

A Simple Algorithm for Best Separable State

Prashanti Anderson
MIT
paanders@mit.edu

Samuel B. Hopkins
MIT
samhop@mit.edu

Amit Rajaraman
MIT
amit_r@mit.edu

August 10, 2026

Abstract

We study the *best separable state* problem (BSS), which asks for the maximum acceptance probability of a quantum measurement over unentangled states. In classical terms, the goal is to maximize $\langle (x \otimes y), M(x \otimes y) \rangle$ over unit vectors x, y where $0 \preceq M \preceq I$; we call this value $\text{BSS}(M)$. We study BSS in the “perfect completeness” regime, where given M such that $\text{BSS}(M) = 1$ the goal is to find the best possible solution x, y – this generalizes the problem of finding a rank-one matrix as close as possible to a given subspace of $\mathbb{R}^{n \times n}$ guaranteed to contain a rank-one matrix.

The strongest known algorithmic guarantees for this problem are: (1) an algorithm which finds a solution with value $1 - \varepsilon$ in time $\exp(\sqrt{n}(\log n)^{O(1)}/\varepsilon^2)$, due to Barak, Kothari, and Steurer, and (2) an algorithm which finds a solution with value q/n in time roughly $n^{O(q)}$, due to Bhattachiprolu, Ghosh, Guruswami, Lee, and Tulsiani.

We give a much simpler approach to rounding the SoS relaxation, generalizing the canonical “global correlation rounding” technique, and obtain a better running time. Given M with $\text{BSS}(M) = 1$, our algorithm finds a solution with value $1 - \varepsilon$ in time $n^{O(\sqrt{n}/\varepsilon)}$, and a solution of value q/n in time $n^{O(\sqrt{q})}$.

Using the same techniques, we prove a new variant of the “pinning lemma,” a measure-decomposition theorem widely used in LP/SDP rounding, high-dimensional probability, and statistical physics, which we believe is of independent interest.

Contents

1	Introduction	1
1.1	Related work	3
2	Technical Overview and Warmup	4
2.1	1 versus $1 - \varepsilon$	5
2.2	1 versus q/n	7
2.3	Removing the $\sqrt{\log n}$ factors	8
2.4	Fine-grained pinning lemma	9
3	Preliminaries	11
4	Best Separable State: 1 vs $1 - \varepsilon$	13
4.1	The rounding success condition	14
4.2	A new pinning lemma	15
4.3	A potential growth lemma	16
4.4	Proof of the main theorem	20
5	Best Separable State: 1 vs q/n	22
5.1	The rounding success condition	23
5.2	A low-soundness reweighing lemma	24
5.3	A low-soundness potential growth lemma	26
5.4	Consequences of affine failure	29
5.5	The proof when $ \zeta $ is large	33
5.6	Proof of the main theorem	34
6	Fine-grained pinning lemma	36

1 Introduction

How can we detect quantum entanglement? In the *best separable state* problem (henceforth, *BSS*), we are asked to take a description of a quantum measurement device and find the maximum probability that the measurement will accept over all un-entangled (“separable”) quantum states. More formally, given a Hermitian matrix $M \in \mathbb{C}^{n^2 \times n^2}$ with $0 \preceq M \preceq I$, the goal is to find or approximate

$$\text{BSS}(M) = \max_{\|x\|=1, \|y\|=1} \langle x \otimes y, M(x \otimes y) \rangle$$

where $x, y \in \mathbb{C}^n$ and $\|\cdot\|$ is the Euclidean norm.¹ BSS also amounts to a fundamental classical algorithmic problem: find the maximum of a degree-4 polynomial over the high-dimensional unit sphere. While this problem is likely computationally intractable for worst-case polynomials, BSS restricts to an interesting special case: polynomials whose representations as matrices have bounded eigenvalues.

Algorithms for BSS are equivalent to classical upper bounds on the complexity class QMA[2], the class of computational problems whose solutions can be verified by a verifier with access to two un-entangled quantum provers.² The relationship of QMA[2] to classical complexity classes is a longstanding mystery in quantum complexity theory: there is a straightforward upper bound $\text{QMA}[2] \subseteq \text{NEXP}$, but $\text{QMA}[2]$ is widely suspected to be contained in smaller complexity classes. Beyond quantum complexity, BSS is intimately intertwined with several other important questions. For example, the 2-to-4 norm is exactly BSS on a structured family of measurement operators, and some reductions in the reverse direction are known [HM13; BBHKSZ12; BH15]; the 2-to-4 norm is also closely related to Small-Set Expansion and the Unique Games Conjecture [Kho02; BBHKSZ12]. And in the special case that M is the projector to a subspace of $n \times n$ matrices, BSS becomes the natural question of finding the closest rank-one matrix to the subspace.

It is a folklore result [NH17] that the top eigenvalue $\lambda_{\max}(M)$ satisfies $\text{BSS}(M) \leq \lambda_{\max}(M) \leq n \cdot \text{BSS}(M)$, and hence gives a polynomial-time n -approximation of $\text{BSS}(M)$. An exponential-time brute-force algorithm based on discretizing the unit sphere solves $\text{BSS}(M)$ exactly (up to numerical errors). But these algorithms are very far from the best known lower bounds, which say only that a constant-factor approximation in $n^{o(\log n)}$ time would disprove ETH [HM13] and that computing $\text{BSS}(M)$ up to additive $1/\text{poly}(n)$ gaps is NP-hard [BT12]. Efficient algorithms closer to these lower bounds would have significant consequences, the most striking of which is that a quasipolynomial-time constant-factor approximation algorithm would imply that $\text{QMA}[2] \subseteq \text{EXP}$.

The first algorithm to go beyond the spectral n -approximation and the brute force exponential-time algorithm, due to Barak, Kothari, and Steurer, uses the sum-of-squares (SoS) hierarchy to solve BSS in the “perfect completeness” case in subexponential time [BKS17]. Here “perfect completeness” refers to the case that the BSS value equals 1.

Theorem 1.1 ([BKS17]). *There is an algorithm which distinguishes between the cases $\text{BSS}(M) = 1$ and $\text{BSS}(M) \leq 1 - \varepsilon$ in time $\exp\left(\sqrt{n} \log(n)^{O(1)} \cdot \frac{1}{\varepsilon^2}\right)$.*

[BKS17] also solves the search version of this problem: given M such that $\text{BSS}(M) = 1$, find x, y such that $\langle x \otimes y, M(x \otimes y) \rangle \geq 1 - \varepsilon$. To prove Theorem 1.1, [BKS17] introduces a sophisticated and challenging-to-analyze rounding scheme for the SoS relaxation of $\text{BSS}(M)$.

¹In this paper, we adopt the classical notation of $x \otimes y$ for the tensor product of the state x with the state y . In bra-ket notation, one would instead write $\max_{x,y} \langle xy | M | xy \rangle$.

²The choice of 2 provers is not essential here, as $\text{QMA}[k] = \text{QMA}[2]$ for any constant k [HM13].

Digesting their algorithm is an ongoing challenge for researchers in quantum information and approximation algorithms. Our main contribution is a very simple approach to rounding the SoS relaxation for $\text{BSS}(M)$, using *global correlation rounding*³ [Mon08; RT12; BRS11], which improves over Theorem 1.1.

Theorem 1.2 (Main result (see Theorem 4.1 and Theorem 5.1)). *There is an algorithm for best separable state with perfect completeness having the following guarantees.*

1. For every $0 < \varepsilon < 1$, the algorithm distinguishes $\text{BSS}(M) = 1$ from $\text{BSS}(M) \leq 1 - \varepsilon$ in time $n^{O(\sqrt{n/\varepsilon})}$.
2. For every $1 \leq q \leq n/2$, the algorithm distinguishes $\text{BSS}(M) = 1$ from $\text{BSS}(M) \leq q/n$ in time $n^{O(\sqrt{q})}$.

Furthermore, in both cases the algorithm solves the corresponding search problem in the same running time.

In the 1 versus $1 - \varepsilon$ regime, our algorithm improves over the state of the art by a $(\log n)^{O(1)}/\varepsilon^{1.5}$ factor in the exponent. This means that our algorithm retains subexponential running time (hence beating brute-force) for any $\varepsilon = n^{-1+\Omega(1)}$, whereas the algorithm of [BKS17] runs in subexponential time only when $\varepsilon = n^{-1/4+\Omega(1)}$.

In the 1 versus q/n regime, the fastest previously known algorithm is the general degree-4 polynomial optimization algorithm of [BGGLT17], which has running time $n^{O(q)}$ (likely improvable to $2^{O(q)}n^{O(1)}$, per [BGGLT17]), while ours has running time $n^{O(\sqrt{q})}$. Thus, our algorithm improves the state of the art for any $q \gg \log^2 n$.

We remark that a simplified proof we present in the technical overview Section 2, which loses a $\sqrt{\log n}$ factor in the exponent, has a simple analysis spanning only a few pages (!) and improves on the guarantees of [BKS17].

We state Theorem 1.2 in the setting of optimization over $\mathbb{C}$, but for simplicity in the proofs, we work over $\mathbb{R}$ as the complex version of the problem reduces to our real-valued version (see [BKS17, Section B]). We also work in the case where M is a projector in the proof sections, as BSS with perfect completeness reduces to the case of projectors via considering the projector to the top eigenspace.

Open problems Many open problems remain, whose solutions would likely introduce new fundamental algorithmic ideas for polynomial optimization. Two especially interesting examples are: can the $\sqrt{n}$ in the exponent of the running time can be reduced, perhaps all the way to $O(\log n)$? And, can $\text{BSS}(M) = 1$ versus $\text{BSS}(M) = 1 - n^{-1+\Omega(1)}$ be distinguished in polynomial time?

AI usage statement The initial analysis of a pinning-based rounding for best separable state was hatched and executed by the authors without AI assistance, as part of a long-running project studying BSS. We initially proved a weaker result than Theorem 1.2, losing a logarithmic and $1/\sqrt{\varepsilon}$ factor in the exponent, and restricted to the “symmetric” case, where the goal is to maximize $\langle (x \otimes x), M(x \otimes x) \rangle$ rather than $\langle (x \otimes y), M(x \otimes y) \rangle$.

AI (GPT 5.5 Pro, GPT 5.6 Sol) was then used to find arguments which remove the logarithmic factor and to generalize the result to the non-symmetric case (involving $\langle (x \otimes y), M(x \otimes y) \rangle$).

³While our first result for distinguishing $\text{BSS}(M) = 1$ from $\text{BSS}(M) \leq 1 - \varepsilon$ directly analyzes the canonical global correlation rounding algorithm, our second result (for distinguishing $\text{BSS}(M) = 1$ and $\text{BSS}(M) = q/n$) also uses a generalization of conditioning in the rounding procedure.

GPT 5.6 Sol also found the version of the loose-by- $\sqrt{\log n}$ argument we present in the technical overview—our original version, preserved in Section 6 because of its interesting connection to pinning lemmas, applies only to the symmetric case. Codex was used to aid in typesetting the paper, filling out related work, and validating citations.

1.1 Related work

Sum of Squares Algorithms for Optimizing Constant-Degree Polynomials Known algorithmic guarantees for approximating constant-degree polynomial optimization depend strongly on the domain and on the class of polynomials. For the simplex, de Klerk, Laurent, and Parrilo give a PTAS for fixed-degree polynomial optimization [KLP06]. For the unit sphere, general guarantees about the convergence of the SoS hierarchy are provided by [DW12; FF21], but the degree of SoS needed by those works for nontrivial guarantees yield exponential-time algorithms in the dimension. Worst-case multiplicative approximation for arbitrary even-degree polynomials over the sphere was studied by Bhattiprolu, Ghosh, Guruswami, Lee, and Tulsiani, who give approximation/runtime tradeoffs based on higher levels of SoS [BGGLT17]. Average-case and planted versions of polynomial optimization over the sphere have also been a central testbed for SoS and related spectral or Kikuchi-type algorithms [BGL17; HSSS16; WEM25]. For the especially important $2 \rightarrow 4$ norm, a particular class of degree-4 polynomials related but not identical to BSS, algorithmic guarantees are known for several families of instances [BBHKSZ12; BH15; HT26].

Global Correlation Rounding “Global correlation rounding” refers to the iterated conditioning-based rounding scheme we use here. It was developed in hierarchy algorithms for constraint satisfaction problems and related SoS rounding frameworks [BRS11; RT12; BKS14; GS12]. Since then, variants of the same idea have been used in scheduling via LP hierarchies [LR21], mean-field and free-energy approximation for Ising models [JKR19], and recent continuous optimization problems such as entrywise low-rank approximation and matrix $p \rightarrow q$ norms [ABH26]. Our use follows this line at the level of the rounding primitive, while the main new ingredient is the fine-grained pinning analysis tailored to the BSS SoS relaxation.

QMA[2] and Best Separable State The literature on QMA[2] is too broad to survey completely here; we highlight the seminal work [HM13] which made the connection between product-state testing, QMA[2], and polynomial optimization, proving $\text{QMA}[k] = \text{QMA}[2]$ for constant k and, assuming ETH, an $n^{\Omega(\log n)}$ lower bound for constant-gap BSS [HM13]. The closest prior work to ours is the perfect-completeness SoS algorithm of Barak, Kothari, and Steurer [BKS17]; their algorithm follows a different rounding strategy and analysis than ours.

Special Cases of BSS Several special cases of BSS admit stronger guarantees than the worst-case problem. When M is the projector onto a subspace, BSS is the problem of finding the closest rank-one matrix to that subspace; this also captures the perfect-completeness setting. If the subspace is generic, we know stronger guarantees: Johnston, Lovitz, and Vijayaraghavan give a polynomial-time algorithm for this case [JLV23]. Another special case is if M is entrywise nonnegative—then Barak, Kelner, and Steurer give a quasipolynomial-time SoS algorithm [BKS14].

Simultaneous Work Simultaneous and independent work by Jeronimo, Wu, and Xu [JWX26a] also obtains an algorithm for 1 versus $1 - \varepsilon$ BSS using $O(\sqrt{n/\varepsilon})$ degree SoS, with rounding

techniques differing substantially from ours. Another simultaneous and independent work of the same authors [JWX26b] shows that $\text{BSS}(M)$ can be solved up to any constant gap with imperfect completeness in time $\exp(\tilde{O}(\sqrt{n}))$.

2 Technical Overview and Warmup

In this section we describe a simple algorithm and analysis which achieves the guarantees Theorem 1.2 up to a $\sqrt{\log n}$ factor in the exponent. We emphasise that these results already improve on the prior known guarantees of [BKS17].

Theorem 2.1 (Warmup result). *Suppose $M \in \mathbb{R}^{n^2 \times n^2}$ has $\text{BSS}(M) = 1$. Then, for every $0 < \varepsilon < 1$, in time $n^{O(\sqrt{n/\varepsilon \cdot \log n})}$ we can find unit vectors $x, y \in \mathbb{R}^n$ such that $\langle (x \otimes y), M(x \otimes y) \rangle \geq 1 - \varepsilon$, and for every $1 \leq q \leq n/2$, in time $n^{O(\sqrt{q \log n})}$ we can find unit $x, y \in \mathbb{R}^n$ such that $\langle (x \otimes y), M(x \otimes y) \rangle \geq q/n$.*

At the end of this section we briefly describe how the $\sqrt{\log n}$ factor in the exponent can be removed.

The convex relaxation Let Π be the projector to the span of eigenvectors having eigenvalue 1 in M . Since $\text{BSS}(M) = 1$, there are nonzero x, y such that $\Pi(x \otimes y) = x \otimes y$. We will use an SoS relaxation of the following system of polynomial equations. (See [BS16] for background on SoS relaxations.)

$$\mathcal{A} = \left\{ \begin{array}{l} \|x\|_2^2 = 1, \quad \|y\|_2^2 = 1, \\ \Pi(x \otimes y) = x \otimes y \end{array} \right\}.$$

By picking a (not too large, $\text{poly}(n)$ -size) discrete set Σ of $\mathbb{R}$, incurring negligible $1/\text{poly}(n)$ -size errors due to discretization, we pass to a discrete assignment problem: find vectors $x, y \in \Sigma^n$ with $\|x\|^2, \|y\|^2 = 1 \pm 1/\text{poly}(n)$ such that $\|\Pi(x \otimes y)\|^2 \geq 1 - 1/\text{poly}(n)$. (For the rest of this overview we ignore this negligible $1/\text{poly}(n)$ error.) This discrete assignment version is a technically convenient departure from the standard SoS relaxation of $\mathcal{A}$ (used by [BKS17]), because it allows the *conditioning* operation described below.

Our algorithm first solves the degree $t = O(\sqrt{n/\varepsilon \cdot \log n})$ or $t = O(\sqrt{q \log n})$ SoS relaxation of this discrete assignment problem, which can be done in time $|\Sigma|^{O(t)} = n^{O(t)}$.

SoS background – pseudodistributions, local distributions, and conditioning We defer formal background to Section 3 and discuss here only what is needed for the technical overview.

A solution to this relaxation is a *degree- t pseudodistribution*. It is often helpful to think of a pseudodistribution as if it were a probability distribution over assignments to (x, y) . We will need two key ways in which pseudodistributions act like such distributions.

First, a degree- t pseudodistribution contains, for every set $S \subseteq [n]$ of size at most t , a probability distribution, called a *local distribution*, over assignments in $\Sigma^{|S|}$ to the variables in S . These local distributions are consistent on their intersections, in that the local distributions on S, S' must have the same marginal on $S \cap S'$.

Second, we can *condition* a pseudodistribution on the value of any single coordinate. For any variable x_i (or y_i), we can draw a sample $\hat{x}_i$ from the local distribution on $\{i\}$ and replace the local distribution on S with the local distribution on $S \cup \{i\}$, conditioned on the event that $x_i = \hat{x}_i$. This yields a degree $t - 1$ pseudodistribution.

Finally, some notation: for a function $f(x, y)$ depending on at most t variables, we write $\tilde{\mathbb{E}} f(x, y)$ for the expected value of f under the appropriate local distribution; we extend the operator $\tilde{\mathbb{E}}$ linearly to all f which can be written as linear combinations of such local functions. Thus, for example, $\tilde{\mathbb{E}} x \in \mathbb{R}^n$ is the vector whose entries are $\mathbb{E} x_i$ with expectation taken under the $\{i\}$ -th local distribution.

2.1 1 versus $1 - \varepsilon$

We describe and analyze a simple rounding procedure which takes a degree $O(\sqrt{n/\varepsilon \cdot \log n})$ pseudodistribution and outputs vectors $\hat{x}, \hat{y}$ such that $\|\Pi(\hat{x} \otimes \hat{y})\|_2 \geq (1 - \varepsilon) \cdot \|\hat{x} \otimes \hat{y}\|_2$. In the next section we describe how to adapt the analysis to the 1 versus q/n setting.

The rounding procedure is simple: enumerate over all possible conditionings of $O(\sqrt{n/\varepsilon \cdot \log n})$ coordinates, for each conditioning compute the conditional means $\tilde{\mathbb{E}} x, \tilde{\mathbb{E}} y$, and output the best solution found in this manner. The step where the algorithm outputs $\tilde{\mathbb{E}} x, \tilde{\mathbb{E}} y$ succeeds if any of the conditioned pseudoexpectations satisfy the following *small covariance* condition:

Lemma 2.2 (Small Covariance Success Condition (Informal)). *Let $\tilde{\mathbb{E}}$ be a pseudodistribution satisfying $\mathcal{A}$. If*

$$\|\widetilde{\mathbf{Cov}}(x, y)\|_F \leq \sqrt{\varepsilon} \cdot \|\tilde{\mathbb{E}} x\| \|\tilde{\mathbb{E}} y\| \quad (1)$$

then

$$\|\Pi((\tilde{\mathbb{E}} x) \otimes (\tilde{\mathbb{E}} y))\|_2 \geq (1 - \varepsilon) \cdot \|(\tilde{\mathbb{E}} x) \otimes (\tilde{\mathbb{E}} y)\|.$$

Here $\widetilde{\mathbf{Cov}}(x, y) = \tilde{\mathbb{E}} xy^\top - (\tilde{\mathbb{E}} x)(\tilde{\mathbb{E}} y)^\top$ and $\|\cdot\|_F$ is the Frobenius norm. The proof of this lemma is straightforward and follows similar ideas to [BKS17]; the main idea is that the upper bound on $\|\widetilde{\mathbf{Cov}}\|_F$ ensures that $\tilde{\mathbb{E}} x \otimes y$ is close to $(\tilde{\mathbb{E}} x) \otimes (\tilde{\mathbb{E}} y)$, and that $\tilde{\mathbb{E}} x \otimes y$, by virtue of the constraints of the SoS relaxation, must lie in the subspace Π projects to. (This last point is where $\text{BSS}(M) = 1$ is crucial; even if $\text{BSS}(M) = 1 - \varepsilon/2$, $\tilde{\mathbb{E}} x \otimes y$ might actually be far from the subspace.) Together, these imply $(\tilde{\mathbb{E}} x)(\tilde{\mathbb{E}} y)^\top$ must lie close to that subspace. We defer the formal proof to Section 4 and instead focus on the meat of the argument: that the “small covariance success condition” (1) holds after conditioning on the values of at most $O(\sqrt{n/\varepsilon \cdot \log n})$ coordinates.

A new pinning lemma The technical heart of our result is a *pinning lemma*, which shows that we can establish the condition in (1) via conditioning on (“pinning”) $O(\sqrt{n/\varepsilon \cdot \log n})$ coordinates.

Lemma 2.3 (Warmup pinning). *Let $0 < \varepsilon \leq 1$, and let (x, y) be jointly-distributed random vectors supported on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$ with finite support. There exists a set $S \subseteq [2n]$ of at most $|S| \leq O(\sqrt{n/\varepsilon \log n})$ coordinates and values $(\hat{x}, \hat{y})_S \in \mathbb{R}^{|S|}$ for those coordinates such that $(\hat{x}, \hat{y})_S$ is in the support of (x, y) restricted to S , and such that*

$$\|\mathbf{Cov}(x, y) \mid (x, y)_S = (\hat{x}, \hat{y})_S\|_F \leq \sqrt{\varepsilon} \cdot \|\mathbb{E} x \mid (x, y)_S = (\hat{x}, \hat{y})_S\| \cdot \|\mathbb{E} y \mid (x, y)_S = (\hat{x}, \hat{y})_S\|.$$

The lemma also applies to pseudodistributions of degree at least $O(\sqrt{n/\varepsilon \cdot \log n})$.

We now outline the proof of Lemma 2.3. We describe a simple algorithm which chooses coordinates $i \in [2n]$ and values $\hat{x}_i$ or $\hat{y}_i$ iteratively, at each step adding one coordinate to the conditioned set S . This proceeds in two phases: after the first phase, both $\|\mathbb{E} x\|^2$ and $\|\mathbb{E} y\|^2$ will be at least $\sqrt{\log n / (\varepsilon n)}$. In the second phase will use these “floors” on the distances of the means from the origin to establish (1).

First phase For the first phase, observe that via greedily choosing coordinates i and values $\hat{x}_i$, we can ensure that the contribution to $\|\mathbb{E} x\|^2$ from pinned coordinates is at least r/n by pinning r coordinates; the same holds for y . Illustrating with the case $r = 1$, since $\sum_{i \leq n} \mathbb{E} x_i^2 = 1$, there is some i such that $\mathbb{E} x_i^2 \geq 1/n$, and hence some value $\hat{x}_i$ in the support of x_i such that $\hat{x}_i^2 \geq 1/n$; we can pin $x_i = \hat{x}_i$ for this choice of $\hat{x}_i$. The argument for general $r > 0$ can be proved by induction. Importantly for what follows, the contribution of pinned coordinates to $\|\tilde{\mathbb{E}} x\|^2$ cannot decrease under subsequent pinning operations, so we can apply this greedy pinning operation to x and then y in sequence.

Second phase We turn to the second phase. We show that if $\tilde{\mathbb{E}}$ has the property that pinned coordinates contribute at least $\sqrt{\log n / (\varepsilon n)}$ to both $\|\tilde{\mathbb{E}} x\|^2, \|\tilde{\mathbb{E}} y\|^2$, then after pinning at most $O(\sqrt{n \log n / \varepsilon})$ additional coordinates, (1) holds. We show this by analyzing the potential function

$$\Phi(\tilde{\mathbb{E}}) = \|\tilde{\mathbb{E}} x\|^2 + \|\tilde{\mathbb{E}} y\|^2.$$

We will show that whenever (1) does not hold that we can “make progress” on the goal via increasing $\Phi(\tilde{\mathbb{E}})$ by conditioning on a coordinate. Note that $0 \leq \Phi(\tilde{\mathbb{E}}) \leq 2$ via our norm constraints on x, y and thus $\Phi(\tilde{\mathbb{E}})$ cannot increase indefinitely. At this level the argument is inspired by the analysis of *global correlation rounding* for CSPs from [BRS11], but the “success condition” we are targeting is significantly stronger than what is needed in the CSP setting.

The main idea is that if the covariance $\|\mathbf{Cov}(x, y)\|_F$ is large, then there is some coordinate x_i such that conditioning on x_i significantly reduces the variance of y , which in turn increases $\|\tilde{\mathbb{E}} y\|^2$ and hence increases $\Phi(\tilde{\mathbb{E}})$ (or, respectively, a coordinate y_i which we can condition on to reduce the variance of x). The quantitative relationship between covariance of two random variables a, b and the reduction in variance in a given by conditioning on b is captured by the following elementary inequality [BRS11]:

$$\mathbf{Var}(a) - \mathbb{E}_b \mathbf{Var}(a|b) \geq \Omega \left(\frac{\mathbf{Cov}^2(a, b)}{\mathbf{Var}(b)} \right). \quad (2)$$

Since x is supported on unit vectors, $\|\mathbb{E} x\|_2^2 = 1 - \text{tr}(\mathbf{Cov}(x))$. Thus, if we condition on a fixed x_j (with value drawn from its marginal distribution) and let the conditioned pseudodistribution be $\tilde{\mathbb{E}}'$, then we have that

$$\begin{aligned} \Phi(\tilde{\mathbb{E}}') - \Phi(\tilde{\mathbb{E}}) &= \sum_{i \in [n]} \left(\mathbf{Var}(x_i) - \mathbb{E}_{x_j} \mathbf{Var}(x_i|x_j) \right) + \sum_{i \in [n]} \left(\mathbf{Var}(y_i) - \mathbb{E}_{x_j} \mathbf{Var}(y_i|x_j) \right) \\ &\geq \frac{\sum_{i \in [n]} \mathbf{Cov}^2(x_i, x_j) + \sum_{i \in [n]} \mathbf{Cov}^2(y_i, x_j)}{\mathbf{Var}(x_j)}. \end{aligned}$$

The same holds for y_j by symmetry, and thus if we pick a random x_j or y_j with probabilities proportional to $\mathbf{Var}(x_j), \mathbf{Var}(y_j)$, via an averaging argument we have

$$\mathbb{E} [\Phi(\tilde{\mathbb{E}}') - \Phi(\tilde{\mathbb{E}})] \geq \frac{\|\mathbf{Cov}(x, x)\|_F^2 + 2\|\mathbf{Cov}(x, y)\|_F^2 + \|\mathbf{Cov}(y, y)\|_F^2}{\text{tr}(\mathbf{Cov}(x)) + \text{tr}(\mathbf{Cov}(y))}. \quad (3)$$

Thus there exists some coordinate (and choice of pinning for that coordinate) such that that gain in the potential is at least $\|\mathbf{Cov}(x, y)\|_F^2$, using non-negativity of norms and the fact that the denominator is at most 2. Now, if (1) fails for $\tilde{\mathbb{E}}$, we have that for this choice of pinning,

$$\Phi(\tilde{\mathbb{E}}') - \Phi(\tilde{\mathbb{E}}) \geq \varepsilon \cdot \|\tilde{\mathbb{E}} x\|_2^2 \cdot \|\tilde{\mathbb{E}} y\|_2^2.$$

We now aim to relate this lower bound to the current value of the potential $\Phi(\tilde{\mathbb{E}})$. Using the lower bound we established in the first phase on $\|\tilde{\mathbb{E}} x\|^2, \|\tilde{\mathbb{E}} y\|^2$, we have that

$$\varepsilon \cdot \|\tilde{\mathbb{E}} x\|_2^2 \|\tilde{\mathbb{E}} y\|_2^2 \geq \frac{1}{2} \cdot \sqrt{\frac{\varepsilon \log n}{n}} \cdot (\|\tilde{\mathbb{E}} x\|_2^2 + \|\tilde{\mathbb{E}} y\|_2^2) = \frac{1}{2} \cdot \sqrt{\frac{\varepsilon \log n}{n}} \cdot \Phi(\tilde{\mathbb{E}}).$$

Thus, whenever (1) fails, we can increase the potential function by a *multiplicative* factor of $1 + \Omega(\sqrt{(\varepsilon \log n)/n}) \approx \exp(\Omega(\sqrt{(\varepsilon \log n)/n}))$. At the start of the second phase, $\Phi(\tilde{\mathbb{E}}) \geq \sqrt{1/(\varepsilon n)}$, and it cannot ever become larger than 2. So after at most $\sqrt{n \log n / \varepsilon}$ pinning operation, this second phase must encounter $\tilde{\mathbb{E}}$ which satisfies (1).

2.2 1 versus q/n

Next we describe how to adapt the foregoing analysis to round $\tilde{\mathbb{E}}$ to $\hat{x}, \hat{y}$ satisfying $\|\Pi(\hat{x} \otimes \hat{y})\|^2 \geq (q/n) \cdot \|\hat{x} \otimes \hat{y}\|^2$ using $O(\sqrt{q \log n})$ pinnings.

Rounding via singular vectors In this regime, rather than using the means $\tilde{\mathbb{E}} x, \tilde{\mathbb{E}} y$, after pinning we output the top left and right singular vectors of $R = \tilde{\mathbb{E}} xy^\top$. The constraints imply $\tilde{\mathbb{E}}(x \otimes y)$ lies in the range of Π . This gives the following analogue of Lemma 2.2.

Lemma 2.4 (Singular-Vector Rounding Success Condition (Informal)). *Let $\tilde{\mathbb{E}}$ be a pseudodistribution satisfying $\mathcal{A}$, and suppose $R = \tilde{\mathbb{E}} xy^\top$ is nonzero. If u, v are top left and right singular vectors of R , then*

$$\|\Pi(u \otimes v)\|_2^2 \geq \frac{\sigma_1(R)^2}{\|R\|_F^2}.$$

Proof. Since $\tilde{\mathbb{E}} x \otimes y$ lies in the range of Π ,

$$\sigma_1(R) = \langle u \otimes v, \text{vec}(R) \rangle = \langle \Pi(u \otimes v), \text{vec}(R) \rangle \leq \|\Pi(u \otimes v)\|_2 \|R\|_F.$$

Squaring proves the lemma. ■

A low-soundness pinning lemma The analogue of Lemma 2.3 is the following.

Lemma 2.5 (Warmup low-soundness pinning). *Let $1 \leq q \leq n/2$, and let (x, y) be jointly-distributed random vectors supported on $\mathbb{S}^{n-1} \times \mathbb{S}^{n-1}$ with finite support. There is a set $S \subseteq [2n]$ of at most $O(\sqrt{q \log n})$ coordinates and values $(\hat{x}, \hat{y})_S$ in the support of $(x, y)_S$ such that, under the resulting conditional distribution,*

$$\|\mathbb{E}[xy^\top \mid (x, y)_S = (\hat{x}, \hat{y})_S]\|_F \leq \sqrt{\frac{n}{q}} \cdot \sigma_1(\mathbb{E}[xy^\top \mid (x, y)_S = (\hat{x}, \hat{y})_S])$$

The lemma also applies to pseudodistributions of degree at least $O(\sqrt{q \log n})$.

Proof sketch. As in the 1 versus $1 - \varepsilon$ case, the argument has two phases. The first is the same as before: pin $O(\sqrt{q \log n})$ coordinates of each of x and y so that pinned coordinates contribute at least $\sqrt{q \log n}/n$ to each of $\|\tilde{\mathbb{E}} x\|^2$ and $\|\tilde{\mathbb{E}} y\|^2$.

Suppose that for some conditioned $\tilde{\mathbb{E}}$, Lemma 2.4 fails, i.e., $\sigma_1(\tilde{\mathbb{E}} xy^\top)^2 \leq \frac{q}{n} \|\tilde{\mathbb{E}} xy^\top\|_F^2$. We will show that some further pinning increases the same potential function $\Phi(\tilde{\mathbb{E}}) = \|\tilde{\mathbb{E}} x\|_2^2 + \|\tilde{\mathbb{E}} y\|_2^2$ as before.

The rows of $\tilde{\mathbb{E}} xy^\top$ corresponding to the pinned coordinates of x equal the vector of pinned x values times $(\tilde{\mathbb{E}} y)^\top$. Since the pinned values contribute at least $\sqrt{q \log n}/n$ to $\|\tilde{\mathbb{E}} x\|_2^2$,

$$\sigma_1(\tilde{\mathbb{E}} xy^\top)^2 \geq \frac{\sqrt{q \log n}}{n} \|\tilde{\mathbb{E}} y\|_2^2,$$

by looking just at the submatrix of rows corresponding to pinned x coordinates. Symmetrically, $\sigma_1(\tilde{\mathbb{E}} xy^\top)^2 \geq \frac{\sqrt{q \log n}}{n} \|\tilde{\mathbb{E}} x\|_2^2$. Taking the average,

$$\sigma_1(\tilde{\mathbb{E}} xy^\top)^2 \geq \frac{\sqrt{q \log n}}{2n} (\|\tilde{\mathbb{E}} x\|_2^2 + \|\tilde{\mathbb{E}} y\|_2^2) = \frac{\sqrt{q \log n}}{2n} \Phi(\tilde{\mathbb{E}}).$$

Failure of singular-vector rounding now gives

$$\|\tilde{\mathbb{E}} xy^\top\|_F^2 \geq \frac{n}{q} \sigma_1(\tilde{\mathbb{E}} xy^\top)^2 \geq \frac{1}{2} \sqrt{\frac{\log n}{q}} \Phi(\tilde{\mathbb{E}}).$$

It remains to turn this lower bound into a lower bound on $\|\mathbf{Cov}(x, y)\|_F^2$, after which we can apply the same strategy as in the 1 versus $1 - \varepsilon$ case (see (2)). Since $(\tilde{\mathbb{E}} x)(\tilde{\mathbb{E}} y)^\top$ has rank one, subtracting it from $\tilde{\mathbb{E}} xy^\top$ reduces the Frobenius norm by no more than subtracting the top singular vectors would:

$$\|\mathbf{Cov}(x, y)\|_F^2 = \|\tilde{\mathbb{E}} xy^\top - (\tilde{\mathbb{E}} x)(\tilde{\mathbb{E}} y)^\top\|_F^2 \geq \|\tilde{\mathbb{E}} xy^\top\|_F^2 - \sigma_1(\tilde{\mathbb{E}} xy^\top)^2 \geq \left(1 - \frac{q}{n}\right) \|\tilde{\mathbb{E}} xy^\top\|_F^2,$$

where the last step again uses that singular-vector rounding fails so $\sigma_1(\tilde{\mathbb{E}} xy^\top)^2 \leq (q/n) \|\tilde{\mathbb{E}} xy^\top\|_F^2$. Putting this together with our lower bound on $\|\tilde{\mathbb{E}} xy^\top\|_F^2$,

$$\|\mathbf{Cov}(x, y)\|_F^2 \geq \frac{1}{4} \sqrt{\frac{\log n}{q}} \Phi(\tilde{\mathbb{E}}).$$

The same variance-decrease calculation as in the 1 versus $1 - \varepsilon$ case now shows that some coordinate and some value for that coordinate increase Φ by at least $\|\mathbf{Cov}(x, y)\|_F^2$. Therefore, whenever singular-vector rounding fails, some pinning satisfies

$$\Phi(\tilde{\mathbb{E}}') \geq \left(1 + \Omega\left(\sqrt{\frac{\log n}{q}}\right)\right) \Phi(\tilde{\mathbb{E}}).$$

This multiplicative growth can therefore occur for at most

$$O\left(\sqrt{\frac{q}{\log n}} \log \frac{n}{\sqrt{q \log n}}\right) \leq O(\sqrt{q \log n})$$

additional pinnings in the second phase. Together with the first phase, this proves Lemma 2.5. ■

2.3 Removing the $\sqrt{\log n}$ factors

We briefly outline how the $\sqrt{\log n}$ factors are removed.

1 versus $1 - \varepsilon$ case It turns out that Lemma 2.3 can be directly improved to remove the $\sqrt{\log n}$ factor, which leads immediately to an algorithm with running time $n^{O(\sqrt{n}/\varepsilon)}$. To obtain this improvement, we use a new potential function $\Phi'(\tilde{\mathbb{E}}) = \|\tilde{\mathbb{E}} x\|^2 + \|\tilde{\mathbb{E}} y\|^2 + (\|\tilde{\mathbb{E}} x\| + \|\tilde{\mathbb{E}} y\|)^2$. The basic intuition behind this potential function is to improve on the roughly $(1 + \min(\|\tilde{\mathbb{E}} x\|^2, \|\tilde{\mathbb{E}} y\|^2))$ rate of multiplicative growth in the potential function Φ we used previously. Note that if $\|\tilde{\mathbb{E}} x\|^2 \approx \|\tilde{\mathbb{E}} y\|^2$, then Φ actually increases much faster than the analysis we gave suggests. Otherwise, roughly speaking, if $\|\tilde{\mathbb{E}} x\|^2 \ll \|\tilde{\mathbb{E}} y\|^2$ (or vice versa), the term $(\|\tilde{\mathbb{E}} x\| + \|\tilde{\mathbb{E}} y\|)^2$ rewards upward movement in $\min(\|\tilde{\mathbb{E}} x\|^2, \|\tilde{\mathbb{E}} y\|^2)$. The details are in Section 4.2.

1 versus q/n case Removing the $\sqrt{\log n}$ factor in the 1 versus q/n case again uses the potential function $\Phi'(\tilde{\mathbb{E}})$ from the 1 versus $1 - \varepsilon$ case, with two additional twists. First, rather than always output the top singular vectors of $\tilde{\mathbb{E}} xy^\top$, two additional possible outputs are considered – one is to output the pair $(\tilde{\mathbb{E}} x) / \|\tilde{\mathbb{E}} x\|, v$, where v is the maximum eigenvector of the quadratic form $v \mapsto \langle (x \otimes v), \Pi(x \otimes v) \rangle$, and the other the same with x and y swapped. The second twist is that in addition to the coordinate-pinning operation, the algorithm uses a reweighing operation on the pseudodistribution, replacing $\tilde{\mathbb{E}} p(x, y)$ with $\tilde{\mathbb{E}} f(x, y)^2 p(x, y) / \tilde{\mathbb{E}} f(x, y)^2$ for a well-chosen degree-2 function f .

2.4 Fine-grained pinning lemma

We conclude this technical overview with a brief discussion of pinning lemmas, and state one new pinning lemma which goes slightly beyond Lemmas 2.3 and 2.5. This section is not essential to the proofs of our main results (and can be safely skipped on first reading), but since pinning lemmas play a crucial role in rounding algorithms for SoS, as well as in statistical physics and high-dimensional probability, the new pinning lemma may be of independent interest.

Background on pinning lemmas Pinning lemmas are measure decomposition theorems which show that high-dimensional probability distributions can be written as a mixture of “few” distributions, most or some of which are “simple,” with various definitions of “few” and “simple,” by conditioning on, or “pinning”, a small number of coordinates.

The original pinning lemma has been widely used across algorithms [RT12; BKS17; ABH26; GS12; BRS11], high-dimensional probability, and statistical physics [Mon08; CKPZ17; JKR19]. It says that every distribution becomes approximately pairwise independent after conditioning on a small number of coordinates. For simplicity in this discussion, we stick to mean-zero distributions on the unit sphere.

Theorem 2.6 (Pinning Lemma (variance version), [BRS11]). *For any distribution μ on the n -dimensional unit sphere and every $\varepsilon > 0$ there is a subset of at most $t \leq O(1/\varepsilon)$ coordinates $S \subseteq [n]$ such that $\mathbb{E}_{x_S \sim \mu_S} \|\mathbf{Cov}(\mu | x_S)\|_F^2 \leq \varepsilon$.*

Here, μ_S denotes μ restricted to the coordinates in S , $\mu | x_S$ denotes the conditional distribution of μ after fixing the coordinates in S to the values in x_S , and $\|\cdot\|_F$ is the Frobenius norm, so that $\|\mathbf{Cov}(\mu | x_S)\|_F^2 = \sum_{i,j \leq n} \mathbf{Cov}(x_i, x_j | x_S)^2$ gives the average squared covariance across pairs of coordinates after conditioning. Notice that the maximum possible value of $\|\mathbf{Cov}\|_F^2$ for a distribution on the unit sphere is 1. If the coordinates of μ are supported in a discrete set of size q (e.g. $q = 2$ for the hypercube), then at most $q^{O(1/\varepsilon)}$ different distributions can appear in the

decomposition $\mu = \mathbb{E}_{x_S \sim \mu_S}(\mu \mid x_S)$; these conditional distributions give us the interpretation of the pinning lemma as decomposing μ into a mixture of not-too-many distributions.

A fine-grained pinning lemma Like Theorem 2.6, Lemmas 2.3 and 2.5 also show that conditioning on the values of a few coordinates of a high-dimensional probability distribution results in small covariance. They differ from Theorem 2.6 in two important ways:

1. They measure “simplicity” of a distribution not just in terms of $\|\mathbf{Cov}\|_F^2$, but in terms of the ratio between $\|\mathbf{Cov}\|_F^2$ and $\|\mathbb{E} x\|^2, \|\mathbb{E} y\|^2$, meaning that a distribution does not count as “simple” unless its covariance is small *compared to its mean*, which is crucial for the application to BSS.
2. To obtain this stronger notion of simplicity, they make two important sacrifices compared to Theorem 2.6:
 - They require conditioning on many more coordinates, and
 - They don’t offer a true measure decomposition, because the *values* for the coordinates that they condition on are chosen greedily rather than being sampled from the marginal distributions of those coordinates.

While having a measure decomposition as opposed to just some set of values to condition on isn’t important for our applications to BSS, this property has been very important in other uses of pinning-style lemmas, especially in the theory of localization schemes [CE25]. This motivates the question of how much of the guarantees of Lemmas 2.3 and 2.5 can be obtained if we insist on a measure decomposition?

It turns out that via more or less the same techniques as we use to prove Lemmas 2.3 and 2.5, one can prove the following fine-grained pinning lemma, which does provide a measure decomposition. The expense, compared to Lemmas 2.3 and 2.5, is that it applies only to a single random vector x rather than a pair (x, y) – this is the reason we cannot use it to prove our main results about BSS.

Theorem 2.7 (Fine-grained pinning lemma, see Section 6). *For every mean-zero distribution μ on the n -dimensional unit sphere and every $0 < \varepsilon < 1$, there is a set of (S, x_S) where $S \subseteq [n]$ and $x_S \in \mathbb{R}^{|S|}$ with $|S| \leq O\left(\frac{\sqrt{n}}{\varepsilon}\right)$, such that $\mathbb{E}_{(S, x_S)}(\mu \mid x_S) = \mu$, and*

$$\mathbb{E}_{(S, x_S)} \|\mathbf{Cov}(\mu \mid x_S)\|_F \leq O(\varepsilon) \cdot \mathbb{E}_{(S, x_S)} \left\| \mathbb{E}_{x \sim \mu \mid x_S} x \right\|^2.$$

See Section 6 for the variant of Theorem 2.7 in the $\varepsilon > 1$ case.

The statement $\mathbb{E}_{(S, x_S)}(\mu \mid x_S) = \mu$ captures the “measure decomposition” aspect of Theorem 2.7. It captures that, in the proof of Theorem 2.7, (S, x_S) is obtained by iteratively choosing one coordinate i at a time, sampling a value x_i from the marginal distribution on that coordinate, and conditioning on that value.

To understand Theorem 2.7, it helps to unpack the different ways in which the condition in Theorem 2.6 can be satisfied. At a high level, the Frobenius norm of the conditional covariance, $\|\mathbf{Cov}(\mu \mid x_S)\|_F$, could be small for two different reasons. One is that most coordinates have become nearly fixed around their means and hence have small variance, and consequently the mean of the conditioned distribution is noticeably far from the origin. The other is that a typical pair of coordinates has become nearly independent, but each individual coordinate still has mean close to zero and retains large variance. Theorem 2.7 tells us that, at least in an

in-expectation sense, the latter does not always happen, $\|\mathbb{E}_{x \sim \mu | x_S} x\|^2$ is lower-bounded, so the average per-coordinate variance cannot be too large. In other words, in expectation the mean drifts noticeably far from the origin.

The cost of this refined guarantee, however, is that we must condition on many more coordinates than in Theorem 2.6, as we must condition on a set of size $\sqrt{n}/\varepsilon$ rather than $1/\varepsilon$. This factor is tight, as can be shown by the simple example of the uniform distribution on $\left\{\pm \frac{1}{\sqrt{n}}\right\}^n$.

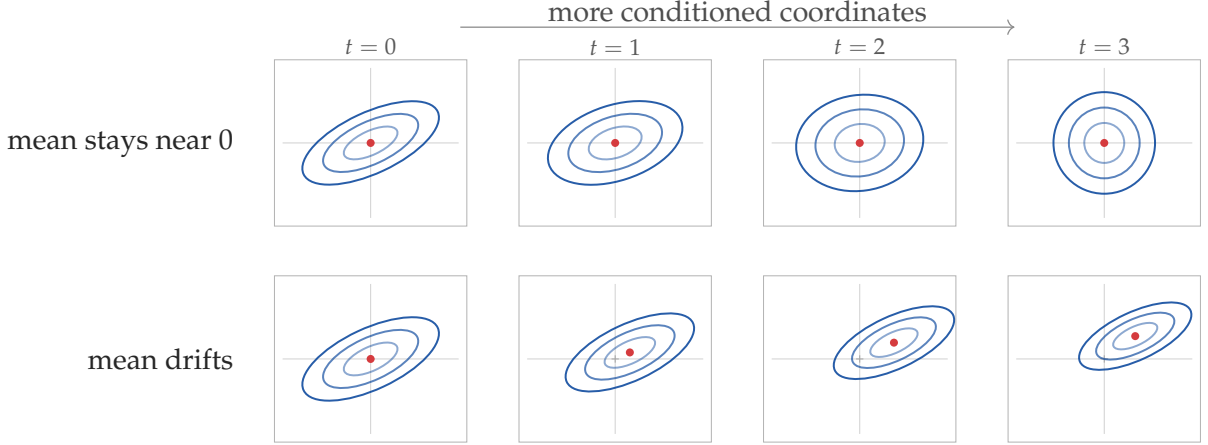


Figure 1: Two ways that $\|\mathbf{Cov}\|_F$ can decrease as we condition on coordinates in a high-dimensional distribution. In the top row, correlations disappear while the conditional mean remains near the origin – the distribution becomes isotropic but retains significant variance. In the bottom row, the overall variance decreases while the mean moves away from the origin. Our fine-grained pinning lemma shows that the second behavior must happen for a noticeable fraction of conditional distributions.

3 Preliminaries

We now provide an overview of the sum-of-squares proof system, taken from the corresponding section in [ABMT25]. We closely follow the exposition as it appears in the lecture notes of Barak and Steurer [BS16].

Pseudo-Distributions. A discrete probability distribution over $\mathbb{R}^m$ is defined by its probability mass function, $D: \mathbb{R}^m \rightarrow \mathbb{R}$, which must satisfy $\sum_{x \in \text{supp}(D)} D(x) = 1$ and $D \geq 0$. We extend this definition by relaxing the non-negativity constraint to merely requiring that D passes certain low-degree non-negativity tests. We call the resulting object a pseudo-distribution.

Definition 3.1 (Pseudo-distribution). A *degree- ℓ pseudo-distribution* is a finitely-supported function $D: \mathbb{R}^m \rightarrow \mathbb{R}$ such that $\sum_x D(x) = 1$ and $\sum_x D(x) p(x)^2 \geq 0$ for every polynomial p of degree at most $\ell/2$, where the summation is over all x in the support of D .

Next, we define the related notion of pseudo-expectation.

Definition 3.2 (Pseudo-expectation). The *pseudo-expectation* of a function f on $\mathbb{R}^m$ with respect to a pseudo-distribution μ , denoted by $\tilde{\mathbb{E}}_{\mu(x)}[f(x)]$, is defined as

$$\tilde{\mathbb{E}}_{\mu(x)}[f(x)] = \sum_x \mu(x) f(x).$$

We use the notation $\tilde{\mathbb{E}}_{\mu(x)} [(1, x_1, x_2, \dots, x_m)^{\otimes \ell}]$ to denote the degree- ℓ moment tensor of the pseudo-distribution μ . In particular, each entry in the moment tensor corresponds to the pseudo-expectation of a monomial of degree at most ℓ in x .

Definition 3.3 (Constrained pseudo-distributions). Let $\mathcal{A} = \{p_1 \geq 0, p_2 \geq 0, \dots, p_r \geq 0\}$ be a system of r polynomial inequality constraints of degree at most d in m variables. Let μ be a degree- ℓ pseudo-distribution over $\mathbb{R}^m$. We say that μ *satisfies* $\mathcal{A}$ at degree $\ell \geq 1$ if for every subset $\mathcal{S} \subset [r]$ and every sum-of-squares polynomial q such that $\deg(q) + \sum_{i \in \mathcal{S}} \max(\deg(p_i), d) \leq \ell$, $\tilde{\mathbb{E}}_{\mu} [q \prod_{i \in \mathcal{S}} p_i] \geq 0$. Further, we say that μ *approximately satisfies* the system of constraints $\mathcal{A}$ if the above inequalities are satisfied up to additive error $\tilde{\mathbb{E}}_{\mu} [q \prod_{i \in \mathcal{S}} p_i] \geq -2^{-n^\ell} \|q\| \prod_{i \in \mathcal{S}} \|p_i\|$, where $\|\cdot\|$ denotes the Euclidean norm of the coefficients of the polynomial, represented in the monomial basis.

Crucially, there's an efficient separation oracle for moment tensors of constrained pseudo-distributions. Below gives the unconstrained statement; the constraint statement follows analogously.

Fact 3.4 ([Sho87; Nes00; Par00; Gri01]). *For any $m, \ell \in \mathbb{N}$, the following convex set has a $m^{\mathcal{O}(\ell)}$ -time weak separation oracle, in the sense of [GLS81]⁴:*

$$\left\{ \tilde{\mathbb{E}}_{\mu(x)} [(1, x_1, x_2, \dots, x_m)^{\otimes \ell}] \mid \mu \text{ is a degree-}\ell \text{ pseudo-distribution over } \mathbb{R}^m \right\}$$

This fact, together with the equivalence of weak separation and optimization [GLS81] forms the basis of the sum-of-squares algorithm, as it allows us to efficiently approximately optimize over pseudo-distributions.

Given a system of polynomial constraints, denoted by $\mathcal{A}$, we say that it is *explicitly bounded* if it contains a constraint of the form $\{\|x\|^2 \leq 1\}$. Then, the following fact follows from Fact 3.4 and [GLS81]:

Theorem 3.5 (Efficient optimization over pseudo-distributions). *There exists an $(m + r)^{\mathcal{O}(\ell)}$ -time algorithm that, given any explicitly bounded and satisfiable system $\mathcal{A}$ of r polynomial constraints in m variables, outputs a degree- ℓ pseudo-distribution that satisfies $\mathcal{A}$ approximately, in the sense of Definition 3.3.⁵*

Conditioning and Reweighting. Finally, we will also require the following *conditioning* operation on pseudo-distributions.

Definition 3.6 (Pseudo-distribution reweighting). Let $\tilde{\mathbb{E}}$ be a pseudo-distribution of degree d and let f be a sum-of-squares polynomial of degree $k \leq d$, with $\tilde{\mathbb{E}}[f] > 0$. Then, $\tilde{\mathbb{E}}'$, the reweighting of $\tilde{\mathbb{E}}$ with f , is defined by

$$\tilde{\mathbb{E}}' g = \frac{\tilde{\mathbb{E}}[fg]}{\tilde{\mathbb{E}}[f]}$$

and is a degree $d - k$ pseudo-distribution. Furthermore, $\tilde{\mathbb{E}}'$ satisfies all constraints that $\tilde{\mathbb{E}}$ satisfies of degree at most $d - k$.

Note that if x is an indicator variable and $f = x^2$, then reweighting by f corresponds to conditioning on $x = 1$.

⁴A separation oracle of a convex set $S \subset \mathbb{R}^M$ is an algorithm that can decide whether a vector $v \in \mathbb{R}^M$ is in the set, and if not, provide a hyperplane between v and S . Roughly, a weak separation oracle is a separation oracle that allows for some η slack in this decision.

⁵Here, we assume that the bit complexity of the constraints in $\mathcal{A}$ is $m^{\mathcal{O}(1)}$.

4 Best Separable State: 1 vs $1 - \varepsilon$

In this section we prove the 1 vs $1 - \varepsilon$ regime of our main result. In particular, we prove the following theorem:

Theorem 4.1. *Let $0 < \varepsilon < 1$ and let $\Pi \in \mathbb{R}^{n^2 \times n^2}$ be a projector to a subspace of $\mathbb{R}^{n^2}$. Then there exists an algorithm that runs in time*

$$n^{O(\sqrt{\frac{n}{\varepsilon}})}$$

and distinguishes between the following cases:

- **YES:** *There exists a nonzero $x, y \in \mathbb{R}^n$ such that $\Pi(x \otimes y) = x \otimes y$.*
- **NO:** *For all $x, y \in \mathbb{R}^n$ we have that $\|\Pi(x \otimes y)\|_2^2 \leq (1 - \varepsilon) \cdot \|x \otimes y\|_2^2$.*

Furthermore, in the YES case, the algorithm produces $\hat{x}, \hat{y}$ such that $\|\Pi(\hat{x} \otimes \hat{y})\|_2^2 > (1 - \varepsilon) \cdot \|\hat{x} \otimes \hat{y}\|_2^2$.

To prove Theorem 4.1, we use the constraint set

$$\mathcal{A} = \left\{ \begin{array}{l} \|x\|_2^2 = 1, \quad \|y\|_2^2 = 1, \\ \Pi(x \otimes y) = x \otimes y \end{array} \right\}.$$

Since we will work over a discretized domain it will be potentially impossible to satisfy the constraints above, even in the YES case. Thus, we instead replace $\mathcal{A}$ with a version which allows $1/\text{poly}(n)$ slack:

$$\mathcal{A}'(\beta) = \left\{ \begin{array}{l} 1 - \frac{1}{n^\beta} \leq \|x\|_2^2 \leq 1 + \frac{1}{n^\beta}, \\ 1 - \frac{1}{n^\beta} \leq \|y\|_2^2 \leq 1 + \frac{1}{n^\beta}, \\ \|(I - \Pi)(x \otimes y)\|_2^2 \leq \frac{1}{n^{2\beta}} \end{array} \right\}. \quad (4)$$

Algorithm 4.2 (Best Separable State: 1 vs $1 - \varepsilon$).

Input: A projector $\Pi \in \mathbb{R}^{n^2 \times n^2}$ and $0 < \varepsilon < 1$.

Operations:

1. If $\varepsilon < 1/n$, search over all pairs u, v in an $(\varepsilon/20)$ -net of $\mathbb{S}^{n-1}$ and output YES iff $\|\Pi(u \otimes v)\|_2^2 > 1 - \varepsilon$ for some pair.
2. Otherwise, let $\gamma = \sqrt{\varepsilon}/4$ and choose $\beta = 4$.
3. Let Σ be a grid of $[-1, 1]$ of mesh size a sufficiently small constant multiple of $n^{-1/2-\beta}$.
4. Let $T = \lceil C\sqrt{n}/\gamma \rceil$ for a sufficiently large absolute constant C . Check feasibility of the degree- $(2T + 8)$ sum-of-squares relaxation over Σ satisfying $\mathcal{A}'(\beta)$.
5. If the program is feasible, output YES. Otherwise, output NO.

In order to show that Algorithm 4.2 is correct, we show that whenever the program is feasible we can round the relaxation to $\hat{x}, \hat{y}$ such that $\|\Pi(\hat{x} \otimes \hat{y})\|_2 > (1 - \varepsilon)\|\hat{x} \otimes \hat{y}\|_2$. This

means that in the NO case, the program cannot be feasible. The rounding algorithm used is a variant of global correlation rounding. We record it below as simply brute forcing over all possible choices of conditionings and indices to condition on; this is because the precise choice of coordinates to condition on and how to sample their values is not straightforward to describe. We defer further discussion of how to show the existence of a good set of coordinates and values to the proof of Lemma 4.5.

Algorithm 4.3 (Best Separable State Rounding).

Input: A degree- $O(T)$ pseudodistribution satisfying $\mathcal{A}'(\beta)$.

Operations: For each set $S_x \subseteq [n]$ and $S_y \subseteq [n]$ of indices of size at most T and pinnings $\hat{x}_{S_x}$ and $\hat{y}_{S_y}$:

1. Let $\tilde{\mathbb{E}}$ be the pseudodistribution obtained by conditioning $\tilde{\mathbb{E}}$ on the event that $x_{S_x} = \hat{x}_{S_x}$ and $y_{S_y} = \hat{y}_{S_y}$.
2. Let $m = \tilde{\mathbb{E}}_r x$ and $q = \tilde{\mathbb{E}}_r y$ at the resulting branch.

Output: $m/\|m\|_2$ and $q/\|q\|_2$ with the highest objective value among all iterations of the loop.

4.1 The rounding success condition

We now describe a sufficient condition for the final mean rounding step to succeed in Algorithm 4.3.

Lemma 4.4 (Mean rounding). *Let $\tilde{\mathbb{E}}$ satisfy $\mathcal{A}'(\beta)$, and suppose that*

$$\|\mathbf{Cov}(x, y)\|_F \leq \gamma \|\tilde{\mathbb{E}} x\|_2 \|\tilde{\mathbb{E}} y\|_2.$$

Write $m = \tilde{\mathbb{E}} x$ and $q = \tilde{\mathbb{E}} y$. Then

$$\|\Pi(m \otimes q)\|_2^2 \geq \left(1 - \left(\gamma + \frac{n^{-\beta}}{\|m\|_2 \|q\|_2}\right)^2\right) \|m \otimes q\|_2^2.$$

Proof. Let $M = \tilde{\mathbb{E}}(x \otimes y)$. By the definition of cross-covariance,

$$M = m \otimes q + \mathbf{Cov}(x, y).$$

Moreover, Jensen's inequality and the residual constraint give

$$\|(I - \Pi)M\|_2 \leq \left(\tilde{\mathbb{E}}\|(I - \Pi)(x \otimes y)\|_2^2\right)^{1/2} \leq n^{-\beta}.$$

Therefore

$$\begin{aligned} \|(I - \Pi)(m \otimes q)\|_2 &\leq \|(I - \Pi)M\|_2 + \|\mathbf{Cov}(x, y)\|_F \\ &\leq n^{-\beta} + \gamma \|m\|_2 \|q\|_2. \end{aligned}$$

Since Π is an orthogonal projector,

$$\|\Pi(m \otimes q)\|_2^2 = \|m \otimes q\|_2^2 - \|(I - \Pi)(m \otimes q)\|_2^2.$$

Substituting the previous bound proves the lemma. ■

4.2 A new pinning lemma

We now establish that we can achieve the desired “win condition” via $O(\sqrt{n/\varepsilon})$ rounds of conditioning.

Lemma 4.5. *Let $0 < \gamma \leq 1$, and let $\tilde{\mathbb{E}}$ be a pseudodistribution of degree $\Omega\left(\frac{\sqrt{n}}{\gamma}\right)$ on random variables (x, y) , satisfying the constraints in (4). Then, there is a pinning of $O\left(\frac{\sqrt{n}}{\gamma}\right)$ coordinates such that denoting by $\tilde{\mathbb{E}}'$ the conditional pseudodistribution obtained from these pinnings*

$$(a) \quad \left\| \widetilde{\mathbf{Cov}}'(x, y) \right\|_F \leq \gamma \cdot \|\tilde{\mathbb{E}}' x\| \cdot \|\tilde{\mathbb{E}}' y\|.$$

$$(b) \quad \|\tilde{\mathbb{E}}' x\| \text{ and } \|\tilde{\mathbb{E}}' y\| \text{ are bounded from below by } \frac{\gamma}{\sqrt{n}}.$$

Remark 4.6. The more important of the two conditions is the first one. The second condition is needed only to control the error from our $1/\text{poly}(n)$ slack constraints. It turns out that it is also naturally established in the process of bound $\|\widetilde{\mathbf{Cov}}'(x, y)\|_F$.

Lemma 4.5 will follow from two lemmas. The first is easy, and establishes that it is possible to pin a small number of coordinates in order to ensure the required lower bounds on the norms of the mean.

Lemma 4.7. *Let $0 < Q \ll 1$. There is a set $I_x \subseteq [n]$ and a pinning a_j of the coordinates x_j for $j \in I_x$ such that $\sum a_j^2 \geq Q$, and $|I_x| = O(nQ)$. The same claim holds for y .*

Proof. Suppose $Q < \frac{1}{4}$, say. We grow the set I_x one coordinate at a time. While $\sum a_j^2 \leq Q$, we have $\sum_{i \notin I_x} \tilde{\mathbb{E}} x_i^2 \geq \frac{1}{2}$, so there exists a choice of coordinate $i \notin I_x$ with $\tilde{\mathbb{E}} x_i^2 \geq \frac{1}{2n}$, and thus a pinning a_i occurring with nonzero probability such that $a_i^2 \geq \frac{1}{2n}$. Repeating this at most $2Qn$ times completes the proof. ■

The second establishes the more interesting claim. Define the potential function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$\Phi(u, v) = 2 (\|u\|^2 + \|v\|^2 + \|u\| \cdot \|v\|) ,$$

and denote by C the cross-covariance

$$C = \tilde{\mathbb{E}} \left((x - \tilde{\mathbb{E}} x) (y - \tilde{\mathbb{E}} y) \right)^\top .$$

Lemma 4.8. *Let $\tilde{\mathbb{E}}$ be a pseudodistribution on variables (x, y) satisfying the constraints (4). Further suppose that all coordinates in some sets of indices $I_x, I_y \subseteq [n]$ have already been pinned such that they contribute squared ℓ_2 mass of at least Q . Let m_x and m_y be the means of this pseudodistribution. Also suppose that*

$$\|C\|_F > \gamma \cdot \|m_x\| \cdot \|m_y\| .$$

Then, there exists a pinning such that if $\tilde{m}_x$ and $\tilde{m}_y$ are the updated means after this pinning, then

$$\Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) \gtrsim \gamma^2 Q^{2/3} \cdot \Phi(m_x, m_y)^{4/3} .$$

We defer a discussion on the choice of potential function and the above lemma to Section 4.3.

Proof of Lemma 4.5. Begin by pinning coordinates of both x and y according to Lemma 4.7, for some Q to be decided. Let m_x^0 and m_y^0 be the means at this step. Following this, while the desideratum $\|C_F\| \leq \gamma \|m_x\| \|m_y\|$ is not satisfied, progressively pin coordinates according to Lemma 4.8, and let m_x^t and m_y^t be the means after t steps of this pinning. We shall show that this can happen for at most $O\left(\frac{1}{Q\gamma^2}\right)$ steps. Set $\Phi_t = \Phi(m_x^t, m_y^t)$. By the potential growth lemma, we have for some constant c that

$$\Phi_{t+1} \geq \Phi_t + c\gamma^2 Q^{2/3} \cdot \Phi_t^{4/3},$$

so

$$\begin{aligned} \Phi_{t+1}^{-1/3} &\leq \Phi_t^{-1/3} \cdot \left(1 + c\gamma^2 Q^{2/3} \cdot \Phi_t^{1/3}\right)^{-1/3} \\ &\leq \Phi_t^{-1/3} - c\gamma^2 Q^{2/3} \\ \Phi_t^{-1/3} - \Phi_{t+1}^{-1/3} &\geq c\gamma^2 Q^{2/3}. \end{aligned}$$

Note now that $\Phi_0 \gtrsim Q$, by the lower bound on the norm of m_x^0 . That is, $\Phi_0^{-1/3} \lesssim Q^{-1/3}$. Since Φ is non-negative, this implies that this can proceed for at most $O\left(\frac{Q^{-1/3}}{\gamma^2 Q^{2/3}}\right) = O\left(\frac{1}{\gamma^2 Q}\right)$ many steps.

Now, we performed $O(Qn)$ rounds of conditioning in the first phases, and $O\left(\frac{1}{\gamma^2 Q}\right)$ many steps of conditioning in the second. Balancing the two by choosing $Q \asymp \frac{1}{\gamma\sqrt{n}}$ completes the proof. $\blacksquare$

4.3 A potential growth lemma

The primary result of this section is the following.

Lemma 4.8. *Let $\tilde{\mathbb{E}}$ be a pseudodistribution on variables (x, y) satisfying the constraints (4). Further suppose that all coordinates in some sets of indices $I_x, I_y \subseteq [n]$ have already been pinned such that they contribute squared ℓ_2 mass of at least Q . Let m_x and m_y be the means of this pseudodistribution. Also suppose that*

$$\|C\|_F > \gamma \cdot \|m_x\| \cdot \|m_y\|.$$

Then, there exists a pinning such that if $\tilde{m}_x$ and $\tilde{m}_y$ are the updated means after this pinning, then

$$\Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) \gtrsim \gamma^2 Q^{2/3} \cdot \Phi(m_x, m_y)^{4/3}.$$

The structure of Φ has two components that behave in different ways: the $\|u\|^2 + \|v\|^2$ part results in growth when some coordinate has large variance, and as we will see shortly, the mixed term rewards movement *orthogonal* to the existing mean. In fact, the only “orthogonal movement” we will use will be the part that is orthogonal to coordinates $I_x \cup I_y$ that have already been pinned.

In particular, the first part, $\|u\|^2 + \|v\|^2$, suffices to establish a potential growth of order $\gamma^2 Q \Phi$ —this will give a bound on the degree of the required SDP of $O\left(\sqrt{\frac{n}{\epsilon}} \cdot \log n\right)$. In order to remove the extraneous log factors here and improve this to $O\left(\sqrt{\frac{n}{\epsilon}}\right)$, we will require the better potential growth above.

Let us start by noting some properties of Φ . The first is a consequence of its convexity.

Lemma 4.9. *Let x and y be (possibly correlated) mean-0 random variables on $\mathbb{R}^n$. Then,*

$$\mathbb{E} \Phi(u + x, v + y) - \Phi(u, v) \geq \mathbb{E} [\|x\|^2 + \|y\|^2].$$

Proof. Note that $(u, v) \mapsto (\|u\| + \|v\|)^2$ is convex. We thus have

$$\begin{aligned}\mathbb{E} \Phi(u + \mathbf{x}, v + \mathbf{y}) &= \mathbb{E} \left[(\|u + \mathbf{x}\| + \|v + \mathbf{y}\|)^2 + \|u + \mathbf{x}\|^2 + \|v + \mathbf{y}\|^2 \right] \\ &\geq (\|u\| + \|v\|)^2 + \mathbb{E} [\|u + \mathbf{x}\|^2 + \|v + \mathbf{y}\|^2] \\ &= \Phi(u, v) + \mathbb{E} [\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2].\end{aligned}\quad \blacksquare$$

The second establishes that if one of the means has grown large, and the other moves upon conditioning in some direction orthogonal to its (before-conditioning) value, then one may obtain stronger growth bounds.

Lemma 4.10. *Let $\mathbf{x}$ and $\mathbf{y}$ be (possibly correlated) mean-0 random variables on $\mathbb{R}^n$, and suppose that almost surely, $\|\Pi_{u^\perp} \mathbf{x}\|_2^2 \geq \theta \cdot \|\mathbf{x}\|^2$. Then,*

$$\mathbb{E} \Phi(u + \mathbf{x}, v + \mathbf{y}) - \Phi(u, v) \geq \theta \|v\| \cdot \mathbb{E} \left[\frac{\|\mathbf{x}\|^2}{\|u + \mathbf{x}\|} \right].$$

In particular, if there is some collection I of indices such that (i) we almost surely have $\mathbf{x}_I = \mathbf{y}_I = 0$, and (ii) $\|u_I\|^2 \geq Q$, then we may take $\theta = \frac{Q}{\|u\|^2}$.

Proof. We have

$$\begin{aligned}\mathbb{E} \Phi(u + \mathbf{x}, v + \mathbf{y}) - \Phi(u, v) &\geq \mathbb{E} \left[(\|u + \mathbf{x}\| + \|v + \mathbf{y}\|)^2 \right] - (\|u\| + \|v\|)^2 \\ &\geq 2(\|u\| + \|v\|) \cdot \mathbb{E} (\|u + \mathbf{x}\| - \|u\| + \|v + \mathbf{y}\| - \|v\|) \\ &\geq 2\|v\| \cdot \mathbb{E} (\|u + \mathbf{x}\| - \|u\|),\end{aligned}$$

where the final inequality uses the convexity of $v \mapsto \|v\|$. Continuing,

$$\begin{aligned}\mathbb{E} (\|u + \mathbf{x}\| - \|u\|) &= \mathbb{E} \left(\|u + \mathbf{x}\| - \left\langle \frac{u}{\|u\|}, u + \mathbf{x} \right\rangle \right) \\ &= \mathbb{E} \frac{\|u + \mathbf{x}\|^2 - \left\langle \frac{u}{\|u\|}, u + \mathbf{x} \right\rangle^2}{\|u + \mathbf{x}\| + \left\langle \frac{u}{\|u\|}, u + \mathbf{x} \right\rangle} \\ &\geq \frac{1}{2} \cdot \mathbb{E} \frac{\|u + \mathbf{x}\|^2 - \left\langle \frac{u}{\|u\|}, u + \mathbf{x} \right\rangle^2}{\|u + \mathbf{x}\|} \\ &= \frac{1}{2} \cdot \mathbb{E} \frac{\|\Pi_{u^\perp} \mathbf{x}\|^2}{\|u + \mathbf{x}\|} \\ &\geq \frac{\theta}{2} \cdot \mathbb{E} \frac{\|\mathbf{x}\|^2}{\|u + \mathbf{x}\|},\end{aligned}$$

as desired. For the second part of the lemma, the assumption on I implies that

$$\|\Pi_{u^\perp} \mathbf{x}\|^2 = \|\mathbf{x}\|^2 - \left\langle \frac{u}{\|u\|}, \mathbf{x} \right\rangle^2 \geq \left(1 - \frac{\|u_{I^c}\|^2}{\|u\|^2} \right) \cdot \|\mathbf{x}\|^2 \geq \frac{Q}{\|u\|^2} \cdot \|\mathbf{x}\|^2,$$

completing the proof. $\blacksquare$

For the remainder of this section, let us assume without loss of generality that $\|m_x\| \leq \|m_y\|$. Our pinning will pin a coordinate of y . For a given pinning $y_j \leftarrow a$, denote $m_x^{j \leftarrow a}$ (resp. $m_y^{j \leftarrow a}$) as the new x -mean (resp. y -mean) after this pinning. Let us also denote

$$\Delta = \max_{j,a} \left\{ \Phi \left(m_x^{j \leftarrow a}, m_y^{j \leftarrow a} \right) - \Phi(m_x, m_y) \right\}.$$

Let us also explicitly write the elements of the cross-covariance C . Its j th column is given by

$$C_j = \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} \left[a \left(m_x^{j \leftarrow a} - m_x \right) \right].$$

The following lemma controls various quantities related to these vectors.

Lemma 4.11. *We have*

$$\left\| \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \left(m_x^{j \leftarrow a} - m_x \right) \right\| = O(n^{-\beta}) \quad (5)$$

and

$$\sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \left\| \left(m_x^{j \leftarrow a} - m_x \right) \right\|^2 \lesssim \Delta + O(n^{-\beta}). \quad (6)$$

The above will be a consequence of the following (slightly mysterious) measure decomposition when the distributions of x and y are supported exactly on the sphere $\mathbb{S}^{n-1}$. Pick a coordinate $j \in [n]$ with probability $\mathbb{E} y_j^2$, and then pin it to a with probability $\frac{\Pr[y_j=a] \cdot a^2}{\mathbb{E} y_j^2}$. Observe that this does *not* lie in the standard canon of pinning algorithms of the form “pick a coordinate according to some distribution, then pin the value of that coordinate according to the associated marginal.”

Let us verify that this is indeed a measure decomposition, before providing intuition. We have

$$\begin{aligned} \sum_{j,a} \mathbb{E} [f(x,y) \mid y_j = a] \cdot \frac{\Pr[y_j = a] \cdot a^2}{\mathbb{E} y_j^2} \cdot \mathbb{E} [y_j^2] &= \sum_{j,a} \mathbb{E} [f(x,y) y_j^2 \mid y_j = a] \cdot \Pr[y_j = a] \\ &= \sum_j \mathbb{E} [f(x,y) y_j^2] \\ &= \mathbb{E} [f(x,y)] \end{aligned} \quad (7)$$

The above is more clearly seen to be a measure decomposition when interpreted in the following way. Sample y , pick a coordinate $j \sim p_y$ (in this case, $p_y(j) = y_j^2$, but this remains a measure decomposition even otherwise), and pin $y_j = y_j$. Indeed, this can be viewed as the “first step” of the measure decomposition which eventually pins all the coordinates of y . When $p_y(j) = y_j^2$, both the marginal probability of j (that is, $\mathbb{E} y_j^2$) and the marginal on a conditioned on the coordinate j being picked (that is, with probability $\frac{\Pr[y_j=a] \cdot a^2}{\mathbb{E} y_j^2}$) have particularly simple interpretations. We now move to the proof of the above lemma.

Proof. By essentially the manipulations in (7), picking $f(x,y) = x$,

$$\frac{\sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} \left[a^2 m_x^{j \leftarrow a} \right]}{\tilde{\mathbb{E}} \|y\|^2} = \frac{\tilde{\mathbb{E}} [x \|y\|^2]}{\tilde{\mathbb{E}} \|y\|^2}.$$

(4) establishes that $\tilde{\mathbb{E}} \models \|y\|^2 = 1 + O(n^{-\beta})$, which yields (5). Towards the second inequality,

$$\Delta \geq \frac{\sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \Phi \left(m_x^{j \leftarrow a}, m_y^{j \leftarrow a} \right)}{\tilde{\mathbb{E}} \|y\|^2} - \Phi(m_x, m_y).$$

Denote $m'_x = \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 m_x^{j \leftarrow a}$, and define m'_y similarly, so $\|m'_x - m_x\| = O(n^{-\beta})$. Consequently, $|\Phi(m'_x, m'_y) - \Phi(m_x, m_y)| = O(n^{-\beta})$. As a result, by Lemma 4.9,

$$\Delta \geq \frac{\sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \Phi \left(m_x^{j \leftarrow a}, m_y^{j \leftarrow a} \right)}{\tilde{\mathbb{E}} \|y\|^2} - \Phi(m'_x, m'_y) - O(n^{-\beta})$$

$$\begin{aligned}
&\geq \frac{\sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} \|m_x^{j \leftarrow a} - m_x\|^2}{\tilde{\mathbb{E}} \|y\|^2} - O(n^{-\beta}) \\
&= \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} \|m_x^{j \leftarrow a} - m_x\|^2 - O(n^{-\beta}). \quad \blacksquare
\end{aligned}$$

We will also use the same lower bound on Δ that was employed in the weaker $O(\sqrt{n \log n / \varepsilon})$ argument:

Lemma 4.12. *We have*

$$\|C\|_F^2 \leq \Delta + O(n^{-\beta}).$$

(For the proof, see the warmup in the technical overview, in particular (3).)

As mentioned, the above bound already yields non-trivial control on the growth of the potential—ignoring the $n^{-\beta}$ terms, it gives $\Delta \gtrsim \gamma^2 \|m_x\|^2 \|m_y\|^2 \gtrsim \gamma^2 Q \Phi(m_x, m_y)$. In our final proof, we will only use this to obtain an upper bound on how large the smaller mean $\|m_x\|$ can be.

With the above in hand, we may prove the main input to Lemma 4.8.

Lemma 4.13. *We have*

$$\|C\|_F^2 \leq \Delta \cdot \frac{\|m_x\|^2}{Q \|m_y\|} \cdot \sqrt{\Delta + \|m_x\|^2}$$

Among the terms within the square-root, the $\|m_x\|^2$ will be the relevant one.

Proof. First off, for any coordinate j ,

$$\tilde{\mathbb{E}}_{a \sim \mu_{y_j}} \left[\frac{\|m_x^{j \leftarrow a} - m_x\|^2}{\|m_x^{j \leftarrow a}\|} \right] \leq \Delta \cdot \frac{\|m_x\|^2}{Q \|m_y\|}. \quad (8)$$

Indeed, this follows immediately from Lemma 4.10 with the observation that $m_x^{j \leftarrow a} - m_x$ is a 0-mean random variable.

Let C_j be the j th column of C . As observed,

$$C_j = \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a \left(m_x^{j \leftarrow a} - m_x \right).$$

By the weighted Cauchy–Schwarz inequality,

$$\begin{aligned}
\|C_j\|^2 &\leq \left(\tilde{\mathbb{E}}_{a \sim \mu_{y_j}} \frac{\|m_x^{j \leftarrow a} - m_x\|^2}{\|m_x^{j \leftarrow a}\|} \right) \cdot \left(\tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \cdot \|m_x^{j \leftarrow a}\| \right) \\
&\leq \Delta \cdot \frac{\|m_x\|^2}{Q \|m_y\|} \cdot \left(\tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \cdot \|m_x^{j \leftarrow a}\| \right)
\end{aligned}$$

Therefore, summing over j ,

$$\begin{aligned}
\|C\|_F^2 &\leq \Delta \cdot \frac{\|m_x\|^2}{Q \|m_y\|} \cdot \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \cdot \|m_x^{j \leftarrow a}\| \\
&\leq \Delta \cdot \frac{\|m_x\|^2}{Q \|m_y\|} \cdot \left(\tilde{\mathbb{E}} \|y\|^2 \cdot \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \cdot \|m_x^{j \leftarrow a}\|^2 \right)^{1/2}
\end{aligned}$$

The constraints in (4) yield that $\tilde{\mathbb{E}} \|y\|^2 = 1 + O(n^{-\beta})$. The numerator of the term within the square-root may be rewritten as

$$\begin{aligned} \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \cdot \|m_x^{j \leftarrow a}\|^2 &= \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \cdot \left\| m_x + (m_x^{j \leftarrow a} - m_x) \right\|^2 \\ &= \left(\tilde{\mathbb{E}} \|y\|^2 \right) \cdot \|m_x\|^2 + 2 \left\langle m_x, \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 m_x^{j \leftarrow a} - m_x \right\rangle \\ &\quad + \sum_j \tilde{\mathbb{E}}_{a \sim \mu_{y_j}} a^2 \left\| m_x^{j \leftarrow a} - m_x \right\|^2 \end{aligned}$$

(5) in Lemma 4.11 bounds the second quantity by $O(n^{-\beta})$, and (6) bounds the third by $O(\Delta + n^{-\beta})$. Putting the pieces together, we thus have

$$\|C\|_F^2 \lesssim \Delta \cdot \frac{\|m_x\|^2}{Q\|m_y\|} \cdot \sqrt{\Delta + \|m_x\|^2}. \quad \blacksquare$$

Finally, we may prove Lemma 4.8.

Proof of Lemma 4.8. Lemma 4.13 coupled with our lower bound assumption on $\|C\|_F$ imply that

$$\Delta \cdot \frac{\|m_x\|^2}{Q\|m_y\|} \cdot \sqrt{\Delta + \|m_x\|^2} \gtrsim \gamma^2 \cdot \|m_x\|^2 \|m_y\|^2.$$

Rearranging,

$$\Delta \cdot \sqrt{\Delta + \|m_x\|^2} \gtrsim \gamma^2 \cdot Q \cdot \|m_y\|^3.$$

On the other hand, Lemma 4.12 yields that

$$\Delta \gtrsim \gamma^2 \cdot \|m_x\|^2 \|m_y\|^2.$$

Indeed, the error of $O(n^{-\beta})$ is negligible in comparison to the Frobenius norm of the covariance, since the failure of the win condition and the mean anchoring imply that $\|C\|_F^2 \gtrsim n^{-O(1)}$. As a result,

$$\|m_x\|^2 \lesssim \frac{\Delta}{\gamma^2 \|m_y\|^2}.$$

Substituting this into the previous string of inequalities,

$$\Delta \cdot \sqrt{\frac{\Delta}{\gamma^2 \|m_y\|^2}} \gtrsim \gamma^2 \cdot Q \cdot \|m_y\|^3,$$

and rearranging,

$$\Delta^{3/2} \geq \gamma^3 Q \cdot \|m_y\|^4 \gtrsim \gamma^3 Q \cdot \Phi(m_x, m_y)^2.$$

Raising both sides to the (2/3)rd power completes the proof. $\blacksquare$

4.4 Proof of the main theorem

We now combine the pinning lemma from the previous section and the rounding success condition to prove the main theorem.

Proof of Theorem 4.1. We first consider the case where $\varepsilon > 1/n$ (we will return to the case where $\varepsilon < 1/n$ later). First, we show that in the YES case the SDP relaxation must be feasible. Normalize a YES witness so that $\|x\|_2 = \|y\|_2 = 1$, and choose $\hat{x}, \hat{y} \in \Sigma^n$ with

$$\|x - \hat{x}\|_2, \|y - \hat{y}\|_2 \leq cn^{-\beta}$$

for a sufficiently small absolute constant c . Then

$$|\|\hat{x}\|_2^2 - 1|, |\|\hat{y}\|_2^2 - 1| \leq n^{-\beta}.$$

Moreover,

$$\|\hat{x} \otimes \hat{y} - x \otimes y\|_2 \leq \|(\hat{x} - x) \otimes \hat{y}\|_2 + \|x \otimes (\hat{y} - y)\|_2 = O(n^{-\beta}).$$

Since $(I - \Pi)(x \otimes y) = 0$, decreasing c if necessary gives

$$\|(I - \Pi)(\hat{x} \otimes \hat{y})\|_2 \leq n^{-\beta}.$$

Thus the point mass at $(\hat{x}, \hat{y})$ satisfies $\mathcal{A}'(\beta)$, and the relaxation is feasible in the YES case.

It remains to prove soundness. Suppose that the relaxation is feasible. Apply Lemma 4.5 to the pseudodistribution induced by the feasible solution. Thus, one of the branches enumerated by Algorithm 4.3 satisfies

$$\|\mathbf{Cov}_r(x, y)\|_F \leq \gamma \|\tilde{\mathbb{E}}_r x\|_2 \|\tilde{\mathbb{E}}_r y\|_2 \quad \text{and} \quad \|\tilde{\mathbb{E}}_r x\|_2, \|\tilde{\mathbb{E}}_r y\|_2 \geq \frac{\gamma}{\sqrt{n}}.$$

Set $m = \tilde{\mathbb{E}}_r x$ and $q = \tilde{\mathbb{E}}_r y$. The second conclusion of the pinning lemma gives

$$\frac{n^{-\beta}}{\|m\|_2 \|q\|_2} \leq \frac{n^{1-\beta}}{\gamma^2}.$$

Recall that $\beta = 4$, $\gamma = \sqrt{\varepsilon}/4$, and $\varepsilon \geq 1/n$. Therefore

$$\frac{n^{1-\beta}}{\gamma^2} = \frac{16}{n^3 \varepsilon} \leq \gamma$$

for all sufficiently large n . Applying Lemma 4.4, we conclude that

$$\|\Pi(m \otimes q)\|_2^2 \geq (1 - (2\gamma)^2) \|m \otimes q\|_2^2 = \left(1 - \frac{\varepsilon}{4}\right) \|m \otimes q\|_2^2 > (1 - \varepsilon) \|m \otimes q\|_2^2.$$

Thus the normalized mean pair violates the NO condition. Since the rounding algorithm tests every conditioning branch of the required length, it finds a pair with at least this objective value. Consequently, the relaxation cannot be feasible in the NO case.

In order to conclude the case where $\varepsilon > 1/n$, it remains to address the runtime. The grid has size $|\Sigma| = n^{O(1)}$, and

$$T = O\left(\frac{\sqrt{n}}{\gamma}\right) = O\left(\sqrt{\frac{n}{\varepsilon}}\right).$$

The degree- $O(T)$ sum-of-squares feasibility problem can therefore be solved in time

$$n^{O(T)} = n^{O(\sqrt{n/\varepsilon})}.$$

Enumerating the conditioning branches in Algorithm 4.3 has the same running time.

We now return to the case where $\varepsilon < 1/n$. In this case, the algorithm searches over all pairs u, v in an $(\varepsilon/20)$ -net of S^{n-1} . We first show completeness. Normalize a YES witness so that $\|x\|_2 = \|y\|_2 = 1$, and choose u, v in the net such that

$$\|x - u\|_2, \|y - v\|_2 \leq \frac{\varepsilon}{20}.$$

Then

$$\begin{aligned} \|u \otimes v - x \otimes y\|_2 &\leq \|(u - x) \otimes v\|_2 + \|x \otimes (v - y)\|_2 \\ &\leq \frac{\varepsilon}{10}. \end{aligned}$$

Since $\Pi(x \otimes y) = x \otimes y$ and Π is a contraction, it follows that

$$\|\Pi(u \otimes v)\|_2 \geq 1 - \frac{\varepsilon}{10}.$$

Thus

$$\|\Pi(u \otimes v)\|_2^2 \geq \left(1 - \frac{\varepsilon}{10}\right)^2 > 1 - \varepsilon,$$

so the algorithm outputs YES. On the other hand, in the NO case every pair of unit vectors u, v satisfies

$$\|\Pi(u \otimes v)\|_2^2 \leq 1 - \varepsilon,$$

and therefore the algorithm outputs NO.

Finally, an $(\varepsilon/20)$ -net of S^{n-1} has size $(O(1/\varepsilon))^n$. Hence, the search over all pairs in the net takes time

$$(O(1/\varepsilon))^{2n}.$$

Since $\varepsilon < 1/n$, we have

$$(O(1/\varepsilon))^{2n} = n^{O(\sqrt{n/\varepsilon})}.$$

This proves the theorem in the remaining regime and completes the proof. ■

5 Best Separable State: 1 vs q/n

Theorem 5.1. *Let $1 < q < n/2$ and let $\Pi \in \mathbb{R}^{n^2 \times n^2}$ be a projector to a subspace of $\mathbb{R}^{n^2}$. Then there exists an algorithm that runs in time*

$$n^{O(\sqrt{q})}$$

and distinguishes between the following cases:

- **YES:** *There exists a nonzero $x, y \in \mathbb{R}^n$ such that $\Pi(x \otimes y) = x \otimes y$.*
- **NO:** *For all $x, y \in \mathbb{R}^n$ we have that $\|\Pi(x \otimes y)\|_2^2 \leq \frac{q}{n} \cdot \|x \otimes y\|_2^2$.*

Furthermore, in the YES case, the algorithm produces $\hat{x}, \hat{y}$ such that $\|\Pi(\hat{x} \otimes \hat{y})\|_2^2 > \frac{q}{n} \cdot \|\hat{x} \otimes \hat{y}\|_2^2$.

As in the previous section, we consider the natural SoS relaxation of BSS. Since the rounding algorithm conditions on coordinate values, we work over a finite grid. Since over this grid it may not be possible to satisfy the canonical constraint set exactly, even in the YES case, we instead use the relaxed constraint set

$$\mathcal{A}'(\beta) = \left\{ \begin{aligned} &1 - n^{-\beta} \leq \|x\|_2^2 \leq 1 + n^{-\beta}, \quad 1 - n^{-\beta} \leq \|y\|_2^2 \leq 1 + n^{-\beta}, \\ &\|(I - \Pi)(x \otimes y)\|_2^2 \leq n^{-2\beta} \end{aligned} \right\}. \quad (9)$$

In our application below we take $\beta = 2$, and let Σ be a grid of $[-1, 1]$ whose mesh is a sufficiently small constant multiple of $n^{-1/2-\beta}$.

The decision algorithm checks feasibility of the degree- $C\sqrt{q}$ SoS relaxation satisfying (9), where C is a sufficiently large absolute constant. This is the same feasibility test as in the 1 versus $1 - \varepsilon$ regime; only the degree and the rounding argument differ.

We once again show that the relaxation must be infeasible in the NO case by providing a rounding algorithm. However, compared to the 1 vs $1 - \varepsilon$ regime, the final rounding step is slightly more involved. We use a combination of coordinate pinnings and reweights, but use multiple different final rounding steps. There are two key differences from the algorithm in Section 4: we require reweights in addition to coordinate pinnings, and we also consider multiple final rounding steps.

Algorithm 5.2 (Low-soundness BSS rounding).

Input: A degree- $C\sqrt{q}$ pseudodistribution satisfying $\mathcal{A}'(\beta)$.

Operations: For every branch obtained by performing at most $C\sqrt{q}$ operations, each of which is either a pinning or an affine-square reweighing as in Lemmas 5.6, 5.10 and 5.13, perform the following rounding steps.

1. Write $m = \tilde{\mathbb{E}} x$, $z = \tilde{\mathbb{E}} y$, and $R = \tilde{\mathbb{E}} xy^\top$.
2. If $m \neq 0$, set $u = m/\|m\|_2$ and let v be the “best response,” given by a top eigenvector of

$$(u \otimes I)^\top \Pi(u \otimes I).$$

Test the product vector $u \otimes v$.

3. If $z \neq 0$, set $v = z/\|z\|_2$ and let u be the “best response,” given by a top eigenvector of

$$(I \otimes v)^\top \Pi(I \otimes v).$$

Test the product vector $u \otimes v$.

4. If $R \neq 0$, compute top left and right singular vectors of R and test the resulting product vector.

Output: The pair with the largest value of $\|\Pi(u \otimes v)\|_2^2$ among all candidates.

5.1 The rounding success condition

We now describe sufficient conditions for at least one of the final rounding steps to succeed.

Lemma 5.3 (Low-soundness rounding). *Let $\tilde{\mathbb{E}}$ satisfy (9), and write*

$$m = \tilde{\mathbb{E}} x, \quad z = \tilde{\mathbb{E}} y, \quad R = \tilde{\mathbb{E}} xy^\top.$$

The two best-response candidates have scores at least

$$\|m\|_2^2 - O(n^{-\beta}) \quad \text{and} \quad \|z\|_2^2 - O(n^{-\beta}),$$

respectively. If $R \neq 0$, the top-singular-vector candidate has score at least

$$\left(\frac{\sigma_1(R) - n^{-\beta}}{\|R\|_F} \right)_+^2.$$

Proof. Suppose first that $m \neq 0$, let $u = m/\|m\|_2$, and set

$$T_u = (u \otimes I)^\top \Pi(u \otimes I).$$

Put $\rho = (I - \Pi)(x \otimes y)$. Using

$$\langle u, x \rangle \|y\|_2^2 = \langle \Pi(u \otimes y) \rangle x \otimes y + \langle u \otimes y \rangle \rho,$$

the norm constraints, the residual constraint, and Cauchy–Schwarz give

$$\tilde{\mathbb{E}} \langle u, x \rangle^2 \leq \tilde{\mathbb{E}} \|\Pi(u \otimes y)\|_2^2 + O(n^{-\beta}).$$

Writing $Y = \tilde{\mathbb{E}} yy^\top$, the first term on the right is $\text{tr}(T_u Y)$. Since $Y \succeq 0$ and $\text{tr}(Y) = 1 + O(n^{-\beta})$, it is at most $\lambda_{\max}(T_u) + O(n^{-\beta})$. Finally, $\|m\|_2^2 \leq \tilde{\mathbb{E}} \langle u, x \rangle^2$. The top eigenvector of T_u is a best response and hence has score at least $\|m\|_2^2 - O(n^{-\beta})$. The same argument with x and y exchanged proves the other best-response bound.

For the singular-vector candidate, let u, v be top left and right singular vectors of R . Jensen’s inequality and the residual constraint give

$$\|(I - \Pi)R\|_F \leq n^{-\beta}.$$

Identifying R with a vector in $\mathbb{R}^n \otimes \mathbb{R}^n$, we have

$$\begin{aligned} \sigma_1(R) &= \langle u \otimes v \rangle R \\ &\leq \langle \Pi(u \otimes v) \rangle R + n^{-\beta} \\ &\leq \|\Pi(u \otimes v)\|_2 \|R\|_F + n^{-\beta}. \end{aligned}$$

Rearranging and using nonnegativity proves the claimed bound. ■

5.2 A low-soundness reweighing lemma

We now prove the key technical ingredient, a “low-soundness” reweighing lemma which will establish one of the desired win conditions in Lemma 5.3.

For the purposes of clarity of exposition we will work with the exact constraint set instead of (4). It is easy to see that the $1/\text{poly}(n)$ error does not impact the correctness of the argument below.

Lemma 5.4. *Let $0 < \theta < c$ for some sufficiently small constant c , and let $\tilde{\mathbb{E}}$ be a pseudodistribution of degree $\Omega(\theta\sqrt{n})$ satisfying (9). Then, there exists a sequence of $O(\theta\sqrt{n})$ operations, each of which is either a pinning (in the sense of the conditioning on the value of some coordinate), or an affine reweigh (for some efficiently computable affine function $p = p(x, y)$ with $\tilde{\mathbb{E}} p^2 > 0$, it reweighs the distribution by p^2), such that the resulting pseudodistribution $\tilde{\mathbb{E}}'$ satisfies the following.*

(a) $\|\tilde{\mathbb{E}}' x\|$ and $\|\tilde{\mathbb{E}}' y\|$ are both bounded from below by $c \frac{\theta}{\sqrt{n}}$.

(b) $\|\tilde{\mathbb{E}}' xy^\top\|_F \gtrsim \frac{\theta}{\sqrt{n}}$.

(c) one of the following win conditions is satisfied:

- $\|\tilde{\mathbb{E}} x\| \gtrsim \theta$,
- $\|\tilde{\mathbb{E}} y\| \gtrsim \theta$, or
- denoting $R = \tilde{\mathbb{E}} xy^\top$, we have $\sigma_1(R) \gtrsim \theta \|R\|_F$.

As in the previous section, we will start by performing the mean anchoring of Lemma 4.7. Define the potential

$$\Phi(u, v) = 2 (\|u\|^2 + \|v\|^2 + \|u\| \cdot \|v\|) .$$

Our goal will be to show that if none of the win conditions above are satisfied, then we may reweigh the distribution to increase the potential $\Phi(m_x, m_y)$.

As we shall see shortly, if $\sigma_1(R) \gtrsim \|\tilde{\mathbb{E}} x\| \|\tilde{\mathbb{E}} y\|$, then we may pin a coordinate as in Lemma 4.8 to grow the potential. To deal with the other scenario, we shall use the following lemma.

Lemma 5.5. *Let $\tilde{\mathbb{E}}$ be a pseudodistribution on variables (x, y) satisfying the constraints (9). Further suppose that all coordinates in some sets of indices $I_x, I_y \subseteq [n]$ have already been pinned such that they contribute squared ℓ_2 mass of at least Q . Denoting $R = \tilde{\mathbb{E}} xy^\top$, suppose that*

$$\sigma_1(R) \leq \frac{1}{2} \cdot \|\tilde{\mathbb{E}} x\| \|\tilde{\mathbb{E}} y\| .$$

Then, there exists an affine function p such that on reweighing the pseudodistribution by p^2 , if $\tilde{m}_x$ and $\tilde{m}_y$ are the updated means, then

$$\Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) \gtrsim Q . \quad (10)$$

Let us put the pieces together and establish Lemma 5.4.

Proof of Lemma 5.4. Start by pinning coordinates as in Lemma 4.7, for some Q to be decided. Let the means of this distribution be m_x^0 and m_y^0 . As in Lemma 4.5, while none of the win conditions are satisfied, we shall sequentially reweigh/pin the distribution. Let m_x^t and m_y^t be the mean after t steps of this, and denote $\Phi_t = \Phi(m_x^t, m_y^t)$.

If

$$\sigma_1(R) \geq \frac{1}{2} \cdot \|\tilde{\mathbb{E}} x\| \cdot \|\tilde{\mathbb{E}} y\| ,$$

then the failure of the final win condition implies that

$$\|R_F\| \gtrsim \frac{1}{\theta} \|\tilde{\mathbb{E}} x\| \cdot \|\tilde{\mathbb{E}} y\| .$$

When θ is sufficiently small, we may relate this to the Frobenius norm of the cross-covariance as

$$\|C\|_F \geq \|R\|_F - \|\tilde{\mathbb{E}} x\| \cdot \|\tilde{\mathbb{E}} y\| \gtrsim \frac{1}{\theta} \|\tilde{\mathbb{E}} x\| \|\tilde{\mathbb{E}} y\| .$$

We then use Lemma 4.8 to grow Φ by $\Omega(\theta^{-2} Q^{2/3} \Phi^{4/3})$.

On the other hand, if

$$\sigma_1(R) \leq \frac{1}{2} \cdot \|\tilde{\mathbb{E}} x\| \cdot \|\tilde{\mathbb{E}} y\| ,$$

then use Lemma 5.5 so

$$\Phi_{t+1} \geq \Phi_t + \Omega(Q) .$$

To summarize, we have that while none of the win conditions are satisfied,

$$\Phi_{t+1} - \Phi_t \gtrsim \min \left\{ Q, \theta^{-2} Q^{2/3} \Phi_t^{4/3} \right\} .$$

Now, by the same proof as Lemma 4.5, the number of increments of the second type is at most $\frac{Q^{-1/3}}{\theta^{-2} Q^{2/3}} = O\left(\frac{\theta^2}{Q}\right)$. The first win condition additionally implies that $\Phi_t \lesssim \theta^2$ at all steps, so the number of increments of the first type is also $O\left(\frac{\theta^2}{Q}\right)$.

Thus, we have performed $O(Qn)$ many rounds of pinning in the first phase, and $O\left(\frac{\theta^2}{Q}\right)$ many rounds of pinning/reweighing in the second. Balancing the two terms by picking $Q \asymp \frac{\theta}{\sqrt{n}}$, the desideratum follows. $\blacksquare$

The two subsequent subsections are dedicated to establishing Lemma 5.5, and we encourage the reader to skip ahead to Section 5.6, where we use the above to prove our main theorem Theorem 5.1.

5.3 A low-soundness potential growth lemma

In this section, we prove the main technical input to the low-soundness result.

Lemma 5.5. *Let $\tilde{\mathbb{E}}$ be a pseudodistribution on variables (x, y) satisfying the constraints (9). Further suppose that all coordinates in some sets of indices $I_x, I_y \subseteq [n]$ have already been pinned such that they contribute squared ℓ_2 mass of at least Q . Denoting $R = \tilde{\mathbb{E}} xy^\top$, suppose that*

$$\sigma_1(R) \leq \frac{1}{2} \cdot \|\tilde{\mathbb{E}} x\| \|\tilde{\mathbb{E}} y\|.$$

Then, there exists an affine function p such that on reweighing the pseudodistribution by p^2 , if $\tilde{m}_x$ and $\tilde{m}_y$ are the updated means, then

$$\Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) \gtrsim Q. \quad (10)$$

Assume without loss of generality that $\|m_x\| \leq \|m_y\|$.

The plan will be as follows. We start by determining the reweighings that change the mean m_x as much as possible; this is the content of Lemma 5.6. However, we need to establish that growing the mean of x does not suddenly make the mean of y decay. This is where our assumption on the top singular value of R will enter the picture; this is the content of Lemma 5.7, and will establish either the norm of m_x or that of m_y must grow. Unfortunately, this plan will not work out exactly as stated, as there will be a (possibly non-negligible) error term in these calculations.

Define the normalized means $u = \frac{m_x}{\|m_x\|}$ and $v = \frac{m_y}{\|m_y\|}$. In the proof, we will bound

$$\Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) \geq \psi(\langle \tilde{m}_x, u \rangle, \langle \tilde{m}_y, v \rangle) - \psi(\|m_x\|, \|m_y\|),$$

for $\psi(w_1, w_2) := 2(w_1^2 + w_2^2 + |w_1 w_2|)$. That is, we only use the contribution to the potential growth coming from changes along the mean.

Let us also introduce the symbolic random variables $X = \langle x - \tilde{\mathbb{E}} x, u \rangle$, $Y = \langle y - \tilde{\mathbb{E}} y, v \rangle$, their variance $\sigma_x^2 = \widetilde{\text{Var}} X$ and $\sigma_y^2 = \widetilde{\text{Var}} Y$, as well as their normalized versions $\bar{X} = \frac{X}{\sigma_x}$ and $\bar{Y} = \frac{Y}{\sigma_y}$.

The affine function we construct will be of the form $a\bar{X} + b$. In the following lemma, we compute how the means change on reweighing by a function of this form.

Lemma 5.6. *Consider the 2×2 matrix*

$$H = \begin{pmatrix} \tilde{\mathbb{E}} \bar{X}^3 & 1 \\ 1 & 0 \end{pmatrix}.$$

Define $\lambda > 0$ by $\lambda - \frac{1}{\lambda} = \tilde{\mathbb{E}} \bar{X}^3$, so $\lambda_1 = \lambda$ and $\lambda_2 = -\frac{1}{\lambda}$ are the two eigenvalues of H . Then, there are reweighings $p_1 = a_1 \bar{X} + b_1$ and $p_2 = a_2 \bar{X} + b_2$, with coefficients equal to the corresponding

eigenvectors of H , such that if the updated means are m_x^1 and m_x^2 , then

$$\frac{\langle m_x^i - m_x, u \rangle}{\sigma_x} = \lambda_i.$$

Furthermore, under these reweighings,

$$\frac{\langle m_y^i - m_y, v \rangle}{\sigma_y} = \tilde{\mathbb{E}} [\bar{Y} \cdot \bar{X}] \cdot \lambda_i + \left(\frac{\lambda_i^2}{1 + \lambda_i^2} \right) \cdot \xi,$$

where

$$\xi = \tilde{\mathbb{E}} [\bar{Y} \cdot \bar{X}^2] - \tilde{\mathbb{E}} [\bar{Y} \cdot \bar{X}] \cdot \tilde{\mathbb{E}} [\bar{X}^2].$$

This quantity ξ , which we will analyse in much more detail shortly, represents the component of $\bar{Y}$ that is not predicted by $\bar{X}$ in a linear fashion.

Proof. This is a routine calculation. We reproduce the details here for completeness.

$$\begin{aligned} \frac{\langle \tilde{m}_x - m_x, u \rangle}{\sigma_x} &= \frac{\tilde{\mathbb{E}} \bar{X} (a\bar{X} + b)^2}{\tilde{\mathbb{E}} (a\bar{X} + b)^2} \\ &= \frac{a^2 \tilde{\mathbb{E}} \bar{X}^3 + 2ab}{a^2 + b^2}. \end{aligned} \quad (11)$$

It is easy to interpret the numerator as the quadratic form $\begin{pmatrix} a \\ b \end{pmatrix}^\top H \begin{pmatrix} a \\ b \end{pmatrix}$. We similarly have, for (a_i, b_i) being an eigenvector of the matrix with eigenvalue λ_i ,

$$\begin{aligned} \frac{\langle \tilde{m}_y - m_y, v \rangle}{\sigma_y} &= \frac{\tilde{\mathbb{E}} [a^2 \bar{Y} \bar{X}^2 + 2ab \bar{X} \bar{Y}]}{a^2 + b^2} \\ &= \tilde{\mathbb{E}} [\bar{X} \bar{Y}] \cdot \frac{a_i^2 \tilde{\mathbb{E}} \bar{X}^3 + 2a_i b_i}{a_i^2 + b_i^2} + \frac{a_i^2}{a_i^2 + b_i^2} \xi \\ &= \tilde{\mathbb{E}} [\bar{X} \bar{Y}] \cdot \lambda_i + \left(\frac{\lambda_i^2}{1 + \lambda_i^2} \right) \cdot \xi. \end{aligned}$$

■

The quantity $\tilde{\mathbb{E}} [\bar{Y} \cdot \bar{X}]$ arising above can easily be bounded from below.

Lemma 5.7. *Under the assumptions of Lemma 5.5,*

$$\tilde{\mathbb{E}} [XY] \leq -\frac{1}{2} \|m_x\| \cdot \|m_y\|.$$

Proof. By the definition of the singular value, the assumption in Lemma 5.5 implies that

$$u^\top R v \leq \frac{1}{2} \|m_x\| \cdot \|m_y\|,$$

so

$$\tilde{\mathbb{E}} [\langle x, u \rangle \cdot \langle y, v \rangle] \leq \frac{1}{2} \|m_x\| \cdot \|m_y\|.$$

Consequently,

$$\begin{aligned} \tilde{\mathbb{E}} [XY] &= \tilde{\mathbb{E}} [\langle x, u \rangle \cdot \langle y, v \rangle] - \|m_x\| \|m_y\| \\ &\leq -\frac{1}{2} \|m_x\| \|m_y\|. \end{aligned}$$

■

If $\xi = 0$ in Lemma 5.6, then we are done. Indeed, consider the random reweigh which chooses p_1 with probability $\frac{1}{1+\lambda^2}$ and p_2 with probability $\frac{\lambda^2}{1+\lambda^2}$ —this choice makes $\mathbb{E}\langle \tilde{m}_y - m_y, u \rangle = \mathbb{E}\langle \tilde{m}_x - m_x, u \rangle = 0$. We may then use a one-dimensional version of Lemma 4.9 to show that

$$\begin{aligned} \mathbb{E} \Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) &\geq \sigma_x^2 + \sigma_y^2 \cdot \tilde{\mathbb{E}}[\bar{Y} \cdot \bar{X}]^2 \\ &\geq \sigma_x^2 + \frac{\tilde{\mathbb{E}}[XY]^2}{\sigma_x^2} \\ &\gtrsim |\tilde{\mathbb{E}}[XY]| \\ &\gtrsim \|m_x\| \cdot \|m_y\| \gtrsim Q. \end{aligned} \quad (12)$$

Let us reintroduce the “residual” ξ , with the same random choice of reweigh from above, showing that if ξ is small, then the desired result follows.

Lemma 5.8. *Suppose that $|\xi| \ll \frac{\|m_x\|}{\sigma_y}$. Then, one of the affine reweighings in Lemma 5.6 satisfies (10).*

Proof. It is no longer true, as in the manipulations in (12), that $\mathbb{E}\langle \tilde{m}_y - m_y, v \rangle = 0$, but we nevertheless have above collection of equations with the extra error terms

$$\begin{aligned} \mathbb{E} \Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) &\geq \sigma_x^2 + \sigma_y^2 \cdot \tilde{\mathbb{E}}[\bar{X}\bar{Y}]^2 \\ &\quad + 2\xi \cdot \frac{\lambda^2}{(1+\lambda^2)^2} \left(2(2\|m_y\| + \|m_x\|)\sigma_y + \left(\lambda - \frac{1}{\lambda}\right) (\sigma_x\sigma_y + 2\tilde{\mathbb{E}}[\bar{X}\bar{Y}]\sigma_y^2) \right) \\ &\quad + 2\xi^2 \cdot \frac{\lambda^2}{(1+\lambda^2)^2} \cdot \sigma_y^2 \\ &\gtrsim \sigma_x^2 + \sigma_y^2 \cdot \tilde{\mathbb{E}}[\bar{X}\bar{Y}]^2 \end{aligned} \quad (13)$$

$$+ 2\xi \cdot \frac{\lambda^2}{(1+\lambda^2)^2} \left(2(2\|m_y\| + \|m_x\|)\sigma_y + \left(\lambda - \frac{1}{\lambda}\right) (\sigma_x\sigma_y + 2\tilde{\mathbb{E}}[\bar{X}\bar{Y}]\sigma_y^2) \right). \quad (14)$$

Let us condense the second term down. First, we have $\frac{\lambda^2}{(1+\lambda^2)^2} \leq \frac{1}{4}$, and $\left| \frac{\lambda^2}{(1+\lambda^2)^2} \cdot \left(\lambda - \frac{1}{\lambda}\right) \right| \leq \frac{1}{2}$. Thus,

$$\begin{aligned} &\left| 2\xi \cdot \frac{\lambda^2}{(1+\lambda^2)^2} \left(2(2\|m_y\| + \|m_x\|)\sigma_y + \left(\lambda - \frac{1}{\lambda}\right) (\sigma_x\sigma_y + 2\tilde{\mathbb{E}}[\bar{Y} \cdot \bar{X}]\sigma_y^2) \right) \right| \\ &\lesssim |\xi| \cdot \left((2\|m_y\| + \|m_x\|)\sigma_y + \left(\sigma_x\sigma_y + 2|\tilde{\mathbb{E}}[\bar{Y} \cdot \bar{X}]|\sigma_y^2) \right) \right) \\ &\lesssim |\xi| \cdot \left(\|m_y\|\sigma_y + \left(\sigma_x\sigma_y + 2\tilde{\mathbb{E}}[\bar{Y} \cdot \bar{X}]\sigma_y^2) \right) \right). \end{aligned}$$

We next show that the parenthesised terms may be discarded. First, Lemma 5.7 yields that

$$\|m_x\| \cdot \|m_y\| \lesssim |\mathbb{E}[XY]| = \sigma_x\sigma_y \cdot |\mathbb{E}[\bar{X}\bar{Y}]| \lesssim \sigma_x^2 + \mathbb{E}[\bar{X}\bar{Y}]^2\sigma_y^2.$$

We may now apply the AM-GM inequality on the parenthesised terms so for some small constant ε ,

$$|\xi| \cdot \left(\sigma_x\sigma_y + \sigma_y^2 |\tilde{\mathbb{E}}[\bar{Y}\bar{X}]| \right) \lesssim \varepsilon \left(\sigma_x^2 + \mathbb{E}[\bar{X}\bar{Y}]^2\sigma_y^2 \right) + \frac{1}{\varepsilon} \xi^2 \sigma_y^2.$$

Therefore, if $|\xi|\sigma_y \ll \|m_x\|$, then we may simplify the above as

$$\mathbb{E} \Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y)$$

$$\begin{aligned}
&\geq \underbrace{(\sigma_x^2 + \sigma_y^2 \cdot \tilde{\mathbb{E}}[\overline{XY}]^2)}_{\gtrsim \|m_x\| \|m_y\|} (1 - C\varepsilon) - \underbrace{|\tilde{\zeta}| \sigma_y \|m_y\|}_{\ll \|m_x\| \|m_y\|} - \underbrace{\tilde{\zeta}^2 \sigma_y^2}_{\ll \|m_x\|^2} \\
&\gtrsim \|m_x\| \|m_y\| \geq Q.
\end{aligned}$$

Here, the second inequality uses the same manipulations from (12). ■

The final part of the proof, which we execute in Section 5.5, will establish that if the $\tilde{\zeta}$ term cannot be absorbed into the first, then it is easy to get a gain of Q .

Lemma 5.9. *Suppose that $|\tilde{\zeta}| \gtrsim \frac{\|m_x\|}{\sigma_y}$. Then, there is an affine reweighing satisfying (10).*

Before doing this, we exhibit numerous consequences of the failure of every affine reweigh.

5.4 Consequences of affine failure

Recall

$$\tilde{\zeta} = \tilde{\mathbb{E}}[\overline{YX}^2] - \tilde{\mathbb{E}}[\overline{YX}] \cdot \tilde{\mathbb{E}}[\overline{X}^3].$$

Let us express $\overline{Y} = \rho \overline{X} + \sqrt{1 - \rho^2} W$, where $\rho = \mathbb{E}[\overline{XY}]$, so $\mathbb{E}[\overline{X}W] = 0$. We then have the far simpler expression

$$\tilde{\zeta} = \sqrt{1 - \rho^2} \tilde{\mathbb{E}}[\overline{X}^2 W]. \quad (15)$$

In this section, we establish the following consequences of the failure of every affine function.

Lemma 5.10. *Suppose that there is no affine reweigh satisfying (10), in the sense that for some small constant δ , for any affine function p , if $\tilde{m}_x$ and $\tilde{m}_y$ are the updated means, then*

$$\Phi(\tilde{m}_x, \tilde{m}_y) - \Phi(m_x, m_y) < \delta Q. \quad (16)$$

Further assume that

$$|\tilde{\zeta}| \gtrsim \frac{\|m_x\|}{\sigma_y}$$

Then,

- (i) $\frac{\|m_x\|}{\|m_y\|} \lesssim \delta$,
- (ii) $\rho = (-\frac{1}{2} + O(\delta)) \frac{\sigma_x}{\sigma_y}$,
- (iii) $\sigma_x \lesssim \sqrt{\delta} \cdot \|m_y\|$,
- (iv) $|\tilde{\zeta}| \lesssim \sqrt{\delta} \cdot \frac{\sigma_x}{\sigma_y}$, and
- (v) $\sigma_x^2 \gtrsim \|m_x\| \|m_y\|$.

Before proceeding, we introduce the following lemma which describes how much an affine reweigh can shift an affine function by.

Lemma 5.11. *For any linear function ℓ , we have*

$$\max_{\substack{\text{affine reweigh } p \\ \tilde{\mathbb{E}}[p^2] = 1}} \left| \tilde{\mathbb{E}}[p^2 \ell] \right| \geq \sqrt{\tilde{\mathbb{E}}[\ell^2]}.$$

Furthermore, there is an efficiently computable affine function p witnessing this inequality.

Proof. Express $\ell = h\sigma + \mu$, with $\tilde{\mathbb{E}} h = 0$ and $\tilde{\mathbb{E}} h^2 = 1$. We restrict our attention to p of the form $p = ah + b$. We then have

$$\begin{aligned} \max_p |\tilde{\mathbb{E}} [p^2 \ell]| &\geq \max_{(a,b)} \left| \begin{pmatrix} a \\ b \end{pmatrix}^\top \begin{pmatrix} \mu & \sigma \\ \sigma & \tilde{\mathbb{E}}[h^2 \ell] \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \right| \\ &= \left\| \begin{pmatrix} \mu & \sigma \\ \sigma & \tilde{\mathbb{E}}[h^2 \ell] \end{pmatrix} \right\|_{\text{op}} \\ &\geq \left\| \begin{pmatrix} \mu & \sigma \\ \sigma & \tilde{\mathbb{E}}[h^2 \ell] \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\| \\ &= \sqrt{\mu^2 + \sigma^2} = \sqrt{\tilde{\mathbb{E}} \ell^2}. \end{aligned} \quad \blacksquare$$

Lemma 5.12. *Under the assumptions of Lemma 5.10,*

$$(i) \quad \sigma_y \sqrt{1 - \rho^2} \lesssim \sqrt{\delta Q} \text{ and}$$

$$(ii) \quad \sigma_x^2 \gtrsim \|m_x\| \cdot \|m_y\|.$$

If either of these inequalities fails, a witness affine function that violates (16) is the affine reweigh described by Lemma 5.11 for the linear function $(2\|m_x\| + \|m_y\|)X + (2\|m_y\| + \|m_x\|)Y$.

Proof. Consider the “tangent” linear function

$$\ell(a, b) = (2\|m_x\| + \|m_y\|)a + (2\|m_y\| + \|m_x\|)b,$$

so $\tilde{\mathbb{E}} \ell(\langle u, x \rangle, \langle v, y \rangle) = \Phi(m_x, m_y)$. Then, Cauchy–Schwarz implies that for any (a, b) , shortening $\Phi = \Phi(m_x, m_y)$,

$$\ell(a, b)^2 \leq \Phi \cdot \psi(a, b). \quad (17)$$

Now, we have

$$\begin{aligned} \Phi \cdot \max_{\substack{\text{affine reweigh } p \\ \tilde{\mathbb{E}} p^2 = 1}} \psi(\tilde{m}_x, \tilde{m}_y) &\stackrel{(17)}{\geq} \max_{\substack{\text{affine reweigh } p \\ \tilde{\mathbb{E}} p^2 = 1}} \left(\tilde{\mathbb{E}} [p^2 \ell(\langle x, u \rangle, \langle y, v \rangle)] \right)^2 \\ &\geq \tilde{\mathbb{E}} \left[\ell(\langle x, u \rangle, \langle y, v \rangle)^2 \right] \\ &= \Phi^2 + \widetilde{\text{Var}} \ell(\langle x, u \rangle, \langle y, v \rangle), \end{aligned}$$

where the second inequality is Lemma 5.11. It follows that

$$\widetilde{\text{Var}} \ell(\langle x, u \rangle, \langle y, v \rangle) \leq \delta Q \Phi.$$

To avoid having to deal with a cross-covariance term, we can write

$$\ell(X, Y) = \ell(X, \sigma_y \cdot \left(\frac{\rho}{\sigma_x} X + \sqrt{1 - \rho^2} W \right)),$$

so

$$\widetilde{\text{Var}} \ell(X, Y) \gtrsim \|m_y\|^2 \cdot \sigma_y^2 \cdot (1 - \rho^2),$$

and thus

$$\sigma_y \sqrt{1 - \rho^2} \lesssim \sqrt{\delta Q},$$

establishing the first inequality. For the second inequality, we have

$$\begin{aligned}
\delta Q \Phi &\gtrsim \mathbf{Var} \ell(X, Y) \\
&\gtrsim \frac{1}{\sigma_x^2} (\mathbb{E} [\ell(X, Y) \cdot X])^2 \\
\sqrt{\delta Q} &\gtrsim \frac{1}{\sigma_x} \left| \frac{2\|m_x\| + \|m_y\|}{\|m_x\| + 2\|m_y\|} \cdot \sigma_x^2 + \mathbb{E}[XY] \right|.
\end{aligned} \tag{18}$$

Now recall Lemma 5.7. This implies that

$$\|m_x\| \cdot \|m_y\| \lesssim -\mathbb{E}[XY] \lesssim \sigma_x^2 + \sigma_x \sqrt{\delta Q} \lesssim \sigma_x^2 + \delta Q,$$

and since the δQ term is negligible in comparison to the left-hand side, we have the desired inequality. $\blacksquare$

We also have the following bound on a third moment quantity that shows up in the definition of ξ .

Lemma 5.13. *Under the assumptions of Lemma 5.10,*

$$\left| \tilde{\mathbb{E}}[\bar{X}^2 W] \right| \lesssim \frac{\|m_y\|}{\sigma_x}.$$

If this inequality fails, a witness affine function that violates (16) is one of the two choices $\frac{\bar{X} \pm W}{\sqrt{2}}$.

Proof. Consider the two reweighs

$$p_{\pm} = \frac{\bar{X} \pm W}{\sqrt{2}}.$$

Due to the orthonormality of $\bar{X}$ and W , we have $\tilde{\mathbb{E}}[p_{\pm}^2] = 1$. Furthermore, their updated means $\tilde{m}_x^{\pm}$ satisfy

$$\langle \tilde{m}_x^+ - \tilde{m}_x^-, u \rangle = \sigma_x \tilde{\mathbb{E}}[(p_+^2 - p_-^2) \bar{X}] = 2\sigma_x \tilde{\mathbb{E}}[\bar{X}^2 W].$$

We further have

$$\Phi(\tilde{m}_x^{\pm}, \tilde{m}_y^{\pm}) \leq \Phi(m_x, m_y) + \delta Q \lesssim \|m_y\|^2,$$

where the final inequality is simply because $\Phi(m_x, m_y) \asymp \|m_y\|^2$. We can lower bound the quantity on the left as

$$\Phi(\tilde{m}_x^{\pm}, \tilde{m}_y^{\pm}) \geq \langle \tilde{m}_x^{\pm}, u \rangle^2.$$

It follows that

$$\sigma_x \left| \tilde{\mathbb{E}}[\bar{X}^2 W] \right| \lesssim \|m_y\|,$$

and rearranging completes the proof. $\blacksquare$

Lemma 5.14. *Under the assumptions of Lemma 5.10,*

$$\frac{\|m_x\|}{\|m_y\|} \lesssim \delta$$

and

$$\mathbb{E}[XY] = \left(-\frac{1}{2} + O(\delta) \right) \sigma_x^2.$$

Proof. Combining the assumption and Lemma 5.13, along with the expression for ξ in (15), we have

$$\sqrt{1 - \rho^2} \cdot \frac{\|m_y\|}{\sigma_x} \gtrsim \frac{\|m_x\|}{\sigma_y}.$$

Substituting Lemma 5.12(i), we have

$$\frac{\|m_x\|}{\|m_y\|} \leq \frac{\sqrt{\delta Q}}{\sigma_x} \quad (19)$$

That is, the x -mean is significantly smaller than the y -mean. On the other hand, Lemma 5.12(ii) yields that

$$\sigma_x^2 \gtrsim \|m_x\| \cdot \|m_y\|.$$

Substituting this back in above yields

$$\frac{\|m_x\|^{3/2}}{\|m_y\|^{1/2}} \lesssim \sqrt{\delta Q},$$

and bounding $\|m_x\| \geq \sqrt{Q}$ yields

$$\frac{\|m_x\|}{\|m_y\|} \lesssim \delta.$$

For the second part, we plug this back into (18), which asserts that

$$\sqrt{\delta Q} \gtrsim \frac{1}{\sigma_x} \left| \frac{2\|m_x\| + \|m_y\|}{\|m_x\| + 2\|m_y\|} \cdot \sigma_x^2 + \mathbb{E}[XY] \right|.$$

Indeed, the first term within the absolute value is now equal to $(\frac{1}{2} + O(\delta)) \cdot \sigma_x^2$, so

$$\mathbb{E}[XY] = \sigma_x^2 \left(-\frac{1}{2} + O(\delta) + O\left(\frac{\sqrt{\delta Q}}{\sigma_x}\right) \right)$$

To conclude, we have

$$\frac{\sqrt{Q}}{\sigma_x} \leq \frac{\|m_x\|}{\sigma_x} \lesssim \sqrt{\frac{\|m_x\|}{\|m_y\|}} \lesssim \sqrt{\delta}. \quad \blacksquare$$

Finally, let us put the above lemmas together to establish Lemma 5.10.

Proof of Lemma 5.10. (i) and (ii) are Lemma 5.14. (iii) follows from (19) and bounding $\sqrt{Q} \leq \|m_x\|$. (v) is Lemma 5.12(ii). For (iv), we have by Lemma 5.13 that

$$|\xi| = \sqrt{1 - \rho^2} \mathbb{E}[\bar{X}^2 W] \lesssim \sqrt{1 - \rho^2} \cdot \frac{\|m_y\|}{\sigma_x}.$$

Now substitute Lemma 5.12(i), we may bound the above further as

$$|\xi| \lesssim \sqrt{\delta Q} \cdot \frac{\|m_y\|}{\sigma_x \sigma_y} \leq \sqrt{\delta} \cdot \frac{\|m_x\| \|m_y\|}{\sigma_x \sigma_y} \lesssim \sqrt{\delta} \cdot \frac{\sigma_x}{\sigma_y},$$

where the final inequality is (v). \blacksquare

5.5 The proof when $|\xi|$ is large

Let us complete the proof.

Lemma 5.9. *Suppose that $|\xi| \gtrsim \frac{\|m_x\|}{\sigma_y}$. Then, there is an affine reweighing satisfying (10).*

Proof. Recall the two reweighs from Lemma 5.6, corresponding to the evaluation of ψ at the two points

$$z_i = \left(\|m_x\| + \sigma_x \lambda_i, \|m_y\| + \rho \sigma_y \lambda_i + \sigma_y \left(\frac{\lambda_i^2}{1 + \lambda_i^2} \right) \xi \right),$$

and $z_0 = (\|m_x\|, \|m_y\|)$, where the λ_i are defined as follows. We start by defining λ by $\lambda - \frac{1}{\lambda} = \tilde{\mathbb{E}} X^3$, then set $\lambda_1 = \lambda$ and $\lambda_2 = -\frac{1}{\lambda}$.

For starters, observe that the failure of these reweighs implies that

$$\sigma_x |\lambda_i| \lesssim \|m_y\|. \quad (20)$$

Indeed,

$$\sigma_x^2 \lambda_i^2 \lesssim \|m_x\|^2 + \psi(z_i) \leq \|m_x\|^2 + \psi(z_0) + \delta Q \lesssim \|m_y\|^2.$$

Let us compute the value of the potential at each of these two values. Define

$$\psi_+(z) = 2(z_1^2 + z_2^2 + z_1 z_2) \text{ and } \psi_-(z) = 2(z_1^2 + z_2^2 - z_1 z_2).$$

Let us start with $\psi_+(z_1) - \psi(z_0)$. Write $\omega_1 = \left(\frac{\lambda_1^2}{1 + \lambda_1^2} \right) \leq 1$. Then, denoting $\beta = \frac{\rho \sigma_y}{\sigma_x} = -\frac{1}{2} + O(\delta)$,

$$\begin{aligned} \frac{1}{2} (\psi_+(z_1) - \psi(z_0)) &= \sigma_x^2 \lambda_1^2 + \sigma_y^2 (\rho \lambda_1 + \omega_1 \xi)^2 + 2\sigma_x \|m_x\| \lambda_1 + 2\sigma_y \|m_y\| (\rho \lambda_1 + \omega_1 \xi) \\ &\quad + (\|m_x\| + \sigma_x \lambda_1) (\rho \sigma_y \lambda_1 + \omega_1 \sigma_y \xi) + \|m_y\| \cdot \sigma_x \lambda_1 \\ &= \sigma_x^2 \lambda_1^2 (1 + \beta^2 + \beta) \\ &\quad + \sigma_x \lambda_1 \underbrace{(2\|m_x\| + \|m_y\| + \beta(\|m_x\| + 2\|m_y\|))}_{\lesssim \delta \|m_y\| \text{ by Lemma 5.10(i,ii)}} \\ &\quad + \omega_1 \xi \sigma_y \left(\underbrace{(2\|m_y\| + \|m_x\|)}_{\gg \lambda_1 \sigma_x \text{ by Lemma 5.10(iii)}} + \lambda_1 \sigma_x \underbrace{(2\beta + 1)}_{O(\delta) \text{ by Lemma 5.10(ii)}} \right) \\ &\quad + \underbrace{\omega_1^2 \xi^2 \sigma_y^2}_{\geq 0} \\ &\geq \frac{1}{2} \cdot \sigma_x^2 \lambda_1^2 - C \omega_1 \sigma_y \|m_y\| \cdot |\xi| \end{aligned} \quad (21)$$

$$\geq \frac{1}{2} \cdot \sigma_x^2 \lambda_1^2 - C \delta \cdot \omega_1 \|m_y\|^2, \quad (22)$$

where the second-to-last inequality absorbs the second term into the first, and the final inequality is Lemma 5.10(iii,iv). Since the above increment is bounded from above by δQ , we have

$$\sigma_x^2 \lambda_1^2 \lesssim \delta Q + \delta \cdot \omega_1 \|m_y\|^2 \lesssim \delta \|m_y\|^2. \quad (23)$$

Similarly, for the other branch, we may compute $\psi_-(z_2) - \psi(z_0)$. Again, write $\omega_2 = \left(\frac{\lambda_2^2}{1 + \lambda_2^2} \right)$, and denote $\beta = \frac{\rho \sigma_y}{\sigma_x} = -\frac{1}{2} + O(\delta)$, so

$$\frac{1}{2} (\psi_-(z_2) - \psi(z_0)) = -2\|m_x\| \|m_y\| + \sigma_x^2 \lambda_2^2 + \sigma_y^2 (\rho \lambda_2 + \omega_2 \xi)^2 + 2\sigma_x \|m_x\| \lambda_2 + 2\sigma_y \|m_y\| (\rho \lambda_2 + \omega_2 \xi)$$

$$\begin{aligned}
& - (\|m_x\| + \sigma_x \lambda_2) (\rho \sigma_y \lambda_2 + \omega_2 \sigma_y \xi) - \|m_y\| \cdot \sigma_x \lambda_2 \\
= & -2\|m_x\| \|m_y\| \\
& + \underbrace{\sigma_x^2 \lambda_2^2 (1 + \beta^2 - \beta)}_{\geq 0} \\
& + \underbrace{\sigma_x \lambda_2 (2\|m_x\| - \|m_y\| + \beta (-\|m_x\| + 2\|m_y\|))}_{\gtrsim \sigma_x |\lambda_2| \|m_y\|} \\
& + \underbrace{\omega_2 \xi \sigma_y ((2\|m_y\| - \|m_x\|) + \lambda_2 \sigma_x (2\beta - 1))}_{\stackrel{(20)}{\gtrsim} -|\omega_2 \xi| \sigma_y \|m_y\|} \\
& + \underbrace{\omega_2^2 \xi^2 \sigma_y^2}_{\geq 0} \\
\geq & \sigma_x |\lambda_2| \|m_y\| - 2\|m_x\| \|m_y\| - C |\omega_2 \xi| \sigma_y \|m_y\|. \tag{24}
\end{aligned}$$

Now, since the increment above in (24) is bounded from above by $O(\|m_x\|^2) = O(\|m_x\| \|m_y\|)$, we have

$$\begin{aligned}
\sigma_x |\lambda_2| & \lesssim \|m_x\| + |\omega_2 \xi| \sigma_y \\
& \lesssim \|m_x\| + \sqrt{\delta} \cdot \sigma_x |\omega_2| \\
& = \|m_x\| + \sqrt{\delta} \cdot \sigma_x \cdot \frac{|\lambda_2|}{|\lambda_1| + |\lambda_2|} \\
& \leq \|m_x\| + \sqrt{\delta} \cdot \sigma_x \cdot \frac{|\lambda_2|}{2\sqrt{|\lambda_1 \lambda_2|}} \\
& = \|m_x\| + \sqrt{\delta} \cdot \frac{\sigma_x |\lambda_2|}{2} \\
|\lambda_2| \sigma_x & \lesssim \|m_x\|.
\end{aligned}$$

Now, $|\lambda_1| = \frac{1}{|\lambda_2|}$, and by Lemma 5.10(v), we have $\sigma_x^2 \gtrsim \|m_x\| \cdot \|m_y\|$. Plugging these in above yield that

$$|\lambda_1| \sigma_x \gtrsim \|m_y\|. \tag{25}$$

However, (23) contradicts (25), completing the proof. $\blacksquare$

5.6 Proof of the main theorem

We are now ready to prove the main result, restated for convenience.

Theorem 5.1. *Let $1 < q < n/2$ and let $\Pi \in \mathbb{R}^{n^2 \times n^2}$ be a projector to a subspace of $\mathbb{R}^{n^2}$. Then there exists an algorithm that runs in time*

$$n^{O(\sqrt{q})}$$

and distinguishes between the following cases:

- **YES:** *There exists a nonzero $x, y \in \mathbb{R}^n$ such that $\Pi(x \otimes y) = x \otimes y$.*
- **NO:** *For all $x, y \in \mathbb{R}^n$ we have that $\|\Pi(x \otimes y)\|_2^2 \leq \frac{q}{n} \cdot \|x \otimes y\|_2^2$.*

Furthermore, in the YES case, the algorithm produces $\hat{x}, \hat{y}$ such that $\|\Pi(\hat{x} \otimes \hat{y})\|_2^2 > \frac{q}{n} \cdot \|\hat{x} \otimes \hat{y}\|_2^2$.

Proof. Let $\beta = 2$ in $\mathcal{A}$. We first prove completeness. Normalize a YES witness so that $\|x\|_2 = \|y\|_2 = 1$, and round both vectors to the grid to obtain $\hat{x}, \hat{y} \in \Sigma^n$ with

$$\|x - \hat{x}\|_2, \|y - \hat{y}\|_2 \leq c_0 n^{-2}$$

for a sufficiently small absolute constant c_0 . Then

$$|\|\hat{x}\|_2^2 - 1|, |\|\hat{y}\|_2^2 - 1| \leq n^{-2},$$

and

$$\|\hat{x} \otimes \hat{y} - x \otimes y\|_2 \leq \|(\hat{x} - x) \otimes \hat{y}\|_2 + \|x \otimes (\hat{y} - y)\|_2 = O(n^{-2}).$$

Since $(I - \Pi)(x \otimes y) = 0$, decreasing c_0 if necessary shows that the point mass at $(\hat{x}, \hat{y})$ satisfies (9). Hence the relaxation is feasible in the YES case.

We next prove soundness. Suppose that the relaxation is feasible, and set

$$\theta = K \sqrt{\frac{q}{n}}$$

for a sufficiently large absolute constant K . We first dispose of the range in which θ is not smaller than the constant required by the pinning lemma. In that range $q = \Omega(n)$, and we apply Theorem 4.1 with $\varepsilon = 1 - q/n$. Since $\varepsilon > 1/2$, its running time is $n^{O(\sqrt{n})} = n^{O(\sqrt{q})}$, and its soundness condition is precisely the NO condition above. We may therefore assume that θ is smaller than the required constant and apply Lemma 5.4. After

$$O(\theta \sqrt{n}) = O(\sqrt{q})$$

coordinate pinnings and affine-square reweightings, one of its three win conditions holds. Fix an absolute constant $c_{\text{round}} > 0$ smaller than the implicit constants in these conditions.

If either mean is large, the corresponding best-response bound in Lemma 5.3 gives

$$\|\Pi(u \otimes v)\|_2^2 \geq c_{\text{round}}^2 K^2 \frac{q}{n} - O(n^{-2}).$$

Here $n^{-2} \leq (q/n)/n$ because $q > 1$.

It remains to consider the singular-value condition. The final conclusion of Lemma 5.4 gives

$$\|R\|_F \geq \frac{\sqrt{q}}{n}, \quad \frac{n^{-2}}{\|R\|_F} \leq \frac{1}{n\sqrt{q}} \leq \frac{1}{\sqrt{n}} \sqrt{\frac{q}{n}}.$$

The singular-vector bound in Lemma 5.3 therefore yields

$$\|\Pi(u \otimes v)\|_2^2 \geq \left(c_{\text{round}} K - \frac{1}{\sqrt{n}}\right)^2 \frac{q}{n}.$$

Choosing K sufficiently large makes both rounding bounds strictly larger than q/n . Thus, in every case,

$$\|\Pi(u \otimes v)\|_2^2 > \frac{q}{n}.$$

Therefore, a feasible relaxation contradicts the NO condition. The algorithm enumerates all coordinate branches of the required depth, and the affine reweightings and all final rounding candidates are computable from the corresponding moment matrices. It therefore finds such a pair whenever the relaxation is feasible.

Finally, the grid has polynomial size, the relaxation has degree $O(\sqrt{q})$, and there are $n^{O(\sqrt{q})}$ coordinate branches. Both the feasibility test and the rounding procedure consequently run in time

$$n^{O(\sqrt{q})}.$$

The efficient computability of the affine reweighs is immediate because each reweigh is one of five explicit reweighs: the two in Lemma 5.6, that in Lemma 5.12, or the two in Lemma 5.13. ■

6 Fine-grained pinning lemma

We prove the following fine-grained pinning lemma:

Lemma 6.1. *Let μ_0 be a distribution on $\mathbb{R}^n$ with finite second moment and let $0 < \gamma \leq \sqrt{n}$. There is a distribution over (S, x_S) where $S \subseteq [n]$ with $|S| \leq O(\sqrt{n}/\gamma)$ and $x_S \in \mathbb{R}^{|S|}$ such that*

$$\mathbb{E}_{S, x_S} \left\| \mathbb{E}_{x \sim \mu_0 | x_S} x - \mathbb{E}_{x \sim \mu_0} x \right\|_2^2 \geq \min \left\{ \frac{1}{\gamma} \cdot \mathbb{E}_{S, x_S} \|\mathbf{Cov}(\mu_0 | x_S)\|_F, \frac{1}{\gamma^2} \cdot \text{tr}(\mathbf{Cov}(\mu_0)) \right\}$$

and furthermore (S, x_S) give a decomposition of μ_0 , in that $\mathbb{E}_{S, x_S}(\mu_0 | x_S) = \mu_0$. The same conclusion holds for a pseudodistribution μ_0 of degree at least $O(\sqrt{n}/\gamma)$.

We first record the estimates used in the proof. Throughout the claims, fix an initial distribution μ_0 on $\mathbb{R}^n$ with finite second moment and $\text{tr}(\mathbf{Cov}(\mu_0)) > 0$. For a conditional distribution ν of μ_0 , let

$$m_\nu := \mathbb{E}_{x \sim \nu} x \quad \text{and} \quad \Sigma_\nu := \mathbf{Cov}(\nu).$$

Let $m_0 := \mathbb{E}_{x \sim \mu_0} x$. For a distribution π_ℓ over conditional distributions obtained from pinning coordinates of μ_0 , define

$$s_\ell := \frac{\mathbb{E}_{\nu \sim \pi_\ell} \|m_\nu - m_0\|_2^2}{\text{tr}(\mathbf{Cov}(\mu_0))}.$$

As in other analyses of global correlation rounding, we will track the evolution of a potential function, in this case s_ℓ , as we pin coordinates. We construct our pinning of μ_0 via the following process: given a current set of pinned coordinates S pinned to values x_S , choose a coordinate i on each branch with weights proportional to $\mathbf{Var}_\nu(x_i)$. Then sample a value x_i from the marginal distribution on coordinate i conditioned on x_S , and update $S \leftarrow S \cup \{i\}$. In the following claims, π_ℓ is the distribution over pinnings resulting from ℓ steps of this process.

Claim 6.2. We have

$$s_{\ell+1} - s_\ell \geq \frac{1}{\text{tr}(\mathbf{Cov}(\mu_0))} \mathbb{E}_{\nu \sim \pi_\ell} \frac{\|\Sigma_\nu\|_F^2}{\text{tr}(\Sigma_\nu)},$$

where the ratio is interpreted as 0 when $\text{tr}(\Sigma_\nu) = 0$. Also,

$$s_\ell + \frac{\mathbb{E}_{\nu \sim \pi_\ell} \text{tr}(\Sigma_\nu)}{\text{tr}(\mathbf{Cov}(\mu_0))} = 1.$$

We first state a simple and standard fact that will be necessary in the proof of the claim.

Fact 6.3 (Variance drop from conditioning (see [BRS11])). *For any two real-valued random variables x, y with $\mathbf{Var}(y) > 0$,*

$$\mathbf{Var}(x) - \mathbb{E}_y \mathbf{Var}(x | y) = \mathbf{Var}(\mathbb{E}[x | y]) \geq \Omega(1) \cdot \frac{\mathbf{Cov}(x, y)^2}{\mathbf{Var}(y)}.$$

We now return to the proof of Claim 6.2.

Proof. The covariance identity

$$\mathbf{Cov}(\mu_0) = \mathbb{E}_{\nu \sim \pi_\ell} \Sigma_\nu + \mathbb{E}_{\nu \sim \pi_\ell} (m_\nu - m_0)(m_\nu - m_0)^\top$$

follows because sampling $\nu \sim \pi_\ell$ and then $x \sim \nu$ gives a sample from μ_0 . Taking traces and dividing by $\text{tr}(\mathbf{Cov}(\mu_0))$ gives the second display.

Fix a branch ν with $\text{tr}(\Sigma_\nu) > 0$; trace-zero branches contribute 0 and are left unchanged. By Fact 6.3, conditioning on coordinate i decreases the variance of x_j by at least

$$\text{Var}_\nu(x_j) - \mathbb{E}_{x_i} \text{Var}_{\nu|x_i}(x_j) \geq \frac{\text{Cov}_\nu(x_i, x_j)^2}{\text{Var}_\nu(x_i)},$$

where zero-variance coordinates are omitted. If i is sampled with probability proportional to $\text{Var}_\nu(x_i)$, then averaging over the choice of i gives

$$\text{tr}(\Sigma_\nu) - \mathbb{E}_{i, x_i} \text{tr}(\Sigma_{\nu|x_i}) \geq \frac{\|\Sigma_\nu\|_F^2}{\text{tr}(\Sigma_\nu)}.$$

We then pass to the conditional distribution $\nu \mid x_i$.

It remains to translate the trace drop to mean drift. For each fixed coordinate i , since $\mathbb{E}_{x_i} m_{\nu|x_i} = m_\nu$,

$$\begin{aligned} \mathbb{E} \|m_{\nu|x_i} - m_0\|_2^2 - \|m_\nu - m_0\|_2^2 &= \mathbb{E} \|m_{\nu|x_i} - m_\nu\|_2^2 \\ &= \mathbb{E}_{x_i} \left[\mathbb{E}_{x \sim \nu|x_i} \|x - m_\nu\|_2^2 - \mathbb{E}_{x \sim \nu|x_i} \|x - m_{\nu|x_i}\|_2^2 \right] \\ &= \text{tr}(\Sigma_\nu) - \mathbb{E} \text{tr}(\Sigma_{\nu|x_i}). \end{aligned}$$

The second equality uses $x - m_\nu = (x - m_{\nu|x_i}) + (m_{\nu|x_i} - m_\nu)$ and the fact that $\mathbb{E}[x - m_{\nu|x_i} \mid x_i] = 0$. Averaging over the sampled coordinate i and over $\nu \sim \pi_\ell$ proves the claim. $\blacksquare$

Note that Claim 6.2 gives an unconditional lower bound on the increase in our potential function s_ℓ . We now translate this to a concrete lower bound on the increase whenever the mean rounding success condition fails.

Claim 6.4. If

$$\mathbb{E}_{\nu \sim \pi_\ell} \|\Sigma_\nu\|_F > \gamma \mathbb{E}_{\nu \sim \pi_\ell} \|m_\nu - m_0\|_2^2,$$

then

$$s_{\ell+1} - s_\ell \geq \max \left\{ \frac{1 - s_\ell}{n}, \frac{\gamma^2 s_\ell^2}{1 - s_\ell} \right\}.$$

Proof. First, $\|\Sigma_\nu\|_F^2 / \text{tr}(\Sigma_\nu) \geq \text{tr}(\Sigma_\nu) / n$ for every branch, so Claim 6.2 gives

$$s_{\ell+1} - s_\ell \geq \frac{1 - s_\ell}{n}.$$

Second, by Cauchy-Schwarz and the identities in Claim 6.2,

$$\begin{aligned} \frac{1}{\text{tr}(\text{Cov}(\mu_0))} \mathbb{E}_{\nu \sim \pi_\ell} \frac{\|\Sigma_\nu\|_F^2}{\text{tr}(\Sigma_\nu)} &\geq \frac{1}{\text{tr}(\text{Cov}(\mu_0))} \cdot \frac{(\mathbb{E}_{\nu \sim \pi_\ell} \|\Sigma_\nu\|_F)^2}{\mathbb{E}_{\nu \sim \pi_\ell} \text{tr}(\Sigma_\nu)} \\ &\geq \frac{\gamma^2 s_\ell^2}{1 - s_\ell}. \end{aligned}$$

Combining the two bounds proves the claim. $\blacksquare$

We now first use this to show as a warmup that $O\left(\frac{\sqrt{n}}{\gamma} \log n\right)$ steps of conditioning suffices to fulfill the mean rounding success condition. This follows via a simple analysis of the recurrence outlined in the previous claim.

Claim 6.5. Let $0 < \gamma \leq \sqrt{n}$. Suppose $s_0 = 0$ and, until reaching $s_{\text{end}} = 1/\gamma^2$, a sequence satisfies the recurrence in Claim 6.4. Then $s_\ell \geq s_{\text{end}}$ for some

$$\ell \leq O\left(\frac{\sqrt{n}}{\gamma} \log n\right).$$

Proof. The first term of the recurrence gives $s_1 \geq 1/n$. Until the endpoint is reached, the recurrence implies

$$s_{\ell+1} - s_\ell \geq \sqrt{\frac{1 - s_\ell}{n} \cdot \frac{\gamma^2 s_\ell^2}{1 - s_\ell}} = \frac{\gamma}{\sqrt{n}} s_\ell.$$

Hence $s_{\ell+1} \geq (1 + \gamma/\sqrt{n})s_\ell$ until the endpoint is reached. Starting from $s_1 \geq 1/n$, this takes at most $O((\sqrt{n}/\gamma) \log n)$ additional rounds. ■

We now give a slightly refined analysis of the recurrence to show that in fact $O\left(\frac{\sqrt{n}}{\gamma}\right)$ steps of conditioning suffices.

Claim 6.6. Let $1/\sqrt{n} \leq \gamma \leq \sqrt{n}$. Suppose $s_0 = 0$ and, until reaching $s_{\text{end}} = 1/\gamma^2$, a sequence satisfies the recurrence in Claim 6.4. Then $s_\ell \geq s_{\text{end}}$ for some

$$\ell \leq O\left(\frac{\sqrt{n}}{\gamma}\right).$$

Proof. First we use the first term in the recurrence to reach $s_\ell \geq 1/(2\gamma\sqrt{n})$. Since $\gamma \geq 1/\sqrt{n}$, this threshold is at most $1/2$, and

$$1 - s_{\ell+1} \leq \left(1 - \frac{1}{n}\right)(1 - s_\ell).$$

Hence this first regime takes at most

$$O\left(n \log\left(\frac{1}{1 - 1/(2\gamma\sqrt{n})}\right)\right) \leq O\left(\frac{\sqrt{n}}{\gamma}\right)$$

rounds. Next use the simpler consequence

$$s_{\ell+1} \geq s_\ell(1 + \gamma^2 s_\ell).$$

Consider a bucket $s_\ell \in [a, 2a]$ with $a \leq 1/\gamma^2$. While the sequence remains in this bucket,

$$s_{\ell+1} \geq s_\ell(1 + \gamma^2 a),$$

so the bucket is crossed in $O(1/(\gamma^2 a))$ rounds. Summing over geometrically increasing buckets starting at $1/(2\gamma\sqrt{n})$, this second regime reaches either s_{end} within

$$O\left(\frac{1}{\gamma^2 \cdot 1/(2\gamma\sqrt{n})}\right) = O\left(\frac{\sqrt{n}}{\gamma}\right)$$

additional rounds, concluding the proof. ■

We now combine the previous claims to prove the pinning lemma.

Proof of Lemma 6.1. If $\text{tr}(\mathbf{Cov}(\mu_0)) = 0$, then μ_0 is supported on a point and the conclusion holds without any conditioning. If $\gamma < 1/\sqrt{n}$, condition on all coordinates; this uses $n \leq \sqrt{n}/\gamma$ conditionings and leaves zero covariance. Otherwise let π_0 be the point mass on μ_0 .

If

$$\mathbb{E}_{v \sim \pi_\ell} \|\Sigma_v\|_F > \gamma \mathbb{E}_{v \sim \pi_\ell} \|m_v - m_0\|_2^2$$

for every $\ell < T = O\left(\frac{\sqrt{n}}{\gamma}\right)$, then Claims 6.4 and 6.6 imply that some s_ℓ reaches $s_{\text{end}} = 1/\gamma^2$ by time T , which corresponds to the second term in the minimum, completing the proof. It is easy to see that the entire proof also goes through for pseudodistributions of appropriately high degree, since we only use the Cauchy–Schwarz inequality and conditioning. ■

References

- [ABH26] Prashanti Anderson, Ainesh Bakshi, and Samuel B. Hopkins. *Entrywise Low-Rank Approximation and Matrix $p \rightarrow q$ Norms via Global Correlation Rounding*. 2026. DOI: [10.48550/arXiv.2604.22699](https://doi.org/10.48550/arXiv.2604.22699). arXiv: [2604.22699](https://arxiv.org/abs/2604.22699) [cs.DS]. URL: <https://arxiv.org/abs/2604.22699> (pages 3, 9).
- [ABMT25] Prashanti Anderson, Ainesh Bakshi, Mahbod Majid, and Stefan Tiegel. “Sample-Optimal Private Regression in Polynomial Time”. In: *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*. STOC ’25. Prague, Czechia: Association for Computing Machinery, 2025, pp. 2341–2349. ISBN: 9798400715105. DOI: [10.1145/3717823.3718218](https://doi.org/10.1145/3717823.3718218). URL: <https://doi.org/10.1145/3717823.3718218> (page 11).
- [BBHKSZ12] Boaz Barak, Fernando G. S. L. Brandão, Aram W. Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. “Hypercontractivity, Sum-of-Squares Proofs, and Their Applications”. In: *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, May 19, 2012, pp. 307–326. DOI: [10.1145/2213977.2214006](https://doi.org/10.1145/2213977.2214006). arXiv: [1205.4484](https://arxiv.org/abs/1205.4484) [cs.CC]. URL: <https://doi.org/10.1145/2213977.2214006> (pages 1, 3).
- [BGGLT17] Vijay V. S. P. Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani. “Weak Decoupling, Polynomial Folds and Approximate Optimization over the Sphere”. In: *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, Oct. 2017, pp. 1008–1019. DOI: [10.1109/FOCS.2017.97](https://doi.org/10.1109/FOCS.2017.97). arXiv: [1611.05998](https://arxiv.org/abs/1611.05998) [cs.DS]. URL: <https://doi.org/10.1109/FOCS.2017.97> (pages 2, 3).
- [BGL17] Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee. “Sum-of-Squares Certificates for Maxima of Random Tensors on the Sphere”. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017)*. Vol. 81. Leibniz International Proceedings in Informatics (LIPIcs). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Aug. 11, 2017, 31:1–31:20. DOI: [10.4230/LIPIcs.APPROX-RANDOM.2017.31](https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2017.31). arXiv: [1605.00903](https://arxiv.org/abs/1605.00903) [cs.CC]. URL: <https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.APPROX-RANDOM.2017.31> (page 3).
- [BH15] Fernando G. S. L. Brandão and Aram W. Harrow. *Estimating operator norms using covering nets*. Sept. 16, 2015. DOI: [10.48550/arXiv.1509.05065](https://doi.org/10.48550/arXiv.1509.05065). arXiv: [1509.05065](https://arxiv.org/abs/1509.05065) [quant-ph]. URL: <https://arxiv.org/abs/1509.05065> (pages 1, 3).

- [BKS14] Boaz Barak, Jonathan A. Kelner, and David Steurer. “Rounding sum-of-squares relaxations”. In: *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, 2014, pp. 31–40. DOI: [10.1145/2591796.2591886](https://doi.org/10.1145/2591796.2591886). arXiv: [1312.6652](https://arxiv.org/abs/1312.6652) [cs.DS]. URL: <https://doi.org/10.1145/2591796.2591886> (page 3).
- [BKS17] Boaz Barak, Pravesh K. Kothari, and David Steurer. “Quantum entanglement, sum of squares, and the log rank conjecture”. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. Association for Computing Machinery, June 19, 2017, pp. 975–988. DOI: [10.1145/3055399.3055488](https://doi.org/10.1145/3055399.3055488). arXiv: [1701.06321](https://arxiv.org/abs/1701.06321) [quant-ph]. URL: <https://doi.org/10.1145/3055399.3055488> (pages 1–5, 9).
- [BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. “Rounding Semidefinite Programming Hierarchies via Global Correlation”. In: *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*. IEEE Computer Society, 2011, pp. 472–481. DOI: [10.1109/FOCS.2011.95](https://doi.org/10.1109/FOCS.2011.95). arXiv: [1104.4680](https://arxiv.org/abs/1104.4680) [cs.DS]. URL: <https://doi.org/10.1109/FOCS.2011.95> (pages 2, 3, 6, 9, 36).
- [BS16] Boaz Barak and David Steurer. *Proofs, Beliefs, and Algorithms through the Lens of Sum-of-Squares*. <https://www.sumofsquares.org/public/index.html>. Lecture notes, work in progress. 2016. (Visited on 08/06/2026) (pages 4, 11).
- [BT12] Hugue Blier and Alain Tapp. “A Quantum Characterization of NP”. In: *Computational Complexity* 21.3 (2012), pp. 499–510. DOI: [10.1007/s00037-011-0016-2](https://doi.org/10.1007/s00037-011-0016-2) (page 1).
- [CE25] Yuansi Chen and Ronen Eldan. “Localization schemes: A framework for proving mixing bounds for Markov chains”. In: *Duke Mathematical Journal* 174.8 (June 1, 2025), pp. 1431–1510. DOI: [10.1215/00127094-2024-0063](https://doi.org/10.1215/00127094-2024-0063). arXiv: [2203.04163](https://arxiv.org/abs/2203.04163) [math.PR]. URL: <https://doi.org/10.1215/00127094-2024-0063> (page 10).
- [CKPZ17] Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborova. “Information-theoretic thresholds from the cavity method”. In: *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. STOC ’17. Association for Computing Machinery, June 19, 2017, pp. 146–157. DOI: [10.1145/3055399.3055420](https://doi.org/10.1145/3055399.3055420). arXiv: [1611.00814](https://arxiv.org/abs/1611.00814) [cs.DM]. URL: <https://doi.org/10.1145/3055399.3055420> (page 9).
- [DW12] Andrew C. Doherty and Stephanie Wehner. “Convergence of SDP hierarchies for polynomial optimization on the hypersphere”. In: *CoRR* abs/1210.5048 (2012). DOI: [10.48550/arXiv.1210.5048](https://arxiv.org/abs/1210.5048). arXiv: [1210.5048](https://arxiv.org/abs/1210.5048) [math.OC]. URL: <https://arxiv.org/abs/1210.5048> (page 3).
- [FF21] Kun Fang and Hamza Fawzi. “The sum-of-squares hierarchy on the sphere and applications in quantum information theory”. In: *Mathematical Programming* 190.1–2 (2021), pp. 331–360. DOI: [10.1007/s10107-020-01537-7](https://doi.org/10.1007/s10107-020-01537-7). arXiv: [1908.05155](https://arxiv.org/abs/1908.05155) [math.OC]. URL: <https://doi.org/10.1007/s10107-020-01537-7> (page 3).
- [GLS81] Martin Grötschel, László Lovász, and Alexander Schrijver. “The ellipsoid method and its consequences in combinatorial optimization”. In: *Combinatorica* 1 (1981), pp. 169–197 (page 12).
- [Gri01] Dima Grigoriev. “Complexity of Positivstellensatz proofs for the knapsack”. In: *computational complexity* 10.2 (2001), pp. 139–154 (page 12).

- [GS12] Venkatesan Guruswami and Ali Kemal Sinop. “Faster SDP Hierarchy Solvers for Local Rounding Algorithms”. In: *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20–23, 2012*. IEEE Computer Society, 2012, pp. 197–206. DOI: [10.1109/FOCS.2012.58](https://doi.org/10.1109/FOCS.2012.58). arXiv: [1207.4372 \[cs.DS\]](https://arxiv.org/abs/1207.4372). URL: <https://doi.org/10.1109/FOCS.2012.58> (pages 3, 9).
- [HM13] Aram W. Harrow and Ashley Montanaro. “Testing Product States, Quantum Merlin-Arthur Games and Tensor Optimization”. In: *Journal of the ACM* 60.1 (Feb. 2013), 3:1–3:43. DOI: [10.1145/2432622.2432625](https://doi.org/10.1145/2432622.2432625). arXiv: [1001.0017 \[quant-ph\]](https://arxiv.org/abs/1001.0017). URL: <https://doi.org/10.1145/2432622.2432625> (pages 1, 3).
- [HSS16] Samuel B. Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. “Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors”. In: *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. STOC ’16. Association for Computing Machinery, June 19, 2016, pp. 178–191. DOI: [10.1145/2897518.2897529](https://doi.org/10.1145/2897518.2897529). URL: <https://doi.org/10.1145/2897518.2897529> (page 3).
- [HT26] Samuel B. Hopkins and Stefan Tiegel. *Algorithms with Polynomially-Improved Approximation Factors for the $2 \rightarrow q$ Norm, and Applications*. 2026. DOI: [10.48550/arXiv.2605.25303](https://arxiv.org/abs/2605.25303). arXiv: [2605.25303 \[cs.DS\]](https://arxiv.org/abs/2605.25303). URL: <https://arxiv.org/abs/2605.25303> (page 3).
- [JKR19] Vishesh Jain, Frederic Koehler, and Andrej Risteski. “Mean-field approximation, convex hierarchies, and the optimality of correlation rounding: a unified perspective”. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. Association for Computing Machinery, June 23, 2019, pp. 1226–1236. DOI: [10.1145/3313276.3316299](https://doi.org/10.1145/3313276.3316299). arXiv: [1808.07226 \[cs.LG\]](https://arxiv.org/abs/1808.07226). URL: <https://doi.org/10.1145/3313276.3316299> (pages 3, 9).
- [JLV23] Nathaniel Johnston, Benjamin Lovitz, and Aravindan Vijayaraghavan. “Computing linear sections of varieties: quantum entanglement, tensor decompositions and beyond”. In: *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Society, Nov. 6, 2023, pp. 1316–1336. DOI: [10.1109/FOCS57990.2023.00079](https://doi.org/10.1109/FOCS57990.2023.00079). arXiv: [2212.03851 \[cs.DS\]](https://arxiv.org/abs/2212.03851). URL: <https://ieeexplore.ieee.org/document/10353209> (page 3).
- [JWX26a] Fernando Granha Jeronimo, Pei Wu, and Haochen Xu. *An Argmax Principle for Sum-of-Squares Relaxations on the Sphere*. 2026. arXiv: [2608.02594 \[cs.CC\]](https://arxiv.org/abs/2608.02594). URL: <https://arxiv.org/abs/2608.02594> (page 3).
- [JWX26b] Fernando Granha Jeronimo, Pei Wu, and Haochen Xu. *Optimal Quantum de Finetti Theorems via Argmax Rounding*. 2026. arXiv: [2608.02590 \[quant-ph\]](https://arxiv.org/abs/2608.02590). URL: <https://arxiv.org/abs/2608.02590> (page 4).
- [Kho02] Subhash Khot. “On the power of unique 2-prover 1-round games”. In: *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, May 19, 2002, pp. 767–775. DOI: [10.1145/509907.510017](https://doi.org/10.1145/509907.510017). URL: <https://doi.org/10.1145/509907.510017> (page 1).
- [KLP06] Etienne de Klerk, Monique Laurent, and Pablo A. Parrilo. “A PTAS for the minimization of polynomials of fixed degree over the simplex”. In: *Theoretical Computer Science* 361.2–3 (Sept. 2006), pp. 210–225. ISSN: 0304-3975. DOI: [10.1016/j.tcs.2006.05.011](https://doi.org/10.1016/j.tcs.2006.05.011) (page 3).

- [LR21] Elaine Levey and Thomas Rothvoss. “A $(1 + \epsilon)$ -Approximation for Makespan Scheduling with Precedence Constraints Using LP Hierarchies”. In: *SIAM Journal on Computing* 50.3 (2021), STOC16-201–STOC16-217. DOI: [10.1137/16M1105049](https://doi.org/10.1137/16M1105049). arXiv: [1509.07808](https://arxiv.org/abs/1509.07808) [cs.DS]. URL: <https://doi.org/10.1137/16M1105049> (page 3).
- [Mon08] Andrea Montanari. “Estimating random variables from random sparse observations”. In: *European Transactions on Telecommunications* 19.4 (June 2008), pp. 385–403. DOI: [10.1002/ett.1289](https://doi.org/10.1002/ett.1289). arXiv: [0709.0145](https://arxiv.org/abs/0709.0145) [cs.IT] (pages 2, 9).
- [Nes00] Yurii Nesterov. “Squared functional systems and optimization problems”. In: *High performance optimization*. Springer, 2000, pp. 405–440 (page 12).
- [NH17] Anand Natarajan and Aram W. Harrow. *Semidefinite programming and computational aspects of entanglement. Lecture 1: November 3*. IHP Fall 2017 lecture notes. Nov. 3, 2017. URL: <https://www.mit.edu/~aram/teaching/sdp/lect1.pdf> (page 1).
- [Par00] Pablo A Parrilo. “Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization”. PhD thesis. California Institute of Technology, 2000 (page 12).
- [RT12] Prasad Raghavendra and Ning Tan. “Approximating CSPs with Global Cardinality Constraints Using SDP Hierarchies”. In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2012, pp. 373–387. DOI: [10.1137/1.9781611973099.33](https://doi.org/10.1137/1.9781611973099.33). URL: <https://doi.org/10.1137/1.9781611973099.33> (pages 2, 3, 9).
- [Sho87] Naum Zuselevich Shor. “An approach to obtaining global extremums in polynomial mathematical programming problems”. In: *Cybernetics* 23.5 (1987), pp. 695–700 (page 12).
- [WEM25] Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. “The Kikuchi Hierarchy and Tensor PCA”. In: *Journal of the ACM* 72.5 (Oct. 9, 2025), 35:1–35:40. DOI: [10.1145/3762806](https://doi.org/10.1145/3762806). arXiv: [1904.03858](https://arxiv.org/abs/1904.03858) [cs.DS]. URL: <https://doi.org/10.1145/3762806> (page 3).