

Quantitative Tsirelson’s Theorems via Approximate Schur’s Lemma and Probabilistic Stampfli’s Theorems

Xiangling Xu^{1,2,3*}, Marc-Olivier Renou^{1,2,3}, Igor Klep^{4,5,6}

¹Inria Paris-Saclay, Bâtiment Alan Turing, 1 rue Honoré d’Estienne d’Orves, 91120 Palaiseau, France

²CPHT, Ecole polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau, France

³LIX, Ecole polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau, France

⁴Faculty of Mathematics and Physics, University of Ljubljana

⁵Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska

⁶Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia

*xu.xiangling@inria.fr

Abstract

Whether an almost-commuting pair of operators must be close to a commuting pair is a central question in operator and matrix theory. We investigate this problem for pairs of C^* -subalgebras \mathcal{A} and \mathcal{B} of $M_d(\mathbb{C})$, showing that each operator in \mathcal{B} is $O(d^2\varepsilon)$ -close in operator norm to an operator in the commutant \mathcal{A}' under two complementary formulations of “ ε -almost commutation.” One formulation is probabilistic, requiring that the operators of \mathcal{B} have small commutators for most Haar-random unitaries acting on \mathcal{A} . This first formulation leads to two novel probabilistic generalizations of Stampfli’s theorem, which relates an operator’s distance from the scalars to the norm of its inner derivation. The second formulation is deterministic, requiring small commutators between the generators of \mathcal{A} and \mathcal{B} ; we analyze this using an approximate Schur’s lemma formulated in terms of Weyl-Heisenberg (clock-and-shift) matrices. As an application of our results to quantum information theory, we obtain a quantitative Tsirelson’s theorem: in dimension d , every ε -almost quantum commuting observable model is well approximated by a quantum tensor-product model with error $O(d^2\varepsilon)$.

1 Introduction

A long-standing research theme in matrix theory asks when pairs of matrices that *almost* commute must be close to a genuinely commuting pair. Initiated by [Hal76; Ros69] for the normalized Hilbert-Schmidt norm and the operator norm, respectively, many results have been developed for these two norms under various structural assumptions [BH74; Cho88; FK10; FR96; Gle10; Ioa24; Lin96; Lin24; LT70; PS79; Voi83]. (See Sec. 4.3 for more details.) Our perspective and setup are motivated by quantum information theory: Tsirelson’s theorem [Tsi06] establishes an equivalence between the tensor-product model and the commuting operator model in finite-dimensional quantum systems. This naturally prompts the quantitative question: to what extent does the equivalence persist when the operators only approximately commute?

Keywords. Almost-commuting matrices, commutator estimates, approximate Schur’s lemma, probabilistic Stampfli theorem, Weyl-Heisenberg (clock-and-shift) matrices, finite-dimensional C^* -algebras, tensor-product factorization, quantitative Tsirelson’s theorem.

2020 Mathematics Subject Classification. Primary 15A27, 47B47, 81P40; Secondary 15A60, 47A55, 60B20.

1.1 Contribution and structure of the paper

Let $\mathcal{A}, \mathcal{B} \subset M_d(\mathbb{C})$ be finite-dimensional C^* -subalgebras with self-adjoint generating families $\{A_{a|x}\}_{a,x}$ and $\{B_{b|y}\}_{b,y}$, respectively. Under an “ ε -almost commuting” hypothesis for the generators—formulated in two complementary senses made precise in Thms. 2.5 and 3.7—we show that each generator $B_{b|y}$ of \mathcal{B} is $O(d^2\varepsilon)$ -close (in operator norm) to a self-adjoint operator $B'_{b|y}$ lying in the commutant \mathcal{A}' . In finite dimensions, semi-simplicity of \mathcal{A} implies that $\mathcal{A} \simeq \bigoplus_{l=1}^L B(\mathcal{H}_A^l) \otimes \mathbb{1}_B^l$, for some Hilbert spaces \mathcal{H}_A^l and \mathcal{H}_B^l with $\mathbb{1}_B^l$ acting on \mathcal{H}_B^l as the identity. Consequently, $B'_{b|y} \in \mathcal{A}'$ is equivalent to $B'_{b|y} = \bigoplus_l \mathbb{1}_A^l \otimes B''_{b|y}$, so our bounds yield an explicit approximate tensor-product factorization for the pair $(\mathcal{A}, \mathcal{B})$.

The first route, detailed in Sec. 2, is deterministic. This approach controls commutator bounds with respect to the Weyl-Heisenberg clock-and-shift matrices, which form a canonical generating set for $M_d(\mathbb{C})$. The key technical tool is an approximate Schur’s lemma (Lem. 2.1), which quantitatively shows that an operator almost commuting with the clock-and-shift matrices must be close to a scalar, and its consequence in bipartite systems (Lem. 2.2). Under assumptions on how efficiently the clock-and-shift matrices can be expressed using the generators $\{A_{a|x}\}$ of \mathcal{A} (quantified by algebraic complexity constants c_1, c_2, c_3), this leads to our first approximate Tsirelson’s theorem (Thm. 2.5), establishing the $O(d^2\varepsilon)$ estimate in the operator norm. We also discuss the scaling of the constants c_1, c_2, c_3 (Rem. 2.6).

Sec. 3 presents the second, probabilistic route. Here, we relax the requirement of uniform commutator bounds. We instead demand small commutators only for most Haar-random unitaries within two-dimensional subspaces (aka, single-qubits), and explain how this can be generalized to arbitrary d -dimensional subspaces in Rem. 3.4. This leads to a probabilistic Stampfli’s theorem (Thm. 3.3), relating these probabilistic commutator bounds to an operator’s distance from scalar multiples of the identity. We further extend this to a doubly probabilistic version (Thm. 3.5) by also randomizing the two-dimensional subspaces. While the former (Thm. 3.3) holds for possibly infinite-dimensional Hilbert spaces, the doubly probabilistic variant (Thm. 3.5) holds only in finite dimensions. The doubly probabilistic Stampfli’s theorem allows us to derive our second main approximate Tsirelson’s theorem (Thm. 3.7), again achieving an $O(d^2\varepsilon)$ error estimate.

Finally, Sec. 4 explores the applications and context of our work. We detail the construction of an approximating tensor-product quantum strategy based on our main theorems (Prop. 4.1, see the following subsection). Furthermore, we explore the interplay between our results, the Navascués-Pironio-Acín (NPA) hierarchy [NPA08; PNA10]—a convergent semidefinite programming (SDP) hierarchy characterizing commuting quantum correlations—and computational complexity (Rem. 4.2). We situate our results within the literature on approximating almost commuting matrices (Sec. 4.3) and conclude with a broader discussion (Sec. 4.4).

1.2 Motivation: application to quantum information theory

Quantum information theory studies how information is represented and processed in quantum systems. In the finite-dimensional setting relevant here, one works with states (density matrices) and observables (self-adjoint operators/POVMs) on Hilbert spaces and asks how structural assumptions—such as subsystem independence—constrain observable correlations.

Quantum information theory offers two different axioms for composing independent subsystems. The standard composition axiom [NC10] postulates that the Hilbert space \mathcal{H}_T describing a joint system $T = \{A, B\}$ is the tensor product $\mathcal{H}_T := \mathcal{H}_A \otimes \mathcal{H}_B$. Each local subsystem’s observables act on its respective tensor factors and are identity over the other subsystem: e.g., unitaries U_A, V_B local to subsystems A, B act on the global system T as $U_A \otimes \mathbb{1}_B$ and $\mathbb{1}_A \otimes V_B$, respectively. An

alternative axiomatization, common in algebraic quantum field theory [Lan17], does not introduce the tensor product, but models independence by postulating that the observables of different parties act on the same Hilbert space and commute. In other words, U_A, V_B act on the same global Hilbert space \mathcal{H}_T with only $[U_A, V_B] = 0$.

Whether these two axiomatizations result in the same physical predictions was a long-standing problem. In particular, Tsirelson's problem asks whether the *tensor-product model* and the *commuting operator model* yield the same set of bipartite quantum correlations—a concept that has been central to both the foundational understanding and practical applications of quantum theory since Bell's groundbreaking work [Bel64]. Tsirelson demonstrated the equivalence of these two models in finite-dimensional settings [Doh+08; SW08; Tsi06], formally:

Theorem (Tsirelson's theorem). *Let $\{A_{a|x}\}, \{B_{b|y}\} \subset B(\mathcal{H})$ be two sets of positive operator-valued measures (POVMs) on a finite-dimensional Hilbert space \mathcal{H} such that $[A_{a|x}, B_{b|y}] = 0$ for all a, b, x, y . Then there exists a decomposition of $\mathcal{H} \simeq \bigoplus_l (\mathcal{H}_A^l \otimes \mathcal{H}_B^l)$, such that operators $A_{a|x}$ and $B_{b|y}$ take the form*

$$A_{a|x} = \bigoplus_l (A_{a|x}^l \otimes \mathbb{1}_B^l) \quad \text{and} \quad B_{b|y} = \bigoplus_l (\mathbb{1}_A^l \otimes B_{b|y}^l),$$

where $A_{a|x}^l \in B(\mathcal{H}_A^l)$ and $B_{b|y}^l \in B(\mathcal{H}_B^l)$. Consequently, any commuting-operator correlations

$$p(ab|xy) = \text{Tr}(\rho \cdot A_{a|x} \cdot B_{b|y})$$

obtained from a state ρ on \mathcal{H} can be reproduced as tensor-product correlations using POVMs $\tilde{A}_{a|x} \in B(\bigoplus_l \mathcal{H}_A^l)$ and $\tilde{B}_{b|y} \in B(\bigoplus_l \mathcal{H}_B^l)$ on a state $\tilde{\rho}$ over $(\bigoplus_l \mathcal{H}_A^l) \otimes (\bigoplus_l \mathcal{H}_B^l)$, i.e.,

$$p(ab|xy) = \text{Tr}(\tilde{\rho} \cdot (\tilde{A}_{a|x} \otimes \tilde{B}_{b|y})).$$

The statement extends inductively to multipartite cases.

Tsirelson conjectured that the above theorem can be generalized to infinite dimensions. However, this was recently shown to be false by the seminal work [Ji+21]: the commuting operator model can produce correlations unattainable by any tensor-product quantum strategy. This result has far-reaching implications including a disproof of Connes' embedding conjecture [Con76] and Kirchberg's conjecture [Kir93]; see [Oza13a] for a nice survey.

By contrast, in finite dimensions, our results lead to a quantitative robustness version of Tsirelson's theorem: if a d -dimensional bipartite strategy is ε -almost commuting in the sense of Thms. 2.5 and 3.7, then its correlations are approximated by those of a genuine tensor-product strategy with error $O(d^2\varepsilon)$ (Prop. 4.1). This is a quantitative counterpart to Ozawa's asymptotic result [Oza13b], further demonstrating that, in finite dimensions, the tensor-product model remains a sound effective description when subsystem independence holds only approximately. The multipartite extension follows by induction, as in the original Tsirelson's theorem.

2 Weyl-Heisenberg (clock-and-shift) formulation

In this section, we present our first main result, a deterministic version of Tsirelson's theorem in Thm. 2.5. Recall the key ideas for proving Tsirelson's theorem (see, e.g. [Doh+08, App. A]): any finite-dimensional C^* -algebra \mathcal{A} generated by Alice's observables $A_{a|x}$ decomposes as a direct sum

of simple blocks $\mathcal{A} \simeq \bigoplus_l B(\mathcal{H}_A^l) \otimes \mathbb{1}_B^l$, and by Schur's lemma, Bob's commuting algebra \mathcal{B} must lie in $\bigoplus_l \mathbb{1}_A^l \otimes B(\mathcal{H}_B^l)$.

Our approximate analogue replaces “commutes” by “almost commutes”. In particular, to formulate our results, we pick Sylvester's *clock* Σ_3 and *shift* Σ_1 unitaries as generators: almost commuting with this pair already controls an operator in every direction.

With these generators we prove an approximate Schur's lemma (Lem. 2.1) and its bipartite version (Lem. 2.2). Next, we show an approximate Tsirelson's theorem for simple algebras (Lem. 2.4), and then by the same block-decomposition argument, a general approximate Tsirelson's theorem (Thm. 2.5). We finish with a discussion of scaling (Rem. 2.6).

2.1 Clock-and-shift matrices

Recall Sylvester's clock-and-shift matrices [App05], which generalize the Pauli matrices to a d -dimensional Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$. Also known as the Weyl-Heisenberg matrices, they are fundamental in finite-dimensional quantum mechanics due to their connection to Weyl's formulation of the canonical commutation relations. These matrices serve as analogs of position and momentum operators in finite-dimensional quantum systems.

Let $\omega = e^{2\pi i/d}$ be the d th root of unity. Using Dirac's notation, denote by $\{|i\rangle \mid i = 0, \dots, d-1\}$ the standard basis of \mathcal{H} , and $|i+j\rangle$ is understood up to mod d . Then the shift matrix $\Sigma_1 \in B(\mathcal{H})$ is defined by $\Sigma_1 : |i\rangle \mapsto |i+1\rangle$ and the clock matrix is defined by $\Sigma_3 : |i\rangle \mapsto \omega^i |i\rangle$. More explicitly:

$$\Sigma_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad \Sigma_3 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \omega & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega^{d-1} \end{pmatrix}. \quad (1)$$

The notation comes from the fact that the shift matrix Σ_3 (resp. clock matrix Σ_1) is a generalization of the Pauli Z -matrix σ_3 (resp. X -matrix σ_1) when $d = 2$. The clock and shift Σ_3, Σ_1 satisfy a generalized algebraic relation of the Pauli matrices in the sense that

$$\begin{aligned} \Sigma_1^d &= \Sigma_3^d = \mathbb{1}, \\ \Sigma_3 \Sigma_1 &= \omega \Sigma_1 \Sigma_3. \end{aligned} \quad (2)$$

Also note that both Σ_1, Σ_3 are unitary and traceless, but no longer Hermitian when $d > 2$. Lastly, they give rise to an orthogonal basis of $B(\mathcal{H})$ (w.r.t Hilbert-Schmidt inner product) composed of unitary matrices

$$\{\sigma_{k,l} := \Sigma_1^k \Sigma_3^l = \sum_{j=0}^{d-1} \omega^{jl} |j+k\rangle \langle j| \mid 0 \leq k, l \leq d-1\}, \quad (3)$$

where $|j+k\rangle \langle j| := (|j+k\rangle)^* |j\rangle$.

2.2 Approximate Schur's Lemma and its bipartite version

Elementary linear algebraic arguments lead to an approximate version of Schur's Lemma for finite-dimensional Hilbert spaces. Write $\|\cdot\|_{\text{op}}$ for the operator (spectral) norm and $\|\cdot\|_{\text{max}}$ for the max norm, i.e., $\|A\|_{\text{max}} = \max_{i,j} |A_{ij}|$.

Lemma 2.1 (Approximate Schur's Lemma). *Let \mathcal{H} be a d -dimensional Hilbert space, $d < \infty$. Consider a fixed matrix $C \in B(\mathcal{H}) \simeq M_d(\mathbb{C})$ and suppose there exists an $\varepsilon > 0$ such that, for both $i = 1, 3$,*

$$\|[C, \Sigma_i]\|_{\text{op}} \leq \varepsilon. \quad (4)$$

Then there exists $c \in \mathbb{C}$ such that

$$\|C - c\mathbf{1}\|_{\text{op}} \leq (d-1)\varepsilon. \quad (5)$$

Proof. Note that for any matrices X, Y, Z we have

$$\|[XY, Z]\|_{\text{op}} = \|X[Y, Z] + [X, Z]Y\|_{\text{op}} \leq \|X\|_{\text{op}}\|[Y, Z]\|_{\text{op}} + \|[X, Z]\|_{\text{op}}\|Y\|_{\text{op}}.$$

Then, by induction and the fact that $\|\Sigma_i\|_{\text{op}} = 1$, we have

$$\|[C, \Sigma_i^k]\|_{\text{op}} \leq k\varepsilon.$$

Next, note that $C\Sigma_1$ is the matrix where each column of C is cyclically shifted leftward, and $\Sigma_1 C$ is the matrix where each row of C is cyclically shifted downward. That is, for all i , the $(i, i-1)$ -entry of $C\Sigma_1$ is C_{ii} , while the $(i, i-1)$ -entry of $(\Sigma_1 C)$ is $C_{i-1, i-1}$. Then the assumption imposes that

$$|C_{i,i} - C_{i-1, i-1}| = |(C\Sigma_1 - \Sigma_1 C)_{i, i-1}| \leq \|C\Sigma_1 - \Sigma_1 C\|_{\text{max}} \leq \|[C, \Sigma_1]\|_{\text{op}} \leq \varepsilon,$$

Consequently,

$$|C_{ii} - \frac{1}{d} \text{Tr}(C)| = \frac{1}{d} \left| \sum_j (C_{ii} - C_{jj}) \right| \leq \frac{1}{d} (0 + \varepsilon + \dots + (d-1)\varepsilon) = \frac{d-1}{2} \varepsilon$$

by repeated uses of the triangle inequality.

Moreover, observe that $\text{diag}(C) = 1/d \sum_{k=0}^{d-1} \Sigma_3^k C \Sigma_3^{-k}$, since

$$\left(\frac{1}{d} \sum_{k=0}^{d-1} \Sigma_3^k C \Sigma_3^{-k} \right)_{ij} = \frac{1}{d} \underbrace{\sum_{k=0}^{d-1} \omega^{k(i-j)} C_{ij}}_{d\delta_{i,j}} = \delta_{i,j} C_{ij}.$$

It follows that

$$\begin{aligned} \|C - \text{diag}(C)\|_{\text{op}} &= \frac{1}{d} \left\| \sum_{k=0}^{d-1} (C - \Sigma_3^k C \Sigma_3^{-k}) \right\|_{\text{op}} \leq \frac{1}{d} \sum_{k=0}^{d-1} \|[\Sigma_3^k, C] \Sigma_3^{-k}\|_{\text{op}} \\ &\leq \frac{1}{d} \sum_{k=0}^{d-1} \|[\Sigma_3^k, C]\|_{\text{op}} \|\Sigma_3^{-k}\|_{\text{op}} \leq \frac{1}{d} \sum_{k=0}^{d-1} k\varepsilon \cdot 1^k = \frac{d-1}{2} \varepsilon. \end{aligned}$$

Finally, let $c = 1/d \text{Tr}(C)$, we compute

$$\begin{aligned} \|C - c\mathbf{1}\|_{\text{op}} &\leq \|C - \text{diag}(C)\|_{\text{op}} + \|\text{diag}(C) - \frac{1}{d} \text{Tr}(C) \mathbf{1}\|_{\text{op}} \\ &\leq \frac{d-1}{2} \varepsilon + \max_i |C_{ii} - \frac{1}{d} \text{Tr}(C)| \leq \frac{d-1}{2} \varepsilon + \frac{d-1}{2} \varepsilon = (d-1)\varepsilon. \end{aligned}$$

□

Via manipulation of the Kronecker tensor product formula, we quickly obtain a bipartite version of approximate Schur's Lemma.

Lemma 2.2. *Consider two Hilbert spaces \mathcal{H}_1 with dimension d_1 and \mathcal{H}_2 with dimension d_2 . Suppose for the matrix $C \in B(\mathcal{H}_1 \otimes \mathcal{H}_2)$, there exists some $\varepsilon > 0$ such that*

$$\|[C, \mathbb{1}_1 \otimes \Sigma_1]\|_{\text{op}}, \|[C, \mathbb{1}_1 \otimes \Sigma_3]\|_{\text{op}} \leq \varepsilon. \quad (6)$$

Then the matrix $C' = 1/d_2 \text{Tr}_{\mathcal{H}_2}(C) \in B(\mathcal{H}_1)$, where $\text{Tr}_{\mathcal{H}_2}$ denotes the partial trace $B(\mathcal{H}_1 \otimes \mathcal{H}_2) \rightarrow B(\mathcal{H}_1)$, satisfies

$$\|C - C' \otimes \mathbb{1}_2\|_{\text{op}} \leq d_1 d_2^2 \varepsilon. \quad (7)$$

In addition, if C is positive semidefinite then so is C' .

Proof. Note that

$$C = \begin{pmatrix} C_{(11)} & \cdots & C_{(1d_1)} \\ \vdots & \ddots & \vdots \\ C_{(d_11)} & \cdots & C_{(d_1d_1)} \end{pmatrix},$$

where each $C_{(ij)}$ is some $d_2 \times d_2$ matrix in $B(\mathcal{H}_2)$. Similarly, we have that

$$\mathbb{1}_1 \otimes \Sigma_i = \begin{pmatrix} \Sigma_i & & \\ & \ddots & \\ & & \Sigma_i \end{pmatrix}$$

are block matrices with only Σ_i on the diagonal. Since the operator norm of a matrix upper-bounds the operator norm of its blocks, the condition $\|[C, \mathbb{1}_1 \otimes \Sigma_i]\|_{\text{op}} \leq \varepsilon$ implies that, for all k, l ,

$$\|[C_{(kl)}, \Sigma_1]\|_{\text{op}}, \|[C_{(kl)}, \Sigma_3]\|_{\text{op}} \leq \varepsilon.$$

Then, applying the approximate version of Schur's Lemma 2.1, for each k, l we check $c_{kl} := 1/d_2 \text{Tr}(C_{(kl)})$ satisfies

$$\|C_{(kl)} - c_{kl} \mathbb{1}_2\|_{\text{max}} \leq \|C_{(kl)} - c_{kl} \mathbb{1}_2\|_{\text{op}} \leq (d_2 - 1)\varepsilon \leq d_2 \varepsilon.$$

Defining $C' = (c_{kl}) \in B(\mathcal{H}_1)$, it follows that

$$C' \otimes \mathbb{1}_2 = \begin{pmatrix} c_{11} \mathbb{1}_2 & \cdots & c_{1d_1} \mathbb{1}_2 \\ \vdots & \ddots & \vdots \\ c_{d_11} \mathbb{1}_2 & \cdots & c_{d_1d_1} \mathbb{1}_2 \end{pmatrix}.$$

Hence,

$$\|C - C' \otimes \mathbb{1}_2\|_{\text{op}} \leq d_1 d_2 \|C - C' \otimes \mathbb{1}_2\|_{\text{max}} \leq d_1 d_2 \max_{kl} (\|C_{(kl)} - c_{kl} \mathbb{1}_2\|_{\text{max}}) \leq d_1 d_2^2 \varepsilon.$$

(Note that the above operator norm upper bound of $C - C' \otimes \mathbb{1}_2$ via its block is the tightest general bound, e.g., consider the matrix whose entries are all 1 and take each entry as a block.)

Lastly, observe that C' is in fact the normalized partial trace of C , since

$$\mathrm{Tr}_{\mathcal{H}_2}(C) = \begin{pmatrix} \mathrm{Tr}(C_{(11)}) & \cdots & \mathrm{Tr}(C_{(1d_1)}) \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}(C_{(d_11)}) & \cdots & \mathrm{Tr}(C_{(d_1d_1)}) \end{pmatrix} = d_2 \begin{pmatrix} c_{11} & \cdots & c_{1d_1} \\ \vdots & \ddots & \vdots \\ c_{d_11} & \cdots & c_{d_1d_1} \end{pmatrix} = d_2 C'.$$

This implies that if C is positive semidefinite, then so is C' , due to complete positivity of the partial trace [Bla06, Ch. II.6.10] \square

2.3 Approximate Tsirelson's theorem from clock-and-shift matrices

Before presenting the approximate version of Tsirelson's theorem, we recall the Artin-Wedderburn decomposition of a finite-dimensional C^* -algebra [Bla06, II.1.6.4, II.8.3.2(iv), and III.1.5.3].

Lemma 2.3. *Every finite-dimensional C^* -algebra \mathcal{A} is semi-simple. That is, there exists an Artin-Wedderburn decomposition*

$$\mathcal{A} = \bigoplus_k \mathcal{A}_k,$$

such that each \mathcal{A}_k is simple, i.e. contains no non-trivial closed two-sided ideals.

Furthermore, if $\mathcal{A} \subset B(\mathcal{H})$ is simple, then there exists a bipartite partition of \mathcal{H} such that $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$ and $\mathcal{A} \simeq B(\mathcal{H}_1) \otimes \mathbb{1}_2$.

This structural result gives a road map: first prove the simple version of approximate Tsirelson's theorem, then the general case follows. For the following approximate Tsirelson's theorem, let us impose extra assumptions on the “generating power of the strategy”, represented by the constants c_1, c_2, c_3 below.

Lemma 2.4 (Approximate Tsirelson's theorem, simple case). *Let \mathcal{H} be a d -dimensional Hilbert space \mathcal{H} . Let $\mathcal{A} \subset B(\mathcal{H})$ be generated by contractive self-adjoint operators $\{A_{a|x}\}$ and $\mathcal{B} \subset B(\mathcal{H})$ be generated by contractive self-adjoint operators $\{B_{b|y}\}$. Assume that there exists an $\varepsilon > 0$, such that for all a, b, x, y ,*

$$\|[A_{a|x}, B_{b|y}]\|_{\mathrm{op}} \leq \varepsilon. \quad (8)$$

Suppose that \mathcal{A} is simple, i.e. there exists a bipartition $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$ such that $\mathcal{A} \simeq B(\mathcal{H}_A) \otimes \mathbb{1}_B$ and $A_{a|x} = A'_{a|x} \otimes \mathbb{1}_B$ for all a, x .

Suppose that the clock-and-shift matrices $\Sigma_3, \Sigma_1 \in B(\mathcal{H}_A)$ are generated by some polynomials P_3, P_1 in $\{A'_{a|x}\}$. Assume, moreover, that the maximal absolute value of their coefficients is bounded by c_1 , the maximal degree is bounded by c_2 , and the maximal number of terms is bounded by c_3 . Then there exist operators $B'_{b|y} \in B(\mathcal{H}_B)$ such that, for all b, y ,

$$\|B_{b|y} - \mathbb{1}_A \otimes B'_{b|y}\|_{\mathrm{op}} \leq c_1 c_2 c_3 d^2 \varepsilon. \quad (9)$$

In addition, if $B_{b|y}$ is positive then so is $B'_{b|y}$

Proof. Note that for any matrices X, Y, Z we have

$$\|[XY, Z]\|_{\mathrm{op}} = \|X[Y, Z] + [X, Z]Y\|_{\mathrm{op}} \leq \|X\|_{\mathrm{op}}\|[Y, Z]\|_{\mathrm{op}} + \|[X, Z]\|_{\mathrm{op}}\|Y\|_{\mathrm{op}}. \quad (10)$$

Then for any monomial α in $\{A_{a|x}\}$ of degree k , one can use the fact that $\|A_{a|x}\|_{\text{op}} \leq 1$ to inductively compute

$$\|[\alpha, B_{b|y}]\|_{\text{op}} \leq k \max_{a,x} \|A_{a|x}\|_{\text{op}} \| [A_{a|x}, B_{b|y}] \|_{\text{op}} \leq k\varepsilon.$$

Then, for polynomials $\Sigma_i = P_i(\{A_{a|x}\})$, we have

$$\|[\Sigma_i, B_{b|y}]\|_{\text{op}} \leq c_1 c_2 c_3 \max_{a,x} (\| [A_{a|x}, B_{b|y}] \|_{\text{op}}) \leq c_1 c_2 c_3 \varepsilon,$$

and we are done by Lem. 2.2. \square

Remark that the “contraction” requirement is not necessary and one can reproduce the same result by replacing ε by $\varepsilon/\|A_{a|x}\|_{\text{op}}$. The simple version can be readily generalized to the general finite-dimensional case with Lem. 2.3.

Theorem 2.5 (Approximate Tsirelson’s theorem, general case). *Let \mathcal{A} be generated by contractive self-adjoint operators $\{A_{a|x}\} \subset B(\mathcal{H})$ and \mathcal{B} be generated by contractive self-adjoint operators $\{B_{b|y}\} \subset B(\mathcal{H})$ for some d -dimensional Hilbert space \mathcal{H} . Assume that there exists an $\varepsilon > 0$, such that for all a, b, x, y ,*

$$\| [A_{a|x}, B_{b|y}] \|_{\text{op}} \leq \varepsilon. \quad (11)$$

Suppose also that \mathcal{A} admits the Artin-Wedderburn decomposition

$$\mathcal{A} = \bigoplus_{l=1}^L \mathcal{A}_l \simeq \bigoplus_{l=1}^L B(\mathcal{H}_A^l) \otimes \mathbb{1}_B^l \text{ and } A_{a|x} = \bigoplus_{l=1}^L A_{a|x}^l \otimes \mathbb{1}_B^l,$$

with the corresponding orthogonal projectors Π_l to the direct summands. Denote by $\Sigma_3^l, \Sigma_1^l \in B(\mathcal{H}_A^l)$ the clock-and-shift operators in \mathcal{H}_A^l .

Furthermore, suppose that there exist polynomials P_l, Q_1^l, Q_3^l , for all $l = 1, \dots, L$ such that

$$\Pi_l = P_l(\{A_{a|x}\}), \Sigma_1^l = Q_1^l(\{\Pi_l A_{a|x} \Pi_l\}), \text{ and } \Sigma_3^l = Q_3^l(\{\Pi_l A_{a|x} \Pi_l\}).$$

Assume that their absolute values of the maximal coefficients are bounded by the constant c_1 , the degrees are bounded by the constant c_2 , and the maximal number of terms is bounded by the constant c_3 . Then there exist operators $B'_{b|y} \in \bigoplus_{l=1}^L \mathbb{1}_A^l \otimes B(\mathcal{H}_B^l) = \mathcal{A}'$ such that, for all b, y ,

$$\|B_{b|y} - B'_{b|y}\|_{\text{op}} \leq 2c_1 c_2 c_3 (c_1 c_2 c_3 + 1) d^2 \varepsilon. \quad (12)$$

In addition, if $B_{b|y}$ is positive then so is $B'_{b|y}$.

Proof. First, we wish to apply Lem. 2.4 to each $\Pi_l A_{a|x} \Pi_l \in \mathcal{A}_l$ and the corresponding $\Pi_l B_{b|y} \Pi_l$. To this end, note that, by straightforward calculations, one has

$$\| [B_{b|y}, \Pi_l] \|_{\text{op}} \leq c_1 c_2 c_3 \| [A_{a|x}, B_{b|y}] \|_{\text{op}} \leq c_1 c_2 c_3 \varepsilon$$

and

$$[\Pi_l A_{a|x} \Pi_l, \Pi_l B_{b|y} \Pi_l] = \Pi_l [\Pi_l, B] A_{a|x} \Pi_l + \Pi_l [A_{a|x}, B_{b|y}] \Pi_l + \Pi_l A_{a|x} [\Pi_l, B_{b|y}] \Pi_l.$$

It follows from $\|A_{a|x}\|_{\text{op}}, \|\Pi_l\|_{\text{op}} \leq 1$ and the Cauchy-Schwarz inequality that

$$\|[\Pi_l A_{a|x} \Pi_l, \Pi_l B_{b|y} \Pi_l]\|_{\text{op}} \leq \|[\Pi_l, B]\|_{\text{op}} + \|[A_{a|x}, B_{b|y}]\|_{\text{op}} + \|[\Pi, B_{b|y}]\|_{\text{op}} \leq (2c_1 c_2 c_3 + 1)\varepsilon.$$

Therefore, by Lem. 2.4, there exists for each l some positive semidefinite $B_{b|y}^l \in B(\mathcal{H}_B^l)$ such that

$$\|\Pi_l B_{b|y} \Pi_l - \mathbb{1}_A^l \otimes B_{b|y}^l\|_{\text{op}} \leq c_1 c_2 c_3 (2c_1 c_2 c_3 + 1) d_l^2 \varepsilon,$$

where $d_l = \dim(\mathcal{H}_A^l \otimes \mathcal{H}_B^l)$.

Now, $B_{b|y}$ does not admit the same direct decomposition as \mathcal{A} due to $[\Pi_l, B_{b|y}] \neq 0$. Thus, we need to also estimate

$$\begin{aligned} \|B_{b|y} - \sum_l \Pi_l B_{b|y} \Pi_l\|_{\text{op}} &\leq \left\| \sum_{l,l'} \Pi_l B_{b|y} \Pi_{l'} - \sum_l \Pi_l B_{b|y} \Pi_l \right\|_{\text{op}} \\ &\leq \left\| \sum_{l \neq l'} \Pi_l B_{b|y} \Pi_{l'} \right\|_{\text{op}} \\ &\leq \sum_{l \neq l'} \|\Pi_l \Pi_{l'} B_{b|y} + \Pi_l [B_{b|y}, \Pi_{l'}]\|_{\text{op}} \\ &\leq \sum_{l \neq l'} \|\Pi_l\|_{\text{op}} \| [B_{b|y}, \Pi_{l'}] \|_{\text{op}} \leq L(L-1) c_1 c_2 c_3 \varepsilon, \end{aligned}$$

where the completeness and orthogonality of Π_l are used.

Finally, one sees that

$$\begin{aligned} \|B_{b|y} - \bigoplus_{l=1}^L \mathbb{1}_A^l \otimes B_{b|y}^l\|_{\text{op}} &\leq \|B_{b|y} - \sum_{l=1}^L \Pi_l B_{b|y} \Pi_l\|_{\text{op}} + \left\| \sum_{l=1}^L \Pi_l B_{b|y} \Pi_l - \bigoplus_{l=1}^L \mathbb{1}_A^l \otimes B_{b|y}^l \right\|_{\text{op}} \\ &\leq L(L-1) c_1 c_2 c_3 \varepsilon + \sum_{l=1}^L c_1 c_2 c_3 (2c_1 c_2 c_3 + 1) d_l^2 \varepsilon \leq c_1 c_2 c_3 (L(L-1) + (2c_1 c_2 c_3 + 1) d^2) \varepsilon. \end{aligned}$$

Note $L \leq d$, so we are done by defining $B'_{b|y} := \bigoplus_{l=1}^L \mathbb{1}_A^l \otimes B_{b|y}^l$. \square

While all results in this section are formulated in terms of the Weyl-Heisenberg (clock-and-shift) matrices Σ_1, Σ_3 , we observe that any full generating set (e.g., the matrix units $E_{kl} = |k\rangle\langle l|$) would still work. We finish the section with a remark on the factors c_1, c_2, c_3 .

Remark 2.6. *Note that the bound Eq. (12) has an $O(d^2 \varepsilon)$ scaling for fixed c_1, c_2, c_3 . We finish the section with some comments on these factors c_1, c_2, c_3 , which are generally example-specific.*

1. *The generating polynomial degree c_2 is related to the length of algebras with known dependence on the dimension d . The conjectured bound is $O(d)$ according to [Paz84], while the best proven bound is $O(d \log(d))$ due to [Shi19]. The bound is $O(\log d)$ when a “generic” assumption is met as detailed in [KS16].*
2. *The number of terms in generating polynomials c_3 is related to c_2 . In the worst case scenario, c_3 is the number of possible monomials of $\{A_{a|x}\}$ up to degree c_2 , which grows exponentially in c_2 (i.e. $c_3 \leq \sum_{k=0}^{c_2} (|\{A_{a|x}\}|)^k$).*
3. *The coefficient magnitude c_1 can be challenging to bound generally. While specific algebraic structures might lead to large c_1 , work by [Pas19] offers a systematic approach. It involves*

constructing a matrix P from the POVM generators, whose properties (e.g., its singular values or entry magnitudes) can serve as an indicator for the likely behavior of c_1 .

3 Haar-random single-qubit unitary formulation

In this section, we present our second main result, a probabilistic version of Tsirelson’s theorem in Thm. 3.7. Since this will have a matrix generator independent formulation, we start by looking for a uniform version of Schur’s lemma. Such a result is given by Stampfli [Sta70, Thm. 4 & Cor. 1].

Theorem 3.1 (Stampfli’s theorem). *Let $C \in B(\mathcal{H})$ for some Hilbert space \mathcal{H} of possibly infinite dimensions. Then*

$$\sup_{B \in B(\mathcal{H}): \|B\|_{\text{op}}=1} \|[C, B]\|_{\text{op}} = \inf_{c \in \mathbb{C}} 2\|C - c\mathbb{1}\|_{\text{op}}. \quad (13)$$

Moreover, if C is a normal operator, then

$$\inf_{c \in \mathbb{C}} \|C - c\mathbb{1}\|_{\text{op}} = R(\sigma(C)), \quad (14)$$

where $R(\sigma(C))$ is the radius of the minimum enclosing disk of the compact set $\sigma(C) \subset \mathbb{C}$. Note that $R(\sigma(C))$ is not the same as the spectral radius of C .

In particular, if $\|[C, B]\|_{\text{op}} \leq \varepsilon$ for all B of norm 1, then C is $\varepsilon/2$ -close to some scalar operator $c\mathbb{1}$ in operator norm. Note that $\|[C, U]\|_{\text{op}} \leq \varepsilon$ for all unitaries U is equivalent to the assumption that $\|[C, B]\|_{\text{op}} \leq \varepsilon$ for all B of norm 1 due to the Russo-Dye theorem [Bla06, Cor. II.3.2.15].

While mathematically pleasing, Stampfli’s premise is too strong in physical scenarios, since testing commutators with all operators B satisfying $\|B\|_{\text{op}} = 1$ would require probing an uncountable family of observables. Therefore, in this section, we revisit Stampfli’s theorem through a physically motivated probabilistic approach—the Haar-random single-qubit unitary formulation. Here, “single-qubit” in physics jargon refers to \mathbb{C}^2 -subspaces (more generally, “qudits” are \mathbb{C}^d for any $d \geq 2$).

We first show that the demanding requirement “ C almost commutes with *every* unitary” can be relaxed to “ C almost commutes with *most* single-qubit unitaries taken at random”, yielding our probabilistic Stampfli theorem (Thm. 3.3). Because the Haar measure is unavailable in infinite dimensions, the randomization is implemented by sampling Haar-random unitaries inside every two-dimensional subspace. However, checking every two-dimensional subspace in an infinite-dimensional space is still unrealistically demanding.

Hence, we then push the idea further: by also randomizing these two-dimensional subspaces, we obtain a doubly probabilistic Stampfli’s theorem (Thm. 3.5) that is better aligned with realistic experiments. Though, due to the technicality of randomization over subspaces, our result necessarily restricts to finite dimensions. Finally, we develop another approximate Tsirelson’s theorem (Thm. 3.7) based on this doubly probabilistic Haar-random single-qubit unitary formulation.

We observe that the above results can also be formulated with d -dimensional subspaces for arbitrary d (Rem. 3.4).

3.1 Probabilistic Stampfli’s theorem

The first probabilistic relaxation of commutation can be written as follows: for Haar-random unitaries U , there are $\varepsilon, \delta > 0$, such that the probability of having a small commutator ($\leq \varepsilon$) with U is high ($\geq 1 - \delta$).

For notational convenience, from now on denote by $\text{Gr}(2, \mathcal{H})$ the Grassmannian of \mathcal{H} , this is a manifold whose elements are exactly two-dimensional subspaces $\mathcal{K} \subset \mathcal{H}$. Let $\mu_{\mathcal{K}}$ denote the Haar probability measure on the unitary group $U(\mathcal{K})$ in $B(\mathcal{K})$.

We first consider self-adjoint operators, and the general case follows from the standard decomposition into real plus imaginary parts.

Lemma 3.2. *Let \mathcal{H} be a Hilbert space of possibly infinite dimension, and let $C \in B(\mathcal{H})$ be a self-adjoint operator. Given $\varepsilon > 0$ and $\delta \in [0, 1]$, suppose that*

$$\Pr_{U \sim \mu_{\mathcal{K}}} \{ \| [P_{\mathcal{K}} C P_{\mathcal{K}}, U] \|_{\text{op}} \leq \varepsilon \} \geq 1 - \delta \quad (15)$$

for every subspace $\mathcal{K} \in \text{Gr}(2, \mathcal{H})$ with projector $P_{\mathcal{K}}$. Then

$$\inf_{c \in \mathbb{C}} \| C - c \mathbb{1}_{\mathcal{H}} \|_{\text{op}} \leq \min \left(\frac{\sqrt{2}}{2} \left(\sqrt{1 - \delta} \varepsilon + 2\sqrt{\delta} \| C \|_{\text{op}} \right), \left\lceil \frac{1}{1 - \delta} \right\rceil \frac{\varepsilon}{2} \right). \quad (16)$$

Note that the bound $\lceil \frac{1}{1 - \delta} \rceil \frac{\varepsilon}{2}$ generally behaves better when δ is small (e.g., it reduces to ε when $\delta \leq 1/2$), while the bound $\frac{\sqrt{2}}{2} \left(\sqrt{1 - \delta} \varepsilon + 2\sqrt{\delta} \| C \|_{\text{op}} \right)$ is more stable and does not blow up as $\delta \rightarrow 1$.

Proof. Let us first derive the bound $\frac{\sqrt{2}}{2} \left(\sqrt{1 - \delta} \varepsilon + 2\sqrt{\delta} \| C \|_{\text{op}} \right)$. The central object to bound is $\mathbb{E} \| [P_{\mathcal{K}} C P_{\mathcal{K}}, U] \|_{\text{op}}^2$ —while the upper bound is straightforward, the lower bound requires more work. The main idea is to identify $\mathcal{K} = \text{Span} \{ |\psi_1\rangle, |\psi_2\rangle \}$. Up to infinite-dimensional subtlety, the vector $|\psi_1\rangle$ (resp. $|\psi_2\rangle$) is chosen to approximate the “eigenvector” of C associated with the minimal (resp. maximal) “eigenvalue”. On this \mathcal{K} , one can then lower bound $\| [P_{\mathcal{K}} C P_{\mathcal{K}}, U] \|_{\text{op}}$ by $2 \inf_{c \in \mathbb{C}} \| C - c \mathbb{1}_{\mathcal{H}} \|_{\text{op}}^2$ using the radius of spectrum of C , and then apply Stampfli’s Theorem 3.1.

We begin with the spectral theorem for the bounded self-adjoint operator C [Bla06, Ch. I.6.1]: there exists a unique projective-valued measure E such that

$$C = \int_{\sigma(C)} \lambda dE(\lambda).$$

Note that the spectrum $\sigma(C)$ satisfies

$$\sigma(C) \subset \left[\inf_{\|\psi\|_2=1} \langle \psi | C | \psi \rangle, \sup_{\|\psi\|_2=1} \langle \psi | C | \psi \rangle \right] := [\lambda_{\min}, \lambda_{\max}].$$

We can check in this case that $R(\sigma(C)) = \frac{1}{2}(\lambda_{\max} - \lambda_{\min})$.

If $\lambda_{\min} = \lambda_{\max}$, then C is automatically a scalar operator and the conclusion is trivial. Otherwise, given an $\eta > 0$, we may consider intervals $I_1, I_2 \subset \sigma(C)$ such that

$$I_1 = [\lambda_{\min}, \lambda_{\min} + \eta] \cap \sigma(C), \quad I_2 = [\lambda_{\max} - \eta, \lambda_{\max}] \cap \sigma(C)$$

with the corresponding spectral projections $E(I_1), E(I_2)$. Fix two unit vectors, $|\psi_1\rangle \in E(I_1)\mathcal{H}$ and

$|\psi_2\rangle \in E(I_2)\mathcal{H}$. Direct calculation shows

$$\begin{aligned}\langle \psi_1 | (C - \lambda_{\min} \mathbb{1}_{\mathcal{H}}) | \psi_1 \rangle &= \int_{\sigma(C)} (\lambda - \lambda_{\min}) d\mu_1(\lambda) = \int_{I_1} (\lambda - \lambda_{\min}) d\mu(\lambda) \leq \eta^2, \\ \langle \psi_2 | (\lambda_{\max} \mathbb{1}_{\mathcal{H}} - C) | \psi_2 \rangle &= \int_{\sigma(C)} (\lambda_{\max} - \lambda) d\mu_2(\lambda) = \int_{I_2} (\lambda_{\max} - \lambda) d\mu(\lambda) \leq \eta^2,\end{aligned}$$

where the measures $\mu_i(X) = \langle \psi_i | E(X) | \psi_i \rangle$ are supported on I_i for $i = 1, 2$.

Subsequently, we identify $\mathcal{K} = \text{Span}\{|\psi_1\rangle, |\psi_2\rangle\} \in \text{Gr}(2, \mathcal{H})$ with projector $P_{\mathcal{K}}$. Clearly $C_{\mathcal{K}} = P_{\mathcal{K}} C P_{\mathcal{K}} \in B(\mathcal{K})$ is a two-dimensional self-adjoint operator, so we denote its two eigenvalues by μ_{\min}, μ_{\max} , and $R(\sigma(C_{\mathcal{K}})) = 1/2(\mu_{\max} - \mu_{\min})$. Then, the above two inequalities show that

$$\begin{aligned}\mu_{\min} &\leq \langle \psi_1 | C_{\mathcal{K}} | \psi_1 \rangle = \langle \psi_1 | C | \psi_1 \rangle \leq \lambda_{\min} + \eta^2 \\ \mu_{\max} &\geq \langle \psi_2 | C_{\mathcal{K}} | \psi_2 \rangle = \langle \psi_2 | C | \psi_2 \rangle \geq \lambda_{\max} - \eta^2.\end{aligned}$$

By Stampfli's Theorem 3.1,

$$\inf_{c \in \mathbb{C}} \|C - c \mathbb{1}_{\mathcal{H}}\|_{\text{op}} = R(\sigma(C)) = \frac{1}{2}(\lambda_{\max} - \lambda_{\min}) \leq \frac{1}{2}(\mu_{\max} - \mu_{\min}) + \eta^2 = \inf_{c \in \mathbb{C}} \|C_{\mathcal{K}} - c \mathbb{1}_{\mathcal{K}}\|_{\text{op}} + \eta^2.$$

To upper bound $\inf_{c \in \mathbb{C}} \|C_{\mathcal{K}} - c \mathbb{1}_{\mathcal{K}}\|_{\text{op}}$, we work with the eigenbasis $\{|\mu_{\min}\rangle, |\mu_{\max}\rangle\}$ of $C_{\mathcal{K}}$ associated with $\{\mu_{\min}, \mu_{\max}\}$. In this basis, every unitary $U \in B(\mathcal{K})$ satisfies $\|U |\mu_{\min}\rangle\|_2^2 = |U_{11}|^2 + |U_{21}|^2 = 1$. Moreover, if U is also Haar-random, then the two random variables $|U_{11}|^2$ and $|U_{21}|^2$ are identically distributed. By symmetry it follows that the expectation values $\mathbb{E}|U_{11}|^2 = \mathbb{E}|U_{21}|^2 = \frac{1}{2}$. Then, one checks that

$$\begin{aligned}\mathbb{E}\|C_{\mathcal{K}} - U\|_{\text{op}}^2 &\geq \mathbb{E}\|(C_{\mathcal{K}}U - UC_{\mathcal{K}}) |\mu_{\min}\rangle\|_2^2 \\ &= \mathbb{E}\|C_{\mathcal{K}}U |\mu_{\min}\rangle - U \mu_{\min} |\mu_{\min}\rangle\|_2^2 \\ &= \mathbb{E}\|(C_{\mathcal{K}} - \mu_{\min} \mathbb{1}_{\mathcal{K}})U |\mu_{\min}\rangle\|_2^2 \\ &= \mathbb{E}(|U_{11}|^2(\mu_{\min} - \mu_{\min})^2 + |U_{21}|^2(\mu_{\max} - \mu_{\min})^2) \\ &= 2\frac{1}{4}(\mu_{\max} - \mu_{\min})^2 \\ &= 2R(\sigma(C_{\mathcal{K}}))^2 = 2 \inf_{c \in \mathbb{C}} \|C_{\mathcal{K}} - c \mathbb{1}_{\mathcal{K}}\|_{\text{op}}^2 \geq 2 \inf_{c \in \mathbb{C}} \|C - c \mathbb{1}_{\mathcal{H}}\|_{\text{op}}^2 - \eta^2.\end{aligned}$$

Since $\eta > 0$ is arbitrary, it follows that

$$\begin{aligned}\inf_{c \in \mathbb{C}} \|C - c \mathbb{1}_{\mathcal{H}}\|_{\text{op}} &\leq \inf_{c \in \mathbb{C}} \|C_{\mathcal{K}} - c \mathbb{1}_{\mathcal{K}}\|_{\text{op}} = R(\sigma(C_{\mathcal{K}})), \\ \text{and } 2 \inf_{c \in \mathbb{C}} \|C - c \mathbb{1}_{\mathcal{H}}\|_{\text{op}}^2 &\leq \mathbb{E}\|C_{\mathcal{K}} - U\|_{\text{op}}^2,\end{aligned}\tag{17}$$

using the previous inequality, which will also be used to derive the other upper bound.

On the other hand, Eq. (15) is equivalent to $\Pr_{U \sim \mu_{\mathcal{K}}} \{\|C - U\|_{\text{op}}^2 \leq \varepsilon^2\} \geq 1 - \delta$, which implies

that

$$\begin{aligned}\mathbb{E}\|C_{\mathcal{K}}, U\|_{\text{op}}^2 &= \Pr_{U \sim \mu_{\mathcal{K}}} \{\|C_{\mathcal{K}}, U\|_{\text{op}}^2 \leq \varepsilon^2\} \cdot \mathbb{E}(\|C_{\mathcal{K}}, U\|_{\text{op}}^2 \mid \|C_{\mathcal{K}}, U\|_{\text{op}}^2 \leq \varepsilon^2) \\ &\quad + \Pr_{U \sim \mu_{\mathcal{K}}} \{\|C_{\mathcal{K}}, U\|_{\text{op}}^2 > \varepsilon^2\} \cdot \mathbb{E}(\|C_{\mathcal{K}}, U\|_{\text{op}}^2 \mid \|C_{\mathcal{K}}, U\|_{\text{op}}^2 > \varepsilon^2) \\ &\leq (1 - \delta)\varepsilon^2 + \delta \cdot \|C_{\mathcal{K}}, U\|_{\text{op}}^2 \leq (1 - \delta)\varepsilon^2 + \delta \cdot 4\|C\|_{\text{op}}^2.\end{aligned}$$

The first inequality is justified due to reducing the weight of the smaller conditional expectation ($\leq \varepsilon^2$) while increasing the weight of the larger one ($\geq \varepsilon^2$) can only enlarge the total, and the second one is a basic calculation. It follows from the lower bound Eq. (17) that

$$\inf_{c \in \mathbb{C}} \|C - c\mathbb{1}_{\mathcal{H}}\|_{\text{op}} \leq \frac{1}{\sqrt{2}} \mathbb{E}\|C_{\mathcal{K}}, U\|_{\text{op}} \leq \frac{\sqrt{2}}{2} \left(\sqrt{1 - \delta} \varepsilon + 2\sqrt{\delta} \|C\|_{\text{op}} \right).$$

We now derive the upper bound $\lceil \frac{1}{1-\delta} \rceil \frac{\varepsilon}{2}$ with a Steinhaus-Weil-like argument. Let $A_{\mathcal{K}} = \{U \in U(\mathcal{K}) \mid \|P_{\mathcal{K}} C P_{\mathcal{K}}, U\|_{\text{op}} \leq \varepsilon\}$, then Eq. (15) implies that $\mu_{\mathcal{K}}(A) \geq 1 - \delta$. Also, $A_{\mathcal{K}}^{-1} = A_{\mathcal{K}}^* = A_{\mathcal{K}}$, since for any $U \in A_{\mathcal{K}}$ one has that

$$\|[P_{\mathcal{K}} C P_{\mathcal{K}}, U^{-1}]\|_{\text{op}} = \|U[P_{\mathcal{K}} C P_{\mathcal{K}}, U^*]U\|_{\text{op}} = \|[P_{\mathcal{K}} C P_{\mathcal{K}}, U]\|_{\text{op}}.$$

Let $l = \lceil \frac{1}{1-\delta} \rceil$, we claim that $A_{\mathcal{K}}^l = U(\mathcal{K})$. Indeed, since $U(\mathcal{K})$ is compact and connected, inductive application of Kemperman's theorem [Kem64, Thm. 1.1] implies that

$$\mu_{\mathcal{K}}(A_{\mathcal{K}}^l) \geq \min(1, l \cdot \mu_{\mathcal{K}}(A_{\mathcal{K}})) \geq \lceil \frac{1}{1-\delta} \rceil (1 - \delta) \geq 1 = \mu_{\mathcal{K}}(U(\mathcal{K})),$$

consequently $A_{\mathcal{K}}^l = U(\mathcal{K})$.

Therefore, every $V \in U(\mathcal{K})$ is of the form $V = \prod_{i=1}^l U_i$ for some $U_i \in A_{\mathcal{K}}$. By the definition of $A_{\mathcal{K}}$ and an inductive application of Eq. (10), we have

$$\|C_{\mathcal{K}}, V\|_{\text{op}} = \|C_{\mathcal{K}}, \prod_{i=1}^l U_i\|_{\text{op}} \leq l\varepsilon = \lceil \frac{1}{1-\delta} \rceil \varepsilon. \quad (18)$$

But then, by the original Stampfli's Theorem 3.1 and Eq. (17), we conclude

$$\inf_{c \in \mathbb{C}} \|C - c\mathbb{1}_{\mathcal{H}}\|_{\text{op}} \leq \inf_{c \in \mathbb{C}} \|C_{\mathcal{K}} - c\mathbb{1}_{\mathcal{K}}\|_{\text{op}} = \frac{1}{2} \sup_{V \in U(\mathcal{K})} \|[P_{\mathcal{K}} C P_{\mathcal{K}}, V]\|_{\text{op}} \leq \lceil \frac{1}{1-\delta} \rceil \frac{\varepsilon}{2}.$$

□

Theorem 3.3 (Probabilistic Stampfli's theorem). *Let \mathcal{H} be a Hilbert space of possibly infinite dimension and let $C \in B(\mathcal{H})$. Given $\varepsilon > 0$ and $\delta \in [0, 1]$, suppose that*

$$\Pr_{U \sim \mu_{\mathcal{K}}} \{\|[P_{\mathcal{K}} C P_{\mathcal{K}}, U]\|_{\text{op}} \leq \varepsilon\} \geq 1 - \delta \quad (19)$$

for every subspace $\mathcal{K} \in \text{Gr}(2, \mathcal{H})$ with projector $P_{\mathcal{K}}$. Then

$$\inf_{c \in \mathbb{C}} \|C - c\mathbb{1}_{\mathcal{H}}\|_{\text{op}} \leq \min \left(\sqrt{2} \left(\sqrt{1 - \delta} \varepsilon + 2\sqrt{\delta} \|C\|_{\text{op}} \right), \lceil \frac{1}{1-\delta} \rceil \frac{\varepsilon}{2} \right). \quad (20)$$

Again, the bound $\lceil \frac{1}{1-\delta} \rceil \frac{\varepsilon}{2}$ behaves generally better in the high confidence regime ($\delta \rightarrow 0$) while the bound $\sqrt{2} \left(\sqrt{1-\delta} \varepsilon + 2\sqrt{\delta} \|C\|_{\text{op}} \right)$ is more stable in the low confidence regime ($\delta \rightarrow 1$).

Proof. To derive the first bound $\sqrt{2} \left(\sqrt{1-\delta} \varepsilon + 2\sqrt{\delta} \|C\|_{\text{op}} \right)$, let

$$H = \frac{1}{2}(C + C^*), \quad K = \frac{1}{2i}(C - C^*)$$

be the unique self-adjoint operators such that $C = H + iK$. For any $\mathcal{K} \in \text{Gr}(2, \mathcal{H})$ with projection $P_{\mathcal{K}}$, write $C_{\mathcal{K}} = P_{\mathcal{K}} C P_{\mathcal{K}}$, $H_{\mathcal{K}} = P_{\mathcal{K}} H P_{\mathcal{K}}$, and $K_{\mathcal{K}} = P_{\mathcal{K}} K P_{\mathcal{K}}$. Clearly both $H_{\mathcal{K}}$ and $K_{\mathcal{K}}$ are still self-adjoint. Pick x, y as minimizers such that

$$\begin{aligned} \|H_{\mathcal{K}} - x\mathbb{1}_{\mathcal{K}}\|_{\text{op}} &= \inf_{c \in \mathbb{C}} \|H_{\mathcal{K}} - c\mathbb{1}_{\mathcal{K}}\|_{\text{op}}, \\ \|K_{\mathcal{K}} - y\mathbb{1}_{\mathcal{K}}\|_{\text{op}} &= \inf_{c \in \mathbb{C}} \|K_{\mathcal{K}} - c\mathbb{1}_{\mathcal{K}}\|_{\text{op}}. \end{aligned}$$

(One can check that x, y are actually the average of the largest and smallest eigenvalues of $H_{\mathcal{K}}, K_{\mathcal{K}}$.) Note that both $\|H_{\mathcal{K}}\|_{\text{op}}, \|K_{\mathcal{K}}\|_{\text{op}} \leq \|C_{\mathcal{K}}\|_{\text{op}} \leq \|C\|_{\text{op}}$ by the triangle inequality.

Next, for every unitary $U \in B(\mathcal{K})$ direct calculation shows that

$$\|[C_{\mathcal{K}}^*, U]\|_{\text{op}} = \|U[C_{\mathcal{K}}, U^*]U\|_{\text{op}} = \|[C_{\mathcal{K}}, U]\|_{\text{op}}.$$

Then

$$\|[H_{\mathcal{K}}, U]\|_{\text{op}} \leq \frac{1}{2}(\|[C_{\mathcal{K}}, U]\|_{\text{op}} + \|[C_{\mathcal{K}}^*, U]\|_{\text{op}}) = \|[C_{\mathcal{K}}, U]\|_{\text{op}}$$

and likewise for $\|[K_{\mathcal{K}}, U]\|_{\text{op}} \leq \|[C_{\mathcal{K}}, U]\|_{\text{op}}$. Therefore, for each Haar-random $U \in B(\mathcal{K})$ such that $\|[C_{\mathcal{K}}, U]\|_{\text{op}} \leq \varepsilon$, the same commutator bounds apply to both $H_{\mathcal{K}}, K_{\mathcal{K}}$, i.e.

$$\Pr_{U \sim \mu_{\mathcal{K}}} \{ \|[H_{\mathcal{K}}, U]\|_{\text{op}} \leq \varepsilon \} = \Pr_{U \sim \mu_{\mathcal{K}}} \{ \|[K_{\mathcal{K}}, U]\|_{\text{op}} \leq \varepsilon \} \geq 1 - \delta.$$

It follows from Lem. 3.2 that both

$$\|H - x\mathbb{1}_{\mathcal{K}}\|_{\text{op}}, \|K - y\mathbb{1}_{\mathcal{K}}\|_{\text{op}} \leq \frac{\sqrt{2}}{2} \left(\sqrt{1-\delta} \varepsilon + 2\sqrt{\delta} \|C\|_{\text{op}} \right),$$

which implies

$$\begin{aligned} \inf_{c \in \mathbb{C}} \|C - c\mathbb{1}\|_{\text{op}} &\leq \|(H + iK) - (x + iy)\mathbb{1}\|_{\text{op}} \\ &\leq \|H - x\mathbb{1}\|_{\text{op}} + \|i(K - y\mathbb{1})\|_{\text{op}} \leq \sqrt{2} \left(\sqrt{1-\delta} \varepsilon + 2\sqrt{\delta} \|C\|_{\text{op}} \right). \end{aligned}$$

Finally, the bound $\lceil \frac{1}{1-\delta} \rceil \frac{\varepsilon}{2}$ directly follows from Lem. 3.2 since the proof does not rely on the self-adjointness. \square

Observe that when the commutator is smaller than ε with high confidence ($\delta < 1/2$), we recover the original Stampfli's Theorem 3.1 up to a factor of 2 (ε instead of $\varepsilon/2$), which shows up to compensate for the fact that we need to consider $A_{\mathcal{K}}^2$ (from the proof of Lem. 3.2).

Remark 3.4. One can generalize the setting of Thm. 3.3 to Haar-random unitaries $U \in B(\mathcal{K})$ when $2 \leq \dim(\mathcal{K}) < \infty$, at the cost of having a slightly worse constant factor:

$$\inf_{c \in \mathbb{C}} \|C - c\mathbf{1}_{\mathcal{H}}\|_{\text{op}} \leq \min \left(2\sqrt{2} \left(\sqrt{1-\delta}\varepsilon + 2\sqrt{\delta}\|C\|_{\text{op}} \right), \left\lceil \frac{1}{1-\delta} \right\rceil \frac{\varepsilon}{2} \right). \quad (21)$$

The second bound $\lceil \frac{1}{1-\delta} \rceil \frac{\varepsilon}{2}$ can be shown the same way as that of Lem. 3.2 as the argument only relies on the compactness and connectedness of finite-dimensional unitary groups. We now give a sketch of the proof for the first bound.

For simplicity assume $\mathcal{K} = \mathcal{H}$ and $C = C^* \in B(\mathcal{K})$. Let λ_{\max} be the maximal eigenvalue of C with eigenvector $|\lambda_{\max}\rangle$ and λ_{\min} be the minimal eigenvalue with eigenvector $|\lambda_{\min}\rangle$, and let $R = (\lambda_{\max} - \lambda_{\min})/2$. By the pigeonhole principle, at least $\lceil d/2 \rceil$ of the eigenvalues lie in the interval $[\lambda_{\max} - R, \lambda_{\max}]$ or in $[\lambda_{\min}, \lambda_{\min} + R]$. Without loss of generality we assume the former so that there are $\geq d/2$ of them are in $[\lambda_{\max} - R, \lambda_{\max}]$.

In the eigenbasis $\{|\lambda_{\min}\rangle, \dots, |\lambda_{\max}\rangle\}$ of C , the vector $U|\lambda_{\min}\rangle$ is the first column of a Haar-random unitary $U \in U(d)$. Hence by the same symmetry argument that $\mathbb{E}|U_{i1}|^2 = 1/d$ for each i . A direct calculation shows

$$\begin{aligned} \mathbb{E}\| [C, U] \|_{\text{op}}^2 &\geq \mathbb{E}\| (C - \lambda_{\min}\mathbf{1})U|\lambda_{\min}\rangle \|_2^2 = \mathbb{E} \sum_i |\lambda_i - \lambda_{\min}|^2 |U_{i1}|^2 \\ &\geq \mathbb{E} \sum_{\lambda_i \in [\lambda_{\max}-R, \lambda_{\max}]} |\lambda_i - \lambda_{\min}|^2 |U_{i1}|^2 \geq \frac{d}{2} R^2 \frac{1}{d} = \frac{R^2}{2}. \end{aligned}$$

The exact same proof then leads to $\sqrt{2}$ factor for the self-adjoint case and consequently $2\sqrt{2}$ for the general case.

3.2 Doubly probabilistic Stampfli's theorem

While Thm. 3.3 is a proper generalization of the original Stampfli's theorem, we note that the Haar-random single-qubit assumption is still not physical enough. Indeed, it requires verifications of almost commutation over all single-qubit subspaces $\mathcal{K} \in \text{Gr}(2, \mathcal{H})$, which is unrealistic.

We therefore consider a doubly probabilistic generalization: also randomly sample two-dimensional subspaces $\mathcal{K} \in \text{Gr}(2, \mathcal{H})$ and then check the almost commutation for Haar-random unitaries in $B(\mathcal{K})$. This is far more reasonable in physical implementations.

However, the random sampling of two-dimensional subspaces in infinite-dimensional \mathcal{H} does not make sense. In fact, it is well-known that the Grassmannian $\text{Gr}(2, \mathcal{H})$ does not admit a non-trivial, σ -finite, $U(\mathcal{H})$ -invariant Borel measure when $\dim(\mathcal{H}) = \infty$. Hence, we consider the finite-dimensional setting for the doubly probabilistic generalization.

Now, $\text{Gr}(2, \mathcal{H})$ does admit a probability measure $\nu_{\text{Gr}(2, \mathcal{H})}$ when $\dim(\mathcal{H}) = d < \infty$. In particular, the notion of a random two-dimensional subspace \mathcal{K} is equivalent to the following:

- (a) Fix two orthonormal vectors $|v_1\rangle, |v_2\rangle \in \mathcal{H}$ as the reference two-dimensional subspace.
- (b) There exists some Haar-random $U(d)$ -unitary $V \in B(\mathcal{H})$ such that $\mathcal{K} = V\text{Span}\{|v_1\rangle, |v_2\rangle\}$.

Thanks to the invariance of Haar measures, $|v_1\rangle$ and $|v_2\rangle$ can be chosen arbitrarily. This allows us to formulate and prove another generalization of Stampfli's theorem.

Theorem 3.5 (Doubly probabilistic Stampfli's theorem). *Let \mathcal{H} be a d -dimensional Hilbert space and let $C \in B(\mathcal{H})$. Given $\varepsilon > 0$ and $\delta, \eta \in [0, 1]$, suppose that*

$$\Pr_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} \left\{ \Pr_{U \sim \mu_{\mathcal{K}}} \{ \| [P_{\mathcal{K}} C P_{\mathcal{K}}, U] \|_{\text{op}} \leq \varepsilon \} \geq 1 - \delta \right\} \geq 1 - \eta, \quad (22)$$

where $P_{\mathcal{K}}$ denotes the projection onto \mathcal{K} . Then

$$\inf_{c \in \mathbb{C}} \|C - c \mathbf{1}_{\mathcal{H}}\|_{\text{op}} \leq 2\sqrt{\frac{d^2 - 1}{6}} \min \left(\sqrt{(1 - \eta)(1 - \delta)} \varepsilon + 2\|C\|_{\text{op}} \sqrt{\delta(1 - \eta) + \eta}, \right. \\ \left. \sqrt{1 - \eta} \left\lceil \frac{1}{1 - \delta} \right\rceil \varepsilon + 2\|C\|_{\text{op}} \sqrt{\eta} \right). \quad (23)$$

and this upper bound necessarily depends on the dimension d . In addition, the leading factor can be reduced to $\sqrt{(d^2 - 1)}/6$ when C is self-adjoint.

Depending on the specific values of $(\varepsilon, \delta, \eta, \|C\|_{\text{op}})$, one bound might be tighter than the other.

Proof. Let us begin with the first bound $2\sqrt{\frac{d^2 - 1}{6}} \left(\sqrt{(1 - \eta)(1 - \delta)} \varepsilon + 2\|C\|_{\text{op}} \sqrt{\delta(1 - \eta) + \eta} \right)$. It is sufficient to show the case when C is self-adjoint, as the general case follows by the same argument in the proof of Thm. 3.3. Analogous to Lem. 3.2, here we instead try to bound $\mathbb{E}_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} \mathbb{E}_{U \sim \mu_{\mathcal{K}}} \| [C_{\mathcal{K}}, U] \|_{\text{op}}$ for $C_{\mathcal{K}} = P_{\mathcal{K}} C P_{\mathcal{K}}$.

By the definition of expectation values and the trivial commutator bound, the upper bound is straightforward:

$$\mathbb{E}_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} \mathbb{E}_{U \sim \mu_{\mathcal{K}}} \| [C_{\mathcal{K}}, U] \|_{\text{op}}^2 \leq (1 - \eta) \cdot ((1 - \delta)\varepsilon^2 + 4\delta\|C\|_{\text{op}}^2) + \eta \cdot 4\|C\|_{\text{op}}^2. \quad (24)$$

For the lower bound, Eq. (17) already shows that

$$\mathbb{E}_{U \sim \mu_{\mathcal{K}}} \| [C_{\mathcal{K}}, U] \|_{\text{op}}^2 \geq 2R(\sigma(C_{\mathcal{K}}))^2,$$

where $R(\sigma(C_{\mathcal{K}}))$ is the radius of the spectrum of $C_{\mathcal{K}}$. Thus, the rest of the proof amounts to calculating $\mathbb{E}_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} R(\sigma(C_{\mathcal{K}}))^2$.

We first calculate $R(\sigma(C_{\mathcal{K}}))^2$. To this end, denote by λ_i the eigenvalues of C with the corresponding eigenvectors $|\lambda_i\rangle$. By the discussion preceding the theorem, there exists some Haar-random $U(d)$ -unitary $V \in B(\mathcal{H})$ such that $\mathcal{K} = V \text{Span}\{|\lambda_1\rangle, |\lambda_2\rangle\}$. It follows from $C = \sum_k \lambda_k |\lambda_k\rangle \langle \lambda_k|$ that

$$(C_{\mathcal{K}})_{ij} = \langle \lambda_i | V^* C V | \lambda_j \rangle = \sum_k \lambda_k \bar{V}_{ki} V_{kj}$$

in the basis $\{V|\lambda_1\rangle, V|\lambda_2\rangle\}$. Then

$$R(\sigma(C_{\mathcal{K}}))^2 = \frac{1}{4} (\text{Tr}(C_{\mathcal{K}})^2 - 4 \det(C_{\mathcal{K}})) \\ = \sum_{k, l} \lambda_k \lambda_l (\bar{V}_{k1} V_{k1} \bar{V}_{l1} V_{l1} + \bar{V}_{k2} V_{k2} \bar{V}_{l2} V_{l2} - 2\bar{V}_{k1} V_{k1} \bar{V}_{l2} V_{l2} + 4\bar{V}_{k1} V_{k2} \bar{V}_{l2} V_{l1}),$$

using the convenient trace-determinant formula for 2×2 matrices.

To compute $\mathbb{E}_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} (V_{k_1 a_1} V_{k_2 a_2} \bar{V}_{l_1 b_1} \bar{V}_{l_2 b_2})$ for a Haar-random unitary $V \in B(\mathcal{H})$, we can use

Weingarten calculus [CS06, Cor. 2.4]. One checks that

$$\begin{aligned}\mathbb{E}_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} R(\sigma(C_{\mathcal{K}}))^2 &= \frac{1}{4} \sum_{k,l} \lambda_k \lambda_l \frac{6\delta_{kl} - 6}{d(d^2 - 1)} \\ &= \frac{3}{2d(d^2 - 1)} \left(d \text{Tr}(C^2) - \text{Tr}(C)^2 \right) \geq \frac{3}{(d^2 - 1)} R(\sigma(C))^2.\end{aligned}$$

The last inequality with factor $(d^2 - 1)$ is in fact sharp. To see this, observe that $(d \text{Tr}(C^2) - \text{Tr}(C)^2)/d^2$ is the variance of $\{\lambda_i\}$ with uniform distribution. Given a fixed $R(\sigma(C)) = (\lambda_{\max} - \lambda_{\min})/2$, the variance is minimized when all non-extremal eigenvalues $\lambda_i = (\lambda_{\max} + \lambda_{\min})/2$, whence $d \text{Tr}(C^2) - \text{Tr}(C)^2 = 2dR(\sigma(C))^2$. Another sanity check is to consider $d = 2$ which gives the trivial constant 1.

It follows that

$$\mathbb{E}_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} \mathbb{E}_{U \sim \mu_{\mathcal{K}}} \| [C_{\mathcal{K}}, U] \|_{\text{op}}^2 \geq \frac{3R(\sigma(C))^2}{(d^2 - 1)},$$

and we are done by Eq. (24).

Finally, we show the second bound $\sqrt{1 - \eta} \left\lceil \frac{1}{1 - \delta} \right\rceil \varepsilon + 2\|C\|_{\text{op}} \sqrt{\eta}$. Recall Eq. (18) from the proof of Lem. 3.2, every $V \in U(\mathcal{K})$ satisfies that $\|[C_{\mathcal{K}}, V]\|_{\text{op}} \leq \left\lceil \frac{1}{1 - \delta} \right\rceil \varepsilon$. That is,

$$\Pr_{\mathcal{K} \sim \nu_{\text{Gr}(2, \mathcal{H})}} \left\{ \Pr_{V \sim \mu_{\mathcal{K}}} \{ \|[P_{\mathcal{K}} C P_{\mathcal{K}}, V]\|_{\text{op}} \leq \left\lceil \frac{1}{1 - \delta} \right\rceil \varepsilon \} \geq 1 - 0 \right\} \geq 1 - \eta.$$

Then we are done by setting $\varepsilon \rightarrow \left\lceil \frac{1}{1 - \delta} \right\rceil \varepsilon$ and $\delta \rightarrow 0$ into the first bound. \square

Observe that if one has the power to verify the ε -almost commutation directly over the unitary group $U(\mathcal{H})$ of \mathcal{H} beyond just the two-dimensional qubit unitaries, then Rem. 3.4 applies and the conclusion of Thm. 3.5 simplifies to Eq. (21).

3.3 Approximate Tsirelson's theorem from Haar-random single-qubit unitary

We refer to the method of randomly sampling a single-qubit subspace $\mathcal{K} \in \text{Gr}(2, \mathcal{H})$ and then certifying commutation with Haar-random $U(2)$ -unitaries in $B(\mathcal{K})$ as *Haar-random single-qubit unitary* sampling. Thus, with the doubly probabilistic Stampfli's theorem as in Thm. 3.5, we can analogously formulate and prove an approximate version of Tsirelson's theorem, resulting in another version of the quantitative Ozawa's result [Oza13b].

Lemma 3.6. *Consider two Hilbert spaces \mathcal{H}_1 with dimension d_1 and \mathcal{H}_2 with dimension d_2 , and let $C \in B(\mathcal{H}_1 \otimes \mathcal{H}_2)$. Given $\varepsilon > 0$ and $\delta, \eta \in [0, 1]$, suppose that*

$$\Pr_{\mathcal{K} \sim \mu_{\text{Gr}(2, \mathcal{H}_2)}} \left\{ \Pr_{U \sim \mu_{\mathcal{K}}} \{ \|[(\mathbf{1}_1 \otimes P_{\mathcal{K}}) C (\mathbf{1}_1 \otimes P_{\mathcal{K}}), \mathbf{1}_1 \otimes U] \|_{\text{op}} \leq \varepsilon \} \geq 1 - \delta \right\} \geq 1 - \eta, \quad (25)$$

where $P_{\mathcal{K}}$ denotes the projection onto $\mathcal{K} \subset \mathcal{H}_2$.

Then $C' = 1/d_2 \text{Tr}_{\mathcal{H}_2}(C) \in B(\mathcal{H}_1)$ satisfies

$$\|C - C' \otimes \mathbb{1}_2\|_{\text{op}} \leq 2d_1 \sqrt{\frac{d_2^2 - 1}{6}} \min \left(\sqrt{(1-\eta)(1-\delta)} \varepsilon + 2\|C\|_{\text{op}} \sqrt{\delta(1-\eta) + \eta}, \right. \\ \left. \sqrt{1-\eta} \left\lceil \frac{1}{1-\delta} \right\rceil \varepsilon + 2\|C\|_{\text{op}} \sqrt{\eta} \right). \quad (26)$$

Consequently, if C is positive semidefinite then so is C' . Note the leading constant $2d_1$ can be improved to d_1 when C is self-adjoint.

Proof. This proof is almost identical to that of Lem. 2.2. Adopting the same notation, we point out the only difference: Thm. 3.5 implies that

$$\|C_{(kl)} - c_{kl} \mathbb{1}_2\|_{\text{op}} \leq 2 \sqrt{\frac{d_2^2 - 1}{6}} \min \left(\sqrt{(1-\eta)(1-\delta)} \varepsilon + 2\|C\|_{\text{op}} \sqrt{\delta(1-\eta) + \eta}, \right. \\ \left. \sqrt{1-\eta} \left\lceil \frac{1}{1-\delta} \right\rceil \varepsilon + 2\|C\|_{\text{op}} \sqrt{\eta} \right),$$

where we use the fact that $\|C\|_{\text{op}} \geq \|C_{(kl)}\|_{\text{op}}$. Hence,

$$\|C - C' \otimes \mathbb{1}_2\|_{\text{op}} \leq d_1 \max_{k,l} \|C_{(kl)} - c_{kl} \mathbb{1}_2\|_{\text{op}}$$

gives the desired bound. Note that the above inequality is sharp when all c_{kl} are the same, meaning that the dimension scaling d_1 is also unavoidable. \square

For the non-simple version of approximate Tsirelson's theorem, similarly to Thm. 2.5, one needs to assume additional commutator bounds with the simple projections.

Theorem 3.7 (Approximate Tsirelson's theorem, Haar-random unitary case). *Let \mathcal{A} be generated by finitely-many contractive self-adjoint operators $\{A_{a|x}\} \subset B(\mathcal{H})$ and \mathcal{B} be generated by finitely-many contractive self-adjoint operators $\{B_{b|y}\} \subset B(\mathcal{H})$ for some d -dimensional Hilbert space \mathcal{H} . Suppose that \mathcal{A} admits the Artin-Wedderburn decomposition*

$$\mathcal{A} = \bigoplus_{l=1}^L \mathcal{A}_l \simeq \bigoplus_{l=1}^L B(\mathcal{H}_A^l) \otimes \mathbb{1}_B^l \text{ and } A_{a|x} = \bigoplus_{l=1}^L A_{a|x}^l \otimes \mathbb{1}_B^l,$$

with the corresponding orthogonal projectors Π_l onto the direct summands.

Given $\varepsilon > 0$ and $\delta, \eta \in [0, 1]$. Suppose that,

$$\Pr_{\mathcal{K}^l \sim \nu_{\text{Gr}(2, \mathcal{H}_A^l)}} \left\{ \Pr_{U \sim \mu_{\mathcal{K}^l}} \left\{ \|[[U_l \otimes \mathbb{1}_B^l, (P_{\mathcal{K}^l} \otimes \mathbb{1}_B^l) \Pi_l B_{b|y} \Pi_l (P_{\mathcal{K}^l} \otimes \mathbb{1}_B^l)]]\|_{\text{op}} \leq \varepsilon \right\} \geq 1 - \delta \right\} \geq 1 - \eta, \quad (27)$$

where $P_{\mathcal{K}^l}$ denotes the projection onto $\mathcal{K}^l \subset \mathcal{H}_A^l$. Furthermore, assume that for all b, y, l

$$\|[\Pi_l, B_{b|y}]\|_{\text{op}} \leq \varepsilon. \quad (28)$$

Then there exist operators $B'_{b|y} \in \bigoplus_{l=1}^L \mathbb{1}_A^l \otimes B(\mathcal{H}_B^l) = \mathcal{A}'$ such that, for all b, y ,

$$\|B_{b|y} - B'_{b|y}\|_{\text{op}} \leq d(d-1)\varepsilon + d\sqrt{\frac{d^2-1}{6}} \min\left(\sqrt{(1-\eta)(1-\delta)}\varepsilon + 2\sqrt{\delta(1-\eta)+\eta}, \sqrt{1-\eta} \left\lceil \frac{1}{1-\delta} \right\rceil \varepsilon + 2\sqrt{\eta}\right). \quad (29)$$

In addition, if $B_{b|y}$ is positive then so is $B'_{b|y}$. Note that the bound has an $O(d^2\varepsilon)$ scaling.

Proof. The proof is analogous to that of Thm. 2.5 so we only give a sketch. Since $B_{b|y}$ is self-adjoint and $\|B_{b|y}\|_{\text{op}} \leq 1$, by Lem. 3.6, there exists for each l some positive $B_{b|y}^l \in B(\mathcal{H}_B^l)$ such that

$$\|\Pi_l B_{b|y} \Pi_l - \mathbb{1}_A^l \otimes B_{b|y}^l\|_{\text{op}} \leq d_l \sqrt{\frac{d_l^2-1}{6}} \min\left(\sqrt{(1-\eta)(1-\delta)}\varepsilon + 2\|C\|_{\text{op}}\sqrt{\delta(1-\eta)+\eta}, \sqrt{1-\eta} \left\lceil \frac{1}{1-\delta} \right\rceil \varepsilon + 2\|C\|_{\text{op}}\sqrt{\eta}\right),$$

where $d_l = \dim(\mathcal{H}_A^l \otimes \mathcal{H}_B^l)$. The commutation assumption of $B_{b|y}$ with Π_l implies

$$\|B_{b|y} - \sum_l \Pi_l B_{b|y} \Pi_l\|_{\text{op}} \leq L(L-1)\varepsilon \leq d(d-1)\varepsilon.$$

Define $B'_{b|y} := \bigoplus_{l=1}^L \mathbb{1}_A^l \otimes B_{b|y}^l$, and we are done by the triangle inequality. \square

4 Applications and outlook

Our main results, Thm. 2.5 and Thm. 3.7, establish quantitative approximate versions of Tsirelson's theorem from two distinct perspectives on almost commutation. The first approach (Thm. 2.5) provides an error guarantee contingent on potentially hard-to-determine algebraic complexity parameters c_1, c_2, c_3 (discussed in Rem. 2.6), scaling roughly as $O(c_1^2 c_2^2 c_3^2 d^2 \varepsilon)$. The second, doubly probabilistic formulation (Thm. 3.7) yields a bound scaling as $O(d^2 \varepsilon)$ with prefactors dependent on probabilistic confidence parameters (δ, η) of random unitary sampling and assumptions about commutation with simple-block projectors. This offers a trade-off, making the latter potentially advantageous when algebra generation is difficult, but probabilistic checks are feasible. Both methods confirm the overall $O(d^2 \varepsilon)$ error bounds, and these discussions can be generalized to multipartite scenarios by induction, akin to the original Tsirelson's theorem.

In this last section, we begin with an application to quantum information theory, as teased in Sec. 1.2: the construction of an approximating tensor-product strategy (Prop. 4.1) based on our main theorems. Next, we discuss the implications of our results in the context of the NPA hierarchy and the known computational complexity results, giving scenarios with non-negligible approximation error (Rem. 4.2). Then, we connect our works to the broader historical context on approximating almost commuting matrices with genuinely commuting ones (Sec. 4.3). Lastly, we finish by discussing the implications of our work and outlining future possible research directions.

4.1 Constructing tensor-product approximation

For clarity, a *tensor-product quantum strategy* is a triple $(\{A_{a|x}\}, \{B_{b|y}\}, \rho)$ on a Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, where for each x the family $\{A_{a|x}\} \subset B(\mathcal{H}_A)$ is a POVM (i.e., $A_{a|x} \geq 0$ and $\sum_a A_{a|x} = \mathbb{1}_A$), for each y the family $\{B_{b|y}\} \subset B(\mathcal{H}_B)$ is a POVM, and $\rho \in B(\mathcal{H})$ is a positive operator with $\text{Tr}(\rho) = 1$. The associated *correlations* are

$$p(ab|xy) = \text{Tr}(\rho \cdot (A_{a|x} \otimes B_{b|y})).$$

By contrast, a *commuting operator quantum strategy* is a triple $(\{A_{a|x}\}, \{B_{b|y}\}, \rho)$ on a single Hilbert space \mathcal{H} , where $\{A_{a|x}\}, \{B_{b|y}\} \subset B(\mathcal{H})$ are POVMs such that $[A_{a|x}, B_{b|y}] = 0$ for all a, b, x, y with correlations

$$p(ab|xy) = \text{Tr}(\rho \cdot A_{a|x} \cdot B_{b|y}).$$

As recalled in Sec. 1.2, these two models are equivalent when \mathcal{H} is finite-dimensional by Tsirelson's theorem.

In what follows, we relax exact commutation to ε -almost commutation (in the senses made precise in Theorems 2.5 and 3.7) and show how our results yield a *constructive* procedure: from any d -dimensional ε -almost commuting strategy, one can build a tensor-product strategy whose correlations approximate the original ones up to $O(d^2\varepsilon)$.

Proposition 4.1. *Given a quantum strategy $(A_{a|x}, B_{b|y}, \rho)$ on a d -dimensional Hilbert space \mathcal{H} that is ε -almost commuting (in the sense of Thms. 2.5 or 3.7). Then there exist Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B$ and a tensor-product strategy $(\tilde{A}_{a|x}, \tilde{B}_{b|y}, \tilde{\rho})$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ such that its correlations are $O(d^2\varepsilon)$ -close to those of the original ε -almost commuting strategy $(A_{a|x}, B_{b|y}, \rho)$.*

Proof. The Artin-Wedderburn decomposition of the algebra generated by $\{A_{a|x}\}$ implies that $\mathcal{H} = \bigoplus_l \mathcal{H}_A^l \otimes \mathcal{H}_B^l$ with

$$A_{a|x} = \bigoplus_l A_{a|x}^l \otimes \mathbb{1}_B^l \in \bigoplus_l B(\mathcal{H}_A^l) \otimes \mathbb{1}_B^l.$$

Let Π_l be the orthogonal projectors to the l -th summand $B(\mathcal{H}_A^l) \otimes \mathbb{1}_B^l$. By Thm. 2.5 or 3.7, for all b, y , the operators $B_{b|y}$ can be approximated within $O(d^2\varepsilon)$ in operator norm by positive operators

$$B'_{b|y} = \bigoplus_l \mathbb{1}_A^l \otimes \frac{1}{d_A^l} \text{Tr}_{\mathcal{H}_A^l}(\Pi_l B_{b|y} \Pi_l) \in \bigoplus_l \mathbb{1}_A^l \otimes B(\mathcal{H}_B^l).$$

They satisfy $\sum_b B'_{b|y} = \mathbb{1}$ since the normalized partial trace is unital and $\sum_b B_{b|y} = \mathbb{1}$, hence forming POVMs. This yields an exactly commuting strategy $(A_{a|x}, B'_{b|y}, \rho)$ on \mathcal{H} which is $O(d^2\varepsilon)$ -close to the original strategy.

We construct the equivalent tensor-product strategy $(\tilde{A}_{a|x}, \tilde{B}_{b|y}, \tilde{\rho})$ as follows: Let $\mathcal{H}_A = \bigoplus_l \mathcal{H}_A^l$ and $\mathcal{H}_B = \bigoplus_l \mathcal{H}_B^l$. Define new operators

$$\tilde{A}_{a|x} = \bigoplus_l A_{a|x}^l \in B(\mathcal{H}_A), \quad \tilde{B}_{b|y} = \bigoplus_l \frac{1}{d_A^l} \text{Tr}_{\mathcal{H}_A^l}(\Pi_l B_{b|y} \Pi_l) \in B(\mathcal{H}_B),$$

and the new state

$$\tilde{\rho} = \iota(\rho) \in B(\mathcal{H}_A \otimes \mathcal{H}_B)$$

via the natural embedding $\iota : B(\bigoplus_l \mathcal{H}_A^l \otimes \mathcal{H}_B^l) \rightarrow B(\mathcal{H}_A \otimes \mathcal{H}_B)$. Thanks to the block structure of $\tilde{\rho}$, one can directly check that the correlations of $(A_{a|x}, B'_{b|y}, \rho)$ are preserved by $(\tilde{A}_{a|x}, \tilde{B}_{b|y}, \tilde{\rho})$, consequently $O(d^2\varepsilon)$ -close to those of the original strategy. \square

We note that our constructions recover the standard Tsirelson's theorem asymptotically as $\varepsilon \rightarrow 0$, indicating our result is a quantitative version of [Oza13b].

4.2 NPA hierarchy and computational complexity

We comment on an interesting consequence of our approximate Tsirelson's theorem in relation to the NPA hierarchy [NPA08; PNA10], an important tool in the studies of quantum correlations. This connection has implications for understanding when the approximation error from our theorem must necessarily be significant.

Remark 4.2. *The NPA hierarchy provides a sequence of constraints, indexed by level N , that characterize correlations arising from commuting observable strategies. This hierarchy is complete in the limit $N \rightarrow \infty$. At a finite level N , an NPA strategy S_N can be realized in a d_N -dimensional Hilbert space and involves observables that are $O(1/\sqrt{N})$ -almost commuting [CV15, Thm. 23].*

Our Thm. 2.5 states that such an $O(1/\sqrt{N})$ -almost commuting strategy S_N can be approximated by a genuine tensor-product strategy with an operator norm error of $O(d_N^2/\sqrt{N})$. We argue that this error term cannot always vanish as $N \rightarrow \infty$ due to computational complexity arguments.

1. *Consider the result $\text{MIP}^* = \text{RE}$ [Ji+21]. This implies there are problems (specifically, RE-hard problems) for which the closure of the set of correlations achievable with tensor-product strategies (C_{qa}) is strictly smaller than the set achievable with commuting observable strategies (C_{qc}), i.e., $C_{qa} \subsetneq C_{qc}$. The NPA strategies S_N generate correlations that converge towards C_{qc} . Our approximation, being a tensor-product strategy, generates correlations within C_{qa} .*

Hence, the approximation error $O(d_N^2/\sqrt{N})$ must be generally non-vanishing in the limit $N \rightarrow \infty$. If not, i.e., the error vanished, it would imply C_{qa} could approximate C_{qc} arbitrarily well, contradicting the known set separation.

2. *A similar line of reasoning applies to the conjecture $\text{MIP}^{\text{co}} = \text{coRE}$ [Ji+21] (more precisely, the gapped decision problem of quantum commuting value is coRE-hard). If this conjecture holds, it would imply the existence of coRE-hard problems where $O(1/\sqrt{N})$ -almost commuting strategies S_N can achieve outcomes (e.g., Bell scores) significantly larger than those achievable by any strictly commuting observable strategy (and thus, by any tensor-product strategy). In such a scenario, these S_N strategies would be inherently “far” from any tensor-product approximation. Consequently, our approximation of S_N by a tensor-product strategy must necessarily result in a non-vanishing error of $O(d_N^2/\sqrt{N})$ to account for this performance gap.*

In essence, these complexity results highlight scenarios where the distinction between almost-commuting and strictly commuting (or tensor-product) models is presented. Our quantitative theorems provide a bound on how well one can bridge this distinction, and these complexity results suggest that our error bound, or indeed any such bound, cannot universally tend to zero.

4.3 Relation to prior works on almost commuting matrices

The question of whether matrices or operators that almost commute are necessarily close to a genuinely commuting pair is a longstanding problem with a rich history, initiated by Rosenthal [Ros69] for the normalized Hilbert-Schmidt norm and Halmos [Hal76] for the operator norm. Early studies

in the operator norm (e.g., [LT70; PS79]) often yielded affirmative answers, though typically with dimension-dependent error bounds.

The search for dimension-independent bounds revealed a crucial dichotomy in the operator norm. Voiculescu [Voi83] showed that almost commuting unitary matrices need not be close to commuting ones (incidentally by considering clock-and-shift matrices Σ_3, Σ_1), and Choi [Cho88] extended this negative result to general matrices. In contrast, Lin’s theorem [FR96; Lin96] provided a positive dimension-independent answer for a pair of self-adjoint matrices. Recently, extending the scope to infinite dimensions and multiple operators, [Lin24] has connected the approximability of self-adjoint operators to spectral properties.

In parallel, the searches for dimension-independent bounds in the normalized Hilbert-Schmidt norm established affirmative results for a pair of normal [Gle10], self-adjoint [FK10], and unitary matrices [HS18]. More recently, Ioana [Ioa24] further confirms approximability if at least one matrix is normal, while showing a negative result for general matrices.

Our work contributes to this area by considering two (thus, inductively, multiple) finite-dimensional C^* -algebras whose generators almost commute, formulated either in terms of operator norm bounds against specific matrix generators (like clock-and-shift matrices) or via a probabilistic formulation involving Haar-random unitaries. For both, we characterize how close these algebras are to having genuinely commuting counterparts (or admitting an approximate tensor-product structure by Prop. 4.1) in the operator norm, deriving bounds that exhibit dependence on the dimension d . Given the discussion in Rem 4.2 based on known separations due to computational complexity results [Ji+21], we do not expect such dimension dependence in the error bounds to be removable.

4.4 Discussions and future directions

Fundamentally, Tsirelson’s theorem connects the tensor-product formalism with the commuting observable formalism for composite finite-dimensional quantum systems. While strict commutation can be conceptually enforced by space-like separation, many physical scenarios or experimental setups might only guarantee approximate independence due to correlated noise, imperfect isolation, or other constraints. Our approximate Tsirelson’s theorems (Thm. 2.5 and Thm. 3.7) show that Tsirelson’s conclusion is robust to such imperfections. They guarantee that ε -almost commuting observables (in either the deterministic or probabilistic sense) necessarily imply that the system’s correlations are $O(d^2\varepsilon)$ -close in operator norm to those of genuine tensor-product quantum correlations. This validates the use of tensor-product formulation as an effective model even when subsystem independence is only approximately satisfied.


As detailed above in Rem. 4.2, our findings interface with the NPA hierarchy. This connection is crucial, as it highlights, through computational complexity results like $\text{MIP}^* = \text{RE}$ [Ji+21], that for certain problems the approximation error $O(d_N^2/\sqrt{N})$ from our theorems cannot be universally negligible. This signifies a fundamental limitation in approximating certain almost-commuting strategies with tensor-product strategies, a limitation our quantitative error bounds necessarily reflect. Conversely, for scenarios without this intrinsic separation, improving our error bounds remains interesting for applications like robust self-testing [ŠB20].

Furthermore, our probabilistic Stampfli’s theorems (Thm. 3.3 and Thm. 3.5) open possibilities beyond Tsirelson’s problem itself. It is natural to explore probabilistic commutation hypotheses in intrinsically infinite-dimensional settings, for instance, by sampling random unitaries using frameworks like free probability theory [MS17]. Viewing Stampfli’s theorem as a generalization of Schur’s lemma, another promising direction involves investigating whether our probabilistic versions can lead to analogous generalizations of other consequences of Schur’s lemma, such as probabilistic Schur-Weyl duality, which could in turn lead to probabilistic formulations of quantum de Finetti

theorems [Ren08].

Acknowledgments

We thank an anonymous reader whose suggestions reduced the d^2 factor in Lemma 2.1 to d and inspired our use of Kemperman’s theorem in developing new versions of probabilistic Stampfli’s theorems. The authors thank Scott McCullough for providing feedback on the manuscript.

X.X. and M.-O.R. acknowledge funding from the INRIA and the CIEDS in the Action Exploratoire project DEPARTURE. X.X. and M.-O.R. acknowledge funding from the ANR through the JCJC grant LINKS (ANR-23-CE47-0003) and T-ERC QNET (ANR-24-ERCS-0008). I.K. was supported by the Slovenian Research Agency program P1-0222 and grants J1-50002, N1-0217, J1-3004, J1-50001, J1-60011, J1-60025. Partially supported by the Fondation de l’École polytechnique as part of the Gaspard Monge Visiting Professor Program. IK thanks École Polytechnique and Inria for hospitality during the preparation of this manuscript. X.X., M.-O.R., and I.K. acknowledge funding from the European Union’s Horizon 2020 Research and Innovation Programme under QuantERA Grant Agreement no. 731473 and 101017733 .

References

- [App05] Marcus Appleby. “Symmetric informationally complete–positive operator valued measures and the extended Clifford group”. In: *Journal of Mathematical Physics* 46.5 (Apr. 2005), p. 052107. ISSN: 0022-2488. DOI: 10.1063/1.1896384. eprint: <https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.1896384/14813871/052107\1\online.pdf>. URL: <https://doi.org/10.1063/1.1896384>.
- [BH74] Joseph J Bastian and Kenneth J Harrison. “Subnormal weighted shifts and asymptotic properties of normal operators”. In: *Proceedings of the American Mathematical Society* 42.2 (1974), pp. 475–479.
- [Bel64] John S Bell. “On the Einstein Podolsky Rosen paradox”. In: *Physics Physique Fizika* 1.3 (1964), p. 195.
- [Bla06] Bruce Blackadar. *Operator Algebras: Theory of C*-Algebras and von Neumann Algebras*. Encyclopaedia of Mathematical Sciences. Springer Berlin Heidelberg, 2006. ISBN: 978-3-540-28517-5. DOI: 10.1007/978-3-540-28517-5.
- [Cho88] Man Duen Choi. “Almost commuting matrices need not be nearly commuting”. In: *Proceedings of the American Mathematical Society* 102.3 (1988), pp. 529–533.
- [CS06] Benoît Collins and Piotr Śniady. “Integration with respect to the Haar measure on unitary, orthogonal and symplectic group”. In: *Communications in Mathematical Physics* 264.3 (2006), pp. 773–795.
- [Con76] Alain Connes. “Classification of injective factors cases II 1, II $^\infty$, III λ , $\lambda \neq 1$ ”. In: *Annals of Mathematics* 104.1 (1976), pp. 73–115.
- [CV15] Matthew Coudron and Thomas Vidick. “Interactive proofs with approximately commuting provers”. In: *Automata, Languages, and Programming: 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I* 42. Springer. 2015, pp. 355–366.

- [Doh+08] Andrew C Doherty, Yeong-Cherng Liang, Ben Toner, and Stephanie Wehner. “The quantum moment problem and bounds on entangled multi-prover games”. In: *2008 23rd Annual IEEE Conference on Computational Complexity*. IEEE. 2008, pp. 199–210.
- [FK10] Nikolay Filonov and Ilya Kachkovskiy. “A Hilbert-Schmidt analog of Huaxin Lin’s Theorem”. In: *arXiv preprint arXiv:1008.4002* (2010).
- [FR96] Peter Friis and Mikael Rordam. “Almost commuting self-adjoint matrices - a short proof of Huaxin Lin’s theorem.” In: *Journal für die reine und angewandte Mathematik* 479 (1996), pp. 121–132. URL: <http://eudml.org/doc/153857>.
- [Gle10] Lev Glebsky. “Almost commuting matrices with respect to normalized Hilbert-Schmidt norm”. In: *arXiv preprint arXiv:1002.3082* (2010).
- [HS18] Don Hadwin and Tatiana Shulman. “Stability of group relations under small Hilbert-Schmidt perturbations”. In: *Journal of Functional Analysis* 275.4 (2018), pp. 761–792.
- [Hal76] Paul R Halmos. “Some unsolved problems of unknown depth about operators on Hilbert space”. In: *Proceedings of the Royal Society of Edinburgh Section A: Mathematics* 76.1 (1976), pp. 67–76.
- [Ioa24] Adrian Ioana. “Almost commuting matrices and stability for product groups”. In: *Journal of the European Mathematical Society* (2024). In press.
- [Ji+21] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. “MIP*=RE”. In: *Communications of the ACM* 64.11 (2021), pp. 131–138.
- [Kem64] Johannes Kemperman. “On products of sets in a locally compact group”. eng. In: *Fundamenta Mathematicae* 56.1 (1964), pp. 51–68. URL: <http://eudml.org/doc/213825>.
- [Kir93] Eberhard Kirchberg. “On non-semisplit extensions, tensor products and exactness of group C*-algebras”. In: *Inventiones mathematicae* 112 (1993), pp. 449–489.
- [KŠ16] Igor Klep and Špela Špenko. “Sweeping words and the length of a generic vector subspace of $M_n(F)$ ”. In: *Journal of Combinatorial Theory, Series A* 143 (2016), pp. 56–65.
- [Lan17] Klaas Landsman. *Foundations of quantum theory: From classical concepts to operator algebras*. Springer Nature, 2017.
- [Lin96] Huaxin Lin. “Almost commuting selfadjoint matrices and applications”. In: *Operator algebras and their applications* (1996), pp. 193–233.
- [Lin24] Huaxin Lin. “Almost commuting self-adjoint operators and measurements”. In: *arXiv preprint arXiv:2401.04018* (2024).
- [LT70] Wilhelmus AJ Luxemburg and RF Taylor. “Almost commuting matrices are near commuting matrices”. In: *Indagationes mathematicae (proceedings)*. Vol. 73. North-Holland. 1970, pp. 96–98.
- [MS17] James A Mingo and Roland Speicher. *Free probability and random matrices*. Vol. 35. Springer, 2017.
- [NPA08] Miguel Navascués, Stefano Pironio, and Antonio Acín. “A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations”. In: *New Journal of Physics* 10.7 (2008), p. 073013.
- [NC10] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

- [Oza13a] Narutaka Ozawa. “About the Connes embedding conjecture: algebraic approaches”. In: *Japanese Journal of Mathematics* 8.1 (2013), pp. 147–183.
- [Oza13b] Narutaka Ozawa. “Tsirelson’s problem and asymptotically commuting unitary matrices”. In: *Journal of Mathematical Physics* 54.3 (Mar. 2013), p. 032202. ISSN: 0022-2488. DOI: 10.1063/1.4795391. eprint: https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.4795391/16053470/032202\1_online.pdf. URL: <https://doi.org/10.1063/1.4795391>.
- [Pas19] James E Pascoe. “An elementary method to compute the algebra generated by some given matrices and its dimension”. In: *Linear Algebra and its Applications* 571 (2019), pp. 132–142.
- [Paz84] Azaria Paz. “An application of the Cayley-Hamilton theorem to matrix polynomials in several variables”. In: *Linear and Multilinear Algebra* 15.2 (1984), pp. 161–170.
- [PS79] Carl Pearcy and Allen Shields. “Almost commuting matrices”. In: *Journal of Functional Analysis* 33.3 (1979), pp. 332–338.
- [PNA10] Stefano Pironio, Miguel Navascués, and Antonio Acin. “Convergent relaxations of polynomial optimization problems with noncommuting variables”. In: *SIAM Journal on Optimization* 20.5 (2010), pp. 2157–2180.
- [Ren08] Renato Renner. “Security of quantum key distribution”. In: *International Journal of Quantum Information* 6.01 (2008), pp. 1–127.
- [Ros69] Peter Rosenthal. “Are almost commuting matrices near commuting matrices?” In: *The American Mathematical Monthly* 76.8 (1969), pp. 925–926.
- [SW08] Volkher B Scholz and Reinhard F Werner. “Tsirelson’s problem”. In: *arXiv preprint arXiv:0812.4305* (2008).
- [Shi19] Yaroslav Shitov. “An improved bound for the lengths of matrix algebras”. In: *Algebra Number Theory* 13.6 (2019), pp. 1501–1507.
- [Sta70] Joseph Stampfli. “The norm of a derivation”. In: *Pacific journal of mathematics* 33.3 (1970), pp. 737–747.
- [ŠB20] Ivan Šupić and Joseph Bowles. “Self-testing of quantum systems: a review”. In: *Quantum* 4, Article 337 (Sept. 2020). ISSN: 2521-327X. DOI: 10.22331/q-2020-09-30-337. URL: <https://doi.org/10.22331/q-2020-09-30-337>.
- [Tsi06] Boris S Tsirelson. *Bell inequalities and operator algebras*. <http://web.archive.org/web/20100904054200/http://www.imaph.tu-bs.de/qi/problems/33.html>. Problem statement from the website of open problems at TU Braunschweig. 2006.
- [Voi83] Dan Voiculescu. “Asymptotically commuting finite rank unitary operators without commuting approximants”. In: *Acta Sci. Math.(Szeged)* 45.1-4 (1983), pp. 429–431.