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

Xiangling Xu[1*], Marc-Olivier Renou[1,2,3], Igor Klep[4,5,6]

[1]Inria Paris-Saclay, Bâtiment Alan Turing, 1 rue Honoré d'Estienne d'Orves, 91120 Palaiseau, France
[2]CPHT, Ecole polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau, France
[3]LIX, Ecole polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau, France
[4]Faculty of Mathematics and Physics, University of Ljubljana
[5]Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska
[6]Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia

[*]xu.xiangling@inria.fr

Abstract

Tsirelson showed that, in finite dimensions, quantum correlations generated by commuting observables—measurements associated with distinct parties whose operators mutually commute—are equivalent to those obtainable from measurements on separate tensor product factors. We generalize this foundational result to the setting of ε-almost commuting observables, establishing two distinct quantitative *approximate Tsirelson's theorems*. Both theorems show that if a d-dimensional bipartite quantum strategy's observables ε-almost commute, then they are within $O(\text{poly}(d)\varepsilon)$ (in operator norm) of observables from a genuine tensor product strategy. This provides a quantitative counterpart to the asymptotic result of [N. Ozawa, *J. Math. Phys.* 54, 032202 (2013)] and justifies the tensor product model as an effective model even when subsystem independence is only approximately satisfied.

Our theorems arise from two different but complementary formulations of almost commutation:
(i) The first approach utilizes deterministic operator norm bounds relative to specific matrix generators (such as clock and shift matrices), leading to an approximate Schur's Lemma from which the first theorem directly follows.
(ii) The second approach employs probabilistic bounds, requiring small commutators only on average against Haar-random single-qubit unitaries. This method yields two novel probabilistic Stampfli's theorems, quantifying distance to scalars based on probabilistic commutation, a result which may be of independent interest. These theorems set the basis for the second approximate Tsirelson's theorem.

1 Introduction

Quantum information theory offers two different axioms for composing independent subsystems. The standard composition axiom [NC10] postulates that the Hilbert space $\mathcal{H}_R$ 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 [Tsi06; SW08; Doh+08], formally:

Theorem (Tsirelson's theorem). *Let $\{A_{a|x}\}, \{B_{b|y}\} \subset B(\mathcal{H})$ be two finite sets of mutually commuting positive operator-valued measures (POVMs) acting on a finite-dimensional Hilbert space $\mathcal{H}$. Then there exists a decomposition of $\mathcal{H}$ into a direct sum of tensor product spaces, $\mathcal{H} \simeq \bigoplus_l (\mathcal{H}_A^l \otimes \mathcal{H}_B^l)$, such that operators $A_{a|x}$ and $B_{b|y}$ take the form*

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

where $A_{a|x}^l \in B(\mathcal{H}_A^l)$ and $B_{b|y}^l \in B(\mathcal{H}_B^l)$. Consequently, any correlations obtained from measurements of operators $A_{a|x}$ and $B_{b|y}$ on any state ρ on $\mathcal{H}$ can be reproduced using operators $\tilde{A}_{a|x} \in B(\bigoplus_l \mathcal{H}_A^l)$ and $\tilde{B}_{b|y} \in B(\bigoplus_l \mathcal{H}_B^l)$ acting on a state $\tilde{\rho}$ on the Hilbert space $(\bigoplus_l \mathcal{H}_A^l) \otimes (\bigoplus_l \mathcal{H}_B^l)$. The statement can be inductively generalized to multipartite cases.

Tsirelson conjectured that the above theorem can be generalized to infinite dimensions. However, it was recently shown by [Ji+21] that the commuting operator model can produce correlations unattainable by any tensor product quantum strategy, with far-reaching implications including a disproof of Connes' embedding conjecture [Con76] and Kirchberg's conjecture [Kir93]. See [Oza13a] for a survey.

1.1 Contributions

This work generalizes Tsirelson's theorem to the physically relevant setting in which finite-dimensional observables only *almost commute*. We present two distinct quantitative generalizations (Thm. 2.6 and 3.7), which establish bounds on how close almost-commuting finite-dimensional operator algebras are to having a tensor product structure. As a direct consequence, we provide a constructive argument (Prop. 4.1) showing that if a d-dimensional bipartite quantum strategy is ε-almost commuting, its correlations can be approximated by those of a genuine tensor product strategy up to an error of $O(\text{poly}(d)\varepsilon)$.

These results provide quantitative counterparts to the asymptotic result of Ozawa [Oza13b], showcasing the robustness of Tsirelson's argument and justifying the use of the tensor product model even when subsystem independence is only approximately satisfied. They also connect to the broader study of approximating almost commuting operators with genuinely commuting ones [Ros69; Hal76; LT70; BH74; PS79; Cho88; Voi83; Lin96; FR96; Gle10; FK10; Ioa24; Lin24]. Analogous to the original Tsirelson's theorem, our results can be straightforwardly extended to multipartite cases.

Our two generalizations stem from different perspectives on almost commutation. This first approach uses the operator norm bounds derived from commutators with the chosen full matrix generators (e.g., Sylvester's clock and shift matrices). This method yields error bounds that depend on the difficulty (quantified by constants c_1, c_2, c_3) of expressing these chosen matrix generators as polynomials in the interested operators. The second employs a probabilistic approach, considering commutation properties against randomly sampled unitaries, resulting in bounds characterized by probabilistic parameters (δ, η) alongside an assumption on the simple block projectors. A key insight from the second method is the development of new variations of Stampfli's theorem [Sta70]: a probabilistic Stampfli's theorem (Thm. 3.3) and a doubly probabilistic Stampfli's theorem(Thm. 3.5). These theorems quantify how close an operator must be to a scalar multiple of the identity given its probabilistic commutation behavior.

1.2 Structure of the paper

We derive approximate Tsirelson's theorems from different but complementary notions of almost commutation.

Sec. 2: *Clock-and-shift method*. This section develops the first generalization based on deterministic commutator bounds with respect to matrix generators.

1. We define almost commutation using Sylvester's clock and shift matrices Σ_1, Σ_3 (introduced in Sec. 2.1), which are generalizations of Pauli matrices σ_1, σ_3 to arbitrary dimensions. Analogous results can be derived for any other full matrix generator sets as well.

2. We derive an approximate Schur lemma (Lem. 2.1) and its bipartite version (Lem. 2.3), quantitatively showing that an operator almost commuting with clock and shift matrices must be close to a scalar (or an operator on the other tensor factor in the bipartite case).

3. Under assumptions on how efficiently the clock and shift matrices can be expressed using the initial operators (quantified by algebraic complexity constants c_1, c_2, c_3), the approximate Schur's lemma leads to our first approximate Tsirelson's theorem (Thm. 2.6). We also discuss the scaling error $O(\mathrm{poly}(d)\varepsilon)$ error scaling involving factors c_1, c_2, c_3 (Rem. 2.7).

Sec. 3: *Haar-random single-qubit unitary method*. This section establishes a probabilistic formulation of almost commutation.

1. We replace uniform commutator bounds with probabilistic ones, requiring small commutators only for *most* Haar-random single-qubit unitaries within two-dimensional subspaces. Note that one may instead consider d-dimensional subspaces for $d \geq 2$ (Rem. 3.4).

2. This leads to a probabilistic Stampfli's theorem (Thm. 3.3), relating these probabilistic bounds to the operator's distance from scalars. We extend this further 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 version (Thm. 3.5) holds only in finite dimensions.

3. Using the doubly probabilistic Stampfli's theorem, we derive its bipartite version (Lem. 3.6) and prove our second approximate Tsirelson's theorem (Thm. 3.7) under both probabilistic assumptions on intra-simple-block commutation (Eq. (25)) and bounds on the commutators with the associated simple block projectors (Eq. (26)). Again we achieve an $O(\mathrm{poly}(d)\varepsilon)$ error guarantee.

Sec. 4: *Applications and outlook*.

1. First, We detail the construction of an approximating tensor product strategy based on our main theorems (Prop. 4.1).

2. Second, we explore the interplay between our results, the NPA hierarchy, and computational complexity, elucidating scenarios with non-vanishing approximation error (Rem. 4.2).

3. Third, we situate our results within the rich line of research on approximating almost commuting matrices (Sec. 4.3).

4. Lastly, the section offers a broader discussion on the overall significance of our work and outlines possible future research directions, including the further development and application of the probabilistic Stampfli's theorems.

2 Sylvester's clock and shift formulation

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 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 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. (Note that any full generating set would still work, and our choice is mostly due to the fact that clock and shift matrices can be seen as generalized Pauli matrices.)

With these generators we prove an approximate Schur's lemma (Lem. 2.1) and its bipartite version (Lem. 2.3). Next, we show an approximate Tsirelson's theorem for simple algebras (Lem. 2.5), and then by the same block-decomposition argument, a general approximate Tsirelson's theorem (Thm. 2.6). We finish with a discussion of scaling (Rem. 2.7).

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 finite-dimensional analogs of position and momentum operators in finite-dimensional quantum systems.

Let $\omega = e^{2\pi i/d}$ be the dth root of unity. Using Dirac's notation, denote by $\{|i\rangle \mid i = 0, \ldots, d-1\}$ 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 is defined by $\Sigma_3 : |i\rangle \mapsto \omega^i |i\rangle$. Or, more explicitly:

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

The notation comes from the fact that the shift matrix Σ_3 (resp. clock matrix Σ_1) is a generalization of 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

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

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

$$\{\sigma_{k,l} := \Sigma_1^k \Sigma_3^l = \sum_{i=0}^{d-1} \omega^{il} |i+k\rangle\langle i|\}_{0 \le k,l \le d-1}, \tag{3}$$

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

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 $\|\cdot\|_{\mathrm{op}}$ for the operator (spectral) norm and $\|\cdot\|_{\max}$ for the max norm, i.e., $\|A\|_{\max} = \max_{i,j}|A_i j|$.

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 $i = 1, 3$,*

$$\|[C, \Sigma_i]\|_{\mathrm{op}} \leq \varepsilon. \tag{4}$$

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

$$\|C - c\mathbb{1}\|_{\mathrm{op}} \leq d\|C - c\mathbb{1}\|_{\max} \leq d^2\varepsilon. \tag{5}$$

Roughly, since Σ_1, Σ_3 together generate the whole $B(\mathcal{H})$, this means that the "approximate center" of $B(\mathcal{H})$ can be approximated by the center of $B(\mathcal{H})$, which is formed by scalars.

Proof. For convenience, let us first rewrite the assumption in the max norm:

$$\|C\Sigma_i - \Sigma_i C\|_{\max} \leq \|[C, \Sigma_i]\|_{\mathrm{op}} \leq \varepsilon,$$

for $i = 1, 3$.

Observe that $C\Sigma_3$ (resp. $\Sigma_3 C$) multiplies the ith column (resp. ith row) of C by ω^{i-1}. Then the assumption implies that $|1 - \omega^{i-j}||C_{ij}| \leq \varepsilon$. Thus, for all off-diagonal terms we have

$$|C_{ij} - 0| \leq 2\varepsilon.$$

In addition, 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 $|C_{i,i} - C_{i-1,i-1}| \leq \varepsilon$, which means that, for all i, j,

$$|C_{ii} - C_{jj}| \leq d\varepsilon,$$

by the triangular inequality.

Finally, taking $c = 1/d\,\mathrm{Tr}(C)$, where $\mathrm{Tr}(C)$ is the trace of C, we have $|c - C_{ii}| \leq d\varepsilon$ and so

$$\|C - c\mathbb{1}\|_{\mathrm{op}} \leq d\|C - c\mathbb{1}\|_{\max} \leq d^2\varepsilon.$$

□

Remark 2.2. *We note that the $O(d^2)$-scaling in Lem. 2.1 and in the subsequent discussion can be reduced to $O(d)$-scaling by having more constraints on the powers of the shift matrix Σ_1, namely,*

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

for all $n = 1, \ldots, d-1$. Similarly, one could also consider commutator bounds for matrix bases different from the clock and shift matrices, for instance, the standard matrix basis $E_{ij} = |i\rangle\langle j|$.

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

Lemma 2.3. *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*

$$\|[C, \mathbb{1}_1 \otimes \Sigma_1]\|_{\mathrm{op}},\ \|[C, \mathbb{1}_1 \otimes \Sigma_3]\|_{\mathrm{op}} \leq \varepsilon \tag{6}$$

for both $\Sigma_1, \Sigma_3 \in B(\mathcal{H}_2)$. Then the matrix $C' = 1/d_2 \operatorname{Tr}_{\mathcal{H}_2}(C) \in B(\mathcal{H}_1)$, where $\operatorname{Tr}_{\mathcal{H}_2}$ denotes the partial trace $B(\mathcal{H}_1 \otimes \mathcal{H}_2) \to B(\mathcal{H}_1)$, satisfies

$$\|C - C' \otimes \mathbb{1}_2\|_{\mathrm{op}} \leq d_1 d_2^2 \varepsilon. \tag{7}$$

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

Proof. Note that

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

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

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

that are block matrices with only Σ_i on the diagonal. Thus by direct calculation, the condition $\|[C, \mathbb{1}_1 \otimes \Sigma_i]\|_{\mathrm{op}} \leq \varepsilon$ implies that, for all k, l,

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

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

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

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

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

Hence,

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

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

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

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 [Tak79, Ch. I §11].

Lemma 2.4. *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 $\mathcal{A} \simeq B(\mathcal{H}_1) \otimes \mathbb{1}_2$.

The structural results give 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 2.5 (Approximate Tsirelson's theorem, simple 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,*

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

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 exists operators $B'_{b|y} \in B(\mathcal{H}_B)$ such that, for all b, y,

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

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

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

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

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

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

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

and we are done by Lem. 2.3. □

Remark that the "contraction" requirement is not necessary and one can reproduce the same result by replacing ε by $\varepsilon/\|A_{a|x}\|_{\mathrm{op}}$. The simple version can be readily generalized to the finite-dimensional case with Lem. 2.4.

Theorem 2.6 (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 $\{B_{b|y}\} \subset B(\mathcal{H})$ in some d-dimensional Hilbert space $\mathcal{H}$. Assume that there exists an $\varepsilon > 0$, such that for all a, b, x, y,*

$$\|[A_{a|x}, B_{b|y}]\|_{\mathrm{op}} \leq \varepsilon. \tag{10}$$

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 $l = 1, \ldots, L$ 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,

$$\|B_{b|y} - B'_{b|y}\|_{\mathrm{op}} \leq 2c_1 c_2 c_3 \left(c_1 c_2 c_3 + 1\right) d^2 \varepsilon. \tag{11}$$

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

Proof. First, we wish to apply Lem. 2.5 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

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

and

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

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

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

Therefore, by Lem. 2.5, there exists for each l some positive semidefinite $B^l_{b|y} \in B(\mathcal{H}^l_B)$ such that

$$\|\Pi_l B_{b|y} \Pi_l - \mathbb{1}^l_A \otimes B^l_{b|y}\|_{\mathrm{op}} \leq c_1 c_2 c_3 (2c_1 c_2 c_3 + 1) d_l^2 \varepsilon,$$

where $d_l = \dim(\mathcal{H}^l_A \otimes \mathcal{H}^l_B)$.

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

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

where the completeness and orthogonality of Π_l are used.

Finally, one sees that

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

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

Remark 2.7. *We finish the section with some comments on the scaling factor from Eq.* (11)*.*

1. *The generating polynomial degree c_2 is related to the length of algebras with known dependence to 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 [KŠ16].*

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.*

4. *As already discussed in Rem. 2.2, one can instead impose constraints on all powers of the shift matrix Σ_1 to improve the d^2 to d. Also, recall that the choice of clock and shift formulation is not mandatory. Thus, it might be of advantage to work with different matrix basis, such as the standard matrix basis $E_{ij} = |i\rangle\langle j|$, depending on the specific example.*

Overall, the coefficients c_1, c_2, c_3 are example-specific, and the upper bound in Eq. (11) *scales as* $O(\mathrm{poly}(d)\varepsilon)$.

3 Haar-random single-qubit unitary formulation

For a matrix generator independent formulation, one might look 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*

$$\sup_{B\in B(\mathcal{H}):\, \|B\|_{\mathrm{op}}=1} \|[C,B]\|_{\mathrm{op}} = \inf_{c\in\mathbb{C}} 2\|C - c\mathbb{1}\|_{\mathrm{op}}. \tag{12}$$

Moreover, if C is a normal operator, then

$$\inf_{c\in\mathbb{C}} \|C - c\mathbb{1}\|_{\mathrm{op}} = R(\sigma(C)), \tag{13}$$

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 $\|[C,B]\|_{\mathrm{op}} \le \varepsilon$ for all B of norm 1, then C is $\varepsilon/2$-close to some scalar operator $c\mathbb{1}$ in operator norm. Note that $\|[C,U]\|_{\mathrm{op}} \le \varepsilon$ for all unitaries U is equivalent to the assumption that $\|[C,B]\|_{\mathrm{op}} \le \varepsilon$ 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 $\|B\|_{\mathrm{op}} = 1$ 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.

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). Specifically, 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 ($\le \varepsilon$) with U is high ($\ge 1 - \delta$).

For notational convenience, from now on denote by $\mathrm{Gr}(2, \mathcal{H})$ 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 self-adjoint decomposition.

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 $\varepsilon, \delta > 0$, suppose that*

$$\Pr_{U\sim\mu_{\mathcal{K}}}\{\|[P_{\mathcal{K}}CP_{\mathcal{K}}, U]\|_{\mathrm{op}} \leq \varepsilon\} \geq 1-\delta \tag{14}$$

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

$$\inf_{c\in\mathbb{C}}\|C - c\mathbb{1}_{\mathcal{H}}\|_{\mathrm{op}} \leq \frac{\sqrt{2}}{2}\left(\sqrt{1-\delta}\,\varepsilon + 2\sqrt{\delta}\|C\|_{\mathrm{op}}\right). \tag{15}$$

Proof. The central object to bound is $\mathbb{E}\|[P_{\mathcal{K}}CP_{\mathcal{K}}, U]\|_{\mathrm{op}}^2$—while the upper bound is straightforward, the lower bound requires more work. The main idea is to identify $\mathcal{K} = \mathrm{Span}\{|\psi_1\rangle, |\psi_2\rangle\}$. 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 $\|[P_{\mathcal{K}}CP_{\mathcal{K}}, U]\|_{\mathrm{op}}$ by $2\inf_{c\in\mathbb{C}}\|C - c\mathbb{1}_{\mathcal{H}}\|_{\mathrm{op}}^2$ 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 its 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

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

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

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

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 $\mathcal{K} = \mathrm{Span}\{|\psi_1\rangle, |\psi_2\rangle\} \in \mathrm{Gr}(2, \mathcal{H})$ with projector $P_{\mathcal{K}}$. Clearly $C_{\mathcal{K}} = P_{\mathcal{K}}CP_{\mathcal{K}} \in B(\mathcal{K})$ is a two-dimensional self-adjoint operator, so we denote its two eigenvalues by

$\mu_{\min}, \mu_{\max}$, 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,

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

To upper bound $\inf_{c \in \mathbb{C}} \| C_{\mathcal{K}} - c\mathbb{1}_{\mathcal{K}} \|_{\mathrm{op}}$, 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 $\mathbb{E}|U_{11}|^2 = \mathbb{E}|U_{21}|^2 = 1/2$. Then, one checks that

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

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

$$2 \inf_{c \in \mathbb{C}} \| C - c\mathbb{1}_{\mathcal{H}} \|_{\mathrm{op}}^2 \leq 2R(\sigma(C_{\mathcal{K}}))^2 \leq \mathbb{E}\| [C_{\mathcal{K}}, U] \|_{\mathrm{op}}^2, \tag{16}$$

using the previous inequality.

On the other hand, Eq. (14) is equivalent to $\Pr_{U \sim \mu_{\mathcal{K}}} \{ \| [C, U] \|_{\mathrm{op}}^2 \leq \varepsilon^2 \} \geq 1 - \delta$, which implies that

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

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. (16) that

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

$\square$

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 $\varepsilon, \delta > 0$, suppose that*

$$\Pr_{U \sim \mu_{\mathcal{K}}} \{\|[P_{\mathcal{K}} C P_{\mathcal{K}}, U]\|_{\mathrm{op}} \leq \varepsilon\} \geq 1 - \delta \tag{17}$$

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

$$\inf_{c \in \mathbb{C}} \|C - c\mathbb{1}\|_{\mathrm{op}} \leq \sqrt{2}\left(\sqrt{1-\delta}\,\varepsilon + 2\sqrt{\delta}\|C\|_{\mathrm{op}}\right). \tag{18}$$

Proof. Let

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

be the unique self-adjoint operators such that $C = H + iK$. For any $\mathcal{K} \in \mathrm{Gr}(2, \mathcal{H})$ 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

$$\|H_{\mathcal{K}} - x\mathbb{1}_{\mathcal{K}}\|_{\mathrm{op}} = \inf_{c \in \mathbb{C}} \|H_{\mathcal{K}} - c\mathbb{1}_{\mathcal{K}}\|_{\mathrm{op}},$$
$$\|K_{\mathcal{K}} - y\mathbb{1}_{\mathcal{K}}\|_{\mathrm{op}} = \inf_{c \in \mathbb{C}} \|K_{\mathcal{K}} - c\mathbb{1}_{\mathcal{K}}\|_{\mathrm{op}}.$$

(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 $\|H_{\mathcal{K}}\|_{\mathrm{op}}, \|K_{\mathcal{K}}\|_{\mathrm{op}} \leq \|C_{\mathcal{K}}\|_{\mathrm{op}} \leq \|C\|_{\mathrm{op}}$ by the triangle inequality.

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

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

Then

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

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

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

It follows from Lem. 3.2 that both

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

which implies

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

□

It is possible that the constant coefficient $\sqrt{2}$ can be improved further considering its asymptotic counterpart, Thm. 3.1, has the factor $1/2$.

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

$$\inf_{c \in \mathbb{C}} \|C - c\mathbb{1}\|_{\mathrm{op}} \le 2\sqrt{2}\left(\sqrt{1-\delta}\,\varepsilon + 2\sqrt{\delta}\|C\|_{\mathrm{op}}\right). \tag{19}$$

We now give a sketch of the proof.

For simplicity assume $\mathcal{K} = \mathcal{H}$ and $C = C^* \in B(\mathcal{K})$. Let $\lambda_{\max}$ be the maximal eigenvalue of C with eigenvector $|\lambda_{\max}\rangle$ and $\lambda_{\min}$ 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 $\ge d/2$ of them are in $[\lambda_{\max} - R, \lambda_{\max}]$.

In the eigenbasis $\{|\lambda_{\min}\rangle, \ldots, |\lambda_{\max}\rangle\}$ 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

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

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 $\mathcal{K} \in \mathrm{Gr}(2, \mathcal{H})$, which is unrealistic.

We therefore consider a doubly probabilistic generalization: also randomly sample two-dimensional subspaces $\mathcal{K} \in \mathrm{Gr}(2, \mathcal{H})$ 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 Grassmanian $\mathrm{Gr}(2, \mathcal{H})$ does not admit a non-trivial, σ-finite, $U(\mathcal{H})$-invariant Borel measure when $\dim(H) = \infty$. Hence, we consider the finite-dimensional setting for the doubly probabilistic generalization.

On the other hand, $\mathrm{Gr}(2, \mathcal{H})$ does admit a probability measure $\nu_{\mathrm{Gr}(2,\mathcal{H})}$ when $\dim(H) = d < \infty$. 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 $\mathcal{K} = V\mathrm{Span}\{|v_1\rangle, |v_2\rangle\}$.

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 $\varepsilon, \delta, \eta > 0$, suppose that*

$$\Pr_{\mathcal{K} \sim \nu_{\mathrm{Gr}(2,\mathcal{H})}} \left\{ \Pr_{U \sim \mu_{\mathcal{K}}} \{\|[P_{\mathcal{K}} C P_{\mathcal{K}}, U]\|_{\mathrm{op}} \le \varepsilon\} \ge 1 - \delta \right\} \ge 1 - \eta, \tag{20}$$

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

$$\inf_{c\in\mathbb{C}}\|C-c\mathbb{1}\|_{\mathrm{op}} \leq 2\sqrt{\frac{d^2-1}{6}}\left(\sqrt{(1-\eta)(1-\delta)}\,\varepsilon + 2\|C\|_{\mathrm{op}}\sqrt{\delta(1-\eta)+\eta}\right), \tag{21}$$

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.

Proof. 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 $\mathbb{E}_{\mathcal{K}\sim\nu_{\mathrm{Gr}(2,\mathcal{H})}}\mathbb{E}_{U\sim\mu_{\mathcal{K}}}\|[C_{\mathcal{K}},U]\|_{\mathrm{op}}$ for $C_{\mathcal{K}}=P_{\mathcal{K}}CP_{\mathcal{K}}$.

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

$$\mathbb{E}_{\mathcal{K}\sim\nu_{\mathrm{Gr}(2,\mathcal{H})}}\mathbb{E}_{U\sim\mu_{\mathcal{K}}}\|[C_{\mathcal{K}},U]\|_{\mathrm{op}}^2 \leq (1-\eta)\cdot\left((1-\delta)\varepsilon^2+4\delta\|C\|_{\mathrm{op}}^2\right)+\eta\cdot 4\|C\|_{\mathrm{op}}^2. \tag{22}$$

For the lower bound, Eq. (16) already shows that

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

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 $\mathbb{E}_{\mathcal{K}\sim\nu_{\mathrm{Gr}(2,\mathcal{H})}}R(\sigma(C_{\mathcal{K}}))$.

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 $\mathcal{K}=V\,\mathrm{Span}\{|\lambda_1\rangle,|\lambda_2\rangle\}$. It follows from $C=\sum_k\lambda_k|\lambda_k\rangle\langle\lambda_k|$ that

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

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

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

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

To compute $\mathbb{E}_{\mathcal{K}\sim\nu_{\mathrm{Gr}(2,\mathcal{H})}}(V_{k_1a_1}V_{k_2a_2}\bar{V}_{l_1b_1}\bar{V}_{l_2b_2})$ for a Haar-random unitary $V\in B(\mathcal{H})$, we can use Weingarten calculus [CŚ06, Cor. 2.4]. One may check that

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

The last inequality with factor (d^2-1) is in fact sharp. To see this, observe that $(d\,\mathrm{Tr}\big(C^2\big)-\mathrm{Tr}(C)^2)/d^2$ 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 $d\,\mathrm{Tr}\big(C^2\big)-\mathrm{Tr}(C)^2=2dR(\sigma(C))$.

It follows that $\mathbb{E}_{\mathcal{K}\sim\nu_{\mathrm{Gr}(2,\mathcal{H})}}\mathbb{E}_{U\sim\mu_{\mathcal{K}}}\|[C_{\mathcal{K}},U]\|_{\mathrm{op}}^2\geq 3R(\sigma(C))^2/(d^2-1)$, and we are done by Eq. (22).

$\square$

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

We refer to the method of randomly sampling a single-qubit subspace $\mathcal{K} \in \mathrm{Gr}(2, \mathcal{H})$ 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 another version of 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 $\varepsilon, \delta, \eta > 0$, suppose that*

$$\Pr_{\mathcal{K} \sim \mu_{\mathrm{Gr}(2,\mathcal{H}_2)}} \left\{ \Pr_{U \sim \mu_{\mathcal{K}}} \{ \| [(\mathbb{1}_1 \otimes P_{\mathcal{K}}) C (\mathbb{1}_1 \otimes P_{\mathcal{K}}), \mathbb{1}_1 \otimes U] \|_{\mathrm{op}} \le \varepsilon \} \ge 1 - \delta \right\} \ge 1 - \eta, \tag{23}$$

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

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

$$\|C - C' \otimes \mathbb{1}_2\|_{\mathrm{op}} \le 2 d_1 \sqrt{\frac{d_2^2 - 1}{6}} \left(\sqrt{(1-\eta)(1-\delta)}\, \varepsilon + 2 \|C\|_{\mathrm{op}} \sqrt{\delta(1-\eta) + \eta} \right). \tag{24}$$

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. 2.3. Adopting the same notation, we point out the only difference: Thm. 3.5 implies that

$$\|C_{(kl)} - c_{kl} \mathbb{1}_2\|_{\mathrm{op}} \le 2 \sqrt{\frac{d_2^2 - 1}{6}} \left(\sqrt{(1-\eta)(1-\delta)}\, \varepsilon + 2 \|C\|_{\mathrm{op}} \sqrt{\delta(1-\eta) + \eta} \right),$$

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

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

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. 2.6, 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 contractive self-adjoint operators $\{A_{a|x}\} \subset B(\mathcal{H})$ and $\mathcal{B}$ be generated by contractive self-adjoint operators $\{B_{b|y}\} \subset B(\mathcal{H})$ in 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 to the direct summands.

Given $\varepsilon, \delta, \eta > 0$. Suppose that,

$$\Pr_{\mathcal{K}^l \sim \nu_{\mathrm{Gr}(2,\mathcal{H}_A^l)}} \left\{ \Pr_{U \sim \mu_{\mathcal{K}^l}} \{ \| [[U_l \otimes \mathbb{1}_B^l, (P_{\mathcal{K}^l} \otimes \mathbb{1}_B^l) \Pi_l B_{b|y} \Pi_l (P_{\mathcal{K}^l} \otimes \mathbb{1}_B^l)] \|_{\mathrm{op}} \leq \varepsilon \} \geq 1 - \delta \right\} \geq 1 - \eta, \quad (25)$$

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

$$\| [\Pi_l, B_{b|y}] \|_{\mathrm{op}} \leq \varepsilon. \quad (26)$$

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,

$$\| B_{b|y} - B'_{b|y} \|_{\mathrm{op}} \leq d(d-1)\varepsilon + d\sqrt{\frac{d^2-1}{6}} \left(\sqrt{(1-\eta)(1-\delta)}\, \varepsilon + 2\sqrt{\delta(1-\eta)+\eta} \right). \quad (27)$$

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

Proof. The proof is analogous to Thm. 2.6 so we only give a sketch. Since $B_{b|y}$ is self-adjoint and $\|B_{b|y}\|_{\mathrm{op}} \leq 1$, by Lem. 3.6, there exists for each l some positive $B_{b|y}^l \in B(\mathcal{H}_B^l)$ such that

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

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

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

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

4 Applications and outlook

Our main results, Thm. 2.6 and Thm. 3.7, establish quantitative approximate versions of Tsirelson's theorem from two distinct perspectives on almost commutation. The first approach (Thm. 2.6) provides an error guarantee contingent on potentially hard-to-determine algebraic complexity parameters c_1, c_2, c_3 (discussed in Rem. 2.7), 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 $O(\mathrm{poly}(d)\varepsilon)$ 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 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

We now explain how our results provide a method to construct a tensor product quantum strategy from a finite-dimensional, almost commuting quantum strategy, such that the tensor correlation approximates the almost commuting one.

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 Thm. 2.6 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 $O(\mathrm{poly}(d)\varepsilon)$-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 i-th summand $B(\mathcal{H}_A^l) \otimes \mathbb{1}_B^l$. By Thm. 2.6 or 3.7, for all b, y, the operators $B_{b|y}$ can be approximated (within $O(\mathrm{poly}(d)\varepsilon)$ in operator norm) by positive operators

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

However, $B'_{b|y}$ may not satisfy $\sum_b B'_{b|y} = \mathbb{1}_B$ exactly. This can be fixed: for every input y, choose a specific output b_0. Define the operators $B'_{b|y}$ normally for $b \neq b_0$ and set

$$B'_{b_0|y} = \mathbb{1}_B - \sum_{b \neq b_0} B'_{b|y}.$$

It is straightforward to verify that $B'_{b_0|y}$ remains positive and is $O(\mathrm{poly}(d)\varepsilon)$-close to $B_{b_0|y}$, since all the other $B'_{b|y}$ are close to $B_{b|y}$.

This yields an exactly commuting strategy $(A_{a|x}, B'_{b|y}, \rho)$ on $\mathcal{H}$ which is $O(\mathrm{poly}(d)\varepsilon)$-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

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

and the new state

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

via the natural embedding $\iota : \bigoplus_l \mathcal{H}_A^l \otimes \mathcal{H}_B^l \to \mathcal{H}_A \otimes \mathcal{H}_B$. 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 $O(\mathrm{poly}(d)\varepsilon)$-close to those of the original strategy. □

We note that our constructions recover the standard Tsirelson's theorem asymptotically as $\varepsilon \to 0$, 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 $N \to \infty$. 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. 2.6 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 $O(\mathrm{poly}(d_N)/\sqrt{N})$. We argue that this error term cannot always vanish as $N \to \infty$ due to computational complexity arguments.

1. *Consider the result* $\mathrm{MIP}^* = \mathrm{RE}$ *[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 $O(\mathrm{poly}(d_N)/\sqrt{N})$ must be generally non-vanishing *in the limit $N \to \infty$. 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* $\mathrm{MIP}^{\mathrm{co}} = \mathrm{coRE}$ *[Ji+21] (more precisely, the gaped 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 $O(\mathrm{poly}(d_N)/\sqrt{N})$ 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 [Lin96; FR96] 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. 2.6 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 $O(\mathrm{poly}(d)\varepsilon)$-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 $\mathrm{MIP}^* = \mathrm{RE}$ [Ji+21], that for certain problems the approximation error $O(\mathrm{poly}(d_N)/\sqrt{N})$ 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

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). 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.

[Bel64] John S Bell. "On the Einstein Podolsky Rosen paradox". In: *Physics Physique Fizika* 1.3 (1964), p. 195.

[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.

[Bla06] Bruce Blackadar. *Operator Algebras: Theory of C*-Algebras and von Neumann Algebras*. Encyclopaedia of Mathematical Sciences. Springer Berlin Heidelberg, 2006. ISBN: 9783540285175. URL: https://books.google.ch/books?id=7-6MZLdRfdAC.

[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.

[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.

[CŚ06] 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.

[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).

[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.

[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.

[Ioa24] Adrian Ioana. “Almost commuting matrices and stability for product groups”. In: *Journal of the European Mathematical Society* (2024).

[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.

[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.

[Lin24] Huaxin Lin. “Almost commuting self-adjoint operators and measurements”. In: *arXiv preprint arXiv:2401.04018* (2024).

[Lin96] Huaxin Lin. “Almost commuting selfadjoint matrices and applications”. In: *Operator algebras and their applications* (1996), pp. 193–233.

[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.

[NC10] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

[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.

[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.

[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.

[PS79] Carl Pearcy and Allen Shields. "Almost commuting matrices". In: *Journal of Functional Analysis* 33.3 (1979), pp. 332–338.

[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.

[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.

[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.

[SW08] Volkher B Scholz and Reinhard F Werner. "Tsirelson's problem". In: *arXiv preprint arXiv:0812.4305* (2008).

[Tak79] Masamichi Takesaki. *Theory of operator algebras I*. Vol. 125. Springer, 1979.

[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.