

Matrix Concentration and Equivalent Operators on Fock Spaces

Afonso S. Bandeira* Dmitriy Kunisky[†] Petar Nizic-Nikolac[‡]

Lucas Pesenti[§] Robert Wang[¶]

October 1, 2026

Abstract

We develop a method for proving matrix concentration inequalities by identifying random matrices with associated deterministic operators acting on suitable Fock spaces and controlling norms of these operators and their restrictions to low-order subspaces. Applying this method, we obtain new strengthenings of the non-commutative Khintchine inequality that improve on the state of the art, in particular sharpening recent inequalities due to Bandeira, Boedihardjo, and van Handel (2023) quantifying intrinsic freeness of random matrices. Our proofs of these results are based on relatively simple operator algebra arguments and involve neither Gaussian interpolation nor explicit combinatorics of tracial moments. Further, our techniques apply equally well to several models of non-commutative random variables of interest in the literature, such as operator series constructed from q -Gaussian and Γ -independent systems of operators, treating all of these objects with the same method. We obtain new norm bounds for such operators both in the style of the non-commutative Khintchine inequality and in the style of Lechner’s operator norm formula.

*Department of Mathematics, ETH Zürich. bandeira@math.ethz.ch

[†]Department of Applied Mathematics & Statistics, Johns Hopkins University. kunisky@jhu.edu

[‡]Department of Mathematics, ETH Zürich. petar.nizic-nikolac@ifor.math.ethz.ch

[§]Department of Mathematics, ETH Zürich. lpesenti@ethz.ch

[¶]Cheriton School of Computer Science, University of Waterloo. robert.wang2@uwaterloo.ca

Contents

1	Introduction	1
1.1	Main results	3
1.2	Proof techniques	6
1.3	Related work	6
1.4	Organization of the paper	7
1.5	Notation	7
1.6	Acknowledgments and statement on AI	8
2	Gaussian matrix concentration via Hermite polynomials	8
2.1	Creation and annihilation operators over Hermite polynomials	8
2.2	Multivariate Gaussians and equivalent operators of random matrices	9
2.3	Proofs of NCK-type inequalities	10
3	Preliminaries on Fock spaces and commutation relations	13
3.1	The full Fock space	14
3.2	General commutation relations	15
3.3	General number operator	16
4	Norm bounds under general commutation relations	18
4.1	A q -deformed NCK inequality	18
4.2	Lehner-type variational norm bounds beyond freeness	20
4.3	Specialization to Gaussian random matrices	21
5	Applications of norm bounds	24
5.1	Gaussian Cayley matrices	24
5.2	NCK-type inequalities for Γ -independent systems	25
5.3	Ground state energy of the SYK model	26
	References	28
A	Scalar optimization estimates	31

1 Introduction

Matrix concentration inequalities are one of the main tools of modern non-asymptotic random matrix theory, giving general-purpose bounds on norms of random matrices and finding applications in several areas of pure and applied mathematics as well as in physics, computer science, and statistics (see, e.g., the surveys [Tro15, vH17, Ban26]). Our goal in this paper is to develop a new perspective on matrix concentration inequalities that leads to improvements and unifications of previous results.

While our results handle a wide range of models, from random matrices to certain deterministic operators arising in non-commutative probability, we focus for now on the *Gaussian series* random matrix model,

$$X = \sum_{i=1}^n g_i A_i, \quad (1)$$

for $g_i \sim \mathcal{N}(0, 1)$ i.i.d. standard Gaussian random variables and $A_i \in M_d(\mathbb{C})_{\text{sa}}$ deterministic self-adjoint $d \times d$ matrix coefficients. Any self-adjoint random matrix with centered jointly Gaussian entries can be expressed in this way for suitable A_i . We focus on the goal of controlling the expectation of the spectral norm $\mathbb{E}\|X\|$; tail bounds can be obtained from these with standard tools (see the surveys cited above for details). Matrices that are not self-adjoint can also be treated by Hermitian dilation.

A classical result in this area is the celebrated Non-Commutative Khintchine (NCK) inequality.

Theorem 1.1 (NCK inequality [Lus86, LP91]). *In the above setting,¹*

$$\sigma(X) \lesssim \mathbb{E}\|X\| \lesssim \sigma(X)(\log d)^{1/2},$$

where we define the matrix variance parameter

$$\sigma(X) := \left\| \sum_{i=1}^n A_i^2 \right\|^{1/2} = \|\mathbb{E}X^2\|^{1/2}.$$

Only the logarithmic factor depending on the dimension d in the upper bound precludes the NCK inequality from giving an estimate of $\mathbb{E}\|X\|$ that is tight up to constants. This factor also corresponds to the dimensional factor appearing in the tail bound of the matrix Bernstein inequality.

A central goal of matrix concentration inequalities is to provide expectation and tail estimates for $\|X\|$, but two standard examples reveal that such estimates cannot be sharp up to constants while depending solely on $\sigma(X)$: for a Wigner matrix, with i.i.d. standard Gaussian entries above the diagonal, the logarithmic factor in NCK is known to be spurious (the lower bound of Theorem 1.1 is tight); on the other hand, for a diagonal matrix with i.i.d. standard Gaussian entries on the diagonal, the logarithmic factor is required (the upper bound of Theorem 1.1 is tight).

Various refinements of the NCK inequality have been proposed in recent years aiming to clarify this matter, starting with the work of Tropp [Tro18]. We focus on the following result of [BBvH23], which makes further progress in this direction using the additional parameter

$$v(X) := \left\| \sum_{i=1}^n \text{vec}(A_i) \text{vec}(A_i)^* \right\|^{1/2} = \|\text{Cov}(\text{vec}(X))\|^{1/2}. \quad (2)$$

¹We write $x \lesssim y$ if $x \leq Cy$ for a universal constant C , and $x \asymp y$ if $x \lesssim y$ and $y \lesssim x$.

The elegant method of that work compares X to a related deterministic operator $\mathcal{X}_{\text{free}}$, defined as

$$\mathcal{X}_{\text{free}} := \sum_{i=1}^n A_i \otimes s_i, \quad (3)$$

for $s_1, \dots, s_n$ a family of freely independent semicircular operators on a suitable Hilbert space (see Section 3.1) and satisfying

$$\sigma(X) \leq \|\mathcal{X}_{\text{free}}\| \leq 2\sigma(X).$$

In a sense, $\mathcal{X}_{\text{free}}$ can be seen as an “optimistic proxy” for X : if $\mathbb{E}\|X\| \approx \|\mathcal{X}_{\text{free}}\|$, then it is the lower bound of the NCK inequality that is tight. Because $\mathcal{X}_{\text{free}}$ also corresponds to the summands in the Gaussian series behaving as though they were freely independent, [BBvH23] refer to this situation as *intrinsic freeness* of X . One of the main results of [BBvH23] gives sufficient conditions for this to be the case.

Theorem 1.2 (Intrinsic freeness NCK inequality [BBvH23]). *In the above setting,*

$$\mathbb{E}\|X\| \leq \|\mathcal{X}_{\text{free}}\| \left(1 + C \left(\frac{v(X)}{\sigma(X)} \right)^{1/2} (\log d)^{3/4} \right) \quad (4)$$

$$\lesssim \sigma(X) + v(X)(\log d)^{3/2}. \quad (5)$$

Provided that $v(X)/\sigma(X) \ll (\log d)^{-3/2}$ (say over a sequence of random matrix models with growing dimension $d \rightarrow \infty$), we find that $\mathbb{E}\|X\| \leq (1 + o(1))\|\mathcal{X}_{\text{free}}\|$, a remarkably precise estimate in terms of the deterministic quantity $\|\mathcal{X}_{\text{free}}\|$. This norm $\|\mathcal{X}_{\text{free}}\|$ admits an explicit variational description due to Lehner [Leh99], which we discuss in Section 4.2; this description may also be formulated as a semidefinite program parameterized by the matrices A_i [Kun26]. Further, [BBvH23] develops a more detailed theory of how, under such assumptions, various other spectral statistics of X resemble those of $\mathcal{X}_{\text{free}}$.

However, a few aspects of this theory remain lacking. First, the logarithmic factors above are larger than the $(\log d)^{1/2}$ appearing in the original NCK inequality, and it is natural to expect that such improved NCK inequalities can be refined to have this smaller polylogarithmic dependence. Second, intuition drawn from the proofs of the NCK inequality and its refinements suggests that the *commutativity* or lack thereof of the A_i should govern whether the polylogarithmic factor should be present or not. In the theory of [BBvH23], this is measured by the parameter $v(X)$; however, it is unclear from its definition in what sense this parameter actually measures commutativity.² Finally and relatedly, the parameter $v(X)$ fails a natural invariance: we have $\|X \otimes I_m\| = \|X\|$ and $\sigma(X \otimes I_m) = \sigma(X)$, while v satisfies

$$v(X \otimes I_m) = v(X) \cdot m^{1/2}. \quad (6)$$

Thus over a sequence of such models with growing m the estimates of [BBvH23] degrade more and more, and accordingly Theorem 1.2 can become, for large enough m in such a construction, inferior even to the original Theorem 1.1.

Our immediate goal in this paper is to provide a different refinement of the NCK inequality that always improves on both Theorem 1.1 and Theorem 1.2, achieves a tight polylogarithmic factor $(\log d)^{1/2}$, depends on a more natural measurement of commutativity of the A_i , and enjoys the above invariance that is failed by $v(X)$ in (6).³

²In the arguments in [BBvH23], $v(\cdot)$ arises indirectly in bounds on the alignment parameter $w(\cdot)$ of Tropp [Tro18].

³It is worth mentioning that [BBvH23] shows that for a parameter to satisfy this invariance and sharply capture the spectral norm of Wigner matrices, it must fail to satisfy the triangle inequality, and indeed this is the case for the parameter $\kappa(X)$ we introduce below.

Our proof technique is quite different from previous work, including [BBvH23], and is operator-theoretic in nature. The general idea is, instead of the random matrix X , to consider the *equivalent operator*, denoted $\mathcal{X}$ below, that acts by matrix multiplication on the space of random vectors whose entries are polynomials in the underlying Gaussian variables $g_1, \dots, g_n$. The equivalent operator, as its name suggests, captures all relevant information about the behavior of X ; for example, we will show in Section 2.2 that the tracial moments $\mathbb{E} \operatorname{tr} X^k$ can be computed in terms of quadratic forms with $\mathcal{X}$. In the basis of Hermite polynomials, the action of $\mathcal{X}$ admits a concise combinatorial description in terms of *creation* and *annihilation operators* on a suitable *Fock space*. We borrow these useful notions from quantum mechanics and quantum field theory (QFT); see [Fol08, DG13] for mathematical references from that perspective. This same structure is used in the standard construction of the operators s_i appearing in the definition of $\mathcal{X}_{\text{free}}$ and Lehner’s calculation of the norm $\|\mathcal{X}_{\text{free}}\|$, and this connection allows us to relate $\mathbb{E}\|X\|$ with $\|\mathcal{X}_{\text{free}}\|$ via $\mathcal{X}$ without appealing to the interpolation argument of [BBvH23].

The same proof technique applies equally well to other operators constructed from creation and annihilation operators satisfying various *commutation relations*, as studied extensively in the literature on QFT and associated operator algebras. For instance, since $\mathcal{X}_{\text{free}}$ can be constructed in this way, our technique gives a simple proof of the upper bound side of Lehner’s formula for $\|\mathcal{X}_{\text{free}}\|$, as well as bounds on the norms of other operator series such as those having Γ -*independent* operators in place of the s_i , as studied in several recent works [CM26, OTY25, BKN⁺26a]. The most general bounds we prove all have the same variational form of Lehner’s formula, and we believe that our approach considerably clarifies this shared structure.

1.1 Main results

Matrix concentration We first present our main results for random matrices. Before describing our inequalities, let us introduce our new matrix concentration parameter:

$$\kappa(X) := \lambda_{\max}(C(X))^{1/2}, \quad (7)$$

$$C(X) := \sum_{i,j=1}^n (e_i e_j^*) \otimes (A_j A_i) = \begin{bmatrix} A_1^2 & A_2 A_1 & \cdots & A_n A_1 \\ A_1 A_2 & A_2^2 & \cdots & A_n A_2 \\ \vdots & \vdots & \ddots & \vdots \\ A_1 A_n & A_2 A_n & \cdots & A_n^2 \end{bmatrix}. \quad (8)$$

This may be understood as follows. Consider the block matrix $R = [A_1 \ A_2 \ \cdots \ A_n] \in M_{d,dn}(\mathbb{C})$. Then, we have $R^* R = \sum_{i,j=1}^n (e_i e_j^*) \otimes (A_i A_j)$, and

$$\lambda_{\max}(R^* R) = \|R\|^2 = \lambda_{\max}(R R^*) = \sigma(X)^2.$$

In contrast, the matrix $C(X)$ whose largest eigenvalue $\kappa(X)^2$ measures is the so-called *partial transpose* of $R^* R$, that is obtained by taking the conjugate transpose of every $d \times d$ block, replacing $A_i A_j$ with $A_j A_i$. The partial transpose operation plays a prominent role in quantum information theory; see, e.g., [Wat18, Section 6.2.3] for details. We emphasize that it can dramatically transform the spectrum of a matrix, for instance mapping a positive semidefinite matrix to an indefinite one, or a low-rank matrix to a full-rank one. For our purposes, we note that this operation is transparently related to the commutativity of the A_i . At the extreme, if the A_i all commute, then $C(X) = R^* R$ and so $\kappa(X) = \sigma(X)$. Indeed, this is the worst case, and we have:

Proposition 1.3. $\kappa(X) \leq \min\{\sigma(X), v(X)\}.$

Proof. Write $C = C(X)$. Suppose $u \in \mathbb{C}^{dn}$ has blocks $u_1, \dots, u_n \in \mathbb{C}^d$. Then, we have

$$|\langle u, Cu \rangle| = \left| \sum_{i,j=1}^n \langle A_j u_i, A_i u_j \rangle \right| \leq \sum_{i,j=1}^n \|A_i u_j\|^2 = \sum_{j=1}^n u_j^* \left(\sum_{i=1}^n A_i^2 \right) u_j \leq \sigma(X)^2 \|u\|^2,$$

showing $\kappa(X)^2 \leq \|C\| \leq \sigma(X)^2$.

For the other bound, define matrices $B_a := \sum_{i=1}^n (u_i)_a A_i \in M_d(\mathbb{C})$ for each $a \in [d]$. We have $\|B_a\|_F^2 = \|\sum_{i=1}^n (u_i)_a \text{vec}(A_i)\|^2 \leq v(X)^2 \sum_{i=1}^n |(u_i)_a|^2$. Then, we may also bound

$$|\langle u, Cu \rangle| = \left| \sum_{a,b,c=1}^d (B_a)_{bc} \overline{(B_b)_{ac}} \right| \leq \sum_{a,b,c=1}^d |(B_a)_{bc}|^2 = \sum_{a=1}^d \|B_a\|_F^2 \leq v(X)^2 \|u\|^2,$$

whereby $\kappa(X)^2 \leq \|C\| \leq v(X)^2$. □

Thus, the amount by which $\kappa(X)$ is smaller than $\sigma(X)$ is a natural measurement of the non-commutativity of the A_i . Further, $\kappa(X)$ is easily verified to have the invariance suggested earlier:

Proposition 1.4. $\kappa(X) = \kappa(X \otimes I_m)$.

Proof. This is immediate since $C(X \otimes I_m) = C(X) \otimes I_m$. □

Our first main result, a new strengthened version of the NCK inequality, is the following bound, resembling the bound (5) of [BBvH23] but with $\kappa(X)$ replacing $v(X)$ and with the logarithmic factor sharpened. Its proof is fully contained in Section 2.

Theorem 1.5 (Strengthened NCK inequality). *In the above setting,*

$$\mathbb{E}\|X\| \leq 2\sigma(X) + 6\kappa(X)(\log d)^{1/2}.$$

In Section 5.1, we give an illustrative example of an application where our bounds give superior estimates to both the NCK inequality (Theorem 1.1) and the estimates of [BBvH23] (Theorem 1.2). This is to the *Gaussian Cayley matrices* of [BKMZ24], a group-structured Gaussian series model proposed in that work as a “challenge case” for matrix concentration inequalities. Indeed, we show that the bounds of [BBvH23] behave identically to the NCK inequality in this case, while our improved NCK inequality (Theorem 1.5) and the parameter $\kappa(X)$ are sensitive to the particular group structure of such a model and in particular the profile of dimensions of irreducible representations of a group, which may be viewed as a measurement of non-commutativity.

Our tools also provide a bound with the leading term $\|\mathcal{X}_{\text{free}}\|$, providing an improvement to the bound (4) of [BBvH23]. As in [BBvH23], this bound depends on the $v(X)$ parameter.

Theorem 1.6 (Strengthened intrinsic freeness NCK inequality). *In the above setting,*

$$\mathbb{E}\|X\| \leq \|\mathcal{X}_{\text{free}}\| \left(1 + 8 \left(\frac{v(X)^2}{\sigma(X)^2} \log d \right)^{1/2} + 8 \left(\frac{v(X)^2}{\sigma(X)^2} \log d \right)^{1/3} \right). \quad (9)$$

In particular, this implies $\mathbb{E}\|X\| \leq (1 + o(1))\|\mathcal{X}_{\text{free}}\|$ whenever $v(X)/\sigma(X) \ll (\log d)^{-1/2}$. We also note that, both in (5) and (9), $v(X)$ can be replaced by $v'(X) := \min\{\sigma(X), v(X)\}$ by simply using the NCK inequality if $v(X) \geq \sigma(X)$. The proof of Theorem 1.6 can be found in Section 4.3.

The authors of [BBvH23] posed the question of whether a bound on Gaussian series norms holds of the form $\mathbb{E}\|X\| \leq \|\mathcal{X}_{\text{free}}\| + C v(X)(\log d)^{1/2}$ for an absolute constant $C \geq 0$, which corresponds to (9) without the last term inside the parentheses. Theorem 1.6 is a partial answer to this question: (9) both provides the same asymptotic sufficient condition $v(X)/\sigma(X) \ll (\log d)^{-1/2}$ to guarantee $\mathbb{E}\|X\| \leq (1 + o(1))\|\mathcal{X}_{\text{free}}\|$, and gives the same bound (up to constants) when instead we have $v(X)/\sigma(X) \gtrsim (\log d)^{-1/2}$. The difference between the claims is only in how quickly the $o(1)$ term in the former case decays.

Lehner-type operator norm bounds Lastly, our technique also implies bounds on certain operators of interest in the study of operator algebras beyond their role in random matrix theory. The bounds we obtain are relatives of Lehner's formula [Leh99] for $\|\mathcal{X}_{\text{free}}\|$ mentioned above. This formula reads:

$$\left\| A_0 \otimes \mathbf{1} + \sum_{i=1}^n A_i \otimes s_i \right\| = \max_{\varepsilon \in \{\pm 1\}} \inf_{Z \succ 0} \lambda_{\max} \left(\varepsilon A_0 + Z^{-1} + \sum_{i=1}^n A_i Z A_i \right), \quad (10)$$

where the s_i are freely independent semicircular elements and $\mathbf{1}$ is the identity operator acting on the same Hilbert space as the s_i . We give a new simple proof of the upper bound of this result, which also generalizes to a broad class of s_i as follows. See Theorem 4.4 for an expanded version.

Theorem 1.7 (Generalized Lehner-type norm bound). *Let $\Gamma \in M_n(\mathbb{R})_{\text{sa}}$ with $\Gamma_{ij} \in [-1, 1]$, $\Gamma_{i,i} < 1$ for $i, j \in [n]$, $A_0, \dots, A_n \in M_d(\mathbb{C})_{\text{sa}}$, $\ell_1, \dots, \ell_n$ be the creation operators on the Γ -Fock space (Definition 3.2), and $\mathcal{X} = A_0 \otimes \mathbf{1} + \sum_{i=1}^n A_i \otimes (\ell_i + \ell_i^*)$. Let N_Γ be the number operator (Definition 3.7) associated to these operators. For each $Z \in M_d(\mathbb{C})_{\text{sa}}$ with $Z \succ 0$, define the block matrix*

$$C_\Gamma(Z) := [\Gamma_{ij} \cdot A_j Z A_i]_{i,j=1}^n \in M_{dn}(\mathbb{C}).$$

Then

$$\|\mathcal{X}\| \leq \max_{\varepsilon \in \{\pm 1\}} \inf_{Z \succ 0} \left\{ \lambda_{\max} \left(\varepsilon A_0 + \sum_{i=1}^n A_i Z A_i + Z^{-1} \right) + \lambda_{\max}^+(C_\Gamma(Z)) \|N_\Gamma\| \right\}.$$

In the case $\Gamma \equiv 0$, the $\ell_i + \ell_i^*$ are freely independent semicircular elements, and so we indeed recover the upper bound of Lehner's formula since $C_\Gamma(Z) = 0$ for all Z in this case. Also, we note that if we have $\Gamma \equiv 1$ then $C_\Gamma(I_d)$ is the same as the matrix $C(X)$ for $X = \sum_{i=1}^n g_i A_i$ featuring in the definition of $\kappa(X)$. The above does not apply to this case and indeed in this case N_Γ is unbounded, but the proof technique is entirely analogous to our proof of Theorem 1.5. For other choices of Γ , we obtain operator bounds relevant to the literature on q -Gaussian systems (when $\Gamma \equiv q \in [-1, 1]$, as in Example 3.5) and Γ -independent systems (when Γ is the adjacency matrix of a simple graph, i.e. has $\Gamma_{i,i} = 0$ and $\Gamma_{i,j} \in \{0, 1\}$, as in Example 3.6).

Finally, we note that we do not derive a single master inequality that implies all of our bounds simultaneously, and it is likely that the tools developed here may be useful in combinations beyond those considered in this paper. For instance, while improved versions of the matrix Bernstein inequality for sums of independent random matrices (see [Tro15, Theorem 1.6.2]) can be derived immediately by combining our Theorem 1.5 with the universality principles of [BvH24, Tro26], in a forthcoming work we will establish further sharpened universality principles and Bernstein-type inequalities by operator-theoretic arguments.

1.2 Proof techniques

Our proof ideas apply broadly to operator series, or so-called matrix *pencils*, of the form

$$\mathcal{X} = \sum_{i=1}^n A_i \otimes (\ell_i + \ell_i^*), \quad (11)$$

for some collection of pairs of adjoint operators ℓ_i, ℓ_i^* on a Hilbert space $\mathcal{H}$ and $A_i \in M_d(\mathbb{C})_{\text{sa}}$. As we discuss below in Section 2, a random matrix with Gaussian coefficients as above is a special case of this construction. In this case, the ℓ_i are the *creation operators* and ℓ_i^* the associated *annihilation operators* on a *symmetric Fock space*, related through the *canonical commutation relations*. Alternatively, they can be viewed as operators on polynomials involving multiplications and differentiations associated to the Hermite polynomials. These operators play an important role in QFT, as we will briefly discuss in Section 2.

We view such a pencil $\mathcal{X}$ as an operator *equivalent* to the corresponding Gaussian random matrix $X = \sum_{i=1}^n g_i A_i$. For example, expected trace powers of the random matrix can be computed in terms of the operator $\mathcal{X}$: for all $p \geq 1$,

$$\mathbb{E} \operatorname{tr} X^p = \operatorname{tr} \otimes \tau(\mathcal{X}^p),$$

where, for x an operator on $\mathcal{H}$, $\tau(x) := \langle \Omega, x\Omega \rangle$ for a suitable *vacuum element* $\Omega \in \mathcal{H}$. Here tr is the normalized trace of matrices, so that the tensor product $\operatorname{tr} \otimes \tau$ is a linear functional on $M_d(\mathbb{C}) \otimes \mathcal{A}$ for suitable algebras $\mathcal{A}$ of operators on $\mathcal{H}$.

Our toolkit centers around bounds in the Loewner ordering on operators of the form (11). On the one hand, applying such bounds to the equivalent operators of random matrices readily leads to bounds on expected norms. On the other hand, other operators of interest such as $\mathcal{X}_{\text{free}}$ in the theory of [BBvH23] also have the form (11), and our techniques apply equally well to those and to several other operators, including the Γ -independent and q -Gaussian models mentioned above, illuminating the analogies between all these settings.

The basic approach to deducing these bounds on $\mathcal{X}$ is to partition the pencil into its creation and annihilation parts:

$$\mathcal{X} = \mathcal{L} + \mathcal{L}^*, \quad \mathcal{L} = \sum_{i=1}^n A_i \otimes \ell_i. \quad (12)$$

To utilize commutation relations between the two parts, we analyze the products $\mathcal{L}\mathcal{L}^*$ or $\mathcal{L}^*\mathcal{L}$, or variants such as $\mathcal{L}^*\mathcal{Z}\mathcal{L} + \mathcal{Z}^{-1}$ with an auxiliary weight operator $\mathcal{Z} \succ 0$. Expanding such products, in the simplest case $\mathcal{L}^*\mathcal{L}$, then leads to products $(A_i \otimes \ell_i)^*(A_j \otimes \ell_j) = (A_i A_j) \otimes (\ell_i^* \ell_j)$. This in turn allows us to use the algebraic structure of the ℓ_i to simplify and in particular to take advantage of various notions of non-commutativity.

1.3 Related work

Other recent progress on norm bounds The idea of analyzing the norm of an operator by decomposing it as $\mathcal{L} + \mathcal{L}^*$ (as in (12)) and then exploiting relations between $\mathcal{L}\mathcal{L}^*$ and $\mathcal{L}^*\mathcal{L}$ has appeared very recently in two works: in Caspers and Chen's construction of strongly convergent matrix models for q -Gaussian systems [CC26, Lemma 3.10], and in Basu, Kothari and Midha's bound on the norm of the Sachdev-Ye-Kitaev (SYK) Hamiltonian [BKM26, Lemma 4.9]. Both examples can be viewed as special instantiations of our framework (see Theorem 4.2), corresponding to the particular commutation relations arising in those applications.

Further, we note that the original problem studied in [BKM26] is itself a random matrix problem, and that their argument similarly proceeds by representing the random matrix by a deterministic operator. However, the deterministic representation obtained in [BKM26] is specific to the SYK model, and differs from the natural representation arising through our approach; see Remark 5.6. The general strategy of representing a random matrix by a deterministic operator and analyzing the latter also appears in Heidary’s recent work on the Nelson-Nguyen conjecture [Hei26].

Γ -independent and q -Gaussian systems Two important special cases of the general commutation relations considered in this work are Γ -independent⁴ semicircular systems [Mlo04, SW16] and q -Gaussian systems [BS91, BS92, BKS97]; see Examples 3.5 and 3.6. Noncommutative Khintchine inequalities for Γ -independent semicircular systems were established by Collins and Miyagawa [CM26] and Oliveira Santos, Tripathi, and Youssef [OTY25]. The argument of [CM26] also uses the decomposition (12). Building on a bound from [OTY25] (Lemma 3.9), our Corollary 5.2 gives a sharper noncommutative Khintchine inequality in this setting. A different application of the same framework yields Corollary 5.3, which is a key ingredient in the proof of strong convergence of the Kikuchi matrix to a q -Gaussian variable in our companion work [BKN⁺26a].

Fock space methods The representation of Gaussian random variables in terms of creation and annihilation operators is classical in the literature on Wiener chaos [Itô51] and Fock spaces [Seg56]. This is the basis of the approach that we use in Section 2 to represent a Gaussian random matrix by an equivalent deterministic operator. In our setting, these operators appear with matrix rather than scalar coefficients; our main new ingredient is to introduce inequalities in the Loewner order that exploit the noncommutativity of these coefficients (see, e.g., Theorem 2.1).

1.4 Organization of the paper

The remainder of the paper is organized as follows. Section 2 introduces the method of equivalent operators within the more familiar setting of Gaussian random variables and linear operators defined on polynomials. It proves the strengthened NCK inequality (Theorem 1.5) and we hope for it also to serve as an accessible motivation for the role of notions of non-commutative probability in our setting while focusing on classical objects. Sections 3 and 4 introduce important notions of non-commutative probability and derive inequalities for several classes of non-commutative variables (realized as operators) using our method. While these objects are less familiar, we aim for these sections to be accessible to a reader without background in non-commutative probability. Section 5 develops some applications of these results.

1.5 Notation

We use the Kronecker notation $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. The standard basis of $\mathbb{R}^d$ or $\mathbb{C}^d$ is denoted $e_1, \dots, e_d$, where $(e_i)_j = \delta_{ij}$. We write $\chi = \sum_{i=1}^d e_i \in \mathbb{R}^d$ for the all-ones vector; its dimension is clear from context when it appears. We write $M_d(\mathbb{F})$ for the set of $d \times d$ matrices with entries in a field $\mathbb{F}$, $M_d(\mathbb{F})_{\text{sa}}$ for the set of self-adjoint such matrices, and $M_{d_1, d_2}(\mathbb{F})$ for the set of $d_1 \times d_2$ matrices with entries in $\mathbb{F}$. For $A \in M_d(\mathbb{C})_{\text{sa}}$, we write $\lambda_{\max}(A)$ for the largest eigenvalue of A and $\lambda_{\max}^+(A) := \max\{0, \lambda_{\max}(A)\}$. Further, $\text{Tr}(A) := \sum_{i=1}^d A_{ii}$ and $\text{tr}(A) := \frac{1}{d} \text{Tr}(A)$ will respectively denote the unnormalized and normalized trace. We write $I_d \in M_d(\mathbb{R})$ for the d -dimensional identity matrix, while $\mathbf{1}$ is the identity operator in various Hilbert spaces that are

⁴Also known as ε -independent or Λ -free.

clear from context. We use calligraphic letters ($\mathcal{X}$, $\mathcal{L}$, ...) for operators on tensor spaces $\mathbb{C}^d \otimes \mathcal{H}$ for a Hilbert space $\mathcal{H}$, and ordinary letters for other matrices and operators.

1.6 Acknowledgments and statement on AI

We thank Joel Tropp for comments on an early draft manuscript including some of the results appearing in this work.

LP acknowledges support from an ETH Zürich Postdoctoral Fellowship (Project 26-1 FEL-092). RW's work on this project was supported by the Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII-GSST). Part of this work was done while RW was an academic guest in the Department of Mathematics at ETH Zürich. PNN and RW thank the Simons Institute for the Theory of Computing and the organizers of the *Spectral Theory Beyond Graphs* program.

AI assistants were used, over the course of several months, in the process of developing the ideas of working with equivalent operators, and later in proofreading the paper. Our original motivation came from our efforts to understand Kikuchi matrices through Γ -independent limits, Feige's hypergraph Moore bound conjecture, and [BKMR25, Conjecture 9] (see our companion work [BKN⁺26a]). We remark, however, that several of what we view as the most interesting outcomes of this research effort were obtained only after both Feige's conjecture and [BKMR25, Conjecture 9] were already solved in [KX26, BKN⁺26b, SH26].

The authors wrote the paper, verified all of its contents, and all errors are theirs.

2 Gaussian matrix concentration via Hermite polynomials

While our technique also provides norm bounds for other operators, including $\mathcal{X}_{\text{free}}$ and a similar construction from Γ -independent families, we start by focusing on the Gaussian series model (1) and proving Theorem 1.5.

As sketched above, the general premise of our proof technique in this setting is to associate to a *random matrix* a deterministic *equivalent operator*, which in particular captures the combinatorics of expected traces of powers of the matrix. In general, these operators in this case and the other cases below can be constructed over Fock spaces from creation and annihilation operators, which we introduce in Section 3. For Gaussian random matrices, however, these operators can be described in a straightforward way as differential operators acting on polynomials, associated to simple identities satisfied by Hermite polynomials and Gaussian integration. We follow this simplified approach in this section for the sake of exposition.

2.1 Creation and annihilation operators over Hermite polynomials

We first describe the operator-theoretic viewpoint on Hermite polynomials and Gaussian combinatorics in one dimension. Let γ denote the standard Gaussian measure on $\mathbb{R}$. We write $\mathbb{E}$ for expectations with respect to $g \sim \gamma$. The space of polynomials $\mathbb{R}[g]$ forms a Hilbert space when equipped with the standard inner product $\langle f, h \rangle := \mathbb{E}[f(g)h(g)]$. The Hermite polynomials, $\{h_k\}_{k=0}^\infty$, form a natural orthonormal basis for $\mathbb{R}[g]$ under this inner product. They can be defined via a three-term recurrence: $h_0 = 1$, and

$$gh_k(g) = \sqrt{k+1}h_{k+1}(g) + \sqrt{k}h_{k-1}(g). \quad (13)$$

From these definitions, it follows that expectations can be viewed as inner products with h_0 , which we will use below:

$$\mathbb{E}[f(g)] = \langle f, h_0 \rangle.$$

Equation (13) implies that the operation of multiplication by g can be described by the action of two linear operators on $\mathbb{R}[g]$: one raises the degree of each Hermite polynomial while the other lowers the degree. We define these operators $\ell, \ell^* : \mathbb{R}[g] \rightarrow \mathbb{R}[g]$ by the following action on the Hermite basis:

$$\ell h_k = \sqrt{k+1} h_{k+1}, \quad \ell^* h_k = \sqrt{k} h_{k-1}, \quad \ell^* h_0 = 0.$$

The operator ℓ is called a *creation operator* and ℓ^* an *annihilation operator*. The operators ℓ and ℓ^* are indeed adjoint as $\langle \ell h_{k_1}, h_{k_2} \rangle = \langle h_{k_1}, \ell^* h_{k_2} \rangle$ for all $k_1, k_2 \geq 0$.

One can also interpret the creation and annihilation operators as differential operators: it follows from standard properties of the Hermite polynomials that $\ell^* = \partial_g$ and $\ell = g - \partial_g$, where “ g ” denotes the operator of multiplication by g on polynomials. The fact that they are adjoint corresponds to the Gaussian integration by parts identity:

$$\int \partial_g f(g) d\gamma(g) = \langle \ell^* f, h_0 \rangle = \langle f, \ell h_0 \rangle = \int g f(g) d\gamma(g).$$

The three-term recurrence (13) can then be thought of as $g = \ell + \ell^* = (g - \partial_g) + \partial_g$.

Below we will be interested in expected traces of powers of Gaussian series models. As a simpler illustration of the same principle, here we consider deriving the moments of a scalar Gaussian random variable. We do so in terms of a linear functional on operators called the *vacuum state*,

$$\tau(a) := \langle h_0, a h_0 \rangle.$$

According to the above identification $g = \ell + \ell^*$, this allows us to express the Gaussian moments in terms of the vacuum state and powers of the operator on the right-hand side:

$$\mathbb{E}[g^p] = \langle h_0, g^p \rangle = \langle h_0, (\ell + \ell^*)^p h_0 \rangle = \tau((\ell + \ell^*)^p).$$

For this reason we call $\ell + \ell^*$ an *equivalent operator* to the scalar random variable g (with respect to the state functional τ).

By linearity of the state functional, we may expand the right-hand side in terms of states of products of ℓ and ℓ^* . The defining feature of ℓ and ℓ^* , also useful in working with such expressions, is their commutation relation. Observe that their commutator acts by $\ell^* \ell h_k - \ell \ell^* h_k = (k+1)h_k - k h_k = h_k$. Thus, we have

$$\ell^* \ell - \ell \ell^* = \mathbf{1}.$$

In QFT, this relation is known as the *bosonic commutation relation* [DG13]. As one application, using this relation repeatedly, one may prove the familiar formula for the moments of g .⁵

2.2 Multivariate Gaussians and equivalent operators of random matrices

The previous construction can easily be extended to a family of independent Gaussian variables. Let γ_n be the standard Gaussian measure on $\mathbb{R}^n$. For each $k_1, \dots, k_n$, write $h_{k_1, \dots, k_n}(g_1, \dots, g_n) := h_{k_1}(g_1) \cdots h_{k_n}(g_n)$. These are the n -variate Hermite polynomials, forming an orthonormal basis of the space of polynomials $\mathbb{R}[g_1, \dots, g_n]$ under the inner product $\langle f, h \rangle = \mathbb{E}[f(g_1, \dots, g_n) h(g_1, \dots, g_n)]$.

Under the same convention as before, we associate to every g_i a multiplication operator acting on $\mathbb{R}[g_1, \dots, g_n]$, and the action of g_i can be decomposed into annihilation and creation parts. Formally, for all $h \in \mathbb{R}[g_1, \dots, g_n]$, we have

$$g_i h = (\ell_i + \ell_i^*) h,$$

⁵From the definitions of τ , ℓ , and ℓ^* we have that, for any operator a , $\tau(a \ell^*) = \tau(\ell a) = 0$. The bosonic commutation relation gives $\tau((\ell + \ell^*)^p) = \tau(\ell^* (\ell + \ell^*)^{p-1}) = (p-1) \tau((\ell + \ell^*)^{p-2})$ upon commuting ℓ^* to the right $p-1$ times. This gives by induction that $\tau((\ell + \ell^*)^p) = (p-1)!!$ for p even and zero otherwise.

where ℓ_i and ℓ_i^* act on the Hermite basis by

$$\begin{aligned}\ell_i h_{k_1, \dots, k_n} &= \sqrt{k_i + 1} h_{k_1, \dots, k_i + 1, \dots, k_n}, \\ \ell_i^* h_{k_1, \dots, k_n} &= \begin{cases} \sqrt{k_i} h_{k_1, \dots, k_i - 1, \dots, k_n} & \text{if } k_i \geq 1, \\ 0 & \text{if } k_i = 0. \end{cases}\end{aligned}$$

These operators satisfy the *multivariate bosonic commutator relations*, given by

$$\ell_i^* \ell_j - \ell_j \ell_i^* = \delta_{i,j} \mathbf{1}. \quad (14)$$

These relations, together with the vacuum state, recover joint moments of the Gaussian variables:

$$\mathbb{E}[g_{i_1} g_{i_2} \dots g_{i_p}] = \tau \left(\prod_{j=1}^p (\ell_{i_j} + \ell_{i_j}^*) \right) = \left\langle h_0, \prod_{j=1}^p (\ell_{i_j} + \ell_{i_j}^*) h_0 \right\rangle \quad (15)$$

where $h_0 = 1$ is the constant function. As before, this can be verified to recover Wick's formula for the left-hand side. An equivalent operator of any Gaussian linear combination, $\sum_i a_i g_i$ for $a_i \in \mathbb{R}$, is therefore given by the operator $\sum_i a_i (\ell_i + \ell_i^*)$. In QFT, this is known as a *Segal field operator*.⁶

To study a self-adjoint Gaussian random matrix, $X = \sum_{i=1}^n g_i A_i$, we define an equivalent operator $\mathcal{X}$ in the space $M_d(\mathbb{C})$ tensored with the space of linear operators over $\mathbb{R}[g_1, \dots, g_n]$:

$$\mathcal{X} := \sum_{i=1}^n A_i \otimes (\ell_i + \ell_i^*) = \mathcal{L} + \mathcal{L}^*,$$

where $\mathcal{L} = \sum_i A_i \otimes \ell_i$. As we will see below, the seemingly innocuous decomposition $\mathcal{X} = \mathcal{L} + \mathcal{L}^*$ plays a central role in our arguments.

We note that these operators act on the space $\mathbb{C}^d \otimes \mathbb{R}[g_1, \dots, g_n]$, which may be viewed as d -dimensional vectors, each of whose entries is a polynomial in $g_1, \dots, g_n$. By (15), $\mathcal{X}$ then captures the trace moments of X via

$$\mathbb{E} \operatorname{tr} X^p = \operatorname{tr} \otimes \tau (\mathcal{X}^p) = \operatorname{tr} \otimes \tau ((\mathcal{L} + \mathcal{L}^*)^p) \quad (16)$$

where $\tau(\mathcal{A}) = \langle h_0, \mathcal{A} h_0 \rangle$ is the vacuum state and tr is the normalized trace, so that $\operatorname{tr} \otimes \tau$ is a state on $M_d(\mathbb{C})$ tensored with the space of linear maps on $\mathbb{R}[g_1, \dots, g_n]$.

2.3 Proofs of NCK-type inequalities

Our goal is to bound $\mathbb{E} \|X\|$ by bounding the traces on the left-hand side. One may show that, for a bounded operator $\mathcal{T}$, we have $|\operatorname{tr} \otimes \tau(\mathcal{T})| \leq \|\mathcal{T}\|$, so one might hope to apply this bound on the right-hand side. While versions of such a strategy are sensible in some applications we discuss later, in the Gaussian setting operators like $\mathcal{X}$ and $\mathcal{L}$ are unbounded. Instead, we bound these and related operators in the *Loewner order* by simpler operators, and then consider the restrictions of those operators to finite-dimensional subspaces of $\mathbb{C}^d \otimes L^2(\gamma_n)$ spanned by low-degree polynomials, on which they are bounded.

A crucial role in the bounds in Loewner order will be played by what is referred to in QFT as the *number operator*,

$$N = \sum_{i=1}^n \ell_i \ell_i^*. \quad (17)$$

⁶We omit the conventional factor of $1/\sqrt{2}$.

The number operator acts on Hermite polynomials by $Nh_{k_1, \dots, k_n} = (k_1 + \dots + k_n)h_{k_1, \dots, k_n}$. In other words, it scales each Hermite polynomial by its total degree.⁷

We note that, since Hermite polynomials can have arbitrary degree, the number operator is unbounded, like $\mathcal{X}$ and $\mathcal{L}$. But, the number operator is bounded on the space of polynomials of degree bounded by some p ; indeed, it is clear from the above description that this restriction has norm at most p . Thus the number operator is easy to use in the above restriction strategy.

Let us start by deriving bounds on some related operators depending on the number operator (17). The parameters of a Gaussian series model $X = \sum_{i=1}^n g_i A_i$ given earlier in Theorem 1.1 and (7) will appear in our statement, which we recall are:

$$\sigma(X)^2 = \left\| \sum_{i=1}^n A_i^2 \right\|, \quad \kappa(X)^2 = \lambda_{\max}(C(X)), \quad C(X) = \sum_{i,j=1}^n (e_i e_j^*) \otimes (A_j A_i).$$

As we will see momentarily, the following are direct analogs of the NCK inequality and our improvement in Theorem 1.5 for operators in the Loewner ordering, and readily give these NCK-type inequalities for $\mathbb{E}\|X\|$ after a short additional argument truncating these operators to spaces of polynomials of bounded degree, on which they are bounded operators. The Loewner ordering is convenient here because it allows us to state related inequalities without explicit truncation. We note also that these bounds may be viewed as variants of ones for scalar linear combinations from the QFT literature; see, e.g., [Coo53] and [RS09, Lemma 2.1].

Theorem 2.1 (Loewner NCK-type inequalities). *Let $X = \sum_{i=1}^n g_i A_i$ be a self-adjoint Gaussian random matrix. For $\mathcal{L} = \sum_{i=1}^n A_i \otimes \ell_i$ we have*

$$\begin{aligned} \mathcal{L}\mathcal{L}^* &\preceq \sigma(X)^2 I_d \otimes N, \\ \mathcal{L}^*\mathcal{L} &\preceq \left(\sum_{i=1}^n A_i^2 \right) \otimes \mathbf{1} + \kappa(X)^2 I_d \otimes N. \end{aligned}$$

Proof. Define the row flattening of the matrix coefficients $R = [A_1 \cdots A_n]$. Note that $\|R^*R\| = \|RR^*\| = \sigma(X)^2$ and that we can write

$$\mathcal{L}\mathcal{L}^* = [I_d \otimes \ell_1 \quad \cdots \quad I_d \otimes \ell_n] (R^*R \otimes \mathbf{1}) \begin{bmatrix} I_d \otimes \ell_1^* \\ \vdots \\ I_d \otimes \ell_n^* \end{bmatrix} \preceq \sigma(X)^2 \sum_{i=1}^n I_d \otimes \ell_i \ell_i^* = \sigma(X)^2 I_d \otimes N,$$

giving the first inequality. For the second inequality, we use the commutation relations (14):

$$\begin{aligned} \mathcal{L}^*\mathcal{L} &= \sum_{i,j=1}^n A_i A_j \otimes \ell_i^* \ell_j \\ &= \sum_{i,j=1}^n A_i A_j \otimes (\delta_{ij} \mathbf{1} + \ell_j \ell_i^*) \\ &= \sum_{i=1}^n A_i^2 \otimes \mathbf{1} + [I_d \otimes \ell_1 \quad \cdots \quad I_d \otimes \ell_n] (C(X) \otimes \mathbf{1}) \begin{bmatrix} I_d \otimes \ell_1^* \\ \vdots \\ I_d \otimes \ell_n^* \end{bmatrix}. \end{aligned}$$

Since the diagonal blocks of $C(X)$ are A_i^2 , we have $\lambda_{\max}(C(X)) \geq 0$ and so the second summand can be upper bounded in Loewner order by $\kappa(X)^2 I \otimes N$, completing the proof. $\square$

⁷In the QFT formulation over Fock spaces, the degree of a Hermite polynomial corresponds to a number of “particles”, in which case the number operator counts the number of particles, justifying its name.

Remark 2.2. We emphasize the key point above that the two orders in which $\mathcal{L}$ and $\mathcal{L}^*$ can be multiplied admit very different bounds in Loewner order. It is precisely from working with the second choice $\mathcal{L}^*\mathcal{L}$ that we obtain a superior inequality, and it is precisely the use of the commutation relations in working with this product that we arrive at the partially-transposed matrix $C(X)$.

We now proceed to bound the norm of the number operator when restricted to low degree polynomials. Below we write Π_p for the orthogonal projection in $\mathbb{R}[g_1, \dots, g_n]$ to polynomials of degree at most p .

Proposition 2.3. *Let $N = \sum_i \ell_i \ell_i^*$ denote the bosonic number operator as defined in (17). Then,*

$$\Pi_p N \Pi_p \preceq p \Pi_p.$$

Proof. Both N and Π_p are diagonalized by the orthonormal Hermite basis, and further $\Pi_p N \Pi_p$ and Π_p have finite rank since the space of polynomials of degree at most p has finite dimension. All Hermite polynomials whose degree is larger than p are in the kernel of both $\Pi_p N \Pi_p$ and $p \Pi_p$. The eigenvalues associated to Hermite polynomials whose degree is some $\Delta \leq p$ are p for $p \Pi_p$ and Δ for $\Pi_p N \Pi_p$, and the result follows. $\square$

We are now ready to prove the main inequalities of this section, the upper bounds of the original NCK inequality (Theorem 1.1) and of our improved NCK inequality (Theorem 1.5). Both come from the following bounds on tracial moments.

Corollary 2.4. *In the setting of Theorem 2.1 we have the following trace moment bounds for $p \geq 1$:*

$$\begin{aligned} (\mathbb{E} \operatorname{tr}(X^{2p}))^{\frac{1}{2p}} &\leq 2\sqrt{p} \sigma(X), \\ (\mathbb{E} \operatorname{tr}(X^{2p}))^{\frac{1}{2p}} &\leq 2\sqrt{\sigma(X)^2 + (p-1)\kappa(X)^2}. \end{aligned}$$

Proof. We define the truncated operators

$$\begin{aligned} \mathcal{L}_p &:= (I_d \otimes \Pi_p) \mathcal{L} (I_d \otimes \Pi_p), \\ \mathcal{X}_p &:= (I_d \otimes \Pi_p) \mathcal{X} (I_d \otimes \Pi_p) = \mathcal{L}_p + \mathcal{L}_p^*. \end{aligned}$$

Note that since $\mathcal{L}^*$ lowers the degree we have $\mathcal{L}^*(I_d \otimes \Pi_p) = (I_d \otimes \Pi_p) \mathcal{L}^*(I_d \otimes \Pi_p) = \mathcal{L}_p^*$. Similarly, since $\mathcal{L}$ raises the degree we have $\mathcal{L}_p = (I_d \otimes \Pi_p) \mathcal{L} (I_d \otimes \Pi_p) = \mathcal{L} (I_d \otimes \Pi_{p-1})$.

Using Theorem 2.1 and Proposition 2.3, we obtain

$$\begin{aligned} \mathcal{L}_p \mathcal{L}_p^* &= (I \otimes \Pi_p) \mathcal{L} \mathcal{L}^* (I \otimes \Pi_p) \\ &\preceq \sigma(X)^2 I \otimes \Pi_p N \Pi_p \preceq p \sigma(X)^2 I \otimes \Pi_p, \\ \mathcal{L}_p^* \mathcal{L}_p &= (I \otimes \Pi_{p-1}) \mathcal{L}^* \mathcal{L} (I \otimes \Pi_{p-1}) \\ &\preceq \sum_i A_i^2 \otimes \Pi_{p-1} + \kappa(X)^2 I \otimes \Pi_{p-1} N \Pi_{p-1} \\ &\preceq (\sigma(X)^2 + (p-1)\kappa(X)^2) (I \otimes \Pi_{p-1}). \end{aligned}$$

Since projections have norm at most 1, we have, by triangle inequality,

$$\|\mathcal{X}_p\| \leq 2 \|\mathcal{L}_p\| \leq 2 \min \left\{ \sqrt{p} \sigma(X), \sqrt{\sigma(X)^2 + (p-1)\kappa(X)^2} \right\}. \quad (18)$$

Lastly, we relate $\mathbb{E} \operatorname{tr} X^{2p}$ with $\|\mathcal{X}_p\|$. Since each multiplication by $\mathcal{X}$ raises the degree of a polynomial by at most 1 (and h_0 has degree 0), we have that $\mathcal{X}^p(e_j \otimes h_0) = \mathcal{X}_p^p(e_j \otimes h_0)$ for all $j \in [d]$. Thus, by (16), we have

$$\mathbb{E} \operatorname{tr} X^{2p} = \operatorname{tr} \otimes \tau(\mathcal{X}^{2p}) = \frac{1}{d} \sum_{j=1}^d \|\mathcal{X}_p^p(e_j \otimes h_0)\|^2 \leq \|\mathcal{X}_p\|^{2p},$$

which together with (18) establishes the trace inequalities. $\square$

Using this, the proofs of the two NCK-type inequalities are immediate.

Corollary 2.5 (Concrete upper bound of Theorem 1.1). *For X as in (1), we have*

$$\mathbb{E} \|X\| \leq 4\sqrt{1 + \log d} \sigma(X).$$

Proof. By Jensen's inequality,

$$\mathbb{E} \|X\| \leq (\mathbb{E} \|X\|^{2p})^{\frac{1}{2p}} \leq (d \mathbb{E} \operatorname{tr} X^{2p})^{\frac{1}{2p}} = d^{\frac{1}{2p}} (\mathbb{E} \operatorname{tr} X^{2p})^{\frac{1}{2p}}. \quad (19)$$

The result then follows by choosing $p = \lceil \log d \rceil$, so that $d^{1/2p} \leq \sqrt{e} \leq 2$. $\square$

Our improvement follows by a slight variation of the same strategy.

Proof of Theorem 1.5. We again start with the bound of (19), which may be written as

$$\mathbb{E} \|X\| \leq \inf_{p \in \mathbb{N}} 2d^{1/2p} \sqrt{\sigma(X)^2 + p\kappa(X)^2}.$$

The result then follows by the result of Lemma A.2 on this optimization and simplifying using that $\log(2d) \leq 2 \log d$. $\square$

In Section 4.3, we will prove a strengthening of the estimate of Theorem 1.5 replacing the leading term $2\sigma(X)$ by $\|\mathcal{X}_{\text{free}}\|$, at the cost of a different error term depending on $v(X)$ instead of $\kappa(X)$. We will do this by introducing a variational component to the argument, akin to the proof of Lechner's formula [Leh99].

Remark 2.6 (Scalar Khintchine inequalities). We note that essentially the same argument, taking $d = 1$ or equivalently considering the Segal field operator $\sum_{i=1}^n a_i(\ell_i + \ell_i^*)$ for scalar coefficients $a_i \in \mathbb{R}$, gives the classical Khintchine inequalities for linear combinations of i.i.d. random variables, such as Rademacher variables [Khi23]. The only difference is that, in this case, one must work with a number operator associated with the orthogonal polynomials of this distribution rather than the Hermite polynomials. Thus the number operator captures all of the salient properties of the law of the creation and annihilation operators for such bounds, just as in our argument above.

3 Preliminaries on Fock spaces and commutation relations

In this section, we introduce a more general and abstract framework for similar calculations to the ones in the previous section that will generalize the algebraic properties of orthogonal polynomials (including Hermite polynomials). We use the framework of non-commutative probability; many of the relevant definitions also appear prominently in applications to QFT.

A *non-commutative probability space* is a pair $(\mathcal{A}, \tau)$, where $\mathcal{A}$ is a unital $*$ -algebra (interpreted as an algebra of random variables but often constructed as an operator algebra), and $\tau : \mathcal{A} \rightarrow \mathbb{C}$ is a linear functional called a *state* (interpreted as a trace or expectation) satisfying $\tau(\mathbf{1}) = 1$ and $\tau(a^*a) \geq 0$. The *law* of a self-adjoint element $x \in \mathcal{A}$ refers to the collection of *moments* $\tau(x^k)$. We say that the law is a certain real-valued probability measure μ if these moments are the moments of μ :

$$\tau(x^k) = \int \lambda^k d\mu(\lambda) \text{ for all } k \geq 1.$$

The creation and annihilation operators ℓ_i and ℓ_i^* discussed above can be viewed, in addition to their concrete realization over Hermite polynomials, as elements of a suitable non-commutative probability space. The bosonic commutation relations that these satisfy correspond to classical independence of scalar Gaussian random variables. Other forms of non-commutative independence, like free independence, can be realized by creation and annihilation operators satisfying different commutation relations (indeed, this is how one defines the operator $\mathcal{X}_{\text{free}}$ mentioned above). In this section, we give constructions of these operators on Hilbert spaces, which provide another way to interpret our previous use of $\mathbb{R}[g_1, \dots, g_n]$ using the Hermite polynomial basis.

3.1 The full Fock space

We begin by reviewing the construction of freely independent semicircular random variables that appeared for example as the s_i in $\mathcal{X}_{\text{free}}$ in (3), following [NS06, Lecture 7]. The *full Fock space* over $\mathbb{C}^n$ is defined as the space $\mathcal{F}(\mathbb{C}^n) := \bigoplus_{k \geq 0} (\mathbb{C}^n)^{\otimes k}$, equipped with the standard Hilbert space inner product, where we take by convention $(\mathbb{C}^n)^{\otimes 0} = \mathbb{C}\Omega$ for Ω a distinguished unit vector called the *vacuum* vector. Algebras of operators on $\mathcal{F}(\mathbb{C}^n)$ are given the structure of a non-commutative probability space by endowing them with the *vacuum state*, the functional

$$\tau(x) := \langle \Omega, x\Omega \rangle. \quad (20)$$

We call each $(\mathbb{C}^n)^{\otimes k}$ the *k-particle space* and the span of Ω the *0-particle space*, this terminology coming from the following interpretation. Let $e_1, \dots, e_n$ be an orthonormal basis of $\mathbb{C}^n$. The Fock space has a natural orthonormal basis $\{\Omega\} \cup \{e_{i_1} \otimes \dots \otimes e_{i_k} : i_1, \dots, i_k \in [n], k \in \mathbb{N}\}$. We can view this basis as the set of all words over the alphabet $\{e_1, \dots, e_n\}$, where Ω represents the empty word. The creation operator ℓ_i adds a letter to the left of a word:

$$\ell_i \Omega = e_i, \quad \ell_i(e_{i_1} \otimes \dots \otimes e_{i_k}) = e_i \otimes e_{i_1} \otimes \dots \otimes e_{i_k}. \quad (21)$$

Its adjoint ℓ_i^* removes the first letter if it is e_i and otherwise annihilates the word:

$$\ell_i^* \Omega = 0, \quad \ell_i^*(e_{i_1} \otimes \dots \otimes e_{i_k}) = \langle e_i, e_{i_1} \rangle e_{i_2} \otimes \dots \otimes e_{i_k}. \quad (22)$$

Then the operators ℓ_i and ℓ_j^* satisfy the *free commutation relations*:⁸

$$\ell_i^* \ell_j = \delta_{ij} \mathbf{1}. \quad (23)$$

With respect to the vacuum state, the operators

$$s_i = \ell_i + \ell_i^*, \quad i = 1, \dots, n,$$

⁸Also known as the *Cuntz-Toeplitz relations*, defined to study C^* -algebras generated by isometries [Cun77].

form a family of standard freely independent semicircular random variables [NS06, Corollary 7.17]. Each s_i has the semicircle law normalized to have variance 1 (in the sense of (20) above). Freeness means that a product of centered polynomials has state zero whenever adjacent factors involve different variables. Precisely, for every $m \geq 1$, $\tau(p_1(s_{i_1}) \cdots p_m(s_{i_m})) = 0$ whenever $p_1, \dots, p_m \in \mathbb{C}[x]$ and $i_1, \dots, i_m \in [n]$ satisfy $\tau(p_j(s_{i_j})) = 0$ for all j and $i_j \neq i_{j+1}$ for $1 \leq j < m$.⁹ The vacuum state satisfies the following properties for all a, b in the C^* -algebra generated by the s_i [NS06, Proposition 7.18]:

1. $\tau(ab) = \tau(ba)$.
2. $\tau(a^*a) \geq 0$, with $\tau(a^*a) = 0$ if and only if $a = 0$.

3.2 General commutation relations

So far, we have seen two examples of commutation relations that operators $\ell_1, \dots, \ell_n$ may satisfy: the bosonic commutation relations (14) and the free commutation relations (23). The following definition generalizes both of these.

Definition 3.1. Given any undirected weighted graph Γ on vertex set $[n]$ with edge weights $\Gamma_{i,j} \in [-1, 1]$, the Γ -canonical commutation relations (Γ -CCR) are

$$\ell_i^* \ell_j - \Gamma_{i,j} \ell_j \ell_i^* = \delta_{i,j} \mathbf{1}. \quad (24)$$

We allow this graph to have loops, so the diagonal weights $\Gamma_{i,i}$ are also defined.

The bosonic and free relations correspond to the cases $\Gamma \equiv 1$ and $\Gamma \equiv 0$, respectively. Bożejko and Speicher [BS94] show how to construct operators satisfying the Γ -CCR using a different inner product on the full Fock space $\mathcal{F}(\mathbb{C}^n)$.

Definition 3.2 (Γ -Fock space [BS94]). Define $\langle \cdot, \cdot \rangle_\Gamma$ on $\mathcal{F}(\mathbb{C}^n)$ as follows:

1. Set $\langle \Omega, \Omega \rangle_\Gamma = 1$.
2. For any $\xi \in (\mathbb{C}^n)^{\otimes k}$, $\xi' \in (\mathbb{C}^n)^{\otimes k'}$ for $k \neq k'$, set $\langle \xi, \xi' \rangle_\Gamma = 0$.
3. Otherwise, for every $\xi \in (\mathbb{C}^n)^{\otimes k-1}$ and $i, j_1, \dots, j_k \in [n]$, define recursively

$$\langle e_i \otimes \xi, e_{j_1} \otimes \cdots \otimes e_{j_k} \rangle_\Gamma = \sum_{\substack{r \in [k] \\ j_r = i}} \langle \xi, e_{j_1} \otimes \cdots \otimes e_{j_{r-1}} \otimes e_{j_{r+1}} \otimes \cdots \otimes e_{j_k} \rangle_\Gamma \prod_{s=1}^{r-1} \Gamma_{i, j_s}.$$

After quotienting $\mathcal{F}(\mathbb{C}^n)$ by the null space of $\langle \cdot, \cdot \rangle_\Gamma$ and taking the completion, we obtain a Hilbert space $\mathcal{F}_\Gamma(\mathbb{C}^n)$ known as the Γ -Fock space. On the Γ -Fock space, the creation operator ℓ_i from (21) and its corresponding adjoint operator ℓ_i^* for $\langle \cdot, \cdot \rangle_\Gamma$ satisfy the Γ -CCR (note that ℓ_i^* is different from the operator in (22)). We endow them with the vacuum state $\tau(x) = \langle \Omega, x\Omega \rangle$.

To interpret this definition, we specialize it to several important special cases.

Example 3.3 (Freely independent semicirculars). When $\Gamma \equiv 0$, the inner product $\langle \cdot, \cdot \rangle_\Gamma$ coincides with the standard inner product on the full Fock space, so we recover the construction of freely independent semicircular random variables from Section 3.1.

⁹Voiculescu made this definition to study free products of operator algebras [Voi85], using the full Fock space construction. The connection with random matrices came later via the related notion of asymptotic freeness [Voi91].

Example 3.4 (Classically independent Gaussians). When $\Gamma \equiv 1$, the k -particle space is obtained from $(\mathbb{C}^n)^{\otimes k}$ by identifying tensors that differ only by a permutation of the indices $\{1, \dots, k\}$. This in turn is equivalent to the construction in Section 2.2 by associating (up to a normalization) the Hermite polynomial $h_{k_1, \dots, k_n}$ with the symmetric Fock space basis vector that contains k_i copies of e_i . The operators $(\ell_1 + \ell_1^*, \dots, \ell_n + \ell_n^*)$ therefore have the joint law with respect to the vacuum state of classically independent standard Gaussian random variables.

Example 3.5 (q -Gaussian system). When $\Gamma \equiv q$ for some $q \in [-1, 1]$, the operators $(\ell_1 + \ell_1^*, \dots, \ell_n + \ell_n^*)$ form a q -Gaussian system, an interpolation between the above examples studied, for instance, by the works [FB70, BS91, BS92, Spe92, BKS97, BKN⁺26a]. Marginally, each $\ell_i + \ell_i^*$ has the q -Gaussian distribution, which we describe in Section 4.1.

Example 3.6 (Γ -independent semicircular family). Suppose Γ is a simple unweighted graph, i.e., $\Gamma_{i,j} \in \{0, 1\}$ for all $i \neq j$, and $\Gamma_{i,i} = 0$. We can interpret each original Fock basis vector as a word over the characters $[n]$. Quotienting the null space of $\langle \cdot, \cdot \rangle_\Gamma$ induces an equivalence relation between basis vectors, where two words are identified whenever one can be obtained from the other by repeatedly swapping adjacent letters that are adjacent in Γ . Indeed, pairs of words in the same equivalence class have unit inner product, while pairs in different classes have 0 inner product. The resulting equivalence classes of words form the *trace monoid* associated with Γ [OTY25].

The operators $(\ell_1 + \ell_1^*, \dots, \ell_n + \ell_n^*)$ have semicircular marginals, and their mixed moments satisfy a generalized notion of independence known as Γ -independence [Mlo04, SW16, CC21, CGvH26]. We refer to [BKN⁺26a, Definitions 1.2 and 1.3] for a more detailed treatment of this example.

Together with the property $\ell_i^* \Omega = 0$, the Γ -CCR uniquely define the joint law of the non-commutative random variables $\ell_1 + \ell_1^*, \dots, \ell_n + \ell_n^*$ with respect to the vacuum state. The diagonal entries Γ_{ii} determine the marginal laws of each $\ell_i + \ell_i^*$, while the off-diagonal entries specify their mixed moments. In particular, $\ell_i + \ell_i^*$ is bounded whenever $\Gamma_{ii} < 1$.

Our goal in this section and the next is to develop tools for bounding the norms of matrix pencils, i.e., operators of the form

$$\mathcal{X} = \sum_{i=1}^n A_i \otimes (\ell_i + \ell_i^*),$$

for the creation operators $\ell_1, \dots, \ell_n$ on the Γ -Fock space, both for general Γ and for the special cases in the examples above.

3.3 General number operator

We next generalize the concept of number operator from Section 2 to the Γ -Fock space.

Definition 3.7. Given $\Gamma \in [-1, 1]_{\text{sym}}^{n \times n}$, define the Γ -number operator to be

$$N_\Gamma := \sum_{i=1}^n \ell_i \ell_i^* \succeq 0,$$

where $\ell_1, \dots, \ell_n$ are the creation operators on the Γ -Fock space from Definition 3.2. For an integer $p \geq 0$, denote by $N_\Gamma|_p$ the restriction of N_Γ to the direct sum of the k -particle spaces for $0 \leq k \leq p$.

For example, in the bosonic setting ($\Gamma \equiv 1$), the number operator scales each basis vector by its particle number. In the free setting ($\Gamma \equiv 0$), the number operator is the projection onto $\Omega^\perp$.

Similarly to Section 2, bounds on the norms of matrix pencils will require a bound on the norm of the associated number operator. We state two simple bounds that will be used below.

Lemma 3.8 (Number of particles bound). *For any $\Gamma \in [-1, 1]_{\text{sym}}^{n \times n}$ and $p \geq 0$, $\|N_\Gamma|_p\| \leq p$.*

Proof. Since N_Γ preserves each particle space, it suffices to consider an eigenvector $x \neq 0$ in the k -particle space for some $k \leq p$, with eigenvalue λ . Choose $y = \ell_{j_1} \cdots \ell_{j_k} \Omega$ maximizing $|\langle x, y \rangle_\Gamma|$ over all $j_1, \dots, j_k \in [n]$ (the maximum is positive, as such vectors span the k -particle space). The effect of $N_\Gamma|_p$ on y is

$$\begin{aligned} N_\Gamma|_p y &= \sum_{i=1}^n \sum_{r: j_r=i} \left(\prod_{s < r} \Gamma_{i, j_s} \right) \ell_i \ell_{j_1} \cdots \ell_{j_{r-1}} \ell_{j_{r+1}} \cdots \ell_{j_k} \Omega \\ &= \sum_{r=1}^k \left(\prod_{s < r} \Gamma_{j_r, j_s} \right) \ell_{j_r} \ell_{j_1} \cdots \ell_{j_{r-1}} \ell_{j_{r+1}} \cdots \ell_{j_k} \Omega, \end{aligned}$$

which is a sum over k vectors of the form $\ell_{i_1} \cdots \ell_{i_k} \Omega$, each scaled by a coefficient of magnitude at most 1. Since $N_\Gamma|_p$ is self-adjoint, this implies

$$|\lambda \langle x, y \rangle_\Gamma| = |\langle N_\Gamma|_p x, y \rangle_\Gamma| = |\langle x, N_\Gamma|_p y \rangle_\Gamma| \leq k |\langle x, y \rangle_\Gamma|.$$

So all eigenvalues of $N_\Gamma|_p$ are at most p ; as $N_\Gamma|_p$ is positive semidefinite, this concludes the proof. $\square$

Lemma 3.9 (Clique bound). *If Γ is a simple graph, then $\|N_\Gamma\| \leq \omega(\Gamma)$, where $\omega(\Gamma)$ denotes the maximum clique size in Γ .*

We note that this does not contradict our previous observation that N_Γ is unbounded in the bosonic case $\Gamma \equiv 1$: in that case we also have $\Gamma_{i,i} = 1$, while here we assume $\Gamma_{i,i} = 0$.

Lemma 3.9 also appears implicitly in the proof of [OTY25, Theorem 1.5]. For completeness, we include a proof below.

Proof. Let $\pi_i := \ell_i \ell_i^*$. Since $\ell_i^* \ell_i = \mathbf{1}$ by the CCR, each π_i is an orthogonal projection. If i, j are adjacent, then by the CCR, one verifies that $(\ell_i \ell_j - \ell_j \ell_i)^* (\ell_i \ell_j - \ell_j \ell_i) = 0$, so ℓ_i and ℓ_j (and ℓ_i and ℓ_j^*) commute, and hence so do π_i and π_j . If i, j are not adjacent in Γ , then $\pi_i \pi_j = \ell_i (\ell_i^* \ell_j) \ell_j^* = 0$, so again π_i and π_j commute. For $S \subseteq [n]$, define

$$\rho_S := \prod_{i \in S} \pi_i \prod_{i \in [n] \setminus S} (\mathbf{1} - \pi_i).$$

We have $\pi_i = \pi_i \prod_{j \neq i} (\pi_j + (\mathbf{1} - \pi_j)) = \sum_{S \subseteq [n]: i \in S} \rho_S$, and therefore $\sum_{i=1}^n \pi_i = \sum_{S \subseteq [n]} |S| \rho_S$. Further, $\rho_S = 0$ unless S is a clique of Γ , so that

$$N_\Gamma = \sum_{i=1}^n \pi_i \preceq \omega(\Gamma) \sum_{S \subseteq [n]} \rho_S = \omega(\Gamma) \mathbf{1},$$

giving the desired bound. $\square$

This proof of Lemma 3.9 only uses the CCR relations abstractly. Alternatively, one may also argue using the trace monoid interpretation in Example 3.6, as follows. On a basis element e_w corresponding to a word w , we have $N_\Gamma e_w = \lambda e_w$, where λ is the number of letters that are “left-removable” from w . In particular, all of the basis elements are eigenvectors of N_Γ and thus the eigenvalues are these values of λ for each word w . But, the set of such left-removable letters from any word forms a clique in Γ , and so $\lambda \leq \omega(\Gamma)$.

4 Norm bounds under general commutation relations

Next, we use the same approach as in Section 2 to derive norm bounds for matrix pencils of operators satisfying the general Γ -CCR, as defined in Section 3. Our main theorems in this section are as follows.

- Theorem 4.2 (Section 4.1) establishes an “intrinsic q -Gaussian” norm bound, akin to an improved NCK inequality with an error term associated to the q -commutation relations.
- Theorem 4.4 (Section 4.2) gives a variational norm bound for intrinsic freeness generalizing the upper bound of Lehner’s formula to operators satisfying general Γ -CCR.
- Theorems 4.5 and 4.6 (Section 4.3) specialize this variational approach to the Gaussian matrix setting, obtaining a norm bound on Gaussian matrices given by a deformation of Lehner’s variational principle and deriving Theorem 1.6 by similar calculations.

4.1 A q -deformed NCK inequality

Theorem 2.1 shows how non-commutativity of coefficients can sharpen eigenvalue estimates of random matrices. Using the same approach, we can also study the behavior of matrices and operators with *partially commutative* structure. We consider the case where this structure is encoded by the q -Gaussian distribution. Let $q \in [0, 1)$ and consider the creation and annihilation operators ℓ and ℓ^* satisfying

$$\ell^* \ell - q \ell \ell^* = 1.$$

The q -Gaussian distribution is the unique distribution γ_q whose moments are given by the vacuum states of powers of $\ell + \ell^*$. They can be characterized combinatorially as

$$\int_{-\infty}^{\infty} t^p d\gamma_q(t) = \langle \Omega, (\ell + \ell^*)^p \Omega \rangle = \sum_{\pi \in P_2([p])} q^{c(\pi)},$$

where $P_2([p])$ is the set of matchings (equivalently, partitions into parts of size exactly 2) of the set $[p]$ and $c(\pi)$ is the number of crossings in such a matching π . For $q \in [0, 1)$, this distribution smoothly interpolates between the semicircular and Gaussian distributions (the limit as $q \rightarrow 1$), is symmetric and compactly supported, and its edge is given by

$$\|\ell + \ell^*\| = \frac{2}{\sqrt{1-q}}.$$

The following theorem extends Theorem 2.1 by comparing a general Γ -CCR system with the corresponding operator model satisfying the q -Gaussian commutation relation. The error is controlled by the extent to which the matrix coefficients fail to transform the Γ -commutation relation into the q -commutation relation.

Proposition 4.1. *Let $A_1, \dots, A_n \in M_d(\mathbb{C})_{\text{sa}}$, $\ell_1, \dots, \ell_n$ be operators satisfying the Γ -CCR relations, and $\mathcal{L} := \sum_i A_i \otimes \ell_i$. Define the partial commutator*

$$C_q := [\Gamma_{i,j} A_j A_i - q A_i A_j]_{i,j=1}^n \in M_{dn}(\mathbb{C}).$$

Then

$$\mathcal{L}^* \mathcal{L} - q \mathcal{L} \mathcal{L}^* \preceq \sum_{i=1}^n A_i^2 \otimes \mathbf{1} + \lambda_{\max}^+(C_q) I_d \otimes N_{\Gamma}.$$

Proof. The proof goes simply by expanding and using the Γ -CCR relations:

$$\begin{aligned}
\mathcal{L}^* \mathcal{L} &= \sum_{i,j=1}^n A_i A_j \otimes \ell_i^* \ell_j \\
&= \sum_{i=1}^n A_i^2 \otimes \mathbf{1} + \sum_{i,j=1}^n \Gamma_{i,j} A_i A_j \otimes \ell_j \ell_i^* \\
&= \sum_{i=1}^n A_i^2 \otimes \mathbf{1} + q \sum_{i,j=1}^n A_i A_j \otimes \ell_i \ell_j^* + \sum_{i,j=1}^n (\Gamma_{i,j} A_j A_i - q A_i A_j) \otimes \ell_i \ell_j^* \\
&= \sum_{i=1}^n A_i^2 \otimes \mathbf{1} + q \mathcal{L} \mathcal{L}^* + [I \otimes \ell_1 \quad \dots \quad I \otimes \ell_n] (C_q \otimes \mathbf{1}) \begin{bmatrix} I \otimes \ell_1^* \\ \vdots \\ I \otimes \ell_n^* \end{bmatrix} \\
&\preceq \sum_{i=1}^n A_i^2 \otimes \mathbf{1} + q \mathcal{L} \mathcal{L}^* + \lambda_{\max}^+(C_q) \cdot I \otimes N_\Gamma. \quad \square
\end{aligned}$$

As before, Proposition 4.1 then leads to the following bounds on moments and norms.

Theorem 4.2 (Intrinsic q -Gaussianity). *Let $A_1, \dots, A_n \in M_d(\mathbb{C})_{\text{sa}}$, $\ell_1, \dots, \ell_n$ be the creation operators on the Γ -Fock space, and $\mathcal{X} = \sum_{i=1}^n A_i \otimes (\ell_i + \ell_i^*)$. Then for all $p \geq 1$ and $q \in [0, 1)$,*

$$(\text{tr} \otimes \tau(\mathcal{X}^{2p}))^{\frac{1}{2p}} \leq \frac{2}{\sqrt{1-q}} \cdot \sqrt{\left\| \sum_{i=1}^n A_i^2 \right\| + \lambda_{\max}^+(C_q) \cdot \|N_\Gamma\|_p}.$$

Further, if $\Gamma_{ii} < 1$ for all i , then we have

$$\|\mathcal{X}\| \leq \frac{2}{\sqrt{1-q}} \cdot \sqrt{\left\| \sum_{i=1}^n A_i^2 \right\| + \lambda_{\max}^+(C_q) \cdot \|N_\Gamma\|}.$$

We note that the factor outside the square root on the right-hand side is precisely the edge of the q -Gaussian distribution, while the first term inside the square root is $\sigma(X)^2$ for $X = \sum_{i=1}^n g_i A_i$ the associated Gaussian series model.

Proof. Let Π_p be the projection onto the space of at most p particles and $\mathcal{L}_p := (I \otimes \Pi_p) \mathcal{L} (I \otimes \Pi_p)$. As in the proof of Corollary 2.4, we have

$$\text{tr} \otimes \tau(\mathcal{X}^{2p}) = \frac{1}{d} \sum_{i=1}^d \|\mathcal{X}^p(e_i \otimes \Omega)\|^2 = \frac{1}{d} \sum_{i=1}^d \|(\mathcal{L}_p + \mathcal{L}_p^*)^p(e_i \otimes \Omega)\|^2 \leq \|\mathcal{L}_p + \mathcal{L}_p^*\|^{2p} \leq 2^{2p} \|\mathcal{L}_p \mathcal{L}_p^*\|^p.$$

Furthermore, by Proposition 4.1,

$$\begin{aligned}
\|\mathcal{L}_p^* \mathcal{L}_p\| &\leq \|\Pi_p \mathcal{L}^* \mathcal{L} \Pi_p\| \leq \left\| (I \otimes \Pi_p) \left(\sum_{i=1}^n A_i^2 \otimes \mathbf{1} + q \mathcal{L} \mathcal{L}^* + \lambda_{\max}^+(C_q) \cdot I \otimes N_\Gamma \right) (I \otimes \Pi_p) \right\| \\
&\leq \left\| \sum_{i=1}^n A_i^2 \right\| + \lambda_{\max}^+(C_q) \|N_\Gamma\|_p + q \|(I \otimes \Pi_p) \mathcal{L} \mathcal{L}^* (I \otimes \Pi_p)\| \\
&= \left\| \sum_{i=1}^n A_i^2 \right\| + \lambda_{\max}^+(C_q) \|N_\Gamma\|_p + q \|\mathcal{L}_p \mathcal{L}_p^*\|.
\end{aligned}$$

Using $\|\mathcal{L}_p \mathcal{L}_p^*\| = \|\mathcal{L}_p^* \mathcal{L}_p\|$ and rearranging the inequality concludes the proof of the first statement.

For the second statement, we use directly $\|\mathcal{X}\| \leq 2\|\mathcal{L}\mathcal{L}^*\|^{\frac{1}{2}}$ and proceed with the exact same argument as in the truncated setting, replacing Π_p by $\mathbf{1}$. $\square$

4.2 Lehner-type variational norm bounds beyond freeness

Next, we return to the question of intrinsic freeness, now for the more general class of operators satisfying the Γ -commutation relations introduced in Section 3. To that end, we introduce a variational refinement of the operator inequalities developed above.

Our model for such a bound is Lehner's variational formula [Leh99] (see (10)) for the norm of pencils of *freely independent* semicircular variables. Here, we give a new proof of just the *upper bound* in (10) which also extends to arbitrary CCR relations.

Proposition 4.3. *In the setting of Proposition 4.1, we have for any $Z \succ 0$,*

$$\mathcal{L} + \mathcal{L}^* \preceq \left(Z^{-1} + \sum_{i=1}^n A_i Z A_i \right) \otimes \mathbf{1} + \lambda_{\max}^+(C_\Gamma(Z)) I \otimes N_\Gamma,$$

where

$$C_\Gamma(Z) := [\Gamma_{i,j} A_j Z A_i]_{i,j=1}^n$$

is the partial commutator scaled by Z .

Applying Proposition 4.3 to $-A_i$ yields a similar lower bound. In the case $\Gamma \equiv 0$, we recover the upper bound in Lehner's formula.

Proof. Let us denote $\mathcal{Z} := Z \otimes \mathbf{1}$. We have

$$0 \preceq \left(\mathcal{Z}^{\frac{1}{2}} \mathcal{L} - \mathcal{Z}^{-\frac{1}{2}} \right)^* \left(\mathcal{Z}^{\frac{1}{2}} \mathcal{L} - \mathcal{Z}^{-\frac{1}{2}} \right) = \mathcal{L}^* \mathcal{Z} \mathcal{L} + \mathcal{Z}^{-1} - \mathcal{L} - \mathcal{L}^*,$$

which implies $\mathcal{L} + \mathcal{L}^* \preceq \mathcal{Z}^{-1} + \mathcal{L}^* \mathcal{Z} \mathcal{L}$. Applying the non-commutative relations,

$$\mathcal{L}^* \mathcal{Z} \mathcal{L} = \sum_{i,j=1}^n A_i Z A_j \otimes \ell_i^* \ell_j = \sum_{i=1}^n A_i Z A_i \otimes \mathbf{1} + \sum_{i,j=1}^n \Gamma_{ij} A_i Z A_j \otimes \ell_j \ell_i^*.$$

It only remains to bound the cross terms using the partial commutator:

$$\begin{aligned} \sum_{i,j=1}^n \Gamma_{ij} A_i Z A_j \otimes \ell_j \ell_i^* &= \sum_{i,j=1}^n \Gamma_{ij} A_j Z A_i \otimes \ell_i \ell_j^* \\ &= [I \otimes \ell_1 \quad \cdots \quad I \otimes \ell_n] (C_\Gamma(Z) \otimes \mathbf{1}) \begin{bmatrix} I \otimes \ell_1^* \\ \vdots \\ I \otimes \ell_n^* \end{bmatrix} \\ &\preceq \lambda_{\max}^+(C_\Gamma(Z)) I \otimes \sum_{i=1}^n \ell_i \ell_i^* = \lambda_{\max}^+(C_\Gamma(Z)) I \otimes N_\Gamma. \end{aligned} \quad \square$$

In the general situation where the number operator is unbounded, we may still apply Proposition 4.3 to obtain bounds on the trace moments using the same projection idea as in the previous sections.

Theorem 4.4 (Variational intrinsic freeness bound). *Let $A_0, \dots, A_n \in M_d(\mathbb{C})_{\text{sa}}$, $\ell_1, \dots, \ell_n$ be the creation operators on the Γ -Fock space, and $\mathcal{X} = A_0 \otimes \mathbf{1} + \sum_{i=1}^n A_i \otimes (\ell_i + \ell_i^*)$. Then for any $p \geq 1$*

$$(\text{tr} \otimes \tau(\mathcal{X}^{2p}))^{\frac{1}{2p}} \leq \max_{\varepsilon \in \{\pm 1\}} \inf_{Z \succ 0} \left\{ \lambda_{\max} \left(\varepsilon A_0 + \sum_{i=1}^n A_i Z A_i + Z^{-1} \right) + \lambda_{\max}^+(C_\Gamma(Z)) \|N_\Gamma|_p\| \right\}.$$

Proof. As in the proof of Theorem 4.2, let Π_p be the projection on the space of at most p particles. Then we have $\text{tr} \otimes \tau(\mathcal{X}^{2p}) \leq \|\mathcal{X}_p\|^{2p}$, where $\mathcal{X}_p = A_0 \otimes \Pi_p + \mathcal{L}_p + \mathcal{L}_p^*$ and $\mathcal{L}_p = (I \otimes \Pi_p) \mathcal{L} (I \otimes \Pi_p)$. For each $\varepsilon \in \{\pm 1\}$, applying Proposition 4.3 to εA_i and conjugating by $I \otimes \Pi_p$ gives

$$\varepsilon \mathcal{X}_p \preceq \left(\varepsilon A_0 + Z^{-1} + \sum_{i=1}^n A_i Z A_i \right) \otimes \Pi_p + \lambda_{\max}^+(C_\Gamma(Z)) I \otimes N_\Gamma|_p.$$

Taking the infimum over Z and using $\|\mathcal{X}_p\| = \max_{\varepsilon \in \{\pm 1\}} \lambda_{\max}(\varepsilon \mathcal{X}_p)$ concludes the proof. $\square$

Proof of Theorem 1.7. The proof is identical to the previous one, except that no truncation is needed. Applying Proposition 4.3 to εA_i for $\varepsilon \in \{\pm 1\}$ directly gives

$$\varepsilon \mathcal{X} \preceq \left(\varepsilon A_0 + Z^{-1} + \sum_{i=1}^n A_i Z A_i \right) \otimes \mathbf{1} + \lambda_{\max}^+(C_\Gamma(Z)) I \otimes N_\Gamma,$$

and we conclude in the exact same way. $\square$

Picking Z to be a scalar multiple of the identity in Theorem 1.7 recovers a bound of the form of Corollary 2.4. More precisely, when $A_0 = 0$ and $Z = cI$ for an optimized choice of $c > 0$ and N_Γ is bounded, we get

$$\|\mathcal{X}\| \leq 2 \sqrt{\left\| \sum_{i=1}^n A_i^2 \right\| + \lambda_{\max}^+(C_\Gamma(I)) \|N_\Gamma\|}.$$

For example, when Γ is a simple graph and $\mathcal{X}$ is a matrix pencil of Γ -independent operators, we obtain an NCK-like inequality with an error term depending on the clique number $\omega(\Gamma)$.

4.3 Specialization to Gaussian random matrices

In this section, we specialize the variational bound of Theorem 4.4 to the case of Gaussian random matrices,

$$X = A_0 + \sum_{i=1}^n g_i A_i.$$

In that case, the partial commutator takes the form

$$C_X(Z) := [A_j Z A_i]_{i,j=1}^n.$$

Combining the weighted moment bound with the trace method gives the following deformation of Lehner's variational formula involving $C_X(Z)$.

Theorem 4.5 (Deformed Lehner bound). *Let X and $C_X(Z)$ be as above. Then*

$$\mathbb{E}\|X\| \lesssim \max_{\varepsilon \in \{\pm 1\}} \inf_{Z \succ 0} \left\{ \lambda_{\max} \left(\varepsilon A_0 + Z^{-1} + \sum_{i=1}^n A_i Z A_i \right) + \log(2d) \lambda_{\max}(C_X(Z)) \right\}.$$

Comparing with Lehner's formula (10) for $\|\mathcal{X}_{\text{free}}\|$, we see that this is precisely Lehner's formula, but with an additional term in the objective function.

Proof. By Jensen's inequality, Theorem 4.4, and the number of particles bound (Lemma 3.8), we have, for every $p \geq 1$,

$$\begin{aligned} \mathbb{E}\|X\| &\leq d^{\frac{1}{2p}} \left(\mathbb{E} \operatorname{tr} X^{2p} \right)^{\frac{1}{2p}} \\ &\leq d^{\frac{1}{2p}} \max_{\varepsilon \in \{\pm 1\}} \inf_{Z \succ 0} \left\{ \lambda_{\max} \left(\varepsilon A_0 + Z^{-1} + \sum_i A_i Z A_i \right) + p \lambda_{\max}(C_X(Z)) \right\}. \end{aligned}$$

For $p = \lfloor \log(2d) \rfloor$, we have $d^{\frac{1}{2p}} \leq e$. Since $\operatorname{Tr} C_X(Z) = \sum_i \operatorname{Tr}(A_i Z A_i) \geq 0$, the maximal eigenvalue of $C_X(Z)$ is nonnegative. Replacing p by $\log(2d)$ concludes the proof. $\square$

Theorem 4.4 can also be used to prove an upper bound for the norm of Gaussian random matrices akin to our strengthened NCK inequality in Theorem 1.5, but involving $\|\mathcal{X}_{\text{free}}\|$ instead of $\sigma(X)$. We first prove the trace moment version; we will deduce Theorem 1.6 as a corollary. Recall that we define $v'(X) := \min\{\sigma(X), v(X)\}$.

Theorem 4.6. *Let $X = A_0 + \sum_i g_i A_i$ be a self-adjoint Gaussian random matrix. Define*

$$\mathcal{X}_{\text{free}} := A_0 \otimes \mathbf{1} + \sum_{i=1}^n A_i \otimes s_i,$$

where $s_1, \dots, s_n$ are freely independent semicircular elements. Then for any $p \geq 1$,

$$\left(\mathbb{E} \operatorname{tr} X^{2p} \right)^{\frac{1}{2p}} \leq \|\mathcal{X}_{\text{free}}\| + 2v'(X)\sqrt{p}.$$

To prove Theorem 4.6, we use the following bound for the norm of the partial commutator as a function of the v' parameter.

Lemma 4.7. *For any self-adjoint Gaussian random matrix $X = \sum_{i=1}^n g_i A_i$ and $Z \succ 0$, we have $\|C_X(Z)\| \leq v'(X)^2 \|Z\|$.*

Proof. We prove the upper bounds with $\sigma(X)$ and $v(X)$ separately.

Define $A : \mathbb{C}^d \rightarrow \mathbb{C}^n \otimes \mathbb{C}^d$ by $Ax = \sum_{i=1}^n e_i \otimes A_i x$. Then

$$\|A\|^2 = \|A^* A\| = \left\| \sum_{i=1}^n A_i^2 \right\| = \sigma(X)^2.$$

Let S_{12} swap the two factors of $\mathbb{C}^n \otimes \mathbb{C}^n$. We have the factorization

$$C_X(Z) = (I_n \otimes A^*)(S_{12} \otimes I_d)(I_n \otimes Z)(I_n \otimes A).$$

Indeed, the right-hand side sends $e_j \otimes x$ to $\sum_i e_i \otimes A_j Z A_i x$. Since S_{12} is unitary, taking norms gives $\|C_X(Z)\| \leq \sigma(X)^2 \|Z\|$.

For the other bound, define $V : \mathbb{C}^n \rightarrow \mathbb{C}^d \otimes \mathbb{C}^d$ by $V e_i = \operatorname{vec}(A_i)$, using the convention $\operatorname{vec}(A_i) = \sum_{a,b} (A_i)_{ab} e_a \otimes e_b$. By (2),

$$\|V\| = v(X).$$

Let F_{13} swap the first and third factors of $(\mathbb{C}^d)^{\otimes 3}$. We claim that

$$C_X(Z) = (V^* \otimes I_d) F_{13} (I_d \otimes Z^\top \otimes I_d) (V \otimes I_d), \quad (25)$$

where the transpose is taken in the basis used for vectorization. Applying the right-hand side to $e_j \otimes e_b$, the coefficient of $e_i \otimes e_a$ is

$$\sum_{u,c} (A_j)_{au} Z_{uc} \overline{(A_i)_{bc}} = \sum_{u,c} (A_j)_{au} Z_{uc} (A_i)_{cb} = (A_j Z A_i)_{ab},$$

where we used the self-adjointness of A_i . This proves (25). Since F_{13} is unitary and $\|Z^\top\| = \|Z\|$, taking norms gives $\|C_X(Z)\| \leq v(X)^2 \|Z\|$. Combining the two bounds gives the result. $\square$

Proof of Theorem 4.6. We first apply Theorem 4.4 with $\Gamma \equiv 1$:

$$\begin{aligned} (\mathbb{E} \operatorname{tr} X^{2p})^{\frac{1}{2p}} &\leq \max_{\varepsilon \in \{\pm 1\}} \inf_{Z \succ 0} \left\{ \lambda_{\max} \left(\varepsilon A_0 + \sum_{i=1}^n A_i Z A_i + Z^{-1} \right) + \|C_X(Z)\| \|N_\Gamma|_p\| \right\} \\ &\leq \max_{\varepsilon \in \{\pm 1\}} \inf_{Z \succ 0} \left\{ \lambda_{\max} \left(\varepsilon A_0 + \sum_{i=1}^n A_i Z A_i + Z^{-1} \right) + p v'(X)^2 \|Z\| \right\}, \end{aligned} \quad (26)$$

where we used Lemmas 3.8 and 4.7 in the second inequality.

Given any $Z \succ 0$, consider $Z_t = (tI + Z^{-1})^{-1}$ for some parameter $t > 0$ to be fixed later. Then we have $\|Z_t\| \leq 1/t$ and $Z_t \preceq Z$, so

$$\sum_{i=1}^n A_i Z_t A_i + Z_t^{-1} \preceq \sum_{i=1}^n A_i Z A_i + Z^{-1} + tI.$$

Hence, any candidate $Z \succ 0$ to Lehner's formula (10) can be mapped to a candidate $Z_t \succ 0$ of (26) achieving objective value at most

$$\lambda_{\max} \left(\varepsilon A_0 + \sum_{i=1}^n A_i Z A_i + Z^{-1} \right) + t + \frac{p v'(X)^2}{t}.$$

Optimizing the right-hand side over t and taking the infimum over Z yields the trace bound. $\square$

Proof of Theorem 1.6. By Jensen's inequality and Theorem 4.6, for every integer $p \geq 1$,

$$\mathbb{E} \|X\| \leq d^{\frac{1}{2p}} (\mathbb{E} \operatorname{tr} X^{2p})^{\frac{1}{2p}} \leq d^{\frac{1}{2p}} (\|\mathcal{X}_{\text{free}}\| + 2v'(X)\sqrt{p}).$$

Applying Lemma A.1 with $a = \|\mathcal{X}_{\text{free}}\|$ and $b = v'(X)$, and using $\log(2d) \leq 2 \log d$, gives

$$\begin{aligned} \mathbb{E} \|X\| &\leq \|\mathcal{X}_{\text{free}}\| + 8 \|\mathcal{X}_{\text{free}}\|^{1/3} v'(X)^{2/3} (\log d)^{1/3} + 8 v'(X) (\log d)^{1/2} \\ &\leq \|\mathcal{X}_{\text{free}}\| \left(1 + 8 \left(\frac{v(X)^2}{\sigma(X)^2} \log d \right)^{1/2} + 8 \left(\frac{v(X)^2}{\sigma(X)^2} \log d \right)^{1/3} \right), \end{aligned}$$

where the last step uses $\sigma(X) \leq \|\mathcal{X}_{\text{free}}\|$ and $v'(X) \leq v(X)$. $\square$

5 Applications of norm bounds

5.1 Gaussian Cayley matrices

Given a finite group G on d elements, let $\{\rho_x\}_{x \in G}$ denote the left regular representation of $x \in G$ (i.e., ρ_x is the $d \times d$ permutation matrix corresponding to left multiplication by x). Consider the asymmetric Gaussian random matrix $Y^{(0)} = \sum_{x \in G} g_x \rho_x$, where g_x are i.i.d. drawn from $\mathcal{N}(0, 1)$, and define its Hermitian dilation

$$X^{(0)} = \begin{bmatrix} 0 & Y^{(0)} \\ Y^{(0)\top} & 0 \end{bmatrix}.$$

Note that $\|Y^{(0)}\| = \|X^{(0)}\|$, so we may consider these interchangeably, but we focus on using our tools on $X^{(0)}$ since it is Hermitian.

As we will see, it is helpful to apply a simple reduction before applying matrix concentration inequalities. The all-ones vector χ of suitable dimension is an eigenvector of both $X^{(0)}$ and $Y^{(0)}$ with eigenvalue $\lambda = \sum_{x \in G} g_x$, whose law is $\mathcal{N}(0, |G|)$. Thus, we may further define modified versions of these matrices where we subtract off this component,

$$Y := Y^{(0)} - \frac{\lambda}{|G|} \chi \chi^\top,$$

$$X := \begin{bmatrix} 0 & Y \\ Y^\top & 0 \end{bmatrix}.$$

Then X is again the Hermitian dilation of Y and we have $\|X\| = \|Y\|$, and further for the original matrices we have

$$\|Y^{(0)}\| = \max\{\|X\|, |\lambda|\},$$

where the second term is with high probability $\tilde{O}(|G|^{1/2})$ and will turn out to be essentially negligible in our calculations. We thus focus on bounding $\|X\|$ accurately.

While the spectral norm of X is known to depend heavily on the group structure and the dimensions of its irreducible representations [BKMZ24], the parameters $\sigma(X)$ and $v(X)$ depend only on $d = |G|$: straightforward calculations give that $\sigma(X) = \sqrt{d}$ and $v(X) = \sqrt{2d}$. Thus the original NCK inequality (Theorem 1.1) gives the estimate

$$\mathbb{E}\|X\| \lesssim \sqrt{d \log d} \tag{27}$$

while the improved NCK inequality of [BBvH23] (Theorem 1.2) gives the same with an inferior logarithmic factor. The estimate in (27) is tight for some groups, for instance, for the cyclic group $G = \mathbb{Z}/d$. However, for other groups we have $\mathbb{E}\|X\| \asymp \sqrt{d}$. In fact, this was precisely the motivation for the study of these matrices in [BKMZ24]: it is one where the tools of [BBvH23] fail, while the spectrum can be understood with representation theory (see [BKMZ24] for details). We show here that our parameter $\kappa(X)$ can witness the representation-theoretic properties that determine the spectrum of X .

Proposition 5.1. *In the above setting, let $r_{\min}(G)$ be the smallest dimension of a nontrivial complex irreducible representation of G . Then,*

$$\kappa(X) = \sqrt{\frac{d}{r_{\min}(G)}} = \frac{\sigma(X)}{\sqrt{r_{\min}(G)}}.$$

Proof. Set $P = d^{-1}\chi\chi^\top$. We have $X = \sum_{x \in G} g_x A_x$ for coefficient matrices

$$A_x = \begin{bmatrix} 0 & \rho_x - P \\ \rho_x^\top - P & 0 \end{bmatrix}.$$

Since $\rho_x P = P \rho_x = P$, permuting the tensor factors identifies $C(X) = [A_y A_x]_{x,y \in G}$ with the direct sum $T_+ \oplus T_-$, where

$$T_+ = [\rho_{yx^{-1}} - P]_{x,y \in G}, \quad T_- = [\rho_{y^{-1}x} - P]_{x,y \in G}.$$

The permutation $x \mapsto x^{-1}$ of the block indices swaps T_+ with T_- . Thus these two have the same largest eigenvalue and $\kappa(X)^2 = \lambda_{\max}(T_+)$.

Decomposing the regular representation into irreducible representations over $\mathbb{C}$ shows that, after conjugation by a suitable unitary matrix, T_+ is a direct sum of a zero block and the matrices

$$T_\pi = [\pi(y)\pi(x)^*]_{x,y \in G},$$

each repeated $\dim \pi$ times, where π ranges over the nontrivial irreducible unitary representations of G (see, e.g., [BKMZ24, Lemma 7] for a similar decomposition). We show that $\lambda_{\max}(T_\pi) = d / \dim \pi$.

Fix such a representation π of dimension r , and let $V_\pi \in M_{r^2, d}(\mathbb{C})$ have columns $\text{vec}(\pi(x))$, indexed by $x \in G$. The Peter-Weyl decomposition gives $V_\pi V_\pi^* = (d/r)I_{r^2}$. Also, by Schur's lemma,

$$(T_\pi^2)_{x,z} = \sum_{y \in G} \pi(y)\pi(x)^*\pi(z)\pi(y)^* = \frac{d}{r} \text{Tr}(\pi(x)^*\pi(z))I_r.$$

Consequently,

$$T_\pi^2 = \frac{d}{r}(V_\pi^* V_\pi) \otimes I_r.$$

Every nonzero eigenvalue of $V_\pi^* V_\pi$ equals d/r , so every nonzero eigenvalue of T_π equals d/r or $-d/r$. Since $\text{Tr}(T_\pi) = dr > 0$, we have $\lambda_{\max}(T_\pi) = d/r$. Taking the maximum over the nontrivial irreducible representations proves the claim. $\square$

In particular, over a sequence of groups $G = G_n$ such that $r_{\min}(G_n) \gg \log |G_n|$, giving rise to a sequence $X = X_n$, our Theorem 1.5 together with Proposition 5.1 establishes $\mathbb{E}\|X\| \asymp \sigma(X) = \sqrt{|G|}$, showing that the logarithmic factor incurred by the other NCK-type inequalities is spurious in this case. Such groups are called (sometimes subject to other quantitative assumptions about how $r_{\min}(G)$ depends on $|G|$) *quasirandom groups* in the literature; see [BKMZ24] for further discussion on this connection.

5.2 NCK-type inequalities for Γ -independent systems

Despite the remarkable success of the theory of intrinsic freeness, there are important cases of Gaussian random matrices for which $\mathcal{X}_{\text{free}}$ fails to describe the spectrum and the norm — for example, for the so-called *Kikuchi matrices*, which arise in the analysis of algorithms for tensor PCA [WEM19, BCSvH26, KX26]. The parallel work [BKN⁺26a, Section 1.2] suggests that the right operator proxy for these matrices is not $\mathcal{X}_{\text{free}}$, but rather $\mathcal{X}_\Gamma := \sum_i A_i \otimes s_i$ for an appropriate graph Γ , where the s_i are Γ -independent semicircles (Example 3.6).

This example motivates the development of a theory for “intrinsic Γ -independence” that would generalize the notion of intrinsic freeness. A natural first step for this program is to understand

the norm of the pencils $\mathcal{X}_\Gamma$, just as Lehner's formula (and the coarser bound $\|\mathcal{X}_{\text{free}}\| \leq 2\sigma(X)$) has allowed the development of intrinsic freeness.

Such operator-valued non-commutative Khintchine inequalities depending on combinatorial properties of the graph Γ have independently emerged in recent years. To compare these results, let us fix some notation for the rest of this section. Let $A_1, \dots, A_n \in M_d(\mathbb{C})_{\text{sa}}$ and set

$$\sigma^2 := \left\| \sum_{i=1}^n A_i^2 \right\|.$$

Let $s_1, \dots, s_n$ be Γ -independent semicircles (i.e., $s_i = \ell_i + \ell_i^*$, where ℓ_i are solutions to the Γ -CCR for a simple unweighted graph Γ). Let $\mathcal{X} := \sum_i A_i \otimes s_i$. Finally, we denote by $\lambda_1(\Gamma) \geq \dots \geq \lambda_n(\Gamma)$ the eigenvalues of Γ and $\omega(\Gamma)$ its maximum clique size.

In this notation, Collins and Miyagawa show that $\|\mathcal{X}\| \leq 2\sigma\sqrt{\lambda_1(\Gamma)+1}$ [CM26]; and Oliveira Santos, Tripathi, and Youssef show that $\|\mathcal{X}\| \leq 2\sigma\sqrt{\omega(\Gamma)}$ [OTY25]. The latter bound is stronger, since $\omega(\Gamma) \leq \lambda_1(\Gamma) + 1$. Meanwhile, combining Theorem 4.2 with Lemma 3.9 directly gives:

Corollary 5.2. *Let $C_\Gamma := [\Gamma_{ij}A_jA_i]_{1 \leq i,j \leq n}$ be the partial commutator of $A_1, \dots, A_n$. Then*

$$\|\mathcal{X}\| \leq 2\sqrt{\sigma^2 + \lambda_{\max}^+(C_\Gamma)\omega(\Gamma)}.$$

Since $\lambda_{\max}^+(C_\Gamma) \leq \sigma^2$ (by the same argument as in Proposition 1.3), the bound in Corollary 5.2 is always at least as good, up to a universal constant factor, as the previous estimates. Further, whenever $\lambda_{\max}^+(C_\Gamma) \ll \sigma^2/\omega(\Gamma)$, our bound is $(2 + o(1))\sigma$, which substantially improves over these.

Another estimate that the preceding inequalities fail to capture is the following, which plays a key role in proving that independent Kikuchi matrices converge to a q -Gaussian system [BKN⁺26a]. It follows directly from Theorem 4.2, applied with scalar coefficients $1/\sqrt{n}$.

Corollary 5.3. *We have*

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n s_i \right\| \leq \inf_{q \in [0,1]} \frac{2}{\sqrt{1-q}} \sqrt{1 + \frac{\lambda_{\max}^+(\Gamma - q\chi\chi^\top)\omega(\Gamma)}{n}}.$$

The $q = 0$ case of this inequality also appears in [CM26, Corollary 3.5].

5.3 Ground state energy of the SYK model

The Sachdev-Ye-Kitaev (SYK) model is a model from quantum physics that has attracted significant attention; we refer the reader to [Plu19, BKM26] for background. Let $0 \leq k \leq n$ be even integers. The problem of finding the ground state energy of the SYK model reduces to computing the largest eigenvalue of the random matrix

$$H_{n,k} = \binom{n}{k}^{-\frac{1}{2}} \sum_{S \in \binom{[n]}{k}} g_S \psi_S,$$

where the g_S 's are independent standard Gaussians, and $\psi_S \in M_{2^{n/2}}(\mathbb{C})_{\text{sa}}$ satisfy the relations

$$\psi_S^2 = I, \quad \psi_S \psi_T = (-1)^{|S \cap T|} \psi_T \psi_S. \quad (28)$$

As $n \rightarrow \infty$ for $k = k(n)$ with $k^2/n \rightarrow \alpha \in [0, \infty)$, the spectral distribution of $H_{n,k}$ converges to the $e^{-2\alpha}$ -Gaussian distribution [FTW19]. The limiting spectral distribution is Gaussian if $k \ll \sqrt{n}$, semicircular if $k \gg \sqrt{n}$, and q -Gaussian in the intermediate regime. Very recently, Basu, Kothari and Midha [BKM26] proved sharp asymptotics for the norm of $H_{n,k}$ in the regime $1 \ll k \ll \sqrt{n}$. As a corollary of Theorem 4.2, we obtain a new proof of their upper bound.

Corollary 5.4. *For every even $2 \leq k < \frac{\sqrt{n}}{4}$, we have*

$$\mathbb{E} \|H_{n,k}\| \leq \frac{\sqrt{2n}}{k} + O(1).$$

Proof. Think of $H_{n,k}$ as a pencil in Gaussian variables, with the matrices ψ_S playing the role of the coefficients. By Theorem 4.2 with $A_S = \binom{n}{k}^{-1/2} \psi_S$, we have, for every even integer p and $q \in [0, 1)$,

$$\left(\mathbb{E} \operatorname{tr} H_{n,k}^p \right)^{\frac{1}{p}} \leq \frac{2}{\sqrt{1-q}} \sqrt{1 + \frac{p \lambda_{\max}^+(C_q)}{\binom{n}{k}}}, \quad \text{where } C_q := \left[(\psi_T \psi_S - q \psi_S \psi_T)_{S,T \in \binom{[n]}{k}} \right]. \quad (29)$$

Let $\Gamma_{n,k}$ be the $\binom{n}{k} \times \binom{n}{k}$ -dimensional matrix with entries $\Gamma_{n,k}[S, T] := (-1)^{|S \cap T|}$. The commutation relations (28) give the decomposition

$$\begin{aligned} C_q &= \operatorname{diag}(\{\psi_S\}_S) \left((\Gamma_{n,k} - q \chi \chi^\top) \otimes I_{2^{n/2}} \right) \operatorname{diag}(\{\psi_S\}_S) \\ &= \operatorname{diag}(\{\psi_S\}_S) \left((\Gamma_{n,k} - q \chi \chi^\top) \otimes I_{2^{n/2}} \right) \operatorname{diag}(\{\psi_S\}_S)^{-1}, \end{aligned}$$

where the second equality follows from $\psi_S = \psi_S^{-1}$ (from (28)). Therefore, C_q has the same spectrum as the inner matrix, and $\lambda_{\max}(C_q) = \lambda_{\max}(\Gamma_{n,k} - q \chi \chi^\top)$. The remainder of the proof is a simple estimate on the top two eigenvalues of the matrix $\Gamma_{n,k}$.

Lemma 5.5. *The top eigenvalue of $\Gamma_{n,k}$ is*

$$\binom{n}{k} \cdot q_{n,k}, \quad \text{with } q_{n,k} = 1 - \frac{2k^2}{n} + O\left(\frac{k^4}{n^2}\right),$$

with eigenvector χ , and

$$\lambda_{\max} \left(\Gamma_{n,k} - q_{n,k} \chi \chi^\top \right) \lesssim \frac{k^2}{n^2} \binom{n}{k}.$$

See [BKM26, Lemma 5.1] for a proof of Lemma 5.5. Plugging this choice of $q_{n,k}$ into (29), we obtain

$$\left(\mathbb{E} \operatorname{tr} H_{n,k}^p \right)^{\frac{1}{p}} \leq \frac{\sqrt{2n}}{k} \left(1 + O\left(\frac{k^2}{n}\right) \right) \sqrt{1 + O\left(\frac{pk^2}{n^2}\right)}.$$

Picking $p = \frac{n^{3/2}}{k}$ and noting that $H_{n,k}$ has dimension $2^{n/2}$,

$$\mathbb{E} \|H_{n,k}\| \leq 2^{\frac{n}{2p}} \left(\mathbb{E} \operatorname{tr} H_{n,k}^p \right)^{\frac{1}{p}} \leq \frac{\sqrt{2n}}{k} + O(1). \quad \square$$

Remark 5.6 (Comparison with [BKM26]). In summary, our proof of Corollary 5.4 proceeds by representing $H_{n,k}$ as the matrix pencil

$$\binom{n}{k}^{-\frac{1}{2}} \sum_S \psi_S \otimes (\ell_S + \ell_S^*), \quad (30)$$

where ℓ_S are the bosonic creation operators, and then applying our non-commutative Khintchine inequality in the bosonic setting, with the matrices ψ_S as coefficients. Instead, [BKM26] first shows that the operator (30) has the same distribution as the operator

$$\binom{n}{k}^{-\frac{1}{2}} \sum_S (a_S + a_S^*), \quad (31)$$

where a_S are the creation operators for the $\Gamma_{n,k}$ -CCR by comparing their moments (where $\Gamma_{n,k}$ is defined in the above proof). Then, they analyze the trace moments of (31) using techniques closely related to those in our proof of Theorem 4.2. We note that applying Theorem 4.2 directly to (31) also recovers Corollary 5.4 without any combinatorial analysis of the SYK moments.

References

- [Ban26] Afonso S. Bandeira. Random matrices, intrinsic freeness, and sharp non-asymptotic inequalities. In *Proceedings of the International Congress of Mathematicians (ICM)*, pages 289–309, 2026.
- [BBvH23] Afonso S. Bandeira, March T. Boedihardjo, and Ramon van Handel. Matrix concentration inequalities and free probability. *Inventiones Mathematicae*, 234(1):419–487, 2023.
- [BCSvH26] Afonso S. Bandeira, Giorgio Cipolloni, Dominik Schröder, and Ramon van Handel. Matrix concentration inequalities and free probability II. Two-sided bounds and applications. *Communications of the American Mathematical Society*, 6:896–946, 2026.
- [BKM26] Arpon Basu, Pravesh K. Kothari, and Siddhant Midha. Sharp bounds on ground state energy of the SYK model. *arXiv:2607.27185*, 2026.
- [BKMR25] Afonso S. Bandeira, Anastasia Kireeva, Antoine Maillard, and Almut Rödger. Randomstrasse101: Open problems of 2024. *arXiv:2504.20539*, 2025.
- [BKMZ24] Afonso S. Bandeira, Dmitriy Kunisky, Dustin G. Mixon, and Xinmeng Zeng. On the concentration of Gaussian Cayley matrices. *Applied and Computational Harmonic Analysis*, 73:101694, 2024.
- [BKN⁺26a] Afonso S. Bandeira, Dmitriy Kunisky, Petar Nizić-Nikolac, Lucas Pesenti, and Robert Wang. Convergence of Kikuchi matrices to Γ -independent and q -Gaussian limits. *arXiv*, 2026.
- [BKN⁺26b] Afonso S. Bandeira, Dmitriy Kunisky, Petar Nizić-Nikolac, Lucas Pesenti, and Robert Wang. The hypergraph Moore bound. *arXiv:2607.14068v2*, 2026.
- [BKS97] Marek Bożejko, Burkhard Kümmerer, and Roland Speicher. q -Gaussian processes: Non-commutative and classical aspects. *Communications in Mathematical Physics*, 185:129–154, 1997.
- [BS91] Marek Bożejko and Roland Speicher. An example of a generalized Brownian motion. *Communications in Mathematical Physics*, 137(3):519–531, 1991.

- [BS92] Marek Bożejko and Roland Speicher. An example of a generalized Brownian motion II. In *Quantum Probability and Related Topics*, volume 7, pages 67–77. World Scientific, 1992.
- [BS94] Marek Bożejko and Roland Speicher. Completely positive maps on Coxeter groups, deformed commutation relations, and operator spaces. *Mathematische Annalen*, 300:97–120, 1994.
- [BvH24] Tatiana Brailovskaya and Ramon van Handel. Universality and sharp matrix concentration inequalities. *Geometric and Functional Analysis*, 34(6):1734–1838, 2024.
- [CC21] Ian Charlesworth and Benoît Collins. Matrix models for ε -free independence. *Archiv der Mathematik*, 116:585–600, 2021.
- [CC26] Martijn Caspers and Enli Chen. Strongly convergent matrix models for q -Gaussian algebras. *arXiv:2607.15100*, 2026.
- [CGvH26] Chi-Fang Chen, Jorge Garza-Vargas, and Ramon van Handel. A new approach to strong convergence II. The classical ensembles. *Geometric and Functional Analysis*, 36:831–890, 2026.
- [CM26] Benoît Collins and Akihiro Miyagawa. Operator-valued Khintchine inequality for ε -free semicircles. *Bulletin of the London Mathematical Society*, 58(1):e70229, 2026.
- [Coo53] Joseph M. Cook. The mathematics of second quantization. *Transactions of the American Mathematical Society*, 74(2):222–245, 1953.
- [Cun77] Joachim Cuntz. Simple C^* -algebras generated by isometries. *Communications in Mathematical Physics*, 57(2):173–185, 1977.
- [DG13] Jan Dereziński and Christian Gérard. *Mathematics of Quantization and Quantum Fields*. Cambridge University Press, 2013.
- [FB70] Uriel Frisch and Richard Bourret. Parastochastics. *Journal of Mathematical Physics*, 11:364–390, 1970.
- [Fol08] Gerald B. Folland. *Quantum Field Theory: A Tourist Guide for Mathematicians*. AMS, 2008.
- [FTW19] Renjie Feng, Gang Tian, and Dongyi Wei. Spectrum of SYK model. *Peking Mathematical Journal*, 2:41–70, 2019.
- [Hei26] Diar Heidary. SparseStack is an optimal oblivious subspace embedding. *arXiv:2609.02978*, 2026.
- [Itô51] Kiyosi Itô. Multiple Wiener integral. *Journal of the Mathematical Society of Japan*, 3(1):157–169, 1951.
- [Khi23] Aleksandr Khintchine. Über dyadische Brüche. *Mathematische Zeitschrift*, 18:109–116, 1923.
- [Kun26] Dmitriy Kunisky. Lehner’s operator norm formulas, semidefinite programming, and spiked matrix models. *arXiv:2606.14687*, 2026.

- [KX26] Pravesh K. Kothari and Jeff Xu. Smooth trade-off for tensor PCA via sharp bounds for Kikuchi matrices. In *Symposium on Discrete Algorithms (SODA)*, pages 2617–2632, 2026.
- [Leh99] Franz Lehner. Computing norms of free operators with matrix coefficients. *American Journal of Mathematics*, 121(3):453–486, 1999.
- [LP91] Françoise Lust-Piquard and Gilles Pisier. Non commutative Khintchine and Paley inequalities. *Arkiv för Matematik*, 29:241–260, 1991.
- [Lus86] Françoise Lust-Piquard. Inégalités de Khintchine dans C_p ($1 < p < \infty$). *Comptes Rendus de l’Académie des Sciences de Paris, Série I*, 303(7):289–292, 1986.
- [Mho04] Wojciech Młotkowski. Λ -free probability. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 7(1):27–41, 2004.
- [NS06] Alexandru Nica and Roland Speicher. *Lectures on the combinatorics of free probability*. Cambridge University Press, 2006.
- [OTY25] Patrick Oliveira Santos, Raghavendra Tripathi, and Pierre Youssef. Khintchine inequalities, trace monoids and Turán-type problems. *arXiv:2506.02517*, 2025.
- [Plu19] Miguel Angel Pluma Rodriguez. *The SYK model and matrices with q -Gaussian entries*. PhD thesis, Universität des Saarlandes, 2019.
- [RS09] Igor Rodnianski and Benjamin Schlein. Quantum fluctuations and rate of convergence towards mean field dynamics. *Communications in Mathematical Physics*, 291(1):31–61, 2009.
- [Seg56] Irving E. Segal. Tensor algebras over Hilbert spaces. I. *Transactions of the American Mathematical Society*, 81(1):106–134, 1956.
- [SH26] Alexander Schmidhuber and Matthew B. Hastings. A spectral proof of the hypergraph Moore bound. *arXiv:2607.26028*, 2026.
- [Spe92] Roland Speicher. A non-commutative central limit theorem. *Mathematische Zeitschrift*, 209:55–66, 1992.
- [SW16] Roland Speicher and Janusz Wysoczański. Mixtures of classical and free independence. *Archiv der Mathematik*, 107:445–453, 2016.
- [Tro15] Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends in Machine Learning*, 8(1-2):1–230, 2015.
- [Tro18] Joel A. Tropp. Second-order matrix concentration inequalities. *Applied and Computational Harmonic Analysis*, 44(3):700–736, 2018.
- [Tro26] Joel A. Tropp. Universality laws for random matrices via exchangeable counterparts. *arXiv:2603.05803*, 2026.
- [vH17] Ramon van Handel. Structured random matrices. *Convexity and Concentration*, pages 107–156, 2017.

- [Voi85] Dan Voiculescu. Symmetries of some reduced free product C^* -algebras. In *Operator Algebras and Their Connections with Topology and Ergodic Theory*, volume 1132, pages 556–588. Springer, 1985.
- [Voi91] Dan Voiculescu. Limit laws for random matrices and free products. *Inventiones mathematicae*, 104(1):201–220, 1991.
- [Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.
- [WEM19] Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi hierarchy and tensor PCA. In *Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468, 2019.

A Scalar optimization estimates

We include the elementary estimates used to pass from trace moments to the operator norm. The constants are not optimized.

Lemma A.1. *For every $d \geq 2$ and $a, b \geq 0$,*

$$\inf_{p \geq 1} d^{\frac{1}{2p}}(a + 2b\sqrt{p}) \leq a + 5 \left[(ab^2 \log(2d))^{\frac{1}{3}} + b \log^{\frac{1}{2}}(2d) \right].$$

Proof. Set $\ell = \log(2d) > 1$. If $b = 0$, take $p \rightarrow \infty$; otherwise, set $p := \left\lceil \ell + \left(\frac{a\ell}{b}\right)^{\frac{2}{3}} \right\rceil$. As $p \geq \ell$, the bound $e^u \leq 1 + 2u$ for $0 \leq u \leq \frac{1}{2}$ gives $d^{\frac{1}{2p}} \leq e^{\frac{\ell}{2p}} \leq 1 + \frac{\ell}{p}$. Therefore,

$$\begin{aligned} d^{\frac{1}{2p}}(a + 2b\sqrt{p}) &\leq a + \frac{a\ell}{p} + 2b\sqrt{p} + \frac{2b\ell}{\sqrt{p}} \\ &\leq a + (a\ell)^{\frac{1}{3}}b^{\frac{2}{3}} + 2b\sqrt{2\ell + 2\left(\frac{a\ell}{b}\right)^{\frac{2}{3}}} + 2b\sqrt{\ell}. \end{aligned}$$

Using $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ and bounding $2 + 2\sqrt{2} < 5$ concludes the proof. $\square$

Lemma A.2. *For every $d \geq 2$ and $0 \leq \kappa \leq \sigma$,*

$$\inf_{p \geq 1} d^{\frac{1}{2p}} \sqrt{\sigma^2 + p\kappa^2} \leq \sigma + 2\kappa \log^{\frac{1}{2}}(2d).$$

Proof. Let $\ell := \log(2d) > 1$. If $\kappa = 0$, take $p \rightarrow \infty$; otherwise, set $p := \left\lceil \ell + \frac{\sigma\sqrt{\ell}}{\kappa} \right\rceil$. By the same argument as in the previous lemma, $d^{\frac{1}{2p}} \leq 1 + \frac{\ell}{p}$. Moreover, since $\kappa \leq \sigma$ and $\ell > 1$,

$$\sigma^2 + p\kappa^2 \leq \sigma^2 + \ell\kappa^2 + \sigma\kappa\sqrt{\ell} + \kappa^2 \leq (\sigma + \kappa\sqrt{\ell})^2.$$

Finally, $\frac{\ell}{p}(\sigma + \kappa\sqrt{\ell}) \leq \kappa\sqrt{\ell}$, so

$$d^{\frac{1}{2p}} \sqrt{\sigma^2 + p\kappa^2} \leq \left(1 + \frac{\ell}{p}\right)(\sigma + \kappa\sqrt{\ell}) \leq \sigma + 2\kappa\sqrt{\ell}. \quad \square$$