{[m]}\|_2 \\ &\geq \tau\rho - c'\sqrt{N_2}\rho(\gamma + 1)\|\alpha_{[m]}\|_2 \\ &\geq \frac{\tau}{2}. \end{aligned}$$

In case **(b)** $\|\alpha_{[m]}\|_2 \geq \tau/(2c'(1+\gamma)\sqrt{N_2})$. We use the anticoncentration from the Gaussian r.v. $Y_2\alpha_{[m]}$. Let $\beta = Q_1\alpha + \gamma Y_1 \alpha_{[m]} \in \mathbb{R}^{N_2}$. For matrices $Q, Y_2 \in \mathbb{R}^{N_2 \times m}$, let $Q(i), Y_2(i) \in \mathbb{R}^m$ represent the ith rows. We have for some absolute constant $c_7 > 0$

$$\begin{aligned} \mathbb{P}\Big[\|Q\alpha\|_2 \leq \varepsilon\Big] &\leq \mathbb{P}\Big[\forall i \in [N_2], |\langle Q(i), \alpha\rangle| \leq \varepsilon\Big] \\ &= \mathbb{P}\Big[\forall i \in [N_2], |\beta_i + \langle Y_2(i), \alpha_{[m]}\rangle| \leq \varepsilon\Big] \\ &= \mathop{\mathbb{P}}_{\substack{g_1,\ldots,g_{N_2} \sim iid \\ N(0,\lambda_2\|\alpha_{[m]}\|)}} \Big[\forall i \in [N_2],\ |\beta_i + g_i| \leq \varepsilon\Big] \\ &\leq \left(\frac{\varepsilon \cdot (1+\gamma)\sqrt{N_2}}{c_7 \lambda_2 \tau}\right)^{N_2} \leq N_2^{-CN_2}, \end{aligned}$$

by setting ε appropriately as $\varepsilon = c_7 \tau \lambda_2/((1+\gamma)N_2^{C+1})$, which is still inverse polynomial in N_2. Now by doing a standard union bound argument over a net of $\alpha \in \mathbb{R}^{N_2}$, the claim follows. □

Claim E.7. *Let* $U = [B + Z_1\ ,\ Z_2]$. *Suppose the matrix* $W \in \mathbb{R}^{R\times(2m\cdot N_2)}$ *is obtained by random modal contraction applied to* $\mathsf{Sym}_{2\to 4}$ *(along the second mode). Formally, suppose* $\Phi_1, \Phi_2, \ldots, \Phi_{N_2} \in \mathbb{R}^{R\times N_2}$ *represent the matrix slices of* Φ, *and suppose* $\forall i \in [N_2]$ $W_i = \Phi_i U$ *where random matrix* $U = [B + Z_1\ ,\ Z_2] \in \mathbb{R}^{N_2\times(2m)}$. *Then the matrix* $W = [W_1 \mid W_2 \mid \ \ldots\ \mid W_{N_2}]$ *satisfies with probability at least* $1 - \exp(-\Omega(N_2))$ *that* $\sigma_{(2mN_2)}(W) \geq c'\rho/n^c$ *for absolute constants* $c, c' > 0$

Proof. We will use Claim E.5 with the matrix Φ. Let $T \subseteq [N_2]$ be any subset that satisfies (41). Set $s = N_2$, $\varepsilon = \delta/4$, $k = |T|$. We will set the matrix $A_1 = \Phi_{1,T}, \ldots, A_{N_2} = \Phi_{N_2,T}$. Let $\tilde{U} = U_T$ (rows restricted to T) and let $U' = U_{[N_2]\setminus T}$; note that they are mutually independent. For each $i \in [N_2]$, $C_i = \Phi_{i,[N_2]\setminus T} U'$. For each $i \in [N_2]$, $W_i = C_i + A_i \tilde{U}$. Hence applying Lemma 5.5 we get the claim. □

With Claims E.6 and E.7 in hand we can now establish (42) complete the proof of Lemma E.4. Consider the matrix $\widehat{M} \in \mathbb{R}^{R\times(2mN_2)}$ given by

$$\widehat{M} = \Phi\Big(\big[B + Z_1\ ,\ Z_2\big] \otimes \big[(B + Z_1 + 2Z_2)\ ,\ F\big]\Big), \tag{45}$$

$$= \Phi\Big(U \otimes Q\Big),$$

$$\text{where } U = [B + Z_1\ ,\ Z_2] \in \mathbb{R}^{N_2\times(2m)},$$

$$\text{and } Q = [B + Z_1 + 2Z_2\ ,\ F] \in \mathbb{R}^{N_2\times N_2},$$

as defined in the earlier claims. Note that $M' \in \mathbb{R}^{R\times mN_2}$ is a submatrix of $\widehat{M} \in \mathbb{R}^{R\times(2mN_2)}$. Hence a least singular value bound on $\widehat{M}$ (note we show it has full column rank) implies the same least singular value bound for M'. Suppose $W \in \mathbb{R}^{R\times(2mN_2)}$ is the matrix defined in Claim E.7 and $W^{(1)}, \ldots, W^{(2m)} \in \mathbb{R}^{R\times N_2}$ are corresponding blocks of W that correspond to the matrix slices taken along the second mode of the tensor corresponding to W. Then up to rearranging columns

$$\widehat{M} \coloneqq \Big(W^{(1)}Q\ ,\ W^{(2)}Q\ ,\ \ldots\ ,W^{(2m)}Q\Big) \tag{46}$$

$$= \Big(W^{(1)}\ ,\ W^{(2)}\ ,\ \ldots\ ,W^{(2m)}\Big)\Big(I_{2m\times 2m} \otimes Q\Big). \tag{47}$$

We condition on the success events in both Claim E.6 and Claim E.7; by a union bound, they both hold with probability at least $1 - 2\exp(-\Omega(n))$. From Claim E.7, the matrix W defined in Claim E.7 has

$$\sigma_{2mN_2}\Big(W^{(1)}\ ,\ W^{(2)}\ ,\ \ldots\ ,W^{(2m)}\Big) = \sigma_{2mN_2}\big(W\big) \geq \frac{c'\rho}{n^{O(1)}}. \tag{48}$$

From Claim E.6,

$$\sigma_{2mN_2}\Big(I_{2m\times 2m} \otimes Q\Big) = \sigma_{N_2}(Q) \geq \frac{c''\rho}{n^{O(1)}}. \tag{49}$$

Combining (47), (48) and (49), we finish the proof of the lemma. □

Received 13-NOV-2023; accepted 2024-02-11