Electronic Colloquium on Computational Complexity, Report No. 13 (2021)

Positive spectrahedrons: Geometric properties, Invariance principles and Pseudorandom generators

Srinivasan Arunachalam
IBM Quantum.
IBM T.J. Watson Research Center
Yorktown Heights, USA
Srinivasan.Arunachalam@ibm.com

Penghui Yao
State Key Laboratory for
Novel Software Technology,
Nanjing University
pyao@nju.edu.cn

January 20, 2021

Abstract

In a recent work, O'Donnell, Servedio and Tan (STOC 2019) gave explicit pseudorandom generators (PRGs) for arbitrary m-facet polytopes in n variables with seed length poly-logarithmic in m, n, concluding a sequence of works in the last decade, that was started by Diakonikolas, Gopalan, Jaiswal, Servedio, Viola (SICOMP 2010) and Meka, Zuckerman (SICOMP 2013) for fooling linear and polynomial threshold functions, respectively. In this work, we consider a natural extension of PRGs for intersections of positive spectrahedrons. A positive spectrahedron is a Boolean function $f(x) = [x_1 A^1 + \cdots + x_n A^n \preceq B]$ where the A^is are $k \times k$ positive semidefinite matrices. We construct explicit PRGs that δ-fool "regular" width-M positive spectrahedrons (i.e., when none of the A^is are dominant) over the Boolean space with seed length $\mathrm{poly}(\log k, \log n, M, 1/\delta)$.

Our main technical contributions are the following: We first prove an invariance principle for positive spectrahedrons via the well-known Lindeberg method. As far as we are aware such a generalization of the Lindeberg method was unknown. Second, we prove various geometric properties of positive spectrahedrons such as their noise sensitivity, Gaussian surface area and a Littlewood-Offord theorem for positive spectrahedrons. Using these results, we give applications for constructing PRGs for positive spectrahedrons, learning theory, discrepancy sets for positive spectrahedrons (over the Boolean cube) and PRGs for intersections of structured polynomial threshold functions.

ISSN 1433-8092

Contents

1 Introduction

Constructing explicit pseudorandom generators (PRG) for a class of interesting Boolean functions has received tremendous attention in the last few decades. One particular class of functions that has seen a flurry of works is the class of halfspaces. A *halfspace* is a Boolean function $f : \{-1,1\}^n \to \{0,1\}$ that can be expressed as $f(x) = \text{sign}(a_1x_1 + \cdots + a_nx_n - b)$ for some real values $a_1, \ldots, a_n, b \in \mathbb{R}$. Halfspaces arise naturally in a many areas of theoretical computer science including machine learning, communication complexity, circuit complexity and pseudorandomness. A successful line of work [Ser06, DHK+10, MZ13, KM15, GKM18] resulted in PRGs that ε-fool halfspaces with seed length poly-logarithmic in (n/ε) over the Boolean space.

Given the success in designing PRGs for single halfspaces (or linear threshold function), two alternate lines of work received a lot of attention, *polynomial threshold functions* and *intersections* of halfspaces. A degree-d polynomial threshold function (PTF) is simply a function $f(x) = \text{sign}(p(x))$ where p is a degree-d polynomial. In this direction, there have been a sequence of works [DGJ+10, DHK+10, Kan10, Kan11a, Kan11b, Kan11c, Kan14b, OST20] that produced PRGs with seed length exponential in d over the Boolean space and quasi-polynomial in d over the Gaussian space. Alternatively, another line of work considered *intersections* of halfspaces (i.e., a polytope). In this direction, a sequence of works [GOWZ10, HKM13, ST17, CDS19, OST19] produced a PRG for m-facet polytopes in n variables with seed length poly-logarithmic in m, n.[1]

In this work, we initiate the construction of PRGs for spectrahedrons: a natural generalization of halfspaces, polytopes and PTFs in one framework. A *spectrahedron* $S \subseteq \mathbb{R}^n$ is a feasible region of a *semidefinite program*. Namely,

$$S = \left\{ x \in \mathbb{R}^n : \sum_i x_i A^i \preceq B \right\}$$

for some symmetric matrices $A^1, \ldots, A^n, B$, where $\preceq$ is the standard Löwner ordering.[2] We say S is a *positive spectrahedron* if either all A^is are positive semidefinite (PSD) or all A^is are negative semidefinite. Spectrahedrons are important basic objects in polynomial optimization and algebraic geometry [BPT12, Sch18]. Mathematically, spectrahedrons have rich and complicated structures and include well-known geometric objects like polytopes, cylinders, polyhedrons, elliptopes. Computationally, semidefinite programming has found many applications in theoretical computer science in the field of optimization [AK07], approximation theory [GW95, GM12], algorithms [AHK05, JLL+20], SoS hierarchy [BHK+19], extension complexity [FMP+15, LRS15]. The class of semidefinite programs that consists of only PSD matrices is an important class of SDPs, termed as *positive semidefinite programs*, which has been used to characterize various quantum interactive proof systems [JUW09, JJUW11, GW13]. Their computational complexity has also received a lot of attention in the past decade [JY11, PT12, AZLO16, JLL+20]. But in several ways, our understanding of spectrahedrons is at an early stage and seriously lags behind our understanding of polytopes. Many basic geometric properties of spetrahedrons, such as dimensions, numbers of connected components, matrix ranks [Viz17] are not well understood, even basic properties such as proving the membership of spetrahedrons for some geometric objects is highly non-trivial [NPS08].

Our main result in this work is PRGs for regular positive spectrahedrons, which we define in Section 1.2. Before stating our main results, we briefly discuss the techniques developed by prior works to construct PRGs for polytopes before discussing the challenges we need to handle here.

[1] We remark that there is still room for improvement in the seed length of the PRG in [OST19].

[2] In this ordering, we say $A \preceq B$ if $B - A$ is positive semidefinite, i.e., all the eigenvalues of $B - A$ are non-negative.

1.1 Prior work and conceptual challenges

1.1.1 Prior work

One of the earliest works that considered fooling threshold functions was by Meka-Zuckerman [MZ13] and [DGJ+10]. A powerful technique that Meka-Zuckerman introduced was a general recipe to construct PRGs for functions f via *invariance principles*. Roughly speaking, an invariance principle for a function $f : \{-1,1\}^n \to \{0,1\}$ states that, the expected value of $f(\mathcal{U}^n)$ (where the input is uniformly random in $\{-1,1\}^n$) is close to the expected value of $f(\mathcal{G}^n)$ (where the input is a standard $\mathcal{G}^n = \mathcal{N}(0,1)^n$ Gaussian). Invariance theorems are generalizations of the classic Berry-Esseen central limit theorem, generally proven using the well-known Lindeberg method [Lin22]. The versatile framework of [MZ13] allows one to use invariance principles along with a few more ingredients to construct PRGs, so the technical challenge is in establishing invariance principles.

Using this framework, Harsha, Klivans and Meka [HKM13] proved an *invariance principle for regular polytopes* (i.e., when the coefficients in (all) the halfspaces are "regular"). The main novelty in their work was the *poly-logarithmic* (in the input parameters) error dependence. In order to prove this, they first proved a general invariance principle for smooth functions (over polytopes). Subsequently they instantiate their invariance principle for the so-called *Bentkus mollifier* [Ben90],[3] crucially relying on the fact that the mollifier has derivatives that scale poly-logarithmic in the input size. Finally in order to go from invariance principles (for the mollifier) to fooling regular polytopes, they need to prove an *anti-concentration of polytopes* in the Gaussian space. For this, they use (as a black-box) a well-known result of Nazarov [Naz03, KOS08], which bounds the *Gaussian surface area* (GSA) of polytopes. Putting together the invariance principle for smooth functions, Bentkus mollifier and Nazarov's bound on GSA, [HKM13] obtained their main results for regular polytopes. We discuss this proof idea in more detail in Section 1.3.

Subsequently, Servedio and Tan [ST17] improved the results of [HKM13] by considering "low-weight" polytopes, which removes the regularity condition (albeit, with the seed length of the PRG in [ST17] depending on the weight). Finally, O'Donnell, Servedio and Tan [OST19] showed how to fool arbitrary polytopes. In [OST19] they still proved a "Boolean-invariance principle" for the Bentkus mollifier, however they bypass the entire Gaussian space (in fact it is a *necessity* to avoid this Gaussian space since standard invariance principles do not hold for non-regular polytopes). Although they bypass the Gaussian intermediate (which is standard in invariance principles), their proof techniques still use the Lindeberg method. Additionally, a crucial tool introduced by them was the Boolean anti-concentration of polytopes, since they can no longer use the GSA bound of Nazarov which used by [HKM13, ST17, CDS19] for *Gaussian* anti-concentration.

1.1.2 PRGs for spectrahedrons: Conceptual challenges

There are two straightforward approaches to constructing PRGs for positive spectrahedrons. The first is to write a spectrahedron as a linear program. Naturally one can approximate a positive-semidefinite constraint $X \succeq 0$ of a $k \times k$ symmetric matrix with exponentially many constraints $z^T X z \geq 0$ for $z \in \mathbb{R}^k$. However the results of [HKM13, OST19] would be moot here since the seed-lengths of their PRGs are poly-logarithmic in the number of constraints, which are polynomial in the dimension k, while our goal it to have seed lengths *poly-logarithmic* in k. The second approach is to use Sylvester's criterion to write out k polynomials of degree at most k (corresponding to the

[3]The Bentkus *mollifier* is a function which provides a "smooth" continuous approximation to the the discrete multivariate indicator function (also referred to as *orthant functions*).

k determinantal representation of the k minors) and one could potentially use PRGs for polynomial threshold functions (PTF). However, finding optimal PRGs for PTFs has remained open for years and the best-known PRGs we have for degree-k PTFs over the Boolean space depends *exponentially* in k [MZ13]. This naturally motivates us to use the "eigenstructure" of $X \succeq 0$ crucially in understanding spectrahedrons. The next line of approach is to use the existing invariance-principle framework of [MZ13] which we overviewed in the previous section, but this opens up a few challenges:

1. **Invariance principles:** Since a spectrahedron naturally deals with eigenvalues of matrices, it is unclear if we could use known invariance principles for spectrahedrons. In fact, we are not even aware of a generalization of the Lindeberg-type argument to show an invariance principle for *spectral functions* (i.e., functions that act on the eigenspectrum of matrices).
2. **Geometric properties:** Prior works of [KOS08, HKM13, ST17, CDS19] crucially used the work of Nazarov [Naz03] which bounds the Gaussian surface area of polytopes in order to prove their anti-concentration. However, spectrahedrons are very poorly understood, and even more basic questions about their average sensitivity, noise sensitivity, surface area are unknown.
3. **Anti-concentration:** An important technique for constructing PRGs using invariance principles requires one to prove *anti-concentration*, i.e., when moving from the smooth mollifier to the orthant functions a crucial ingredient is anti-concentration. It is far from clear if spectrahedrons enjoy such nice properties in either Boolean spaces or Gaussian spaces.

As far as we are aware, none of these questions have been considered for any class of spectrahedrons except polytopes. Our main contribution is to make significant progress in all these questions for the class of positive spectrahedrons.

1.2 Our main result

In order to state our main result we first define PRGs and (τ, M)-regular spectrahedrons. A *pseudorandom generator* is a function $G : \{-1,1\}^r \to \{-1,1\}^n$ and is said to ε-fool a class of functions $\mathcal{F} \subseteq \{f : \{-1,1\}^n \to \{0,1\}\}$ with *seed length* r if it satisfies the following: for every $f \in \mathcal{F}$, we have

$$\left| \Pr_{\boldsymbol{x}\sim\mathcal{U}_n}[f(\boldsymbol{x})=1] - \Pr_{\boldsymbol{y}\sim\mathcal{U}_r}[f(G(\boldsymbol{y}))=1] \right| \leq \varepsilon,$$

where $\mathcal{U}_n$ (resp. $\mathcal{U}_r$) corresponds to uniform distribution over $\{-1,1\}^n$ (resp. $\{-1,1\}^r$). We next define the class of regular positive spectrahedrons. Given $\tau, M > 0$, we say a sequence of $k \times k$ positive semidefinite matrices $(A^1, \ldots, A^n)$ is (τ, M)-*regular* if

$$\mathbb{I} \preceq \sum_{i=1}^{n} \left(A^i\right)^2 \preceq M \cdot \mathbb{I} \quad \text{and} \quad A^i \preceq \tau \cdot \mathbb{I} \text{ for every } i \in [n]. \tag{1}$$

This regularity assumption is a very natural assumption, it says that the *width* of a semidefinite program defined by these matrices is bounded. We remark that our regularity condition naturally extends (and is in fact *less* restrictive) the regularity condition that was used in prior works on fooling halfspaces and polytopes [GOWZ10, DGJ+10, MZ13, HKM13]. In Section 1.5.2 we discuss more about why this notion of regularity is necessary and sufficient for our proof techniques.

A *spectrahedron* $S \subseteq \mathbb{R}^n$ is a feasible region of the convex set $S = \{x \in \mathbb{R}^n : \sum_i x_i A^i \preceq B\}$.[4] We say S is a *positive spectreheron* if either all A^is are positive semidefinite (PSD) or all A^is are negative semidefinite. We say S is a (τ, M)*-regular positive spectrahedron* if $(A^1, \ldots, A^n)$ are (τ, M) regular. It is also natural to consider an *intersection* of positive spectrahedrons $S_1, \ldots, S_t$. However, without loss of generality one can assume that $t = 2$ since one can "pack" all the S_is with PSD matrices into a larger block diagonal matrix with dimension $t \cdot k$ and similarly all the negative semidefinite matrices, so we can always assume we are working with an intersection of two positive spectrahedrons.[5] For simplicity, in the introduction we assume that we are working with a single regular positive spectrahedron here and state our main theorem.

Result 1 (PRG for positive spectrahedrons)**.** *There exists a* PRG $G : \{0,1\}^r \to \{-1,1\}^n$ *with seed length*

$$r = O(\log n \cdot \log k \cdot M \cdot 1/\delta)$$

that δ-fools (τ, M)-regular positive spectrahedrons for $\tau \leq \mathrm{poly}(\delta/(M \cdot \log k))$.

Typically, handling the "regular case" is the first step towards obtaining optimal results in pseudorandom generators for geometric objects and we have accomplished that here for the first time. To prove this theorem, we follow the well-known three-step approach and prove the following:

1. An invariance principle for the Bentkus mollifier of *arbitrary* regular spectrahedrons.
2. Boolean and Gaussian anti-concentration for *positive* regular spectrahedrons.
3. An invariance principle for *positive* regular spectrahedrons

Before proving these statements, we first overview the [HKM13, OST19] approach to proving invariance principles (since our high-level ideas are inspired by their works).

1.3 Sketch of the [HKM13] invariance principle for polytopes

First recall that a polytope is the feasible region of the set $\{x \in \mathbb{R}^n : Wx \leq b\}$ for a fixed $W \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$.[6] We say a polytope is τ-regular if each row W^i satisfies $\|W^i\|_2 = 1$ and $\|W^i\|_4 \leq \tau$. At a high-level the [HKM13] invariance principle states the following:

$$\left| \Pr_{\boldsymbol{x} \sim \mathcal{U}_n}[W\boldsymbol{x} \leq b] - \Pr_{\boldsymbol{g} \sim \mathcal{G}^n}[W\boldsymbol{g} \leq b] \right| \leq \mathrm{poly}(\log n, \tau). \quad (2)$$

To show this, they first express the *orthant* function above (which we denote $\mathcal{O} : \mathbb{R}^n \to \{0,1\}$), as $[W\boldsymbol{x} \leq b] = [W^1\boldsymbol{x} \leq b_1] \cdots [W^n\boldsymbol{x} \leq b_n]$. Given this structure, they now use the well-known Lindeberg method [Lin22] (see [O'D14, Tao10] for a detailed exposition) to move from the uniform distribution over a Boolean space to the Gaussian space. To establish Eq. (2), they follow a three-step approach: (1) First, they prove a version of Eq. (2) for *smooth* functions $\widetilde{\mathcal{O}} : \mathbb{R}^n \to \mathbb{R}$ (i.e., functions who have bounded multivariate derivatives). In particular, they use the Lindeberg method to show that the expected value of $\widetilde{\mathcal{O}}(W\boldsymbol{x})$ for $x \sim \mathcal{U}_n$, is "close" to the expected value of $\widetilde{\mathcal{O}}(W\boldsymbol{g})$ for $\boldsymbol{g} \sim \mathcal{G}^n$. To understand this closeness, they write out $\widetilde{\mathcal{O}}(W\boldsymbol{z})$ using the standard multivariate Taylor

[4]For simplicity in exposition, we assume here that $\|B\| \leq M$ (our main theorems depend on the norm of B).

[5]Crucially we remark that the seed length of our PRG has dependence only logarithmic in k, so even with an intersection of t positive spectrahedrons, the dependence would be logarithmic in t as well.

[6]For simplicity, we assume that the number of constraints and variables are equal. Their analysis is more general.

expansion and bound the distance between $\widetilde{\mathcal{O}}(W\boldsymbol{x})$ and $\widetilde{\mathcal{O}}(W\boldsymbol{g})$ by the higher-order derivatives of the smooth function $\widetilde{\mathcal{O}}$. (2) Second, they observe that a result of Bentkus [Ben90] provides exactly an approximator $\widetilde{\mathcal{O}} : \mathbb{R}^n \to \mathbb{R}$ (which we refer to as the *Bentkus mollifier*) which serves as a *smooth* approximation to the $\{0,1\}$-valued orthant function $\mathcal{O}(x) = [W\boldsymbol{x} \leq b]$. Additionally this mollifier *crucially* satisfies the property that $\|\widetilde{\mathcal{O}}^{(\ell)}\|_1 \leq O\left(\log^\ell n\right)$.[7] (3) So far they established that the Bentkus mollifier (which served as a proxy for $[W\boldsymbol{x} \leq b]$) satisfies an approximate version of Eq. (2). In order to go from being close with respect to this Bentkus mollifier to multidimensional CDF closeness, they prove *Gaussian* anti-concentration of polytopes. For this, they use a result of Nazarov [Naz03] (as a black-box) which shows that the Gaussian surface area of a polytope is $O(\sqrt{\log n})$. These three steps allow them to prove Eq. (2).

1.4 First contribution: Invariance principle for Bentkus mollifier

We begin by defining spectral functions. Let $f : \mathbb{R}^k \to \mathbb{R}$, we say $\psi : \mathsf{Sym}_k \to \mathbb{R}$ is a *spectral function* if $\psi(M) = f(\lambda(M))$ for all $M \in \mathsf{Sym}_k$ where $\lambda(M) = (\lambda_1, \ldots, \lambda_k)$ are the k eigenvalues of M. In other words, a spectral function $\psi(\cdot)$ depends on a function ψ applied to the eigenvalues of its argument. We say f satisfies an invariance principle if

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_n}\left[\psi\left(\sum_i \boldsymbol{x}_i A^i - B\right)\right] \approx_\varepsilon \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{G}^n}\left[\psi\left(\sum_i \boldsymbol{g}_i A^i - B\right)\right],$$

for symmetric matrices $A_1, \ldots, A_n, B$. A conceptual challenge in proving an invariance principle even for smooth spectral functions is that standard Lindeberg-style proofs of invariance theorems use multivariate Taylor series of the mollifier function cannot be used here, since our functions act on the *eigenvalues* of matrices. In the past, there have been various invariance principles [MOO05, Mos08, HKM13, Yao19] but none of them apply here; as far as we are aware invariance principles with non-diagonal A^i, B have not been studied. In this work, we overcome this challenge and adapt the Lindeberg-style proofs of probabilistic invariance principles to prove its analogue for spectral functions.

To this end, recall that we are concerned with spectrahedrons whose feasible regions are given by $\{x \in \mathbb{R}^n : \sum_i x_i A^i \preceq B\}$, which can alternatively be written as $\{x : \lambda_{\max}\left(\sum_i x_i A^i - B\right) \leq 0\}$. So we let our spectral function $f : \mathbb{R}^k \to \mathbb{R}$ to be $f(\lambda) = [\max_i \lambda_i \leq 0]$ (recall that although our spectrahedron acts on n bits on which we want to prove an invariance principle, our spectral function acts only on the k eigenvalues). For this function, we can still use the Bentkus mollifier $\widetilde{\mathcal{O}} : \mathbb{R}^k \to \mathbb{R}$ as a smooth approximation to f.[8] So our first main contribution is to prove an invariance principle for the Bentkus mollifier applied to the spectrum of matrices. We remark that in contrast to [HKM13], we do not prove a general invariance principle for spectral functions, instead our spectral function is tailored for the Bentkus mollifier (which is also the case for [OST19]).

Fréchet derivatives. Since our Bentkus mollifier is acting on the eigenspectrum of matrices, instead of multivariate Taylor expansion, we adopt *Fréchet derivatives*, a notion of derivatives that is studied in Banach spaces. Unfortunately, Fréchet series (in contrast to standard multivariate series) are still not well understood. In fact even basic properties such as continuity, Lipschitz continuity, differentiability, continuous differentiability, were only proven in the last three decades [BSS98, Lew96, BS99, CQT03], which have been well-known for centuries in standard calculus. In particular,

[7] Here $\|f^{(\ell)}\|_1$ is the 1-norm of the coefficients in the ℓ-th derivative. In [HKM13], they care about $\|f^{(4)}\|_1 = \max_x \sum_{p,q,r,s} |\partial_p\partial_q\partial_r\partial_s f(x)|$.

[8] In fact our analysis can allow arbitrary orthant functions which can be approximated by a Bentkus mollifier.

even a succinct representation of high-order Fréchet derivatives [Sen07, AS10, AS12, AS16] for *spectral* functions only appeared in the last decade.

Fortunately for us, Sendov [Sen07] provided a tensorial representation of high-order Fréchet series for spectral functions which we employ to analyze the Fréchet derivatives of the Bentkus mollifier. The challenge is in bounding the 3-tensors that appears in Sendov's theorem, which produce 6 terms corresponding to different permutations of the tensors after simplification. Three of these 6 terms can simply be upper bounded by $\|\widetilde{\mathcal{O}}^{(3)}\|_1$ which we know to be small for the Bentkus mollifier. We remark that these are exactly, and the only, terms that appear in the standard invariance principle proofs for linear forms. Intuitively this is not surprising since the first three terms simply correspond to the case when the A^i, B are *diagonal* which reduces a spectrahedron to a polytope. However, bounding the remaining terms is highly non-trivial and one of our technical contributions is in showing these remaining terms are bounded for the Bentkus mollifier.

Bounding derivatives and obtaining invariance principle. Bounding these last three terms of the 3-tensors significantly deviates from the analysis of [HKM13] since we need to deal with off-diagonal entries of matrices which is unique to the matrix-spectrahedron case and is not faced in [HKM13, ST17, OST19]. To bound this, we use several properties of Fréchet derivatives such as, mean value theorems for Fréchet derivatives, divided differences representations of Fréchet derivatives [BLZ05], and Dyson's theorem [Bha13] which provides a useful integral expression for Fréchet derivatives (using the structure of the mollifier). More importantly, since we work with the Bentkus mollifier [Ben90], we completely open up the Bentkus black-box and show various analytic properties of this mollifier $\widetilde{\mathcal{O}}$ in order to prove that our Fréchet derivatives are bounded.

In order to go from bounded third-order Fréchet derivatives to a final invariance principle, we still need to borrow some results from random matrix theory to upper bound the moments of $\sum_i \boldsymbol{x}_i A^i$. Although, the concentration of $\sum_i \boldsymbol{x}_i A^i$ for uniformly random $\boldsymbol{x} \sim \mathcal{U}_n$ is well-studied by standard matrix Chernoff bounds [Tro15], we need better concentration of this random matrix variable at higher Schatten norms. For the diagonal polytope case [HKM13] used the standard hypercontractivity and [OST19] used Rosenthal's inequality. Fortunately for us, a matrix-version of Rosenthal's inequality [MJC+14] was proven a few years back and we use it to conclude our proof (in fact we also crucially rely on this inequality to construct our PRG). Putting everything together we obtain our main invariance principle for the Bentkus mollifier applied as a spectral function

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_n}\left[\widetilde{\mathcal{O}}\left(\sum_{i=1}^n \boldsymbol{x}_i A^i - B\right)\right] - \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{G}^n}\left[\widetilde{\mathcal{O}}\left(\sum_{i=1}^n \boldsymbol{g}_i A^i - B\right)\right]\right| \leq \mathrm{poly}(\log k, M, \tau). \tag{3}$$

We remark that the invariance principle above does not assume the positivity of the matrices. We believe this is a necessity for future work on fooling arbitrary spectrahedrons.

1.5 Second contribution: Geometric properties of positive spectrahedrons

Even with an invariance principle in hand, we are faced with the same challenges as [HKM13, ST17, OST19] to show an anti-concentration statement. Recall that our goal is to show that for a (τ, M)-regular positive spectrahedron S, the expected value of the *indicator function* $[x \in S]$ for $x \sim \mathcal{U}_n$ is close to the expected value of $[\boldsymbol{g} \in S]$ for $\boldsymbol{g} \sim \mathcal{G}^n$. This is "almost" what we showed in the previous section except that the Bentkus mollifier $\widetilde{\mathcal{O}}$ in Eq. (3) is replaced by the orthant indicator function $f(x) = [\max_i x_i \leq 0]$. In order to move from the smooth function distance to CDF distance, one particular approach taken by [HKM13, ST17, CDS19] is to use geometric properties of polytopes, and as far as we are aware this is widely open for spectrahedrons.

1.5.1 Geometric properties

A well-known theorem of Ball [Bal93] shows an upper bound of $O(n^{1/4})$ on the Gaussian surface area (GSA) of arbitrary n-dimensional convex object. Crucially in the works of [HKM13, ST17] they used an improvement of Ball's theorem by Nazarov [Naz03] who showed that the GSA of k-facet polytopes is $O(\sqrt{\log k})$. This logarithmic-upper bound on GSA allows [KOS08, HKM13, ST17, CDS19] to obtain invariance principles, learning algorithms and pseudorandom generators that depend poly-logarithmic in k. In contrast, for our setting it is unclear what is the GSA of spectrahedrons. Clearly, Ball's theorem gives an upper bound of $O(n^{1/4})$ on GSA for us. Apart from that, spectrahedrons are poorly understood. Below we prove an upper bound of $O(1)$ on the GSA of *positive* spectrahedrons.

Result 2 (Geometric properties of positive spectrahedrons)**.** *Let S be a positive spectrahedron and consider $F : \{-1,1\}^n \to \{0,1\}$ defined as $F(x) = [x \in S]$. The average sensitivity of F is $O(\sqrt{n})$, the ε-Boolean noise sensitivity of F is $O(\sqrt{\varepsilon})$, and the Gaussian surface area is $O(1)$.*

We remark that the noise-sensitivity statement we have above can be viewed as a "positive-matrix-analogue" version of the well-known Peres's theorem [Per04]. In order to prove this statement, we first observe that the average sensitivity of F being $O(\sqrt{n})$ immediately follows by the observation that positive spectrahedrons correspond to *unate functions* and Kane [Kan14a] showed $\mathsf{AS}(f) \le O(\sqrt{n})$ if f is unate (and a similar statement is known to be false for noise sensitivity). One issue we need to handle when translating between noise sensitivity and average sensitivity is the following: in the standard technique of [Per04, DGJ+10, Kan14a], one upper bounds the ε-noise sensitivity of a function f by "bucketing" the input variables into $m = O(1/\varepsilon)$ buckets $B_1, \ldots, B_m$ and reduces the function $f : \{-1,1\}^n \to \{-1,1\}$ to a function $g : \{-1,1\}^m \to \{-1,1\}$ defined as $g(b) = \sum_{\ell=1}^m b_i \sum_{i \in B_\ell} z_i A^i$ (for uniformly random z). One then upper bounds $\mathsf{NS}_\varepsilon(f)$ using $\mathsf{AS}(g)$ (up to a factor ε). Clearly when using this technique to bound ε-noise sensitivity of halfspaces, both f, g are intersections of halfspaces and one can upper bound the average sensitivity of g using Kane's result [Kan14a] to be $O(\sqrt{m})$. However in our setting if f is an indicator of a *positive* spectrahedron, then g no longer needs to be an indicator of a *positive* spectrahedron since $\sum_{i \in B_\ell} z_i A^i$ need not even be either a positive semidefinite matrix or a negative semidefinite matrix. We overcome this by modifying the bucketing procedure of [DGJ+10] to ensure g is an indicator of a unate function. However, in the process case we end up upper bounding $\mathsf{NS}_\varepsilon(f)$ by the "average 2-sensitivity" of g. We extend the results of Kane [Kan14a] by showing that even the "average 2-sensitivity" of g is small for our setting. Finally, to move from an upper bound on ε-noise sensitivity to Gaussian surface area, we use standard folklore results [DHK+10, Kan11a, Bal13].

1.5.2 Boolean anti-concentration

Gaussian anti-concentration of polytopes directly follows from the fact that the Gaussian surface area of polytopes is bounded since its surface has only finite normed vectors. This is crucially used in [HKM13, ST17, CDS19]. However, it is not clear how to obtain Gaussian anti-concentration of positive spectrahedrons even with bounded Gaussian surface area (as proven in Result 2) due to its complicated geometric structures. Here, to move from mollifier-closeness to CDF closeness, we prove a *Boolean* anti-concentration for positive spectrahedrons, which is in fact *stronger* than Gaussian anti-concentration, inspired by the Boolean anti-concentration for polytopes in [OST19].

Regularity condition. Before explaining the Boolean anti-concentration, we need to revisit the *regularity* condition, which is also used for polytopes. In [HKM13, ST17], it is assumed that every halfspace (or row in the matrix W) satisfies $\|W^i\|_2 = 1$ and $\|W^i\|_4 \leq \tau$. One important question is: what is a regularity assumption for spectrahedrons and for which assumptions can we *show* anti-concentration? A natural possibility is to see if Nazarov's result [Naz03] holds for spectrahedrons (i.e., show anti-concentration in the weaker *Gaussian* setting). To the best of our knowledge, this has firstly not been studied in literature. Moreover, it is not hard to see that, in order for the proof of Nazarov to work for spectrahedrons, one can make a very *strong* assumption that *every* A^i satisfies $\lambda_{\min}(A^i) \geq 1$. However, this seems to significantly restrict the class of spectrahedrons.

In order to resolve this, we propose (τ, M)-regularity as defined in Eq. (1) and prove a stronger statement, i.e., Boolean anti-concentration for (τ, M)-regular positive spectrahedrons. We use this statement to go from closeness between the mollifier $\widetilde{\mathcal{O}}\left(\sum_i \boldsymbol{x}_i A^i - B\right)$ and $\widetilde{\mathcal{O}}\left(\sum_i \boldsymbol{g}_i A^i - B\right)$ (which we already established in Eq. (3)) to closeness between $\left[\sum_i \boldsymbol{x}_i A^i \preceq B\right]$ and $\left[\sum_i \boldsymbol{g}_i A^i \preceq B\right]$. In this direction, we prove a Littlewood-Offord type theorem for positive spectrahedrons.

Result 3 (Littlewood-Offord for positive spectrahedrons)**.** *If $(A^1, \ldots, A^n)$ are (τ, M)-regular. Then every Λ, we have*

$$\Pr_{\boldsymbol{x}\sim\mathcal{U}_n}\left[\lambda_{\max}\left(\sum_i \boldsymbol{x}_i A^i - B\right) \in [-\Lambda, \Lambda]\right] \leq O(\Lambda).$$

The classic Littlewood-Offord theorem [LO39, Erd45] anti-concentration inequality for a half-space $w \in \mathbb{R}^n$ (satisfying $|w_i| \geq 1$) and $\alpha \in \mathbb{R}$ proves a bound on the probability that $\sum_i w_i \boldsymbol{x}_i \in [\alpha, \alpha + 2]$ (where $\boldsymbol{x} \sim \mathcal{U}_n$). In [OST19] they generalized this for *intersections* of halfspaces and in the result above we show a matrix-version of Littlewood-Offord theorem. Intuitively, our statement shows the largest eigenvalue of a positive spectrahedron cannot all be very-concentrated in a small region (i.e., small eigenvalue regions have small measure over the Boolean cube).

The proof of our result is similar to the proofs in [Kan14a, OST19] which show anti-concentration for intersections of unate functions (which is the case for positive spectrahedrons). There are a couple of subtleties for us: in [OST19], they perform random "bucketing" of the coordinates in a polytope and show that with high probability, each bucket has "significant" weight, which follows immediately from the Paley-Zygmund inequality. However, for us, firstly random bucketing does not produce a positive spectrahedron (the same issue which we faced in Theorem 2), so instead we need to bucket in a non-standard way to go from a positive spectrahedron to a bucket which corresponds to a unate function. Next, to show that each bucket has significant weight (which in our case corresponds to large smallest eigenvalue), we invoke the matrix Chernoff bound for negatively correlated variables. We remark that higher-dimensional extensions of the Littlewood-Offord theorem [FF88, TV12] do not talk of eigenspectrum of matrices and differs from our result.

Using the standard bits-to-Gaussians trick, this also gives us Gaussian anti-concentration (i.e., the positive spectrahedrons analogue of Nazarov's result [Naz03] which is unknown as far as we are aware). Putting this together with our invariance principle statement we obtain our main result.

Result 4 (Fooling positive spectrahedrons)**.** *For every (τ, M)-regular positive spectrahedron S,*

$$\Big|\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mathcal{U}_n}[\boldsymbol{x} \in S] - \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{G}^n}[\boldsymbol{g} \in S]\Big| \leq \mathrm{poly}(M, \log k, \tau). \tag{4}$$

Apart from the applications of constructing pseudorandom generators (which we discuss in the next section) we believe that our invariance principle for the Bentkus mollifier of *arbitrary* spectrahedrons, opening up the Bentkus mollifier (i.e., understanding the Bentkus functions which were almost used as a black-box in [HKM13, ST17, OST19]), the Littlewood-Offord theorem for positive spectrahedrons, Gaussian surface area of positive spectrahedrons, could be of independent interest.

1.6 Applications

1.6.1 Pseudorandom generators

We now briefly discuss how to use the invariance principle to obtain our pseudorandom generator. Our construction is based on the Meka-Zuckerman [MZ13] PRG construction for fooling halfspaces. We note in the passing that this same PRG (with different parameters) was also used by [HKM13, ST17] and slight modification of it by [OST19]. We omit the details of the PRG construction here referring the interested reader to Section 6.3 for an explicit construction.

One subtlety in order to go from invariance principle to fooling the MZ-generator is the following: recall that our invariance principles showed that expected value under the uniform distribution was close to the expected value under the Gaussian distribution. However, in order to fool the MZ-generator one needs to show that the invariance principle proofs holds also for k-wise independent distributions. In this direction, we use a neat trick from [OST19] that shows that in order to show invariance principles for k-wise independent distributions, it suffices to show just Boolean anti-concentration, and second we crucially use the fact that the matrix Rosenthal inequality can be derandomized by analyzing its the original proof. Put together, this shows that our invariance principle proof holds for k-wise independent distributions and gives us our main PRG result.

Result 5 (PRG for positive spectrahedrons)**.** *Let S be a (τ, M)-regular positive spectrahedron. There exists a* PRG $G : \{0,1\}^r \to \{-1,1\}^n$ *with* $r = (\log n) \cdot \mathrm{poly}(\log k, M, 1/\delta)$ *that δ-fools S with respect to the uniform distribution for every* $\tau \leq \mathrm{poly}(\delta/(\log k \cdot M))$.

1.6.2 Learning theory

Learning geometric objects is a fundamental problem in computational learning theory. An application of upper bounding noise sensitivity or Gaussian surface area of spectrahedrons (in Theorem 2) is in agnostic learning. The agnostic learning framework introduced by [KSS94, Hau92] is the following: let $\mathcal{C} \subseteq \{c : \{-1,1\}^n \to \{0,1\}\}$ be a concept class and $\mathcal{D} : \{-1,1\}^n \times \{0,1\} \to [0,1]$ be a distribution. Define $\mathsf{opt}(\mathcal{C}) = \min_{c\in\mathcal{C}} \Pr_{(x,b)\sim\mathcal{D}}[c(x) \neq b]$, i.e., what is the *best* approximation to $\mathcal{D}$ from within the concept class. The goal of an agnostic learner is the following: given many samples $(x,b) \sim \mathcal{D}$, the goal of a learner is to produce a hypothesis $h : \{-1,1\}^n \to \{0,1\}$ which satisfies

$$\Pr_{(x,b)\sim\mathcal{D}}[h(x) \neq b] \leq \mathsf{opt}(\mathcal{C}) + \varepsilon.$$

Note that if $\mathsf{opt}(\mathcal{C}) = 0$, this is the standard PAC learning framework and agnostic learning models learnability under adversarial noise. A natural restriction of this model is when the marginal of $\mathcal{D}$ on the first n bits is the uniform distribution on $\{0,1\}^n$. It is a folklore result [KOS04] that a function f having low noise sensitivity can be approximated by low-degree polynomials (see [HKM13, Lemma 2.7] for an explicit statement). Furthermore, the well-known L1-polynomial regression algorithm [KKMS08] shows how to learn low-degree polynomials in the agnostic framework. Putting these two connections together gives us the following theorem.

Result 6 (Learning positive spectrahedrons)**.** *The concept class of positive spectrahedrons (in n variables with $k \times k$ symmetric matrices) can be agnostically learned under the uniform distribution in time $n^{O(\log k)}$ for every constant error parameter.*

The previous best known result [KOS08] for learning positive spectrahedrons even in the PAC model was $2^{O(n^{1/4})}$ (as far as we are aware); our result provides a substantially better complexity.

1.6.3 Discrepancy sets for spectrahedrons

Understanding discrepancy sets for convex objects is a fundamentally important problem in the fields of convex geometry, optimization, and a range of other areas. Prior works of [HKM13, ST17, OST19] constructed such discrepancy sets for polytopes, but a natural question is to extend their construction to spectrahedrons. In our context, one application of our main result can be viewed as the following: consider the set of all possible positive spectrahedrons (over the Boolean cube) $S = \{x \in \{-1,1\}^n : \sum_i x_i A^i \preceq B\}$, then can we construct a *small* subset of the Boolean cube $\{-1,1\}^n$ such that this set δ-approximates the $\{-1,1\}^n$-volume of every positive spectrahedron? One way to construct such a set is to construct a PRG for the class of functions. So an immediate corollary of our PRG for positive spectrahedrons is the following theorem.[9]

Result 7 (Discrepancy set for positive spectrahedrons)**.** *There is a deterministic algorithm which, given a (τ, M)-regular positive spectrahedron S, runs in time $\exp(\log n, \log k, M, 1/\delta)$ and outputs a δ-approximation of the number of points in $\{-1,1\}^n$ contained in S as long as $\tau \le \text{poly}(\delta/(M \log k))$.*

1.6.4 Intersection of (structured) polynomial threshold functions

Constructing PRGs for PTFs has received a lot of attention. However, the best known seed length for fooling a degree-k PTF on n bits scales as $O(\log n \cdot 2^k)$ (over the Boolean space). A simple observation we make is that fooling spectrahedrons (on n bits with $k \times k$ matrices) can be in fact be viewed as the more challenging task of fooling an *intersection* of k many degree-k PTFs.

Recall that a spectrahedron is given by $S = \{x \in \mathbb{R}^n : B - \sum_i x_i A^i \succeq 0\}$. Without loss of generality, we may assume that the measure of x satisfying $\det\left(\sum_i B - x_i A^i\right) = 0$ is zero. Sylvester's criterion implies that a matrix M (which in our case is $B - \sum_i x_i A^i$) is positive definite *if and only if* the determinant of the k principle minors of M are positive. Hence, an alternate characterization of S is the set of $x \in \mathbb{R}^n$ for which

$$S = \bigwedge_{r=1}^{k} \left[\mathsf{det}\left(B - \sum_i x_i A^i\right)_{r \times r} > 0 \right] = \bigwedge_{r=1}^{k} \text{sign}[p_r(x)]$$

modulo a zero-measure set, where $M_{r\times r}$ means the top left $r \times r$ principle minor of M. Clearly each determinantal expression produces a polynomial p_r of degree at most r. So, our main result about fooling S, shows that there is a *structured* class of *intersections* of degree-k PTFs (i.e., the class of polynomials which can be written as in terms of the above) which can be fooled by a PRG with seed length $O(\log n \cdot \log k \cdot M/\delta)$, which is exponentially better than using existing PRGs for PTFs.

We remark that apriori, it is not even clear why should an *arbitrary* polynomial even correspond to a spectrahedron as above? However, a well-known result of [HMV06, GM12] states that an arbitrary degree-d polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$ with real coefficients has a symmetric *determinantal representation*,[10] i.e., there exists symmetric $A^0, A^1, \ldots, A^n$ such that

$$p(x_1, \ldots, x_n) = \mathsf{det}\left(A^0 + \sum_i x_i A^i\right).$$

where $A^i \in \mathsf{Sym}\binom{n+d}{d}$. So, if we could fool arbitrary spectrahedrons that might be a promising avenue to fool PTFs and intersections of PTFs.

[9]We remark that counting integer solutions to positive spectrahedrons is not as *naturally motivated* as that for polytopes, but nevertheless understanding discrepancy sets for geometric objects is a fundamental question.

[10]See [Qua12] for a simple linear algebraic proof of this statement.

1.7 Future work

Our work opens this new line of research into understanding PRGs for spectrahedrons with several novel techniques. This raises several questions for future work.

1. Can we remove regularity for positive spectrahedrons? One of the crucial techniques that Servedio and Tan [ST17] introduced (inspired by a prior work of Servedio [Ser06]) was decomposing a polytope into head and tail variables (i.e., tail coordinates in a halfspace which satisfy regularity and head coordinates are the dominant variables). They express the head variables as CNF, use the result of Bazzi [Baz09] to fool the head variables and invariance principles for tail variables. However, in our setting breaking up a single spectrahedron into head and tail variables is unclear and even if possible, what is the analogue of the CNF for our setting?

2. Can we fool arbitrary spectrahedrons? Besides the difficulty in removing the regularity condition, another fundamental barrier we face here is, anti-concentration. What is the Gaussian surface area of a spectrahedron, even this is unknown (as far as we are aware). Our techniques such as bucketing, using Kane's result [Kan14a], and Boolean anti-concentration [OST19] crucially use the assumption of positivity. Going beyond this, might require new understanding on the geometric structures (like average sensitivity, noise sensitivity) about arbitrary spectrahedrons.

3. A general invariance principle for spectral functions? Here, we showed our invariance principle specifically for the Bentkus mollifier. However, like the result of [HKM13] can we prove a general invariance principle for arbitrary smooth spectral functions? Given the applications of invariance principles, they are now considered to be powerful techniques in computational complexity theory. Having an invariance principle for spectral functions could find more applications such as deciding noisy entangled quantum games [Yao19].

4. Can we fool spectrahedral caps? Let $S_{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ denote the n-dimensional sphere, then a *spectrahedral cap* is the set of S_{n-1} that is "cut" by a spectrahedron, i.e., for a spectrahedron S, we define the spectrahedral cap C_S as $C_S = S_{n-1} \cap S$. In the polytope-setting, fooling spherical caps has received a lot of attention classically [HKM13, KM15] (with almost optimal seed length PRGs). Can we similarly fool spectrahedral caps?

5. Fooling polynomial threshold functions? Can we make progress in finding better PRGs for PTFs using techniques we developed here for fooling arbitrary spectrahedrons?

Acknowledgements. We thank Jop Briët and Minglong Qin for helpful comments. This collaboration earlier faced some bureaucratic issues. We are deeply grateful for the support from Jelani Nelson, Kewen Wu, Yitong Yin and others in the TCS community. P.Y. was supported by the National Key R&D Program of China 2018YFB1003202, National Natural Science Foundation of China (Grant No. 61972191), the Program for Innovative Talents and Entrepreneur in Jiangsu, the Fundamental Research Funds for the Central Universities 0202/14380068 and Anhui Initiative in Quantum Information Technologies Grant No. AHY150100. Part of the work was done when P.Y. and S.A. were participating in the program "Quantum Wave in Computing" held at Simons Institute for the Theory for Computing.

Organization. In Section 2 we introduce all the mathematical aspects which we use in this paper as well as state various lemmas in random matrix theory and multidimensional calculus. In Section 3, we introduce the Bentkus mollifier and discuss various properties. In Section 4 we state our main theorem regarding spectral derivatives of smooth functions and go on to bound the spectral derivatives for the Bentkus function (proving a technical lemma in Appendix A). In

Section 5 we prove an upper bound on the Gaussian surface area of positive spectrahedrons as well as our Littlewood-Offord theorem for this class. In Section 6 we prove our invariance principle theorem and go on to construct a pseudorandom generator for the class of positive spectrahedrons.

2 Preliminaries

For an integer $n \geq 1$, let $[n]$ represent the sets $\{1, \ldots, n\}$. Given a finite set $\mathcal{X}$ and a natural number k, let $\mathcal{X}^k$ be the set $\mathcal{X} \times \cdots \times \mathcal{X}$, the Cartesian product of $\mathcal{X}$, k times. Given $a = (a_1, \ldots, a_k)$ and a set $S \subseteq [k]$, we write a_S and a_{-S} to represent the projections of a to the coordinates specified in S and the on coordinates outside S, respectively. For any $i \in [k]$, a_{-i} represents $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$ and $a_{<i}$ represents $a_1, \ldots, a_{i-1}$. $a_{\leq i}, a_{>i}, a_{\geq i}$ are defined similarly. Let μ be a probability distribution on $\mathcal{X}$, and $\mu(x)$ represent the probability of $x \in \mathcal{X}$ according to μ. Let X be a random variable distributed according to μ. We use the same symbol to represent a random variable and its distribution whenever it is clear from the context. The expectation of a function f on $\mathcal{X}$ is defined as $\mathbb{E}[f(X)] = \mathbb{E}_{\boldsymbol{x} \sim X}[f(\boldsymbol{x})] = \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot f(x) = \sum_x \mu(x) \cdot f(x)$, where $\boldsymbol{x} \sim X$ represents that $\boldsymbol{x}$ is drawn according to X. For any event $\mathcal{E}_x$ on x, $[\mathcal{E}(x)]$ represents the indicator function of $\mathcal{E}$. In this paper, the lower-cased letters in bold $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \cdots$ are reserved for random variables.

Distributions. Throughout, we denote $\mathcal{G}$ (where $\mathcal{G} = \mathcal{N}(0,1)$) to be the standard univariate spherical Gaussian distribution over $\mathbb{R}$ with mean 0 and variance 1. We denote $\mathcal{U}_n$ to be the uniform distribution on $\{-1,1\}^n$. We say a sequence of random variables $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is t-wise uniform if any subset of X of size t is uniformly distributed (observe that the uniform distribution is clearly t-wise independent for every $t \geq 1$). A distribution $\mathcal{H}$ on functions $[n] \to [m]$ is said to be an r-wise uniform hash family if for $\boldsymbol{h} \sim \mathcal{H}$, $(\boldsymbol{h}(1), \ldots, \boldsymbol{h}(n))$ is r-wise uniform.

2.1 Derivatives and multidimensional Taylor expansion

For any $f : \mathbb{R} \to \mathbb{R}$ in $\mathcal{C}^d$, which is the set of all real functions that are d-time differentiable, we use $f^{(d)}$ to denote the d-th derivative of f. Given a function $F : \mathbb{R}^k \to \mathbb{R}$ and a k-dimensional multi-index $\alpha = (\alpha_1, \ldots, \alpha_m) \in \mathbb{N}^k$, $\partial_\alpha F$ denotes the mixed partial derivative taken α_i times in the i-th coordinate.

Fact 1. *Let $k \in \mathbb{N}$ and $f : \mathbb{R}^k \to \mathbb{R}$ be a $\mathcal{C}^d$ function. Then for all $x, y \in \mathbb{R}^k$,*

$$f(x+y) = \sum_{\alpha \in \mathbb{N}^k : |\alpha| \leq d-1} \frac{\partial_\alpha f(x)}{\alpha!} \prod_{i=1}^m y_i^{\alpha_i} + \mathsf{err}(x,y),$$

where $\alpha! = \alpha_1! \cdots \alpha_m!, |\alpha| = \sum_i \alpha_i$ and

$$|\mathsf{err}(x,y)| \leq \sup_{v \in \mathbb{R}^k} \sum_{\alpha \in \mathbb{N}^k : |\alpha| = d} |\partial_\alpha f(v)| \max_i |y_i|^d.$$

For a t-time differentiable function $f : \mathbb{R}^k \to \mathbb{R}$ and $s \leq t$, define

$$\|f^{(s)}\|_1 = \max\left\{ \sum_{p_1, p_2, \ldots, p_s \in [k]} |\partial_{p_1} \cdots \partial_{p_s} f(x)| : x \in \mathbb{R}^k \right\}$$

Definition 2. *Let $f : \mathbb{R} \to \mathbb{R}$. For any distinct inputs $x_1, \ldots, x_n \in \mathbb{R}$, the divided difference is defined recursively as follows.*

$$f^{[0]} = f,$$
$$f^{[i]}(x_1, \ldots, x_{i+1}) = \frac{f^{[i]}(x_1, \ldots, x_{i-1}, x_i) - f^{[i]}(x_1, \ldots, x_{i-1}, x_{i+1})}{x_i - x_{i+1}}.$$

For other values of $x_1, \ldots, x_{i+1}$, $f^{[i]}$ is defined by continuous extension.

Fact 3 (Mean value theorem for divided difference [Boo05])**.** *For any $f \in \mathcal{C}^n$ and any $x_1, \ldots, x_{n+1}$, there exists $\xi \in (\min\{x_1, \ldots, x_{n+1}\}, \max\{x_1, \ldots, x_{n+1}\})$ such that*

$$f^{[n]}(x_1, \ldots, x_{n+1}) = \frac{f^{(n)}(\xi)}{n!}.$$

2.2 Combinatorial properties of Boolean functions

Let $f : \{0,1\}^n \to \{0,1\}$, $g : \mathbb{R}^n \to \{0,1\}$ and S be a Borel set in $\mathbb{R}^n$. We define the following combinatorial properties of Boolean-valued functions f, g.

1. Average sensitivity: $\mathsf{AS}(f) = \sum_{i=1}^n \Pr_{\boldsymbol{x}}[f(\boldsymbol{x}) \neq f(\boldsymbol{x} \oplus e_i)]$, where the probability is taken uniformly in $\{0,1\}^n$.

2. ε-Noise sensitivity: $\mathsf{NS}_\varepsilon(f) = \Pr_{\boldsymbol{x},\boldsymbol{y}}[f(\boldsymbol{x}) \neq f(\boldsymbol{y})]$ where the probability is taken according to the distribution: $\boldsymbol{x}$ is uniformly random in $\{0,1\}^n$ and $\boldsymbol{y}$ is obtained from $\boldsymbol{x}$ by independently flipping each $\boldsymbol{x}_i$ with probability ε.

3. Gaussian noise sensitivity: $\mathsf{GNS}_\varepsilon(g) = \Pr_{\boldsymbol{x},\boldsymbol{z}}[g(\boldsymbol{x}) \neq g(\boldsymbol{y})]$ where $\boldsymbol{x}, \boldsymbol{z}$ are independent and random Gaussian vectors in $\mathcal{G}^n$, and $\boldsymbol{y} = (1-\varepsilon)\boldsymbol{x} + \sqrt{2\varepsilon - \varepsilon^2}\boldsymbol{z}$.

4. Gaussian surface area: $\mathsf{GSA}(S) = \liminf_{\delta \to 0} \frac{\mathcal{G}^n(S_\delta \setminus S)}{\delta}$ where $S_\delta = \{x : \mathrm{dist}(x, S) \leq \delta\}$ denotes the δ-neighborhood of S under Euclidean distance.

We refer interested readers to [O'D14] for more on these parameters and their applications to analysis of Boolean functions.

2.3 Matrix analysis and Random matrices

For any integer $k > 0$, we use Mat_k and Sym_k to represent the set of $k \times k$ real matrices and symmetric matrices, respectively. For any matrix X, $\|X\|_p$ represents the Schattern p-norm of X and $\|X\|$ represents the spectral norm of X. $\mathbb{I}_k$ represents a $k \times k$ identity matrix. The subscript k may be omitted whenever the dimension is clear from the context. We need the following results in matrix analysis.

Fact 4. *[Bha00] For any $k \times k$ real symmetric matrix A, let B be its upper triangle part of A. Namely $B_{i,j} = A_{i,j}$ if $i \leq j$ and is 0 otherwise. Then $\|B\| \leq \frac{\ln k}{\pi}\|A\|$.*

Fact 5. *[Tro12, Theorem 1.1] Let $n, k \geq 1$ be integers and $X_1, \ldots, X_n$ be independent random $k \times k$ real symmetric matrices satisfy $0 \preceq X_i \preceq R$ for $i \in [n]$. Set*

$$\mu = \lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]\right).$$

Then

$$\Pr\left[\lambda_{\min}\left(\sum_{i=1}^{n} X_i\right) \le (1-\delta)\,\mu\right] \le k \cdot \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu/R}$$

for every $\delta \in [0,1)$.

Fact 6. *For every integer* $m \ge 1$ *and* $A_1, \ldots, A_n \in \mathsf{Sym}_k$ *it holds that*

$$\mathbb{E}\left[\|\textstyle\sum_i \boldsymbol{g}_i A^i\|^m\right] \le (1 + 2m\lceil\log k\rceil)^{m/2} \cdot \|\textstyle\sum_i (A^i)^2\|^{m/2}$$

and

$$\mathbb{E}\left[\|\textstyle\sum_i \boldsymbol{x}_i A^i\|^m\right] \le (1 + 2m\lceil\log k\rceil)^{m/2} \cdot \|\textstyle\sum_i (A^i)^2\|^{m/2},$$

where the expectations are taken over $\boldsymbol{x} \sim \mathcal{U}_n$ *and* $\boldsymbol{g} \sim \mathcal{G}^n$. *Additionally, the second inequality still holds if* $\boldsymbol{x}$ *is* $2m\lceil\log k\rceil$*-wise uniform.*

Proof. It suffices to prove the second inequality as the first one follows by the standard bits-to-Gaussians tricks [O'D14, Chapter 11]. Let $B = \sum_i \boldsymbol{x}_i A^i$ where $\boldsymbol{x} \sim \mathcal{U}_n$. The proof closely follows the argument in [Tro16], where Tropp proved the case that $m = 1$. For any integer $p \ge 1$, it is proved in [Tro16, Eqs. (4.9,4.11)] that

$$\mathbb{E}\left[\mathrm{Tr}\ B^{2p}\right] \le k \cdot \left(\frac{2p+1}{e}\right)^p \cdot \|\textstyle\sum_i (A^i)^2\|^p.$$

Thus

$$\mathbb{E}\left[\|B\|^m\right] \le \mathbb{E}\left[\mathrm{Tr}\ B^{2pm}\right]^{1/2p} \le k^{1/2p} \cdot \left(\frac{2pm+1}{e}\right)^{m/2} \cdot \|\textstyle\sum_i (A^i)^2\|^{m/2}.$$

Setting $p = \lceil\log k\rceil$, we conclude the result. □

Fact 7 (Matrix Rosenthal inequality [MJC+14, Corollary 7.4])**.** *Let* $X_1, \ldots, X_n$ *be centered, independent random real symmetric matrices. Then*

$$\left(\mathbb{E}\left[\|\textstyle\sum_i X_i\|_{4p}^{4p}\right]\right)^{\frac{1}{4p}} \le \sqrt{4p-1}\,\|\left(\textstyle\sum_i \mathbb{E}\left[X_i^2\right]\right)^{\frac{1}{2}}\|_{4p} + (4p-1)\left(\sum_i \mathbb{E}\left[\|X_i\|_{4p}^{4p}\right]\right)^{\frac{1}{4p}}.$$

This inequality still holds if $X_1, \ldots, X_n$ *are* $4p$*-wise independent.*

2.4 Matrix functions, spectral functions and Fréchet derivatives

Let $f : \mathbb{R}^k \to \mathbb{R}$ and $\lambda : \mathsf{Sym}_k \to \mathbb{R}^k$ where $\lambda(X) = (\lambda_1(X), \ldots, \lambda_k(X))$ are the eigenvalues of M sorted in a non-increasing order. We refer to $\lambda_{\max} = \lambda_1$ interchangeably. Let $F = f \circ \lambda : \mathsf{Sym}_k \to \mathbb{R}$.

If $f : \mathbb{R} \to \mathbb{R}$ is an analytic function in $\mathbb{R}$, namely its Taylor series converges in $\mathbb{R}$, we define $f(X)$ for general matrices using its Taylor expansion. It is not hard to see that the Taylor series still converges with matrix inputs. If X is symmetric with a spectral decomposition $X = UDU^T$, where $D = \mathrm{diag}\left(\lambda_1(X), \ldots, \lambda_k(X)\right)$, then $f(X) = U\mathrm{diag}\left(f(\lambda_1(X)), \ldots, \lambda_k(X)\right)U^T$.

The Fréchet derivatives are a notion of derivatives defined in Banach space. In this paper, we only concern about the Fréchet derivatives on matrix spaces. Readers may refer to [Col12] for a more thorough treatment. The Fréchet derivatives are the maps that are defined as follows.

Definition 8. *Given integers $m, n \geq 1$, a map $F : \mathsf{Mat}_m \to \mathsf{Mat}_n$ and $P, Q \in \mathsf{Mat}_m$, the Fréchet derivative of F at P with respect to Q is defined to be*

$$DF(P)[Q] = \frac{d}{dt} F(P + tQ)|_{t=0}.$$

The k-th order Fréchet derivative of F at P with respect to $(Q_1, \ldots, Q_k)$ is defined to be

$$D^k F(P)[Q_1, \ldots, Q_k] = \frac{d}{dt} D^{k-1} F(P + tQ_k)[Q_1, \ldots, Q_{k-1}]|_{t=0}.$$

Fréchet derivatives share many common properties with the derivatives in Euclidean spaces, such as linearity, composition rules, Taylor expansions, etc. We refer the interested reader to [Col12] for more. Some basic properties of Fréchet derivatives are summarized in the following fact.

Fact 9. *[Bha13, Chapter X.4] Given $F, G : \mathsf{Mat}_n \to \mathsf{Mat}_m$ and $P, Q_1, \ldots, Q_k \in \mathsf{Mat}_n$, it holds that*

1. $D(F + G)(P)[Q] = DF(P)[Q] + DG(P)[Q]$.
2. $D(F \cdot G)(P)[Q] = DF(P)[Q] \cdot G(P) + F(P) \cdot DG(P)[Q]$.
3. *If $m = n$,* $D(F \circ G)(P)[Q] = (D(G \circ F)(P) \circ DF(P))[Q]$.
4. $D^k F(P)[Q_1, \ldots, Q_k] = D^k F(P)\left[Q_{\sigma(1)}, \ldots, Q_{\sigma(k)}\right]$ *for every $k > 0$ and permutation $\sigma \in S_k$.*

The following fact states that Fréchet derivatives can be expressed as divided differences.

Fact 10. *[BLZ05] Let $f : \mathbb{R} \to \mathbb{R}$ be an analytical function and $X = \operatorname{diag}(x_1, \ldots, x_k)$ be a diagonal matrix whose spectrum is in $\mathbb{R}$. For any matrix A, B, the following holds*[11]

1. $$Df(X)[A] = \left(f^{[1]}(x_{i_1}, x_{i_2}) A_{i_1,i_2}\right)_{1 \leq i_1, i_2 \leq k}. \tag{5}$$

2. $$D^2 f(X)[A, B] = \left(\sum_{j=1}^{k} f^{[2]}(x_{i_1}, x_j, x_{i_2}) A_{i_1,j} B_{j,i_2}\right)_{1 \leq i_1, i_2 \leq k}. \tag{6}$$

Fact 11 (Dyson's expansion [Bha13, Chapter X.4])**.** *Let $f(x) = e^x$. For any $X \in \mathsf{Sym}_k$ and $A, B \in \mathsf{Mat}_k$, it holds*

$$Df(X)[A] = \int_0^1 du\ e^{(1-u)X} A e^{uX}.$$

Lemma 12. *Let $f(x) = e^{-x^2/2}$. It holds that*

$$\begin{aligned}
& D^2 f(X)[A, B] \\
= \ & \frac{1}{4}\int_0^1 du \int_0^1 dv\ \ (1-u)\, e^{-(1-u)(1-v)X^2/2} (XB + BX)\, e^{-(1-u)vX^2/2} (XA + AX)\, e^{-uX^2/2} \\
& + \frac{1}{4}\int_0^1 du \int_0^1 dv\ u e^{-(1-u)X^2/2} (XA + AX)\, e^{-u(1-v)X^2/2} (XB + BX)\, e^{-uvX^2/2} \\
& - \frac{1}{2}\int_0^1 du\ e^{-(1-u)X^2/2} (AB + BA)\, e^{-uX^2/2}.
\end{aligned}$$

[11] In [BLZ05, Lemma 3.8] this fact is proven when $A = B$ is a symmetric matrix and it is not hard to generalize their proof to obtain Eqs. (5), (6).

In particular, if $A = B = H$ is a symmetric matrix ,then

$$
\begin{aligned}
& D^2 f(X)[H,H] \\
= & \frac{1}{4}\int_0^1 du \int_0^1 dv \ \ (1-u)\, e^{-(1-u)(1-v)X^2/2}(XH+HX)\, e^{-(1-u)vX^2/2}(XH+HX)\, e^{-uX^2/2} \\
& +\frac{1}{4}\int_0^1 du \int_0^1 dv \ \ (u)\, e^{-(1-u)X^2/2}(XH+HX)\, e^{-u(1-v)X^2/2}(XH+HX)\, e^{-uvX^2/2} \\
& -\int_0^1 du \ e^{-(1-u)X^2/2} H^2 e^{-uX^2/2}.
\end{aligned}
$$

Note that $f(x) = e^{-x^2/2}$ is analytical in $\mathbb{R}$. Thus it is valid to define f on arbitrary matrices.

Proof. For any $t \in (0,1)$, we define $g(x) = e^{-tx^2}$. By the definition of Fréchet derivative

$$
\begin{aligned}
Dg(X)[A] &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(e^{-t(X+\varepsilon A)^2 - e^{-tA^2}}\right) \\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(e^{-t\left(X^2+\varepsilon(XA+AX)+\varepsilon^2A^2\right)} - e^{-tX^2}\right) \\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left(e^{-t\left(X^2+\varepsilon(XA+AX)\right)} + O\left(\varepsilon^2\right) - e^{-tX^2}\right) \\
&= -t\int_0^1 du \ e^{-(1-u)tX^2}(XA+AX)\, e^{-utX^2},
\end{aligned}
$$

where the second equality is from the fact that $\|e^{X+\varepsilon Y} - e^X\| = O(\varepsilon)$ and the last equality is from Fact 11. Setting $t = \frac{1}{2}$, we have

$$
Df(X)[A] = -\frac{1}{2}\int_0^1 du \ e^{-(1-u)X^2/2}(XA+AX)\, e^{-uX^2/2}.
$$

Taking one more derivative on X with respect to B, we conclude the result. □

2.5 Spectrahedrons and Positive spectrahedrons

Definition 13. *Given $\tau, M > 0$, we say a sequence of $k \times k$ positive semidefinite matrices $(A_1, \ldots, A_n)$ is (τ, M)-regular if*

$$
\mathbb{I} \preceq \sum_{i=1}^{n}\left(A^i\right)^2 \preceq M\cdot\mathbb{I} \text{ and } A^i \preceq \tau\cdot\mathbb{I} \text{ for every } i \in [m] \tag{7}
$$

A *spectrahedron* $S \subseteq \mathbb{R}^k$ is a feasible region of a semidefinite program. Namely, the set $S = \left\{x \in \mathbb{R}^n : \sum_i x_i A^i \preceq B\right\}$ for some symmetric matrices $A_1, \ldots, A_n, B$. We say S is a *positive spectrahedron* if either all A^is are positive semidefinite or all A^is are negative semidefinite (NSD). Moreover, it is (τ, M)-regular if either $(A_1, \ldots, A_n)$ or $(-A_1, \ldots, -A_n)$ is (τ, M)-regular.

We say S is an intersection of positive spetrahedrons if $S = S_1 \cap S_2$ where S_1 and S_2 are positive spectrahedrons whose matrices are all positive semidefinite and negative semidefinite, respectively. Note that it suffices to consider the intersections of two spetrahedrons as one can pack all PSD matrices into one large block-diagonal matrix (looking ahead this will only affect the parameters in our main results by a logarithmic factor). Packing the corresponding B_is, one get a positive spectrahedron. Same for all negative semidefinite matrices.

2.6 Pseudorandomness

Definition 14. *A function $g : \{-1,1\}^r \to \{-1,1\}^n$ with seed length r, is said to δ-fool a function $f : \{-1,1\}^n \to \mathbb{R}$ if*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{s}\sim\mathcal{U}_r} [f(g(\boldsymbol{s}))] - \mathop{\mathbb{E}}_{\boldsymbol{u}\sim\mathcal{U}_n} f(\boldsymbol{u}) \right| \leq \delta.$$

The function g is said to be a efficient pseudorandom generator (PRG) *that δ-fools a class $\mathcal{F}$ of n-variable functions if g is computable by a deterministic uniform poly(n)-time algorithm and g fools all function $f \in \mathcal{F}$.*

2.7 Tensors

For $\ell \geq 1$, let T^ℓ be an ℓ-tensor, i.e., $T^\ell : (\mathbb{R}^k)^{\times \ell} \to \mathbb{R}$. Note that a ℓ-tensor is defined uniquely by the coefficients $\{T_{i_1,\ldots,i_\ell} : i_1,\ldots,i_\ell \in [k]\}$. Below we abuse notation by letting $T(i_1,\ldots,i_\ell) = T_{i_1,\ldots,i_\ell}$. Often we will use the natural bijection between 2ℓ-tensors acting on $\mathbb{R}^k$ and ℓ-tensors acting on Mat_k, i.e., for a 2ℓ-tensor $T : (\mathbb{R}^k)^{\times 2\ell} \to \mathbb{R}$ defined as

$$T(x^1,\ldots,x^{2\ell}) = \sum_{i_1,\ldots,i_{2\ell}\in[k]} T(i_1,\ldots,i_\ell,i_{\ell+1},\ldots,i_{2\ell}) x^1_{i_1}\cdots x^{2\ell}_{i_{2\ell}},$$

we can also view T as $T' : (\mathsf{Mat}_k)^{\times \ell} \to \mathbb{R}$ defined by rearranging the terms above to obtain:

$$T'(X^1,\ldots,X^\ell) = \sum_{i_1,j_1\in[n]} \sum_{i_2,j_2\in[k]} \cdots \sum_{i_\ell,j_\ell\in[k]} T(i_1,\ldots,i_\ell,j_1,\ldots,j_\ell) X^1_{i_1,j_1}\cdots X^\ell_{i_\ell,j_\ell}$$

Finally, we define "permutation folding" operator which takes a (2ℓ)-tensor on $\mathbb{R}^k$ as defined above and produces a permutation to produce an ℓ-tensor on Mat_k.

Definition 15 (Definition of $\operatorname{diag}^\sigma T$)**.** *Let $T : (\mathbb{R}^k)^{\times t} \to \mathbb{R}$ be a k-tensor and $\sigma \in S_t$. Then we define $\operatorname{diag}^\sigma T : (\mathsf{Mat}_k)^{\times t} \to \mathbb{R}$ as the following map*

$$(\operatorname{diag}^\sigma T)((i_1,j_1)\ldots,(i_k,j_k)) = T(i_1,\ldots,i_k) \quad \text{iff } \vec{i} = \sigma\vec{j}, \tag{8}$$

and 0 otherwise.

3 Bentkus mollifier

In this paper, we are interested in smooth approximators of the function $\psi : \mathbb{R}^k \to \mathbb{R}$ defined as

$$\psi(x) = \left[\max_i x_i \leq 0\right]. \tag{9}$$

To this end, we introduce the Bentkus mollifier defined by Bentkus in [Ben90] and establish several new properties. Readers may refer to [Ben90, FK20] for a more thorough treatment.

Definition 16. *[Ben90] Let $g(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$. For every integer $k \geq 1$, define $G : \mathbb{R}^k \to \mathbb{R}$ as*

$$G(x_1,\ldots,x_k) = \prod_{i=1}^k g(x_i).$$

3.1 Properties of the mollifier and its derivatives

It is easy to calculate that

$$g'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{10}$$
$$g''(x) = -\frac{x}{\sqrt{2\pi}} \exp\left(-x^2/2\right) \tag{11}$$
$$g'''(x) = \frac{1}{\sqrt{2\pi}} \left(x^2 - 1\right) \exp\left(-x^2/2\right). \tag{12}$$

In order to simplify many calculations, we introduce the function

$$\bar{g}(x) = \frac{g'(x)}{g(x)}. \tag{13}$$

Fact 17. *[FK20] It holds that*

$$\overline{g}'(u) = -(u + \overline{g}(u)) \cdot \overline{g}(u); \tag{14}$$
$$\overline{g}''(u) = \left(u^2 - 1\right) \overline{g}(u) + 3u\overline{g}(u)^2 + 2\overline{g}(u)^3. \tag{15}$$

Also $\overline{g}$ is positive and monotone decreasing in $\mathbb{R}$. $\overline{g}'$ is negative and monotone increasing in $\mathbb{R}$.

Fact 18. *[Fel68, Section 7.1] For any $x \geq 0$, it holds that*

$$\frac{e^{-x^2/2}}{\sqrt{2\pi}} \left(\frac{1}{x} - \frac{1}{x^3}\right) \leq 1 - g(x) \leq \frac{e^{-x^2/2}}{x\sqrt{2\pi}}.$$

The following lemma immediately follows from Fact 17 and Fact 18.

Lemma 19. *For any $\Delta \geq 1$ and $x \in \mathbb{R}$ with $|x| \leq \Delta$, it holds that*

$$|\overline{g}(x)| \leq 2\Delta, \left|\overline{g}'(x)\right| \leq 3\Delta \left|\overline{g}(x)\right|, \left|\overline{g}''(x)\right| \leq 15\Delta^2 \left|\overline{g}(x)\right|.$$

3.2 Properties of the spectral norm of the mollifier

Fact 20. *[Ben90] It holds that for any integer $t, k \geq 1$*

$$\sup_{x \in \mathbb{R}^k} \|G^{(t)}(x)\|_1 \leq C_t \log^{t/2}(k+1) \tag{16}$$

for some constant C_t only depending on t.

Lemma 21. *For any $x \in \mathbb{R}^k$, if there exist more than $3\log k$ indices satisfying $x_i \leq 0$, then $\|G^{(1)}(x)\|_1 \leq O\left(\frac{1}{k^2}\right)$.*

Proof. Note that $g(z) \le \frac{1}{2}$ if $z \le 0$. Let $T = \{i : x_i \le 0\}$. Then

$$\begin{aligned}
\|G^{(1)}(x)\|_1 &= \sum_{i=1}^{k} \left| g'(x_i) \prod_{j \ne i} g(x_j) \right| \\
&= \sum_{i \in T} \left| g'(x_i) \prod_{j \ne i} g(x_j) \right| + \left| \sum_{i \notin T} g'(x_i) \prod_{j \ne i} g(x_j) \right| \\
&\le \frac{|T|}{2^{|T|-1}} + \frac{1}{2^{|T|}} \left| \sum_{i \notin T} g'(x_i) \prod_{\substack{j \ne i: \\ j \notin T}} g(x_j) \right| \le \frac{|T|}{2^{|T|-1}} + \frac{2\sqrt{2\log k}}{2^{|T|}},
\end{aligned}$$

where the equality used that the terms are all positive and the second inequality is from Fact 20. The upper bound is $O\left(\frac{1}{k^2}\right)$ if $|T| \ge 3\log k$. □

Claim 22. *For any $x > y$, it holds that*

$$\left| \frac{g(x)g'(y) - g'(x)g(y)}{x-y} \right| \le (1+|x|)\exp\left(-\frac{y^2}{2}\right) = (1+|x|)\, g'(y) \cdot \sqrt{2\pi}. \tag{17}$$

Proof.

$$\begin{aligned}
&\left| \frac{g(x)g'(y) - g'(x)g(y)}{x-y} \right| \\
= \;& \frac{1}{2\pi} \left| \int_{-\infty}^{0} \frac{\exp\left(-\frac{1}{2}\left(y^2 + (t+x)^2\right)\right) - \exp\left(-\frac{1}{2}\left(x^2 + (t+y)^2\right)\right)}{x-y} dt \right| \\
\le \;& \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \int_{-\infty}^{0} \left| \exp\left(-\frac{t^2}{2}\right) \frac{\exp(-ty) - \exp(-tx)}{x-y} \right| dt \\
= \;& \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \int_{-\infty}^{0} \left| \exp\left(-\frac{t^2}{2} - tx\right) \frac{1 - \exp(-t(y-x))}{y-x} \right| dt \\
\le \;& \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \int_{-\infty}^{0} \left| \exp\left(-\frac{t^2}{2} - tx\right) t \right| dt \\
= \;& \frac{1}{2\pi} \exp\left(-\frac{y^2}{2}\right) \int_{-\infty}^{0} \left| \exp\left(-\frac{1}{2}(t+x)^2\right) t \right| dt \\
= \;& \frac{1}{2\pi} \exp\left(-\frac{y^2}{2}\right) \left(\exp\left(-\frac{x^2}{2}\right) + \sqrt{2\pi}x - x \int_{x}^{\infty} e^{-t^2/2} dt \right) \\
\le \;& (1+|x|)\exp\left(-\frac{y^2}{2}\right),
\end{aligned}$$

where the second inequality used $|1 - e^{-z}| \le |z|$. □

For every $\theta > 0, \alpha \in \mathbb{R}$, we define the Bentkus mollifier as follows.

$$G_\theta(x) = \Pr_{\boldsymbol{g} \sim \mathcal{G}^k}\left[x + \alpha + \theta \boldsymbol{g} \le 0\right] \tag{18}$$

It is not hard to verify that

$$G_\theta(x) = \prod_{i=1}^{n} \int_{-\infty}^{-\frac{x_i}{\theta}} \frac{1}{\sqrt{2\pi}} e^{-x_i^2/2} = G\left(-\frac{x_1}{\theta}, \cdots, -\frac{x_k}{\theta}\right).$$

The following fact states that $G_\theta(\cdot + \alpha)/G_\theta(\cdot - \alpha)$ is a good approximator of ψ defined in Eq. (9) except a small inner/outer region near the "boundary" which is made precise below.

Fact 23 (Lemma 6.7 and Fact 6.8 in [OST19])**.** *For any $\delta, \theta \in (0,1)$, $x \in \mathbb{R}^k$ there exists $\Lambda = \Theta\left(\theta \cdot \sqrt{\log(k/\delta)}\right)$ and $\alpha = \Theta\left(\theta \cdot \sqrt{\log(k/\delta)}\right)$ such that the following holds.*

1. *$|G_\theta(x + \alpha) - \psi(x)| \le \delta$ if $\max_i x_i \le -\Lambda$.*
2. *$|G_\theta(x - \alpha) - \psi(x)| \le \delta$ if $\max_i x_i \ge \Lambda$.*
3. *$G_\theta(x + \alpha) - \delta \le \psi(x) \le G_\theta(x - \alpha) + \delta$ for all $x \in \mathbb{R}^k$.*

where $x + \alpha = (x_1 + \alpha, \ldots, x_k + \alpha)$

Let $A^i = \operatorname{diag}\left(A_1^i, A_2^i\right)$ and $D = \operatorname{diag}(D_1, D_2)$ be block diagonal matrices. To keep the notations succinct, we set $A(x) = \sum_i x_i A^i - D$.

Fact 24. *[OST19, Lemma 6.9] Let $k, \delta, \theta, \Lambda, \alpha$ be the parameters satisfying Fact 23. Let Ψ, Ψ_θ : $\mathsf{Sym}_k \to \mathbb{R}$ be the functions defined as $\Psi(M) = \psi(\lambda(M))$, $\Psi_\theta(M) = G_\theta(\lambda(M))$, where ψ is defined in Eq. (9) and G_θ is defined in Eq. (18), $\boldsymbol{x}$ and $\boldsymbol{x}'$ be two random variables in $\mathbb{R}^k$ satisfying that*

$$\left|\mathbb{E}\left[\Psi_\theta\left(A(\boldsymbol{x}) + \beta\mathbb{I}\right)\right] - \mathbb{E}\left[\Psi_\theta\left(A\left(\boldsymbol{x}'\right) + \beta\mathbb{I}\right)\right]\right| \le \eta,$$

for both $\beta = \alpha$ and $\beta = -\alpha$. Then, it holds that

$$\left|\mathbb{E}\left[\Psi\left(A(\boldsymbol{x})\right)\right] - \mathbb{E}\left[\Psi\left(A\left(\boldsymbol{x}'\right)\right)\right]\right| \le \eta + 3\delta + \Pr\left[\lambda_{\max}\left(A(\boldsymbol{x})\right) \in (-\Lambda, \Lambda]\right].$$

4 Computing spectral derivatives

4.1 Formulas for spectral derivatives

Before we describe the main theorem of this section, we need the following notation introduced by Sendov in [Sen07] to calculate the high-order Fréchet derivatives of spectral functions.

Definition 25. *[Sen07] Let $t \ge 1$ and $x \in \mathbb{R}^t$. Let $T : (\mathbb{R}^k)^{\times t} \to \mathbb{R}$ be a t-tensor. For every, $\ell \in [t]$, define a $(t+1)$-tensor $T_{\text{out}}^\ell : (\mathbb{R}^k)^{\times(t+1)} \to \mathbb{R}$ as follows*

$$(T_{\text{out}}^\ell)(i_1, \ldots, i_{t+1}) = \begin{cases} 0 & i_\ell = i_{t+1} \\ \frac{T(i_1,\ldots,i_{\ell-1},i_{t+1},i_{\ell+1},\ldots,i_t) - T(i_1,\ldots,i_{\ell-1},i_\ell,i_{\ell+1},\ldots,i_t)}{x_{i_{t+1}} - x_{i_\ell}} & i_\ell \ne i_{t+1}. \end{cases}$$

Finally, for every $\ell \in [t]$, define

$$T_\sigma(x) = \begin{cases} \nabla f(x) & \ell = 1, \sigma = (1) \\ (T(x))_{\text{out}}^\ell & \ell \le t-1 \\ \nabla T_\sigma(x) & \ell = t, \end{cases}$$

where $\sigma(\ell)$ is defined as follows: let σ be a permutation of $[k]$ given in the cycle decomposition, then $\sigma(\ell)$ is a permutation of $[k+1]$ elements whose cycle representation is the same as σ except that the element $k+1$ is inserted after the ℓth element and before the $(\ell+1)$th element in the cycle representation of σ.[12]

We are now ready to state the main theorem for computing spectral derivatives.

Theorem 26. *[Sen07] Let $X \in \mathsf{Sym}_k$ be such that the eigenvalues of X are all distinct. Let $F : \mathsf{Sym}_k \to \mathbb{R}$ be a spectral function (i.e., $F = f \circ \lambda$ for $f : \mathbb{R}^k \to \mathbb{R}$). Then F is t-times differentiable at X* if and only if *f is t-times differentiable at $\lambda(X)$.*

Moreover, for every $\sigma \in S_t, x \in \mathbb{R}^k$, let $T_\sigma(x) : (\mathbb{R}^k)^{\times t} \to \mathbb{R}$ be a t-tensor as defined in Definition 25 (which depends on the function f).[13] *Then, for every $U_1, \ldots, U_t \in \mathsf{Sym}_k$, we have*

$$D^t F(X)[U_1, \ldots, U_t] = \left(\sum_{\sigma \in S_t} \mathrm{diag}^\sigma T_\sigma(\lambda(X))\right)(V^T U_1 V, \ldots, V^T U_t V),$$

where V satisfies $X = V(\mathrm{diag}(\lambda(X))V^T$ and $\mathrm{diag}^\sigma T : (\mathsf{Mat}_k)^t \to \mathbb{R}$ is a t-tensor on the set Sym_k (as defined in Definition 15).

4.2 Understanding spectral derivatives for smooth functions

In this section, we first understand the relevant quantities to compute the spectral derivatives of smooth functions.

Theorem 27. *Let $k, n \geq 1$. Let $f : \mathbb{R}^k \to \mathbb{R}$ be a 3-times differentiable symmetric function and $\lambda : \mathsf{Sym}_k \to \mathbb{R}^k$ be the map $\lambda(M) = (\lambda_1(M), \ldots, \lambda_k(M))$ for every $M \in \mathsf{Sym}_k$. Let $F : \mathsf{Sym}_k \to \mathbb{R}$ be defined as $F(M) = (f \circ \lambda)(M)$ for all $M \in \mathsf{Sym}_k$. Then, for every $P \in \mathsf{Sym}_k$ with* distinct *eigenvalues and $H \in \mathsf{Sym}_k$, let $P = V(\mathrm{diag}(\lambda(P)))V^T$ be a spectral decomposition of P and $H = VQV^T$. Then $D^3 F(P)[Q, Q, Q]$ is the summation of the following terms.*

1. $\sum_{i_1} \nabla^3_{i_1,i_1,i_1} f(x) H^3_{i_1,i_1}$
2. $\sum_{i_1 \neq i_2} \nabla^3_{i_1,i_2,i_1} f(x) H^2_{i_1,i_1} H_{i_2,i_2}$
3. $\sum_{i_1 \neq i_2 \neq i_3} (\nabla^3_{i_1,i_2,i_2} f(x)) \cdot H_{i_1,i_1} H_{i_2,i_2} H_{i_3,i_3}$
4. $\sum_{i_1 \neq i_2} \left(\frac{\nabla^2_{i_2,i_2} - \nabla^2_{i_1,i_2}}{x_{i_2} - x_{i_1}} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2}\right) f(x) H_{i_2,i_2} H^2_{i_2,i_1}$
5. $\sum_{i_1 \neq i_2 \neq i_3} \frac{\nabla^2_{i_2,i_3} - \nabla^2_{i_1,i_3}}{x_{i_2} - x_{i_1}} f(x) H^2_{i_1,i_2} H_{i_3,i_3}$
6. $\sum_{i_1 \neq i_2 \neq i_3} \left(\frac{\nabla_{i_3} - \nabla_{i_1}}{(x_{i_3} - x_{i_2})(x_{i_3} - x_{i_1})} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_3} - x_{i_2})(x_{i_2} - x_{i_1})}\right) f(x) H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1}$
7. $\sum_{i_1 \neq i_2 \neq i_3} \left(\frac{\nabla_{i_2} - \nabla_{i_3}}{(x_{i_3} - x_{i_1})(x_{i_2} - x_{i_3})} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_3} - x_{i_1})(x_{i_2} - x_{i_1})}\right) f(x) H_{i_1,i_3} H_{i_2,i_1} H_{i_3,i_2}$,

[12] For more intuition, consider a simple example: let $\sigma = (12)(3)$ be a permutation on $[3]$, then $\sigma(\cdot)$ is a permutation on $[4]$ defined as follows: $\sigma(1)$ is $(142)(3)$, similarly $\sigma(2) = (124)(3)$, $\sigma(3) = (12)(34)$, $\sigma(4) = (12)(3)(4)$.

[13] Think of $x \in \mathbb{R}^k$ as the eigenvalues of $X \in \mathsf{Sym}_k$, i.e., $x = \lambda(X)$.

where $x = (\lambda_1(P), \ldots, \lambda_k(P))$.

Proof. To prove this theorem, we first apply Theorem 26 for $t = 3$ to obtain

$$D^3 F(P)[Q,Q,Q] = \left(\sum_{\sigma \in S_3} \operatorname{diag}^{\sigma} T_{\sigma}(\lambda(P))\right)(H,H,H). \tag{19}$$

We next carefully express each quantity in the summation using the definition of these tensors and upper bound each term. To this end, we break down all the six elements of S_3 and analyze them separately as follows.

Case 1: $\sigma = (1)(2)(3)$**.** Then $T_\sigma(x) = \nabla^3 f(x)$.

Case 2: $\sigma = (12)(3)$**.** First note that we have for $\sigma = (12)$ and

$$\left(T_{(12)}(x)\right)_{i_1,i_2} = \begin{cases} 0 & i_1 = i_2 \\ \frac{1}{x_{i_2} - x_{i_1}} \cdot \left(\nabla_{i_2} - \nabla_{i_1}\right) f(x) & i_1 \neq i_2 \end{cases}$$

Now, in order to compute $T_{(12)(3)}$, we need to compute $\nabla T_{(12)}(x)$ which can be written as follows

$$\begin{aligned}&\left(T_{(12)(3)}(x)\right)_{i_1,i_2,i_3} \\ &= \begin{cases} 0 & i_1 = i_2 \\ \frac{1}{x_{i_3} - x_{i_1}} \cdot \left(\nabla^2_{i_3,i_3} - \nabla^2_{i_1,i_3}\right) f(x) - \frac{1}{(x_{i_3} - x_{i_1})^2} \cdot \left(\nabla_{i_3} - \nabla_{i_1}\right) f(x) & i_2 = i_3 \neq i_1 \\ \frac{1}{x_{i_2} - x_{i_3}} \cdot \left(\nabla^2_{i_2,i_3} - \nabla^2_{i_3,i_3}\right) f(x) + \frac{1}{(x_{i_2} - x_{i_3})^2} \cdot \left(\nabla_{i_2} - \nabla_{i_3}\right) f(x) & i_1 = i_3 \neq i_2 \\ \frac{1}{x_{i_2} - x_{i_1}} \cdot \left(\nabla^2_{i_2,i_3} - \nabla^2_{i_1,i_3}\right) f(x) & i_1 \neq i_2 \neq i_3 \end{cases}\end{aligned}$$

Case 3: $\sigma = (13)(2)$**.** First note that for $\sigma = (1)(2)$, we have $T_{(1)(2)} = \nabla^2 f$ and $\sigma(1) = (13)(2)$. So, we need to compute $\left(\nabla^2 f\right) f(x)^1_{\text{out}}$ and we get

$$\left(T_{(13)(2)}(x)\right)_{i_1,i_2,i_3} = \begin{cases} 0 & i_1 = i_3 \\ \frac{1}{x_{i_3} - x_{i_1}} \cdot \left(\nabla^2_{i_3,i_2} - \nabla^2_{i_1,i_2}\right) f(x) & i_1 \neq i_3 \end{cases}$$

Case 4: $\sigma = (1)(23)$**.** First note that for $\sigma = (1)(2)$, we have $T_{(1)(2)} = \nabla^2 f$ and $\sigma(2) = (1)(23)$. So, we need to compute $\left(\nabla^2 f\right) f(x)^2_{\text{out}}$ and we get

$$\left(T_{(1)(23)}(x)\right)_{i_1,i_2,i_3} = \begin{cases} 0 & i_2 = i_3 \\ \frac{1}{x_{i_3} - x_{i_2}} \cdot \left(\nabla^2_{i_3,i_1} - \nabla^2_{i_2,i_1}\right) f(x) & i_2 \neq i_3 \end{cases}$$

Case 5: $\sigma = (123)$**.** Let $\sigma = (12)$, then $\sigma(2) = (123)$. So we need to compute $\left(T_{(12)}\right) f(x)^2_{\text{out}}$ and we obtain

$$\begin{aligned}&\left(T_{(123)}(x)\right)_{i_1,i_2,i_3} \\ &= \begin{cases} \frac{1}{(x_{i_2} - x_{i_1})^2} \cdot \left(\nabla_{i_2} - \nabla_{i_1}\right) f(x) & i_2 \neq i_3 = i_1 \\ \frac{1}{(x_{i_3} - x_{i_1})^2} \cdot \left(\nabla_{i_3} - \nabla_{i_1}\right) f(x) & i_1 = i_2 \neq i_3 \\ \frac{1}{(x_{i_3} - x_{i_2})(x_{i_3} - x_{i_1})} \cdot \left(\nabla_{i_3} - \nabla_{i_1}\right) f(x) - \frac{1}{(x_{i_3} - x_{i_2})(x_{i_2} - x_{i_1})} \cdot \left(\nabla_{i_2} - \nabla_{i_1}\right) f(x) & i_1 \neq i_3 \neq i_2 \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

Case 6: $\sigma = (132)$**.** Let $\sigma = (12)$, then $\sigma\tau(1) = (132)$. So we need to compute $\left(T_{(12)}\right) f(x)^1_{\text{out}}$ and we obtain.

$$\left(T_{(132)}(x)\right)_{i_1,i_2,i_3} = \begin{cases} -\frac{1}{(x_{i_2}-x_{i_1})^2} \cdot (\nabla_{i_2} - \nabla_{i_1}) f(x) & i_1 \neq i_3 = i_2 \\ \frac{1}{(x_{i_3}-x_{i_2})^2} \cdot (\nabla_{i_3} - \nabla_{i_2}) f(x) & i_2 = i_1 \neq i_3 \\ \frac{1}{(x_{i_3}-x_{i_1})(x_{i_2}-x_{i_3})} \cdot (\nabla_{i_2} - \nabla_{i_3}) f(x) - \frac{1}{(x_{i_3}-x_{i_1})(x_{i_2}-x_{i_1})} \cdot (\nabla_{i_2} - \nabla_{i_1}) f(x) & i_1 \neq i_3 \neq i_2 \\ 0 & \text{otherwise} \end{cases}$$

Using the above cases we can now rewrite Eq. (19) as

$$\sum_{\sigma \in S_3} T_\sigma(x)(H,H,H) = \sum_\sigma \sum_{i_1,i_2,i_3} (T_\sigma(x))_{i_1,i_2,i_3} H_{i_1,i_{\sigma(1)}} H_{i_2,i_{\sigma(2)}} H_{i_3,i_{\sigma(3)}}$$

Let's write this out as follows: by T_i, we mean $T_{case(i)}$ above

$$\sum_{i_1,i_2,i_3} (T_1)_{i_1,i_2,i_3} H_{i_1,i_1} H_{i_2,i_2} H_{i_3,i_3} + (T_2)_{i_1,i_2,i_3} H_{i_1,i_2} H_{i_2,i_1} H_{i_3,i_3} + (T_3)_{i_1,i_2,i_3} H_{i_1,i_3} H_{i_2,i_2} H_{i_3,i_1}$$
$$+ (T_4)_{i_1,i_2,i_3} H_{i_1,i_1} H_{i_2,i_3} H_{i_3,i_2} + (T_5)_{i_1,i_2,i_3} H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} + (T_6)_{i_1,i_2,i_3} H_{i_1,i_3} H_{i_2,i_1} H_{i_3,i_2}$$

and in particular, assuming H is symmetric the above simplifies to

$$\begin{aligned} &\sum_{i_1,i_2,i_3} (T_1)_{i_1,i_2,i_3} H_{i_1,i_1} H_{i_2,i_2} H_{i_3,i_3} + (T_2)_{i_1,i_2,i_3} H^2_{i_1,i_2} H_{i_3,i_3} + (T_3)_{i_1,i_2,i_3} H^2_{i_1,i_3} H_{i_2,i_2} \\ &\qquad + (T_4)_{i_1,i_2,i_3} H_{i_1,i_1} H^2_{i_2,i_3} + (T_5)_{i_1,i_2,i_3} H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} + (T_6)_{i_1,i_2,i_3} H_{i_1,i_3} H_{i_2,i_1} H_{i_3,i_2} \end{aligned} \tag{20}$$

Now, we will break up this sum into 5 cases as follows which we need to upper bound

Case (i): $i_1 = i_3 \neq i_2$**.** Then Eq. (20) reduces to the following

$$\sum_{i_1,i_2} H^2_{i_1,i_1} H_{i_2,i_2} (T_1 + T_3) + H_{i_1,i_1} H^2_{i_2,i_1} (T_2 + T_4 + T_5 + T_6) \tag{21}$$

Note that when we say T_q above, we mean $(T_q)_{i_1,i_2,i_3} = (T_q)_{i_1,i_2,i_1}$ (since $i_3 = i_1$). Let us now plug in the values of the corresponding T_qs into the formula and rewrite the above as follows

$$\begin{aligned} &\sum_{i_1 \neq i_2} H^2_{i_1,i_1} H_{i_2,i_2} \left(\nabla^3_{i_1,i_2,i_1} f(x) + 0\right) + \\ &\quad + H_{i_1,i_1} H^2_{i_2,i_1} \left(\frac{\nabla^2_{i_2,i_1} - \nabla^2_{i_1,i_1}}{x_{i_2} - x_{i_1}} + \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} + \frac{\nabla^2_{i_1,i_1} - \nabla^2_{i_2,i_1}}{x_{i_1} - x_{i_2}} + \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} \right) f(x) \\ &= \sum_{i_1 \neq i_2} H^2_{i_1,i_1} H_{i_2,i_2} \left(\nabla^3_{i_1,i_2,i_1} f(x)\right) + 2 H_{i_1,i_1} H^2_{i_2,i_1} \left(\frac{\nabla^2_{i_2,i_1} - \nabla^2_{i_1,i_1}}{x_{i_2} - x_{i_1}} + \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} \right) f(x) \end{aligned} \tag{22}$$

Case (ii): $i_1 = i_2 \neq i_3$**.** Then Eq. (20) reduces to

$$\sum_{i_1,i_3} H^2_{i_1,i_1} H_{i_3,i_3} (T_1 + T_2) + H_{i_1,i_1} H^2_{i_3,i_1} (T_3 + T_4 + T_5 + T_6) \tag{23}$$

The above simplies to the following

$$\sum_{i_1 \neq i_3} H_{i_1,i_1}^2 H_{i_3,i_3} \left(\nabla^3_{i_1,i_1,i_3} f(x) + 0\right) +$$
$$+ H_{i_1,i_1} H_{i_3,i_1}^2 \left(\frac{\nabla^2_{i_3,i_1} - \nabla^2_{i_1,i_1}}{x_{i_3} - x_{i_1}} + \frac{\nabla^2_{i_3,i_1} - \nabla^2_{i_1,i_1}}{x_{i_3} - x_{i_1}} + \frac{\nabla_{i_3} - \nabla_{i_1}}{(x_{i_3} - x_{i_1})^2} + \frac{\nabla_{i_3} - \nabla^2_{i_1}}{(x_{i_3} - x_{i_1})^2} \right) f(x) \tag{24}$$
$$= \sum_{i_1 \neq i_3} H_{i_1,i_1}^2 H_{i_3,i_3} \left(\nabla^3_{i_1,i_1,i_3} f(x)\right) + 2 H_{i_1,i_1} H_{i_3,i_1}^2 \left(\frac{\nabla^2_{i_3,i_1} - \nabla^2_{i_1,i_1}}{x_{i_3} - x_{i_1}} + \frac{\nabla_{i_3} - \nabla_{i_1}}{(x_{i_3} - x_{i_1})^2} \right) f(x)$$

Case (iii): $i_2 = i_3 \neq i_1$**.** Then Eq. (20) reduces to

$$\sum_{i_1,i_2} H_{i_2,i_2}^2 H_{i_1,i_1} (T_1 + T_4) + H_{i_2,i_2} H_{i_2,i_1}^2 (T_2 + T_3 + T_5 + T_6) \tag{25}$$

The above simplifies to the following

$$\sum_{i_1 \neq i_2} H_{i_2,i_2}^2 H_{i_1,i_1} \left(\nabla^3_{i_1,i_2,i_2} f(x) + 0\right) +$$
$$+ H_{i_2,i_2} H_{i_2,i_1}^2 \left(\frac{\nabla^2_{i_2,i_2} - \nabla^2_{i_1,i_2}}{x_{i_2} - x_{i_1}} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} + \frac{\nabla^2_{i_2,i_2} - \nabla^2_{i_1,i_2}}{x_{i_2} - x_{i_1}} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} \right) f(x) \tag{26}$$
$$= \sum_{i_1 \neq i_2} H_{i_2,i_2}^2 H_{i_1,i_1} \left(\nabla^3_{i_1,i_2,i_2} f(x)\right) + 2 H_{i_2,i_2} H_{i_2,i_1}^2 \left(\frac{\nabla^2_{i_2,i_2} - \nabla^2_{i_1,i_2}}{x_{i_2} - x_{i_1}} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} \right) f(x)$$

Case (i)+ Case (ii)+ Case (iii). We first upper bound these three cases to get the desired upper bound in the theorem statement. First summing the three cases, we have

$$\sum_{i_1 \neq i_2} H_{i_1,i_1}^2 H_{i_2,i_2} \left(\nabla^3_{i_1,i_2,i_1} + \nabla^3_{i_1,i_2,i_2} + \nabla^3_{i_2,i_1,i_1}\right) f(x)$$
$$+ 6 \sum_{i_1 \neq i_2} H_{i_2,i_2} H_{i_2,i_1}^2 \underbrace{\left(\frac{\nabla^2_{i_2,i_2} - \nabla^2_{i_1,i_2}}{x_{i_2} - x_{i_1}} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_2} - x_{i_1})^2} \right) f(x)}_{(\star)} \tag{27}$$

We now bound $(\star)$ using the following claim.

Case (iv): $i_2 = i_3 = i_1$**.** Then Eq. (20) reduces to

$$\sum_{i_1} H_{i_1,i_1}^3 (T_1 + T_2 + T_3 + T_4 + T_5 + T_6) = \sum_{i_1} H_{i_1,i_1}^3 \nabla^3_{i_1,i_1,i_1} f \tag{28}$$

Case (v): $i_2 \neq i_3 \neq i_1$. Then Eq. (20) stays the same and we get

$$\begin{aligned}
&\sum_{i_1,i_2,i_3} (\nabla^3_{i_1,i_2,i_2} f) \cdot H_{i_1,i_1} H_{i_2,i_2} H_{i_3,i_3} \\
&\quad + \frac{\nabla^2_{i_2,i_3} - \nabla^2_{i_1,i_3}}{x_{i_2} - x_{i_1}} f(x) H^2_{i_1,i_2} H_{i_3,i_3} + \frac{\nabla^2_{i_3,i_2} - \nabla^2_{i_1,i_2}}{x_{i_3} - x_{i_1}} f(x) H^2_{i_1,i_3} H_{i_2,i_2} + \frac{\nabla^2_{i_3,i_1} - \nabla^2_{i_2,i_1}}{x_{i_3} - x_{i_2}} f(x) H_{i_1,i_1} H^2_{i_2,i_3} \\
&\quad + \left(\frac{\nabla_{i_3} - \nabla_{i_1}}{(x_{i_3} - x_{i_2})(x_{i_3} - x_{i_1})} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_3} - x_{i_2})(x_{i_2} - x_{i_1})} \right) f(x) H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} \\
&\quad + \left(\frac{\nabla_{i_2} - \nabla_{i_3}}{(x_{i_3} - x_{i_1})(x_{i_2} - x_{i_3})} - \frac{\nabla_{i_2} - \nabla_{i_1}}{(x_{i_3} - x_{i_1})(x_{i_2} - x_{i_1})} \right) f(x) H_{i_1,i_3} H_{i_2,i_1} H_{i_3,i_2}
\end{aligned} \tag{29}$$

This concludes the proof of the theorem statement. □

4.3 Main theorem: spectral derivatives for Bentkus function

We now state the main theorem we prove using the theorem above. Let $G : \mathbb{R}^k \to \mathbb{R}$ be the Bentkus function given in Definition 16.

Theorem 28. *Let $k \geq 1$ be an integer and $\psi : \mathsf{Sym}_k \to \mathbb{R}$ be a function defined as $\psi(M) = (G \circ \lambda)(M)$ where G is given in Definition 16. Given $\Delta \geq 1$ $X \in \mathsf{Sym}_k$ with eigenvalues $\lambda(X) = (x_1, \ldots, x_k)$ satisfying that $\|X\| \leq \Delta$, it holds that*

$$\left| D^3 \psi(X) [H, H, H] \right| \leq O\left(\Delta^2 \cdot \log^3 k \cdot \|H\|^3 \right).$$

The following corollary simply follows from the definition of G_θ in Eq. 18 and the chain rule of Fréchet derivatives in Fact 9.

Corollary 29. *Let $k \geq 1$ be an integer and $\theta > 0, \alpha \in \mathbb{R}$ and $\Psi_\theta : \mathsf{Sym}_k \to \mathbb{R}$ be a function defined as $\Psi_\theta(M) = (G_\theta \circ \lambda)(M + \alpha\mathbb{I})$, where G_θ is given in Eq. (18). Given $\Delta \geq 1$, $X \in \mathsf{Sym}_k$ with eigenvalues $\lambda(X) = (x_1, \ldots, x_k)$ satisfying that $\|X\| \leq \Delta$, it holds that*

$$\left| D^3 \Psi_\theta (X + \alpha\mathbb{I}) [H, H, H] \right| \leq O\left(\frac{\Delta^2 + \alpha^2}{\theta^3} \cdot \log^3 k \cdot \|H\|^3 \right).$$

In order to prove the theorem above, We upper bound all the terms listed in Theorem 27 individually in the following sections (in increasing order of difficulty). Given the calculations are fairly technical we break down the analysis in the following sections for modularity and reader convenience. In Section 4.4.1 we bound the first three terms in Theorem 27 (this is the easy case since the analysis is very similar to what happens in [HKM13] and requires new properties of the Bentkus function), in Section 4.4.2 and 4.4.3 we bound the fourth and fifth term (this already deviates from the analysis of [HKM13]) and finally in Section 4.5 we bound the sixth and seventh term (this calculation is fairly involved and deviates significantly from prior works, since we need to deal with properties of Fréchet derivatives, new properties of Bentkus function and the *non-diagonal* entries of the matrices H which is unique to the matrix-spectrahedron case and is not faced in [HKM13, ST17, OST19]).

As spectral functions and spectral norms are unitary invariant, we assume that $X = \text{diag}(x_1, \ldots, x_k)$ is diagonal without loss of generality. To adopt Theorem 27, we assume that all the $x_1, \ldots, x_n$ are distinct. We further conclude the result by continuity.

4.4 Bounding terms (1)-(5) in Theorem 27 for Bentkus function

Let $G:\mathbb{R}^k\to\mathbb{R}$ be the Bentkus function given in Definition 16. Recall that $G(x)=\prod_i g(x_i)$, where $g(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-x}e^{-t^2/2}dt$. Recall the notation $g'(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $\overline{g}(x)=g'(x)/g(x)$.

4.4.1 Bounding terms $(1,2,3)$ in Theorem 27

Lemma 30 (Bounding terms $(1,2,3)$)**.** *The following three terms*

$$\left|\sum_{i_1}\nabla^3_{i_1,i_1,i_1}G(x)\,H^3_{i_1,i_1}\right|,\quad \left|\sum_{i_1\neq i_2}\nabla^3_{i_1,i_2,i_1}G(x)\,H^2_{i_1,i_1}H_{i_2,i_2}\right|,\quad \left|\sum_{i_1\neq i_2\neq i_3}\nabla^3_{i_1,i_2,i_2}G(x)\cdot H_{i_1,i_1}H_{i_2,i_2}H_{i_3,i_3}\right|$$

can be upper bound by $O(\log^{1.5}k\cdot\|H\|^3)$.

Proof. The first upper bound is straightforward. Observe that

$$\left|\sum_{i_1}\nabla^3_{i_1,i_1,i_1}G(x)\,H^3_{i_1,i_1}\right|\le\max_i|H_{i,i}|^3\cdot\sum_{i_1}\left|\nabla^3_{i_1,i_1,i_1}G(x)\right|\le\max_i|H_{i,i}|^3\cdot\|G^{(3)}(x)\|_1\le\|H\|^3\cdot\log^{1.5}k,$$

where the second inequality follows by definition of $\|G^{(3)}\|_1$ and the last inequality used $\max_{i,j}|H_{i,j}|\le\|H\|$ (the latter being the spectral norm of H) and Fact 20 to conclude $\|G^{(3)}\|_1\le O\left(\log^{1.5}k\right)$. Similarly, the remaining two terms can also be bounded exactly as above (by observing that $\sum_{i_1\neq i_2}\nabla^3_{i_1,i_2,i_1}G$ and $\sum_{i_1\neq i_2\neq i_3}(\nabla^3_{i_1,i_2,i_2}G)$ appear in the expression of $\|G^{(3)}\|_1$. □

4.4.2 Bounding term (4) in Theorem 27

In order to bound the remaining terms in Theorem 27, we need the following claim.

Claim 31. *It holds that*

1. $\sum_{i_1\neq i_2}\overline{g}(x_{i_1})\left|G(x)\,H_{i_2,i_2}H^2_{i_1,i_2}\right|\le O\left(\sqrt{\log k}\cdot\|H\|^3\right)$.
2. $\sum_{i_1\neq i_2}\overline{g}(x_{i_2})\left|G(x)\,H_{i_2,i_2}H^2_{i_1,i_2}\right|\le O\left(\sqrt{\log k}\cdot\|H\|^3\right)$.
3. $\sum_{i_1\neq i_2\neq i_3}\left|\overline{g}(x_{i_2})\,\overline{g}(x_{i_3})\,G(x)\,H^2_{i_1,i_2}H_{i_3,i_3}\right|\le O\left(\log k\cdot\|H\|^3\right)$.

Proof. For Item 1, we have

$$\sum_{i_1\neq i_2}\overline{g}(x_{i_1})\left|G(x)\,H_{i_2,i_2}H^2_{i_1,i_2}\right|\le\sum_{i_1}\overline{g}(x_{i_1})\,G(x)\cdot\max_{i_1}\sum_{i_2}\left|H_{i_2,i_2}H^2_{i_1,i_2}\right|\le\|G^{(1)}(x)\|_1\|H\|^3$$

where the last inequality is because

$$\max_{i_1}\sum_{i_2}\left|H_{i_2,i_2}H^2_{i_1,i_2}\right|\le\|H\|\max_{i_1}\left(H^2\right)_{i_1,i_1}\le\|H\|^3,\tag{30}$$

using the fact that $\max_{ij}|H_{ij}|\le\|H\|$. Using Fact 20 shows the first inequality. Item 2 follows by the same reason.

For Item 3, we have

$$\begin{aligned}\sum_{i_1\neq i_2\neq i_3} \left|\overline{g}\left(x_{i_2}\right)\overline{g}\left(x_{i_3}\right)G\left(x\right)H^2_{i_1,i_2}H_{i_3,i_3}\right| &= \sum_{i_2\neq i_3}\left|\overline{g}\left(x_{i_2}\right)\overline{g}\left(x_{i_3}\right)G\left(x\right)\right|\max_{i_2,i_3}\sum_{i_1}\left|H^2_{i_1,i_2}H_{i_3,i_3}\right| \\ &\leq O\left(\log k\cdot\|H\|^3\right)\end{aligned}$$

where the inequality is from Fact 20 and the fact that

$$\sum_{i_1}\left|H^2_{i_1,i_2}H_{i_3,i_3}\right| \leq \|H\|\cdot\left(H^2\right)_{i_2,i_2} \leq \|H\|^3. \tag{31}$$

□

Lemma 32 (Bounding terms (4) in Theorem 27)**.** *We have*

$$\sum_{i_1\neq i_2} H_{i_2,i_2}H^2_{i_2,i_1}\left(\frac{\nabla^2_{i_2,i_2}G-\nabla^2_{i_1,i_2}G}{x_{i_2}-x_{i_1}} - \frac{\nabla_{i_2}G-\nabla_{i_1}G}{(x_{i_2}-x_{i_1})^2}\right) \leq O\left(\Delta^2\cdot\sqrt{\log k}\|H\|^3\right).$$

Proof. First observe that

$$\nabla_{i_2}G(x) = g'(x_{i_2})\prod_{j\neq i_2}G(x_j) = \overline{g}(x_{i_2})\cdot G(x),$$

and similarly we have

$$\nabla_{i_2,i_2}G(x) = \overline{g}(x_{i_2})\nabla_{i_2}G(x)+G(x)\nabla_{i_2}\overline{g}(x_{i_2}) = \left(\overline{g}(x_{i_2})^2-(x_{i_2}+\overline{g}(x_{i_2}))\overline{g}(x_{i_2})\right)G(x) = -x_{i_2}\overline{g}(x_{i_2})G(x),$$

where we used Fact 17. Now, let us start upper bounding the lemma statement as follows

$$\begin{aligned}&\left|\sum_{i_1\neq i_2} H_{i_2,i_2}H^2_{i_2,i_1}\left(\frac{\nabla^2_{i_2,i_2}G-\nabla^2_{i_1,i_2}G}{x_{i_2}-x_{i_1}} - \frac{\nabla_{i_2}G-\nabla_{i_1}G}{(x_{i_2}-x_{i_1})^2}\right)\right| \\ &\leq \sum_{i_1\neq i_2}\left|-\frac{\overline{g}(x_{i_1})\overline{g}(x_{i_2})+x_{i_2}\overline{g}(x_{i_2})}{x_{i_2}-x_{i_1}} - \frac{\overline{g}(x_{i_2})-\overline{g}(x_{i_1})}{(x_{i_2}-x_{i_1})^2}\right|\cdot\left|G(x)\cdot H_{i_2,i_2}H^2_{i_2,i_1}\right| \\ &= \sum_{i_1\neq i_2}\left|-\frac{\overline{g}(x_{i_1})\overline{g}(x_{i_2})+x_{i_2}\overline{g}(x_{i_2})}{x_{i_2}-x_{i_1}} - \frac{\overline{g}'(\xi_{i_1,i_2})}{x_{i_2}-x_{i_1}}\right|\cdot\left|G(x)\cdot H_{i_2,i_2}H^2_{i_2,i_1}\right| \\ &= \sum_{i_1\neq i_2}\left|\frac{\overline{g}(x_{i_1})\overline{g}(x_{i_2})+x_{i_2}\overline{g}(x_{i_2})}{x_{i_2}-x_{i_1}} - \frac{(\xi_{i_1,i_2}+\overline{g}(\xi_{i_1,i_2}))\overline{g}(\xi_{i_1,i_2})}{x_{i_2}-x_{i_1}}\right|\cdot\left|G(x)\cdot H_{i_2,i_2}H^2_{i_2,i_1}\right| \\ &\leq \sum_{i_1\neq i_2}\underbrace{\left|\frac{x_{i_2}\overline{g}(x_{i_2})-\xi_{i_1,i_2}\overline{g}(\xi_{i_1,i_2})}{x_{i_2}-x_{i_1}}\right|\cdot\left|G(x)\cdot H_{i_2,i_2}H^2_{i_2,i_1}\right|}_{:=(1)} + \underbrace{\left|\frac{\overline{g}(x_{i_1})\overline{g}(x_{i_2})-\overline{g}(\xi_{i_1,i_2})^2}{x_{i_2}-x_{i_1}}\right|\cdot\left|G(x)\cdot H_{i_2,i_2}H^2_{i_2,i_1}\right|}_{:=(2)},\end{aligned}$$

where the first equality used the mean-value theorem to obtain a $\xi_{i_1,i_2}\in[x_{i_1},x_{i_2}]$, second equality used Eq. (14). We now bound both these terms separately as follows.

Term 1 upper bound. Note that ξ_{i_1,i_2} is between x_{i_1} and x_{i_2}. The first term is upper bounded by

$$\begin{aligned}
& \sum_{i_1 \neq i_2} \left| \frac{x_{i_2}\overline{g}(x_{i_2}) - \xi_{i_1,i_2}\overline{g}(\xi_{i_1,i_2})}{x_{i_2} - \xi_{i_1,i_2}} \right| \cdot \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right| \\
= & \sum_{i_1 \neq i_2} \left| \left(1 - \eta_{i_1,i_2}^2\right) \overline{g}\left(\eta_{i_1,i_2}\right) - \eta_{i_1,i_2} \overline{g}\left(\eta_{i_1,i_2}\right)^2 \right| \cdot \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right| \\
\leq & \; 2\Delta^4 \sum_{i_1 \neq i_2} \overline{g}\left(\eta_{i_1,i_2}\right) \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right|
\end{aligned}$$

for some η_{i_1,i_2} between x_{i_2} and ξ_{i_1,i_2}, where we apply a mean value theorem for the function $x\overline{g}(x)$ for the equality and Lemma 19 for the inequality. Note that $\overline{g}(\cdot)$ is nonnegative and monotone decreasing by Fact 17. Thus the first term is upper bounded by

$$2\Delta^2 \sum_{i_1 \neq i_2} \max\left\{\overline{g}\left(x_{i_1}\right), \overline{g}\left(x_{i_2}\right)\right\} \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right|$$

which, in turn, is upper bounded by $O\left(\Delta^4 \cdot \sqrt{\log k} \cdot \|H\|^3\right)$ from Fact 20 and Eqs (30), (31).

Term 2 upper bound. By triangle inequality we upper bound the second term by

$$\begin{aligned}
& \sum_{i_1 \neq i_2} \left| \frac{\overline{g}(x_{i_1})\overline{g}(x_{i_2}) - \overline{g}(\xi_{i_1,i_2})^2}{x_{i_2} - x_{i_1}} \right| \cdot \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right| \\
& \leq \sum_{i_1 \neq i_2} \left| \frac{\overline{g}(x_{i_1})\overline{g}(x_{i_2}) - \overline{g}(x_{i_1})^2}{x_{i_2} - x_{i_1}} \right| \cdot \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right| + \sum_{i_1 \neq i_2} \left| \frac{\overline{g}(x_{i_1})^2 - \overline{g}(\xi_{i_1,i_2})^2}{x_{i_2} - x_{i_1}} \right| \cdot \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right|.
\end{aligned} \tag{32}$$

We first upper bound the first quantity in Eq. (32) first as follows.

$$\begin{aligned}
& \sum_{i_1 \neq i_2} \left| \frac{\overline{g}(x_{i_1})\overline{g}(x_{i_2}) - \overline{g}(x_{i_1})^2}{x_{i_2} - x_{i_1}} \right| \cdot \left| G(x) \cdot H_{i_2,i_2} H_{i_2,i_1}^2 \right| \\
& = \sum_{i_1 \neq i_2} \left| \frac{\overline{g}(x_{i_2}) - \overline{g}(x_{i_1})}{x_{i_2} - x_{i_1}} \right| \cdot |G(x)| \cdot |\overline{g}(x_{i_1})| \cdot \left| H_{i_2,i_2} H_{i_2,i_1}^2 \right| \\
& = \sum_{i_1 \neq i_2} |\overline{g}'(\zeta_{i_1,i_2})| \cdot |G(x)| \cdot |\overline{g}(x_{i_1})| \cdot \left| H_{i_2,i_2} H_{i_2,i_1}^2 \right| \\
& \leq 6\Delta^2 \cdot \sum_{i_1 \neq i_2} |G(x)| \cdot |\overline{g}(x_{i_1})| \cdot \left| H_{i_2,i_2} H_{i_2,i_1}^2 \right| \\
& \leq 6\Delta^2 \|G^{(1)}\|_1 \|H\|^3. \\
& \leq O\left(\Delta^2 \cdot \sqrt{\log k} \cdot \|H\|^3\right),
\end{aligned} \tag{33}$$

where ζ_{i_1,i_2} between x_{i_1} and x_{i_2}, first inequality uses Fact 19, the second inequality uses Eqs. (30), (31) and the last inequality is from Fact 20.

We now bound the second term in Eq. (32) as follows

$$\sum_{i_1\neq i_2}\left|\frac{\overline{g}(x_{i_1})^2-\overline{g}(\xi_{i_1,i_2})^2}{x_{i_2}-x_{i_1}}\right|\cdot|G(x)\cdot H_{i_2,i_2}H_{i_2,i_1}^2|$$
$$\leq\sum_{i_1\neq i_2}\left|\frac{\overline{g}(x_{i_1})^2-\overline{g}(\xi_{i_1,i_2})^2}{\xi_{i_1,i_2}-x_{i_1}}\right|\cdot|G(x)\cdot H_{i_2,i_2}H_{i_2,i_1}^2| \qquad \text{(for } \xi \text{ is between } x_{i_1} \text{ and } x_{i_2}\text{)}$$
$$=2\sum_{i_1\neq i_2}\left|\overline{g}\left(\eta_{i_1,i_2}\right)\overline{g}'\left(\eta_{i_1,i_2}\right)\right|\cdot\left|G\left(x\right)\right|\cdot\left|H_{i_2,i_2}H_{i_2,i_1}^2\right| \qquad \text{(for some } \eta_{i_1,i_2} \text{ between } x_{i_1} \text{ and } \xi_{i_1,i_2}\text{)}$$
$$\leq 12\Delta^2\sum_{i_1\neq i_2}\left|\overline{g}\left(\eta_{i_1,i_2}\right)\right|\cdot\left|G\left(x\right)\right|\cdot\left|H_{i_2,i_2}H_{i_2,i_1}^2\right| \qquad \text{(Fact 19)}$$
$$\leq 12\Delta^2\sum_{i_1\neq i_2}\max\left\{\overline{g}\left(x_{i_1}\right),\overline{g}\left(x_{i_2}\right)\right\}\cdot\left|G\left(x\right)\right|\cdot\left|H_{i_2,i_2}H_{i_1,i_2}^2\right|. \qquad \text{(Fact 17 and Lemma 19)}$$

Further applying Fact 20 and putting together Eqs. (31)(30), we conclude that it can be upper bounded by $O\left(\Delta^2\sqrt{\log k}\|H\|^3\right)$. □

4.4.3 Bounding term (5) in Theorem 27

Lemma 33 (Bounding terms (5) in Theorem 27)**.** *We have*

$$\left|\sum_{i_1\neq i_2\neq i_3}\frac{\nabla^2_{i_2,i_3}G\left(x\right)-\nabla^2_{i_1,i_3}G\left(x\right)}{x_{i_2}-x_{i_1}}H_{i_1,i_2}^2H_{i_3,i_3}\right|\leq O\left(\Delta\cdot\log k\cdot\|H\|^3\right).$$

Proof.

$$\left|\sum_{i_1\neq i_2\neq i_3}\frac{\nabla^2_{i_2,i_3}G\left(x\right)-\nabla^2_{i_1,i_3}G\left(x\right)}{x_{i_2}-x_{i_1}}H_{i_1,i_2}^2H_{i_3,i_3}\right|$$
$$=\left|\sum_{i_1\neq i_2\neq i_3}\frac{\overline{g}\left(x_{i_3}\right)\left(\overline{g}\left(x_{i_2}\right)-\overline{g}\left(x_{i_1}\right)\right)}{x_{i_2}-x_{i_1}}G\left(x\right)H_{i_1,i_2}^2H_{i_3,i_3}\right|$$
$$=\left|\sum_{i_1\neq i_2\neq i_3}\overline{g}'\left(\xi_{i_1,i_2}\right)\overline{g}\left(x_{i_3}\right)G\left(x\right)H_{i_1,i_2}^2H_{i_3,i_3}\right| \qquad \text{(for some } \xi_{i_1,i_2} \text{ between } x_{i_1} \text{ and } x_{i_2}\text{)}$$
$$\leq 3\Delta\sum_{i_1\neq i_2\neq i_3}\left|\max\left\{\overline{g}\left(x_{i_1}\right),\overline{g}\left(x_{i_2}\right)\right\}\overline{g}\left(x_{i_3}\right)G\left(x\right)H_{i_1,i_2}^2H_{i_3,i_3}\right|$$
$$\leq O\left(\Delta\cdot\log k\cdot\|H\|^3\right),$$

where the last inequality is from Fact 20 and Eqs. (30)(31). □

4.5 Bounding terms (6, 7) in Theorem 27 for Bentkus function

Let $G:\mathbb{R}^k\to\mathbb{R}$ be the Bentkus function given in Definition 16. Recall that $G(x)=\prod_i g(x_i)$, where $g\left(x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-x}e^{-t^2/2}dt$. Recall the notation $g'\left(x\right)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ and $\overline{g}\left(x\right)=g'(x)/g(x)$. Restating the terms for convenience.

Lemma 34 (Bounding terms (6, 7) in Theorem 27)**.**

$$\left| \sum_{i_1 \neq i_2 \neq i_3} \frac{\frac{\bar{g}(x_{i_1}) - \bar{g}(x_{i_3})}{x_{i_3} - x_{i_1}} - \frac{\bar{g}(x_{i_1}) - \bar{g}(x_{i_2})}{x_{i_2} - x_{i_1}}}{x_{i_3} - x_{i_2}} G\left(x\right) H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} \right| \leq O\left(\Delta \log^2 k \|H\|^3\right) \tag{34}$$

This is the most involved part. Note that the left hand side is unchanged if we zero out all diagonal entries of H. And further note that $\|H - \mathrm{diag}\left(H\right)\| \leq 2\|H\|$ where $\mathrm{diag}\left(H\right)$ is a diagonal matrix obtained by diagonalizing H. Thus, we may assume that the diagonal elements in H are zero without loss of generality. We break down the analysis into two cases (the first one being the simpler case).

4.5.1 Case 1: Many negative x_is.

The simpler case is when the number of negative x_is is "large".

Lemma 35. *If $|\{i : x_i < 0\}| > 3 \log k$, then the quantity in Eq. (34) is upper bounded by $O\left(\Delta^2 \frac{\sqrt{\log k}}{k} \cdot \|H\|^3\right)$.*

Proof. Applying Fact 3 a mean value theorem of divided difference and Lemma 19, the term in Eq. (34) is upper bounded by

$$\begin{aligned}
& O\left(\Delta^2 \sum_{i_1 \neq i_2 \neq i_3} \bar{g}(\zeta_{i_1,i_2,i_3})\right) G\left(x\right) |H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1}| \\
\leq\ & O\left(\Delta^2 \sum_{i_1 \neq i_2 \neq i_3} \max\left\{\bar{g}\left(x_{i_1}\right), \bar{g}\left(x_{i_2}\right), \bar{g}\left(x_{i_3}\right)\right\}\right) G\left(x\right) |H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1}| \\
\leq\ & O\left(\Delta^2 \|G^{(1)}\|_1 \max_{i_1} \sum_{i_2,i_3} |H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1}|\right) \\
\leq\ & O\left(\Delta^2 \cdot k \cdot \|G^{(1)}\left(x\right)\|_1 \cdot \max_{i_1,i_2} \sum_{i_3} |H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1}|\right) \\
\leq\ & O\left(\Delta^2 \cdot k \cdot \|G^{(1)}\left(x\right)\|_1 \cdot \|H\|^3\right) \\
\leq\ & O\left(\frac{\Delta^2 \sqrt{\log k}}{k} \cdot \|H\|^3\right)
\end{aligned}$$

where the first inequality is from the positivity and monotonicity of $\bar{g}\left(\cdot\right)$ due to Fact 17 to conclude that $|\bar{g}(\zeta_{i_1,i_2,i_3})| \leq \max\{|\bar{g}(x_{i_1})|, |\bar{g}(x_{i_2})|, |\bar{g}(x_{i_3})|\}$; the second last inequality is from the following fact

$$\sum_{i_3} |H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1}| \leq \|H\| \sqrt{\left(\sum_{i_3} H_{i_1,i_3}^2\right)\left(\sum_{i_3} H_{i_2,i_3}^2\right)} \leq \|H\|^3; \tag{35}$$

the last inequality is from Lemma 21 (which uses that the number of negative x_is is $\leq 3 \log k$). $\square$

4.5.2 Case 2: A few negative x_is

We now assume that $|\{i : x_i < 0\}| \le 3 \log k$ and this case the most complicated and upper bounding it is the most technical. We push this proof to Appendix A.

Proof of Theorem 28. Combining Theorem 27 and Lemmas 30, 32, 33, 34, we obtain our result. ☐

5 Geometric properties of positive spectrahedrons

5.1 Noise sensitivity and Gaussian surface area

In this section we prove certain combinatorial and geometric properties of positive spectrahedrons. Understanding the surface area of a convex object is a fundamental question in convex geometry. In the context of theoretical computer science, one of the earliest works Klivans, O'Donnell and Servedio [KOS04] related learnability of geometric convex objects (in the PAC and agnostic setting) to a natural complexity measure of of *Gaussian surface area* (GSA). Recall that for a convex object $S \subseteq \mathbb{R}^n$, we have

$$\mathsf{GSA}\,(S) = \liminf_{\delta \to 0} \frac{\mathcal{G}^n\,(S_\delta \setminus S)}{\delta}$$

where $S_\delta = \{x : \mathrm{dist}\,(x, S) \le \delta\}$ denotes the δ-neighborhood of S under Euclidean distance.

In some sense the work of [KOS04] showed that, GSA of convex objects *characterizes* learnability of these objects under the Gaussian distribution. This remarkable connection has provided further motivation to understand what is the GSA of basic well-studied convex sets. In this direction, a well-known result of Ball gives an upper bound on the surface area of arbitrary convex objects.

Theorem 36. *[Bal93] The surface area of every convex set on n coordinates is at most $O(n^{1/4})$.*[14]

For our setting it is unclear what the Gaussian surface area of spectrahedrons is. Clearly, since spectrahedrons are convex objects, one can use Theorem 36 to show an upper bound of $O(n^{1/4})$. Below we show that one can in fact prove an upper bound of $O(1)$ on the Gaussian surface area of *positive* spectrahedrons. We make this formal in the theorem below.

Theorem 37 (Matrix version of Peres theorem)**.** *Let S be a positive spectrahedron defined as*

$$S = \Big\{x \in \mathbb{R}^n : \sum_i x_i A^i \preceq B,\ \ A^1, \ldots, A^n, B \in \mathsf{Sym}_k \text{ and } A^i \text{ is } \mathsf{PSD} \text{ for } i \in [n]\Big\}.$$

Then the Gaussian surface area of S is $O(1)$ (independent of k, n). Let $f(x) = [x \in S]$ for $x \in \{-1, 1\}^n$. Then the ε-noise sensitivity of f is $\mathsf{NS}_\varepsilon(f) = O(\sqrt{\varepsilon})$.

Corollary 38. *Let S^1, S^2 be 2 distinct positive spectrahedrons specified by $\{A_j^1, \ldots, A_j^n, B_j\}_{j\in[2]}$ respectively, where $A_1^i \succeq 0$ and $A_2^i \preceq 0$ for all i. Let*

$$F(x) = \bigwedge_{j=1}^{2} \left[\sum_i x_i A_j^i \preceq B_j\right]$$

be an intersection of positive spectrahedrons. Then

$$\mathsf{AS}(F) \le O(\sqrt{n}), \qquad \mathsf{GSA}(S^1 \cap S^2) = O(1)$$

[14] Subsequently it was shown that this bound was optimal for a convex body formed by $\exp(n^{1/4})$ randomly intersecting halfspaces.

The proof of this theorem follows closely the proof of Kane [Kan14a] who showed that polytopes have Gaussian surface area $O(\sqrt{\log k})$ (thereby reproving Nazarov [Naz03]). Before stating the Kane's result, we need to introduce the following notion.

Definition 39 (Unate function). *A function $f : \{-1,1\}^n \to \{0,1\}$ is* unate *if it satisfies the following: for every $i \in [n]$, f is either increasing or decreasing with respect to the ith coordinate, i.e., for every $i \in [n]$, either $f(x_1,\ldots,x_{i-1},-1,x_{i+1},\ldots,x_n) \le f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n)$ for all x or $f(x_1,\ldots,x_{i-1},-1,x_{i+1},\ldots,x_n) \ge f(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n)$ for all x.*

In particular, Kane proved the following stronger statement.

Theorem 40. *[Kan14a] Let $f_1,\ldots,f_k : \{-1,1\}^n \to \{0,1\}$ be unate functions and let $F : \{-1,1\}^n \to \{0,1\}$ be defined as $F(x) = \bigwedge_i f_i(x)$. Then the average sensitivity of F satisfies $\mathsf{AS}(F) \le O(\sqrt{n\log(k+1)})$.*[15]

It is not hard to see that a positive spectrahedron is a unate function so Theorem 40 holds for us as well for $k = 1$. To be precise we have

Corollary 41. *Let S be as defined in Theorem 37. Let $F : \{-1,1\}^n \to \{0,1\}$ be defined as $F(x) = 1$ if and only if $x \in S$. Then $\mathsf{AS}(F) \le O(\sqrt{n})$.*

In order to translate Theorem 40 to obtain the corollary above: for a positive spectrahedron S, let $F(x) = [x \in S]$ for $x \in \{-1,1\}^n$, then one can rewrite F as $F(x) = \bigwedge_{j=1}^k \left[\lambda_j \left(\sum_i x_i A^i - B\right) \ge 0\right]$ which is an AND of k unate functions by the Weyl's inequality [Bha13, Theorem III 2.1](the inner functions f_i are unate since all the A^is are promised to be PSD).

Recall that we are interested in the Gaussian surface area of such bodies (not just the average sensitivity) which is closely related to the *noise sensitivity* of positive spectrahedrons. In the same paper, Kane [Kan14a] adapts the well-known techniques of [DGJ+10] to show that the ε-noise sensitivity of *intersections* of halfspaces is at most $O(\sqrt{\varepsilon \log k})$ and remarks that such a bound *does not* hold for the intersections of unate functions. Below, we show that one can modify the proof of [DGJ+10] to also show that the noise sensitivity of positive spectrahedrons is can be bounded by a the "average 2-sensitivity" of positive spectrahedrons which we show is $O(\sqrt{\varepsilon})$ by modifying Kane's proof in Theorem 40. This proves our Theorem 37.

Proof of Theorem 37. In order to prove the theorem, we first show that for a function $f : \{-1,1\}^n \to \{0,1\}$ defined as

$$f(x) = \left[\sum_{i=1}^n x_i A^i \preceq B\right]$$

for $A^1,\ldots,A^n, B \in \mathsf{Sym}_k$ and A^i is PSD for $i \in [n]$, the ε-noise sensitivity of f satisfies

$$\mathsf{NS}_\varepsilon(f) = \Pr_{\substack{(\boldsymbol{x},\boldsymbol{y})\\ \varepsilon-\text{correlated}}}[f(\boldsymbol{x}) \ne f(\boldsymbol{y})] \le O(\sqrt{\varepsilon}).$$

For simplicity let us assume that $\varepsilon = 1/m$, for some integer m which divides n (since NS_ε is a non-decreasing function in ε, we can even round ε down to satisfy this condition).

In order to analyze $\mathsf{NS}_\varepsilon(f)$ we first observe that one can generate an ε-correlated pair of strings $(x,y) \in \{-1,1\}^n$ as follows:

[15] There is a $+1$ compared to Kane's result to ensure that the result is valid for $k = 1$.

1. Pick a uniformly random string $\boldsymbol{z} \sim \mathcal{U}_n$.

2. Randomly partition $[n]$ into m disjoint buckets $C_1, \ldots, C_m \subseteq [n]$ such that $\cup_i C_i = [n]$. Furthermore, for $z \in \{-1,1\}^n$ (picked in step 1), split each bucket as follows: for every $\ell \in [m]$, split C_ℓ into $C_{\ell,1}$ and $C_{\ell,-1}$ such that $C_{\ell,1}$ corresponds to the *positive* coordinates in z_{B_ℓ} and $C_{\ell,-1}$ corresponds to the *negative* coordinates in z_{C_ℓ}. So overall there are $2m$ disjoint buckets $\{C_{\ell,s} : \ell \in [m], s \in \{-1,1\}\}$ such that $\cup_{\ell,s} C_{\ell,s} = [n]$. Set $\tilde{C}_\ell = C_{\ell,1}$ if $\ell \leq n$ and $\tilde{C}_\ell = C_{\ell-n,-1}$ if $\ell > n$.

3. Corresponding to each bucket $\tilde{C}_\ell$, pick a uniformly random bit $\mathbf{b}_\ell \sim \mathcal{U}_1$.

4. Obtain $\boldsymbol{x}$ as follows: for every $\ell \in [2m]$, obtain $\boldsymbol{x}$ from $\boldsymbol{z}$ by multiplying all the bits in $\boldsymbol{z}_{\tilde{C}_\ell}$ by $\mathbf{b}_\ell$.

5. We obtain $\boldsymbol{y}$ as follows: pick a uniformly random $\ell \in [m]$ and flip the signs of $\boldsymbol{x}_i$ (obtained in step 4) for all the indices i in C_ℓ, i.e., $\boldsymbol{y}_i = -\boldsymbol{x}_i$ if $i \in C_\ell$ and $\boldsymbol{y}_i = \boldsymbol{x}_i$ otherwise.

Observe that the the $(\boldsymbol{x}, \boldsymbol{y})$ obtained in step $(4,5)$ are uniform and ε-correlated. To see this, first observe that the probability of obtaining $x \in \{-1,1\}^n$ is given by

$$\begin{aligned}
\Pr_{\substack{\boldsymbol{z}\sim\mathcal{U}_n,\\ \{C_k\},\boldsymbol{b}\sim\mathcal{U}_m}}[\boldsymbol{x} = x] &= \Pr_{\boldsymbol{z},C,\boldsymbol{b}}[\boldsymbol{z}_{C_1} \cdot \boldsymbol{b}_1 = x_{C_1}, \ldots \boldsymbol{z}_{C_m} \cdot \boldsymbol{b}_m = x_{C_m}] \\
&= \sum_{i=1}^{m} \Pr_{\boldsymbol{z},C,\boldsymbol{b}}\left[\boldsymbol{z}_{C_i} \cdot \boldsymbol{b}_i = x_{C_i} | \boldsymbol{z}_{C_{<i}} \cdot \boldsymbol{b}_{<i} = x_{C_{<i}}\right] \\
&= \sum_{i=1}^{m} \Pr_{\boldsymbol{z},C,\boldsymbol{b}}[\boldsymbol{z}_{C_i} \cdot \boldsymbol{b}_i = x_{C_i}] = \sum_{i=1}^{m} \frac{1}{2^{|C_i|}} = \frac{1}{2^n},
\end{aligned}$$

where the third equality is because $\boldsymbol{z}, \boldsymbol{b}$ are uniformly random and final equality is because $\cup_k C_k = [n]$. In order to see $(\boldsymbol{x}, \boldsymbol{y})$ are ε-correlated, observe that for a fixed $i \in [n]$ the probability $\boldsymbol{x}_i$ differs from $\boldsymbol{y}_i$ is exactly the probability i lies in the bucket $C_\ell = C_{\ell,1} \cup C_{\ell,-1}$ picked in Step (5) above. The probability of picking a bucket C_ℓ is exactly $1/m = \varepsilon$. This event happens independently over all the coordinates $i \in [m]$, hence y is ε-correlated with x.

Now that we have shown $(\boldsymbol{x}, \boldsymbol{y})$ are ε-correlated, we next observe that for a fixed z and buckets $\tilde{C}_1, \ldots, \tilde{C}_{2m}$, we can write $f : \{-1,1\}^n \to \{0,1\}$ as a function $g : \{-1,1\}^{2m} \to \{0,1\}$ defined as

$$g(b) = \left[\sum_{q=1}^{2m} b_q \sum_{j \in \tilde{C}_q} z_j A^j \preceq B\right]. \tag{36}$$

Similarly, one can define $f(y)$ as $g(b')$ where b' is obtained from b by picking a uniformly random $\boldsymbol{\ell} \sim [m]$ and flipping $b_{\boldsymbol{\ell}}, b_{\boldsymbol{\ell}+m}$ where $C_{\boldsymbol{\ell}}$ is the bucket chosen in Step (5). Furthermore, observe that

$$\mathsf{NS}_\varepsilon(f) = \Pr_{\substack{(\boldsymbol{x},\boldsymbol{y})\\ \varepsilon-\text{correlated}}}[f(\boldsymbol{x}) \neq f(\boldsymbol{y})] = \Pr_{\substack{\boldsymbol{b}\sim\mathcal{U}_{2m},\\ \boldsymbol{\ell}\sim[m]}}[g(\boldsymbol{b}) \neq g(\boldsymbol{b}^{\boldsymbol{\ell},\boldsymbol{\ell}+m})],$$

where $\boldsymbol{b}^{i,j}$ is obtained by flipping the i, jth coordinates in $\boldsymbol{b}$ and $\boldsymbol{\ell}$ are chosen uniformly random in $[m]$. We can further upper bound the quantity above by

$$\begin{aligned}
\mathsf{NS}_\varepsilon(f) &= \Pr_{\substack{\boldsymbol{b}\sim\mathcal{U}_{2m},\\ \boldsymbol{\ell}\sim[m]}}[g(\boldsymbol{b}) \neq g(\boldsymbol{b}^{\boldsymbol{\ell},\boldsymbol{\ell}+m})] \\
&\leq \Pr_{\substack{\boldsymbol{b}\sim\mathcal{U}_{2m},\\ \boldsymbol{\ell}\sim[m]}}[g(\boldsymbol{b}) \neq g(\boldsymbol{b}^{\boldsymbol{\ell}})] + \Pr_{\substack{\boldsymbol{b}\sim\mathcal{U}_{2m},\\ \boldsymbol{\ell}\sim[m]}}[g(\boldsymbol{b}^{\boldsymbol{\ell}}) \neq g(\boldsymbol{b}^{\boldsymbol{\ell},\boldsymbol{\ell}+m})] \\
&= \Pr_{\substack{\boldsymbol{b}\sim\mathcal{U}_{2m},\\ \boldsymbol{\ell}\sim[m]}}[g(\boldsymbol{b}) \neq g(\boldsymbol{b}^{\boldsymbol{\ell}})] + \Pr_{\substack{\boldsymbol{b}\sim\mathcal{U}_{2m},\\ \boldsymbol{\ell}\sim[m]}}[g(\boldsymbol{b}) \neq g(\boldsymbol{b}^{\boldsymbol{\ell}+m})] = \frac{1}{m}\mathsf{AS}(g),
\end{aligned} \tag{37}$$

where the second equality used the fact that $\boldsymbol{b}, \boldsymbol{\ell}$ are uniform over their respective domains and the last equality used the definition of $\mathsf{AS}(g)$ to obtain

$$\mathsf{AS}(g) = \sum_{\ell=1}^{2m} \Pr[g(\boldsymbol{x}) \neq g(\boldsymbol{x}^\ell)] = \sum_{\ell=1}^{m} \Pr[g(\boldsymbol{x}) \neq g(\boldsymbol{x}^\ell)] + \sum_{\ell=m+1}^{2m} \Pr[g(\boldsymbol{x}) \neq g(\boldsymbol{x}^\ell)].$$

We now finally upper bound the average sensitivity of g. Observe that $\sum_{j\in\tilde{C}_q} \boldsymbol{z}_j A^j$ is either PSD or NSD (since all the $\boldsymbol{z}_j$ in the bucket $\tilde{C}_q$ have the same sign and A^js are all PSD by definition). From Eq. (36), it is not hard to verify that g is a unate function. Hence, we have

$$\mathsf{NS}_\varepsilon(f) \leq \frac{1}{m}\mathsf{AS}(g) \leq O\left(\sqrt{\frac{1}{m}}\right) = O(\sqrt{\varepsilon}), \tag{38}$$

where the first inequality is by Eq. (37), second inequality uses Theorem 40 and the last equality used the definition of $m = 1/\varepsilon$.

In order to conclude the proof, we use the following two well-known results. Using the standard bits-to-Gaussians tricks [O'D14, Chapte 11], observe that

$$\mathsf{GNS}_\varepsilon(f) \leq \mathsf{NS}_\varepsilon(f). \tag{39}$$

For an explicit proof of this statement, see [DGJ+10, Proposition 9.2].[16] The second result we use is by Ball [Bal13][17] who showed that: if a Boolean function f is an indicator function of a convex set S, i.e., $f^{-1}(1) = S$ and if S has a smooth boundary, then the Gaussian surface area of S can be bounded as

$$\mathsf{GSA}(S) \leq \lim_{\varepsilon\to 0} \frac{\mathsf{GNS}_\varepsilon(f)}{\sqrt{\varepsilon}}. \tag{40}$$

Putting together Eq. (40) and Eq. (39), we get

$$\mathsf{GSA}(S) \leq \lim_{\varepsilon\to 0} \frac{\mathsf{NS}_\varepsilon(f)}{\sqrt{\varepsilon}} \leq O(1),$$

where the final inequality used the upper bound we derived earlier in Eq. (38). This concludes the proof of the theorem. □

[16] We note that [DGJ+10, Proposition 9.2] shows this statement with *equality* asymptotically (i.e., when we taken k Bernoulli's to approximate a Gaussian for $k \to \infty$) for f being a degree-d polynomial threshold function, and the same proof holds true when f is an intersection of spectrahedrons.

[17] We remark that one can also obtain this bound [Kan11a, Section 3].

We now prove Corollary 38 which bounds the Gaussian surface area of intersections of positive spectrahedrons.

Proof of Corollary 38. The proof is very similar to the proof of the theorem above. Let $m = \lceil 1/\varepsilon \rceil$. We follow the same bucketing steps $(1) - (5)$ in Theorem 37 to obtain a $g : \{-1,1\}^{2m} \to \{0,1\}$ given by

$$g(b) = \left[\sum_{q=1}^{2m} b_q \sum_{j \in C_q} z_1 A_1^j \preceq B_1 \right] \cdot \left[\sum_{q=1}^{2m} b_q \sum_{j \in C_q} z_1 A_2^j \preceq B_2 \right].$$

Observe that g is an intersection of positive spectrahedrons and by definition each positive spectrahedron is a unate function. So, by Theorem 40, we have

$$\mathsf{AS}(g) \leq O(\sqrt{m}) = \sqrt{1/\varepsilon}.$$

Repeating the same steps after Eq. (38), we get that

$$\mathsf{GSA}(S^1 \cap S^2) \leq \lim_{\varepsilon \to 0} \frac{\mathsf{GNS}_\varepsilon(F)}{\sqrt{\varepsilon}} \leq \lim_{\varepsilon \to 0} \frac{\mathsf{NS}_\varepsilon(F)}{\sqrt{\varepsilon}} \leq \lim_{\varepsilon \to 0} \sqrt{\varepsilon \cdot \mathsf{AS}(g)} \leq O(1).$$

This concludes the proof of the corollary. □

5.2 Boolean Anti-concentration: Littlewood Offord for spectrahedrons

We now prove the main lemma which shows that the largest eigenvalues of positive spectrahedrons cannot be very concentrated. In particular, we show that for a uniformly random $\boldsymbol{x} \sim \mathcal{U}_n$, suppose we consider a spectrahedron $D = \sum_i \boldsymbol{x}_i A^i - B$ then the measure (over the Boolean cube) that D has largest eigenvalue in a small interval is fairly small. This anti-concentration statement will be crucial in our invariance principle proof when we move from the Bentkus mollifier to our CDF function. In the passing we remark that, prior to this work, we aren't even aware if the weaker *Gaussian* analogue of this statement was known (in particular, the results of [HKM13, ST17] only require Gaussian anti-concentration for which they use a result of Nazarov [Naz03] as a black-box).

We remark that the proof of our main theorem (stated below) follows the result of [OST19, Kan14a] closely since they are able to handle intersections of unate functions which is the case for positive spectrahedrons. However, there are two subtleties.

- (i) In [OST19] they bucket the set of halfspaces (which form the polytope) and show that each bucket has significant weight. Crucially for them, they use the fact that intersections of halfspaces are still unate functions. But this is not the case for positive spectrahedrons. For this, we need to modify the bucketing procedure (akin to what happens in the proof of Theorem 37) so that this bucketing of positive spectrahedrons still results in a unate function.
- (ii) In [OST19] they prove an analogue of Lemma 46 which shows that each bucket has "significant weight". However our proof deviates significantly from the proof in [OST19]. For them, proving the statement in the lemma (for diagonal matrices), follows directly from Paley-Zygmund inequality, but as far as we are aware, we do not have a matrix-version of this inequality. Due to this difficulty, we modify their proof and use the matrix Chernoff bound to prove the statement above.

Theorem 42. *Let $k \geq 0$ be an integer and $\tau \leq \frac{1}{100\sqrt{\log k}}$. Let $\{B_1, B_2\} \subseteq \mathsf{Sym}_k$, $\{A_1^i\}_{i\in[n]}$ and $\{A_2^i\}_{i\in[n]}$ be sequences of* PSD *and* NSD *matrices, respectively. They satisfy that for all $i \in [n], j \in [2]$, $A_1^i \preceq \tau \cdot \mathbb{I}, A_2^i \succeq -\tau\mathbb{I}$ and $\sum_{i=1}^n (A_j^i)^2 \succeq \mathbb{I}$. Then for every $\Lambda \geq 20\tau \log k$, we have*

$$\Pr_{\boldsymbol{x}\sim\mathcal{U}_n}\left[\exists j \in [2] \ s.t. \ \lambda_{\max}\left(\sum_i \boldsymbol{x}_i A_j^i - B_j\right) \in (-\Lambda, \Lambda]\right] \leq O(\Lambda).$$

Again using the standard bits-to-Gaussians trick, we have the following corollary.

Corollary 43. *Let $k \geq 0$ be an integer and $\tau \leq \frac{1}{\log k}$. Let $\{B_1, B_2\} \subseteq \mathsf{Sym}_k$, $\{A_1^i\}_{i\in[n]}$ and $\{A_2^i\}_{i\in[n]}$ be sequences of* PSD *and* NSD *matrices, respectively. They satisfy that for all $i \in [n], j \in [2]$, $A_1^i \preceq \tau \cdot \mathbb{I}, A_2^i \succeq -\tau\mathbb{I}$ and $\sum_i (A_j^i)^2 \succeq \mathbb{I}$. Then for every $\Lambda \geq 20\tau \log k$, we have*

$$\Pr_{\boldsymbol{g}\sim\mathcal{G}^n}\left[\exists j \in [2] \ s.t. \ \lambda_{\max}\left(\sum_i \boldsymbol{g}_i A_j^i - B_j\right) \in (-\Lambda, \Lambda]\right] \leq O(\Lambda).$$

In order to prove this theorem we will use the following two lemmas by [OST19]. Before stating these lemmas, we introduce a few definitions from [OST19] (adapted to our setting of positive spectrahedrons). For the rest of the section, we let $F : \{-1,1\}^n \to \{0,1\}$ be the indicator of an intersection of positive spectrahedrons, i.e., for every $j \in [2]$, let $F_j(x) = \left[\sum_{i=1}^n x_i A_j^i \preceq B_j\right]$, where $\{A_j^i\}_i$ satisfy Eq. (7) and

$$F(x) = \bigwedge_{j=1}^{2} F_j(x) = \bigwedge_{j=1}^{2}\left[\sum_{i=1}^n x_i A_j^i \preceq B_j\right]. \tag{41}$$

1. For a set $S \subseteq \{-1,1\}^n$, let $\mathcal{E}(S)$ be the fraction of $n \cdot 2^{n-1}$ edges which have one endpoint in S and one endpoint in S^c (i.e., complement of S).
2. We let $H_j \subseteq \{-1,1\}^n$ be the *indicator-set* for F_j, i.e., $x \in H_j$ if and only if $F_j(x) = 1$. Additionally, suppose we have sets $\{\bar{H}_1, \bar{H}_2\}$ such that $H_j \subseteq \bar{H}_j$ such that $\bar{H}_j$ are also the indicator-sets of unate functions. Let $\partial H_j = \bar{H}_j \backslash H_j$.
3. For $\alpha \in [0,1]$, we say ∂H_j is *α-semi thin* if for every $x \in H_j$, at least an α-fraction of its hypercube-neighbours (i.e., set of $y \in \{-1,1\}^n$ for which $d(x,y) = 1$) are outside ∂H_j.
4. We now define a few sets: let

$$F = \bar{H}_1 \cap \bar{H}_2, \qquad F^\circ = H_1 \cap H_2, \qquad \partial F = F \backslash F^\circ$$

With this terminology, we have the following lemma that bounds the number of edges that cross F.

Lemma 44 ([OST19, Theorem 7.18])**.** *For $j \in [2]$, let H_j be as defined above. Suppose H_j is α-semi thin, then*

$$\mathsf{vol}(\partial F) \leq O\left(\frac{1}{\alpha\sqrt{n}}\right)$$

Using this lemma, we get the following theorem (which is the analogue of [OST19, Theorem 7.19]).

Theorem 45. *Let $\lambda > 0, \alpha \in [0,1], \{B_1, B_2\} \subseteq \mathsf{Sym}_k$. Let $\{A^i_j\}_{i\in[n],j\in[2]} \subseteq \mathsf{Sym}_k$ satisfy that $A^i_1 \succeq 0, A^i_2 \preceq 0$ for all $i \in [n]$. At least α-fraction of $i \in [n]$ satisfy that $A^i_1 \succeq \lambda \cdot \mathbb{I}$ and $A^i_2 \preceq -\lambda \cdot \mathbb{I}$. Then, we have*

$$\Pr_{\boldsymbol{x}\sim\mathcal{U}_n}\left[\exists j \in [2] \text{ s.t. } \lambda_{\max}\left(\sum_i \boldsymbol{x}_i A^i_j - B_j\right) \in (-2\lambda, 0]\right] \leq O\left(\frac{1}{\alpha\sqrt{n}}\right).$$

Proof. Let $\{A^i_j\}, \{B_j\}$ be as in the theorem statement. Let

$$H_j = \left\{x \in \{-1,1\}^n : \lambda_{\max}\left(\sum_i x_i A^i_j - B_j\right) \leq -2\lambda\right\}, \quad \bar{H}_j = \left\{x \in \{-1,1\}^n : \lambda_{\max}\left(\sum_i x_i A^i_j - B_j\right) \leq 0\right\}.$$

Clearly we then have that

$$\partial H_j = \left\{x \in \{-1,1\}^n : \lambda_{\max}\left(\sum_i x_i A^i_j - B_j\right) \in (-2\lambda, 0]\right\}$$

and

$$\partial F = \left\{x \in \{-1,1\}^n : \exists j \in [2] \text{ s.t. } \lambda_{\max}\left(\sum_i x_i A^i_j - B_j\right) \in (-2\lambda, 0]\right\}.$$

Since we assumed that at least an α-fraction of is satisfied $A^i_1 \succeq \lambda \cdot \mathbb{I}$ and $A^i_2 \preceq -\lambda \cdot \mathbb{I}$, it follows that H_j is α-semi thin, hence we can apply Lemma 44 to obtain the theorem statement. □

Using this theorem, we are now ready to prove our main technical lemma which says that we can always "randomly bucket" our positive spectrahedron so that many of these buckets have "pretty large" smallest eigenvalue.

Lemma 46. *Let $\{A^i\}_{i\in[n]} \subseteq \mathsf{Sym}_k$ be a sequence of positive semidefinite matrices which is (τ, M)-regular with $\tau \leq \frac{1}{100\sqrt{\log k}}$. Let $m \geq \frac{1}{10\tau^2 \log k}$ and $\pi : [n] \to [m]$ be a random hash function that independently assigns each $i \in [n]$ to a uniformly random bucket in $[m]$. For $c \in [m]$, let*

$$\sigma_c = \sum_{j\in\pi^{-1}(c)} A^j$$

and we say the bucket $c \in C$ is good *if $\sigma_c \succeq \frac{1}{2\tau m} \cdot \mathbb{I}$. Then,*

$$\Pr\left[\text{at most } 3m/4 \text{ buckets } c \in [m] \text{ are good }\right] \leq \exp\left(-m/4\right).$$

Proof. Let $\boldsymbol{z}_i \in \{0,1\}$ be a random variable satisfying $\Pr[\boldsymbol{z}_i = 1] = 1/m$. Let $Z_i = \boldsymbol{z}_i \cdot A^i$, hence one can write $\sigma_c = \sum_i Z_i$. In particular, this implies

$$\mathbb{E}\left[\sigma_c\right] = \frac{1}{m}\sum_i A^i \succeq \frac{1}{\tau \cdot m}\sum_i \left(A^i\right)^2 \succeq \frac{1}{\tau m}.$$

Applying Fact 5 (for $\delta = 1/2$, $\mu = 1/\tau m$, $R = \tau$) we have

$$\Pr\left[\sum_i Z_i \succeq \frac{1}{2\tau m}\mathbb{I}\right] \geq 1 - k \cdot \left(\frac{2}{e}\right)^{\frac{1}{2\tau^2 m}} \geq \frac{9}{10}$$

For $j \in [n]$ and $c \in [m]$ define random variables

$$Y_{c,j} = \begin{cases} 1 \text{ if } \pi(j) = c \\ 0 \text{ otherwise,} \end{cases} \qquad \text{and} \qquad X_j = \left[\sum_{c=1}^{m} Y_{c,j}\sigma_c \succeq \frac{1}{2\tau m}\mathbb{I} \right].$$

Using the Claim 47 below, $X_1, \ldots, X_n$ are negatively associated. Thus we may apply the Chernoff bound to $\sum_{i=1}^{m} X_i$ which has mean at least $3m/4$, which gives us the lemma statement.

Claim 47. *The random variables $X_1, \ldots, X_n$ are negatively associated.*

Proof. From [DP09, Page 35, Example 3.1], the set of random variables $\{Y_{c,j}\}_{1\leq c\leq m}$ are negatively associated for $j \in [n]$. Note that $\{Y_{1,j}, \ldots, Y_{m,j}\}_{j\in[n]}$ are n independent families of random variables. By [DP09, Page 35], $\{Y_{c,j}\}_{c\in[m],j\in[n]}$ are negatively associated. Given $\sigma_1, \ldots, \sigma_m$, $\left[\sum_{c=1}^{m} Y_{c,j}\sigma_c \succeq \frac{1}{2\tau m}\mathbb{I}\right]$ is a monotone non-decreasing function of $Y_{c,1}, \ldots, Y_{c,n}$. Thus from [DP09, Page 35], $X_1, \ldots, X_m$ are negatively associated. □

The proof of this claim concludes the proof of the lemma. □

We are now ready to proof our main theorem.

Proof of Theorem 42. For $j \in [2]$, let $f_j(x) = \sum_{i=1}^{n} x_i A_j^i$. Let $\pi : [n] \to [2m]$ be a random hash function that independently assigns each $i \in [n]$ to uniformly random bucket in $[2m]$. Let $C_1, \ldots, C_{2m} \subseteq [n]$ be the buckets and $z \in \{-1,1\}^{2m}$ be uniformly random. Consider the function $g_j : \{-1,1\}^{2k} \to \mathsf{Sym}_k$ defined as

$$g_j(z) = \sum_{q=1}^{2m} z_q \cdot \sum_{i\in C_q} A_j^i.$$

For $q \in [2m]$, define $\bar{A}_j^q = \sum_{i\in C_q} A_j^i$, so $g_j(z) = \sum_q z_q \bar{A}_j^q$. Observe that distribution of f_j and g_j are the same, i.e., for every $D \in \mathsf{Sym}_k$ we have

$$\Pr_{\boldsymbol{z}\sim\mathcal{U}_{2m},\{C_i\}}[g_j(\boldsymbol{z}) = D] = \Pr_{\boldsymbol{x}\sim\mathcal{U}_n}[f_j(\boldsymbol{x}) = D]. \tag{42}$$

In order to see this we argue that the n-bit string $w \in \{-1,1\}^n$ defined as $w_i = z_q$ iff $i \in C_q$, is uniformly random. To show this, we first prove the following: for $z \in \{-1,1\}^{2m}$, let $S = \{q \in [2m] : z_q = 1\}$ and $T = \cup_{q\in S} C_q$. Then, observe that for every $T \subseteq [n]$, we have $\Pr_{\boldsymbol{z},\{C_q\}}[\mathbf{T} = T] = 2^{-n}$ (for every $i \in [n]$, the probability of $i \in C_q$ is $1/(2m)$ and the probability C_q is included in T is $1/2$ since z_q is a uniformly random bit, hence for every $i \in [n]$, we have $\Pr_{\boldsymbol{z},\{C_q\}}[i \in T] = \sum_{i=1}^{2m}(1/2m) \cdot (1/2) = 1/2$ and this is independent for every $i \in [n]$ by construction). It is now easy to see that w is uniformly random because

$$\Pr_{\boldsymbol{z},\{C_j\}}[W = w] = \sum_T \Pr[\mathbf{T} = T] \cdot \Pr[W = w|\mathbf{T} = T] = \frac{1}{2^n}\sum_T \Pr[W = w|\mathbf{T} = T] = 2^{-n},$$

where the last equality used the fact that once we fix T, then all the bits of w which are 1 are fixed.

For $m = \frac{1}{20\tau^2 \log k}$, let $\pi : [n] \to [2m]$ be a random hash that buckets these n variables (jointly for $j \in [2]$). By Lemma 46, we argued that, with probability at least $1 - e^{-m/2}$, at least $9m/5$ of the $2m$ buckets are good for $j = 1$, i.e., a good bucket $q \in [2m]$ for $j = 1$ satisfies $\sum_{i\in\pi^{-1}(q)} A_1^i \succeq \frac{1}{4\tau m} \cdot \mathbb{I}$. For the same reason, with probability at least $1 - e^{-m/2}$, at least $9m/5$ of the $2m$ buckets are good for $j=2$, i.e., a good bucket $q \in [2m]$ for $j=2$ satisfies $\sum_{i\in\pi^{-1}(q)} A_2^i \preceq -\frac{1}{4\tau m}\cdot \mathbb{I}$. Applying a union bound, at least $8m/5$ of $2m$ buckets are good for every $j\in[2]$ with probability at least $1-2\cdot e^{-m/2}$.

By the argument in the start of the proof, we know that after bucketing, we can convert each f_j into a function $g_j : \{-1,1\}^{2m} \to \mathsf{Sym}_k$ such that f_j and g_j have the same distribution. Now we can invoke Theorem 45 as follows: we know that a 4/5-fraction of $q\in[2m]$ satisfy $\bar{A}_1^q \succeq \frac{1}{4\tau m}\cdot\mathbb{I}$ and $\bar{A}_2^q \preceq -\frac{1}{4\tau m}\cdot \mathbb{I}$, so we have

$$\Pr_{\boldsymbol{z}\sim\mathcal{U}_m}\left[\exists j\in[2] \text{ s.t. } \lambda_{\max}\left(\sum_{q=1}^m \boldsymbol{z}_q\bar{A}_j^q - B_j\right)\in(-1/2\tau m, 0]\right] \le O\left(\sqrt{\frac{1}{m}}\right) + 2e^{-m/2}.$$

We now prove the main theorem statement. In order to do so, first observe that, we can partition the bound on the LHS into $\lceil 2\Lambda\tau m\rceil$ intervals as $\Lambda \ge 1/2\tau m$ from our choice of parameters.[18] and by a union bound we have

$$\begin{aligned}&\Pr_{\boldsymbol{x}\sim\mathcal{U}_n}\left[\exists j\in[2] \text{ s.t. } \lambda_{\max}\left(\sum_i \boldsymbol{x}_i A_j^i - B_j\right)\in(-\Lambda,\Lambda]\right]\\ \le\ & O\left(\Lambda\cdot\tau\cdot m\left(\sqrt{\frac{1}{m}}+\exp(-\Omega(m/2))\right)\right)\end{aligned}$$

From the choice of the parameters, the first term above dominates. And thus

$$\Pr_{\boldsymbol{x}\sim\mathcal{U}_n}\left[\exists j\in[2] \text{ s.t. } \lambda_{\max}\left(\sum_i \boldsymbol{x}_i A_j^i - B_j\right)\in[-\Lambda,0]\right] \le O(\Lambda).$$

Similarly one can also show when the LHS of the equation above is replaced with $(0,\Lambda]$. Hence we get our theorem statement. □

6 Invariance principle for positive spectrahedrons

In this section, we establish our main invariance principles.

6.1 Invariance principle for smooth spectral functions

We now prove our main lemma which is an invariance principle for the Bentkus mollifier. We remark that our analysis is the standard Lindeberg-style argument for proving invariance principles, but when applied to the spectral Bentkus mollifier. We first write out the Fréchet series for the Bentkus mollifier, which we then upper bound using our main Theorem 28. In order to understand the error terms in the Fréchet series, we use the matrix Rosenthal inequality (in Fact 7) in order to understand the moments of random matrices (we remark that this inequality will also be useful in our PRG construction). Superficially, our proof techniques resemble the previous invariance principle proofs used in [HKM13, ST17, OST19], but the quantities we need to bound are very different from their analysis.

[18]To be precise, for a vector $v\in\mathbb{R}^k$, observe that the event $[\forall i\in[k] : v_i \le b_i+\Lambda,$ and $\exists j\in[k] : v_j \ge b_j - \Lambda]$ can be broken down into the intersections of $\Lambda/2\tau m$ events given by $\bigwedge_{\ell=0}^{2\Lambda\tau m-1}[\forall i\in[k] : v_i\le b_i+\Lambda-\ell/2\tau m,$ and $\exists j\in[\ell] : v_j > b_j-\Lambda-(\ell+1)/2\tau m]$.

Lemma 48. *Let $k \geq 1, \theta, \tau \in (0,1)$ and $\Psi_\theta : \mathsf{Sym}_k \to \mathbb{R}$ be defined as $\Psi_\theta(M) = (G_\theta \circ \lambda)(M)$ where G_θ is the Bentkus mollifier defined in Eq. (18). Let S_1, S_2 be (τ, M)-regular positive spectrahedrons specified by matrices $\{A_1^1, \ldots, A_1^n, B_1\}$ and $\{A_2^1, \ldots, A_2^n, B_2\}$ respectively. Let $A^i = \operatorname{diag}(A_1^i, A_2^i)$ and $B = \operatorname{diag}(B_1, B_2)$ be block diagonal matrices. Then*

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_n} \left[\Psi_\theta \left(\sum_{i=1}^n \boldsymbol{x}_i A^i - B \right) \right] - \mathop{\mathbb{E}}_{\boldsymbol{g} \sim \mathcal{G}^n} \left[\Psi_\theta \left(\sum_{i=1}^n \boldsymbol{g}_i A^i - B \right) \right] \right| \leq O\left(\frac{\log^7 k}{\theta^3} \cdot (M + \|B\|^2) \cdot (M \cdot \tau)^{1.5} \right).$$

This inequality holds if $\boldsymbol{x}$ is $(10 \log k)$-wise uniform.

Proof. Let $t = \lceil 1/\tau \rceil$. Let $\mathcal{H} = \{h : [n] \to [t]\}$ be a family of $(10 \log k)$-wise uniform hashing functions, i.e., for every subset $I \subseteq [n]$ of size at most $10 \log k$, and $b \in [t]^I$, we have

$$\Pr_{\boldsymbol{h} \in \mathcal{H}} [\boldsymbol{h}(i) = b_i] = \frac{1}{t^{|I|}},$$

where the probability is taken over a uniformly random function $h \in \mathcal{H}$. Fix an $h \in \mathcal{H}$ (think of h as a partition of $[n]$ into t blocks $S_1, \ldots, S_t \subseteq [n]$, where $S_i = h^{-1}(i)$ for all $i \in [t]$). For $\boldsymbol{x} \sim \mathcal{U}_n$ and $\boldsymbol{y} \sim \mathcal{G}^n$ let us divide $\boldsymbol{x}, \boldsymbol{y}$ into blocks $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^t$ and $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^t$ according to h. It is not hard to see that $\boldsymbol{x}^i \sim_{\text{uniform}} \{-1,1\}^{|h^{-1}(i)|}$ and $\boldsymbol{y}^i \sim \mathcal{G}^{|h^{-1}(i)|}$. We now upper bound the quantity

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_n} \left[\Psi_\theta \left(\sum_{i=1}^n \boldsymbol{x}_i A^i - B \right) \right] - \mathop{\mathbb{E}}_{\boldsymbol{y} \in \mathcal{G}^n} \left[\Psi_\theta \left(\sum_{i=1}^n \boldsymbol{y}_i A^i - B \right) \right] \right| \tag{43}$$

by the standard hybrid argument. Let $\{Z^0, \ldots, Z^t\}$ be a set of random variable on n coordinates such that Z^0 is the uniform distribution on $\{-1,1\}^n$ and Z^t is uniform in $\mathcal{G}^n$. To this end, define Z^ℓ as follows: for $j \in [\ell]$, let $Z^\ell_{|h^{-1}(j)} = \boldsymbol{y}^j$ and for $\ell < j \leq t$ let $Z^\ell_{|h^{-1}(j)} = \boldsymbol{x}^j$. It is easy to see that $Z^0 \sim \mathcal{U}_n$ and $Z^t \sim \mathcal{G}^n$. We now can upper bound Eq. (43) as

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_n} \left[\Psi_\theta \left(\sum_{i=1}^n \boldsymbol{x}_i A^i - B \right) \right] - \mathop{\mathbb{E}}_{\boldsymbol{y} \sim \mathcal{G}^n} \left[\Psi_\theta \left(\sum_{i=1}^n \boldsymbol{y}_i A^i - B \right) \right] \right| \\
&= \left| \sum_{\ell=1}^t \mathop{\mathbb{E}}_{\substack{\boldsymbol{x} \sim \mathcal{U}_n \\ \boldsymbol{y} \sim \mathcal{G}^n}} \left[\Psi_\theta \left(\sum_{i=1}^n Z_i^\ell A^i - B \right) \right] - \mathop{\mathbb{E}}_{\substack{\boldsymbol{x} \sim \mathcal{U}_n \\ \boldsymbol{y} \sim \mathcal{G}^n}} \left[\Psi_\theta \left(\sum_{i=1}^n Z_i^{\ell-1} A^i - B \right) \right] \right| \qquad (44) \\
&\leq \sum_{\ell=1}^t \left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{x} \sim \mathcal{U}_n \\ \boldsymbol{y} \sim \mathcal{G}^n}} \left[\Psi_\theta \left(\sum_{i=1}^n Z_i^\ell A^i - B \right) \right] - \mathop{\mathbb{E}}_{\substack{\boldsymbol{x} \sim \mathcal{U}_n \\ \boldsymbol{y} \sim \mathcal{G}^n}} \left[\Psi_\theta \left(\sum_{i=1}^n Z_i^{\ell-1} A^i - B \right) \right] \right|
\end{aligned}$$

We now upper bound each of the t quantities on the RHS of Eq. (44). Fix $\ell \in [t]$ and let us assume for simplicity that $h^{-1}(\ell) = [m]$. By definition of Z^ℓ we observe that $Z_j^\ell = Z_j^{\ell+1}$ for all $j \in \{m+1, \ldots, n\}$ and in fact we have

$$Z^\ell = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m, Z_{m+1}, \ldots, Z_n), \quad Z^{\ell+1} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m, Z_{m+1}, \ldots, Z_n),$$

where $\boldsymbol{x}_i \sim \mathcal{U}_1$ and $y_i \in \mathcal{G}$ is uniform in their respective domains. Crucially note that $Z_{m+1}, \ldots, Z_n$ is independent of the $\boldsymbol{x}_i$s or $\boldsymbol{y}_i$s by definition of $Z^\ell, Z^{\ell+1}$. Rewriting the ℓ-th term in Eq. (44), we get

$$\left| \mathop{\mathbb{E}}_{\substack{\boldsymbol{x} \sim \mathcal{U}_n \\ \boldsymbol{y} \sim \mathcal{G}^n}} \left[\Psi_\theta \left(\underbrace{\sum_{i=1}^m \boldsymbol{x}_i A^i}_{Q} + \underbrace{\sum_{i=m+1}^n Z_i A^i - B}_{P} \right) \right] - \mathop{\mathbb{E}}_{\substack{\boldsymbol{x} \sim \mathcal{U}_n \\ \boldsymbol{y} \sim \mathcal{G}^n}} \left[\Psi_\theta \left(\underbrace{\sum_{i=1}^m \boldsymbol{y}_i A^i}_{R} + \underbrace{\sum_{i=m+1}^n Z_i A^i - B}_{P} \right) \right] \right| \tag{45}$$

Let us analyze both these quantities separately. We can first write the Fréchet series for both these expressions as

$$\Psi_\theta(Q+P) = \Psi_\theta(P) + D\Psi_\theta\left(P\right)[Q] + \frac{1}{2}D^2\Psi_\theta\left(P\right)[Q,Q] + \frac{1}{6}D^3\Psi_\theta\left(P'\right)[Q,Q,Q] \tag{46}$$

where $P' = P + \xi Q$ for some $\xi \in [0,1]$.[19]

$$\Psi_\theta(R+P) = \Psi_\theta(P) + D\Psi_\theta\left(P\right)[R] + \frac{1}{2}D^2\Psi_\theta\left(P\right)[R,R] + \frac{1}{6}D^3\Psi_\theta\left(P''\right)[R,R,R], \tag{47}$$

where $P'' = P + \xi' R$ for some $\xi \in [0,1]$.

Now, observe that since the first moment and the second moment of $\boldsymbol{x}$ match with the standard normal distributions. Thus we have that

$$\begin{aligned}\mathop{\mathbb{E}}_{\substack{\boldsymbol{x}\sim\mathcal{U}_n\\ \boldsymbol{y}\sim\mathcal{G}^n}}\left[D\Psi_\theta\left(P\right)[R]\right] &= \mathop{\mathbb{E}}_{\substack{\boldsymbol{x}\sim\mathcal{U}_n\\ \boldsymbol{y}\sim\mathcal{G}^n}}\left[D\Psi_\theta\left(P\right)[Q]\right]\\ \mathop{\mathbb{E}}_{\substack{\boldsymbol{x}\sim\mathcal{U}_n\\ \boldsymbol{y}\sim\mathcal{G}^n}}\left[D^2\Psi_\theta\left(P\right)[R,R]\right] &= \mathop{\mathbb{E}}_{\substack{\boldsymbol{x}\sim\mathcal{U}_n\\ \boldsymbol{y}\sim\mathcal{G}^n}}\left[D^2\Psi_\theta\left(P\right)[Q,Q]\right].\end{aligned} \tag{48}$$

So by taking the difference of Eq. (47) and Eq. (46), only the third order spectral derivatives remain to be bounded. For this, we now use the Corollary 29 and obtain

$$\left|D^3\Psi_\theta\left(P'\right)[Q,Q,Q]\right| \le O\left(\frac{\Delta_1^2}{\theta^3}\log^3 k\cdot\|Q\|^3\right) \tag{49}$$

$$\left|D^3\Psi_\theta\left(P''\right)[R,R,R]\right| \le O\left(\frac{\Delta_2^2}{\theta^3}\log^3 k\cdot\|R\|^3\right). \tag{50}$$

where $\Delta_1 = \|P'\|$ and $\Delta_2 = \|P''\|$.

Thus, the absolute value of Eq. (45) is upper bounded by

$$\frac{\log^3 k}{\theta^3}\mathbb{E}\left[\Delta_1^2\|Q\|^3 + \Delta_2^2\|R\|^3\right] \le \frac{\log^3 k}{\theta^3}\left(\mathbb{E}\left[\|P'\|^4\right]^{1/2}\mathbb{E}\left[\|Q\|^6\right]^{1/2} + \mathbb{E}\left[\|P''\|^4\right]^{1/2}\mathbb{E}\left[\|R\|^6\right]^{1/2}\right), \tag{51}$$

where the inequality is by Cauchy-Schwarz inequality.

Using Fact 6 and the fact that $\sum_i (A^i)^2 \preceq M\cdot\mathbb{I}$, we have

$$\mathbb{E}\left[\|P'\|^4\right] \le O\left(\log^2 k\cdot M^2 + \|B\|^4\right), \quad \mathbb{E}\left[\|P''\|^4\right] \le O\left(\log^2 k\cdot M^2 + \|B\|^4\right) \tag{52}$$

We now upper bound the last term in Eq. (51) using the following claim.

Claim 49. *It holds that* $\mathbb{E}\left[\|Q\|^6\right] \le O\left(\log^6 k\cdot\tau^3\cdot M^3\right)$, $\quad\mathbb{E}\left[\|R\|^6\right] \le O\left(\log^6 k\cdot\tau^3\cdot M^3\right)$.

Before proving this claim, observe that combining Claim 49 with Eq. (52), (51), we can upper bound Eq. (51) (and in turn Eq. (45)) by

$$O\left(\frac{\log^3 k}{\theta^3}\cdot\left(M\log k + \|B\|^2\right)\cdot\left(\log^3 k\cdot\tau^{1.5}\cdot M^{1.5}\right)\right) \le O\left(\frac{\log^7 k}{\theta^3}\cdot\left(M + \|B\|^2\right)\cdot(M\cdot\tau)^{1.5}\right)$$

[19] This follows directly from the mean value theorem for Fréchet derivatives [AP95].

Putting together this inequality with Eq. (44), we finally get

$$\Big| \mathop{\mathbb{E}}_{x\sim\mathcal{U}_n}\left[\Psi_\theta\left(\sum_{i=1}^n x_i A^i\right)\right] - \mathop{\mathbb{E}}_{y\sim\mathcal{G}^n}\left[\Psi_\theta\left(\sum_{i=1}^n y_i A^i\right)\right]\Big| \le O\left(\frac{\log^7 k}{\theta^3}\cdot(M+\|B\|^2)\cdot(M\cdot\tau)^{1.5}\right),$$

concluding the theorem proof. We now prove the claim above.

Proof of Claim 49. Note that $Q = \sum_{i=1}^n \boldsymbol{x}_i A^i$, where $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)$ is i.i.d. with $\Pr[\boldsymbol{x}_i = 1] = \Pr[\boldsymbol{x}_i = -1] = \frac{1}{2t}$ and $\Pr[\boldsymbol{x}_i = 0] = 1 - 1/t$. Then using Fact 7, we have

$$\begin{aligned}\mathbb{E}\left[\|Q\|_{8p}^{8p}\right]^{1/8p} &\le \sqrt{8p-1}\Big\|\left(\frac{1}{t}\sum_i (A^i)^2\right)^{1/2}\Big\|_{8p} + (8p-1)\left(\frac{1}{t}\sum_i \|A^i\|_{8p}^{8p}\right)^{1/8p}\\ &\le \sqrt{8p-1}\cdot\sqrt{\frac{M}{t}}\cdot k^{\frac{1}{8p}} + (8p-1)\left(\frac{\tau^{8p-2}\cdot k\cdot M}{t}\right)^{1/8p}\end{aligned}$$

where the second inequality used $\sum_i (A^i)^2 \preceq M\cdot\mathbb{I}$ for both terms and $0 \preceq A^i \preceq \tau\mathbb{I}$ for upper bounding the second term. Setting $p = 10\log k$, $t = 1/\tau$ we have

$$\mathbb{E}\left[\|Q\|_{8p}^{8p}\right]^{1/8p} \le O\left(\sqrt{\log k}\cdot\sqrt{\tau}\cdot\sqrt{M} + \log k\cdot\tau\cdot(M/\tau)^{1/(80\log k)}\right) = O\left(\log k\cdot\sqrt{\tau}\cdot\sqrt{M}\right).$$

Thus, we have

$$\mathbb{E}\left[\|Q\|^6\right] \le \mathbb{E}\left[\|Q\|_{8p}^{8p}\right]^{\frac{3}{4p}} \le O\left(\log^6 k\cdot\tau^3\cdot M^3\right),$$

where in the first inequality note that the LHS is the spectral norm and the RHS is the $(8p)$-Schatten norm. This proves the first inequality in the claim statement. The second inequality in the claim follows by the exact same argument (since Fact 7 applies to even $\sum_i \boldsymbol{g}_i A^i$). □

The proof of this claim concludes the proof of the theorem. □

6.2 Invariance principle for positive spectrahedrons

We are now ready to prove our main theorem now, which involves combining our anti-concentration Theorem 42 and our invariance principle for Bentkus mollifier in Lemma 48.[20]

Theorem 50. *Let $k \ge 1$, $M \ge 1, \gamma \ge 1, \tau \in [0,1], \delta \in [0,1]$. Let S_1, S_2 be (τ, M)-regular positive spectrahedrons specified by matrices $\{A_1^1,\dots,A_1^n,B_1\} \in \mathsf{Sym}_k$ and $\{A_2^1,\dots,A_2^n,B_2\} \in \mathsf{Sym}_k$ respectively satisfying $\|B_1\|,\|B_2\| \le \gamma$. Let $S = S_1\cap S_2$. If μ is a $(10\log k)$-wise uniform distribution over $\{-1,1\}^n$, then*

$$\left|\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[\boldsymbol{x}\in S] - \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{G}^n}[\boldsymbol{g}\in S]\right| \le C\cdot\left(M+\gamma^2\right)^{1/5}\cdot\log^{7/5} k\cdot M^{3/10}\cdot\tau^{3/10},$$

for some universal constant $C > 0$.

[20] We remark that our theorem statements should also hold true for a larger class of *proper distributions* as considered in [HKM13], which requires one to extend our main Theorem 19 to show that even the 4th order spectral derivatives can be bounded by $\|f^{(4)}\|_1$. We believe this should be possible and leave this to be made rigorous for future work.

Proof. Again for notational simplicity, let $A^i = \mathrm{diag}\left(A_1^i, A_2^i\right)$ and $B = \mathrm{diag}\left(B_1, B_2\right)$ be block diagonal matrices. We conclude the result by combining Fact 24, Lemma 48 and Corollary 42 as follows: first Lemma 48 implies

$$\left|\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}\left[\Psi_\theta\left(\sum_{i=1}^n \boldsymbol{x}_i A^i - B\right)\right] - \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{G}^n}\left[\Psi_\theta\left(\sum_{i=1}^n \boldsymbol{g}_i A^i - B\right)\right]\right| \le O\left(\frac{\log^7 k}{\theta^3}\cdot\left(M+\|B\|^2\right)\cdot(M\cdot\tau)^{1.5}\right),$$

In particular, using Fact 24 (for $D = B - \beta\cdot\mathbb{I}$ and $D = B + \beta\cdot\mathbb{I}$), the "if" condition of Fact 24 is satisfied with

$$\eta = O\left(\frac{\log^7 k}{\theta^3}\cdot\left(M+(\gamma+\beta)^2\right)\cdot(M\cdot\tau)^{1.5}\right)$$

where $\beta = O(\theta\cdot\sqrt{\log k/\delta})$. In particular, Fact 24 and Corollary 43 now together imply that

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}\left[\Psi\left(\sum_{i=1}^n \boldsymbol{x}_i A^i - B\right)\right] - \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{G}^n}\left[\Psi\left(\sum_{i=1}^n \boldsymbol{g}_i A^i - B\right)\right]\right| \\
&\le \gamma + 3\delta + \Pr_{\boldsymbol{g}\sim\mathcal{G}^n}\left[\lambda_{\max}\left(\sum_{i=1}^n \boldsymbol{g}_i A^i - B\right)\in[-\Lambda,\Lambda]\right] \\
&= O\left(\frac{\log^7 k}{\theta^3}\cdot\left(M+\left(\gamma+\theta\cdot\sqrt{\log(k/\delta)}\right)^2\right)\cdot(M\cdot\tau)^{1.5}+\delta+\Lambda\right) \\
&\le O\left(\frac{\log^7 k}{\theta^3}\cdot\left(M+\left(\gamma+\sqrt{\log(k/\delta)}\right)^2\right)\cdot(M\cdot\tau)^{1.5}+\delta+\Lambda\right)
\end{aligned}$$

Let us fix

$$\theta \leftarrow \delta, \quad \theta \leftarrow \Lambda, \quad \left((M\cdot\tau)^{1.5}\cdot\log^7 k\cdot\left(M+\left(\gamma+\sqrt{\log k}\right)^2\right)\right)^{1/5} \leftarrow \theta.$$

This gives us

$$\left|\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}\left[\Psi\left(\sum_{i=1}^n \boldsymbol{x}_i A^i - B\right)\right] - \mathop{\mathbb{E}}_{\boldsymbol{g}\sim\mathcal{G}^n}\left[\Psi\left(\sum_{i=1}^n \boldsymbol{g}_i A^i - B\right)\right]\right| \le \left((M\cdot\tau)^{1.5}\cdot\log^7 k\cdot(M+\gamma^2)\right)^{1/5}.$$

□

6.3 Application: Pseudorandom generators for positive spectrahedrons.

We are now ready to describe our pseudorandom generator for fooling positive spectrahedrons. Our PRG is based on the well-known construction of Meka and Zuckerman [MZ13] which we describe now. We remark that the same PRG (with minor modifications and different parameter settings) was used in [MZ13, HKM13, ST17] in order to obtain PRGs for polytopes.

Meka-Zuckerman PRG. We begin by describing the Meka-Zuckerman PRG. Let us fix the parameters $\delta\in(0,1)$, $\tau = \Omega(\delta^{10/3}/(\log^5 k\cdot M\cdot(M+\gamma^2)))$ so that we have $\left(M+\gamma^2\right)^{1/5}\cdot\log^{7/5} k\cdot M^{3/10}\cdot\tau^{3/10} = \delta$ (where the LHS of this equality is the upper bound obtained in our invariable principle proof). Let $t = \lceil 1/\tau\rceil$ and consider the family of $(2\log k)$-wise uniform functions $\mathcal{H} = \{h : [n]\to[t]\}$, i.e., for every for every subset $I\subseteq[n]$ of size at most $5\log k$, and $b\in[t]^I$, we have

$$\Pr_{\boldsymbol{h}\in\mathcal{H}}\left[\boldsymbol{h}(i) = b_i\right] = \frac{1}{t^{|I|}},$$

where the probability is taken over a uniformly random function $h \in \mathcal{H}$. Efficient constructions of such hash function families are known with $|\mathcal{H}| = O(n^{5 \log k})$. For simplicity (as in the proof of [MZ13, HKM13]), we also assume that for every $j \in [t]$, we have $|h^{-1}(j)| = n/t$. Let $m = n/t$ and $G_0 : \{0,1\}^s \to \{-1,1\}^m$ generate a $(10 \log k)$-wise uniform distribution over $\{-1,1\}^m$, i.e., for every $I \subseteq [n]$ of size at most $5 \log k$ and $b \in \{-1,1\}^I$, we have

$$\Pr_{\substack{\boldsymbol{z} \in \{0,1\}^s \\ \boldsymbol{x} = G_0(\boldsymbol{z})}} [\boldsymbol{x}_i = \boldsymbol{b}_i \text{ for all } i \in I] = \frac{1}{2^{|I|}},$$

where the probability is taken over uniformly random $z \in \{0,1\}^s$. It is well-known by [NN93] that efficient constructions of generators G_0 are known for $s = O(\log k \log n)$. Finally, we are ready to describe the Meka-Zuckerman generator: for a given hash function family $\mathcal{H}$ and generator G_0, define $G : \mathcal{H} \times (\{0,1\}^s)^t \to \{-1,1\}^n$ by

$$G(h, z^1, \ldots, z^t) = x, \qquad \text{where } x_{|h^{-1}(i)} = G_0(z^i) \text{ for } i \in [t].$$

Clearly the seed length of this generator is

$$O\left((\log n)(\log k) + (\log n)(\log k)\frac{1}{\tau}\right) = O((\log n)(\log k)/\tau) = (\log n) \cdot \operatorname{poly}(\log k, M, 1/\delta, \gamma),$$

where the first term is the logarithm of the number of elements of the hash function family $|\mathcal{H}|$, the second term because we have $s = O((\log n)(\log k))$ and recall that we picked $t = O(1/\tau)$ and the final equality used the bound on τ we fixed at the start of the proof.

We now restate our main theorem and prove it.

Theorem 51. *Let $\delta \in (0,1)$, $k, n, M \geq 1$ and $\tau \leq \delta^{10/3}/(\log^5 k \cdot M \cdot (M + \gamma^2))$. Let S_1, S_2 be (τ, M)-regular positive spectrahedrons specified by matrices $\{A_1^1, \ldots, A_1^n, B_1\} \in \mathsf{Sym}_k$ and $\{A_2^1, \ldots, A_2^n, B_2\} \in \mathsf{Sym}_k$ with $\|B_1\|, \|B_2\| \leq \gamma$. Let $S = S_1 \cap S_2$. There exists a* PRG *$G : \{0,1\}^r \to \{-1,1\}^n$ with*

$$r = (\log n) \cdot \operatorname{poly}(\log k, M, 1/\delta, \gamma)$$

that δ-fools S with respect to the uniform distribution.

The proof of this theorem is a generic statement that allows one to go from invariance principles proven using the proof techniques to construct PRGs. The proof uses the same proof ideas of Harsha, Klivans and Meka [HKM13, Section 7.2] (except that now we directly proved *Boolean* anti-concentration instead of the weaker *Gaussian* anti-concentration as proven by [HKM13]). We provide the proof below for completeness.

Proof. Again for notational simplicity, let $A^i = \operatorname{diag}\left(A_1^i, A_2^i\right)$ and $B = \operatorname{diag}(B_1, B_2)$ be block diagonal matrices. The PRG G will be the Meka-Zuckerman PRG defined above, so the seed length $r = (\log n) \cdot \operatorname{poly}(\log k, M, 1/\delta, \gamma)$ immediately follows.

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_r} \left[\Psi_\theta \left(\sum_{i=1}^n (G(\boldsymbol{x}))_i A^i - B \right) \right] - \mathop{\mathbb{E}}_{\boldsymbol{g} \sim \mathcal{G}^n} \left[\Psi_\theta \left(\sum_{i=1}^n \boldsymbol{g}_i A^i - B \right) \right] \right| \leq O\left(\frac{\log^7 k}{\theta^3} \cdot (M + \|B\|^2) \cdot (M \cdot \tau)^{1.5} \right), \tag{53}$$

where we used the fact that $G(x)$ for uniformly random $x \in \{0,1\}^r$ generates a $(10 \log k)$-wise uniform distribution and Lemma 48 holds for every $(10 \log k)$-wise uniform distribution μ. Repeating the same calculation that we did in the proof of Theorem 50, we get

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_r} \left[\Psi \left(\sum_{i=1}^{n} (G(\boldsymbol{x}))_i A^i - B \right) \right] - \mathop{\mathbb{E}}_{\boldsymbol{g} \sim \mathcal{G}^n} \left[\Psi \left(\sum_{i=1}^{n} \boldsymbol{g}_i A^i - B \right) \right] \right| \\
&\leq \gamma + 3\delta + \Pr_{\boldsymbol{g} \sim \mathcal{G}^n} \left[\lambda_{\max} \left(A \left(\boldsymbol{g} \right) \right) \in (-\Lambda, \Lambda] \right] \\
&= O\left(\frac{\log^7 k}{\theta^3} \cdot (M + \|B\|^2) \cdot (M \cdot \tau)^{1.5} + \delta + \Lambda \right),
\end{aligned}$$

and using our assumption on τ (and the same parameters as in Theorem 50), this implies that

$$\left| \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathcal{U}_r} [G(\boldsymbol{x}) \in S] - \mathop{\mathbb{E}}_{\boldsymbol{g} \sim \mathcal{G}^n} [\boldsymbol{g} \in S] \right| \leq \delta,$$

hence proving our theorem statement. □

References

[AHK05] Sanjeev Arora, Elad Hazan, and Satyen Kale. Fast algorithms for approximate semidefinite programming using the multiplicative weights update method. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 339–348. IEEE, 2005. 1

[AK07] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 227–236, 2007. 1

[AP95] Antonio Ambrosetti and Giovanni Prodi. *A primer of nonlinear analysis*, volume 34. Cambridge University Press, 1995. 41

[AS10] Brendan P.W. Ames and Hristo S. Sendov. Asymptotic expansions of the ordered spectrum of symmetric matrices. *Nonlinear Analysis: Theory, Methods & Applications*, 72(11):4288 – 4297, 2010. 6

[AS12] Brendan P.W. Ames and Hristo S. Sendov. A new derivation of a formula by Kato. *Linear Algebra and its Applications*, 436(3):722 – 730, 2012. 6

[AS16] Brendan P.W. Ames and Hristo S. Sendov. Derivatives of compound matrix valued functions. *Journal of Mathematical Analysis and Applications*, 433(2):1459 – 1485, 2016. 6

[AZLO16] Zeyuan Allen-Zhu, Yin Tat Lee, and Lorenzo Orecchia. Using optimization to obtain a width-independent, parallel, simpler, and faster positive SDP solver. In *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1824–1831, 2016. 1

[Bal93] Keith Ball. The reverse isoperimetric problem for Gaussian measure. *Discrete & Computational Geometry*, 10(4):411–420, 1993. 7, 31

[Bal13] Keith Ball. Talk: Noise sensitivity and Gaussian surface area, 2013. https://www.mi.fu-berlin.de/en/math/groups/discgeom/dates/2nd_ERC_Workshop/slides/ball_talk_beamer.pdf. 7, 34

[Baz09] Louay MJ Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009. 11

[Ben90] Vidmantas Bentkus. Smooth approximations of the norm and differentiable functions with bounded support in Banach space ℓ_∞^k. *Lithuanian Mathematical Journal*, 30(3):223–230, 1990. 2, 5, 6, 17, 18

[Bha00] Rajendra Bhatia. Pinching, trimming, truncating, and averaging of matrices. *The American Mathematical Monthly*, 107(7):602–608, 2000. 13

[Bha13] Rajendra Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013. 6, 15, 32

[BHK$^+$19] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019. 1

[BLZ05] Jan Brinkhuis, Zhi-Quan Luo, and Shuzhong Zhang. Matrix convex functions with applications to weighted centers for semidefinite programming, 2005. 6, 15

[Boo05] Carl de Boor. *Divided differences*. Surv. Approx. Theory 1, 2005. 13

[BPT12] Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, 2012. 1

[BS99] Rajendra Bhatia and Kalyan B. Sinha. Derivations, derivatives and chain rules. *Linear Algebra and its Applications*, 302-303:231 – 244, 1999. 5

[BSS98] Rajendra Bhatia, Dinesh Singh, and Kalyan B. Sinha. Differentiation of operator functions and perturbation bounds. *Communications in Mathematical Physics*, 191:603–611, 1998. 5

[CDS19] Eshan Chattopadhyay, Anindya De, and Rocco A Servedio. Simple and efficient pseudorandom generators from Gaussian processes. In *34th Computational Complexity Conference (CCC 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019. 1, 2, 3, 6, 7

[Col12] Rodney Coleman. *Calculus on normed vector spaces*. Springer Science & Business Media, 2012. 14, 15

[CQT03] Xin Chen, Houduo Qi, and Paul Tseng. Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems. *SIAM Journal on Optimization*, 13(4):960–985, 2003. 5

[DGJ$^+$10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010. 1, 2, 3, 7, 32, 34

[DHK+10] Ilias Diakonikolas, Prahladh Harsha, Adam Klivans, Raghu Meka, Prasad Raghavendra, Rocco A Servedio, and Li-Yang Tan. Bounding the average sensitivity and noise sensitivity of polynomial threshold functions. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 533–542, 2010. 1, 7

[DP09] Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, 2009. 38

[Erd45] Paul Erdös. On a lemma of Littlewood and Offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945. 8

[Fel68] Willliam Feller. *An introduction to probability theory and its applications, vol 1*. New York: Wiley, 1968. 18

[FF88] Péter Frankl and Z Furedi. Solution of the Littlewood-Offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988. 8

[FK20] Xiao Fang and Yuta Koike. High-dimensional central limit theorems by Stein's method. *arXiv preprint arXiv:2001.10917*, 2020. 17, 18

[FMP+15] Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Exponential lower bounds for polytopes in combinatorial optimization. *Journal of the ACM (JACM)*, 62(2):1–23, 2015. 1

[GKM18] Parikshit Gopalan, Daniel M Kane, and Raghu Meka. Pseudorandomness via the discrete Fourier transform. *SIAM Journal on Computing*, 47(6):2451–2487, 2018. 1

[GM12] Bernd Gärtner and Jiri Matousek. *Approximation algorithms and semidefinite programming*. Springer Science & Business Media, 2012. 1, 10

[GOWZ10] Parikshit Gopalan, Ryan O'Donnell, Yi Wu, and David Zuckerman. Fooling functions of halfspaces under product distributions. In *2010 IEEE 25th Annual Conference on Computational Complexity*, pages 223–234. IEEE, 2010. 1, 3

[GW95] Michel X. Goemans and David P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42(6):1115–1145, 1995. 1

[GW13] Gus Gutoski and Xiaodi Wu. Parallel approximation of min-max problems. *Computational Complexity*, 22:385 – 428, 2013. 1

[Hau92] David Haussler. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and computation*, 100(1):78–150, 1992. 9

[HKM13] Prahladh Harsha, Adam Klivans, and Raghu Meka. An invariance principle for polytopes. *Journal of the ACM (JACM)*, 59(6):1–25, 2013. , 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 25, 35, 39, 42, 43, 44

[HMV06] J William Helton, Scott A McCullough, and Victor Vinnikov. Noncommutative convexity arises from linear matrix inequalities. *Journal of Functional Analysis*, 240(1):105–191, 2006. 10

[JJUW11] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. QIP = PSPACE. *Journal of the ACM*, 58(6), 2011. 1

[JLL+20] Arun Jambulapati, Yin Tat Lee, Jerry Li, Swati Padmanabhan, and Kevin Tian. Positive semidefinite programming: Mixed, parallel, and width-independent. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 789–802, 2020. 1

[JUW09] R. Jain, S. Upadhyay, and J. Watrous. Two-message quantum interactive proofs are in PSPACE. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 534–543, 2009. 1

[JY11] R. Jain and P. Yao. A parallel approximation algorithm for positive semidefinite programming. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 463–471, 2011. 1

[Kan10] Daniel M. Kane. k-independent Gaussians fool polynomial threshold functions. *arXiv preprint arXiv:1012.1614*, 2010. 1

[Kan11a] Daniel M. Kane. The Gaussian surface area and noise sensitivity of degree-d polynomial threshold functions. *computational complexity*, 20(2):389–412, 2011. 1, 7, 34

[Kan11b] Daniel M. Kane. k-independent Gaussians fool polynomial threshold functions. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity, CCC*, pages 252–261. IEEE Computer Society, 2011. 1

[Kan11c] Daniel M. Kane. A small PRG for polynomial threshold functions of Gaussians. In Rafail Ostrovsky, editor, *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS*, pages 257–266. IEEE Computer Society, 2011. 1

[Kan14a] Daniel Kane. The average sensitivity of an intersection of half spaces. *Research in the Mathematical Sciences*, 1(1):13, 2014. 7, 8, 11, 32, 35

[Kan14b] Daniel M. Kane. A pseudorandom generator for polynomial threshold functions of Gaussian with subpolynomial seed length. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 217–228. IEEE, 2014. 1

[KKMS08] Adam Tauman Kalai, Adam R Klivans, Yishay Mansour, and Rocco A Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008. 9

[KM15] Pravesh K. Kothari and Raghu Meka. Almost optimal pseudorandom generators for spherical caps. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 247–256, 2015. 1, 11

[KOS04] Adam R Klivans, Ryan O'Donnell, and Rocco A Servedio. Learning intersections and thresholds of halfspaces. *Journal of Computer and System Sciences*, 68(4):808–840, 2004. 9, 31

[KOS08] Adam R Klivans, Ryan O'Donnell, and Rocco A Servedio. Learning geometric concepts via Gaussian surface area. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 541–550. IEEE, 2008. 2, 3, 7, 9

[KSS94] Michael J Kearns, Robert E Schapire, and Linda M Sellie. Toward efficient agnostic learning. *Machine Learning*, 17(2-3):115–141, 1994. 9

[Lew96] A. S. Lewis. Derivatives of spectral functions. *Mathematics of Operations Research*, 21(3):576–588, 1996. 5

[Lin22] J.W. Lindeberg. Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 15:211–225, 1922. 2, 4

[LO39] John Edensor Littlewood and Albert C Offord. On the number of real roots of a random algebraic equation. ii. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 35, pages 133–148. Cambridge University Press, 1939. 8

[LRS15] James R Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 567–576, 2015. 1

[MJC+14] Lester Mackey, Michael I. Jordan, Richard Y. Chen, Brendan Farrell, and Joel A. Tropp. Matrix concentration inequalities via the method of exchangeable pairs. *Ann. Probab.*, 42(3):906–945, 05 2014. 6, 14

[MOO05] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 21–30. IEEE, 2005. 5

[Mos08] Elchanan Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 156–165. IEEE Computer Society, 2008. 5

[MZ13] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM Journal on Computing*, 42(3):1275–1301, 2013. 1, 2, 3, 9, 43, 44

[Naz03] Fedor Nazarov. On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric aspects of functional analysis*, pages 169–187. Springer, 2003. 2, 3, 5, 7, 8, 32, 35

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM journal on computing*, 22(4):838–856, 1993. 44

[NPS08] Jiawang Nie, Pablo A. Parrilo, and Bernd Sturmfels. *Semidefinite Representation of the k-Ellipse*, pages 117–132. Springer New York, New York, NY, 2008. 1

[O'D14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014. 4, 13, 14, 34

[OST19] Ryan O'Donnell, Rocco A Servedio, and Li-Yang Tan. Fooling polytopes. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 614–625, 2019. 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 20, 25, 35, 36, 39

[OST20] Ryan O'Donnell, Rocco A Servedio, and Li-Yang Tan. Fooling Gaussian PTFs via local hyperconcentration. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1170–1183, 2020. 1

[Per04] Yuval Peres. Noise stability of weighted majority. *arXiv math/0412377*, 2004. 7

[PT12] Richard Peng and Kanat Tangwongsan. Faster and simpler width-independent parallel algorithms for positive semidefinite programming. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Parallelism in Algorithms and Architectures*, SPAA '12, page 101–108. Association for Computing Machinery, 2012. 1

[Qua12] Ronan Quarez. Symmetric determinantal representation of polynomials. *Linear algebra and its applications*, 436(9):3642–3660, 2012. 10

[Sch18] Claus Scheiderer. Spectrahedral shadows. *SIAM Journal on Applied Algebra and Geometry*, 2(1):26–44, 2018. 1

[Sen07] Hristo S Sendov. The higher-order derivatives of spectral functions. *Linear algebra and its applications*, 424(1):240–281, 2007. 6, 20, 21

[Ser06] Rocco A Servedio. Every linear threshold function has a low-weight approximator. In *21st Annual IEEE Conference on Computational Complexity (CCC'06)*, pages 18–32. IEEE, 2006. 1, 11

[ST17] Rocco A Servedio and Li-Yang Tan. Fooling intersections of low-weight halfspaces. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 824–835. IEEE, 2017. 1, 2, 3, 6, 7, 8, 9, 10, 11, 25, 35, 39, 43

[Tao10] Terence Tao. 254a notes: Topics in random matrix theory., 2010. https://terrytao.wordpress.com/tag/lindeberg-replacement-trick/. 4

[Tro12] Joel A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12:389–434, 2012. 13

[Tro15] Joel A Tropp. An introduction to matrix concentration inequalities. *arXiv preprint arXiv:1501.01571*, 2015. 6

[Tro16] Joel A. Tropp. The expected norm of a sum of independent random matrices: An elementary approach. In Christian Houdré, David M. Mason, Patricia Reynaud-Bouret, and Jan Rosiński, editors, *High Dimensional Probability VII*, pages 173–202, Cham, 2016. Springer International Publishing. 14

[TV12] Terence Tao and Van Vu. The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi. *Combinatorica*, 32(3):363–372, 2012. 8

[Viz17] Cynthia Vizant. Spectrahedra., 2017. https://clvinzan.math.ncsu.edu/slides/MSRI_SpectrahedraSlides.pdf. 1

[Yao19] Penghui Yao. A doubly exponential upper bound on noisy EPR states for binary games. *arXiv:1904.08832*, 2019. 5, 11

A Proof of Lemma 34: Case 2

Recall that the goal is to prove the following inequality

$$\left| \sum_{i_1 \neq i_2 \neq i_3} \frac{\frac{\overline{g}(x_{i_1}) - \overline{g}(x_{i_3})}{x_{i_3} - x_{i_1}} - \frac{\overline{g}(x_{i_1}) - \overline{g}(x_{i_2})}{x_{i_2} - x_{i_1}}}{x_{i_3} - x_{i_2}} G\left(x\right) H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} \right| \leq O\left(\Delta \log^2 k \|H\|^3\right) \tag{54}$$

First observe that the LHS of the inequality above can be rephrased as follows.

$$\left|\sum_{i_1\neq i_2\neq i_3}\frac{\frac{\bar{g}(x_{i_1})-\bar{g}(x_{i_3})}{x_{i_3}-x_{i_1}}-\frac{\bar{g}(x_{i_1})-\bar{g}(x_{i_2})}{x_{i_2}-x_{i_1}}}{x_{i_3}-x_{i_2}}G\left(x\right)H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|$$

$$=\left|2\sum_{\substack{i_1\neq i_2\neq i_3\\ x_{i_3}>x_{i_2}}}\frac{\frac{g'(x_{i_1})g(x_{i_3})-g(x_{i_1})g'(x_{i_3})}{x_{i_3}-x_{i_1}}g\left(x_{i_2}\right)-\frac{g'(x_{i_1})g(x_{i_2})-g(x_{i_1})g'(x_{i_2})}{x_{i_2}-x_{i_1}}g\left(x_{i_3}\right)}{x_{i_3}-x_{i_2}}G\left(x_{-\{i_1,i_2,i_3\}}\right)H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right| \tag{55}$$

Providing an upper bound on this consists of several lemmas and the result is concluded by combing all of them via triangle inequalities. To keep the expressions short, we use the following notations to represent Eq. (54), which are clear in the context.

$$\left|2\sum_{\substack{i_1\neq i_2\neq i_3\\ x_{i_3}>x_{i_2}}}\frac{\frac{\langle i_1\rangle'\langle i_3\rangle-\langle i_1\rangle\langle i_3\rangle'}{[i_3-i_1]}\langle i_2\rangle-\frac{\langle i_1\rangle'\langle i_2\rangle-\langle i_1\rangle\langle i_2\rangle'}{[i_2-i_1]}\langle i_3\rangle}{[i_3-i_2]}\right|, \tag{56}$$

where we implicitly hide the $G\left(x_{-\{i_1,i_2,i_3\}}\right)H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}$ term. We first give a sketch of how we are going to upper bound this inequality and break it into subsections.

$$(56)=\underbrace{\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_1\rangle'\langle i_3\rangle}{[i_3-i_1]}\cdot\frac{\langle i_3\rangle-\langle i_2\rangle}{[i_3-i_2]}}_{\textit{Section A.1, Lemma 52}}-\underbrace{\frac{\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_1\rangle'\langle i_3\rangle}{[i_3-i_1]}-\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_1\rangle'\langle i_2\rangle}{[i_2-i_1]}}{[i_3-i_2]}\langle i_3\rangle}_{(\star)\ \textit{Section A.1, Remark 1}}.$$

We now break up Remark 1 into two cases

$$(\star)=\underbrace{\textit{Remark }1\cdot\mathbb{I}[\min\{x_{i_1},x_{i_3}\}>x_{i_2}]}_{(\dagger)}+\underbrace{\textit{Remark }1\cdot\mathbb{I}[x_{i_1}<x_{i_2}<x_{i_3}]}_{(\dagger\dagger)}.$$

Note that there are the only two cases we need to handle since by symmetry between i_2 and i_3, we can assume $x_{i_3}>x_{i_2}$, without loss of generality. Now we bound these two terms, separately.

$$(\dagger)=\underbrace{\frac{\frac{\langle i_3\rangle'-\langle i_1\rangle'}{[i_3-i_1]}-\frac{\langle i_2\rangle'-\langle i_1\rangle'}{[i_2-i_1]}}{[i_3-i_2]}\langle i_1\rangle\langle i_3\rangle}_{\textit{Section A.2, Lemma 57}}-\underbrace{\frac{\frac{\langle i_3\rangle-\langle i_1\rangle}{[i_3-i_1]}-\frac{\langle i_2\rangle-\langle i_1\rangle}{[i_2-i_1]}}{[i_3-i_2]}\langle i_1\rangle'\langle i_3\rangle}_{\textit{Section A.2, Lemma 58}}.$$

and

$$(\dagger\dagger)=\underbrace{\frac{\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_3\rangle\langle i_3\rangle'}{[i_3-i_1]}-\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_2\rangle\langle i_2\rangle'}{[i_2-i_1]}}{[i_3-i_2]}\langle i_3\rangle}_{\textit{Section A.3, Lemma 59}}+\underbrace{\frac{\frac{\langle i_3\rangle'-\langle i_1\rangle'}{[i_3-i_1]}\langle i_3\rangle-\frac{\langle i_2\rangle'-\langle i_1\rangle'}{[i_2-i_1]}\langle i_2\rangle}{[i_3-i_2]}\langle i_3\rangle}_{(\P)\ \textit{Section A.3, Remark 2}},$$

and

$$(\P)=\underbrace{\frac{\langle i_3\rangle'-\langle i_1\rangle'}{[i_3-i_1]}\cdot\frac{\langle i_3\rangle-\langle i_2\rangle}{[i_3-i_2]}\cdot\langle i_3\rangle}_{\textit{Section A.4, Lemma 60}}+\underbrace{\frac{\frac{\langle i_3\rangle'-\langle i_1\rangle'}{[i_3-i_1]}-\frac{\langle i_2\rangle'-\langle i_1\rangle'}{[i_2-i_1]}}{[i_3-i_2]}\langle i_2\rangle\langle i_3\rangle}_{\textit{Section A.4, Lemma 61}}$$

Finally in order to upper bound Eq. (56), we simply bound each of these terms by $O\left(\Delta\log^3 k\|H\|^3\right)$ in the respective sections (as underbraced by the terms).

A.1 Case 2.1

Lemma 52.

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}}}\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_1\rangle'\langle i_3\rangle}{[i_3-i_1]}\cdot\frac{\langle i_3\rangle-\langle i_2\rangle}{[i_3-i_2]}\right|\leq O\left(\Delta^2\cdot\log^2 k\cdot\|H\|^3\right).$$

Remark 1. *Using the triangle inequality, it suffices to upper bound*

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}}}\frac{\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_1\rangle'\langle i_3\rangle}{[i_3-i_1]}-\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_1\rangle'\langle i_2\rangle}{[i_2-i_1]}}{[i_3-i_2]}\langle i_3\rangle\right|$$

Proof of Lemma 52. We apply Claim 22 to the first sum and obtain $O\left(\Delta\max\left\{g'\left(x_{i_1}\right),g'\left(x_{i_3}\right)\right\}\right)$ (note that we have $\max\{\cdot,\cdot\}$ to compensate for the fact that $x_{i_1}\geq x_{i_3}$ or $x_{i_3}\geq x_{i_1}$). Therefore, the left hand side in Lemma 52 can be upper bounded by

$$\begin{aligned}&O\left(\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}}}\Delta\left|\max\left\{g'\left(x_{i_1}\right),g'\left(x_{i_3}\right)\right\}\cdot\frac{g\left(x_{i_3}\right)-g\left(x_{i_2}\right)}{x_{i_3}-x_{i_2}}\cdot G\left(x_{-\{i_1,i_2,i_3\}}\right)H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right)\\ \leq\;&O\left(\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}\geq 0,x_{i_1}\geq 0}}(\cdots)+\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}\geq 0,x_{i_1}<0}}(\cdots)+\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2},x_{i_2}<0,x_{i_1}\geq 0}}(\cdots)+\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2},x_{i_1}<0,x_{i_2}<0}}(\cdots)\right)\end{aligned}\tag{57}$$

First term in Eq. (57). Note that $g(x)\geq\frac{1}{2}$ if $x\geq 0$. Since g' is monotone decreasing in the interval $[0,\infty)$, the first summation is upper bounded by

$$\begin{aligned}&O\left(\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}\geq 0,x_{i_1}\geq 0}}\Delta\left|\max\left\{g'\left(x_{i_1}\right)g'\left(x_{i_2}\right)G\left(x_{-i_3}\right),g'\left(x_{i_3}\right)g'\left(x_{i_2}\right)G\left(x_{-i_1}\right)\right\}H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right|\right)\\ \leq\;&O\left(\Delta\cdot\|G^{(2)}\|_1\cdot\max_{i_1,i_2}\sum_{i_3}|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}|\right) &(58)\\ \leq\;&O\left(\Delta\cdot\log k\cdot\max_{i_1,i_2}\sum_{i_3}|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}|\right)\leq O\left(\Delta\cdot\log k\cdot\|H^3\|\right) &(59)\end{aligned}$$

where the first inequality used that $|g'(\zeta_{i_1,i_2})|\leq\max\{|g'(x_{i_1})|,|g'(x_{i_2})|\}$ and the last inequality follows by Eq. (35).

Second term in Eq. (57). The second summation is upper bounded as follows, again by the mean value theorem observe that

$$
\begin{aligned}
& O\left(\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}\geq 0, x_{i_1}<0}} \Delta\left|\max\left\{g'\left(x_{i_1}\right)g'\left(x_{i_2}\right), g'\left(x_{i_3}\right)g'\left(x_{i_2}\right)\right\} G\left(x_{-i_1}\right) H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right) \\
\leq\ & O\left(\Delta\cdot\sum_{i_1:x_{i_1}<0} \|G^{(1)}\left(x_{-i_1}\right)\|_1 \max_{i_2}\sum_{i_3}\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right| + \|G^{(2)}\left(x_{-i_1}\right)\|_1 \max_{i_2,i_3}\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right) \\
\leq\ & O\left(\Delta\cdot\log^{1.5} k\cdot\|H\|^3\right),
\end{aligned}
$$

where the last inequality is from Fact 20 and the assumption that $|\{i : x_i \leq 0\}| \leq 3\log k$.

Third term in Eq. (57). Using the fact that $g'(\cdot)$ is bounded by a constant, the third summation is upper bounded by

$$
\begin{aligned}
& O\left(\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}, x_{i_2}<0, x_{i_1}\geq 0}} \Delta\left|\max\left\{g'\left(x_{i_1}\right), g'\left(x_{i_3}\right)\right\}\cdot G\left(x_{-\{i_2,i_3\}}\right) H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right) \\
=\ & O\left(\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}, x_{i_1}\geq 0, x_{i_2}<0, x_{i_3}\geq 0}} (\cdots) + \sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}, x_{i_1}\geq 0, x_{i_2}<0, x_{i_3}<0}} (\cdots)\right). \qquad (60)
\end{aligned}
$$

For the first summation in Eq. (60), using the fact that $g\left(x\right) \geq \frac{1}{2}$ when $x \geq 0$, it is upper bounded by

$$
\begin{aligned}
& O\left(\Delta\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}, x_{i_1}\geq 0, x_{i_2}<0, x_{i_3}\geq 0}} \left|\max\left\{g'\left(x_{i_1}\right), g'\left(x_{i_3}\right)\right\}\cdot G\left(x_{-\{i_2\}}\right) H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right) \\
\leq\ & O\left(\Delta\sum_{i_2:x_{i_2}<0} \|G^{(1)}\|_1 \max_{i_1}\sum_{i_3}\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right) \\
\leq\ & O\left(\Delta\sum_{i_2:x_{i_2}<0} \sqrt{\log k}\max_{i_1}\sum_{i_3}\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right) \\
\leq\ & O\left(\Delta\cdot\log^{1.5} k\cdot\|H\|^3\right),
\end{aligned}
$$

where the second inequality is from Fact 20, and the last inequality used Eq. (35) and the assumption that $|\{i : x_i \leq 0\}| \leq 3\log k$.

In order to upper bound the second summation in Eq. (60), first observe that both $g\left(\cdot\right)$ and $G\left(\cdot\right)$ are positive and upper bounded by 1. Thus, Eq. (60) can be bounded as

$$
O\left(\Delta\sum_{\substack{i_2\neq i_3:\\ x_{i_2}<0, x_{i_3}<0}} \sum_{i_1}\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right) \leq O\left(\Delta\cdot\log^2 k\cdot\|H\|^3\right).
$$

where we again use Eq. (35) and the assumption that $|\{i : x_i \le 0\}| \le 3\log k$.

Fourth term in Eq. (57). The last summation is upper bounded by $O\left(\Delta \cdot \log^2 k \cdot \|H\|^3\right)$ using the same arguments to upper bound the second summation in Eq. (60). □

A.2 Case 2.2

We upper bound the quantity in Remark 1 in two cases that $x_{i_1} > x_{i_2}$ and $x_{i_2} > x_{i_1}$. In order to prove this lemma we need the following lemmas and claims.

Claim 53. *For integer $k \ge 1$, $X \in \mathsf{Sym}_k$ and $H \in \mathsf{Mat}_k$ it holds that*

$$\left\| (XH + HX)\, e^{-\frac{X^2}{2}} \right\|_2 \le 2\|X\| \cdot \|He^{-X^2/2}\|_2$$

Proof. As the Schattern norm is unitarily invariant, we assume that $X = \mathrm{diag}\,(x_1, \ldots, x_n)$ is diagonal without loss of generality. Then

$$\|(XH + HX)\, e^{-X^2/2}\|_2^2 = \sum_{i,j} H_{i,j}^2 \left(x_i + x_j\right)^2 e^{-x_j^2} \le 4\|X\|^2 \cdot \sum_{i,j} H_{i,j}^2 e^{-x_j^2} = 4\Delta^2 \|He^{-X^2/2}\|_2^2.$$

□

Lemma 54. *Given an integer $k \ge 1$, $u_1, u_2, u_3 \ge 0$ satisfying $u_1 + u_2 + u_3 = 1$ and $X \in \mathsf{Sym}_k, H_1, H_2, H_3 \in \mathsf{Mat}_k$, if $u_1, u_3 \le \frac{1}{2}$, then it holds that*

$$\left|\mathrm{Tr}\left[e^{-u_1 X^2} H_1 e^{-u_2 X^2} H_2 e^{-u_3 X^2} H_3\right]\right| \le \|H_1 e^{-\frac{1}{2}X^2}\|_2 \cdot \|H_2 e^{-\frac{1}{2}X^2}\|_2 \cdot \|H_3\|$$

Proof. Using the inequality $|\mathrm{Tr} ABC| \le \|A\|_2 \cdot \|B\|_2 \cdot \|C\|$ (where $\|\cdot\|_2$ is the standard Frobenius norm and $\|\cdot\|$ is the spectral norm), we have

$$\begin{aligned}
&\left|\mathrm{Tr}\left[e^{-u_1 X^2} H_1 e^{-u_2 X^2} H_2 e^{-u_3 X^2} H_3\right]\right| \\
\le\ & \|e^{-u_1 X^2} H_1 e^{\left(u_1 - \frac{1}{2}\right)X^2}\|_2 \cdot \|e^{-u_3 X^2} H_2 e^{\left(u_3 - \frac{1}{2}\right)X^2}\|_2 \cdot \|H_3\| \\
\le\ & \|H_1 e^{-\frac{1}{2}X^2}\|_2 \cdot \|H_2 e^{-\frac{1}{2}X^2}\|_2 \cdot \|H_3\|
\end{aligned}$$

The last inequality is from Lemma 55 □

Lemma 55. *Given diagonal matrices $A = \mathrm{diag}\,(a_1, \ldots, a_k)\,, B = \mathrm{diag}\,(b_1, \ldots, b_k)$ with $a_1 \ge \cdots \ge a_k \ge 0$ and $b_1 \ge \cdots \ge b_k \ge 0$ and an arbitrary matrix H, it holds that*

$$\|AHB\|_2 \le \|HAB\|_2.$$

Proof. Note that

$$\begin{aligned}
\|HAB\|_2^2 - \|AHB\|_2^2 = \sum_{i,j} H_{ij}^2 \left(a_j^2 b_j^2 - a_i^2 a_j^2\right) &= \sum_{i,j} H_{i,j}^2 \left(\frac{a_i^2 b_i^2 + a_j^2 b_j^2 - a_i^2 b_j^2 - a_j^2 b_i^2}{2}\right) \\
&= \frac{1}{2} \sum_{i,j} \left(a_i^2 - a_j^2\right)\left(b_i^2 - b_j^2\right) \ge 0,
\end{aligned}$$

where the first equality is from the symmetry. □

Lemma 56. *Given an integer $k \geq 1$, matrices $A, B, C \in \mathsf{Mat}_k$ and $X \in \mathsf{Sym}_k$ with $\|X\| \leq \Delta$, it holds that*

$$\begin{aligned}
& \left|\mathrm{Tr}\left[D^2\left(e^{-X^2/2}\right)[A,B]\,C\right]\right| \\
\leq\ & 4\Delta^2 \cdot \max\left\{\begin{array}{c}\|Ae^{-X^2/2}\|_2 \cdot \|Be^{-X^2/2}\|_2 \cdot \|C\|, \|Ae^{-X^2/2}\|_2 \cdot \|Ce^{-X^2/2}\|_2 \cdot \|B\|, \\ \|Be^{-X^2/2}\|_2 \|Ce^{-X^2/2}\|_2 \cdot \|A\| \end{array}\right\}.
\end{aligned}$$

Proof. Combining Lemma 12, Lemma 54 and the inequality that

$$\|(XA+AX)\,e^{-X^2/2}\|_2 \leq 2\|Ae^{-X^2/2}\|_2, \quad \|XA+AX\| \leq 2\Delta\|A\|,$$

we conclude the result. □

Lemma 57.

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_1}>x_{i_2}, x_{i_3}>x_{i_2}}} \frac{\frac{\langle i_3\rangle' - \langle i_1\rangle'}{[i_3-i_1]} - \frac{\langle i_2\rangle' - \langle i_1\rangle'}{[i_2-i_1]}}{[i_3-i_2]} \langle i_1\rangle\langle i_3\rangle\right| \leq O\left(\Delta^2 \cdot \log^{2.5} k\|H\|^3\right)..$$

Proof of Lemma 57. We break the summation into two summations

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3\\ x_{i_1}>x_{i_3}>x_{i_2}}} (\cdots)\right| + \left|\sum_{\substack{i_1\neq i_2\neq i_3\\ x_{i_3}>x_{i_1}>x_{i_2}}} (\cdots)\right| \tag{61}$$

For the first summation, we define

$$A_{i,j} = \begin{cases} H_{i,j}, & \text{if } x_i < x_j \\ 0, & \text{otherwise.}\end{cases}$$

and Then $\|A\| \leq \log k \cdot \|H\|$ by Fact 4 (without loss of generality, we may assume that x_is are sorted in increasing order. Further notice that all the diagonal entries of H are zeros. Thus A is the upper triangle part of H). We first bound the first term in Eq. (61). In this direction, we first rewrite it as

$$\frac{1}{\sqrt{2\pi}}\sum_{i_2} G\left(x_{-i_2}\right)\left(\left(D^2\left(e^{-X^2/2}\right)[A,A^T]H\right)_{i_2,i_2}\right) = \frac{1}{\sqrt{2\pi}}\sum_{i_2<0}(\cdots) + \frac{1}{\sqrt{2\pi}}\sum_{i_2\geq 0}(\cdots), \tag{62}$$

where $X = \mathrm{diag}\left(x_1,\ldots,x_k\right)$ and we implicitly used that we are summing over terms with $x_{i_2} < x_{i_3}$. For the first summation in Eq. (62),

$$\begin{aligned}
& \left|\mathrm{Tr}\left(D^2\left(e^{-X^2/2}\right)[A,A^T]HE_{i_2,i_2}\right)\right| \\
\leq\ & 4\Delta^2 \log k \cdot \max\left\{\|Ae^{-X^2/2}\|_2^2 \cdot \|HE_{i_2,i_2}\|, \|Ae^{-X^2/2}\|_2 \cdot \|He^{-X^2/2}\|_2 \cdot \|AE_{i_2,i_2}\|\right\} \\
\leq\ & 4\Delta^2 \log k\|He^{-X^2/2}\|_2^2 \cdot \|H\|. \qquad (63)
\end{aligned}$$

where the first inequality is by Lemma 56 and the second inequality is because X and E_{i_2,i_2} are diagonal and A is a submatrix of H. Thus, the first summation in Eq. (62) is upper bounded by

$$
\begin{aligned}
&\frac{\Delta^2 \cdot \log k}{\sqrt{2\pi}} \sum_{i_2 : x_{i_2} < 0} G\left(x_{-i_2}\right) \|He^{-X^2/2}\|_2^2 \cdot \|H\| \\
= & \left(\Delta^2 \log k \cdot \sum_{i_2 : x_{i_2} < 0} G\left(x_{-i_2}\right) \sum_{i_1, i_3} e^{-x_{i_3}^2} H_{i_1,i_3}^2 \|H\|\right) \\
\leq & \left(\Delta^2 \log^2 k \cdot \max_{i_2} \sum_{i_1 \neq i_3} g'\left(x_{i_3}\right) \cdot G\left(x_{-i_2}\right) \cdot H_{i_1,i_3}^2 \|H\|\right) \\
\leq & \; O\left(\Delta^2 \log^{2.5} k \|H\|^3\right), \qquad (64)
\end{aligned}
$$

where the first inequality is from the assumption that $|\{i : x_i < 0\}| \leq 3 \log k$ and the second inequality is from Fact 20.

For the second summation in Eq. (62), we define

$$
\tilde{H}_{i,j} = \begin{cases} \frac{H_{i,j}}{g(x_j)}, & \text{if } x_j \geq 0 \\ 0, & \text{otherwise.} \end{cases}
$$

Then $\|\tilde{H}\| \leq 2\|H\|$ as $g\left(x_i\right) \geq \frac{1}{2}$ if $x_i \geq 0$. It is easy to verify that the second summation in Eq. (62) is equal to

$$
\left| \frac{1}{\sqrt{2\pi}} G\left(x\right) \operatorname{Tr} D^2 \left(e^{-X^2/2}\right) [A, A^T] \tilde{H} \right| \leq \frac{\Delta^2 \log k}{\sqrt{2\pi}} G\left(x\right) \|He^{-X^2/2}\|_2^2 \|H\| \leq O\left(\Delta^2 \cdot \log^{1.5} k \|H\|^3\right).
$$

where the first inequality is from Lemma 56 and the second inequality is from Fact 20.

Finally, the second summation in Eq. (61) can be upper bounded using the verbatim same arguments by $O\left(\Delta^2 \cdot \log^{2.5} k \cdot \|H\|^3\right)$. This proves the lemma statement. □

Lemma 58.

$$
\left| \sum_{\substack{i_1 \neq i_2 \neq i_3 : \\ x_{i_3} > x_{i_2}, x_{i_1} > x_{i_2}}} \frac{\frac{\langle i_3 \rangle - \langle i_1 \rangle}{[i_3 - i_1]} - \frac{\langle i_2 \rangle - \langle i_1 \rangle}{[i_2 - i_1]}}{[i_3 - i_2]} \langle i_1 \rangle' \langle i_3 \rangle \right| \leq O\left(\Delta \cdot \log^{1.5} k \cdot \|H\|^3\right)
$$

Proof.

$$
\left| \sum_{\substack{i_1 \neq i_2 \neq i_3 : \\ x_{i_3} > x_{i_2}, x_{i_1} > x_{i_2}}} \frac{\frac{\langle i_3 \rangle - \langle i_1 \rangle}{[i_3 - i_1]} - \frac{\langle i_2 \rangle - \langle i_1 \rangle}{[i_2 - i_1]}}{[i_3 - i_2]} \langle i_1 \rangle' \langle i_3 \rangle \right| = \left| \sum_{\substack{i_1 \neq i_2 \neq i_3 : \\ x_{i_3} > x_{i_2}, x_{i_1} > x_{i_2} \geq 0}} (\cdots) + \sum_{\substack{i_1 \neq i_2 \neq i_3 : \\ x_{i_3} > x_{i_2}, x_{i_1} > 0, x_{i_2} < 0}} (\cdots) \right| \qquad (65)
$$

To upper bound the first summation in Eq. (65), we apply Fact 3 and upper bound the first summation by

$$\begin{aligned}
& O\left(\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2},x_{i_1}>x_{i_2}\geq 0}} \left|g''\left(\xi_{i_1,i_2,i_3}\right)g'\left(x_{i_1}\right)G\left(x_{-\{i_1,i_2\}}\right)H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right)\\
\leq\ & O\left(\left|\Delta\cdot\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2},x_{i_1}>x_{i_2}\geq 0}} g'\left(x_{i_2}\right)g'\left(x_{i_1}\right)G\left(x_{-\{i_1,i_2\}}\right)H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right)\\
\leq\ & O\left(\|G^{(2)}\left(x\right)\|_2\max_{i_1,i_2}\sum_{i_3}\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right)\\
\leq\ & O\left(\Delta\cdot\log k\cdot\|H\|^3\right)
\end{aligned}$$

where the last inequality is from Fact 20 and Eq. (35). Note that $|g''(\xi)|\leq\Delta$ for any $\xi\in[x_{i_2},\max\{x_{i_1},x_{i_3}\}]$ by Eq. (11). Applying Fact 3, the second summation in Eq. (65) is upper bounded by

$$\begin{aligned}
& O\left(\Delta\sum_{i_2:x_{i_2}<0}\sum_{i_1,i_3}g'\left(x_{i_1}\right)G\left(x_{-\{i_1,i_2\}}\right)\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right)\\
\leq\ & O\left(\Delta\cdot\log k\cdot\max_{i_2}\|G^{(1)}\left(x_{-i_2}\right)\|_1\cdot\max_{i_1}\sum_{i_3}\left|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|\right)\\
\leq\ & O\left(\Delta\cdot\log^{1.5}k\cdot\|H\|^3\right)
\end{aligned}$$

where the first inequality is from the assumption that $|\{i:x_i<0\}|\leq 3\log k$ and the second inequality is from Fact 20 and Eq. (35). □

A.3 Case 2.3

We now bound the second case of Remark 1 when $x_{i_3}>x_{i_2}>x_{i_1}$. Recall that the goal is to upper bound the following lemma.

Lemma 59.

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}}\frac{\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_3\rangle\langle i_3\rangle'}{[i_3-i_1]}-\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_2\rangle\langle i_2\rangle'}{[i_2-i_1]}}{[i_3-i_2]}\langle i_3\rangle\right|\leq O\left(\Delta\cdot\log^{1.5}k\cdot\|H\|^3\right)$$

Remark 2. *Combining with Remark 1, it suffices to upper bound*

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}}\frac{\frac{\langle i_3\rangle'-\langle i_1\rangle'}{[i_3-i_1]}\langle i_3\rangle-\frac{\langle i_2\rangle'-\langle i_1\rangle'}{[i_2-i_1]}\langle i_2\rangle}{[i_3-i_2]}\langle i_3\rangle\right|$$

Proof of Lemma 59.

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}} \frac{\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_3\rangle\langle i_3\rangle'}{[i_3-i_1]}-\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_2\rangle\langle i_2\rangle'}{[i_2-i_1]}}{[i_3-i_2]}\langle i_3\rangle\right|$$

$$\leq \sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}} \left|\frac{\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_3\rangle\langle i_2\rangle'}{[i_3-i_1]}-\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_2\rangle\langle i_2\rangle'}{[i_2-i_1]}}{[i_3-i_2]}\langle i_3\rangle\right| + \sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}} \left|\frac{\frac{\langle i_1\rangle\langle i_3\rangle'-\langle i_3\rangle\langle i_3\rangle'}{[i_3-i_1]}-\frac{\langle i_1\rangle\langle i_2\rangle'-\langle i_3\rangle\langle i_2\rangle'}{[i_3-i_1]}}{[i_3-i_2]}\langle i_3\rangle\right|$$

$$= \sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}} \left|\frac{\frac{\langle i_1\rangle-\langle i_3\rangle}{[i_3-i_1]}-\frac{\langle i_1\rangle-\langle i_2\rangle}{[i_2-i_1]}}{[i_3-i_2]}\langle i_2\rangle'\langle i_3\rangle\right| + \sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}} \left|\frac{\langle i_1\rangle-\langle i_3\rangle}{[i_3-i_1]}\cdot\frac{\langle i_3\rangle'-\langle i_2\rangle'}{[i_3-i_2]}\cdot\langle i_3\rangle\right| \tag{66}$$

The first term is upper bounded by $O\left(\Delta\cdot\log^{1.5}k\cdot\|H\|^3\right)$ using the same argument in Lemma 58. The second term can be rephrased as

$$\left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1}}} \frac{g(x_{i_1})-g(x_{i_3})}{x_{i_3}-x_{i_1}}\cdot\frac{g'(x_{i_3})-g'(x_{i_2})}{x_{i_3}-x_{i_2}} g(x_{i_3})\, G(x_{-i_1})\, H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right|$$

$$\leq \left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1},x_{i_1}\geq 0}}(\cdots)\right| + \left|\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1},x_{i_1}<0}}(\cdots)\right| \tag{67}$$

For the first summation in Eq. (67), we apply the mean value theorem for both g and g'. From Eq. (11) it is upper bounded by

$$\begin{aligned}
&\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1},x_{i_1}\geq 0}} \Delta\left|g'(x_{i_1})\, g'(x_{i_2})\, g(x_{i_3})\, G(x_{-i_1})\, H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right| \\
\leq\;& O\left(\Delta\cdot\|G^{(2)}(x)\|_1\|H\|^3\right) \\
\leq\;& O\left(\Delta\cdot\log k\cdot\|H\|^3\right).
\end{aligned}$$

For the second term in Eq. (67), it is not hard to verify that

$$\left|\frac{g'(x_{i_3})-g'(x_{i_2})}{x_{i_3}-x_{i_2}}\right| \leq \Delta\max\left\{g'(x_{i_3}),g(x_{i_2})\right\} \tag{68}$$

Further notice that $|g'(\cdot)|\leq 1$. Applying the mean value theorem to g, we upper bound the second summation in 67 by

$$\begin{aligned}
& O\left(\Delta\sum_{\substack{i_1\neq i_2\neq i_3:\\ x_{i_3}>x_{i_2}>x_{i_1},x_{i_1}<0}} \max\left\{g'(x_{i_3}),g'(x_{i_2})\right\} g(x_{i_3})\, G(x_{-i_1})\, |H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}|\right) \\
\leq\;& O\left(\Delta\cdot\log k\cdot\max_{i_1}\cdot\|G^{(1)}(x_{-i_1})\|_1\cdot\max_{i_2}\sum_{i_3}|H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}|\right) \\
\leq\;& O\left(\Delta\cdot\log^{1.5}k\cdot\|H\|^3\right)
\end{aligned}$$

where the first inequality is from the assumption that $|\{i : x_i < 0\}| \le 3\log k$ and the second inequality is from Fact 20 Eq. (35). ☐

A.4 Case 2.4

In this section, we want to upper bound Remark 2. To do so we write it as the sum of two terms (the first one is easy to bound).

Lemma 60.

$$\left| \sum_{\substack{i_1 \ne i_2 \ne i_3: \\ x_{i_3} > x_{i_2} > x_{i_1}}} \frac{\langle i_3 \rangle' - \langle i_1 \rangle'}{[i_3 - i_1]} \cdot \frac{\langle i_3 \rangle - \langle i_2 \rangle}{[i_3 - i_2]} \cdot \langle i_3 \rangle \right| \le O\left(\Delta \cdot \log^2 k \cdot \|H\|^3\right)$$

Proof of Lemma 60. We split the summation into two cases that $x_{i_2} \ge 0$ and $x_{i_2} < 0$. For the case that $x_{i_2} \ge 0$, we apply the mean value theorem to $g(\cdot)$ and Eq. (68), it is upper bounded by $O\left(\Delta \cdot \log k \cdot \|H\|^3\right)$. For the case that $x_{i_2} < 0$, we have $x_{i_1} < 0$. Note that $|g'(\cdot)| \le 1$. Thus it is upper bounded by

$$O\left(\sum_{\substack{i_1 \ne i_2: \\ x_{i_1} < x_{i_2} < 0}} \sum_{i_3} |H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1}| \right) \le O\left(\log^2 k \cdot \|H\|^3\right).$$

☐

Next, our goal is to prove the following lemma which upper bounds the Remark 2 statement.

Lemma 61.

$$\left| \sum_{\substack{i_1 \ne i_2 \ne i_3: \\ x_{i_3} > x_{i_2} > x_{i_1}}} \frac{\frac{\langle i_3 \rangle' - \langle i_1 \rangle'}{[i_3 - i_1]} - \frac{\langle i_2 \rangle' - \langle i_1 \rangle'}{[i_2 - i_1]}}{[i_3 - i_2]} \langle i_2 \rangle \langle i_3 \rangle \right| \le O\left(\sqrt{\log k} \cdot \|H\|^3\right).$$

Before we prove this lemma, we first prove a "simpler" proposition which will be crucial in upper bound the above.

Proposition 62.

$$\left| \sum_{i_1 \ne i_2 \ne i_3} \frac{\frac{\langle i_1 \rangle' - \langle i_3 \rangle'}{[i_3 - i_1]} - \frac{\langle i_1 \rangle' - \langle i_2 \rangle'}{[i_2 - i_1]}}{[i_3 - i_2]} \right| \le O\left(\Delta^2 \cdot \sqrt{\log k} \cdot \|H\|^3\right)$$

Proof. Using Fact 10,

$$\left| \sum_{i_1 \ne i_2 \ne i_3} \frac{\frac{g'(x_{i_3}) - g'(x_{i_1})}{x_{i_3} - x_{i_1}} - \frac{g'(x_{i_2}) - g'(x_{i_1})}{x_{i_2} - x_{i_1}}}{x_{i_3} - x_{i_2}} G(x) H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} \right| = \frac{1}{\sqrt{2\pi}} \left| \mathrm{Tr}\left[D^2 \left(e^{-\frac{X^2}{2}} \right) [H, H] \cdot H \right] \right| G(x),$$

where $X = \mathrm{diag}(x_1, \ldots, x_n)$. Using Lemma 12, it suffices to upper bound

$$G(x)\left|\mathrm{Tr}\left[e^{-\frac{uX^2}{2}}(XH+HX)e^{-\frac{v(1-u)X^2}{2}}(XH+HX)e^{-\frac{(1-v)(1-u)X^2}{2}}H\right]\right| \tag{69}$$

and

$$G(x)\left|\mathrm{Tr}\left[e^{\frac{(u-1)X^2}{2}}H^2e^{-\frac{uX^2}{2}}H\right]\right| \tag{70}$$

Note that $u + v(1-u) + (1-v)(1-u) = 1$. At least two of these three quantities are at most $\frac{1}{2}$. We upper bound Eq. 69 in the following three cases.

If $u \le \frac{1}{2}$ and $(1-u)(1-v) \le \frac{1}{2}$, using Claim 53 and Lemma 54, Eq. (69) is upper bounded by

$$\|(XH+HX)e^{-\frac{X^2}{2}}\|_2^2\|H\| \le \Delta^2\|He^{-X^2/2}\|_2^2 \cdot \|H\|$$

If $u \le \frac{1}{2}$ and $v(1-u) \le \frac{1}{2}$, then the Eq. (69) is upper bounded by

$$\begin{aligned}\|(XH+HX)e^{-\frac{X^2}{2}}\|_2 \cdot \|He^{-\frac{X^2}{2}}\|_2\|XH+HX\| &\le 2\Delta\|He^{-X^2/2}\|_2^2 \cdot \|XH+HX\| \\ &\le 4\Delta^2\|He^{-X^2/2}\|_2^2 \cdot \|H\|.\end{aligned}$$

where the second last inequality is by Claim 53. The case that $u(1-v) \le \frac{1}{2}$ and $v(1-u) \le \frac{1}{2}$ follows similarly.

Also Eq. (70) can be upper bounded with similar arguments. Thus

$$G(x) \cdot \left|\mathrm{Tr}\ e^{(u-1)X^2/2}H^2e^{-uX^2/2}H\right| \le 4\Delta^2 G(x)\|H\| \cdot \|He^{-X^2/2}\|_2^2. \tag{71}$$

Therefore,

$$\left|\sum_{i_1\neq i_2\neq i_3}\frac{\frac{g'(x_{i_3})-g'(x_{i_1})}{x_{i_3}-x_{i_1}}-\frac{g'(x_{i_2})-g'(x_{i_1})}{x_{i_2}-x_{i_1}}}{x_{i_3}-x_{i_2}}G(x)H_{i_1,i_2}H_{i_2,i_3}H_{i_3,i_1}\right| \tag{72}$$

$$\le \left(G(x)\cdot \mathrm{Tr}\left[D^2\left(e^{-\frac{X^2}{2}}\right)[H,H]\cdot H\right]\right) \tag{73}$$

$$\begin{aligned}
&\le O\left(\Delta^2 G(x)\|H\|\|He^{-X^2/2}\|_2^2\right)\\
&= O\left(\Delta^2\sum_{i_1,i_2}e^{-x_{i_2}^2}H_{i_1,i_2}^2G(x)\cdot\|H\|\right)\\
&\le O\left(\Delta^2\sum_{i_1,i_2}g'(x_{i_2})G(x)H_{i_1,i_2}^2\cdot\|H\|\right)\\
&\le O\left(\Delta^2\sum_{i_1,i_2}g'(x_{i_2})G(x_{-i_2})H_{i_1,i_2}^2\cdot\|H\|\right)\\
&\le O\left(\Delta^2\|G^{(1)}(x)\|_1\cdot\left(\max_{i_2}\sum_{i_1}H_{i_1,i_2}^2\right)\cdot\|H\|\right)\\
&\le O\left(\Delta^2\|G^{(1)}(x)\|_1\cdot\max_{i_2}\left(H^2\right)_{i_2,i_2}\cdot\|H\|\right)
\end{aligned}$$

$$\le O\left(\Delta^2\cdot\sqrt{\log k}\cdot\|H\|^3\right), \tag{74}$$

where the second inequality used $e^{-x_i^2/2} \leq 1$, third inequality used $g(x) \in [0,1]$ and the last inequality is from Fact 20. ∎

We are now ready to prove the main lemma. Note that end of the day we need to bound the case (in Remark 2) when the quantity in Lemma 60 contains $G(x_{-i_1})$ instead of $G(x_{-\{i_1,i_2,i_3\}})$. Observe that the inequality in Lemma 61 can be written as

$$\left| \sum_{\substack{i_1 \neq i_2 \neq i_3 \\ i_1 < i_2, i_1 < i_3}} \frac{\frac{g'(x_{i_1}) - g'(x_{i_3})}{x_{i_3} - x_{i_2}} - \frac{g'(x_{i_1}) - g'(x_{i_2})}{x_{i_2} - x_{i_1}}}{x_{i_3} - x_{i_2}} G\left(x_{-\{i_1\}}\right) H_{i_1,i_2} H_{i_2,i_3} H_{i_3,i_1} \right| \leq O\left(\Delta^2 \cdot \log^{2.5} k \cdot \|H\|^3\right) \tag{75}$$

Observe that in this section we are concerned with $x_{i_1} < x_{i_2} < x_{i_3}$ so the summation in this lemma and the equation above are over the same indices.

Proof of Lemma 61. By the paragraph above, proving this lemma is equivalent to proving Eq. (75). For the first summation above, let

$$\left(A^{i_1}\right)_{i,j} = \begin{cases} H_{i,i_1}, & \text{if } j = i_1 \text{ and } i > i_1 \\ 0, & \text{otherwise.} \end{cases}$$

The left hand side of the claim statement can be expressed as

$$\left| \frac{1}{\sqrt{2\pi}} \sum_{i_1} G\left(x_{-i_1}\right) \left(\mathrm{Tr}\ D^2 \left(e^{-X^2/2}\right) \left[A^{i_1}, \left(A^{i_1}\right)^T\right] H\right) \right| = \left| \frac{1}{\sqrt{2\pi}} \sum_{i_1 : x_{i_1} < 0} (\cdots) + \frac{1}{\sqrt{2\pi}} \sum_{i_1 : x_{i_1} \geq 0} (\cdots) \right| \tag{76}$$

Using the same arguments in Lemma 62 and the fact that $\|A^{i_1} e^{-X^2/2}\|_2 \leq \|He^{-X^2/2}\|_2, \|A^{i_1}\| \leq \|H\|$, we can upper bound the first summation in Eq. (76) by

$$\frac{1}{\sqrt{2\pi}} \sum_{i_1 : x_{i_1} < 0} G\left(x_{-i_1}\right) \|He^{-X^2/2}\|_2^2 \cdot \|H\| \leq O\left(\Delta^2 \cdot \log^{2.5} k \|H\|^3\right),$$

where the inequality follows from the argument in Eq. (64).

For the second summation, define

$$B_{i_1,i_2} = \begin{cases} \frac{H_{i_1,i_2}}{\sqrt{g(x_{i_1})}}, & \text{if } x_{i_1} \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\|Be^{-X^2/2}\|_2 \leq \sqrt{2}\|He^{-X^2/2}\|_2$ and $\|B\| \leq \sqrt{2}\|H\|$ (since $g(x) \geq 1/2$ for $x \geq 0$). Then the second summation in Eq. (76) is equal to

$$\left| G(x) \frac{1}{\sqrt{2\pi}} \mathrm{Tr}\ D^2 \left(e^{-X^2/2}\right) \left[B, B^T\right] A \right| \leq \frac{1}{\sqrt{\pi}} G(x) \|He^{-X^2/2}\|_2 \|H\| \leq O\left(\Delta^2 \cdot \log^{1.5} k \cdot \|H\|^3\right)$$

where the inequality follows from the argument in Eq. (64). ∎

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il