

# A MOMENT MAJORIZATION PRINCIPLE FOR RANDOM MATRIX ENSEMBLES

STEVEN HEILMAN

ABSTRACT. We prove a moment majorization principle for matrix-valued functions with domain  $\{-1, 1\}^m$ ,  $m \in \mathbb{N}$ . The principle is an inequality between higher-order moments of a non-commutative multilinear polynomial with different random matrix ensemble inputs, where each variable has small influence and the variables are instantiated independently.

This technical result can be interpreted as a noncommutative generalization of one of the two inequalities of the seminal invariance principle of Mossel, O’Donnell and Oleszkiewicz. Applications to noncommutative noise stability and noncommutative anticoncentration are given.

## 1. INTRODUCTION

**1.1. A noncommutative moment majorization theorem.** We study matrix-valued functions  $f$  with domain  $\{-1, 1\}^m$  within the context of probability theory and Fourier analysis. More specifically, we study functions  $f$  such that, for every  $\sigma \in \{-1, 1\}^m$ , the operator norm of  $f(\sigma)$  is at most 1. The special case when  $f$  is valued in the two-point space  $\{-1, 1\}$  has been studied extensively within theoretical computer science [KKL88], but also in diverse areas such as combinatorics, isoperimetry [Tal94], or social choice theory [Kal02, MOO10, MN15]. (For a more comprehensive list of references and discussion, see e.g. the survey [O’D14b].) In applications to theoretical computer science, a function  $f: \{-1, 1\}^m \rightarrow \{-1, 1\}$  can be used to represent an instance of a combinatorial optimization problem. That is, the function  $f$  can be thought of as a list of elements of  $\{-1, 1\}$ , seen as a Boolean assignment to the  $2^m$  variables of some constraint satisfaction problem. Functions with domain  $\{-1, 1\}^m$  and range the simplex  $\{(x_1, \dots, x_n) \in \mathbb{R}^n: \sum_{i=1}^n x_i = 1, x_1 \geq 0, \dots, x_n \geq 0\}$  have also been considered [KN09, KN13, IM12]. Projecting  $f$  onto each coordinate gives a family of functions with range  $[0, 1]$ , so that similar tools to the Boolean case can be applied.

In many of the applications listed above a standard manipulation is to extend  $f$  to a multilinear polynomial, so that the distribution of  $f$  can be studied under different distributions on its domain, such as the standard Gaussian distribution. In our setting it is natural (and, as we will see, for our purposes necessary) to investigate the behavior of matrix-valued functions under distributions on their domain that allow the possibility for matrix variables. For any set  $S$ , let  $M_n(S)$  denotes the  $n \times n$  matrices with entries in  $S$ . Any  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$  can be extended to a multilinear polynomial in  $m$  *noncommutative* variables with matrix coefficients. Consider for instance the case  $m = 2$  and the polynomial  $f(\sigma_1, \sigma_2) = \sigma_1 \sigma_2$ , where  $\sigma_1, \sigma_2 \in \{-1, 1\}$ . Since the variables  $\sigma_1, \sigma_2$  commute, it is not necessary to specify the order in which the product of the variables is taken in  $f$ . However, once  $f$  is extended

---

2010 *Mathematics Subject Classification.* 68Q17,60E15,47A50.

*Key words and phrases.* invariance principle, moment majorization, Lindeberg replacement.

to matrix variables  $X_1, X_2$ , an ordering needs to be specified. We adopt the convention of ordering matrix variables by increasing order, e.g.  $f(X_1, X_2) = X_1 X_2$ .

Let  $d, m, n$  be positive integers. For us, a noncommutative multilinear polynomial of degree  $d$  in  $m$  variables can be expressed as

$$Q(X_1, \dots, X_m) = \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \widehat{Q}(S) \prod_{i \in S} X_i,$$

where  $\widehat{Q}(S)$  is an  $n \times n$  complex matrix for every  $S \subseteq \{1, \dots, m\}$ ,  $X_1, \dots, X_m$  are noncommutative  $n \times n$  matrix variables, and the product  $\prod_{i \in S} X_i$  is always taken in increasing order. For example,  $\prod_{i \in \{1, 2\}} X_i = X_1 X_2$ . Noncommutative polynomials appear in many other contexts, most notably, within free probability [Voi91, Theorem 3.3]. In addition there is a general theory of so-called nc-functions [KVV14], but this theory does not seem to apply to the noncommutative polynomials we consider here. (An nc function  $h$  is a function defined on matrices of any dimension, such that, for any  $n \geq 1$ , and for any  $n \times n$  matrices  $A, B, C$  such that  $C$  is invertible,  $h(CAC^{-1}) = Ch(A)C^{-1}$  and  $h(A \oplus B) = h(A) \oplus h(B)$ . Neither property is satisfied by a general matrix-valued non-commutative polynomial as defined below.)

Our main goal consists in bounding the moments of  $Q$  for different random matrix distributions in the domain, when all partial derivatives of  $Q$  are small (i.e. when  $Q$  has small influences). In particular, we would like to say that the moments of polynomials  $Q$  with small influences under Gaussian random matrix inputs are close to the moments of  $Q$  under uniform  $\{-1, 1\}^m$  inputs. Unfortunately, this task is in general impossible. For example, consider the linear polynomial  $Q(X_1, \dots, X_m) = (X_1 + \dots + X_m)/\sqrt{m}$ . For any square matrix  $A$ , let  $|A| = (AA^*)^{1/2}$ . Let  $b_1, \dots, b_m$  be i.i.d. uniform random variables in  $\{-1, 1\}$ , and let  $I$  denote the  $n \times n$  identity matrix. Then  $\mathbb{E}(1/n)\text{Tr} |Q(b_1 I, \dots, b_m I)|^4 = 3 - 2/m$ . On the other hand, let  $G_1, \dots, G_m$  be  $n \times n$  independent Wigner matrices with real Gaussian entries. In this case  $Q(G_1, \dots, G_m)$  is equal in distribution to  $G_1$ , and in particular  $\lim_{n \rightarrow \infty} (1/n)\text{Tr} |Q(G_1, \dots, G_m)|^4 = 2$ , by the semicircle law. Thus even though the first and second moments of the input distributions match, i.e.  $\mathbb{E}b_1 = 0$ ,  $\mathbb{E}b_1^2 = 1$ ,  $\mathbb{E}G_1 = 0$  and  $\mathbb{E}G_1 G_1^* = I$ , the associated moments of  $Q$  can be very different.

In summary, a general invariance principle cannot hold in this noncommutative setting. We could instead try to prove a weaker statement such as: the moments of  $Q$  under noncommutative inputs with  $\mathbb{E}G_1 = 0$  and  $\mathbb{E}G_1 G_1^* = I$  are bounded by the moments of  $Q$  under Boolean inputs. We call such a statement a *moment majorization theorem*. Unfortunately, this is also not true in general, as we now show. For any  $1 \leq i \leq n$ , let  $A_i$  be the  $n \times n$  matrix with a 1 in the first row and  $i^{\text{th}}$  column, and zeros in all other entries. Then for any  $1 \leq j, k \leq n$  with  $j \neq k$ , we have  $A_j A_k^* = 0$ . Consider the linear polynomial

$$Q(X_1, \dots, X_n) = \sum_{i=0}^{n-1} B^i A X_i,$$

which is such that  $\mathbb{E}\text{Tr} |Q(b_1, \dots, b_n)|^4 = n$ . Now let  $H_1$  be a uniformly random Haar distributed  $n \times n$  unitary matrix, and let  $H_2, \dots, H_n$  be independent copies of  $H_1$ . Then

$$\mathbb{E}\text{Tr} |Q(H_1, \dots, H_n)|^4 = n\text{Tr} |A_1|^4 + n(n-1)\mathbb{E}\text{Tr} (|A_1|^2 H_1 |A_1|^2 H_1^*) = n + n(n-1)/n = 2n - 1.$$

Since a general noncommutative majorization principle cannot hold we instead establish a limited moment majorization theorem, which will nevertheless be sufficient for our applications. We make two changes. We first increase the dimension of the random matrix inputs  $G_1, \dots, G_m$ . That is, we allow the variables of  $Q$  to take values in the set of  $p \times p$  matrices with  $p > n$  by defining the  $p \times p$  matrix

$$\iota(A) = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix}$$

for any  $n \times n$  matrix  $A$ , and

$$Q^\iota(X_1, \dots, X_m) = \sum_{S \subseteq \{1, \dots, m\}} \iota(\widehat{Q}(S)) \prod_{i \in S} X_i,$$

where  $X_1, \dots, X_m$  are noncommutative  $p \times p$  matrix variables. Second, we randomly rotate  $G_1, \dots, G_m$  by  $p \times p$  Haar-distributed random unitary matrices  $H_1, \dots, H_m$ .

We state one particular variant of our noncommutative moment majorization theorem. Let  $n$  be a positive integer and let  $p$  be a positive multiple of  $n$ . We write  $H \sim \mathcal{H}$  to denote a  $p \times p$  Haar-distributed random unitary matrix,  $b \sim \mathcal{B}$  for a uniformly random  $b \in \{-1, 1\}$ , and  $G \sim \mathcal{G}$  for any random variable taking values in  $M_n(\mathbb{C})$  such that  $\mathbb{E}G = 0$  and  $\mathbb{E}GG^* = I$ . We also write  $G' \sim \mathcal{G} \otimes I$  to denote  $G' = G \otimes I$  with  $G \sim \mathcal{G}$ , so that  $G'$  is a  $p \times p$  matrix. We use the succinct notation  $G_i \sim \mathcal{G}$  to denote a collection  $G_1, \dots, G_m$  of independent random matrices with distribution  $\mathcal{G}$ , and denote  $Q(G_1, \dots, G_m)$  as  $Q\{G_i\}$ . The operator norm of a matrix  $A$  is denoted  $\|A\|$ .

**Theorem 1.1 (Noncommutative Fourth Moment Majorization).** *Let  $Q$  be a noncommutative multilinear polynomial of degree  $d$  in  $m$  variables such that  $\|Q(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$ . Let  $\tau := \max_{i=1, \dots, m} \sum_{S \subseteq \{1, \dots, m\}: i \in S} \text{Tr}(\widehat{Q}(S)\widehat{Q}(S)^*)$  be the maximum influence of  $Q$ . Let  $G_i \sim \mathcal{G}$  and  $c_2 \geq 1$  such that  $\|\mathbb{E}(G_1 G_1^*)^2\| \leq c_2$ . Then*

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^\iota\{G_i H_i\}|^4 \leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} |Q\{b_i\}|^4 + 8(8c_2)^{4d} n^4 \tau^{1/4} + O_{m,n}(p^{-1/2}). \quad (1)$$

**Remark 1.2.** This Theorem is a special case of Theorem 3.9 below. The term  $8(8c_2)^{4d} n^4 \tau^{1/4}$  on the right-hand side of (1) can be replaced by  $(8c_2)^{4d} \tau$  by additionally assuming that  $\mathbb{E}a_1 a_2 a_3 = 0$  for any  $a_1, a_2, a_3$  which are (possibly repeated) entries of  $G_1$ . We omit the proof of this strengthened statement, since the details are essentially identical to the proof of Theorem 3.5.

**Remark 1.3.** Note that, although  $Q^\iota$  takes values in the set of  $p \times p$  matrices, the trace is normalized by  $1/n$ , and  $Q^\iota$  still “acts like” an  $n \times n$  matrix. In particular the moments of  $Q^\iota$  do not become arbitrarily small in general; for instance it holds (see Lemma 2.5 below) that

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^\iota\{G_i H_i\}|^2 = \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} |Q\{b_i\}|^2.$$

Theorem 1.1 shows that the fourth moment of  $Q$  with appropriate random matrix inputs is bounded by the fourth moment of  $Q$  with Boolean inputs. We provide a variant of this majorization principle for higher order moments (Theorem 3.9) and for increasing test functions (Theorem 3.5). These majorization principles all have dependence on the degree of

the polynomial, but we also give degree-independent bounds for polynomials whose higher order coefficients decay at an exponential rate (Corollary 3.8).

Although majorization principles such as Theorem 1.1 involve the trace norm of a polynomial, we can obtain bounds on the operator norm of  $Q$  in the following way. Let  $t \in \mathbb{R}$  and consider the function  $t \mapsto (\max(0, |t| - 1))^2$ . This function can be applied to self-adjoint matrices via spectral calculus, and  $A \mapsto (\max(0, (AA^*)^{1/2} - 1))^2 = 0$  if the singular values of  $A$  are all bounded by 1. In particular, if  $\|Q(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$  then  $(\max(0, (Q(\sigma)Q(\sigma)^*)^{1/2} - 1))^2 = 0$  for all  $\sigma \in \{-1, 1\}^m$ . The following majorization principle gives control on the operator norm of  $Q$  when we substitute appropriate random matrices into the domain of  $Q$ .

**Theorem 1.4 (Noncommutative Operator Norm Majorization).** *Let  $Q$  be a noncommutative multilinear polynomial of degree  $d$ . Suppose  $\|Q(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$ . Let  $\tau := \max_{i=1, \dots, m} \sum_{S \subseteq \{1, \dots, m\}: i \in S} \text{Tr}(\widehat{Q}(S)\widehat{Q}(S)^*)$  be the maximum influence of  $Q$ . Let  $G_i \sim \mathcal{G}$  and  $c_2, c_3 \geq 1$  such that  $\|\mathbb{E}(G_1 G_1^*)^2\| \leq c_2$  and  $\|\mathbb{E}(G_1 G_1^*)^3\| \leq c_3$ . Then*

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr}(\max(0, |Q^t\{G_i H_i\}| - 1))^2 \leq (8c_2 c_3)^{9d} n^{1/2} \tau^{1/6} + O_{m,n}(\tau^{-1/3} p^{-1/2}). \quad (2)$$

This Theorem appears below in Theorem 3.7. We will use this theorem to argue that under the proper normalization condition a low-influence polynomial typically maps random Gaussian matrix inputs to matrices of norm not much larger than 1. The ensemble that will be of most interest for us is the following.

**Example 1.5.** Let  $N > 0$ . Let  $V_1, \dots, V_N$  be  $n \times n$  complex matrices such that  $\sum_{i=1}^N V_i V_i^* = 1$ . Let  $g_1, \dots, g_N$  be i.i.d. standard complex Gaussian random variables (so that  $\mathbb{E}g_1 = 0$  and  $\mathbb{E}|g_1|^2 = 1$ .) Define  $G_1 = \sum_{i=1}^N g_i V_i$ , and let  $G_2, \dots, G_m$  be independent copies of  $G_1$ . Then it follows from [HT99, Corollary 2.8] that the random matrices  $G_1, \dots, G_m$  satisfy the hypothesis of Theorem 1.1 with  $c_2 = 2$ . In the case that  $V_1, \dots, V_N$  are real matrices and  $g_1, \dots, g_N$  are i.i.d. standard real Gaussians, the hypothesis of Theorem 1.1 is satisfied with  $c_2 = 4 \cdot 2 = 8$ , as follows from the complex case and the inequality  $\|\mathbb{E}(\Re(G_1)\Re(G_1)^*)^2\| \leq \|\mathbb{E}(G_1 G_1^*)^2\| \leq 2$  when  $g_1, \dots, g_N$  are complex.

Here are some other examples of random matrix ensembles satisfying the hypothesis of Theorem 1.4.

**Example 1.6.** Example 1.5 specifically applies to Gaussian Wigner matrices as follows. Let  $U_1, \dots, U_{n^2}$  be  $n \times n$  matrices such that these matrices are the standard orthonormal basis of  $\mathbb{C}^{n^2}$ . Then  $G_1$  is a Wigner matrix and the hypothesis of Theorem 1.1 is satisfied with  $c_2 = 2$ . Similarly, let  $U_1, \dots, U_{n(n+1)/2}$  be  $n \times n$  matrices such that these matrices are the standard orthonormal basis of symmetric  $n \times n$  matrices. Then  $G_1$  is a Wigner matrix and the hypothesis of Theorem 1.1 is satisfied with  $c_2 = 2$  (see [DS01, Theorem II.11] or [Ver12, Theorem 5.32]).

**Example 1.7.** Let  $G_1, \dots, G_m$  be i.i.d.  $n \times n$  Haar-distributed random unitary matrices. Then the random matrices  $G_1, \dots, G_m$  satisfy the hypothesis of Theorem 1.1 with  $c_2 = 1$ .

We give a brief overview of the strategy of the proof of Theorem 1.4 and its generalization, Theorem 3.9. In order to prove the majorization principle, we first establish some basic facts

about Fourier analysis of matrix-valued functions  $f$  in Section 2. In particular, starting from a function  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$ , we extend  $f$  to a noncommutative multilinear polynomial  $Q = Q_f$  of  $m$  variables.

For these polynomials  $Q$ , we consider a few different inner products, norms, derivatives, Plancherel identities, and we also define the Ornstein-Uhlenbeck semigroup. We then prove a noncommutative hypercontractive inequality for such polynomials  $Q$ . This hypercontractive inequality is developed in Section 3 and proven in Theorem 3.1. It can also be considered a polynomial generalization of the matrix Khintchine inequality.

To prove the noncommutative majorization principle, we use the Lindeberg replacement method from [MOO10, Cha06], but in our (matrix-valued) polynomials, we are replacing one random variable with one random matrix, one at a time. The “second order” terms in this replacement have a noncommutative nature which introduces an error. Instead of trying to bound this error (which seems difficult or impossible in general), we choose particular random matrices such that the “second order” terms are small. This is accomplished by replacing an  $n \times n$  random matrix  $X$  with a much larger  $p \times p$  matrix  $\begin{pmatrix} X & 0 \\ 0 & 0 \end{pmatrix} H$ , where  $H$  is a uniformly random  $p \times p$  unitary matrix and  $p > n$ . When  $p \rightarrow \infty$ , the noncommutative “second order” errors in the Lindeberg replacement vanish.

**1.2. Applications.** The commutative invariance principle [Rot79, Cha06, MOO10] implies that if  $Q$  is a commutative multilinear polynomial with small derivatives (i.e. small influences), then the distribution of  $Q$  on i.i.d. uniform inputs in  $\{-1, 1\}$  is close to the distribution of  $Q$  on i.i.d. standard Gaussian random variables. A more general statement can be made for more general functions and distributions; for details see e.g. [MOO10]. An invariance principle can also be considered as a concentration inequality, generalizing the central limit theorem with error bounds (i.e. the Berry-Esséen Theorem). For other variants of invariance principles, see [Mos10, IM12].

The form of the invariance principle given in [MOO10] is proven by a combination of the Lindeberg replacement argument and the hypercontractive inequality [Bon70, Nel73, Gro75]. That is, one replaces one argument of  $Q$  at a time, adding up the resulting errors and controlling them via the hypercontractive inequality. One version of hypercontractivity says that a higher  $L_q$  norm of a polynomial is bounded by a lower  $L_p$  norm of that polynomial, where  $q > p$ , with a bound dependent on the degree of the polynomial  $Q$ . For example, if  $Q$  has degree  $d$ , then the  $L_4$  norm of  $Q$  is bounded by  $9^d$  times the  $L_2$  norm of  $Q$ .

The commutative invariance principle has seen many applications [O’D14b, O’D14a] in recent years. Here is a small sample of such applications and references: isoperimetric problems in Gaussian space and in the hypercube [MOO10, IM12], social choice theory, Unique Games hardness results [KKMO07, IM12], analysis of algorithms [BR15], random matrix theory [MP14], free probability [NPR10], optimization of noise sensitivity [Kan14]. The Lindeberg replacement argument itself has many applications, e.g. in proving the universality of eigenvalue statistics for Wigner matrices [TV11, Theorem 15].

We anticipate that our noncommutative majorization principle will find similar applications. Even though it is impossible to prove a noncommutative invariance principle in general, most applications of the commutative invariance principle only involve one direction of the inequality. That is, most applications of the invariance principle are really just applications of a majorization principle such as Theorem 1.1 or 1.4.

To demonstrate further applications of our majorization principle, we show in Section 4 that one of the two main parts of the proof of the Majority is Stablest Theorem from [MOO10] can be extended to the noncommutative setting. Then, in Section 5, we demonstrate a (probably sub-optimal) anti-concentration estimate for noncommutative multilinear polynomials. Both of these results proceed as in [MOO10] by replacing their invariance principle with our majorization principle.

Since majorization principles such as Theorems 1.1 and 1.4 show the closeness of one distribution to another, these statements could be fit into the “concentration of measure” paradigm. The paper [MS12] proves a concentration inequality for noncommutative polynomials, but these methods seem insufficient to prove a majorization principle. An invariance principle has been proven in the free probability setting [DN14], but the details of exactly what polynomials can be dealt with, and which distributions can be handled, seem incomparable to our majorization principle.

**Remark 1.8.** We remark that, although it may be tempting to try to prove Theorem 1.4 from the commutative invariance principle of [MOO10], there seems to be no straightforward way to accomplish this task. For example, we could interpret each entry in the output of a noncommutative polynomial  $Q$  as a commutative polynomial function of the inputs. But then in order to control  $\text{Tr} |Q|^4$ , we would need information on the joint distribution of the entries of  $Q$ , which is not provided by the invariance principle of [MOO10].

## 2. DEFINITIONS, BACKGROUND AND NOTATION

**2.1. Matrices.** For  $n \in \mathbb{N}$  we denote the set of  $n$  by  $n$  matrices by  $M_n(\mathbb{C})$ . We use  $M_{n \times n}(\mathbb{C})$  in place of  $M_{n^2}$  to denote  $n^2$  by  $n^2$  matrices when we wish to emphasize that a specific tensor decomposition of the space  $\mathbb{C}^{n^2} = \mathbb{C}^n \otimes \mathbb{C}^n$  on which the matrix acts has been fixed.

For  $A \in M_n(\mathbb{C})$ ,  $\|A\|$  is the operator norm of  $A$  (the largest singular value). We denote by  $A^*$  the conjugate-transpose. The absolute value is  $|A| = (AA^*)^{1/2}$ . We use  $I$  to denote the identity matrix.

Any real function  $f : \mathbb{R} \rightarrow \mathbb{R}$  can be applied to a Hermitian matrix  $A$  by diagonalizing  $A$  and applying  $f$  to the eigenvalues of  $A$ . Define  $\text{Chop} : \mathbb{R} \rightarrow \mathbb{R}$  as  $\text{Chop}(t) = t$  if  $|t| \leq 1$ ,  $\text{Chop}(t) = 1$  if  $t \geq 1$ , and  $\text{Chop}(t) = -1$  if  $t \leq -1$ .

**2.2. Asymptotic Notation.** Let  $a, b, c \in \mathbb{R}$ . We write  $a = O_c(b)$  if there exists a constant  $C(c) > 0$  such that  $|a| \leq C(c) |b|$ . We write  $a \leq O_c(b)$  if there exists a constant  $C(c) > 0$  such that  $a \leq C(c)b$ .

**2.3. Random variables and expectations.** The following notational conventions will be useful when working with functions of multiple variables. Let  $m, n \in \mathbb{N}$  and let  $S, T$  be arbitrary sets. Let  $f : S^m \rightarrow T$  and let  $X_1, \dots, X_n$  be independent random variables with the same distribution  $\mathcal{X}$  taking values in  $S$ . Then we will denote  $\mathbb{E}f(X_1, \dots, X_m)$  by  $\mathbb{E}_{X_j \sim \mathcal{X}} f\{X_j\}$ ; more generally the curly bracket notation  $f\{A_j\}$  will be used to denote  $f(A_1, \dots, A_m)$ .

We will use the following ensembles. Let  $n, N \in \mathbb{N}$ , and let  $p$  be a positive multiple of  $n$ .  $H \sim \mathcal{H}$  denotes a  $p \times p$  Haar-distributed random unitary matrix, where the dimension  $p$  will always be clear from context.  $b \sim \mathcal{B}$  denotes a uniformly random  $b \in \{\pm 1\}$ .  $G \sim \mathcal{G}$  denotes any random variable taking values in  $M_n(\mathbb{C})$  such that  $\mathbb{E}G = 0$  and  $\mathbb{E}GG^* = I$ . And  $G' \sim \mathcal{G} \otimes I$  denotes  $G' = G \otimes I_{p/n} \in M_p(\mathbb{C})$ , where  $G \sim \mathcal{G}$ .  $G \sim \mathcal{V}$  denotes a random variable

distributed as  $G = \sum_{i=1}^N g_i V_i$ , where  $g_1, \dots, g_N$  are i.i.d. standard complex Gaussian random variables and  $V_1, \dots, V_N$  are  $n \times n$  complex matrices satisfying  $\sum_{i=1}^N V_i V_i^* = \sum_{i=1}^N V_i^* V_i = I_n$ . And  $G' \sim \mathcal{V} \otimes I$  denotes  $G' = G \otimes I_{p/n} \in M_p(\mathbb{C})$ , where  $G \sim \mathcal{V}$ . Whenever  $\mathcal{V}$  is used the matrices  $V_1, \dots, V_N$  will be clear from context. We also sometimes write  $G_j \sim \mathcal{D}$  to mean  $G_1, \dots, G_m$  are independent random variables with distribution  $\mathcal{D}$ , where again  $m$  will always be clear from context.

**2.4. Fourier expansions.** Let  $n, m \in \mathbb{N}$  and  $f, h: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$ . We consider the inner product

$$\langle f, h \rangle := \mathbb{E}_{b_j \sim \mathcal{B}} \text{Tr}(f\{b_j\}(h\{b_j\})^*) = 2^{-m} \sum_{\sigma \in \{-1, 1\}^m} \text{Tr}(f(\sigma)h(\sigma)^*).$$

Given  $S \subseteq \{1, \dots, m\}$  and  $\sigma = (\sigma_1, \dots, \sigma_m) \in \{-1, 1\}^m$ , define

$$W_S(\sigma) := \prod_{i \in S} \sigma_i.$$

The set of functions  $\{W_S\}_{S \subseteq \{1, \dots, m\}}$  forms an orthonormal basis for the space of functions from  $\{-1, 1\}^m$  to  $M_n(\mathbb{C})$ , when it is viewed as a vector space over  $\mathbb{C}$  with respect to the inner product  $\langle \cdot, \cdot \rangle$ .

Let  $\hat{f}(S) := 2^{-m} \sum_{\sigma \in \{-1, 1\}^m} f(\sigma) W_S(\sigma)$  be the Fourier coefficient of  $f$  associated to  $S$ . Note that  $\hat{f}(S) \in M_n(\mathbb{C})$ . Then  $f = \sum_{S \subseteq \{1, \dots, m\}} \hat{f}(S) W_S$ , and

$$\begin{aligned} 2^{-m} \sum_{\sigma \in \{-1, 1\}^m} f(\sigma)(g(\sigma))^* &= 2^{-m} \sum_{\sigma \in \{-1, 1\}^m} \sum_{S, S' \subseteq \{1, \dots, m\}} \hat{f}(S)(\hat{g}(S'))^* W_S(\sigma) W_{S'}(\sigma) \\ &= \sum_{S, S' \subseteq \{1, \dots, m\}} \hat{f}(S)(\hat{g}(S'))^* \cdot 2^{-m} \sum_{\sigma \in \{-1, 1\}^m} W_S(\sigma) W_{S'}(\sigma) \\ &= \sum_{S \subseteq \{1, \dots, m\}} \hat{f}(S)(\hat{g}(S))^*. \end{aligned} \quad (3)$$

**2.5. Noncommutative polynomials.** Let  $n, m \in \mathbb{N}$  be integers. We consider noncommutative multilinear polynomials  $Q \in M_n(\mathbb{C})[X_1, \dots, X_m]$ , where  $X_1, \dots, X_m$  are noncommutative indeterminates. Monomials are always ordered by increasing order of the index, e.g.  $X_1 X_2$  and not  $X_2 X_1$ . Any such polynomial can be expanded as

$$Q(X_1, \dots, X_m) = \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \hat{Q}(S) \prod_{i \in S} X_i, \quad (4)$$

where  $\hat{Q}(S) \in M_n(\mathbb{C})$  for all  $S \subseteq \{1, \dots, m\}$  and  $0 \leq d \leq m$  is the degree of  $Q$ , defined as  $\max\{|S| : \hat{f}(S) \neq 0, S \subseteq \{1, \dots, m\}\}$ .

Let  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$ . Define the (non-commutative) *multilinear polynomial*  $Q_f \in M_n(\mathbb{C})[X_1, \dots, X_m]$  associated to  $f$  by

$$Q_f(X_1, \dots, X_m) := \sum_{S \subseteq \{1, \dots, m\}} \hat{f}(S) \prod_{i \in S} X_i. \quad (5)$$

**2.6. Partial Derivatives.** Let  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$ . Let  $i \in \{1, \dots, m\}$ . Define the  $i^{\text{th}}$  partial derivative of  $f$  by

$$\partial_i f(\sigma) := \sum_{S \subseteq \{1, \dots, m\}: i \in S} \widehat{f}(S) W_S(\sigma) = \frac{1}{2}(f(\sigma) - f(\sigma_1, \dots, -\sigma_i, \dots, \sigma_m)), \quad (6)$$

and the  $i^{\text{th}}$  influence of  $f$  by

$$\text{Inf}_i f := \sum_{S \subseteq \{1, \dots, m\}: i \in S} \text{Tr}(\widehat{f}(S) \widehat{f}(S)^*). \quad (7)$$

Note that by (3),  $\text{Inf}_i f = \langle \partial_i f, \partial_i f \rangle$  and  $\sum_{i=1}^m \text{Inf}_i f = \sum_{S \subseteq \{1, \dots, m\}} |S| \text{Tr}(\widehat{f}(S) (\widehat{f}(S))^*)$ .

**2.7. Ornstein-Uhlenbeck semigroup.** For  $f, h: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$  define their convolution  $f * h$  by

$$f * h(\sigma) := 2^{-m} \sum_{\omega \in \{-1, 1\}^m} f(\sigma \cdot \omega^{-1}) h(\omega) = \sum_{S \subseteq \{1, \dots, m\}} \widehat{f}(S) \widehat{h}(S) W_S(\sigma), \quad \forall \sigma \in \{-1, 1\}^m. \quad (8)$$

Here  $\sigma \cdot \omega$  denotes the componentwise product of  $\sigma$  and  $\omega$  and  $\omega^{-1}$  denotes the multiplicative inverse of  $\omega$ , so that  $\omega \cdot \omega^{-1} = (1, \dots, 1)$ .

Let  $0 < \rho < 1$ . For any  $\sigma = (\sigma_1, \dots, \sigma_m) \in \{-1, 1\}^m$ , let

$$R_\rho(\sigma) := \prod_{j=1}^m (1 + \rho \sigma_j) = \sum_{S \subseteq \{1, \dots, m\}} \rho^{|S|} W_S(\sigma). \quad (9)$$

Let  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$ . Define the Ornstein-Uhlenbeck semigroup  $T_\rho f$  by

$$T_\rho f(\sigma) := \sum_{S \subseteq \{1, \dots, m\}} \rho^{|S|} \widehat{f}(S) W_S(\sigma) = f * R_\rho(\sigma), \quad \forall \sigma \in \{-1, 1\}^m. \quad (10)$$

**2.8. Truncation of Fourier Coefficients (or Littlewood-Paley Projections).** Let  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$ , and  $Q_f \in M_n(\mathbb{C})[X_1, \dots, X_m]$  the multilinear polynomial associated to  $f$ . Let  $d \in \mathbb{N}$ . Let  $P_d$  denote projection onto the level- $d$  Fourier coefficients. That is,  $\forall \sigma \in \{-1, 1\}^m, \forall X_1, \dots, X_m \in M_n(\mathbb{C})$ ,

$$P_d f(\sigma) := \sum_{S \subseteq \{1, \dots, m\}: |S|=d} \widehat{f}(S) W_S(\sigma), \quad P_d Q_f(X_1, \dots, X_m) := \sum_{S \subseteq \{1, \dots, m\}: |S|=d} \widehat{f}(S) \prod_{i \in S} X_i.$$

Let  $P_{\leq d} := \sum_{i \leq d} P_i$  denote projection onto the Fourier coefficients of degree at most  $d$ . Denote  $P_{> d} f := f - P_{\leq d} f$ ,  $P_{> d} Q_f := Q_f - P_{\leq d} Q_f$ .

**2.9. Embeddings.**

**Definition 2.1 (Matrix embedding).** Let  $A \in M_n(\mathbb{C})$  and  $p \geq n$ . Define the embedding  $\iota_p: M_n(\mathbb{C}) \rightarrow M_p(\mathbb{C})$  by

$$\iota_p(A) := \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \in M_p(\mathbb{C}).$$

If  $B = \sum_{k=1}^N C_k \otimes D_k \in M_{n \times n}(\mathbb{C})$ , extend this definition to

$$\iota_p(B) := \sum_{k=1}^N \iota_p(C_k) \otimes \iota_p(D_k) \in M_{p \times p}(\mathbb{C}).$$



Lastly, if  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$  define  $\iota_p(f): \{-1, 1\}^m \rightarrow M_p(\mathbb{C})$  by

$$\iota_p(f)(\sigma) = \iota_p(f(\sigma)) \quad \forall \sigma \in \{-1, 1\}^m.$$

We will sometimes denote the same quantities by  $A^\iota$ ,  $B^\iota$  and  $f^\iota$  respectively, leaving the dependence of  $\iota$  on  $p$  and  $n$  implicit for clarity of notation.

Note that if  $Q = \sum_{S \subseteq \{1, \dots, m\}} \widehat{Q}(S) \prod_{i \in S} X_i \in M_n(\mathbb{C})[X_1, \dots, X_m]$  is a noncommutative polynomial the last item in Definition 2.1 is equivalent to defining

$$Q^\iota = \sum_{S \subseteq \{1, \dots, m\}} \iota_p(\widehat{Q}(S)) \prod_{i \in S} X_i \in M_p(\mathbb{C})[X_1, \dots, X_m].$$

If moreover  $Q_f$  is defined from  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$  as in (5) then

$$Q_f^\iota = \sum_{S \subseteq \{1, \dots, m\}} \iota_p(\widehat{f}(S)) \prod_{i \in S} X_i \in M_p(\mathbb{C})[X_1, \dots, X_m].$$

## 2.10. Coordinate projections.

**Definition 2.2.** Let  $U \in M_n(\mathbb{C}^N)$  with  $UU^* = U^*U = I$  and  $Q \in M_n(\mathbb{C})[X_1, \dots, X_m]$  a noncommutative polynomial. We denote the *Gaussian  $L_2$  norm* of  $Q$  associated to the ensemble  $\mathcal{G}$  by

$$\|Q\|_{L_2, \mathcal{G}} := \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \\ H_k \sim \mathcal{H}_p}} \text{Tr} |Q^\iota \{G_k H_k\}|^2 \right)^{1/2}.$$

Let  $\mathcal{N}(0, 1)$  denote the standard complex Gaussian distribution for a random variable. For integers  $m, N$  let  $\mathbb{E}_{g_{ij} \sim \mathcal{N}(0, 1)}$  denote expectation with respect to  $g_{1,1}, \dots, g_{m,N}$  i.i.d.  $\mathcal{N}(0, 1)$ .

**2.11. Bounds on random polynomials.** The key difference between the random matrices  $G_i \sim \mathcal{G}$  and  $(G_i \otimes I)H_i$  where  $H_i \sim \mathcal{H}_p$  is that the matrices  $(G_i \otimes I)H_i$  behave well with respect to matrix products. This property is exploited in Corollary 2.4 below.

**Lemma 2.3.** *Let  $p$  be a positive multiple of  $n$ , let  $A, B \in M_p(\mathbb{C})$ . Let  $I = I_n$  be the  $n \times n$  identity matrix. Then*

$$\mathbb{E}_{H \sim \mathcal{H}_p} \|I^\iota A H B I^\iota\|^2 \leq \frac{n^2}{p} \|A\|^2 \|B\|^2.$$

*Proof.* The nonzero eigenspace  $K$  of  $H B I^\iota B^* H^*$  is a uniformly distributed subspace of dimension at most  $n$  of  $\mathbb{C}^p$ . Given a unit vector  $x$ , the squared norm of the projection of  $x$  on  $K$  has expectation at most  $n/p$ . Applying this to the eigenvectors of  $A^* I^\iota A$ ,

$$\mathbb{E}_{H \sim \mathcal{H}_p} \text{Tr}(I^\iota A H B I^\iota B^* H^* A^*) \leq \frac{n}{p} \|B\|^2 \text{Tr}(A^* I^\iota A) \leq \frac{n^2}{p} \|B\|^2 \|A\|^2.$$

□

**Corollary 2.4.** *Let  $R, S \in M_n(\mathbb{C})[X_1, \dots, X_m]$  be multilinear polynomials not depending on the  $j$ -th variable such that  $\mathbb{E}_{b_j \sim \mathcal{B}} \frac{1}{n} \text{Tr} |R\{b_j\}|^2 \leq 1$  and  $\mathbb{E}_{b_j \sim \mathcal{B}} \frac{1}{n} \text{Tr} |S\{b_j\}|^2 \leq 1$ . Then*

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \left\| (S^\iota \{G_i H_i\}) G_k H_k (R^\iota \{G_i H_i\})^* \right\| = O_{n,m}(p^{-1/2}),$$

where the implicit constant may depend on  $n$  and  $m$ .

*Proof.* Write  $S^\iota = S_0 + \sum_{k \neq j} S_k X_k$  and  $R^\iota = R_0 + \sum_{k \neq j} R_k X_k$ , where  $\forall k \in \{1, \dots, m\} \setminus \{j\}$ ,  $S_k, R_k \in M_n(\mathbb{C})[X_1, \dots, X_m]$  depend neither on the  $k$ -th or the  $j$ -th variable, and  $S_0, R_0 \in M_n(\mathbb{C})$ . Then for any  $k \neq j$ , since  $S = I^\iota S$  and  $R^* = R^* I^\iota$ ,

$$\begin{aligned} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \|S^\iota G_k H_k H_k^* G_k^* (R_k^\iota)^*\| &= \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \|I^\iota S^\iota G_k H_k H_k^* G_k^* (R_k^\iota)^* I^\iota\| \\ &\leq \frac{n}{\sqrt{p}} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \|S^\iota G_j\| \|G_k^* (R_k^\iota)^*\| \\ &\leq \frac{n}{\sqrt{p}} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \|S^\iota G_j\|^2 \right)^{1/2} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \|G_k^* (R_k^\iota)^*\|^2 \right)^{1/2} \\ &\leq \frac{n^2}{\sqrt{p}} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \|S^\iota\|^2 \right)^{1/2} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \|R_k^\iota\|^2 \right)^{1/2} \end{aligned}$$

where the last inequality used  $\mathbb{E}_{G \sim \iota(\mathcal{G})} \|G\|^2 = \mathbb{E}_{G \sim \mathcal{G}} \|G\|^2 \leq \mathbb{E}_{G \sim \mathcal{G}} \text{Tr}(GG^*) = n$  and the first inequality used Lemma 2.3. To conclude, we use the normalization assumption on  $R, S$ , and (11) and (14) from Lemma 2.5 below.  $\square$

**Lemma 2.5.** *Let  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$  and  $Q \in M_n(\mathbb{C})[X_1, \dots, X_m]$  the multilinear polynomial associated to  $f$ . Let  $0 \leq k \leq m$ . Let  $G_i \sim \mathcal{G}$  and  $b_i \sim \mathcal{B}$ . Let  $\mathcal{X} = (G_1, \dots, G_k)$  and  $\mathcal{Y} = (b_{k+1}, \dots, b_m)$ . Then*

$$\mathbb{E}_{x_i \sim \mathcal{X}, y_j \sim \mathcal{Y}} |Q\{x_i, y_j\}|^2 = \mathbb{E}_{b_i \sim \mathcal{B}} |Q\{b_i\}|^2. \quad (11)$$

$$\text{Inf}_i(Q^\iota) = \text{Inf}_i(Q), \quad \forall i \in \{1, \dots, m\}. \quad (12)$$

$$\frac{1}{n} \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} |Q^\iota\{b_i\}|^2 = \frac{1}{n} \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} |Q\{b_i\}|^2. \quad (13)$$

$$\frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |Q^\iota\{G_i H_i\}|^2 = \frac{1}{n} \mathbb{E}_{G_i \sim \mathcal{G}} \text{Tr} |Q\{G_i\}|^2. \quad (14)$$

*Proof.* Recall the variables  $G_i \sim \mathcal{G}$  are independent,  $\mathbb{E}G_i = 0$  and  $\mathbb{E}G_i G_i^* = I$  for all  $1 \leq i \leq k$ . Similarly, the  $b_i \sim \mathcal{B}$  are independent with  $\mathbb{E}b_i = 0$  and  $\mathbb{E}b_i b_i^* = 1$  for all  $k+1 \leq i \leq m$ . So,

$$\begin{aligned} &\mathbb{E}_{x \sim \mathcal{X}, y \sim \mathcal{Y}} (Q(x, y)(Q(x, y))^*) \\ &= \mathbb{E} \sum_{S \subseteq \{1, \dots, m\}} \left( \widehat{Q}(S) \left( \prod_{i \in S: i \leq k} G_i \prod_{i \in S: i > k} b_i \right) \left( \prod_{i \in S: i \leq k} G_i \prod_{i \in S: i > k} b_i \right)^* (\widehat{Q}(S))^* \right) \\ &= \sum_{S \subseteq \{1, \dots, m\}} \widehat{Q}(S) (\widehat{Q}(S))^* \\ &\stackrel{(3)}{=} \mathbb{E}_{b_i \sim \mathcal{B}} |Q\{b_i\}|^2. \end{aligned}$$

Then (11) follows. Equation (12) follows from (7), Definition 2.1, and Lemma 2.5. Equalities (13) and (14) follow from Definition 2.1.  $\square$

**2.12. Unique Games Conjecture.** The Unique Games Conjecture is a commonly assumed conjecture in complexity theory, though its current status is unresolved.

**Definition 2.6 (Unique Games).** Let  $m \in \mathbb{N}$ . Let  $G = G(\mathcal{S}, \mathcal{W}, \mathcal{E})$  be a bipartite graph with vertex sets  $\mathcal{S}$  and  $\mathcal{W}$  and edge set  $\mathcal{E} \subseteq \mathcal{S} \times \mathcal{W}$ . For all  $(v, w) \in \mathcal{E}$ , let  $\pi_{vw}: \{1, \dots, m\} \rightarrow \{1, \dots, m\}$  be a permutation. An instance of the Unique Games problem is  $\mathcal{L} = (G(\mathcal{S}, \mathcal{W}, \mathcal{E}), m, \{\pi_{vw}\}_{(v,w) \in \mathcal{E}})$ .  $m$  is called the alphabet size of  $\mathcal{L}$ . A labeling of  $\mathcal{L}$  is a function  $\eta: \mathcal{S} \cup \mathcal{W} \rightarrow \{1, \dots, m\}$ . An edge  $(v, w) \in \mathcal{E}$  is satisfied if and only if  $\eta(v) = \pi_{vw}(\eta(w))$ . The goal of the Unique Games problem is to maximize the fraction of satisfied edges of the labeling, over all such labelings  $\eta$ . We call the maximum possible fraction of satisfied edges  $\text{OPT}(\mathcal{L})$ .

**Definition 2.7 (Unique Games Conjecture, [Kho02, KKMO07]).** For every  $0 < \beta < \alpha < 1$  there exists  $m = m(\alpha, \beta) \in \mathbb{N}$  and a family of Unique Games instances  $(\mathcal{L}_n)_{n \geq 1}$  with alphabet size  $m$  such that no polynomial time algorithm can distinguish between  $\text{OPT}(\mathcal{L}_n) < \beta$  or  $\text{OPT}(\mathcal{L}_n) > \alpha$ .

### 3. MAJORIZATION PRINCIPLE

**3.1. A noncommutative hypercontractive inequality.** One of the main tools used in the proof of our majorization principle is a hypercontractive inequality for noncommutative multilinear polynomials. The inequality bounds the  $2K$ -norm, for  $K \geq 1$  an integer, of a polynomial  $Q$  by the 2-norm of  $Q$ . We refer to this inequality as a  $(2K, 2)$  hypercontractive inequality; it can be considered as a polynomial generalization of the noncommutative Khintchine inequality between the  $2K$  norm and the 2 norm. (The Khintchine inequality corresponds to the polynomial  $Q(X_1, \dots, X_m) = X_1$  and  $G_1 \sim \mathcal{V}$ ; see [MJC<sup>+</sup>14, Corollary 7.3] or [DR13].)

Recall the definition of the ensembles  $\mathcal{G}$  and  $\mathcal{V}$  in Section 2.3.

**Theorem 3.1 (( $2K, 2$ ) Hypercontractivity).** *Let  $K \in \mathbb{N}$ . Let  $Q$  be a noncommutative multilinear polynomial of degree  $d \in \mathbb{N}$ , as in (4). Let  $G_j \sim \mathcal{G}$ , where  $\mathcal{G}$  is such that  $\|\mathbb{E}(G_1 G_1^*)^K\| \leq c_K$  for some  $c_K \geq 1$ . Then*

$$\mathbb{E}_{G_j \sim \mathcal{V}} \text{Tr} |Q\{G_j\}|^{2K} \leq (2K - 1)^{dK} c_K^d \left( \mathbb{E}_{G_j \sim \mathcal{V}} \text{Tr} |Q\{G_j\}|^2 \right)^K.$$

Or, more generally,

$$\mathbb{E}_{G_j \sim \mathcal{G}} \text{Tr} |Q\{G_j\}|^{2K} \leq (2K - 1)^{2dK} c_K^d \left( \mathbb{E}_{G_j \sim \mathcal{G}} \text{Tr} |Q\{G_j\}|^2 \right)^K.$$

**Remark 3.2.** As mentioned in [MOO10, Theorem 3.13] or [Nel73, Theorem 4], the best possible constant in this hypercontractive inequality is  $(2K - 1)^{dK}$  in the case that  $G_j \sim \mathcal{V}$  are replaced with  $b_i \sim \mathcal{B}$ . So, we achieve the optimal constant in this case, since we can use  $c_K = 1$  for all  $K \in \mathbb{N}$  in this case.

The result that hypercontractivity also holds for the variables  $G_1 = b_1 \sim \mathcal{B}$  generalizes a result of [Gro72, Lemma 6.1].

*Proof.* We begin with the first inequality. It suffices to prove the following hypercontractive estimate: if  $\rho \geq 0$  satisfies  $\rho \leq (2K - 1)^{-1/2} c_K^{-1/(2K)}$ , then

$$\mathbb{E}_{G_j \sim \mathcal{V}} \text{Tr} |T_\rho Q\{G_j\}|^{2K} \leq \left( \mathbb{E}_{G_j \sim \mathcal{V}} \text{Tr} |Q\{G_j\}|^2 \right)^K. \quad (15)$$

To see that (15) implies the first inequality of the theorem, choose  $\rho = (2K - 1)^{-1/2} c_K^{-1/(2K)}$ , and observe that

$$\begin{aligned} \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |Q\{G_j\}|^{2K} &\stackrel{(10)}{=} \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} \left| T_\rho \left( \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \rho^{-|S|} \widehat{Q}(S) \prod_{i \in S} G_i \right) \right|^{2K} \\ &\stackrel{(15)}{\leq} \left( \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} \left| \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \rho^{-|S|} \widehat{Q}(S) \prod_{i \in S} G_i \right|^2 \right)^K \stackrel{(11) \wedge (3)}{=} \left( \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \rho^{-2|S|} \operatorname{Tr} |\widehat{Q}(S)|^2 \right)^K \end{aligned} \quad (16)$$

$$\leq \rho^{-2dK} \left( \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \operatorname{Tr} |\widehat{Q}(S)|^2 \right)^K \stackrel{(11) \wedge (3)}{=} (2K - 1)^{dK} c_K^d \left( \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |Q\{G_j\}|^2 \right)^K. \quad (17)$$

The proof of (15) is by induction on the number  $m$  of variables of  $Q$ . If  $m = 0$  then there are no variables and the inequality follows from the elementary inequality  $\sum_{i=1}^n (\lambda_i(Q))^{2K} \leq (\sum_{i=1}^n (\lambda_i(Q))^2)^K$  applied to the singular values of the (deterministic) matrix  $Q$ . To establish the inductive step, write  $Q = R_0 + R_1 X_m$ , where  $R_0, R_1$  depend on at most  $m - 1$  variables each (for clarity we suppress this dependence from the notation). Note that  $T_\rho Q = T_\rho R_0 + \rho(T_\rho R_1) X_m$ . We begin with a binomial expansion

$$|T_\rho Q(X_1, \dots, X_m)|^{2K} = \sum_{(a_1, \dots, a_{2K}) \in \{0, 1\}^{2K}} \prod_{i=1}^K \rho^{a_{2i-1}} T_\rho R_{a_{2i-1}} X_m^{a_{2i-1}} (X_m^*)^{a_{2i}} \rho^{a_{2i}} (T_\rho R_{a_{2i}})^*. \quad (18)$$

Any term in the sum for which  $a_j = 0$  for an odd number of elements  $j \in \{1, \dots, 2K\}$  has expectation zero. Applying Hölder's inequality,

$$\mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |T_\rho Q|^{2K} \leq \mathbb{E} \operatorname{Tr} |T_\rho R_0|^{2K} + \sum_{\substack{(a_1, \dots, a_{2K}) \in \{0, 1\}^{2K}: \\ \frac{1}{2} \sum_{i=1}^{2K} a_i \text{ is a positive integer}}} \prod_{i=1}^{2K} \rho^{a_i} \left[ \mathbb{E} \operatorname{Tr} |(T_\rho R_{a_i}) G_m^{a_i}|^{2K} \right]^{\frac{1}{2K}}. \quad (19)$$

Let  $(a_1, \dots, a_{2K}) \in \{0, 1\}^{2K}$  such that  $\ell := \frac{1}{2} \sum_{i=1}^{2K} a_i$  is a positive integer. The number of times the term  $\prod_{i=1}^{2K} \rho^{a_i} \left[ \mathbb{E} \operatorname{Tr} |(T_\rho R_{a_i}) G_m^{a_i}|^{2K} \right]^{\frac{1}{2K}}$  is repeated in the sum in (19) is  $\binom{2K}{2\ell} = \frac{(2K)!}{(2K-2\ell)!(2\ell)!}$ . That is, (19) can be rewritten as

$$\mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |T_\rho Q|^{2K} \leq \mathbb{E} \operatorname{Tr} |T_\rho R_0|^{2K} + \sum_{\ell=1}^K \rho^{2\ell} \binom{2K}{2\ell} \left[ \mathbb{E} \operatorname{Tr} |(T_\rho R_1) G_m|^{2K} \right]^{\frac{\ell}{K}} \left[ \mathbb{E} \operatorname{Tr} |T_\rho R_0|^{2K} \right]^{\frac{K-\ell}{K}}. \quad (20)$$

For any  $A, B \in M_n(\mathbb{C})$  it holds that  $\operatorname{Tr} | |B| A^* A |B| |^K \leq \operatorname{Tr} | |B|^K (A^* A)^K |B|^K |$  (see e.g. [Bha97, Theorem IX.2.10]), hence

$$\operatorname{Tr} (A B B^* A^*)^K = \operatorname{Tr} (A^* A B B^*)^K \leq \operatorname{Tr} ((B B^*)^K (A^* A)^K). \quad (21)$$

Starting from (20) and applying (21) to each of the inner terms,

$$\begin{aligned} \mathbb{E}_{G_j \sim \nu} \text{Tr} |T_\rho Q|^{2K} &\leq \mathbb{E} \text{Tr} |T_\rho R_0|^{2K} + \sum_{\ell=1}^K \rho^{2\ell} \binom{2K}{2\ell} c_K^{\ell/K} \left[ \mathbb{E} \text{Tr} |T_\rho R_1|^{2K} \right]^{\frac{\ell}{K}} \left[ \mathbb{E} \text{Tr} |T_\rho R_0|^{2K} \right]^{\frac{K-\ell}{K}} \\ &\leq [\mathbb{E} \text{Tr} |R_0|^2]^K + \sum_{\ell=1}^K \rho^{2\ell} \binom{2K}{2\ell} c_K^{\ell/K} \left[ \mathbb{E} \text{Tr} |R_1|^2 \right]^\ell \left[ \mathbb{E} \text{Tr} |R_0|^2 \right]^{K-\ell}, \end{aligned} \quad (22)$$

where the second inequality is obtained by applying the inductive hypothesis. For any odd integer  $J$  we denote  $J!! = \prod_{i=0}^{(J-1)/2} (J-2i)$ , or  $J!! = 1$  if  $(J-1)/2 < 1$ . Now if  $1 \leq \ell \leq K$ ,

$$\begin{aligned} \binom{2K}{2\ell} \binom{K}{\ell}^{-1} &= \frac{(2K)!}{K!} \frac{\ell!}{(2\ell)!} \frac{(K-\ell)!}{(2(K-\ell))!} = \frac{2^K (2K-1)!!}{2^\ell (2\ell-1)!! 2^{K-\ell} (2(K-\ell)-1)!!} \\ &= \frac{(2K-1)!!}{(2\ell-1)!! (2(K-\ell)-1)!!} = \prod_{i=0}^{\ell-1} \frac{2K-2i-1}{2\ell-2i-1} \leq (2K-1)^\ell. \end{aligned}$$

Using this inequality and  $0 \leq \rho \leq (2K-1)^{-1/2} c_K^{-1/(2K)}$  we get

$$\rho^{2\ell} \binom{2K}{2\ell} c_K^{\ell/K} \leq \rho^{2\ell} (2K-1)^\ell c_K^{\ell/K} \binom{K}{\ell} \leq \binom{K}{\ell}, \quad \forall 1 \leq \ell \leq K.$$

Applying this inequality to (22),

$$\begin{aligned} \mathbb{E}_{G_j \sim \nu} \text{Tr} |T_\rho Q|^{2K} &\leq \sum_{\ell=0}^K \binom{K}{\ell} \left[ \mathbb{E}_{G_j \sim \nu} \text{Tr} |R_1|^2 \right]^\ell \left[ \mathbb{E}_{G_j \sim \nu} \text{Tr} |R_0|^2 \right]^{K-\ell} \\ &= \left( \sum_{i=0}^1 \mathbb{E}_{G_j \sim \nu} \text{Tr} (R_i R_i^*) \right)^K = \left[ \mathbb{E}_{G_j \sim \nu} \text{Tr} (Q Q^*) \right]^K, \end{aligned}$$

where the last equality follows from  $\mathbb{E} G_1 G_1^* = I$ .

We now prove the second inequality of the Theorem. We repeat the above proof, though now the sum in (19) ranges over all  $(a_1, \dots, a_{2K}) \in \{0, 1\}^{2K}$  such that  $\sum_{i=1}^{2K} a_i \geq 2$ . If  $\ell > 0$  satisfies  $2\ell + 1 = \sum_{i=1}^{2K} a_i$ , then the corresponding term appears  $\binom{2K}{2\ell+1}$  times in (19), and we bound this term in (19) with the arithmetic-mean geometric-mean inequality as follows.

$$\begin{aligned} \prod_{i=1}^{2K} \left[ \mathbb{E} \text{Tr} |(T_\rho R_{a_i}) G_m^{a_i}|^{2K} \right]^{\frac{1}{2K}} &= \left[ \mathbb{E} \text{Tr} |(T_\rho R_1) G_m|^{2K} \right]^{\frac{2\ell+1}{2K}} \left[ \mathbb{E} \text{Tr} |T_\rho R_0|^{2K} \right]^{\frac{2K-2\ell-1}{2K}} \\ &\leq \frac{1}{2} \left[ \mathbb{E} \text{Tr} |(T_\rho R_1) G_m|^{2K} \right]^{\frac{\ell}{K}} \left[ \mathbb{E} \text{Tr} |T_\rho R_0|^{2K} \right]^{\frac{K-\ell}{K}} \\ &\quad + \frac{1}{2} \left[ \mathbb{E} \text{Tr} |(T_\rho R_1) G_m|^{2K} \right]^{\frac{\ell+1}{K}} \left[ \mathbb{E} \text{Tr} |T_\rho R_0|^{2K} \right]^{\frac{K-\ell-1}{K}}. \end{aligned}$$

Then, as above, we have  $\frac{1}{2} \binom{2K}{2\ell+1} \binom{K}{\ell}^{-1} \leq \frac{K-1}{3} (2K-1)^\ell$  and  $\frac{1}{2} \binom{2K}{2\ell+1} \binom{K}{\ell+1}^{-1} \leq \frac{2}{3} (2K-1)^\ell$ . So, if  $0 \leq \rho \leq (2K-1) c_K^{-1/(2K)}$ , then  $\rho^{2\ell} \left( \binom{2K}{2\ell} + \frac{1}{2} \binom{2K}{2\ell+1} + \frac{1}{2} \binom{2K}{2\ell-1} \right) c_K^{\ell/K} \leq \binom{K}{\ell}$ , for all  $1 \leq \ell \leq K$ , and then the proof concludes as above.  $\square$

**Corollary 3.3** (( $2K, 2$ ) Hypercontractivity). *Let  $K \in \mathbb{N}$ . Let  $Q$  be a noncommutative multilinear polynomial of degree  $d \in \mathbb{N}$ , as in (4). Then*

$$\mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |Q\{G_j\}|^{2K} \leq (2K - 1)^{dK} c_K^d (K!)^d \left( \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |Q\{G_j\}|^2 \right)^K.$$

$$\mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \operatorname{Tr} |Q^\ell\{G_j H_j\}|^{2K} \leq (2K - 1)^{dK} c_K^d (K!)^d \left( \mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \operatorname{Tr} |Q^\ell\{G_j H_j\}|^2 \right)^K.$$

$$\mathbb{E}_{b_j \sim \mathcal{B}} \operatorname{Tr} |Q\{b_j\}|^{2K} \leq (2K - 1)^{dK} c_K^d \left( \mathbb{E}_{b_j \sim \mathcal{B}} \operatorname{Tr} |Q\{b_j\}|^2 \right)^K.$$

*Proof.* The first inequality follows from Theorem 3.1 using [HT99, Corollary 2.8] to show that for  $G \sim \mathcal{V}$  and any  $\ell \in \mathbb{N}$ ,  $\|\mathbb{E}(GG^*)^\ell\| \leq \ell!$ .

To prove the second inequality, we follow the proof of Theorem 3.1 using  $\|\mathbb{E}(GG^*)^\ell\| \leq \ell!$  for any  $G \sim \mathcal{V} \otimes I$ , where the  $Q$  used in the proof becomes  $Q^\ell$ . Writing  $Q^\ell = R_0^\ell + R_1^\ell X_m$ , there are only two required changes. First, the equalities (16) and (17) are justified by combining (14) with (11) and (3). For example, (16) is justified by

$$\begin{aligned} & \left( \mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \operatorname{Tr} \left| \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \rho^{-|S|} \widehat{Q}^\ell(S) \prod_{i \in S} G_i^\ell H_i \right|^2 \right)^K \\ & \stackrel{(14)}{=} \left( \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} \left| \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \rho^{-|S|} \widehat{Q}(S) \prod_{i \in S} G_i \right|^2 \right)^K \\ & \stackrel{(11) \wedge (3)}{=} \left( \sum_{S \subseteq \{1, \dots, m\}: |S| \leq d} \rho^{-2|S|} \operatorname{Tr} |\widehat{Q}(S)|^2 \right)^K. \end{aligned}$$

And (17) is justified in the same way. Similarly, the last inequality in the proof is justified as

$$\begin{aligned} \mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \operatorname{Tr} |T_\rho Q^\ell|^{2K} & \leq \sum_{\ell=0}^K \binom{K}{\ell} \left[ \mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \operatorname{Tr} |R_1^\ell|^2 \right]^\ell \left[ \mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \operatorname{Tr} |R_0^\ell|^2 \right]^{K-\ell} \\ & \stackrel{(14)}{=} \sum_{\ell=0}^K \binom{K}{\ell} \left[ \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |R_1|^{2\ell} \right]^\ell \left[ \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |R_0|^{2(K-\ell)} \right]^{K-\ell} \\ & = \left( \mathbb{E}_{G_j \sim \mathcal{V}} \operatorname{Tr} |Q|^2 \right)^K \\ & \stackrel{(14)}{=} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \operatorname{Tr} |Q^\ell|^2 \right)^K. \end{aligned}$$

The last inequality in the Corollary follows directly from Theorem 3.1.  $\square$

In summary,  $Q$  is hypercontractive when we substitute into  $Q$  the noncommutative random variables  $G_i H_i$ . Since  $Q$  is also hypercontractive when we substitute into  $Q$  commutative random variables distributed uniformly in  $\{-1, 1\}$ , we get the standard consequence that  $Q$  is hypercontractive when we substitute into it a mixture of commutative and noncommutative random variables.

**Corollary 3.4 ((2K, 2) Hypercontractivity for mixed inputs).** *Let  $G_j \sim \mathcal{G}$  be i.i.d. random  $n \times n$  matrices and let  $Q$  be a noncommutative multilinear polynomial of degree  $d \in \mathbb{N}$  such that  $Q$  satisfies (2K, 2) hypercontractivity for some  $K \in \mathbb{N}$ . That is, assume there exists  $c_K \geq 1$  such that*

$$\mathbb{E}_{G_j \sim \mathcal{G}} \text{Tr} |Q\{G_j\}|^{2K} \leq (2K - 1)^{2dK} c_K^d (\mathbb{E}_{G_j \sim \mathcal{G}} \text{Tr} |Q\{G_j\}|^2)^K.$$

Let  $\mathcal{X} = (G_1, \dots, G_k)$ . Let  $b_j \sim \mathcal{B}$  and let  $\mathcal{Y} = (b_{k+1}, \dots, b_m)$ . Then

$$\mathbb{E}_{x \sim \mathcal{X}, y \sim \mathcal{Y}} \text{Tr} |Q(x, y)|^{2K} \leq (2K - 1)^{2dK} c_K^d (\mathbb{E}_{x \sim \mathcal{X}, y \sim \mathcal{Y}} \text{Tr} |Q(x, y)|^2)^K.$$

*Proof.* For any  $p \geq 1$ , let  $\|\cdot\|_{p, \mathcal{Z}}$  denote the norm  $\|Q\|_{p, \mathcal{Z}} = (\mathbb{E}_{z \sim \mathcal{Z}} \text{Tr} |Q(z)|^p)^{1/p}$ .

$$\begin{aligned} \|Q\|_{2K, \mathcal{X} \cup \mathcal{Y}} &= \left\| \sum_{S \subseteq \{1, \dots, m\}} \widehat{Q}(S) \prod_{j \in S: j \leq k} G_j \prod_{j \in S: j > k} b_j \right\|_{2K, \mathcal{X} \cup \mathcal{Y}} \\ &= \left\| \left\| \sum_{S \subseteq \{1, \dots, m\}} \widehat{Q}(S) \prod_{j \in S: j \leq k} G_j \prod_{j \in S: j > k} b_j \right\|_{2K, \mathcal{X}} \right\|_{2K, \mathcal{Y}} \\ &\leq (2K - 1)^k c_K^{k/(2K)} \left\| \left\| \sum_{S \subseteq \{1, \dots, m\}} \left( \widehat{Q}(S) \prod_{j \in S: j > k} b_j \right) \prod_{j \in S: j \leq k} G_j \right\|_{2, \mathcal{X}} \right\|_{2K, \mathcal{Y}}, \end{aligned}$$

by Theorem 3.1. Next, using Minkowski's inequality, from the above we get

$$\begin{aligned} \|Q\|_{2K, \mathcal{X} \cup \mathcal{Y}} &\leq (2K - 1)^k c_K^{k/(2K)} \left\| \left\| \sum_{S \subseteq \{1, \dots, m\}} \left( \widehat{Q}(S) \prod_{j \in S: j > k} b_j \right) \prod_{j \in S: j \leq k} G_j \right\|_{2K, \mathcal{Y}} \right\|_{2, \mathcal{X}} \\ &= (2K - 1)^k c_K^{k/(2K)} \left\| \left\| \sum_{S \subseteq \{1, \dots, m\}} \left( \widehat{Q}(S) \prod_{j \in S: j \leq k} G_j \right) \prod_{j \in S: j > k} b_j \right\|_{2K, \mathcal{Y}} \right\|_{2, \mathcal{X}} \\ &\leq (2K - 1)^d c_K^{d/(2K)} \left\| \left\| \sum_{S \subseteq \{1, \dots, m\}} \left( \widehat{Q}(S) \prod_{j \in S: j \leq k} G_j \right) \prod_{j \in S: j > k} b_j \right\|_{2, \mathcal{Y}} \right\|_{2, \mathcal{X}} \\ &= (2K - 1)^d c_K^{d/(2K)} \left\| \sum_{S \subseteq \{1, \dots, m\}} \widehat{Q}(S) \prod_{j \in S: j \leq k} G_j \prod_{j \in S: j > k} b_j \right\|_{2, \mathcal{X} \cup \mathcal{Y}} \\ &= (2K - 1)^d c_K^{d/(2K)} \|Q\|_{2, \mathcal{X} \cup \mathcal{Y}}, \end{aligned}$$

where the third line is by Theorem 3.1. □

### 3.2. Majorization principle.

**Theorem 3.5 (Noncommutative Majorization Principle for Increasing Test Functions).** *Let  $Q \in M_n(\mathbb{C})[X_1, \dots, X_m]$  be a noncommutative multilinear polynomial of degree  $d$  such that  $\mathbb{E}_{b_j \sim \mathcal{B}} \frac{1}{n} \text{Tr}(Q\{b_j\}Q\{b_j\}^*) \leq 1$ . Let  $\tau := \max_{1 \leq j \leq m} \text{Inf}_j Q$ . Let  $G_i \sim \mathcal{V} \otimes I$ . Assume  $\|\mathbb{E}(G_1 G_1^*)^2\| \leq c_2$  and  $\|\mathbb{E}(G_1 G_1^*)^3\| \leq c_3$  with  $c_2, c_3 \geq 1$ .*

*Let  $\psi: [0, \infty) \rightarrow \mathbb{R}$  be a function with three continuous derivatives such that  $\psi'(t) \geq 0$  for all  $t \geq 0$ . Let  $a_2 = \sup_{t \geq 0} |\psi''(t)|$  and  $a_3 = \sup_{t \geq 0} |\psi'''(t)|$ . Then*

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |Q'\{G_i H_i\}|^2 \leq \mathbb{E}_{b_j \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |Q\{b_j\}|^2 + a_3 n^{3/2} (5^6 c_2 c_3)^d \tau^{1/2} + O_{n,m}(a_2 p^{-1/2}).$$

*Proof.* We show the bound using the Lindeberg replacement method, replacing the  $m$  variables  $G_k H_k$  by  $b_j$  one at a time, starting from the last, for each  $j \in \{1, \dots, m\}$ . Suppose variables  $j+1, j+2, \dots, m$  have already been replaced, and write  $Q^t = R + SX_j$  where  $R, S$  do not depend on the  $j^{\text{th}}$  variable.

Any three times continuously differentiable  $F: [0, \infty) \rightarrow \mathbb{R}$  has a Taylor expansion

$$F(1) = F(0) + F'(0) + \frac{1}{2}F''(0) + \frac{1}{2} \int_0^1 (1-s)^2 F'''(s) ds. \quad (23)$$

Let  $F(t) = \text{Tr}\psi((R + tSX)(R + tSX)^*)$  for  $t \in [0, 1]$ . Then

$$F(0) = \text{Tr}\psi(RR^*), \quad (24)$$

$$F'(0) = \text{Tr}(\psi'(RR^*)(SXR^* + RX^*S^*)), \quad (25)$$

$$F''(0) = \text{Tr}(\psi''(RR^*)2SXX^*S^* + \psi''(RR^*)(SXR^* + RX^*S^*)^2), \quad (26)$$

$$\begin{aligned} F'''(t) &= \text{Tr}(\psi'''((R + tSX)(R + tSX)^*)((SXR^* + RX^*S^*) + 2tSXX^*S^*)(SXX^*S^*)) \\ &\quad + \text{Tr}(\psi''((R + tSX)(R + tSX)^*)2((SXR^* + RX^*S^*) + 2tSXX^*S^*)(SXX^*S^*)) \\ &\quad + \text{Tr}(\psi'''((R + tSX)(R + tSX)^*)2((SXR^* + RX^*S^*) + 2tSXX^*S^*)^3). \end{aligned} \quad (27)$$

For any  $t \in [0, 1]$ , let  $F_1(t) = \text{Tr}\psi((R + tSb_j)(R + tSb_j)^*)$  and  $F_2(t) = \text{Tr}\psi((R + tSG_jH_j)(R + tSG_jH_j)^*)$ . From (23),

$$\mathbb{E}F_2(1) - \mathbb{E}F_1(1) = \mathbb{E}F_2''(0) - \mathbb{E}F_1''(0) + \mathbb{E}\frac{1}{2} \int_0^1 (1-s)^2 F_2'''(s) ds - \mathbb{E}\frac{1}{2} \int_0^1 (1-s)^2 F_1'''(s) ds, \quad (28)$$

where we used that  $\mathbb{E}F_2(0) = \mathbb{E}F_1(0)$  and  $\mathbb{E}F_2'(0) = \mathbb{E}F_1'(0)$ . We bound the two differences on the right-hand side of (28) separately.

For the first, using that  $\psi'(RR^*)$  is positive semidefinite the first term can be bounded as

$$\begin{aligned} \mathbb{E}\text{Tr}(\psi'(RR^*)2SG_jH_jH_j^*G_jS^*) &= \mathbb{E}\text{Tr}(\psi'(RR^*)2S \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} S^*) \\ &\leq \mathbb{E}\text{Tr}(\psi'(RR^*)2SS^*) \\ &= \mathbb{E}\text{Tr}(\psi'(RR^*)2Sb_jb_j^*S^*). \end{aligned} \quad (29)$$

The second term in (26) is readily bounded using Corollary 2.4, from which it follows that  $\mathbb{E}\|SG_kH_kR^* + RH_j^*G_j^*S^*\| = O_{n,m}(p^{-1/2})$ , and  $|\psi''(t)| \leq a_2$  for all  $t \geq 0$ . Combining the two bounds,

$$\mathbb{E}F_2''(0) - \mathbb{E}F_1''(0) \leq O_{n,m}(a_2 p^{-1/2}).$$

For the second difference on the right-hand side of (28), there are two terms, corresponding to the first two lines of (27) and the third line respectively. For the first two terms we apply the Cauchy-Schwarz inequality, isolating the last factor  $SXX^*S^*$  and using  $|\psi''(t)| \leq a_2$  for



all  $t \geq 0$  to bound them by

$$\begin{aligned}
& 3a_2 \left( \mathbb{E} \text{Tr}((SXX^*S^*)^2) \right)^{1/2} \left( \mathbb{E} \text{Tr}((SXR^* + RX^*S^* + 2tSXX^*S^*)^2) \right)^{1/2} \\
& \leq 3a_2(3^4c_2)^d \left( \mathbb{E} \text{Tr}(SS^*) \right) \left( 6\mathbb{E} \text{Tr}(SXR^*RX^*S^*) + 12t^2\mathbb{E} \text{Tr}((SXX^*S^*)^2) \right)^{1/2} \\
& \leq 3a_2(3^4c_2)^d \left( \mathbb{E} \text{Tr}(SS^*) \right) \left( \mathbb{E} \text{Tr}((SXX^*S^*)^2)(36\mathbb{E} \text{Tr}((RR^*)^2 + 12t^2\mathbb{E} \text{Tr}((SXX^*S^*)^2))) \right)^{1/4} \\
& \leq 36a_2(3^4c_2)^d n^{1/2} \left( \mathbb{E} \text{Tr}(SS^*) \right)^{3/2},
\end{aligned}$$

where for the first inequality, the first term is bounded using Corollary 3.4 (first using Corollary 3.3 to show that hypercontractivity holds for  $Q^t$  and then applying Corollary 3.4 with  $Q = Q^t$ ) and  $\mathbb{E}XX^* \leq I$ , and the second term is bounded using  $(A + B + C)(A + B + C)^* \leq 4(AA^* + BB^* + CC^*)$ , the second inequality uses Cauchy-Schwarz, and the last again Corollary 3.4,  $\mathbb{E}XX^* \leq I$ , and  $\mathbb{E}RR^* \leq I$ .

Finally we turn to the second term which appears in the expansion of the second difference on the right-hand side of (28) according to (27), corresponding to the third line of (27). Letting  $P = SXR^* + RX^*S^* + 2tSXX^*S^*$ , the term can be bounded using Hölder's inequality by

$$\begin{aligned}
a_3 \mathbb{E} \text{Tr}|P|^3 &= O(a_3) \left( \mathbb{E} \text{Tr}|RXS^*|^3 + \mathbb{E} \text{Tr}|SXX^*S^*|^3 \right) \\
&= O(a_3) \left( \mathbb{E} \text{Tr}|SX|^6 \right)^{1/2} \left( \left( \mathbb{E} \text{Tr}|R|^6 \right)^{1/2} + \left( \mathbb{E} \text{Tr}|SX|^6 \right)^{1/2} \right) \\
&= O(a_3)n^{3/2}(5^6c_3)^d \left( \mathbb{E} \text{Tr}|S|^2 \right)^{3/2},
\end{aligned}$$

where the last line uses Corollary 3.4 (applied as above) and  $\mathbb{E}SS^* \leq I$ ,  $\mathbb{E}RR^* \leq I$ .

Combining all error estimates and using  $\mathbb{E} \text{Tr}SS^* = \text{Inf}_j(Q)$  we obtain

$$\begin{aligned}
& \mathbb{E} \frac{1}{n} F_2(1) \\
& \leq \mathbb{E} \frac{1}{n} F_1(1) + O_{n,m}(p^{-1/2}) + \mathbb{E} \frac{1}{2n} \int_0^1 (1-s)^2 |F_2'''(s)| ds + \mathbb{E} \frac{1}{2n} \int_0^1 (1-s)^2 |F_1'''(s)| ds \\
& \leq \mathbb{E} \frac{1}{n} F_1(1) + O_{n,m}(a_2p^{-1/2}) + a_3n^{1/2}(5^6c_3)^d (\text{Inf}_j(Q))^{3/2}. \tag{30}
\end{aligned}$$

Iterating over all  $m$  variables,

$$\begin{aligned}
& \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |Q^t \{G_i H_i\}|^2 - \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |Q \{b_i\}|^2 \\
& \leq a_3n^{1/2}(5^6c_2c_3)^d \left( \max_{1 \leq j \leq m} \text{Inf}_j(Q) \right)^{1/2} \left( \sum_{j=1}^m \text{Inf}_j(Q) \right) + O_{n,m}(p^{-1/2}) \\
& \leq a_3n^{3/2}(5^6c_2c_3)^d \left( \max_{1 \leq j \leq m} \text{Inf}_j(Q) \right)^{1/2} + O_{n,m}(a_2p^{-1/2}).
\end{aligned}$$

□

Let  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  be Lipschitz, so that  $\sup_{x \neq y \in \mathbb{R}} \frac{|\psi(x) - \psi(y)|}{|x - y|} \leq 1$ . Let  $x \in \mathbb{R}$  and let  $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ ,  $\phi: \mathbb{R} \rightarrow \mathbb{R}$ . For any  $\lambda > 0$ , define  $\phi_\lambda(x) = \lambda^{-1}\phi(x/\lambda)$ . Define

$$\psi_\lambda(x) = \psi * \phi_\lambda(x) = \int_{\mathbb{R}} \psi(t)\phi_\lambda(x - t)dt. \quad (31)$$

Then  $|\psi(x) - \psi_\lambda(x)| < \lambda$  for all  $x \in \mathbb{R}$ , and  $|\frac{d^k}{dx^k}\phi_\lambda(x)| \leq \lambda^{1-k}$  for all  $x \in \mathbb{R}$ , so that

$$\left| \frac{d^k}{dx^k}\psi_\lambda(x) \right| \leq 3\lambda^{1-k}, \quad \forall x \in \mathbb{R}, \quad \forall 1 \leq k \leq 3.$$

**Lemma 3.6.** *Let  $\lambda > 0$ . If  $\psi$  is convex, then  $\psi_\lambda$  is convex, and  $\psi(x) \leq \psi_\lambda(x)$  for all  $x \in \mathbb{R}$ .*

*Proof.* The first property is a standard differentiation argument for convolutions. Since  $\psi(x+h) + \psi(x-h) - 2\psi(x) \geq 0$  for all  $x, h \in \mathbb{R}$ , we also have  $\psi_\lambda(x+h) + \psi_\lambda(x-h) - 2\psi_\lambda(x) \geq 0$ . The second property follows from Jensen's inequality.  $\square$

Let  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  be defined by

$$\psi(t) = \max(0, |t| - 1), \quad \forall t \in \mathbb{R}. \quad (32)$$

**Theorem 3.7.** *Let  $Q \in M_n(\mathbb{C})[X_1, \dots, X_m]$  be a noncommutative multilinear polynomial of degree  $d$  with  $\mathbb{E}_{b_j \sim \mathcal{B}} \frac{1}{n} \text{Tr}(|Q\{b_j\}|^2) \leq 1$ , and let  $\tau = \max_{i=1, \dots, m} \text{Inf}_i(Q)$ . Let  $G_i \sim \mathcal{G} \otimes I$ . Assume  $\|\mathbb{E}(G_1 G_1^*)^2\| \leq c_2$  and  $\|\mathbb{E}(G_1 G_1^*)^3\| \leq c_3$  with  $c_2, c_3 \geq 1$ . Let  $\psi$  be as in (32). Then*

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |Q^t \{G_i H_i\}|^2 \leq \mathbb{E}_{b_j \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |Q\{b_j\}|^2 + n^{1/2} (5^6 c_2 c_3)^d \tau^{1/6} + O_{m,n}(\tau^{-1/3} p^{-1/2}).$$

*Proof.* Let  $\lambda > 0$ , and define  $\psi_\lambda$  as in (31). From Lemma 3.6,  $\psi_\lambda(x) \geq \psi(x) \geq 0$  for all  $x \in \mathbb{R}$ . So,

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |Q^t \{G_i H_i\}|^2 \leq \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi_\lambda |Q^t \{G_i H_i\}|^2. \quad (33)$$

From Theorem 3.5,

$$\begin{aligned} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi_\lambda |Q^t \{G_i H_i\}|^2 &\leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi_\lambda |Q\{b_i\}|^2 \\ &+ \lambda^{-2} (O(n^{3/2})(5^6 c_2 c_3)^d \tau^{1/2} + O_{n,m}(p^{-1/2})). \end{aligned} \quad (34)$$

Using  $\psi_\lambda(x) \leq \psi(x) + \lambda$  for all  $x \geq 0$ ,

$$\frac{1}{n} \text{Tr} \psi_\lambda |Q(\sigma)|^2 \leq \frac{1}{n} \text{Tr} \psi |Q(\sigma)|^2 + \lambda, \quad \forall \sigma \in \{-1, 1\}^m. \quad (35)$$

Combining (33), (34) and (35) completes the proof, with a choice of  $\lambda$  such that  $\lambda^3 = \Theta(n^{3/2} \tau^{1/2})$ .  $\square$

Recall the definition of  $T_\rho$  in (10) and the function  $\text{Chop}: \mathbb{R} \rightarrow \mathbb{R}$ ,  $\text{Chop}(t) = t$  if  $|t| \leq 1$ ,  $\text{Chop}(t) = 1$  if  $t \geq 1$ , and  $\text{Chop}(t) = -1$  if  $t \leq -1$ .

**Corollary 3.8 (Smoothed Version of Theorem 3.7).** *Suppose  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$  with  $\|f(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$ . Let  $0 < \rho < 1$  and let  $\tau := \max_{i=1, \dots, m} \text{Inf}_i f$ . Assume  $\|\mathbb{E}(G_1 G_1^*)^2\| \leq c_2$  and  $\|\mathbb{E}(G_1 G_1^*)^3\| \leq c_3$  for some  $c_2, c_3 \geq 1$ . Then  $\|Q_f\|_{L_2, \mathcal{G}} \leq 1$  and*

$$\frac{1}{n} \|T_\rho Q_f^t - \text{Chop} T_\rho Q_f^t\|_{L_2, \mathcal{G}}^2 \leq 10n^{1/2} \tau^{\frac{1-\rho}{30(c_2 c_3)}} + O_{m,n}(\tau^{-1/3} p^{-1/2}). \quad (36)$$

*Proof.* Using the elementary inequality  $[\max(0, t-1)]^2 \leq \psi(t^2)$  for all  $t \geq 0$ , where  $\psi$  is defined in (32), applied to the singular values of  $T_\rho Q_f^t$ ,

$$\frac{1}{n} \|T_\rho Q_f^t - \text{Chop} T_\rho Q_f^t\|_{L_2, \mathcal{G}}^2 \leq \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |T_\rho Q_f^t \{G_i H_i\}|^2. \quad (37)$$

We first apply Theorem 3.7 to  $P_{\leq d}(T_\rho Q_f^t)$ , where  $d \in \mathbb{N}$  is to be determined later. Since  $0 < \rho < 1$ , (10), (7) imply that

$$\max_{i=1, \dots, m} \text{Inf}_i P_{\leq d} T_\rho Q_f \leq \max_{i=1, \dots, m} \text{Inf}_i P_{\leq d} Q_f \leq \max_{i=1, \dots, m} \text{Inf}_i Q_f,$$

and we get by Theorem 3.7 and (12)

$$\begin{aligned} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |P_{\leq d} T_\rho Q_f^t \{G_i H_i\}|^2 &\leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |P_{\leq d} T_\rho Q_f \{b_i\}|^2 \\ &+ n^{1/2} (5^6 c_2 c_3)^d \tau^{1/6} + O_{m,n}(\tau^{-1/3} p^{-1/2}). \end{aligned} \quad (38)$$

For any  $a, b \in \mathbb{R}$ ,  $|\psi((a+b)^2) - \psi(a^2)| \leq 2|a||b| + 2|b|^2$  follows by  $|\psi((a+b)^2) - \psi(a^2)| \leq |b| \max_{t \in [a, a+b]} \left| \frac{d}{dt} \psi(t^2) \right| \leq |b| 2(|a| + |b|)$ . Combining with the Cauchy-Schwarz inequality,

$$\begin{aligned} \text{Tr} \psi |T_\rho Q_f^t|^2 - \text{Tr} \psi |P_{\leq d} T_\rho Q_f^t|^2 &= \text{Tr} \psi |T_\rho P_{\leq d} Q_f^t + T_\rho P_{> d} Q_f^t|^2 - \text{Tr} \psi |T_\rho P_{\leq d} Q_f^t|^2 \\ &\leq 2(\text{Tr} |T_\rho P_{\leq d} Q_f^t|^2)^{1/2} (\text{Tr} |T_\rho P_{> d} Q_f^t|^2)^{1/2} + 2 \text{Tr} |T_\rho P_{> d} Q_f^t|^2. \end{aligned}$$

Taking expectation values and using (14), (11) and (10), which imply that

$$\begin{aligned} \mathbb{E}_{\substack{G_k \sim \mathcal{V} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |T_\rho P_{> d} Q_f^t|^2 &= \mathbb{E}_{G_j \sim \mathcal{G}} \text{Tr} |T_\rho P_{> d} Q_f^t|^2 \\ &= \mathbb{E}_{b_j \sim \mathcal{B}} \text{Tr} |T_\rho P_{> d} Q_f^t|^2 \\ &= \sum_{S \subseteq \{1, \dots, m\}: |S| > d} \rho^{2|S|} \text{Tr}(\widehat{f}(S)(\widehat{f}(S))^*) \\ &\leq \rho^{2d} \sum_{S \subseteq \{1, \dots, m\}} \text{Tr}(\widehat{f}(S)(\widehat{f}(S))^*) \\ &\leq n \rho^{2d}, \end{aligned}$$

we get

$$\frac{1}{n} \left| \mathbb{E}_{b_i \sim \mathcal{B}} \left( \text{Tr} \psi |T_\rho Q_f^t \{b_i\}|^2 - \text{Tr} \psi |P_{\leq d} T_\rho Q_f^t \{b_i\}|^2 \right) \right| \leq 4\rho^d, \quad (39)$$

$$\frac{1}{n} \left| \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \left( \text{Tr} \psi |T_\rho Q_f^t \{G_i H_i\}|^2 - \text{Tr} \psi |P_{\leq d} T_\rho Q_f^t \{G_i H_i\}|^2 \right) \right| \leq 4\rho^d. \quad (40)$$

Using  $t \leq |t|$  for any  $t \in \mathbb{R}$ ,

$$\begin{aligned}
& \frac{1}{n} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \psi |T_\rho Q_f^t \{G_i H_i\}|^2 - \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} \psi |T_\rho Q_f^t \{b_i\}|^2 \right) \\
& \leq \frac{1}{n} \left| \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \psi |T_\rho Q_f^t \{G_i H_i\}|^2 - \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \psi |T_\rho P_{\leq d} Q_f^t \{G_i H_i\}|^2 \right| \\
& \quad + \frac{1}{n} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \psi |T_\rho P_{\leq d} Q_f^t \{G_i H_i\}|^2 - \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} \psi |T_\rho P_{\leq d} Q_f^t \{b_i\}|^2 \right) \\
& \quad + \frac{1}{n} \left| \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} \psi |T_\rho P_{\leq d} Q_f^t \{b_i\}|^2 - \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} \psi |T_\rho Q_f^t \{b_i\}|^2 \right| \\
& \leq n^{1/2} (5^6 c_2 c_3)^d \tau^{1/6} + 8\rho^d + O_{m,n}(\tau^{-1/3} p^{-1/2}),
\end{aligned} \tag{41}$$

where the last inequality uses (40) to bound the first term, (38) for the second and (39) for the third. From (9) and (10) we get  $\|T_\rho Q_f(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$ , so by definition of  $\psi$  we have  $\mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} \psi |T_\rho Q_f^t \{b_i\}|^2 = 0$ . Combining (37) with (41) and choosing  $d = \min(\max(1, -\frac{\log(\tau)}{30(c_2 c_3)}), m)$  completes the proof (using  $-\log \rho \geq 1 - \rho$  for all  $0 < \rho < 1$ ).  $\square$

**3.3. Moment Majorization.** Theorem 3.5 implies that the even moments of a noncommutative multilinear polynomial follow a majorization principle. Although we will not make use of Theorem 3.9 for the applications in this paper, we include it as the statement could be of independent interest; the theorem has analogues in both the commutative [MOO10] and free probability settings [DN14].

**Theorem 3.9 (Noncommutative Majorization Principle for  $2K^{\text{th}}$  Moments).** *Let  $Q$  be a noncommutative multilinear polynomial of degree  $d$  in  $m$  variables, as in (4). Suppose  $\|Q(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$ . Let  $p > n$ , and let  $Q^t$  be the zero-padded extension of  $Q$ , as defined in Definition 2.1. Let  $\tau := \max_{i=1, \dots, m} \text{Inf}_i Q$ . Let  $G_i \sim \mathcal{G}$ . Assume that  $\|\mathbb{E}(G_1 G_1^*)^K\| \leq c_K$ , for some  $K \in \mathbb{N}$  and  $c_K \geq 1$ . Then*

$$\begin{aligned}
& \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^t \{G_i H_i\}|^{2K} \\
& \leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} |Q \{b_i\}|^{2K} + K^3 (2K - 1)^{2dK} c_K^d n^{2K} \tau^{1/4} + O_{m,n}(p^{-1/2} K^2 \tau^{-1/4}).
\end{aligned} \tag{42}$$

*Proof.* We begin with an upper tail estimate for  $Q^t$ . From Markov's inequality,

$$\mathbb{P}_{\substack{G_i \sim \mathcal{G} \otimes I \\ H_i \sim \mathcal{H}_p}} (\text{Tr} |Q^t|^2 > t) = \mathbb{P}_{\substack{G_i \sim \mathcal{G} \otimes I \\ H_i \sim \mathcal{H}_p}} ((\text{Tr} |Q^t|^2)^K > t^K) \leq t^{-K} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} [\text{Tr} |Q^t|^2]^K. \tag{43}$$

Since  $Q^t = I^t Q^t$  (where here  $I$  denotes the  $n \times n$  identity matrix), Hölder's inequality implies

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} (\text{Tr} |Q^t|^2)^K = \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} (\text{Tr} |I^t Q^t|^2)^K \leq n^{K-1} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |Q^t|^2^K. \tag{44}$$

Combining (43) and (44) and applying Theorem 3.1,

$$\begin{aligned} \mathbb{P}_{\substack{G_i \sim \mathcal{G} \otimes I \\ H_i \sim \mathcal{H}_p}} (\text{Tr} |Q^\iota|^2 > t) &\leq t^{-K} n^{K-1} (2K-1)^{2dK} c_K^d \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |Q^\iota|^2 \right)^K \\ &\leq t^{-K} n^{2K-1} (2K-1)^{2dK} c_K^d. \end{aligned} \quad (45)$$

Let  $s > 0$  be a constant to be fixed later. Define  $\psi: [0, \infty) \rightarrow [0, \infty)$  so that  $\psi(t) = t^K$  for any  $0 \leq t \leq s$ , and  $\psi$  is linear with slope  $Ks^{K-1}$  on  $(s, \infty)$ . It is possible to construct such a  $\psi$  with all three derivatives bounded, and in particular the third derivative bounded by some  $a_3 \leq K^3 s^{K-3}$  on the interval  $[s, s+1]$ . From Theorem 3.5,

$$\begin{aligned} &\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |Q^\iota \{G_i H_i\}|^2 \\ &\leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |Q\{b_i\}|^2 + K^3 s^{K-3} n^{3/2} (5^6 c_2 c_3)^{d\tau^{1/2}} + O_{n,m}(p^{-1/2} K^2 s^{K-2}). \end{aligned} \quad (46)$$

Finally, defining  $a = K^3 s^{K-3} n^{3/2} (5^6 c_2 c_3)^{d\tau^{1/2}} + O_{n,m}(p^{-1/2} K^2 s^{K-2})$  and letting  $D$  be the event that  $\text{Tr} \psi |Q^\iota \{G_i H_i\}|^2 \leq s$ , we have

$$\begin{aligned} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^\iota \{G_i H_i\}|^{2K} &= \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^\iota \{G_i H_i\}|^{2K} \cdot 1_D + \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^\iota \{G_i H_i\}|^{2K} \cdot 1_{D^c} \\ &= \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi |Q^\iota \{G_i H_i\}|^2 + \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^\iota \{G_i H_i\}|^{2K} \cdot 1_{D^c}. \end{aligned}$$

Using (46) to bound the first term and Cauchy-Schwarz for the second, the above can be bounded as

$$\begin{aligned} &\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} |Q^\iota \{G_i H_i\}|^{2K} \\ &\leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |Q\{b_i\}|^2 + a + \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \left( \frac{1}{n} \text{Tr} |Q^\iota \{G_i H_i\}|^{2K} \right)^2 \right)^{1/2} \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} 1_{D^c} \right)^{1/2} \\ &\leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |Q\{b_i\}|^2 + a + \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |Q^\iota \{G_i H_i\}|^{4K} \right)^{1/2} \left( \mathbb{P}_{\substack{G_i \sim \mathcal{G} \otimes I \\ H_i \sim \mathcal{H}_p}} (\text{Tr} |Q^\iota|^2 > s) \right)^{1/2} \\ &\leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi |Q\{b_i\}|^2 + a + (2K-1)^{2dK} c_K^d \left( \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |Q^\iota \{G_i H_i\}|^2 \right)^K s^{-K/2} n^{K-1}, \end{aligned}$$

using Theorem 3.1 to bound the first term inside a square root, and (45) for the second. Finally, using Lemma 2.5 and  $\|Q(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$ ,  $\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |Q^\iota \{G_i H_i\}|^2 \leq 1$ .

Choosing  $s = \tau^{-1/(4K)}$  finishes the proof.  $\square$

#### 4. MAXIMIZING NONCOMMUTATIVE NOISE STABILITY

By adapting the proof of the Majority is Stablest Theorem from [MOO10], we can get the following consequence of Corollary 3.8.

**Corollary 4.1.** *Let  $0 \leq \rho < 1$  and let  $\varepsilon > 0$ . Let  $\delta = 20n^{1/2}\tau^{\frac{1-\rho}{30(c_2c_3)}} + O_{m,n}(\tau^{-1/3}p^{-1/2})$ . Then there exists  $\tau > 0$  such that, if  $f: \{-1, 1\}^m \rightarrow M_n(\mathbb{C})$  satisfies  $\|f(\sigma)\| \leq 1$  for all  $\sigma \in \{-1, 1\}^m$ ,  $\mathbb{E}_{b_i \sim \mathcal{B}} f\{b_i\} = 0$ , and  $\max_{i=1, \dots, m} \text{Inf}_i(f) < \tau$ , then*

$$\frac{1}{n} \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} |T_\rho Q_f\{b_i\}|^2 \leq \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |\text{Chop} T_\rho Q_f\{G_i H_i\}|^2 + O(\varepsilon + \delta). \quad (47)$$

Moreover,  $|\frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \text{Chop} T_\rho Q_f\{G_i H_i\}| \leq \delta$ .

**Remark 4.2.** In the case  $n = p = 1$ , it is known from [MOO10] that the right-hand side of (47) is  $\frac{2}{\pi} \arcsin \rho + \varepsilon$ . For larger  $n$ , the left side of (47) can be interpreted as the noise stability of  $Q_f$  with discrete inputs, and the right side as the noise stability of a function with operator norm pointwise bounded by 1 under random Gaussian matrix inputs. Eq. (47) can thus be thought of as a matrix-valued version of one of the two main steps in the proof of the Majority is Stablest Theorem. However, for larger  $n$ , there seems to be no version of Borell's isoperimetric inequality that describes what the right-hand side of (47) should be. (Recall that Borell's isoperimetric inequality states that the noise stability of a subset of Euclidean space of fixed Gaussian measure is maximized when the set is a half space.)

*Proof.* Since  $\mathbb{E}_{b_i \sim \mathcal{B}} f\{b_i\} = 0$ , we have  $\mathbb{E}_{b_i \sim \mathcal{B}} Q_f\{b_i\} = 0$ . Using the Cauchy-Schwarz inequality and Corollary 3.8,

$$\begin{aligned} & \left| \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |T_\rho Q_f\{G_i H_i\}|^2 - \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |\text{Chop} T_\rho Q_f\{G_i H_i\}|^2 \right| \\ &= \left| \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |T_\rho Q_f\{G_i H_i\}|^2 - \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \left( \text{Chop} T_\rho Q_f\{G_i H_i\} [T_\rho Q_f\{G_i H_i\}]^* \right) \right| \\ & \quad + \left| \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \left( \text{Chop} T_\rho Q_f\{G_i H_i\} [T_\rho Q_f\{G_i H_i\}]^* \right) - \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |\text{Chop} T_\rho Q_f\{G_i H_i\}|^2 \right| \\ &\leq \left( \|T_\rho Q_f\{G_i H_i\}\|_{2,\nu} + \|\text{Chop} T_\rho Q_f\{G_i H_i\}\|_{2,\nu} \right) \cdot \|T_\rho Q_f\{G_i H_i\} - \text{Chop} T_\rho Q_f\{G_i H_i\}\|_{2,\nu} \\ &\leq 20n^{1/2}\tau^{\frac{1-\rho}{30(c_2c_3)}} + O_{m,n}(\tau^{-1/3}p^{-1/2}). \end{aligned}$$

Therefore,

$$\begin{aligned} & \frac{1}{n} \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} |T_\rho Q_f\{b_i\}|^2 \stackrel{(11) \wedge (14)}{=} \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |T_\rho Q_f\{G_i H_i\}|^2 \\ & \leq \frac{1}{n} \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} |\text{Chop} T_\rho Q_f\{G_i H_i\}|^2 + 20n^{1/2}\tau^{\frac{1-\rho}{30(c_2c_3)}} + O_{m,n}(\tau^{-1/3}p^{-1/2}), \end{aligned}$$

proving (47).

Using  $\mathbb{E}_{b_i \sim \mathcal{B}} Q'_f \{b_i\} = \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} T_\rho Q'_f \{G_i H_i\} = 0$  and the Cauchy-Schwarz inequality,

$$\begin{aligned} \frac{1}{n} \left| \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr Chop} T_\rho Q'_f \{G_i H_i\} \right| &= \frac{1}{n} \left| \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr Chop} T_\rho Q'_f \{G_i H_i\} - \mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} T_\rho Q'_f \{G_i H_i\} \right| \\ &\leq 20n^{1/2} \tau^{\frac{1-\rho}{30(c_2 c_3)}} + O_{m,n}(\tau^{-1/3} p^{-1/2}), \end{aligned}$$

using Corollary 3.8 again.  $\square$

## 5. AN ANTI-CONCENTRATION INEQUALITY

As in [MOO10], we can use our invariance principle to prove anti-concentration estimates of polynomials.

**Corollary 5.1 (An Anti-Concentration Estimate).** *There exists a constant  $C > 0$  such that the following holds. Let  $Q: (M_n(\mathbb{C}))^m \rightarrow M_n(\mathbb{C})$  be a noncommutative multilinear polynomial of degree  $d$ . Assume  $\mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} |Q\{b_i\}|^2 \leq 1$ . Let  $\tau = \max_{1 \leq j \leq m} \text{Inf}_j(Q)$ . Define  $\text{Var}(Q) = \mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} |Q\{b_i\} - (\mathbb{E}_{b_j \sim \mathcal{B}} Q\{b_j\})|^2$ . Then, for any  $t \in \mathbb{R}$ ,*

$$\begin{aligned} \frac{1}{n} \mathbb{P}_{\substack{G_i \sim \mathcal{G} \otimes I \\ H_i \sim \mathcal{H}}} (\|Q'\{G_i H_i\}\| > t) &\leq \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > t) + O(n^3 c_3^d \tau^{1/100}) \\ &\quad + Cd(4\tau^{1/100} n / [\text{Var}(Q)]^{1/2})^{1/d} + O_{m,n}(\tau^{-1/100} p^{-1/2}). \end{aligned}$$

**Remark 5.2.** The  $\frac{1}{n}$  term on the left side of the inequality seems to be an artifact of our proof method. It comes from (50) below, where we bound the normalized trace of a matrix by its operator norm.

*Proof.* Define  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  by  $\phi(x) = c \cdot \exp\left(\frac{-1}{1-x^2}\right)$  if  $|x| < 1$  and  $\phi(x) = 0$  for all other  $x \in \mathbb{R}$ . The constant  $1/2 < c < 4$  is chosen so that  $\int_{\mathbb{R}} \phi(x) dx = 1$ . It is well-known that  $\phi$  is an infinitely differentiable function with bounded derivatives.

Fix  $r, s > 0$ . Define  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  by

$$\psi(x) = \begin{cases} 0 & \text{if } x \leq r - s \\ \frac{s-r+x}{2s} & \text{if } r - s \leq x \leq r + s \\ 1 & \text{if } x > r + s. \end{cases} \quad (48)$$

Define  $\psi_\lambda(x) = \psi * \phi_\lambda(x) = \int_{\mathbb{R}} \psi(y) \phi_\lambda(x - y) dy$ . Then  $\psi_\lambda(x) = \psi(x)$  for any  $x \in \mathbb{R}$  with  $x > r + s + \lambda$  or  $x < r - s - \lambda$ . So,

$$\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \text{Tr} \psi_\lambda |Q'\{G_i H_i\}|^2 \geq \mathbb{P}_{\substack{G_i \sim \mathcal{G} \otimes I \\ H_i \sim \mathcal{H}}} (\|Q'\{G_i H_i\}\| > r + s + \lambda). \quad (49)$$

$$\mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi_\lambda |Q\{b_i\}|^2 \leq \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > r - s - \lambda). \quad (50)$$

Note that  $|\psi'''(x)| \leq 10^{10} \lambda^{-2} s$ . Applying Theorem 3.5,

$$\begin{aligned} &\mathbb{E}_{\substack{G_k \sim \mathcal{G} \otimes I \\ H_k \sim \mathcal{H}_p}} \frac{1}{n} \text{Tr} \psi_\lambda |Q'\{G_i H_i\}|^2 \\ &\leq \mathbb{E}_{b_i \sim \mathcal{B}} \frac{1}{n} \text{Tr} \psi_\lambda |Q\{b_i\}|^2 + s \lambda^{-2} n^{3/2} 10^{10} (5^6 c_3)^d \tau^{1/4} + O_{m,n}(\lambda^{-2} s p^{-1/2}). \end{aligned} \quad (51)$$

Combining (49), (50) and (51),

$$\begin{aligned}
& \frac{1}{n} \mathbb{P}_{\substack{G_i \sim \mathcal{G} \otimes I \\ H_i \sim \mathcal{H}}} (\|Q'\{G_i H_i\}\| > r + s + \lambda) \\
& \leq \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q'\{b_i\}\| > r - s - \lambda) + s\lambda^{-2} 10^{10} (5^6 c_3)^d \tau^{1/4} + O_{m,n}(s\lambda^{-2} p^{-1/2}) \\
& = \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > r + s + \lambda) + \mathbb{P}_{b_i \sim \mathcal{B}} (r + s + \lambda > \|Q\{b_i\}\| > r - s - \lambda) \\
& \quad + s\lambda^{-2} n^{3/2} 10^{10} (5^6 c_3)^d \tau^{1/4} + O_{m,n}(s\lambda^{-2} p^{-1/2}). \tag{52}
\end{aligned}$$

It remains to show that the second term in (52) is small. To this end we apply the anti-concentration result of [CW01, Theorem 8] (with  $q = 2d$  in their notation) to get: there exists an absolute constant  $C' > 0$  such that, if  $g_1, \dots, g_m$  are i.i.d. standard real Gaussian random variables, and if  $Q$  is any noncommutative multilinear polynomial, then for all  $\varepsilon > 0$ ,

$$\mathbb{P}_{g_1, \dots, g_m} (\|Q\{g_i\}\| < \varepsilon) \leq C' d(\varepsilon / [\mathbb{E}_{g_1, \dots, g_m} \|Q\{g_i\}\|^2]^{1/2})^{1/d}.$$

Since  $\mathbb{E}_{g_1, \dots, g_m} \|Q\{g_i\}\|^2 \geq \mathbb{E}_{g_1, \dots, g_m} \frac{1}{n} \text{Tr} |Q\{g_i\}|^2 \geq \mathbb{E}_{g_1, \dots, g_m} \frac{1}{n} \text{Tr} |Q\{g_i\} - \widehat{Q}(\emptyset)|^2$ , we conclude that, for any  $r \in \mathbb{R}$ , we have the following ‘‘small ball’’ probability estimate.

$$\mathbb{P}_{g_1, \dots, g_m} (|\|Q\{g_i\}\| - r| < \varepsilon) \leq C' d(\varepsilon n / [\mathbb{E}_{g_1, \dots, g_m} \text{Tr} |Q\{g_i\} - \widehat{Q}(\emptyset)|^2]^{1/2})^{1/d}. \tag{53}$$

Now, applying the invariance principle [IM12, Theorem 3.6] with the function  $\Psi: M_n(\mathbb{C}) \rightarrow \mathbb{R}$  defined by

$$\Psi(A) = \begin{cases} 0 & \text{if } \|A\| \leq r - 2s \\ \frac{2s - r + \|A\|}{s} & \text{if } r - 2s \leq \|A\| \leq r - s \\ 1 & \text{if } r - s \leq \|A\| \leq r + s \\ \frac{2s + r - \|A\|}{s} & \text{if } r + s \leq \|A\| \leq r + 2s \\ 0 & \text{if } \|A\| > r + 2s. \end{cases}$$

We get

$$|\mathbb{E}_{b_i \sim \mathcal{B}} \Psi(Q\{b_i\}) - \mathbb{E}_{g_1, \dots, g_m} \Psi(Q\{g_i\})| \leq \frac{2}{s} n^3 C'' \tau^{1/50}. \tag{54}$$

So, applying the definition of  $\Psi$  to (54), we get

$$\begin{aligned}
\mathbb{P}_{b_i \sim \mathcal{B}} (|\|Q\{b_i\}\| - r| < s) & \leq \frac{2}{s} C'' n^3 \tau^{1/50} + \mathbb{P}_{g_1, \dots, g_m} (|\|Q\{g_i\}\| - r| < 2s) \\
& \stackrel{(53)}{\leq} \frac{2}{s} C'' n^3 \tau^{1/50} + C' d(2sn / [\mathbb{E}_{g_1, \dots, g_m} \text{Tr} |Q\{g_i\} - \widehat{Q}(\emptyset)|^2]^{1/2})^{1/d} \\
& = \frac{2}{s} C'' n^3 \tau^{1/50} + C' d(2sn / [\mathbb{E}_{b_i \sim \mathcal{B}} \text{Tr} |Q\{b_i\} - \widehat{Q}(\emptyset)|^2]^{1/2})^{1/d}.
\end{aligned}$$

Finally, substitute the last inequality into (52) and set  $s = \lambda = \tau^{1/100}$ .  $\square$

**Remark 5.3.** The theorem [IM12, Theorem 3.6] used in (54) provides an extra multiplicative factor of  $2^{n^2}$  in (54). However, this constant can be removed in the following way. Using their notation, they define a function  $\phi: \mathbb{R}^k \rightarrow \mathbb{R}$  so that  $\phi(x) = \exp(\frac{-1}{1-\|x\|_2^2})$  if  $\|x\|_2 < 1$  and  $\phi(x) = 0$  otherwise. (In the present paper, we set  $k = n^2$ .) This is the function they use in their convolution formula. If we instead use a function  $\phi$  which is a product of



one-dimensional functions, each of which is supported in the interval  $[-1, 1]$ , e.g.  $\phi(x) = \prod_{i=1}^k e^{\frac{-1}{1-x^2}}$ , then the factor  $2^k$  no longer appears in their proof.

**Remark 5.4.** The stronger, though more restrictive anti-concentration inequality

$$\begin{aligned} \sup_{t \in \mathbb{R}} \left| \mathbb{P}_{g_1, \dots, g_m} (\|Q\{g_i\}\| > t) - \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > t) \right| \\ \leq O(n^3 c_3^d \tau^{1/100}) + Cd(2\tau^{1/100} n / [\text{Var}(Q)]^{1/2})^{1/d}, \end{aligned} \quad (55)$$

follows more directly from [CW01, Theorem 8] and [IM12, Theorem 3.6] by repeating the argument above. For example, if we use  $\psi$  defined in (48), then [IM12, Theorem 3.6] implies that

$$|\mathbb{E}_{b_i \sim \mathcal{B}} \psi(Q\{b_i\}) - \mathbb{E}_{g_1, \dots, g_m} \psi(Q\{g_i\})| \leq \frac{2}{s} n^3 C'' \tau^{1/50}.$$

Applying the definition of  $\psi$  to this inequality, we get

$$\left| \mathbb{P}_{g_1, \dots, g_m} (\|Q\{g_i\}\| > r + s + \lambda) - \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > r - s - \lambda) \right| \leq \frac{2}{s} n^3 C'' \tau^{1/50}. \quad (56)$$

Therefore,

$$\begin{aligned} & \left| \mathbb{P}_{g_1, \dots, g_m} (\|Q\{g_i\}\| > r + s + \lambda) - \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > r + s + \lambda) \right| \\ &= \left| - \mathbb{P}_{g_1, \dots, g_m} (r + s + \lambda > \|Q\{g_i\}\| > r - s - \lambda) \right. \\ & \quad \left. + \mathbb{P}_{g_1, \dots, g_m} (\|Q\{g_i\}\| > r + s + \lambda) - \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > r - s - \lambda) \right| \\ &\leq \left| \mathbb{P}_{g_1, \dots, g_m} (r + s + \lambda > \|Q\{g_i\}\| > r - s - \lambda) \right| \\ & \quad + \left| \mathbb{P}_{g_1, \dots, g_m} (\|Q\{g_i\}\| > r + s + \lambda) - \mathbb{P}_{b_i \sim \mathcal{B}} (\|Q\{b_i\}\| > r - s - \lambda) \right|. \end{aligned}$$

The second term is bounded by (56) and the first term is bounded by (53), setting  $s = \lambda = \tau^{1/100}$ .

**Remark 5.5.** It would be desirable to upgrade Corollary 5.1 and (55) to the stronger inequalities presented in [MNV16]. We leave this research direction to future investigations.

**Acknowledgements.** Thanks to Todd Kemp, Elchanan Mossel, Assaf Naor, Krzysztof Oleszkiewicz, Dimitri Shlyakhtenko and Thomas Vidick for helpful discussions. Thanks also to the anonymous reviewer for several helpful comments and for finding several typos and mistakes.

## REFERENCES

- [Bha97] R. Bhatia, *Matrix analysis*, Graduate Texts in Mathematics, Springer New York, 1997.
- [Bon70] Aline Bonami, *Étude des coefficients de Fourier des fonctions de  $L^p(G)$* , Ann. Inst. Fourier (Grenoble) **20** (1970), no. fasc. 2, 335–402 (1971). MR 0283496 (44 #727)
- [BR15] Jonah Brown-Cohen and Prasad Raghavendra, *Combinatorial optimization algorithms via polymorphisms*, Electronic Colloquium on Computational Complexity (ECCC) **22** (2015), 7.
- [Cha06] Sourav Chatterjee, *A generalization of the Lindeberg principle*, Ann. Probab. **34** (2006), no. 6, 2061–2076. MR 2294976 (2008c:60028)

- [CW01] Anthony Carbery and James Wright, *Distributional and  $L^q$  norm inequalities for polynomials over convex bodies in  $\mathbb{R}^n$* , Math. Res. Lett. **8** (2001), no. 3, 233–248. MR 1839474
- [DN14] Aurélien Deya and Ivan Nourdin, *Invariance principles for homogeneous sums of free random variables*, Bernoulli **20** (2014), no. 2, 586–603. MR 3178510
- [DR13] Sjoerd Dirksen and Éric Ricard, *Some remarks on noncommutative Khintchine inequalities*, Bull. Lond. Math. Soc. **45** (2013), no. 3, 618–624. MR 3065031
- [DS01] Kenneth R. Davidson and Stanislaw J. Szarek, *Local operator theory, random matrices and Banach spaces*, Handbook of the geometry of Banach spaces, Vol. I, North-Holland, Amsterdam, 2001, pp. 317–366. MR 1863696
- [Gro72] Leonard Gross, *Existence and uniqueness of physical ground states*, J. Functional Analysis **10** (1972), 52–109. MR 0339722 (49 #4479)
- [Gro75] ———, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. MR 0420249 (54 #8263)
- [HT99] U. Haagerup and S. Thorbjørnsen, *Random matrices and  $K$ -theory for exact  $C^*$ -algebras*, Doc. Math. **4** (1999), 341–450 (electronic). MR 1710376 (2000g:46092)
- [IM12] Marcus Isaksson and Elchanan Mossel, *Maximally stable Gaussian partitions with discrete applications*, Israel J. Math. **189** (2012), 347–396. MR 2931402
- [Kal02] Gil Kalai, *A Fourier-theoretic perspective on the Condorcet paradox and Arrow’s theorem*, Adv. in Appl. Math. **29** (2002), no. 3, 412–426. MR 1942631
- [Kan14] Daniel M. Kane, *The correct exponent for the Gotsman-Linial conjecture*, Comput. Complexity **23** (2014), no. 2, 151–175. MR 3212596
- [Kho02] Subhash Khot, *On the power of unique 2-prover 1-round games*, Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing (New York), ACM, 2002, pp. 767–775 (electronic). MR MR2121525
- [KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial, *The influence of variables on boolean functions*, Proc. of 29th Annual IEEE Symposium on Foundations of Computer Science, 1988, pp. 68–80.
- [KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell, *Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?*, SIAM J. Comput. **37** (2007), no. 1, 319–357. MR 2306295 (2008d:68035)
- [KN09] Subhash Khot and Assaf Naor, *Approximate kernel clustering*, Mathematika **55** (2009), no. 1-2, 129–165. MR 2573605 (2011c:68166)
- [KN13] ———, *Sharp kernel clustering algorithms and their associated grothendieck inequalities*, Random Structures & Algorithms **42** (2013), no. 3, 269–300.
- [KVV14] Dmitry S. Kaliuzhnyi-Verbovetskyi and Victor Vinnikov, *Foundations of free noncommutative function theory*, Mathematical Surveys and Monographs, vol. 199, American Mathematical Society, Providence, RI, 2014. MR 3244229
- [MJC<sup>+</sup>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** (2014), no. 3, 906–945. MR 3189061
- [MN15] Elchanan Mossel and Joe Neeman, *Robust optimality of Gaussian noise stability*, J. Eur. Math. Soc. (JEMS) **17** (2015), no. 2, 433–482. MR 3317748
- [MNV16] Raghu Meka, Oanh Nguyen, and Van Vu, *Anti-concentration for polynomials of independent random variables*, Theory Comput. **12** (2016), Paper No. 11, 16. MR 3542863
- [MOO10] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, Ann. of Math. (2) **171** (2010), no. 1, 295–341. MR 2630040 (2012a:60091)
- [Mos10] Elchanan Mossel, *Gaussian bounds for noise correlation of functions*, Geom. Funct. Anal. **19** (2010), no. 6, 1713–1756. MR 2594620 (2011b:60080)
- [MP14] Shahar Mendelson and Grigoris Paouris, *On the singular values of random matrices*, J. Eur. Math. Soc. (JEMS) **16** (2014), no. 4, 823–834. MR 3191978
- [MS12] Mark W. Meckes and Stanislaw J. Szarek, *Concentration for noncommutative polynomials in random matrices*, Proc. Amer. Math. Soc. **140** (2012), no. 5, 1803–1813. MR 2869165 (2012j:60048)

- [Nel73] Edward Nelson, *The free Markoff field*, J. Functional Analysis **12** (1973), 211–227. MR 0343816 (49 #8556)
- [NPR10] Ivan Nourdin, Giovanni Peccati, and Gesine Reinert, *Invariance principles for homogeneous sums: universality of Gaussian Wiener chaos*, Ann. Probab. **38** (2010), no. 5, 1947–1985. MR 2722791 (2011g:60043)
- [O'D14a] Ryan O'Donnell, *Analysis of Boolean functions*, Cambridge University Press, 2014.
- [O'D14b] ———, *Social choice, computational complexity, gaussian geometry, and boolean functions*, Proceedings of the ICM, 2014.
- [Rot79] V. I. Rotar', *Limit theorems for polylinear forms*, J. Multivariate Anal. **9** (1979), no. 4, 511–530. MR 556909 (81m:60039)
- [Tal94] Michel Talagrand, *On Russo's approximate zero-one law*, Ann. Probab. **22** (1994), no. 3, 1576–1587. MR 1303654 (96g:28009)
- [TV11] Terence Tao and Van Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), no. 1, 127–204. MR 2784665 (2012d:60016)
- [Ver12] Roman Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, Compressed sensing, Cambridge Univ. Press, Cambridge, 2012, pp. 210–268. MR 2963170
- [Voi91] Dan Voiculescu, *Limit laws for random matrices and free products*, Invent. Math. **104** (1991), no. 1, 201–220. MR 1094052 (92d:46163)

DEPARTMENT OF MATHEMATICS, UCLA, LOS ANGELES, CA 90095-1555  
*E-mail address:* `stevenmheilman@gmail.com`