The suboptimality ratio of projective measurements restricted to low-rank subspaces

Albert Senen-Cerda

The suboptimality ratio of projective measurements restricted to low-rank subspaces

Albert Senen–Cerda

IRIT, LAAS–CNRS, and Université Toulouse Paul Sabatier, France
albert.senen-cerda@irit.fr

Abstract

Limitations in measurement instruments can hinder the implementation of some quantum algorithms. Understanding the suboptimality of such measurements with restrictions may then lead to more efficient measurement policies. In this paper, we theoretically examine the suboptimality arising from a Procrustes problem for minimizing the average distance between two fixed quantum states when one of the states has been measured by a Projective Measurement (PM). Specifically, we compare optima when we can only use PMs that are aligned with a low-rank subspace supported by the quantum states, and the case that we can measure with the full set of PMs. For this problem, we show that the suboptimality ratio is independent of the dimension of the full space, and is at most polylogarithmic in the dimension of the low-rank subspace. In the proof of this result, we use a probabilistic approach and the main techniques include trace inequalities related to projective measurements, and operator norm bounds for equipartitions of Parseval frames, which are of independent interest.

1 Introduction

Projective Measurements (PMs), and the larger set of Positive Operator Valued Measures (POVMs) play a key role in quantum tomography [7], state discrimination [11], quantum key distribution [18, 24], and measurement-based quantum computing [9]. In practical applications, however, due to restrictions in measurement instruments and resources, not all optimal PMs or POVMs can be accurately implemented. For example, this occurs when only a limited selection of high-fidelity PMs are available to the practitioner or when no ancilla states are possible; those typically used to implement POVMs. From a theoretical perspective, these restrictions are circumvented by using a smaller set of measurements plus additional resources, e.g., POVMs can be instead simulated from PMs with additional postselection [25].

If we focus on optimization problems with measurements, it is unclear which problems have PMs as optima, and more precisely how suboptimal they become if additional restrictions exist. In this paper, we focus on this latter issue and bound the suboptimality arising from a Procrustes problem with PMs that are aligned with a low-rank structure of the measured states.

Upon measuring a quantum state with an observable on an ambient space $\mathcal{H}$, the probability that a certain outcome is observed depends on the alignment between the measured state and the state that would correspond to that outcome. A projective measurement $\mathcal{P} = \{P_1, \dots, P_n\}$ consists of n orthogonal projectors that characterize the outcomes of a measurement. After measurement of a mixed quantum state ρ by $\mathcal{P}$, the random outcome of the measurement X is P_i with probability $\mathrm{Tr}[\rho P_i] \geq 0$ for all $i \in [n]$. The Procrustes problem for PMs can be formulated as finding a PM $\mathcal{P}$ of ρ that in expectation brings the random outcome state $X = X(\mathcal{P}, \rho)$ closest to another mixed quantum state τ. In Frobenious norm, the problem is equivalent to finding

$$\min_{\mathcal{P} \in P(\mathcal{H})} \mathbb{E}\Big[\big\| \tau - X(\mathcal{P}, \rho) \big\|_{\mathrm{F}}^2 \Big], \tag{1}$$

where $P(\mathcal{H})$ is the set of PMs. Up to constants, (1) is equal to the following maximization problem over PMs

$$\max_{\mathcal{P}\in P(\mathcal{H})} \sum_{i=1}^{n} \mathrm{Tr}\big[\tau P_i \rho P_i\big]. \tag{2}$$

This optimization problem is nonconvex, and for our setting we consider additional restrictions. Firstly, we assume that τ, and ρ have a low rank structure and are only composed of states supported in a subspace $\mathcal{S} \subset \mathcal{H}$ of dimension $r \ll n$. Secondly, we model constraints of the measurement instrument by considering only PMs $\mathcal{Q}$ that are aligned with this subspace $\mathcal{S}$. Formally, this means that r out of n orthogonal projectors generate also a projective measurement on $\mathcal{S}$. The set of such PMs will generate a strict subset $P(\mathcal{S}) \subset P(\mathcal{H})$, and we can thus also define a constrained maximization problem analogous to (2). A relevant question is how suboptimal is the restriction to measurements over the low-rank structure compared to using the full set of PMs. We can define the suboptimality ratio as the smallest constant $K_{n,r} \geq 1$ such that for any quantum states ρ and τ with support in a subspace $\mathcal{S} \subset \mathcal{H}$ of dimension r,

$$\max_{\mathcal{P}\in P(\mathcal{H})} \sum_{i=1}^{n} \mathrm{Tr}\big[\tau P_i \rho P_i\big] \leq K_{n,r} \max_{\mathcal{Q}\in P(\mathcal{S})} \sum_{i=1}^{r} \mathrm{Tr}\big[\tau Q_i \rho Q_i\big]. \tag{3}$$

The constant $K_{n,r}$ encodes the suboptimality of restricting to measurements in the low-rank subspace, and a low value would imply that we can use the smaller set of measurements aligned with the subspace of quantum states to approximate (2). Constants that encode information on maximal suboptimality are common in quantum information. We can namely mention Tsirelson's bound that is perhaps the most well-known and is fundamental in bounding violations of Bell's inequalities [12]. For the maximization in (2), there are some results in the context of quantum control [26]. In particular, for the constant $K_{n,r}$, when τ and ρ are pure states ($r = 2$) then it is known that $K_{n,2} = 1$ for all $n \geq 2$ [36]. For general mixed states ($r \geq 2$), the magnitude of $K_{n,r}$ is however unknown.

Our main contribution is to show that there exists $c > 0$ such that for any $n \geq r \geq 2$,

$$K_{n,r} \leq c \log(r)^4. \tag{4}$$

The bound (4) indicates that the suboptimality does not depend on the dimension n of the ambient space $\mathcal{H}$ and is at most polylogarithmic in the dimension r of the low-rank subspace $\mathcal{S}$. As a consequence, we may approximate the solution of (2) up to a polylogarithmic factor by optimizing over the smaller space $\mathcal{S}$ where both ρ and τ are supported.

Our second contribution is technical and relates to the methods used in the proof of (4). We use a probabilistic approach to upper bound the trace of certain Kraus maps by randomized operators while preserving their PM structure. Restricted to $\mathcal{S}$, the projectors P_i induce vectors $|v_i\rangle \in \mathcal{S}$ that generate a Parseval frame and satisfy

$$\sum_{i=1}^{n} |v_i\rangle\langle v_i| = \mathbb{1}_{\mathcal{S}}. \tag{5}$$

The measurement $\mathcal{P} \in P(\mathcal{H})$ in (2) can be understood on $\mathcal{S}$ as sum of a PM $\mathcal{Q} \in P(\mathcal{S})$ that is weighted by $T = \lceil n/r \rceil$ matrices L_t obtained from some equipartition of the frame vectors

$$\sum_{t=1}^{T} \mathrm{Tr}\Big[\tau L_t \mathcal{Q}\big[L_t^\dagger \rho\ L_t\big] L_t^\dagger\Big] \quad \text{where} \quad \mathcal{Q}[a] = \sum_{i=1}^{r} Q_i a Q_i. \tag{6}$$

Similar to (5), the weight matrices satisfy

$$\sum_{t=1}^{T} L_t L_t^\dagger = \mathbb{1}_{\mathcal{S}}. \tag{7}$$

The difficulty of the analysis lies in decoupling the dimension n from (6) while preserving the PMs structure of the measurements involved. To tackle this problem we use a probabilistic approach and

find a single random measurement that in expectation behaves as the sum of the measurements in (6). Specifically, we consider Gaussian random weight matrices of the type

$$\hat{L} = \sum_{t=1}^{T} g_t L_t, \quad \text{where} \quad g_t \sim \mathrm{N}(0,1) \quad \text{are i.i.d. for all } t \in [T]. \tag{8}$$

By using this random weight in a measurement, we transfer the dependence on the dimension n from the sum of weighted PM in (6) to the variance of (8).

Additional key steps in the proof include showing the existence of a partition π of the frames that satisfies variance constraints *independent* of n. The probabilistic method from combinatorics is used in this step to bound the operator norm of a frame variance matrix. In this way, with high probability we obtain

$$\|\hat{L}\|_{\mathrm{op}} < c \log(r). \tag{9}$$

We also introduce new trace inequalities to extract the weights L_t from the traces in (6) while preserving the PMs structure of the measurements. These inequalities are based on complex interpolation results.

Remarkably, our proof method directly relates the polylogarithmic term in (4) with the additional dimensional factor that appears in well-known random matrix concentration inequalities [34] compared to their univariate counterparts. This fact suggests that our method cannot be further improved without using additional information about the optima of (2).

1.1 Related Literature

Procrustes problems have been previously studied in the literature. We can namely mention Procrustes problems for orthogonal matrices [30], weighted versions with quadratic constraints [16], conic constraints [4, 14], and problems with manifold constraints [6, 1].

The maximization problem in (2) appears in the context of quantum control with nonselective measurements in [26], and for general Kraus maps in [27]. Special cases are studied in [36], including the case $r = 2$ and $n \geq 2$. In particular $K_{n,2} = 1$ is shown using the first order stationary conditions; see (16) in Section 2.1. By relaxing in (2) the optimization to general Kraus maps, no local maxima exist for $n = 2$ [27], but this may not be the case for projective measurements.

From a theoretical perspective, the analysis of the suboptimality in optimization problems is commonly conducted to approximate solutions of computationally hard problems. Related to the quantum setting, for example, we can mention semi-definite program relaxations for generalized orthogonal Procrustes problems [5, 37] that are related to the well-known Grothendieck's inequality [8]. Grothendieck's constant [19] characterizes maximal inequality bounds and can be interpreted as a suboptimality ratio between optimization problems with different constraints on the dimensions [10], which is similar to (3). We remark, however, that these constants are universally bounded and do not grow indefinitely with the intrinsic dimension of the problem—the dimension of $\mathcal{S}$ in our setting—except if there are additional constraints, such as those induced by a graph structure [3].

We use key properties of Parseval frames [35] such as (5). In particular, we partition the frame into equal-sized sets and try to impose an orthogonal structure on each partition. Orthogonalization has been studied before in [13] for POVMs, where it is shown that if a POVM is close to being orthogonal, then there exists a PM that is actually close to the POVM. Differently in our problem, the frame in (5) may be far from orthogonal, and we have additional constraints such as the dimension of the PMs that we seek.

In the proof of (4), we use the probabilistic method [2] to prove the existence of operators with certain properties by showing that they occur with positive probability. In quantum information, for example, this method can be used to find low-rank approximations of Kraus maps [21] by using random Kraus operators. In our case, we use the random matrix (8) to randomize, and upper bound (6) while preserving the PM structure of the operators. Concentration inequalities for Gaussian random matrices from [34] are then used to estimate the operator norm provided that its variance can be controlled. Operator norm bounds for random matrices have been used in the quantum setting to understand quantum expanders [20, 17], and other random quantum channels [15, 22]. In our setting, to show that the variance of (8) is small we consider a random partition

π and bound the operator norm of a random variance matrix of $\hat{L}$ induced by this partition. We use the method of moments to estimate the operator norm [33], but differently from the typical concentration bounds that assume independence, in our case the entries of the matrix have strong dependencies induced by the frame.

Finally, we also use operator trace inequalities such as trace convexity [23], and in Section 4.5 we show new trace inequalities involving projective measurements using complex interpolation techniques. We can mention [32] where interpolation inequalities for traces of matrix products are shown using a similar approach.

2 Preliminaries

Let $\mathcal{H}$ be a complex finite-dimensional Hilbert space with inner product $\langle\cdot|\cdot\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{C}$, and such that $\dim(\mathcal{H}) = n$. Let $B(\mathcal{H})$ be the space of linear maps of $\mathcal{H}$, and let

$$B^{\mathrm{sa}}(\mathcal{H}) = \big\{A \in B(\mathcal{H}) \mid A = A^\dagger\big\},$$

be the set of self-adjoint operators of $\mathcal{H}$. We also let $B^{\mathrm{sa}}_{\geq}(\mathcal{H})$, and $B^{\mathrm{sa}}_{>}(\mathcal{H})$ be the sets of nonnegative selfadjoint, and positive selfadjoint operators, respectively. The set of density matrices of $\mathcal{H}$ is defined by

$$D(\mathcal{H}) = \big\{\rho \in B^{\mathrm{sa}}_{\geq}(\mathcal{H}), \mathrm{Tr}[\rho] = 1\big\}. \tag{10}$$

The set of (nondegenerate) PMs in $\mathcal{H}$ is composed of sets of n orthogonal rank-one projectors

$$P(\mathcal{H}) = \big\{(P_1, \ldots, P_n) \in B(\mathcal{H})^{\times n} : P_i = P_i^\dagger,\ P_iP_j = \delta_{ij}P_i,\ \mathrm{rk}(P_i) = 1 \text{ for } i \in [n]\big\}. \tag{11}$$

For any $\mathcal{P} \in P(\mathcal{H})$, the following partition of unity property for the identity operator $\mathbb{1}_{\mathcal{H}} \in B(\mathcal{H})$ holds

$$\sum_{i=1}^{n} P_i = \mathbb{1}_{\mathcal{H}}. \tag{12}$$

When a mixed quantum state encoded by a density matrix $\rho \in D(\mathcal{H})$ is measured by a PM $\mathcal{P} \in P(\mathcal{H})$, the outcome state X is $P_i \in D(\mathcal{H})$ with probability $\mathrm{Tr}[P_i\rho] \geq 0$. Let $\mu(\rho, \mathcal{P})$ be the distribution of the random density matrix X describing the outcome state after measuring ρ with $\mathcal{P} \in P(\mathcal{H})$. The expected state after measurement is then given by

$$\mathcal{P}[\rho] = \mathbb{E}_{X\sim\mu(\rho,\mathcal{P})}[X] = \sum_{i=1}^{n} \mathrm{Tr}\big[\rho P_i\big]P_i = \sum_{i=1}^{n} P_i\rho P_i. \tag{13}$$

2.1 Procrustes problem for projective measurements

Given a target quantum state $\tau \in D(\mathcal{H})$, we can ask what is the expected distance between τ, and the outcome state of the measurement of ρ by $\mathcal{P}$. The Procrustes problem for PM consists of finding for which measurement $\mathcal{P} \in P(\mathcal{H})$ the expected distance is minimized. In the Frobenious norm, $\big\|A\big\|_{\mathrm{F}}^2 = \mathrm{Tr}[AA^\dagger]$, the Procrustes problem is

$$\begin{aligned}
\min_{\mathcal{P}\in P(\mathcal{H})} \mathbb{E}_{X\sim\mu(\rho,\mathcal{P})}\Big[\big\|\tau - X\big\|_{\mathrm{F}}^2\Big] &= \min_{\mathcal{P}\in P(\mathcal{H})} \sum_{i=1}^{n} \mathrm{Tr}[P_i\rho]\big\|\tau - P_i\big\|_{\mathrm{F}}^2 \\
&= \min_{\mathcal{P}\in P(\mathcal{H})} -\sum_{i=1}^{n} 2\mathrm{Tr}\big[P_i\rho\big]\mathrm{Tr}\big[P_i\tau\big] + C(\rho,\tau) \\
&= -2\max_{\mathcal{P}\in P(\mathcal{H})} \mathrm{Tr}\big[\rho\mathcal{P}[\tau]\big] + C(\rho,\tau),
\end{aligned} \tag{14}$$

where the last equality holds since projectors are rank-one and so for all $i \in [n]$, $\mathrm{Tr}[P_i\rho]\mathrm{Tr}[P_i\tau] = \mathrm{Tr}[P_i\rho P_i\tau]$. Both τ and ρ are fixed, so both (14) and (2) are equivalent up to a constant.

The maximization problem (2) is symmetric over ρ and τ, and can be posed as an optimization over the unitary group $U(\mathcal{H}) = \{U \in B(\mathcal{H}) : UU^\dagger = U^\dagger U = \mathbb{1}_{\mathcal{H}}\}$. We let the measurement

$\tilde{\mathcal{P}}(U) = \{U\tilde{P}_1U^\dagger, \dots, U\tilde{P}_nU^\dagger\}$ be dependent on $U \in U(\mathcal{H})$ for some $\tilde{\mathcal{P}} \in P(\mathcal{H})$ fixed. Then we can consider the equivalent maximization problem

$$\max_{U\in U(\mathcal{H})} \operatorname{Tr}\big[\rho\tilde{\mathcal{P}}(U)[\tau]\big]. \tag{15}$$

In [36], the first order stationary conditions for critical points of (15) were computed: a critical point $U^\star$ with measurement $\mathcal{P}^\star = \mathcal{P}(U^\star)$ must satisfy

$$\big[\rho, \mathcal{P}^\star[\tau]\big] = \big[\mathcal{P}^\star[\rho], \tau\big], \tag{16}$$

where $[A, B] = AB - BA$ is the commutator. This problem (15) is nonconvex, and to our knowledge it is unknown how complex the optimization landscape is, e.g., how many critical points exist satisfying (16). In Section 5, we will use manifold optimization [1] to try to numerically find (15).

2.2 Restriction to aligned projective measurements

For the maximization problem in (2), we impose some restrictions in the set of measurements as well as some additional structure that we can exploit. To model the limitations of measurement instruments, we assume that we can only use projective measurements aligned with a subspace smaller than $\mathcal{H}$. Furthermore, we assume there is an additional low-rank structure in the quantum states that is also aligned with the aforementioned subspace. We formalize these ideas as follows.

Firstly, we assume that both τ and ρ belong to a low-rank subspace $\mathcal{S} \subset \mathcal{H}$ with $\dim(\mathcal{S}) = r$. Specifically, for any vector $|v\rangle \in S^\perp$—the orthogonal subspace in $\mathcal{H}$ of $\mathcal{S}$—$\rho|v\rangle = \tau|v\rangle = 0$, and all eigenvectors with positive eigenvalue of ρ and τ belong to $\mathcal{S}$. Secondly, to model the restriction of certain measurements, we will consider the maximization only over a subset of PMs that commute with the unique orthogonal projector $\Pi_{\mathcal{S}} : \mathcal{H} \to \mathcal{S}$ from $\mathcal{H}$ onto $\mathcal{S}$ and is related to the set of PMs of $\mathcal{S}$. As in (11), we define

$$P(\mathcal{S}) = \big\{(Q_1, \dots, Q_r) \in B(\mathcal{S})^{\times r} : Q_i = Q_i^\dagger,\ Q_iQ_j = \delta_{ij}Q_i,\ \operatorname{rk}(Q_i) = 1 \text{ for } i \in [r]\big\}. \tag{17}$$

There is an injective map $\iota : P(\mathcal{S}) \to P(\mathcal{H})$ given by completing the set of r orthogonal projectors of $\mathcal{Q} \in P(\mathcal{S})$ in $\mathcal{H}$ with orthogonal projectors from a fixed $\mathcal{Q}^\perp \in P(\mathcal{S}^\perp)$. Since ρ, and $\tau \in D(\mathcal{S})$ are invariant under $\Pi_{\mathcal{S}}$, we have the equality

$$\operatorname{Tr}\big[\tau\mathcal{Q}[\rho]\big] = \operatorname{Tr}\big[\tau\iota(\mathcal{Q})[\rho]\big] \text{ for any } \mathcal{Q} \in P(\mathcal{S}). \tag{18}$$

We consider then the set of restricted measurements to be $P(\mathcal{S})$, which is independent of $\mathcal{H}$, and define the restricted optimization problem as the maximization

$$\max_{\mathcal{Q}\in P(\mathcal{S})} \operatorname{Tr}\big[\tau\mathcal{Q}[\rho]\big]. \tag{19}$$

3 Suboptimality ratio of projective measurements

On the Hilbert space $\mathcal{H}$ with $\dim(\mathcal{H}) = n$, the suboptimality ratio $K_{n,r}$ is defined as the smallest constant such that for any ρ, and τ supported on the subspace $\mathcal{S} \subseteq \mathcal{H}$ of fixed dimension $r \leq n$, we have

$$\max_{\mathcal{P}\in P(\mathcal{H})} \operatorname{Tr}\big[\tau\mathcal{P}[\rho]\big] \leq K_{n,r} \max_{\mathcal{Q}\in P(\mathcal{S})} \operatorname{Tr}\big[\tau\mathcal{Q}[\rho]\big]. \tag{20}$$

The constant $K_{n,r} \geq 1$ encodes how suboptimal restricted measurements are when there is an aligned low-rank structure present in the quantum states compared to using a larger set of PMs.

Our main result in Theorem 1 bounds the suboptimality ratio $K_{n,r}$ when the quantum states τ, and ρ are supported in $\mathcal{S}$, and the measurements are also restricted to this subspace. Its proof can be found in Section 4.

Theorem 1. *Let $\mathcal{H}$ be a complex Hilbert space with $\dim(\mathcal{H}) = n$, and let $\mathcal{S} \subseteq \mathcal{H}$ be a Hilbert subspace with $\dim(\mathcal{S}) = r \leq n$. There exists $c > 0$ such that for any $n \geq r \geq 2$, and any ρ, $\tau \in D(\mathcal{S})$,*

$$\max_{\mathcal{P}\in P(\mathcal{H})} \operatorname{Tr}\big[\tau\mathcal{P}[\rho]\big] \leq c\log(r)^4 \max_{\mathcal{Q}\in P(\mathcal{S})} \operatorname{Tr}\big[\tau\mathcal{Q}[\rho]\big]. \tag{21}$$

From Theorem 1, we obtain that the bound of $K_{n,r}$ is independent of the dimension of the ambient Hilbert space $\mathcal{H}$, and at most polylogarithmic in the dimension of the subspace $\mathcal{S}$. By choosing restricted measurements in $\mathcal{S}$, we lose at most a polylogarithmic factor in (20). Remarkably, as can be seen in the proof of Theorem 1, the polylogarithmic dependence in the bound of $K_{n,r}$ arises from a connection to random matrix theory. Specifically, in random matrix concentration results there is typically an additional dimensional factor—the matrix size r—that appears compared to their counterparts in univariate concentration inequalities. In our case, this factor determines the logarithmic dependence and is a consequence of our probabilistic approach to showing the result; see Remark 3 in Section 4.4.

For the Procrustes problem in (1) we can approximate the optimum by restricting to the subspace $\mathcal{S}$. Theorem 1, and (14) immediately yield a rough approximation inequality in Corollary 1.

Corollary 1. *Suppose $\tau, \rho \in D(\mathcal{S}) \subset D(\mathcal{H})$, where $\mathcal{S} \subset \mathcal{H}$ is a r-dimensional subspace of $\mathcal{H}$. Then, there exists a universal constant $c > 0$ such that*

$$\min_{\mathcal{Q}\in P(\mathcal{S})} \mathbb{E}_{X\sim\mu(\rho,\mathcal{Q})}\Big[\big\|\tau - X\big\|_{\mathrm{F}}^2\Big] - (c\log(r)^4 - 1)\mathrm{Tr}\big[\tau\mathcal{Q}[\rho]\big] \leq \min_{\mathcal{P}\in P(\mathcal{H})} \mathbb{E}_{X\sim\mu(\rho,\mathcal{P})}\Big[\big\|\tau - X\big\|_{\mathrm{F}}^2\Big]. \quad (22)$$

We remark that Corollary 1 is not sharp since the bound of $K_{n,r}$ does not include information about the relationship between *absolute* values in (20), only their *relative* values. In particular, there may be cases when the inequality in (22) is trivial since the left-hand side may be negative. This fact raises the question if the bound of $K_{n,r}$ in Theorem 1 can be furhter improved.

In Section 5, we use numerical simulations to examine the constant $K_{n,r}$ for different values of r, and n. The simulations suggest that $K_{n,r} \simeq 1$ for small values of r, but a more complex behavior cannot be ruled out.

4 Proof of Theorem 1

We present a sketch of the proof of Theorem 1 in the following steps (i)–(vi), each one corresponding to Sections 4.1–4.6, respectively.

(i) *Reduction to weighted PMs.* In Section 4.1, we fix a PM $\mathcal{P} \in P(\mathcal{H})$ and we examine the effect that the measurement has restricted to the subspace $\mathcal{S}$ where ρ and τ are supported. We show that the restriction of $\mathcal{P}$ can be decomposed as $T = \lceil n/r \rceil$ 'weighted' measurements over $\mathcal{S}$

$$\mathrm{Tr}\big[\rho\mathcal{P}[\tau]\big] = \sum_{i=1}^{T} \mathrm{Tr}\big[\rho\mathcal{Q}_{L_i^\pi}[\tau]\big]. \quad (23)$$

The matrices $\{L_t^\pi\}_{t=1}^T$ are composed of some equipartition of vectors $|v_i\rangle_{i=1}^n$ from a Parseval frame induced by the orthogonal projectors in $\mathcal{P}$. Moreover, the equipartition of the vectors can be chosen with a permutation π of their labels.

(ii) *A good partition of Parseval frames exists.* In Section 4.2, we find a partition of the Parseval frame from the previous decomposition such that not only the matrices $\{L_t^\pi\}_{t=1}^T$ satisfy the normalization condition corresponding to frames

$$\sum_{t=1}^{T} L_t^\pi (L_t^\pi)^\dagger = \mathbb{1}_{\mathcal{S}}, \quad (24)$$

but also

$$\mathcal{L}^\pi = \sum_{t=1}^{T} (L_t^\pi)^\dagger L_t^\pi \simeq \mathbb{1}_{\mathcal{S}}, \quad (25)$$

as much as possible. We use the probabilistic method from combinatorics with $\mathcal{L}^\pi$, and the method of moments from random matrix theory to bound the operator norm $\|\mathcal{L}^\pi\|_{\mathrm{op}} = \sup_{|x|=1} |\mathcal{L}^\pi x|$. We show that there exists a permutation π of the vectors, and its corresponding equipartition such that $\|\mathcal{L}^\pi\|_{\mathrm{op}} \leq c\log(r)$.

(iii) *Randomization.* In order to avoid the fact that we have $T = \lceil n/r \rceil$ summands in (23), in Section 4.3 we use a random weight matrix

$$\hat{L} = \sum_{t=1}^{T} g_t L_t^{\pi}, \tag{26}$$

with g_t for $t \in [T]$ i.i.d. $\mathrm{N}(0,1)$ Gaussian random variables. Moreover, we show using operator trace inequalities that

$$\sum_{i=1}^{T} \mathrm{Tr}\big[\rho \mathcal{Q}_{L_i^{\pi}}[\tau]\big] \leq 3\mathbb{E}\Big[\mathrm{Tr}\big[\rho \mathcal{Q}_{\hat{L}}[\tau]\big]\Big], \tag{27}$$

where $\mathcal{Q}_{\hat{L}}$ is a random measurement on $\mathcal{S}$ weighted by $\hat{L}$.

(iv) *Concentration.* In Section 4.4 we use well-known concentration inequalities for Gaussian random matrices to show concentration of the operator norm $\|\hat{L}\|_{\mathrm{op}}$. The variance of $\hat{L}$ crucially becomes independent of n due to (24), and (25).

(v) *Interpolation.* We extract the weights from the weighted measurement $\mathcal{Q}_{\hat{L}}$ in (27). Specifically, in Section 4.5 we show by using complex interpolation inequalities that for any weighted measurement $\mathcal{Q}_L$ with $L \in B(\mathcal{H})$, there exists $\mathcal{Z} \in P(\mathcal{S})$ such that

$$\mathrm{Tr}\big[\rho \mathcal{Q}_L[\tau]\big] \leq \|L\|_{\mathrm{op}}^4 \mathrm{Tr}\big[\rho \mathcal{Z}[\tau]\big]. \tag{28}$$

(vi) *Combining all steps.* Finally, in Section 4.6 the previous interpolation inequality implies that there exists a constant $C > 0$ and $\mathcal{Y} \in P(\mathcal{S})$ such that

$$\mathbb{E}\Big[\mathrm{Tr}\big[\rho \mathcal{Q}_{\hat{L}}[\tau]\big]\Big] \leq C\mathbb{E}\big[\|\hat{L}\|_{\mathrm{op}}^4\big] \mathrm{Tr}\big[\rho \mathcal{Y}[\tau]\big], \tag{29}$$

where $\hat{L}$ is the random matrix in (26). A polylogarithmic bound in r for $\mathbb{E}\big[\|\hat{L}\|_{\mathrm{op}}^4\big]$ is obtained by using the concentration properties from step (iv). Finally, from (23), (27), and (29), the claim of Theorem 1 follows.

4.1 Reduction to weighted PMs

We first show that the projections of a fixed measurement $\mathcal{P} \in P(\mathcal{H})$ in (14) to $\mathcal{S}$ is equivalent to a sum over *weighted* PMs in $P(\mathcal{S})$.

Definition 2 (Weighted PMs)**.** *Let $\mathcal{Q} = (Q_1, \ldots, Q_r) \in P(\mathcal{S})$ and $L \in B(\mathcal{S})$. A weighted PM is a map $\mathcal{Q}_L : D(\mathcal{S}) \to B(\mathcal{S})$ defined for any $\rho \in D(\mathcal{S})$ as*

$$\mathcal{Q}_L[\rho] = \sum_{i=1}^{r} L Q_i L^{\dagger} \rho L Q_i L^{\dagger} = L\mathcal{Q}\big[L^{\dagger}\tau L\big]L^{\dagger}. \tag{30}$$

From Definition 2, given $\mathcal{Q} \in P(\mathcal{S})$ and $L \in B(\mathcal{S})$ we can write for $\tau, \rho \in D(\mathcal{H})$

$$\mathrm{Tr}\big[\rho \mathcal{Q}_L[\tau]\big] = \mathrm{Tr}\big[\rho L\mathcal{Q}[L^{\dagger}\tau L]L^{\dagger}\big]. \tag{31}$$

The following lemma poses the optimization problem in (2) as a maximum over a sum of weighted PMs in $\mathcal{S}$ that are dependent. This is a special case of the well-known Naimark dilation theorem.

Lemma 1. *Let $\mathcal{S} \subseteq \mathcal{H}$ be an inclusion of finite-dimensionial Hilbert spaces with $\dim(\mathcal{S}) = r$, and $\dim(\mathcal{H}) = n$. Let $\rho, \tau \in D(\mathcal{S}) \subseteq D(\mathcal{H})$ and $\mathcal{P} \in P(\mathcal{H})$. There exists $\mathcal{Q} \in P(\mathcal{S})$, and $L_1, \ldots, L_T \in B(\mathcal{S})$ with $T = \lceil n/r \rceil$ such that*

$$\mathrm{Tr}\big[\rho \mathcal{P}[\tau]\big] = \sum_{i=1}^{T} \mathrm{Tr}\big[\rho \mathcal{Q}_{L_i}[\tau]\big], \tag{32}$$

and

$$\sum_{t=1}^{T} L_t L_t^\dagger = \mathbb{1}_{\mathcal{S}}. \tag{33}$$

Proof. We have $\mathcal{P} = (P_1, \ldots, P_n) \in P(\mathcal{H})$ with $P_i = |u_i\rangle\langle u_i|$ for $i \in [n]$, and $\{|u_i\rangle\}_{i=1}^n$ is an Orthonormal Basis (ONB) of $\mathcal{H}$. Denote the orthogonal projection from $\mathcal{H}$ to $\mathcal{S}$ by $\Pi_S : \mathcal{H} \to \mathcal{S}$, and define $|v_i\rangle = \Pi_S|u_i\rangle \in \mathcal{S}$ for all $i \in [n]$. By using the fact that (i) $\rho = \Pi_S \rho \Pi_S^\dagger$ and similarly for τ, the following equality holds,

$$\begin{aligned}
\mathrm{Tr}\big[\rho\mathcal{P}[\tau]\big] &= \sum_{i=1}^{n} \mathrm{Tr}\big[\rho P_i \tau P_i\big] \\
&\overset{\text{(i)}}{=} \sum_{i=1}^{n} \mathrm{Tr}\big[\rho \Pi_S P_i \Pi_S^\dagger \tau \Pi_S P_i \Pi_S^\dagger\big], \\
&= \sum_{i=1}^{n} \mathrm{Tr}\big[\rho |v_i\rangle\langle v_i|\tau|v_i\rangle\langle v_i|\big].
\end{aligned} \tag{34}$$

Out of the n vectors, we consider T disjoint sets of r vectors, where we complete with the null vector if necessary. If $m = \lceil n/r \rceil r$ is the total number of vectors, we may chose the equipartition $\{|v_i\rangle\}_{i=1}^r, \{|v_i\rangle\}_{i=r+1}^{2r}, \ldots, \{|v_i\rangle\}_{i=T(r-1)+1}^{Tr}$. Let $\{|e_i\rangle\}_{i=1}^r$ be a ONB of $\mathcal{S}$ and let $\mathcal{E} = (E_1, \ldots, E_r) \in P(\mathcal{S})$ be the PM in this basis, that is, $E_i = |e_i\rangle\langle e_i|$ for $i \in [r]$. We define the operators

$$L_t = |v_{r(t-1)+1}\rangle\langle e_1| + \ldots + |v_{rt}\rangle\langle e_r| = \sum_{l=1}^{r} |v_{r(t-1)+l}\rangle\langle e_l| \in B(\mathcal{S}) \text{ for } t \in [T], \tag{35}$$

and note that $L_t E_l L_t^\dagger = |v_{r(t-1)+l}\rangle\langle v_{r(t-1)+l}|$ for all $l \in [r]$, and $t \in [T]$. In the notation of (31), we have now

$$\begin{aligned}
\sum_{i=1}^{n} \mathrm{Tr}\big[\rho P_i \tau P_i\big] &= \sum_{t=1}^{T}\sum_{i=1}^{r} \mathrm{Tr}\big[\rho L_t E_i L_t^\dagger \tau L_t E_i L_t^\dagger\big] \\
&= \sum_{t=1}^{T} \mathrm{Tr}\big[\rho \mathcal{E}_{L_t}[\tau]\big].
\end{aligned} \tag{36}$$

Moreover, from the condition that $\sum_{i=1}^n P_i = \mathbb{1}_{\mathcal{H}}$, and $\sum_{l=1}^r E_l = \mathbb{1}_{\mathcal{S}} = \Pi_{\mathcal{S}}$ we obtain that

$$\mathbb{1}_{\mathcal{S}} = \Pi_S \mathbb{1}_{\mathcal{H}} \Pi_S^\dagger = \sum_{i=1}^{n} \Pi_S P_i \Pi_S^\dagger = \sum_{i=1}^{m} |v_i\rangle\langle v_i| = \sum_{t=1}^{T}\sum_{i=1}^{r} L_t E_i L_t^\dagger = \sum_{t=1}^{T} L_t L_t^\dagger. \tag{37}$$

□

The vectors $\{|v_i\rangle\}_{i=1}^m$ in the proof of Lemma 1 constitute a Parseval frame of $\mathcal{S}$, since

$$\sum_{i=1}^{m} |v_i\rangle\langle v_i| = \mathbb{1}_{\mathcal{S}}, \tag{38}$$

which implies that $|v_i| \leq 1$, and

$$\sum_{i=1}^{m} |v_i|^2 = r. \tag{39}$$

4.2 Equipartitions of Parseval frames

In Section 4.1 we have restricted a PM $\mathcal{P} \in P(\mathcal{H})$ to $\mathcal{S}$, and we have show in Lemma 1 that it is equivalent to a sum over weighted PMs in $\mathcal{S}$. To later introduce randomization in Section 4.3

without adding a large variance, we additionally need the weight matrices to be approximately symmetric. In particular, while we have the partition of unity property in (33), its counterpart

$$\mathcal{L} = L_1^\dagger L_1 + \cdots + L_T^\dagger L_T, \qquad \text{where}$$
$$(\mathcal{L})_{ij} = \sum_{t=1}^{T} \langle v_{r(t-1)+i} \mid v_{r(t-1)+j} \rangle \tag{40}$$

can be very far away from the identity $\mathbb{1}_{\mathcal{S}}$. To overcome this issue, we will exploit two invariant properties of the decomposition that we have not used in the definition of $\{L_t\}_{t\in[T]}$ and Lemma 1. Without loss of generality, we will assume that $T = n/r \in \mathbb{N}$, since we can always complete the frame with the null vector to satisfy the constraint.

- For any permutation $\pi \in S_n$, we can define similar matrices $\{L_t^\pi\}_{t\in[T]}$ by using the permuted set of vectors $\{|v_{\pi(i)}\rangle\}_{i\in m}$ which satisfy Lemma 1, but do not leave (40) invariant.
- We can add a phase modifier $\Theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$ to the frame $\{|v_i\rangle\}_{i=1}^n$ by defining for $j \in [n]$, $|v_j\rangle(\Theta) = \exp(\mathrm{i}\theta_j)|v_j\rangle$ such that Lemma 1 also holds for this frame, but (40) is not invariant.

For a permutation $\pi \in S_n$, and a frame modifier $\Theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$, we define for $t \in [T]$

$$L_t^\pi(\Theta) = \mathrm{e}^{\mathrm{i}\theta_{\pi(r(t-1)+1)}} |v_{\pi(r(t-1)+1)}\rangle\langle e_1| + \ldots + \mathrm{e}^{\mathrm{i}\pi(\theta_{rt})} |v_{\pi(rt)}\rangle\langle e_r| \in B(\mathcal{S}). \tag{41}$$

The matrices $\{L_i^\pi(\Theta)\}_{i=1}^T$ also satisfy Lemma 1 for any $\pi \in S_n$ and $\Theta \in \mathbb{R}^n$. The generalization of (40) for $\pi \in S_n$ and $\Theta \in \mathbb{R}^n$ is then

$$\mathcal{L}^\pi(\Theta) = L_1^\pi(\Theta)^\dagger L_1^\pi(\Theta) + \ldots + L_T^\pi(\Theta)^\dagger L_T^\pi(\Theta). \tag{42}$$

Our aim is to find $\pi \in S_n$ and $\Theta \in \mathbb{R}^n$ such that $\|\mathcal{L}^\pi(\Theta)\|_{\mathrm{op}}$ is as small as possible. Using the probabilistic method from combinatorics we can show the following.

Proposition 1. *Let $\mathcal{L}^\pi(\Theta)$ be defined in (42). There exists a constant $c > 0$ such that for any $r \geq 2$ and $T \geq T(r) \geq 1$, there exists a permutation $\pi^\star \in S_n$, and a choice of frame modified by $\Theta^\star \in \mathbb{R}^n$ such that*

$$\left\|\mathcal{L}^{\pi^\star}(\Theta^\star)\right\|_{\mathrm{op}} < c\log(r). \tag{43}$$

Proposition 1 contains a dimension-dependent term $\log(r)$, which is small enough for our purposes and crucially holds for T—that is, for n—large enough. To show Proposition 1 we consider the random matrix $\mathcal{L}^\pi(\Theta)$ with a random permutation $\pi \sim \mathrm{Unif}(S_m)$, and a random phase $\exp(\mathrm{i}\Theta_i) \sim \mathrm{Unif}(S^1)$ for all $i \in [n]$. The method of moments is used to estimate the operator norm of $\mathcal{L}^\pi(\Theta)$, and is a well-known method in random matrix theory [33]. This typically involves using that for any positive-semidefinite Hermitian matrix A, $\|A\|_{\mathrm{op}}^k \leq \mathrm{Tr}[A^k]$ for any $k \geq 1$. Therefore, we can compute the expectation of the operator norm by estimating the expected trace

$$\mathbb{E}\big[\|\mathcal{L}^\pi(\Theta)\|_{\mathrm{op}}^k\big] \leq \mathbb{E}\big[\mathrm{Tr}[\mathcal{L}^\pi(\Theta)^k]\big] = F(k). \tag{44}$$

By showing that $F(k)$ is small enough, from (44) we deduce that for some permutation $\pi \in S_n$, and frame modifier Θ the operator norm is also small enough. Differently to common results in random matrix theory, the entries of $\mathcal{L}^\pi(\Theta)$ are not independent, but strongly correlated depending on the frame. Nonetheless, we are able to obtain a bound of the operator norm by leveraging the frame properties. The following bound characterizes the expectation (44).

Lemma 2. *Under the assumptions of Proposition 1, there exists $c > 0$ such that for any $r \geq 2$, $T \geq 2$ and $r \geq k \geq 1$,*

$$F(k) \leq c\left(rk^k + \frac{k!k^{2k}r^k}{T}\right). \tag{45}$$

Let us show first that Lemma 2 implies Proposition 1.

Proof (of Proposition 1). Fix $k = \lfloor \log(r) \rfloor$, and choose $T = T(r) \geq 1$ large enough such that

$$F(k) \leq 2crk^k. \tag{46}$$

Then, there exists a constants $c_1, C > 0$ such that for all $r \geq 2$,

$$\begin{aligned} \mathbb{E}\big[\|\mathcal{L}^\pi(\Theta)\|_{\mathrm{op}}\big] &\leq \mathbb{E}\big[\|\mathcal{L}^\pi(\Theta)\|_{\mathrm{op}}^k\big]^{1/k} \\ &\leq F(k)^{1/k} \\ &\leq c_1^{1/\log(r)} r^{1/\log(r)} \log(r) \\ &\leq C \log(r). \end{aligned} \tag{47}$$

From (47), we deduce that there exists a permutation $\pi^\star$ and frame modifier $\Theta^\star$ such that

$$\|\mathcal{L}^{\pi^\star}(\Theta^\star)\|_{\mathrm{op}} \leq C \log(r). \tag{48}$$

□

The proof of Lemma 2 is involved, and is deferred to Appendix A. The intuition behind the proof is that the correlations that the different vectors $|v_i\rangle$ of the frame may have—encoded in the inner products $\langle v_i|v_j\rangle$—either are small when T is large, or they are concentrated in a handful of vectors. In the former case, for a random permutation and frame modifier, the matrix $\mathcal{L}^\pi(\Theta)$ will be approximately homogeneous and thus have operator norm of order $O(1)$. In the latter case, most permutations will 'separate' the significant vectors well so that we will only need to worry about the diagonal of $\mathcal{L}^\pi(\Theta)$. We can then estimate the expectation of the operator norm when significant vectors add up in the diagonal entries. This is quantified by the leading term in (45). It also turns out that when there are only significant vectors (when $L_1 = \mathbb{1}_{\mathcal{S}}$, for example), the order of (48) is almost sharp. Specifically, the expected operator norm in this case is the expected maximum of a multinomial random variable with uniform probability $(1/r, \ldots, 1/r)$ and with r trials, which has order $\log(r)/\log(\log(r))$ [28].

In (47) the choice of $k \simeq \log(r)$ is optimal, since for larger or smaller order of k, we would obtain a worse bound on $F(k)$.

4.3 Randomization

We assume for the time being that T is large enough so that by Proposition 1 we can choose the equipartition of the frame induced by $\mathcal{P}$ in Lemma 1 that satisfies both (33) and

$$\big\|L_1^\dagger L_1 + \ldots + L_T^\dagger L_T\big\|_{\mathrm{op}} \leq c \log(r). \tag{49}$$

We namely omit in the notation the selected permutation $\pi^\star$ and phase $\Theta^\star$ that Proposition 1 guarantees, that is, we will denote $L_t = L_t^{\pi^\star}(\Theta^\star)$ for all $t \in [T]$. Later on in Section 4.6, we will show that we by enlarging the Hilbert space $\mathcal{H}$, the assumption that T is large enough will become superfluous.

Let g_t for $t \in [T]$ be i.i.d. $\mathrm{N}(0,1)$ random variables. Define the operator-valued random variable

$$\hat{L} = \sum_{t=1}^{T} g_t L_t. \tag{50}$$

By using this random matrix as a weight, we can decouple the dimension T from the sum of T weighted measurments in Lemma 1, and transfer its dependence to the variance of $\hat{L}$.

Lemma 3. *In the same setting of Lemma 1, let $\hat{L}$ be defined in* (50)*, and let $\mathcal{Q} \in P(\mathcal{S})$, then*

$$\sum_{t=1}^{T} \mathrm{Tr}\big[\rho \mathcal{Q}_{L_t}[\tau]\big] \leq 3\mathbb{E}\Big[\mathrm{Tr}\big[\rho \mathcal{Q}_{\hat{L}}[\tau]\big]\Big]. \tag{51}$$

Before showing Lemma 3 at the end of this section, we require some intermediate results first. First, we find a proxy that in expectation bounds the sum of the weighted PM on the left side of (51). Let $\hat{B}$ be an independent copy of $\hat{L}$, and $\mathcal{Q} \in P(\mathcal{S})$. We denote by $\mathcal{Q}_{\hat{L},\hat{B}}$ the random biweighted measurement that satisfies

$$\begin{aligned}\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big] &= \sum_{i=1}^{r} \mathrm{Tr}\Big[\rho\hat{L}Q_i\hat{L}^\dagger\Big]\mathrm{Tr}\Big[\tau\hat{B}Q_i\hat{B}^\dagger\Big] \\ &= \sum_{i=1}^{r} \mathrm{Tr}\Big[\rho\hat{L}Q_i\hat{L}^\dagger \otimes \tau\hat{B}Q_i\hat{B}^\dagger\Big] \\ &= \sum_{i=1}^{r} \mathrm{Tr}\Big[(\rho\otimes\tau)(\hat{L}\otimes\hat{B})(Q_i\otimes Q_i)(\hat{L}^\dagger\otimes\hat{B}^\dagger)\Big]. \end{aligned} \tag{52}$$

Note that we also have $\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big] \geq 0$.

Lemma 4. *Let $\rho,\tau \in D(\mathcal{S})$, $\mathcal{Q} \in P(\mathcal{S})$, and $\{L_t\}_{t\in[T]}$ be weight matrices. Let $\hat{L}$, $\hat{B}$ be independent copies of* (50)*. Then,*

$$\sum_{t=1}^{T} \mathrm{Tr}\big[\rho\mathcal{Q}_{L_t}[\tau]\big] \leq \mathbb{E}\Big[\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big]\Big] \tag{53}$$

Proof. From the fact that $\hat{L}$ and $\hat{B}$ are independent, we readily obtain from (52) that

$$\begin{aligned}\mathbb{E}\Big[\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big]\Big] &= \sum_{i=1}^{r} \mathbb{E}\Big[\mathrm{Tr}\Big[(\rho\otimes\tau)(\hat{L}\otimes\hat{B})(Q_i\otimes Q_i)(\hat{L}^\dagger\otimes\hat{B}^\dagger)\Big]\Big] \\ &= \sum_{i=1}^{r}\sum_{t,z=1}^{T} \mathrm{Tr}\Big[(\rho\otimes\tau)(L_t\otimes L_z)(Q_i\otimes Q_i)(L_t^\dagger\otimes L_z^\dagger)\Big] \\ &= \sum_{t,z=1}^{T}\sum_{l=1}^{r} \mathrm{Tr}\Big[(\rho\otimes\tau)(L_t\otimes L_z)(Q_i\otimes Q_i)(L_t^\dagger\otimes L_z^\dagger)\Big] \\ &\geq \sum_{t=1}^{T}\sum_{l=1}^{r} \mathrm{Tr}\Big[(\rho\otimes\tau)(L_t\otimes L_t)(Q_i\otimes Q_i)(L_t^\dagger\otimes L_t^\dagger)\Big] \qquad \text{(all terms positive)} \\ &= \sum_{t=1}^{T}\sum_{l=1}^{r} \mathrm{Tr}[\rho L_t Q_i L_t^\dagger \tau L_t Q_i L_t^\dagger] \qquad (\{Q_i\}_{i=1}^r \text{ are rank one projectors}) \\ &= \sum_{t=1}^{T} \mathrm{Tr}\big[\rho\mathcal{Q}_{L_t}[\tau]\big]. \end{aligned} \tag{54}$$

□

The biweighted measurement $\mathcal{Q}_{\hat{L},\hat{B}}$ in (52) does not have the single weighted measurement structure that we need. To overcome this problem we use the following two results based on trace convexity to upper bound $\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big]$ by several weighted measurements.

Lemma 5. (Adapted from [23, Corollary 1.1]). *Let $A,B \in B_{\geq}^{\mathrm{sa}}(\mathcal{S})$ and let $G_1,G_2 \in B(\mathcal{S})$. For any $\lambda \in [0,1]$, let*

$$G = \lambda G_1 + (1-\lambda)G_2. \tag{55}$$

The following inequality holds

$$\mathrm{Tr}\big[AGBG^\dagger\big] \leq \lambda\mathrm{Tr}\big[AG_1BG_1^\dagger\big] + (1-\lambda)\mathrm{Tr}\big[AG_2BG_2^\dagger\big]. \tag{56}$$

For the following result we require some additional notation. For any $J \in B(\mathcal{S})$, and $M \in B^{\mathrm{sa}}(\mathcal{S} \otimes \mathcal{S})$, we let

$$\mathsf{Ad}_J[M] = (J \otimes J)\ M\ (J \otimes J)^\dagger. \tag{57}$$

We define the set of separable operators of $B^{\mathrm{sa}}_{\geq}(\mathcal{S} \otimes \mathcal{S})$ as

$$\mathrm{Sep}\big(B^{\mathrm{sa}}_{\geq}(\mathcal{S} \otimes \mathcal{S})\big) = \Big\{ \sum_{i=1}^{a} \lambda_i A_1^i \otimes A_2^i \Big| a \in \mathbb{N}, A_1^i, A_2^i \in B^{\mathrm{sa}}_{\geq}(\mathcal{S}),\ \text{and } \lambda_i \geq 0 \text{ for all } i \in [a] \Big\}. \tag{58}$$

Lemma 6. *Let $C, D \in \mathrm{Sep}(B^{\mathrm{sa}}_{\geq}(\mathcal{S} \otimes \mathcal{S}))$ and $X, Y \in B(\mathcal{S})$, then in the notation of* (57)*,*

$$\begin{aligned} &\mathrm{Tr}[C(X \otimes Y)D(X \otimes Y)^\dagger] + \mathrm{Tr}[C(Y \otimes X)D(Y \otimes X)^\dagger] \leq \\ &\frac{1}{2}\Big(\mathrm{Tr}\big[C\mathsf{Ad}_{X+Y}[D]\big] + \mathrm{Tr}\big[C\mathsf{Ad}_{X-Y}[D]\big]\Big) + \mathrm{Tr}\big[C\mathsf{Ad}_X[D]\big] + \mathrm{Tr}\big[C\mathsf{Ad}_Y[D]\big]. \end{aligned} \tag{59}$$

Proof. We will first change variables. Define

$$\mathcal{M}_+ = X \otimes Y + Y \otimes X. \tag{60}$$

We can check that expanding the terms of $\mathcal{M}_+$ gives the following identity

$$\begin{aligned} \mathrm{Tr}\big[C(X \otimes Y)D(X \otimes Y)^\dagger\big] + \mathrm{Tr}\big[C(Y \otimes X)D(Y \otimes X)^\dagger\big] &= \mathrm{Tr}\big[C\mathcal{M}_+ D\mathcal{M}_+^\dagger\big] \\ &\quad - \mathrm{Tr}\big[C(X \otimes Y)D(Y \otimes X)^\dagger\big] - \mathrm{Tr}\big[C(Y \otimes X)D(X \otimes Y)^\dagger\big] \\ &\leq \mathrm{Tr}\big[C\mathcal{M}_+ D\mathcal{M}_+^\dagger\big] + \big|\mathrm{Tr}\big[C(X \otimes Y)D(Y \otimes X)^\dagger\big]\big| + \big|\mathrm{Tr}\big[C(Y \otimes X)D(X \otimes Y)^\dagger\big]\big|. \end{aligned} \tag{61}$$

We examine the term involving $\mathcal{M}_+$ in (61) closely. Note the identity

$$\begin{aligned} \mathcal{M}_+ &= \frac{1}{2}\Big((X+Y) \otimes (X+Y) - (X-Y) \otimes (X-Y)\Big) \\ &= \frac{1}{2}\Big((X+Y) \otimes (X+Y) + \mathrm{i}(X-Y) \otimes \mathrm{i}(X-Y)], \end{aligned} \tag{62}$$

We can use the trace convexity of Lemma 5 with the indentity (62) of $\mathcal{M}_+$. Trace convexity yields the following inequality

$$\mathrm{Tr}\big[C\mathcal{M}_+ D\mathcal{M}_+^\dagger\big] \leq \frac{1}{2}\Big(\mathrm{Tr}\big[C\mathsf{Ad}_{X+Y}[D]\big] + \mathrm{Tr}\big[C\mathsf{Ad}_{X-Y}[D]\big]\Big). \tag{63}$$

We bound the remaining terms of (61). We assume for the time being that C, and D are elementary tensors, that is, $C = C_1 \otimes C_2$, and $D = D_1 \otimes D_2$, where $C_1, C_2, D_1, D_2 \in B^{\mathrm{sa}}_{\geq}(\mathcal{S})$. We will use (i) Cauchy-Schwartz inequality in $\mathrm{Tr}[C_1 X D_1 Y^\dagger] = \mathrm{Tr}[(C_1^{1/2} X D_1^{1/2})(C_1^{1/2} Y D_1^{1/2})^\dagger]$, and similarly with $\mathrm{Tr}[C_2 Y D_2 X^\dagger]$. The inequality reads

$$\begin{aligned} \big|\mathrm{Tr}\big[C(X \otimes Y)D(Y \otimes X)^\dagger\big]\big| &= \Big|\mathrm{Tr}\big[C_1 X D_1 Y^\dagger\big]\mathrm{Tr}\big[C_2 Y D_2 X^\dagger\big]\Big| \\ &\overset{\text{(i)}}{\leq} \mathrm{Tr}\big[C_1 X D_1 X^\dagger\big]^{\frac{1}{2}}\mathrm{Tr}\big[C_1 Y D_1 Y^\dagger\big]^{\frac{1}{2}}\mathrm{Tr}\big[C_2 X D_2 X^\dagger\big]^{\frac{1}{2}}\mathrm{Tr}\big[C_2 Y D_2 Y^\dagger\big]^{\frac{1}{2}} \\ &= \mathrm{Tr}\big[C(X \otimes X)D(X \otimes X)^\dagger\big]^{\frac{1}{2}}\mathrm{Tr}\big[C(Y \otimes Y)D(Y \otimes Y)^\dagger\big]^{\frac{1}{2}} \\ &= \mathrm{Tr}\big[C\mathsf{Ad}_X[D]\big]^{\frac{1}{2}}\mathrm{Tr}\big[C\mathsf{Ad}_Y[D]\big]^{\frac{1}{2}}. \end{aligned} \tag{64}$$

We similarly obtain (64) for the remaining term in (61). In particular, we have

$$\begin{aligned} \big|\mathrm{Tr}\big[C(X \otimes Y)D(Y \otimes X)^\dagger\big]\big| + \big|\mathrm{Tr}\big[C(Y \otimes X)D(X \otimes Y)^\dagger\big]\big| &\leq 2\mathrm{Tr}\big[C\mathsf{Ad}_X[D]\big]^{\frac{1}{2}}\mathrm{Tr}\big[C\mathsf{Ad}_Y[D]\big]^{\frac{1}{2}} \\ &\leq \mathrm{Tr}\big[C\mathsf{Ad}_X[D]\big] + \mathrm{Tr}\big[C\mathsf{Ad}_Y[D]\big]. \end{aligned} \tag{65}$$

Finally, we can substitute the bounds of (63) and (65) into (61) to show the result in the case that C, and D are elementary tensors. To extend the result to all $C, D \in \mathrm{Sep}(B^{\mathrm{sa}}_{\geq}(\mathcal{S} \otimes \mathcal{S}))$, note that if C is separable, we can decompose $C = \sum_{i=1}^{a} \lambda_i C_1^i \otimes C_2^i$ with $C_j^i \in B^{\mathrm{sa}}_{\geq}(\mathcal{S})$ for $j \in [2]$, and $\lambda_i \geq 0$ for $i \in [a]$. We can do a similar decomposition with D. Now, from the bilinearity of both sides of the inequality (59) with respect to C, and D, we can extend (59) to all positive linear combinations of tensors in $B^{\mathrm{sa}}_{\geq}(\mathcal{S}) \otimes B^{\mathrm{sa}}_{\geq}(\mathcal{S})$. □

We are now in position to show Lemma 3.

Proof (of Lemma 3). From Lemma 4, we have the inequality

$$\sum_{t=1}^{T} \mathrm{Tr}\big[\rho \mathcal{Q}_{L_t}[\tau]\big] \leq \mathbb{E}\Big[\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big]\Big]. \tag{66}$$

We symmetrize the expectation in (66) noting that since $\hat{L}$ and $\hat{B}$ are identically and independently distributed we have

$$\begin{aligned}\mathbb{E}\Big[\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big]\Big] &= \frac{1}{2}\mathbb{E}\Big[\mathrm{Tr}\big[\mathcal{Q}_{\hat{L},\hat{B}}[\rho,\tau]\big]\Big] + \frac{1}{2}\mathbb{E}\Big[\mathrm{Tr}\big[\mathcal{Q}_{\hat{B},\hat{L}}[\rho,\tau]\big]\Big] \\ &= \frac{1}{2}\sum_{i=1}^{r}\mathbb{E}\Big[\mathrm{Tr}\Big[(\rho\otimes\tau)(\hat{L}\otimes\hat{B})(Q_i\otimes Q_i)(\hat{L}^\dagger\otimes\hat{B}^\dagger)\Big] \\ &\qquad + \mathrm{Tr}\Big[(\rho\otimes\tau)(\hat{B}\otimes\hat{L})(Q_i\otimes Q_i)(\hat{B}^\dagger\otimes\hat{L}^\dagger)\Big]\Big] \end{aligned} \tag{67}$$

We use now Lemma 6 for each pair of summands in the expectation of (67). We obtain

$$\begin{aligned}(67) &\leq \sum_{i=1}^{r}\frac{1}{4}\mathbb{E}\Big(\mathrm{Tr}\big[(\rho\otimes\tau)\mathsf{Ad}_{\hat{L}+\hat{B}}[Q_i\otimes Q_i]\big] + \mathrm{Tr}\big[(\rho\otimes\tau)\mathsf{Ad}_{\hat{L}-\hat{B}}[Q_i\otimes Q_i]\big]\Big) \\ &\qquad + \frac{1}{2}\mathbb{E}\Big(\mathrm{Tr}\big[(\rho\otimes\tau)\mathsf{Ad}_{\hat{L}}[Q_i\otimes Q_i]\big]\Big) + \frac{1}{2}\mathbb{E}\Big(\mathrm{Tr}\big[(\rho\otimes\tau)\mathsf{Ad}_{\hat{B}}[Q_i\otimes Q_i]\big]\Big) \\ &= \frac{1}{4}\mathbb{E}\Big(\mathrm{Tr}\big[\rho\mathcal{Q}_{\hat{L}+\hat{B}}[\tau]\big] + \mathrm{Tr}\big[\rho\mathcal{Q}_{\hat{L}-\hat{B}}[\tau]\big]\Big) + \mathbb{E}\Big(\mathrm{Tr}\big[\rho\mathcal{Q}_{\hat{L}}[\tau]\big]\Big). \qquad (\hat{B}\text{ distributed like }\hat{L})\end{aligned}$$

From independence and since $\hat{L}$, and $\hat{B}$ have identical symmetric Gaussian distributions, $\hat{L}+\hat{B}$ possesses the same distribution as $\hat{L}-\hat{B}$. Moreover, from its definition in (50), $\hat{L}+\hat{B}$ is a sum of i.i.d. centered Gaussian random matrices with double the variance than $\hat{L}$, that is, $\hat{L}+\hat{B}$ will have the same distribution as $\sqrt{2}\hat{L}$. Using these identities we can write

$$\begin{aligned}\frac{1}{4}\mathbb{E}\Big(\mathrm{Tr}\big[\rho\mathcal{Q}_{\hat{L}+\hat{B}}[\tau]\big] + \mathrm{Tr}\big[\rho\mathcal{Q}_{\hat{L}-\hat{B}}[\tau]\big]\Big) &= \frac{1}{2}\mathbb{E}\Big(\mathrm{Tr}\big[\rho\mathcal{Q}_{\sqrt{2}\hat{L}}[\tau]\big]\Big) \\ &= 2\mathbb{E}\Big(\mathrm{Tr}\big[\rho\mathcal{Q}_{\hat{L}}[\tau]\big]\Big)\end{aligned} \tag{68}$$

Together with the previous inequality, this last bound completes the proof. □

In (51), we have randomized the weighted measurements with $\hat{L}$, and $\hat{B}$, but we are still left with weighted measurements that cannot be directly compared with PMs. Later in Section 4.5, we show that traces of weighted measurements are upper bounded by traces of PM up to a factor depending on the weights.

4.4 Concentration

In the previous section, we have decoupled the dimension n in the number of summands in the decomposition by weighted PM in (23). However, the dependence may still be present in the form of large variance of $\hat{L}$. A key value that characterizes how large $\hat{L}$ can be is its operator norm $\|\hat{L}\|_{\mathrm{op}}$. The order of an operator norm of a random matrix has been thoroughly studied for independent series of Gaussian random matrices [34, 33], which is our case.

We will use the following result that relates the magnitude of the operator norm to that of its expected variance.

Lemma 7. (Adapted from [34, Theorem 4.1.1])
Let $B = \{B_1, \ldots, B_T\}$ be complex matrices of dimension r and define the variance statistic by

$$\nu(B) = \max\left(\Big\|\sum_{t=1}^{T} B_t B_t^\dagger\Big\|_{\mathrm{op}}, \Big\|\sum_{t=1}^{T} B_t^\dagger B_t\Big\|_{\mathrm{op}}\right). \tag{69}$$

Let $g_1, \dots, g_T$ be i.i.d. real $\mathrm{N}(0,1)$ random variables. Then for any $l > 0$, we have

$$\mathbb{P}\Big[\Big\|\sum_{t=1}^{T} g_t B_t\Big\|_{\mathrm{op}} > l\Big] \leq 2r \exp\Big(-\frac{l^2}{\nu(B)}\Big). \tag{70}$$

Recall that we assume for the time being that T is large enough so that $\mathcal{L} = L_1^\dagger L_1 + \cdots + L_T^\dagger L_T$ satisfies Proposition 1. Together with (33), the variance statistic satisfies for some $c > 0$

$$\begin{aligned}\nu(\{L_t\}_{t=1}^T) &= \max\Big(\|\mathbb{1}_{\mathcal{S}}\|_{\mathrm{op}}, \|\mathcal{L}\|_{\mathrm{op}}\Big) \\ &= \max\big(1, c\log(r)\big).\end{aligned} \tag{71}$$

Using Lemma 7, we directly obtain a bound on the operator norm of $\hat{L}$.

Lemma 8. *Suppose $r \geq 2$ and let $\hat{L}$ be defined in (50) with weight matrices $\{L_t\}_{t=1}^T$ that satisfy Proposition 1. There exists $c > 0$ such that if $l \geq 0$, with probability at least $1 - 2r\exp(-l^2/(c\log(r))$,*

$$\|\hat{L}\|_{\mathrm{op}} \leq l. \tag{72}$$

The following moment estimate will be used later on and is key in determining $K_{n,r}$.

Corollary 2. *In the same setting as Lemma 8, if the weight matrices $\{L_t\}_{t=1}^T$ satisfy Proposition 1 there exists $c > 0$ such that*

$$\mathbb{E}\big[\|\hat{L}\|_{\mathrm{op}}^4\big] \leq c\log(r)^4. \tag{73}$$

Proof. By Fubini's theorem for the expectation of a positive random variable X, we have the identity

$$\mathbb{E}[X^4] = \int_0^\infty \mathbb{P}[X > s] 4s^3 \mathrm{d}s. \tag{74}$$

From Lemma 8, for $r \geq 2$ we also have the bound

$$\mathbb{P}\big[\|\hat{L}\|_{\mathrm{op}} > s\big] \leq \min\Big(1, 2r\exp\Big(-\frac{s^2}{c\log(r)}\Big)\Big). \tag{75}$$

The value 1 is attained by the second factor of (75) at $l^\star$, with $c_1\log(r) \leq l^\star \leq c_2\log(r)$ for some $c_1, c_2 > 0$. Hence, using (74) with the bound (75) yields

$$\begin{aligned}\mathbb{E}\big[\|\hat{L}\|_{\mathrm{op}}^4\big] &\leq \int_0^{l^\star} 4s^3 \mathrm{d}s + \int_{l^\star}^\infty 8r\exp\Big(-\frac{s^2}{c\log(r)}\Big) s^3 \,\mathrm{d}s \\ &\leq (c_2)^4\log(r)^4 + 8r\int_{c_1\log(r)}^\infty \exp\Big(-\frac{s^2}{c\log(r)}\Big) s^3 \,\mathrm{d}s.\end{aligned} \tag{76}$$

We can compute the last term in (76) by using the change of variables $s^2 = c\log(r)y$,

$$\begin{aligned}\int_{c_1\log(r)}^\infty \exp\Big(-\frac{s^2}{c\log(r)}\Big) s^3 \mathrm{d}s &= c_4\log(r)^2 \int_{c_3\log(r)}^\infty \exp(-y) y \,\mathrm{d}y \\ &= c_4\log(r)^2\Big(-\exp(-y)y - \exp(-y)\Big)\Big|_{c_3\log(r)}^\infty \\ &\leq c_5\frac{\log(r)^3}{r},\end{aligned} \tag{77}$$

for some constants $c_3, c_4, c_5 > 0$ independent of $r \geq 2$. Finally, substitute (77) in (76). □

Remark 3. *We note that the bound in Proposition 1 is sufficient to obtain a polylogarithmic bound on the expected value of $\|\hat{L}\|_{\mathrm{op}}$, which will eventually determine the order of $K_{n,r}$. However, if the logarithmic bound of Proposition 1 can be improved to $\|\hat{L}\|_{\mathrm{op}} = O(1)$, we still have the intrinsic dimensional dependence on r in the the concentration inequality for random matrices; the r factor in Lemma 7. The polylogarithmic terms in the final bound of $K_{n,r}$ are thus unavoidable unless there is an additional low-rank structure of the matrices $\{L_t\}_{t\in[T]}$ that can be exploited. This fact showcases the limits of this approach, as we are not biasing the probability space towards the optimum of* (19)*, e.g., by considering only Parseval frames that have already a certain low-rank structure.*

4.5 Interpolation bound

In Lemma 3 of Section 4.4, we have upper bounded the sum of T weighted measurements by an expectation over a single random weighted measurement $\mathcal{Q}_{\hat{L}}$ that depends on the Gaussian random matrix $\hat{L}$. In this section, we will 'extract' the weight $\hat{L}$ of the weighted measurement. Specifically, we will show the following inequality for weighted measurements; recall Definition 2.

Proposition 2. *Let $\rho, \tau \in D(\mathcal{S})$ be density matrices, $L \in B(\mathcal{S})$, and $\mathcal{Q} \in P(\mathcal{S})$ be a PM. There exists $\mathcal{Z} \in P(\mathcal{S})$ such that*

$$\mathrm{Tr}\big[\rho \mathcal{Q}_L[\tau]\big] \leq \|L\|_{\mathrm{op}}^4 \mathrm{Tr}\big[\rho \mathcal{Z}[\tau]\big] \tag{78}$$

Proposition 2 shows that we can upper bound traces of weighted measurements by traces of PMs, and at most a multiplicative factor is added depending only on the weight matrices.

To show Proposition 2, we require two intermediate results. The first lemma shows an integral representation of the conjugation by an operator that has norm less than one.

Lemma 9. *Let $K \in B^{\mathrm{sa}}(\mathcal{S})$ be such that $\|K\|_{\mathrm{op}} \leq 1$ and let $A \in B(\mathcal{S})$. There exists a probability measure μ over $\mathbb{R}$ and a unitary $D \in U(\mathcal{S})$ such that if $|K| = \sqrt{KK^\dagger}$,*

$$KAK = \int_{\mathbb{R}} |K|^{-\mathrm{i}t} DAD |K|^{-\mathrm{i}t} \mu(dt). \tag{79}$$

Moreover, if $K \in B^{\mathrm{sa}}_{\geq}(\mathcal{S})$, then $D = \mathbb{1}_{\mathcal{S}}$.

Proof. We assume first that $K \in B^{\mathrm{sa}}_{>}(\mathcal{S})$ so that $K = \sum_{i=1}^r \lambda_i K_i = |K|$, where $1 \geq \lambda_i > 0$ and K_i are rank-one projectors. Ket $K^{-\mathrm{i}t} = \sum_{i=1}^r \mathrm{e}^{-\mathrm{i}x_i t} K_i$, where $x_i = \log(\lambda_i) \leq 0$. We expand the terms K and $K^{-\mathrm{i}t}$ in both sides of (79) in terms of the projectors K_i and examine at the term in the expansion corresponding to $K_i A K_j$. For (79) to hold, we must namely have the following equality for all $i, j \in [r]$

$$\lambda_i \lambda_j = \int_{\mathbb{R}} \mathrm{e}^{-\mathrm{i}(x_i + x_j)t} \mu(dt). \tag{80}$$

Observe that the right-hand side of (80) is the Fourier transform $\hat{\mu}$ of μ at $x_i + x_j$. The following equality for all $i, j \in [r]$ must then hold.

$$\lambda_i \lambda_j = \hat{\mu}(x_j + x_i). \tag{81}$$

Since $x_i \leq 0$ for all $i \in [r]$, a function $\hat{\mu}$ satisfying (81) is $\hat{\mu}(x) = \exp(-|x|)$. By applying the inverse Fourier transform to the Ansatz for $\hat{\mu}(x)$ yields

$$\mu(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \hat{\mu}(x) dx = \frac{1}{2\pi} \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \mathrm{e}^{-|x|} dx = \frac{1}{\pi(1+t^2)}. \tag{82}$$

Moreover, by integrating directly we have $\int_{\mathbb{R}} \mu(t) dt = 1$, so that μ is also a distribution over $\mathbb{R}$. This shows (79) for the case $K \in B^{\mathrm{sa}}_{>}(\mathcal{S})$.

If K is full rank but possesses negative eigenvalues, we can consider the polar decomposition $K = |K| D$, where $|K| = \sqrt{KK^\dagger} \in B^{\mathrm{sa}}_{>}(\mathcal{S})$ and $D \in U(\mathcal{S})$ is such that $|K| D = K$. The unitary D

will encode the signs of the eigenvalues of K, and satisfies $D = D^\dagger$. Apply the same arguments as in the case $K \in B^{\mathrm{sa}}_{>}(\mathcal{S})$, now using $|K|$, and DAD instead. After doing so, we also obtain

$$KAK = \int_{\mathbb{R}} |K|^{-\mathrm{i}t} DAD |K|^{-\mathrm{i}t} \mu(dt). \tag{83}$$

Finally, if K has null eigenvalues, there exists a sequence of operators $Z_1, \ldots, Z_j, \ldots \in B^{\mathrm{sa}}(\mathcal{S})$ such that $Z_j \in B^{\mathrm{sa}}(\mathcal{S})$, $\|Z_j\|_{\mathrm{op}} \leq 1$, $\ker(Z_j) = \emptyset$, and $\|Z_j - K\|_{\mathrm{op}} \to 0$ as $j \to \infty$. From the fact that $\||Z_j|^{-\mathrm{i}t} DAD |Z_j|^{-\mathrm{i}t}\|_{\mathrm{op}} \leq \|A\|_{\mathrm{op}}$ for any $j \in \mathbb{N}$, the integral in (79) is uniformly bounded. By the dominated convergence theorem, the limit as $j \to \infty$ exists and will be equal to KAK. □

The second lemma we need is a commonly used interpolation result from complex analysis:

Lemma 10. (Hadamard Three Lines Theorem, [29, p. 33]) *Let $\Psi(z)$ be a complex-valued function, bounded and continious on the closed strip $\mathcal{B} = \{z \in \mathbb{C} : 0 \leq \mathrm{Re}(z) \leq 1\}$, analytic in the interior of $\mathcal{B}$ and satisfying*

$$|\Psi(z)| \leq M_0 \quad \textit{if} \quad \mathrm{Re}(z) = 0 \tag{84}$$
$$|\Psi(z)| \leq M_1 \quad \textit{if} \quad \mathrm{Re}(z) = 1. \tag{85}$$

Then, for $z \in \mathcal{B}$, $|\Psi(z)| \leq M_0^{\mathrm{Re}(z)} M_1^{1-\mathrm{Re}(z)}$.

We are now in position to show the proposition.

Proof (of Proposition 2). We may assume that $L \in B^{\mathrm{sa}}_{\geq}(\mathcal{H})$ and L is full rank. Indeed, by the polar decomposition, if $L \in B(\mathcal{S})$, there exists $U \in U(\mathcal{S})$ such that $L = |L|U$. We can then consider the PM $\tilde{\mathcal{Q}} = \{UQ_1U^\dagger, \ldots, UQ_rU^\dagger\} \in P(\mathcal{S})$ that satisfies the equality $\mathrm{Tr}[\rho \mathcal{Q}_L(\tau)] = \mathrm{Tr}[\rho \tilde{\mathcal{Q}}_{|L|}(\tau)]$. Additionally, if L has null eigenvalues, we can use a sequence $Z_1, \ldots, Z_j, \cdots \subset B(\mathcal{S})$ with Z_j of full rank for all $j \in \mathbb{N}$ such that $Z_j \to L$ in operator norm in a similar manner as used in Lemma 9 to prove the claim. We assume in the rest of the proof that $L \in B^{\mathrm{sa}}_{>}(\mathcal{H})$.

We set $\tilde{L} = L/\|L\|_{\mathrm{op}}$, so that $\|\tilde{L}\|_{\mathrm{op}} \leq 1$. Therefore,

$$\mathrm{Tr}\big[\rho \mathcal{Q}_L[\tau]\big] = \|L\|^4_{\mathrm{op}} \mathrm{Tr}\big[\rho \mathcal{Q}_{\tilde{L}}[\tau]\big]. \tag{86}$$

We focus on the last term. Define the complex-valued function

$$\Psi(z) = \sum_{i=1}^{r} \mathrm{Tr}\big[\rho L^z Q_i \tilde{L}^{2-z} \tau \tilde{L}^z Q_i L^{2-z}\big]. \tag{87}$$

Note that since $\tilde{L}$ is self-adjoint, $\Psi(1) = \mathrm{Tr}\big[\rho \mathcal{Q}_{\tilde{L}}(\tau)\big]$. The conditions $\|\tilde{L}\|_{\mathrm{op}} \leq 1$ and $\tilde{L} \geq 0$ imply that if $0 \leq \mathrm{Re}(z) \leq 2$, then $\|\tilde{L}^z\|_{\mathrm{op}} \leq 1$ and $\|\tilde{L}^{2-z}\|_{\mathrm{op}} \leq 1$. Therefore, the function Ψ is analytic, and also bounded and continuous on the strip $\mathcal{S} = \{z \in \mathbb{C} : 0 \leq \mathrm{Re}(z) \leq 2\}$. Define

$$M_0 = \sup_{\mathrm{Re}(z)=0} |\Psi(z)| = \sup_{t \in \mathbb{R}} |\Psi(0 + \mathrm{i}t)|,$$
$$M_2 = \sup_{\mathrm{Re}(z)=2} |\Psi(z)| = \sup_{t \in \mathbb{R}} |\Psi(2 + \mathrm{i}t)|. \tag{88}$$

By Lemma 10, the upper bound $\Psi(1) \leq \max(M_2, M_0)$ holds.

Without loss of generality, we assume that the maximum is attained at M_2—the case of M_0 is analogous. For an expression of the supremum M_2 in (88), we can consider the continious map $f : \mathbb{R} \to U(\mathcal{S})$ given by $f(t) = \tilde{L}^{\mathrm{i}t}$, and notice that if $\tilde{\Psi}(t) = \Psi(2 + it)$, then

$$\tilde{\Psi}(t) = \sum_{i=1}^{r} \mathrm{Tr}\big[\rho \tilde{L}^2 f(t) Q_i f(t)^\dagger \tau \tilde{L}^2 f(t) Q_i f(t)^\dagger\big]. \tag{89}$$

By compactness of $U(\mathcal{S})$, $\sup_{t \in \mathbb{R}} |\tilde{\Psi}(t)|$ is attained at some unitary $U \in U(\mathcal{S})$ in the closure of $\mathrm{Im}(f)$, and satisfies

$$M_2 = \Big|\sum_{i=1}^{r} \mathrm{Tr}\big[\rho \tilde{L}^2 U Q_i U^\dagger \tau \tilde{L}^2 U Q_i U^\dagger\big]\Big|. \tag{90}$$

Let $\tilde{\mathcal{Q}} = (UQ_1U^\dagger, \ldots, UQ_rU^\dagger) = (\tilde{Q}_1, \ldots, \tilde{Q}_r) \in P(\mathcal{S})$. Using the inequality $\Psi(1) \leq M_2$, the following holds

$$\begin{aligned}\mathrm{Tr}\big[\rho\mathcal{Q}_{\tilde{L}}(\tau)\big] &\leq \Big|\sum_{i=1}^{r}\mathrm{Tr}\big[\rho\tilde{L}^2UQ_iU^\dagger\tau\tilde{L}^2UQ_iU^\dagger\big]\Big| \\ &= \Big|\sum_{i=1}^{r}\mathrm{Tr}\big[\rho\tilde{L}^2\tilde{Q}_i\tau\tilde{L}^2\tilde{Q}_i\big]\Big|. \end{aligned} \tag{91}$$

We are now in position to apply Lemma 9 to (91).

$$\begin{aligned}\Big|\sum_{i=1}^{r}\mathrm{Tr}\big[\rho\tilde{L}^2\tilde{Q}_i\tau\tilde{L}^2\tilde{Q}_i\big]\Big| &= \Big|\int_{\mathbb{R}}\sum_{i=1}^{r}\mathrm{Tr}\big[\rho\tilde{L}^{-2\mathrm{i}t}\tilde{Q}_i\tau\tilde{L}^{-2\mathrm{i}t}\tilde{Q}_i\big]\mu(dt)\Big| && \text{(Lemma 9)} \\ &\leq \int_{\mathbb{R}}\Big|\sum_{i=1}^{r}\mathrm{Tr}\big[\rho\tilde{L}^{-2\mathrm{i}t}\tilde{Q}_i\tau\tilde{L}^{-2\mathrm{i}t}\tilde{Q}_i\big]\Big|\mu(dt) \\ &\leq \sup_{t\in\mathbb{R}}\Big|\sum_{i=1}^{r}\mathrm{Tr}\big[\rho\tilde{L}^{-2\mathrm{i}t}\tilde{Q}_i\tau\tilde{L}^{-2\mathrm{i}t}\tilde{Q}_i\big]\Big|. \end{aligned} \tag{92}$$

By compactness of $U(\mathcal{S})$, the supremum in (92) is attained at some $V \in U(\mathcal{S})$ in the closure of $\{L^{-2\mathrm{i}t} | t \in \mathbb{R}\}$.

With V at hand, we have a sequence of inequalities that use the following facts: (i) $\tilde{Q}_i$ are rank one for $i \in [r]$, (ii) the Cauchy-Schwartz inequality $|\mathrm{Tr}[AB]| \leq \mathrm{Tr}[AA^\dagger]^{1/2}\mathrm{Tr}[BB^\dagger]^{1/2}$ with $A = \rho^{1/2}V\tilde{Q}_i^{1/2}$, $B = \tilde{Q}_i^{1/2}\rho^{1/2}$ for each $i \in [r]$, and also in the case of τ, respectively. Finally we also use (iii) Cauchy-Schwartz inequality for vectors.

$$\begin{aligned}\Big|\sum_{i=1}^{r}\mathrm{Tr}\big[\rho V\tilde{Q}_i\tau V\tilde{Q}_i\big]\Big| &\leq \sum_{i=1}^{r}\Big|\mathrm{Tr}\big[\rho V\tilde{Q}_i\tau V\tilde{Q}_i\big]\Big| \\ &\leq \sum_{i=1}^{r}\Big|\mathrm{Tr}\big[\rho V\tilde{Q}_i\big]\mathrm{Tr}[\tau V\tilde{Q}_i]\Big| && \text{(i)} \\ &\leq \sum_{i=1}^{r}\mathrm{Tr}\big[\rho V\tilde{Q}_iV^\dagger\big]^{\frac{1}{2}}\mathrm{Tr}\big[\rho\tilde{Q}_i\big]^{\frac{1}{2}}\mathrm{Tr}\big[\tau V\tilde{Q}_iV^\dagger\big]^{\frac{1}{2}}\mathrm{Tr}\big[\tau\tilde{Q}_i\big]^{\frac{1}{2}} && \text{(ii)} \\ &\overset{\text{(iii)}}{\leq} \left(\sum_{i=1}^{r}\mathrm{Tr}\big[\rho V\tilde{Q}_iV^\dagger\big]\mathrm{Tr}\big[\tau V\tilde{Q}_iV^\dagger\big]\right)^{\frac{1}{2}}\left(\sum_{i=1}^{r}\mathrm{Tr}\big[\rho\tilde{Q}_i)\big]\mathrm{Tr}\big[\tau\tilde{Q}_i\big]\right)^{\frac{1}{2}} && \text{(iii)} \\ &\leq \Big(\mathrm{Tr}\big[\rho\mathcal{Y}[\tau]\big]\Big)^{\frac{1}{2}}\Big(\mathrm{Tr}\big[\rho\tilde{\mathcal{Q}}[\tau]\big]\Big)^{\frac{1}{2}} && \text{(i)} \\ &\leq \max\Big(\mathrm{Tr}\big[\rho\mathcal{Y}(\tau)\big], \mathrm{Tr}\big[\rho\tilde{\mathcal{Q}}(\tau)\big]\Big), && \text{(93)}\end{aligned}$$

where $\mathcal{Y} = (V\tilde{Q}_1V^\dagger, \ldots, V\tilde{Q}_rV^\dagger) \in P(\mathcal{S})$. From (86), combining the inequalites in (91), (92), and (93) we conclude that there exists $\mathcal{Z} \in P(\mathcal{S})$ such that

$$\mathrm{Tr}\big[\rho\mathcal{Q}_L[\tau]\big] \leq \|L\|_{\mathrm{op}}^4\mathrm{Tr}\big[\rho\mathcal{Z}[\tau]\big]. \tag{94}$$

□

Remark 4. *We note that Proposition 2 can be generalized to measurements other than PMs. Indeed, we have not used the assumption that $\mathcal{Q} = (Q_1, \ldots, Q_r)$ is a PM, only that each Q_i for $i \in [r]$, is symmetric and rank one. This suggests that we can extend Proposition 2 to symmetric Kraus representations of rank-one POVMs.*

4.6 Combining all steps

For any $\mathcal{P} \in P(\mathcal{H})$, we use Lemma 1 together with Proposition 2 in Lemma 3, to conclude that the following inequalities hold.

$$\begin{aligned}
\operatorname{Tr}\big[\rho\mathcal{P}[\tau]\big] &= \sum_{i=1}^{T} \operatorname{Tr}\big[\rho\mathcal{Q}_{L_i}[\tau]\big] && \text{(Lemma 1)}\\
&\leq 3\mathbb{E}\Big[\operatorname{Tr}\big[\rho\mathcal{Q}_{\hat{L}}[\tau]\big]\Big] && \text{(Lemma 3)}\\
&\leq 3\mathbb{E}\Big[\|\hat{L}\|_{\mathrm{op}}^4 \operatorname{Tr}\big[\rho\mathcal{Z}[\tau]\big]\Big] && \text{(Proposition 2)}\\
&\leq 3\mathbb{E}\Big[\|\hat{L}\|_{\mathrm{op}}^4 \max_{\mathcal{Y}\in P(\mathcal{S})} \operatorname{Tr}\big[\rho\mathcal{Y}[\tau]\big]\Big]\\
&\leq 3\Big(\mathbb{E}\big[\|\hat{L}\|_{\mathrm{op}}^4\big]\Big) \max_{\mathcal{Y}\in P(\mathcal{S})} \operatorname{Tr}\big[\rho\mathcal{Y}[\tau]\big]. && (95)
\end{aligned}$$

We assume first that T—that is, n— is large enough so that Proposition 1 holds. Corollary 2 then implies that there exists $c > 0$ such that for $r \geq 2$,

$$\mathbb{E}\big[\|\hat{L}\|_{\mathrm{op}}^4\big] \leq c\log(r)^4. \tag{96}$$

In this case, we can use (96) in (95) to conclude that there exists a constant $c > 0$ such that for any $\mathcal{P} \in P(\mathcal{H})$,

$$\operatorname{Tr}\big[\rho\mathcal{P}[\tau]\big] \leq c\log(r)^4 \max_{\mathcal{Y}\in P(\mathcal{S})} \operatorname{Tr}\big[\rho\mathcal{Y}[\tau]\big]. \tag{97}$$

Taking the maximum over $\mathcal{P} \in P(\mathcal{H})$ yields Theorem 1 in the case n is large enough.

If T—that is, n—is small compared to r, and the assumptions of Proposition 1 do not hold, we can still show the result of Theorem 1. Indeed, (97) shows that there exists a constant $C > 0$, and $n(r)$ such that when $n \geq n(r)$

$$\max_{\mathcal{P}\in P(\mathcal{H})} \operatorname{Tr}\big[\rho\mathcal{P}[\tau]\big] \leq C\log(r)^4 \max_{\mathcal{Y}\in P(\mathcal{S})} \operatorname{Tr}\big[\rho\mathcal{Y}[\tau]\big]. \tag{98}$$

We need to check that (98) also holds whenever $r \leq n < n(r)$. To do so simply note that we can enlarge the space $\mathcal{H}$ into a larger Hilbert space $\tilde{\mathcal{H}}$ of dimension $n(r) > n$, and consider the same optimization problem in (20) over $P(\tilde{\mathcal{H}})$ instead. In this case, there exists an inclusion $\iota : P(\mathcal{H}) \to P(\tilde{\mathcal{H}})$ that preserves the value of $\operatorname{Tr}\big[\rho\mathcal{P}[\tau]\big]$ as shown for the case of $P(\mathcal{S})$ in (18). From this inclusion, we can conclude

$$\begin{aligned}
\max_{\mathcal{P}\in P(\mathcal{H})} \operatorname{Tr}\big[\rho\mathcal{P}[\tau]\big] &= \max_{\mathcal{P}\in P(\mathcal{H})} \operatorname{Tr}\big[\rho\iota(\mathcal{P})[\tau]\big] && \text{(inclusion } \iota : P(\mathcal{H}) \to P(\tilde{\mathcal{H}})\text{)}\\
&\leq \max_{\tilde{\mathcal{P}}\in P(\tilde{\mathcal{H}})} \operatorname{Tr}\big[\rho\tilde{\mathcal{P}}[\tau]\big]\\
&\overset{(98)}{\leq} C\log(r)^4 \max_{\mathcal{Y}\in P(\mathcal{S})} \operatorname{Tr}\big[\rho\mathcal{Y}[\tau]\big]. && (99)
\end{aligned}$$

This last step shows the claim of Theorem 1 for any $n \geq r \geq 2$.

5 Numerical experiments

We numerically examine the value of $K_{n,r} = K_r$, defined in (20), and observe its dependency on r for different values. To do so, in dimension n, we use the formulation in (15) by setting

$$S_n(U) = \operatorname{Tr}\big[\rho\mathcal{E}(U)[\tau]\big] \qquad U \in U(n), \tag{100}$$

where $\mathcal{E}$ is the measurement in the canonical basis of $\mathbb{C}^n$, and $\mathcal{E}(U) = (UE_1U^\dagger, \ldots, UE_nU^\dagger)$. For $U \in U(n)$, can then find the gradient of $S_n(U)$ restricted to the tangent space of $U(n)$, by projecting

$$P_n(U) = \operatorname{Proj}_{\mathrm{T}_U U(n)}(\nabla S_n(U)). \tag{101}$$

We refer to Appendix B for additional implementation details such as for (101). The projected gradient ascent on the manifold $U(n)$ is given by initializing $U_0 \in U(n)$, and setting

$$U_{t+1} = \exp_{U_t}\Big(\beta_t P_n(U_t)\Big) \quad \text{for} \quad t \in [t_{\max}], \tag{102}$$

where $\exp_{U_t}$ is the exponential map $\exp : \mathrm{T}_{U_t}U(n) \to U(n)$, and $\beta_t > 0$ is the stepsize. Estimating (102) is computationally expensive so we will use instead the following approximation

$$U_{t+1} = \mathrm{Proj}_{U(n)}\Big(U_t + \beta_t P_n(U_t)U_t\Big) \quad \text{for} \quad t \in [t_{\max}], \tag{103}$$

where the projection to $U(n)$ is given by the unitary obtained from the polar decomposition.

We will choose $\rho, \tau \in D(\mathcal{S})$ at random, and $U_0 \in U(n)$ such that $U_0|_{\mathcal{S}} \sim \mathrm{Unif}(U(r))$, and $U_0|_{\mathcal{S}^\perp} = \mathbb{1}_{\mathcal{S}^\perp}$. For optimizing $S_r(V)$ we will use the initialization $V_0 \sim \mathrm{Unif}(U(r))$. After $t_{\max}$ iterations of (103), we denote the empirical estimator of K_r by

$$\hat{K}_r = \frac{\max_{t\in[t_{\max}]} S_n(U_t)}{\max_{t\in[t_{\max}]} S_r(V_t)}. \tag{104}$$

The estimated values of $\hat{K}_r$ can be found in Figure 1. They seem to suggest $K_r \simeq 1$ for $r \leq 20$. Note that the estimator can satisfy $\hat{K}_r < 1$, since the optimization problem over $U(n)$ for $n > r$ commonly requires a larger horizon $t_{\max}$, and smaller stepsize to achieve the same suboptimality due to larger dimension. Similarly, the optimization of $S_r(V)$ appears to have saddle points and local maxima that allow for $\hat{K}_r > 1$ to occur when V_t converges to those. Therefore, we cannot rule out that with other initializations—also for ρ, and τ—a different behavior of $\hat{K}_r$ may be observed.

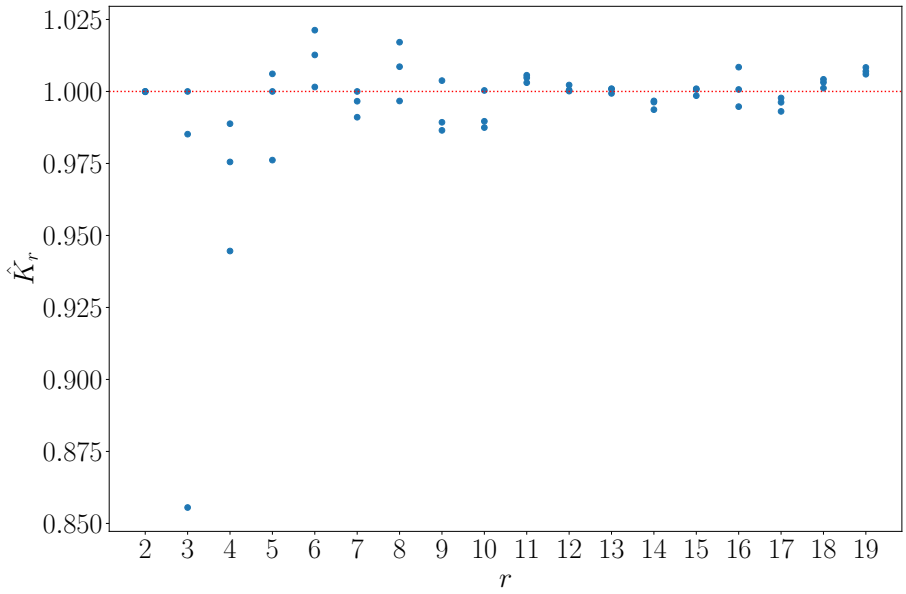

Figure 1: Evaluation of $\hat{K}_r$ for $2 \leq r \leq 20$, and $n \in \{r+1, r+2, r+3\}$ by using (104), $t_{\max} = 10^4$, and $\beta_t = 10^{-2}$.

It would be interesting to understand if there is no benefit in using PMs that are not fully aligned with the subspace $\mathcal{S}$. In this case, (2) could be solved more efficiently when τ, and ρ have low-rank. In this regard, inspired by previous remarks and numerical results we pose the following conjecture relative to (20).

Conjecture 5. *There exists $C \geq 1$ such that $K_{n,r} \leq C$ for all $n, r \geq 1$.*

If Conjecture 5 holds for $C = 1$, it would additionally distance the maximization problems in (20) from those in quantum information that bound maximal correlations of (entangled) quantum states, showcased by, e.g., Grothendieck's inequality [8].

References

[1] P-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2008.

[2] Noga Alon and Joel H. Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[3] Noga Alon, Konstantin Makarychev, Yury Makarychev, and Assaf Naor. Quadratic forms on graphs. *Inventiones mathematicae*, 163(3):499–522, 2006.

[4] Mohit Kumar Baghel, Nicolas Gillis, and Punit Sharma. On the non-symmetric semidefinite procrustes problem. *Linear Algebra and its Applications*, 648:133–159, 2022.

[5] Afonso S Bandeira, Christopher Kennedy, and Amit Singer. Approximating the little grothendieck problem over the orthogonal and unitary groups. *Mathematical programming*, 160:433–475, 2016.

[6] Rajendra Bhatia and Marco Congedo. Procrustes problems in riemannian manifolds of positive definite matrices. *Linear Algebra and its Applications*, 563:440–445, 2019.

[7] Alessandro Bisio, Giulio Chiribella, Giacomo Mauro D'Ariano, Stefano Facchini, and Paolo Perinotti. Optimal quantum tomography. *IEEE Journal of Selected Topics in Quantum Electronics*, 15(6):1646–1660, 2009.

[8] Ron Blei. *The Grothendieck inequality revisited*, volume 232. American Mathematical Society, 2014.

[9] Hans J. Briegel, David E. Browne, Wolfgang Dür, Robert Raussendorf, and Maarten Van den Nest. Measurement-based quantum computation. *Nature Physics*, 5(1):19–26, 2009.

[10] Jop Briët, Fernando Mário de Oliveira Filho, and Frank Vallentin. Grothendieck inequalities for semidefinite programs with rank constraint. *Theory OF Computing*, 10(4):77–105, 2014.

[11] Anthony Chefles. Quantum state discrimination. *Contemporary Physics*, 41(6):401–424, 2000.

[12] Boris S. Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4:93–100, 1980.

[13] Mikael de La Salle. Orthogonalization of positive operator valued measures. *arXiv preprint arXiv:2103.14126*, 2021.

[14] Terézia Fulová and Mária Trnovská. Solving constrained procrustes problems: a conic optimization approach. *arXiv preprint arXiv:2304.14961*, 2023.

[15] Carlos E. González-Guillén, Marius Junge, and Ion Nechita. On the spectral gap of random quantum channels. *arXiv preprint arXiv:1811.08847*, 2018.

[16] John C. Gower and Garmt B. Dijksterhuis. *Procrustes problems*, volume 30. OUP Oxford, 2004.

[17] Matthew B. Hastings. Random unitaries give quantum expanders. *Physical Review A—Atomic, Molecular, and Optical Physics*, 76(3):032315, 2007.

[18] Amir Kalev, Ady Mann, and Michael Revzen. Choice of measurement as the signal. *Physical Review Letters*, 110(26):260502, 2013.

[19] Jean-Louis Krivine. Constantes de grothendieck et fonctions de type positif sur les spheres. *Séminaire Maurey-Schwartz*, pages 1–17, 1978.

[20] Cécilia Lancien. Optimal quantum (tensor product) expanders from unitary designs. *arXiv preprint arXiv:2409.17971*, 2024.

[21] Cécilia Lancien and Andreas Winter. Approximating quantum channels by completely positive maps with small kraus rank. *Quantum*, 8:1320, 2024.

[22] Cécilia Lancien, Patrick Oliveira Santos, and Pierre Youssef. Limiting spectral distribution of random self-adjoint quantum channels. *Mathematical Physics, Analysis and Geometry*, 27(3): 15, 2024.

[23] Elliott H. Lieb. Convex trace functions and the wigner-yanase-dyson conjecture. *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, 19:0–35, 1973.

[24] Bin Liu, Fei Gao, Su-Juan Qin, Wei Huang, Feng Liu, and Qiao-Yan Wen. Choice of measurement as the secret. *Physical Review A*, 89(4):042318, 2014.

[25] Michał Oszmaniec, Filip B. Maciejewski, and Zbigniew Puchała. Simulating all quantum measurements using only projective measurements and postselection. *Physical Review A*, 100 (1):012351, 2019.

[26] Alexander Pechen, Nikolai Il'in, Feng Shuang, and Herschel Rabitz. Quantum control by von neumann measurements. *Physical Review A*, 74(5):052102, 2006.

[27] Alexander Pechen, Dmitrii Prokhorenko, Rebing Wu, and Herschel Rabitz. Control landscapes for two-level open quantum systems. *Journal of Physics A: Mathematical and Theoretical*, 41 (4):045205, 2008.

[28] Martin Raab and Angelika Steger. "Balls into bins"—a simple and tight analysis. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 159–170. Springer, 1998.

[29] Michael Reed and Barry Simon. *II: Fourier Analysis, Self-Adjointness*, volume 2. Elsevier, 1975.

[30] Peter H. Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966.

[31] Richard P. Stanley. Enumerative combinatorics volume 1 second edition. *Cambridge studies in advanced mathematics*, 2011.

[32] David Sutter, Mario Berta, and Marco Tomamichel. Multivariate trace inequalities. *Communications in Mathematical Physics*, 352:37–58, 2017.

[33] Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

[34] Joel A. Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[35] Shayne F.D. Waldron. *An introduction to finite tight frames*. Springer, 2018.

[36] Yaoxiong Wang, Rebing Wu, Xin Chen, Yunjian Ge, Junhui Shi, Herschel Rabitz, and Feng Shuang. Quantum state transformation by optimal projective measurements. *Journal of mathematical chemistry*, 49:507–519, 2011.

[37] Joong-Ho Won, Hua Zhou, and Kenneth L. Lange. Orthogonal trace-sum maximization: Applications, local algorithms, and global optimality. *SIAM journal on matrix analysis and applications : a publication of the Society for Industrial and Applied Mathematics*, 42 2:859–882, 2018.

A Proof of Lemma 2

The proof of Lemma 2 consists of tracking and estimating the moments of the matrix $\mathcal{L}^{\pi}(\Theta)$. In Appendix A.1, we introduce the problem and the objects that will help with counting, such as cycles, and factors. In Appendix A.2 we estimate the chance that certain factors appear in the expansion, and how they depend on T. Next in Appendix A.3 we characterize symmetries in the expansion of $F(k)$ that will simplify the counting later on. In Appendix A.4, we bound some of the products in $F(k)$, and finally in Appendix A.5 we use all previous work to estimate the asymptotic expansion as T becomes large; see Lemma 18.

A.1 Preliminaries

For simplicity we will assume that there are $m = Tr \geq n$ frame vectors exactly—we complete with the null vector, if necessary. Recall (42), and define for all $t \in [T]$, and $i, j \in [r]$,

$$\begin{aligned}\mathcal{L}_t^\pi(\Theta) &= L_t^\pi(\Theta)^\dagger L_t^\pi(\Theta) \qquad \text{and} \\ (\mathcal{L}_t^\pi(\Theta))_{ij} &= \langle v_{\pi(r(t-1)+i)}(\Theta) \mid v_{\pi(r(t-1)+j}(\Theta))\rangle \\ &= \exp\big(\mathrm{i}\theta_{\pi(r(t-1)+j)} - \mathrm{i}\theta_{\pi(r(t-1)+i)}\big)\, \langle v_{\pi(r(t-1)+i)} \mid v_{\pi(r(t-1)+j)}\rangle\,. \end{aligned} \tag{105}$$

In this notation, we have

$$\begin{aligned}\mathcal{L}^\pi(\Theta) &= \sum_{t=1}^T \mathcal{L}_t^\pi(\Theta), \qquad \text{and} \\ (\mathcal{L}^\pi(\Theta))_{ij} &= \sum_{t=1}^T \langle v_{\pi(r(t-1)+i)}(\Theta) \mid v_{\pi(r(t-1)+j}(\Theta))\rangle \\ &= \sum_{t=1}^T \exp\big(\mathrm{i}(\theta_{\pi(r(t-1)+j)} - \theta_{\pi(r(t-1)+i)}\big)\, \langle v_{\pi(r(t-1)+i)} \mid v_{\pi(r(t-1)+j)}\rangle \end{aligned} \tag{106}$$

Recall that

$$F(k) = \mathbb{E}_{\pi,\Theta}\Big[\mathrm{Tr}\big[\mathcal{L}^\pi(\Theta)^k\big]\Big]. \tag{107}$$

For a fixed $t \in [T]$, consider one of the factors appearing in the expansion of $\mathrm{Tr}[\mathcal{L}^\pi(\Theta)^k]$, say

$$\langle v_{\pi(i_1)}(\Theta) \mid v_{\pi(i_2)}(\Theta)\rangle \cdots \langle v_{\pi(i_k)}(\Theta) \mid v_{\pi(i_{k+1})}(\Theta)\rangle. \tag{108}$$

Corresponding to an index $\pi(i_1)$, if the vector $|v_{\pi(i_1)}(\Theta)\rangle$ does not appear the same number of times in the primal and dual variables of of inner product $\langle \cdot \mid \cdot \rangle$ in (108), the expectation over Θ will be zero. This is due to the fact that if $\theta \sim \mathrm{Unif}([0, 2\pi])$,

$$\mathbb{E}_\theta\big[\exp(\mathrm{i}l\theta)\big] = \begin{cases} 0 \text{ if } l \in \mathbb{Z}\backslash\{0\} \\ 1 \text{ if } l = 0. \end{cases} \tag{109}$$

Only vectors with the same number of appearances in the primal and dual arguments will thus avoid the cancellation by the expectation over the phase modifier Θ. For other matrix products appearing in the expansion $\mathrm{Tr}\big[\mathcal{L}^\pi(\Theta)^k\big]$ the same argument applies. This limits the amount of possible factors such as (108) appearing in $F(k)$. For a fixed π, we need first to characterize the type of factors that appear in the expansion

$$F(k, \pi) = \mathbb{E}_\Theta\Big[\mathrm{Tr}\big[\mathcal{L}^\pi(\Theta)^k\big]\Big]. \tag{110}$$

A permutation $\pi \in S_m$ will induce a partition $\mathcal{V}^\pi$ of $[m]$, given by the partition sets

$$\begin{aligned}\mathcal{V}_1^\pi &= \big\{\pi(1), \cdots, \pi(r)\big\} \\ &\cdots \\ \mathcal{V}_T^\pi &= \big\{\pi(r(T-1)+1), \cdots, \pi(Tr)\big\}. \end{aligned} \tag{111}$$

If $i \in \mathcal{V}_t^\pi$, the matrix $L_t^\pi(\Theta)$ will contain the vector $|v_i\rangle$. Then, only when $i, j \in \mathcal{V}_t^\pi$ for some $t \in [T]$, can the inner products $\langle v_i \mid v_j\rangle$ or $\langle v_j \mid v_i\rangle$ appear in the expansion of terms in (110) that contain the matrix $\mathcal{L}_t^\pi(\Theta)$. Similarly, if the term $\langle v_i \mid v_j\rangle \langle v_k \mid v_l\rangle$ appears adjacent in the trace of some factor in (110), but for that permutation π we have that $j \in \mathcal{V}_t^\pi$, and $k \in \mathcal{V}_s^\pi$ belong to different partitions, it implies that j, and k have same position as vectors in the matrices $\mathcal{L}_t^\pi(\Theta)$, and $\mathcal{L}_s^\pi(\Theta)$ respectively, and the matrix product $\mathcal{L}_t^\pi(\Theta)\mathcal{L}_s^\pi(\Theta)$ appears in the trace.

A useful representation for a partition (111) is encoded in its associated graph, depicted in Figure 2. If K_z is the complete graph with z vertices, the graph associated to π is isomorphic to

the Cartesian product $K_r \square K_T$. The Cartesian product satisfies that $(i,t),(j,s) \in K_r \square K_T$ are adjacent if and only if $i = j$, and t is adjacent to s or i is adjacent to j, and $t = s$. In our setting, only the edges corresponding to the component K_r will give factors $\langle v_i \mid v_j \rangle$, while each edge on the component K_T indicate that the partition has changed, that is, the trace contains a product of different matrices, e.g., $\mathcal{L}_t^\pi(\Theta)\mathcal{L}_s^\pi(\Theta)$. The type of factors that appear in $F(k,\pi)$ will depend on cycles on $K_r \square K_T$ that are compatible with the permutation. We characterize such cycles in the following definition.

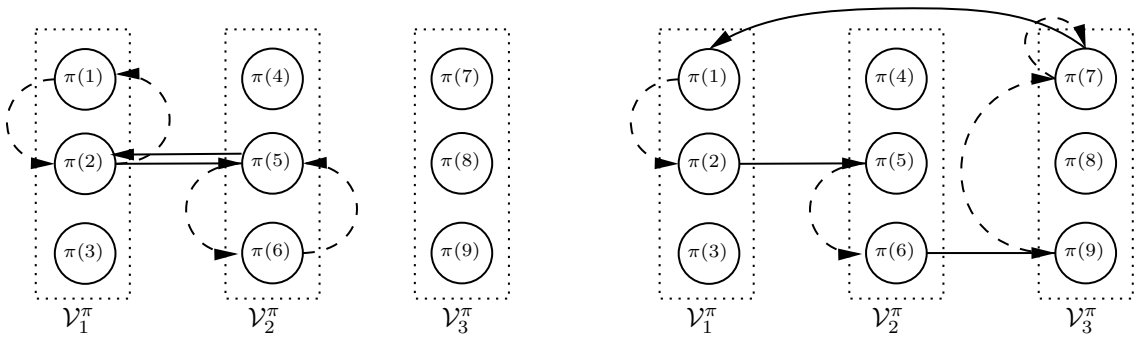

Figure 2: The graph associated to a partition $\mathcal{V}^\pi$ with examples of cycles starting at $\pi(1)$. Solid lines represent the transitions between partition sets, and can only be horizontal. Dashed lines represent transitions within partitions and add a factor $\langle v_i, v_j \rangle$ to a product such as (108). (a) An example of a representation of a nonvanishing cycle $\gamma = \pi(1,2,5,6,6,5,2,1)$. (b) Example of a cycle δ that will not appear in $F(k)$. Indeed, there are edges within partitions that are visited only once. Hence, the product in (108) associated to the cycle will have expectation zero over Θ.

Definition 6. *For a permutation π, a valid* cycle γ *(in our context) is a concatenation of $2k$ indices in $[m]$ up to permutation that generate a cycle in the graph induced by the partition $\mathcal{V}^\pi$, and such that*

(i) Restricted to each partition set $\mathcal{V}_t^\pi$ for $t \in [T]$, the cycle is union of disjoint cycles.

(ii) Transitions of the cycle γ between partition sets only occur between indices, $i, j \in [m]$ such that $\pi^{-1}(i) \equiv \pi^{-1}(j) \mod r$.

If γ is a valid cycle for π, the expectation over Θ of its associated factor

$$V_\gamma(\Theta) = \langle v_{\gamma_1}(\Theta), v_{\gamma_2}(\Theta) \rangle \cdots \langle v_{\gamma_k}(\Theta), v_{\gamma_{k+1}}(\Theta) \rangle, \tag{112}$$

does not vanish because of Definition 6 (i). That is, $V_\gamma(\Theta) = V_\gamma$ is independent of Θ. Moreover, the transitions between partitions of γ will constrain the matrix product in (110) that generated V_γ.

The order of the product elements in a multinomial product of matrices in (107) will determine the constraints of the cycles—the transition between partitions—that can appear from this product. For a fixed permutation π, depending on the matrix multinomials in the expansion of (110) and how 'commutative' they are, a cycle γ may not be in the trace of that product. See Figure 3 for an example. Alternatively, if the permutation π is random, the more a matrix product is 'mixed', the larger amount of constraints imposes to the cycles appearing in its trace, and so generally less cycles will appear. We will thus need to track the exponents of the matrices in the expansion of $F(k)$ to determine how often a cycle appears for a (random) permutation.

A.2 Tracking factors and their chance of occurring

We examine the noncommutative expansion of $F(k)$, and we track the exponents depending on the matrix indices. We define the set $\mathfrak{L}(k)$ of possible indices for the k different matrices appearing on

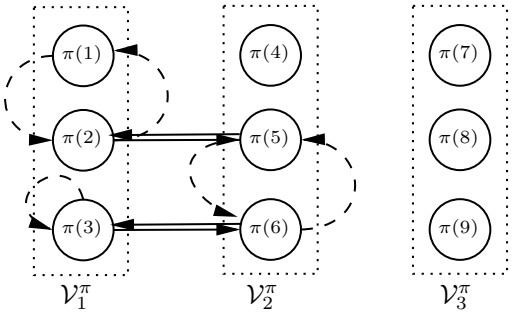

Figure 3: For the permutation π, the depicted cycle appears in the expansion of the trace of (omitting π and Θ) $\mathcal{L}_1^2\mathcal{L}_2\mathcal{L}_1\mathcal{L}_2$, but not in the trace of $\mathcal{L}_1^3\mathcal{L}_2^2$. Indeed, there must be 4 transitions between partition sets.

each term in the expansion of (110). Concretely, the set of exp-indices is defined as

$$\mathfrak{L}(k) = [T]^k. \tag{113}$$

and each $\ell = (\ell_1, \dots, \ell_k) \in \mathfrak{L}(k)$ will track the matrix indices in the expansion of

$$\mathbb{E}_\Theta\Big[\mathrm{Tr}\Big[\prod_{j=1}^k \mathcal{L}_{\ell_j}^\pi(\Theta)\Big]\Big]. \tag{114}$$

For $\ell \in \mathfrak{L}(k)$, we will denote the set of *different* indices appearing in ℓ by $N(\ell) = \cup_{i=1}^k \{\ell_i\}$, and $n(\ell) = |N(\ell)|$ its amount. Recall (112). Given $\ell \in \mathfrak{L}(k)$, for a fixed permutation $\pi \in S_m$, we will denote the set of factors in (137) as

$$B(\ell,\pi) = \Big\{ V_\gamma \;\Big|\; \text{for some valid } \gamma, V_\gamma \text{ appears in } \mathbb{E}_\Theta\big[\mathrm{Tr}[\prod_{j=1}^k \mathcal{L}_{\ell_j}^\pi(\Theta)]\big] \Big\}, \quad \text{and define}$$

$$B(\ell) = \cup_{\pi\in S_m} B(\ell,\pi), \tag{115}$$

as well as the set

$$\mathcal{G}(k) = \bigsqcup_{\ell\in\mathfrak{L}(k)} \{\ell, B(\ell)\}. \tag{116}$$

Definition 7. *We will say that a factor V* appears *in $\Gamma = \{\ell, B(\ell)\} \in \mathcal{G}(k)$ for a permutation $\pi \in S_m$, and denote so by $V \in B(\Gamma,\pi)$, if $V \in B(\ell,\pi)$. We will also simply say that a factor V appears in Γ if $V \in B(\Gamma)$.*

We denote the set of different indices in $\Gamma = \{\ell, B(\ell)\} \in \mathcal{G}(k)$ and their number by $N(\Gamma) = N(\ell)$, and $n(\Gamma) = n(\ell)$, respectively. Similarly, we will denote the vectors appearing in a factor $V \in B(\Gamma)$ by $N(V)$, and their number by $n(V)$. If a factor V appears in Γ, the probability that V appears for a random permutation $\pi \in S_m$ is positive, and will depend on the constraints imposed by Γ, and its valid cycles.

With the notation of Definition 7, if $V \in B(\Gamma,\pi)$ for some permutation π, each index in $N(V) \subset [m]$ must belong to only one of the partition sets from (111) with label in $N(\Gamma) \subset [T]$. Moreover, in this case a cycle γ such that $V_\gamma = V$ will satisfy the constraints that come from the order in the multiplication of the matrices $\{L_t^\pi\}_{t\in[T]}$ associated to Γ, encoded by the exp-index of Γ. With this information, in the following Lemma 11 we obtain necessary conditions for γ to satisfy $V_\gamma \in B(\Gamma,\pi)$, which we will use to bound the chance that this happens when π is random later on.

Lemma 11. *For a permutation $\pi \in S_m$ and $\Gamma \in \mathcal{G}(k)$, if a cycle γ satisfies $V_\gamma \in B(\Gamma,\pi)$, then the following conditions hold.*

(A) *The cycle γ restricted to each partition set of $\mathcal{V}^\pi$ in (111), will be a disjoint union of cycles (or the emptyset). Only partitions with index in $N(\Gamma)$ may have nonempty intersection with γ, and any index in $N(V_\gamma) \cap \mathcal{V}_t^\pi$ for $t \in N(\Gamma)$ will appear as many times in the primal as in the dual arguments of the inner products of V_γ.*

(B) *The cycle γ changes partitions in the graph associated to the permutation π only between indices $i, j \in N(V_\gamma)$ that satisfy $\pi^{-1}(i) \equiv \pi^{-1}(j) \mod r$, that is, if $i \in \mathcal{V}_t^\pi$, $j \in \mathcal{V}_{t'}^\pi$, both indices appear in the same relative position within their respective partitions t, and t'.*

Proof. For a fixed π and $\Gamma \in \mathcal{G}(k)$, condition (A) is implied by the cycle being valid, and not vanishing in $F(k, \pi)$. Thus, in each partition set $\mathcal{V}_t^\pi$ the condition must also hold.

For condition (B), note that transitions between partitions correspond to changes of matrix index $l \in N(\Gamma)$ as they appear in the noncommutative product corresponding to $\prod_{j=1}^k \mathcal{L}_{\ell_j}^\pi(\Theta)$ for $\Gamma = \{\ell, N(\ell)\}$. Indices of γ will change partitions in the graph associated to π if their corresponding vectors belong to different matrices that are consecutive in the matrix product. For two indices $i, j \in N(V_\gamma)$ that change between partitions in the graph, from the multiplication rule of matrices, the locations of these indices in the matrices share row and column number. This implies that the vectors corresponding to these indices are in the same position within their respective partitions. If $i = \pi(d_i)$, and $j = \pi(d_j)$, for some $d_i, d_j \in [m]$, then d_i must be the same as d_j modulo r. □

The rules in Lemma 11 yield necessary conditions for γ to satisfy $V_\gamma \in B(\Gamma, \pi)$. For example, if a product $\langle v_{i_1}, v_{i_2}\rangle\langle v_{i_3}, v_{i_4}\rangle$ appears in V and i_2 and i_3 belong to different partition sets in the graph of γ, then the factor $\langle v_{i_4}, v_{i_1}\rangle$ cannot appear in V_γ by condition (A).

In the following lemmas, we use the previous necessary conditions to bound the probability that a factor V appears for a random permutation π and $\Gamma \in \mathcal{G}(k)$. We take especial interest in the contributions by the diagonal terms of the matrices $\mathcal{L}_t^\pi(\Theta)$—factors V that contain only norms—in the expansion of $F(k)$; they will become leading terms later on.

Lemma 12. *Suppose that V appears in Γ and is such that $n(V) = n(\Gamma)$. Then V is composed only of diagonal elements of $\{\mathcal{L}_t^\pi(\Theta)\}_{i\in[T]}$ and thus $V = \prod_{i\in N(V)} |v_i|^{2k_i}$ where $\mathbf{k} = (k_1, \dots, k_{n(\Gamma)})$ are the multiplicities of the vectors with indices in $N(V)$.*

Proof. If $n(V) = n(\Gamma)$ for $\Gamma = \{\ell, N(\ell)\}$, the product $\prod_{i=1}^k \mathcal{L}_{\ell_i}^\pi(\Theta)$ that has $n(\Gamma)$ distinct matrices—not counting multiplicies—can only have a cycle γ with $n(V)$ different indices if and only if for each index in $N(\Gamma)$, there is exactly one index in $N(V_\gamma)$. This happens only when diagonal elements of $\mathcal{L}_t^\pi(\Theta)$ appear in V_γ. Therefore, the identity $V = V_\gamma = \prod_{i\in N(V)} |v_i|^{2k_i}$ must hold for some γ, where $\mathbf{k} = (k_1, \dots, k_{n(\Gamma)})$ are the multiplicities of the indices $t \in N(\Gamma)$ in ℓ, as well as of $N(V)$. □

Lemma 13. (Adapted from [31, Sec. 1.9]) *Let $d \geq n$. The number of ways to partition d labeled items into n labeled classes such that no class is empty is given by*

$$n!\left\{{d \atop n}\right\} = \sum_{j=0}^{n-1} (-1)^j (n-j)^d \binom{d}{j}, \tag{117}$$

where $\left\{{d \atop n}\right\}$ is the Stirling number of the second kind that satisfies $\left\{{d \atop d}\right\} = \left\{{d \atop 1}\right\} = 1$, and $\left\{{d \atop n}\right\} \leq \frac{n^d}{n!}$.

A cycle γ that satisfies condition (A), on each partition it is decomposed in smaller disjoint cycles $\delta_1, \dots, \delta_s$ for some $s \geq 1$. We can choose the decomposition such that each cycle cannot be further decomposed, and the indices of each subcycle belong to a single partition. We denote the maximal number of disjoint cycles that decompose γ by $d(\gamma)$. This maximal decomposition is independent of partition set $\mathcal{V}^\pi$, and satisfies

$$V_\gamma = V_{\delta_1} \cdots V_{d(\gamma)}. \tag{118}$$

Note that the subcycles $\delta_1, \dots, \delta_{d(\gamma)}$ do not keep track of the transitions between the partitions of γ; this is instead in Γ. Crucially, for any $\Gamma \in \mathcal{G}(k)$ such that condition (A) holds for some γ, we must then have $n(\Gamma) \leq d(\gamma) \leq n(V_\gamma)$.

The number of maximal subcycles allows us to bound in the next lemma how many permutations π allow the expansion term in (110) corresponding to Γ to have a factor V appearing. For the following Lemma, if $\mathbf{k}(\Gamma) = (k_1, \dots, k_{n(\Gamma)})$ are the multiplicities of the exp-index ℓ of Γ, we let z_1 be the different number of times that k_1 appears in $\mathbf{k}(\Gamma)$, z_2 the number of times a different

multiplicity than k_1 appears in $\mathbf{k}(\Gamma)$, and so on. Denote the number different multiplicities of $\mathbf{k}(\Gamma)$ in this way by $Z(\Gamma)$. The vector

$$\mathbf{z}(\Gamma) = (z_1, \ldots, z_{Z(\Gamma)}) \tag{119}$$

will contain the multiplicities of $\mathbf{k}(\Gamma)$, and

$$\sum_{i=1}^{Z(\Gamma)} z_i = n(\Gamma). \tag{120}$$

With slight abuse of notation, the vectors $\mathbf{k}(\Gamma)$, or $\mathbf{z}(\Gamma)$ will also be considered as lists whenever used in a multinomial number, e.g, $\binom{k}{\mathbf{k}(\Gamma)}$.

Lemma 14. *Suppose that $\Gamma \in \mathcal{G}(k)$, and let $V \in B(\Gamma)$. If $\pi \sim \mathrm{Unif}(S_m)$, then*
(i) *If $n(V) > n(\Gamma)$,*

$$C_{V,\Gamma} = \mathbb{P}\big[V \in B(\Gamma, \pi)\big] \leq \frac{n(\Gamma)! \left\{ {d(V) \atop n(\Gamma)} \right\} r^{n(V)-n(\Gamma)+1}}{\binom{Tr}{n(V)} n(V)!}. \tag{121}$$

(ii) *If $n(V) = n(\Gamma)$, let $\mathbf{z}(\Gamma)$ be the multiplicities of $\mathbf{k}(\Gamma)$ defined in* (119)*, then*

$$C_{V,\Gamma} = \mathbb{P}\big[V \in B(\Gamma, \pi)\big] = \frac{r}{\binom{Tr}{n(V)} \binom{n(\Gamma)}{\mathbf{z}(\Gamma)}}. \tag{122}$$

Proof. We count first the number of ways that permutations of $[m]$ can rearrange the indices of V. Out of $m = Tr$ possible total index locations, $n(V)$ must be chosen. By symmetry, there are

$$\frac{(Tr)!}{(Tr - n(V))!} = \binom{Tr}{n(V)} n(V)! \tag{123}$$

possible ordered choices of locations for the $n(V)$ indices that can be attained by a permutation $\pi \in S_m$. We bound how many of these permutations allow the existance of a cycle δ compatible with Γ and the permutation π, such that $V_\delta = V$. We show first part (ii).

Case (ii) If $n(V) = n(\Gamma)$, by Lemma 12, we only need to look at the diagonal elements of the matrices $\mathcal{L}_t^\pi$ for $t \in N(\Gamma)$ and see where the indices of V are sent by π. From the assumption $V \in B(\Gamma)$, by Lemma 12 we must have $V = \prod_{i \in N(V)} |v_i|^{2k_i}$, where k_i coincide with the multiplicities of the ℓ index of Γ. For any permutation π such that $V \in B(\Gamma, \pi)$, each $|v_i\rangle$ with $i \in N(V)$ must belong to a unique partition $\mathcal{V}_t^\pi$ with $t \in N(\Gamma)$, and also satisfy condition (B) from Lemma 11. If there exists a cycle δ such that $V_\delta = V$, then all indices of V must satisfy condition (B) pairwise since they are all adjacent to each other. There are r different values modulo r such that condition (B) can be satisfied simultaneously for all $i \in N(V)$. Furthermore, for each of the choices, we can try to assign differently the indices of $N(\gamma)$ within the partitions available in $N(\Gamma)$. However, the multiplicities of Γ have to match the multiplicity of the vectors in V for the factor to appear for that index assignment. Only if multiplicities coincide will the factor V appear for the permutation. For $i, j \in N(\gamma)$, we can only exchange i with the position of j, if $k_i = k_j$. Let $\mathbf{z}(\Gamma)$ be the vector of multiplicities of $\{k_i\}_{i \in N(\Gamma)}$. From the previous discussion, for a fixed value modulo r, there are exactly $\prod_{i=1}^{z(\Gamma)} z_i!$ ways to permute indices in $N(V)$ within the different partitions corresponding to $N(\Gamma)$ such that the product V appears for that permutation. There are at most $r \prod_{i=1}^{z(\Gamma)} z_i!$ permutations that satisfy $V \in B(\Gamma, \pi)$. Conversely, any permutation satisfying these constraints will also imply that $V \in B(\Gamma, \pi)$. We conclude

$$\mathbb{P}\big[V \in B(\Gamma, \pi)\big] = \frac{r \prod_{i=1}^{z(\Gamma)} z_i!}{\binom{Tr}{n(V)} n(V)!} = \frac{r}{\binom{Tr}{n(V)} \binom{n(\Gamma)}{\mathbf{z}(\Gamma)}}. \tag{124}$$

Case (i) We assume that $n(V) > n(\Gamma)$. To upper bound the amount of permutations that allow $V \in B(\Gamma, \pi)$, we consider cycles that satisfy the necessary conditions (A) and (B) from Lemma 11.

Let $\Gamma = \{\ell, B(\ell)\} \in \mathcal{G}(k)$, and γ a cycle satisfying $V = V_\gamma$. The chance that $V \in B(\Gamma, \pi)$ is at most the chance that there exists γ such that $V = V_\gamma$ and γ satisfies the necessary conditions (A) and (B) in Lemma 11 for π and Γ,

$$\mathbb{P}[V \in B(\Gamma, \pi)] \leq \mathbb{P}[\ \exists \gamma \text{ s.t. } V_\gamma = V, \gamma \text{ satisfies (A) and (B) for } \pi \text{ and } \Gamma] \tag{125}$$

We count how many cycles γ may satisfy all conditions simultaneously. For a permutation π, if a cycle γ exists that satisfies $V = V_\gamma$, for each index $t \in N(\Gamma)$ corresponding to a partition set $\mathcal{V}_t^\pi$, there exist at least one constraint for the position of the indices of $N(V) \cap \mathcal{V}_t^\pi$ when the cycle γ either arrives from, or leaves to another partition. This constraint is imposed by the fact that the positions must be up to π equal modulo r—see Lemma 11.

Denote by $\gamma|_{\mathcal{V}_t^\pi}$ for each $t \in N(\Gamma)$ the maximal set of subcycles of γ that are contained in the partition set $\mathcal{V}_t(\pi)$. Since there are at least two transitions—arrival and departure—of γ for each partition set, we show that there are at least $n(\Gamma) - 1$ indices of V whose positions in the associated graph are fixed by the positions of other indices in $N(V)$ in the graph. Indeed, for each $t \in N(\Gamma)$ at least one index $u_t \in N(V) \cap \mathcal{V}_t^\pi$ must be adjacent in the graph of γ to at least another index $u_{t'}$ belonging to another partition $t' \in N(\Gamma)$. If this was not the case, then for at least one index $t \in N(\Gamma)$, γ would arrive to the partition corresponding to t, but not leave, which would contradict the fact that γ is a cycle. A connected graph with the $n(\Gamma)$ partitions is generated by γ tracking arrivals and departures of partition sets. It is well-known that the the connected graph with minimal amount of edges that we can construct with $n(\Gamma)$ vertices has $n(\Gamma) - 1$ edges. Hence, there will be at least $n(\Gamma) - 1$ indices in $N(V)$ whose positions are determined by other indices of $N(V)$ whenever there is a cycle γ such that $V = V_\gamma$.

We now upper bound how many permutations π allow for a cycle γ to exist with $V = V_\gamma$ satisfying conditions (A), and (B). Conditional on $V = V_\gamma$, consider the maximal decomposition of V by disjoint subcycles described in (118). Let $d(V)$ be the number of disjoint subcycles for this decomposition and denote the set of subcycles by $D(V) = \{\delta_1, \ldots, \delta_{d(V)}\}$. Condition (A) requires that the indices of any subcycle δ_i, $i \in [d(V)]$ belong to the same partition. Any γ satisfying $V = V_\gamma$, and (A) will have these $d(V)$ labeled maximal subcycles distributed among $n(\Gamma)$ labeled partition sets, with at least one subcycle per partition set. By Lemma 13, the different number of ways to distribute the subcycles is

$$n(\Gamma)! \left\{ {d(V) \atop n(\Gamma)} \right\} = \sum_{i=0}^{n(\Gamma)-1} (-1)^j (n(\Gamma) - j)^{d(V)} \binom{d(V)}{j}, \tag{126}$$

where $\left\{ {d(V) \atop n(\Gamma)} \right\}$ is the Stirling number of second kind.

For each fixed distribution of the subcycles, each index $i \in N(V)$ within a partition $t \in N(\Gamma)$ can be assigned to at most r positions in its corresponding partition. However, from the previous discussion, in order to generate a connected cycle γ such that $V_\gamma = V$, there are at least $n(\Gamma) - 1$ indices of V whose locations are determined by other indices in $N(V)$. Therefore, there are at most $r^{n(V)-n(\Gamma)+1}$ possible ways to distribute the indices for each given subcycle assignment, and satisfy (B). Combining the previous bounds, the number of different ways to allow a cycle γ to exists with $V = V_\gamma$, while also satisfying (B) and (A) for Γ is at most

$$n(\Gamma)! \left\{ {d(V) \atop n(\Gamma)} \right\} r^{n(V)-n(\Gamma)+1}. \tag{127}$$

Combining (127) with (123), we obtain that

$$\mathbb{P}[V \in B(\Gamma, \pi)] \leq \frac{n(\Gamma)! \left\{ {d(V) \atop n(\Gamma)} \right\} r^{n(V)-n(\Gamma)+1}}{\binom{Tr}{n(V)} n(V)!}. \tag{128}$$

□

Lemma 14 bounds the probability $C_{V,\Gamma}$ of a factor V appearing for $\Gamma \in \mathcal{G}(k)$ and a random permutation. In the next section, we examine the symmetries of $\mathcal{G}(k)$ that will simplify the computations later, and are key in obtaining a leading order of $F(k)$ independent of T.

A.3 Characterizing symmetries in the expansion

The set $\mathcal{G}(k)$ possesses several additional symmetries that we can exploit by using groups. There are namely two group actions of S_k and S_T on $\mathcal{G}(k)$. Both actions will allow us to decouple the combinatorial problem of counting exp-indices to counting only factors.

First action by S_k. We define the action $\mathfrak{h} : S_k \to \mathrm{Aut}(\mathcal{G}(k))$ that permutes the location of the indices of $\Gamma = \{\ell, B(\ell)\} \in \mathcal{G}(k)$, that is, for $g \in S_k$,

$$\begin{aligned} g(\ell_1, \dots, \ell_k) &= (\ell_{g(1)}, \dots, \ell_{g(k)}) \\ \mathfrak{h}[g]\big(\{\ell, B(\ell)\}\big) &= \{g(\ell), B(g(\ell))\}. \end{aligned} \tag{129}$$

This group action tracks how many $\Gamma' \in \mathcal{G}(k)$ in the expansion of (107) have similar indices to those of Γ, up to reordering. For example, they track if

$$\mathbb{E}_{\Theta,\pi}\Big[\mathrm{Tr}\Big[\big(\mathcal{L}_1^\pi(\Theta)\big)^2\big(\mathcal{L}_2^\pi(\Theta)\big)^2\Big]\Big] \text{ and } \mathbb{E}_{\Theta,\pi}\Big[\mathrm{Tr}\Big[\mathcal{L}_1^\pi(\Theta)\mathcal{L}_2^\pi(\Theta)\mathcal{L}_1^\pi(\Theta)\mathcal{L}_2^\pi(\Theta)\Big]\Big], \tag{130}$$

may share some of the factors. We define the set of orbits of this action on $\mathcal{G}(k)$ as

$$\bar{\mathcal{G}}(k) = \frac{\mathcal{G}(k)}{S_k}. \tag{131}$$

We bound the number of elements of an orbit $\bar{\Gamma} \in \bar{\mathcal{G}}(k)$. Given $\Gamma = \{\ell, B(\ell)\} \in \mathcal{G}(k)$ we denote the multiplicities of ℓ by $\mathbf{k}(\Gamma) = (k_1, \dots, k_{n(\Gamma)})$. A group element $g \in S_k$ will permute the entries of the vector ℓ but not their multiplicities. The orbit of $\Gamma \in \mathcal{G}(k)$ thus contains at most all possible different ways of choosing out of k locations in the exp-indices ℓ, $n(\Gamma)$ labels with their respective multiplicities in $\mathbf{k}(\Gamma)$. If $\bar{\Gamma}$ denotes the orbit of Γ under this action, the number of elements is given by the multinomial number

$$|\bar{\Gamma}| = \binom{k}{k_1, \dots, k_{n(\Gamma)}} = \binom{k}{\mathbf{k}(\Gamma)}. \tag{132}$$

The number, and index labels of $\Gamma \in \mathcal{G}(k)$ are invariant under $\mathfrak{h}$. We can define $n(\bar{\Gamma}) = n(\Gamma)$, and $N(\bar{\Gamma})) = N(\Gamma)$ for any representative $\Gamma \in \bar{\Gamma}$.

Note that the action of S_k is in general not well-defined for factors of Γ since $B(\ell) \neq B(g(\ell))$, as the following example shows.

Example 8. *For $k = 4$, let $\ell_1 = (1, 1, 2, 2)$, and $\ell_2 = (1, 2, 1, 2)$—in (130)—be exp-indices in the same equivalence class under action $\mathfrak{h}$, and consider the following factor corresponding to some cycle γ*

$$V_\gamma = \langle v_1, v_2\rangle\langle v_2, v_1\rangle\langle v_3, v_4\rangle\langle v_4, v_3\rangle \in B(\ell_1), \tag{133}$$

where $\{1, 2\} \subset \mathcal{V}_1$, $\{3, 4\} \subset \mathcal{V}_2$ belong to different partition sets. In order for V_γ to also appear in $B(\ell_2)$, only the following product or its conjugate could appear in the expansion corresponding to $B(\ell_2)$ according to Definition 6,

$$\langle v_1, v_2\rangle\langle v_3, v_4\rangle\langle v_2, v_1\rangle\langle v_4, v_3\rangle. \tag{134}$$

Here, the indices 2 and 3 satisfy constrain (B) of Lemma 11, and appear in the same position for both partitions. However, the same should occur with indices 2 and 4. The constraints (B) cannot be satisfied and so $V_\gamma \notin B(\ell_2)$.

Nonetheless, the action of $\mathfrak{h}$ will be well-defined for some factors; see Remark 9 below. We will use this symmetry to estimate their contribution to $F(k)$, which will become the leading term.

Remark 9. *For a factor $V \in B(\ell)$ composed only of diagonal elements, that is, V is a product of norms of vectors $|v_i\rangle$, the action $\mathfrak{h}$ is well-defined. Indeed, the diagonal parts of the matrices $\{\mathcal{L}_t^\pi(\Theta)\}_{t\in[T]}$ commute for any exp-index, that is, $V \in B(g(\ell))$ for any $g \in S_k$.*

Second action by S_T. We can also define an action $\mathfrak{g} : S_T \to \mathrm{Aut}(\bar{\mathcal{G}}(k))$. A permutation $h \in S_T$ acts on $\ell \in [T]^k$ by permuting its indices

$$h(\bar{\ell}_1, \ldots, \bar{\ell}_k) = (h(\ell_1), \ldots, h(\ell_k)). \tag{135}$$

The action on a representative $[\{\bar{\ell}, B(\bar{\ell})\}] \in \bar{\Gamma}$ is

$$\mathfrak{g}[h]\big([\{\bar{\ell}, B(\bar{\ell})\}]\big) = [\{h(\bar{\ell}), B(h(\bar{\ell}))\}]. \tag{136}$$

Differently to $\mathfrak{h}$, the action $\mathfrak{g}$ is well defined for factors due to symmetry, e.g., we expect the same factors V occurring in

$$\mathbb{E}_{\Theta,\pi}\Big[\mathrm{Tr}\Big[\big(\mathcal{L}_1^\pi(\Theta)\big)^2\big(\mathcal{L}_2^\pi(\Theta)\big)^2\Big]\Big] \text{ and } \mathbb{E}_{\Theta,\pi}\Big[\mathrm{Tr}\Big[\big(\mathcal{L}_3^\pi(\Theta)\big)^2\big(\mathcal{L}_4^\pi(\Theta)\big)^2\Big]\Big]. \tag{137}$$

Indeed, if a factor $V \in B(\Gamma, \pi)$, the effect of permuting the indices of Γ with h for a fixed π is equivalent to permuting the partition sets with label $t \in N(\Gamma)$, while leaving V and π fixed. We can then compensate this change by using another permutation $\pi' \in S_m$ such that $V \in B(h(\Gamma), \pi')$ instead. From the definition of $B(\ell)$ in (115) we obtain that the factors appearing for a representative of $\bar{\Gamma}$ are invariant under the action $\mathfrak{g}$. We denote the set of orbits under this action by

$$G(k) = \frac{\bar{\mathcal{G}}(k)}{S_T} \tag{138}$$

Let $\Delta \in G(k)$ be the orbit of $\bar{\Gamma}$ under $\mathfrak{g}$. The number of different exp-indices of any of its representatives in $\bar{\Gamma}$ is $n(\Delta) = n(\bar{\Gamma})$, and their multiplicities are $\mathbf{k}(\Delta) = \bar{\mathbf{\Gamma}}$. We will also use the notation $\mathbf{z}(\Delta)$ to denote the multiplicities of $\mathbf{k}(\Delta)$ similarly to (119).

We compute the size of Δ. Recall that $\bar{\Gamma}$ contains all permutations of the exp-index ℓ of a representative $\Gamma \in \bar{\Gamma}$. We find a representative Γ such that its exp-indices are equal adjacent to each other

$$(\overbrace{\ell_1, \cdots, \ell_1}^{k_1 \text{times}}, \cdots, \overbrace{\ell_{n(\Gamma)}, \ldots, \ell_{n(\Gamma)}}^{k_{n(\Gamma)} \text{times}}) \tag{139}$$

In this manner, we can identify $\bar{\Gamma}$ with a commutative multinomial of order k in $[T]$ variables

$$\prod_{i=1}^{n(\Gamma)} (x_{\ell_i})^{k_i}. \tag{140}$$

A permutation $h \in S_T$ sends the labels $i \in N(\Gamma)$ to any other different labels in $[T]$ while leaving the multiplicities invariant. If all multiplicities are different, then the amount of different multinomials under this action will be equivalent to the number of injective maps from the set $N(\Gamma)$ to $[T]$, which is given by the multinomial coefficient

$$\frac{T!}{(T - n(\Gamma))!} = \binom{T}{n(\Gamma)} n(\Gamma)! = \binom{T}{n(\Delta)} n(\Delta)!. \tag{141}$$

However, if some of the multiplicities in $\mathbf{k}(\Gamma)$ are equal, some of the previous maps for the labels in $N(\Gamma)$ will be sent to the same multinomial. Let $\mathbf{z}(\Gamma)$ be the vector of size $n(\Gamma)$ with the multiplicities of $\mathbf{k}(\Gamma)$. This vector satisfies

$$n(\Gamma) = \sum_{i=1}^{Z(\Gamma)} z_i. \tag{142}$$

For a fixed multinomial of the type (140) with multiplicities $\mathbf{k}(\Gamma)$ there are $\prod_{i=1}^{Z(\Gamma)} z_i!$ different relabeling of its indices that leave the multinomial invariant. Therefore, the number of commutative monomials of order k in $[T]$ variables with multiplicities $\mathbf{k}(\Gamma)$ is given by

$$|\Delta| = \binom{T}{n(\Gamma)} \binom{n(\Gamma)}{\mathbf{z}(\Gamma)} = \binom{T}{n(\Delta)} \binom{n(\Delta)}{\mathbf{z}(\Delta)}. \tag{143}$$

A.4 Combinatorial inequalities

We have examined exp-indices and their symmetries. We now define sets that will help us with the counting of factors.

$$\mathcal{M}(k,l) = \Big\{(k_1,\dots,k_m) \in \big([k]\cup\{0\}\big)^{\times m} \Big| \sum_{i=1}^{m} k_i = k, k_i > 0 \text{ for } \textit{exactly } l \text{ indices } i \in [m]\Big\}. \quad (144)$$

The set $\mathcal{M}(k,l)$ can be identified with the set of different multinomials of order k with l variables out of m. The set of different multinomials of order k with m variables $\mathcal{M}(k)$ is then one-to-one with

$$\mathcal{M}(k) = \bigsqcup_{1\le l\le k} \mathcal{M}(k,l). \quad (145)$$

For a vector $\mathbf{k} \in \mathcal{M}(k)$ (of dimension m), we denote its number of nonzero entries by $n(\mathbf{k})$.

Definition 10. *Let a vector $\mathbf{k} \in \mathcal{M}(k)$, we denote its frequency by $f(\mathbf{k}) = (k_{[1]},\dots,k_{[k]})$, where $k_{[i]}$ is the ith largest entry of $\mathbf{k}$, and we complete with zeros if necessary. Similarly, for any $\Gamma \in \bar{\Gamma} \in \Delta \in G(k)$, we define $f(\Delta) = f(\bar{\Gamma}) = f(\Gamma) = (\mathbf{k}(\Gamma), 0_{k-n(\Gamma)})$, which is the vector of multiplicities of Γ in order with added zeros if necessary. We denote the set of frequencies with at most k different items by $\mathfrak{F}(k) = f(\mathcal{M}(k))$.*

Note that any $f \in \mathfrak{F}(k)$ can be represented by a decreasing vector $(f_1,\dots,f_k)$, where we complete with zeros if necessary. If $\Delta \in \mathcal{G}(k)$ and $n(\Delta) = l$, we can define the subset of (144) that have the same frequencies as Δ.

$$\mathcal{M}(k,l)[\Delta] = \Big\{\mathbf{k} \in \mathcal{M}(k,l) \;\Big|\; f(\mathbf{k}) = f(\Delta)\Big\}. \quad (146)$$

The set $\mathcal{M}(k,l)[\Delta]$ is one-to-one with the set of multinomials of order k with l variables out of m, such that their vectors of multiplicities is $\mathbf{k}(\Delta)$ up to reordering. Crucially, the frequencies of different elements in $G(k)$ are also different as shown in the following lemma.

Lemma 15. *Let $\Delta_1, \Delta_2 \in G(k)$. Then $f(\Delta_1) = f(\Delta_2)$, or equivalently $\mathbf{k}(\Delta_1)$ is $\mathbf{k}(\Delta_2)$ up to reordering, if and only if $\Delta_1 = \Delta_2$.*

Proof. If $f(\Delta_1) = f(\Delta_2)$, then Δ_1 has a representative $\Gamma_1 = \{\ell_1, B(\ell_1)\}$ that satisfies $i_1 \in \ell_1$ k_1 times, $i_2 \in \ell_1$ k_2 times, …, and $i_{n(\Gamma_1)} \in \ell_1$ $k_{n(\Gamma_1)}$ times for some $i_1,\dots,i_{n(\Gamma_1)} \in N(\Gamma_1)$. Similarly for Δ_2 with a representative $\Gamma_2 = \{\ell_2, B(\ell_2)\}$ that satisfies $j_1 \in \ell_2$ k_1 times, $j_2 \in \ell_2$ k_2 times, …, and $j_{n(\Gamma_1)} \in \ell_2$ $k_{n(\Gamma_1)}$ times for some $j_1,\dots,j_{n(\Gamma_2)} \in N(\Gamma_2)$. Since $n(\Delta_1) = n(\Gamma_1) = n(\Gamma_2) = n(\Delta_2)$, there exists a permutation $h \in S_T$ that takes $i_1,\dots,i_{n(\Gamma_1)}$ to $j_1,\dots,j_{n(\Gamma_2)}$. In particular, Δ_1, and Δ_2 have the same representatives, and so they must be equal. □

If we consider the disjoint union of (146) over all $\Delta \in G(k)$, we recover (144)

$$\mathcal{M}(k,l) = \bigsqcup_{\substack{\Delta\in G(k)\\ n(\Delta)=l}} \mathcal{M}(k,l)[\Delta]. \quad (147)$$

We introduce a partial order over frequencies in $\mathfrak{F}(k)$.

Definition 11. *A frequency $f = (f_1,\dots,f_k)$ is dominated by $g = (g_1,\dots,g_k)$ in $\mathfrak{F}(k)$, denoted by $g \prec f$, if for a partition $\{\mathcal{A}_i\}_{i=1}^{l}$ of $[k]$ with some $l \ge 1$, we have for all $j \in [k]$ such that $f_j \neq 0$,*

$$f_j = \sum_{i\in\mathcal{A}_j} g_i. \quad (148)$$

For $\Delta \in G(k)$ such that $d > n(\Delta)$, we define

$$\mathcal{M}(k,d)[\Delta] = \Big\{\mathbf{k} \in \mathcal{M}(k,d) \;\Big|\; f(\mathbf{k}) \prec f(\Delta)\Big\}, \quad (149)$$

We use the previous definitions to simplify the computations with the following results.

Lemma 16. *In the notation of this section, let $\Gamma \in \bar{\Gamma} \in \Delta \in G(k)$ be such that $n(\Gamma) = l$.*

$$\sum_{\substack{V \in B(\Gamma) \\ n(V)=l}} C_{V,\Gamma} V = \frac{r}{\binom{Tr}{n(\Delta)}\binom{n(\Delta)}{\mathbf{z}(\Delta)}} \sum_{\mathbf{k} \in \mathcal{M}(k,l)[\Delta]} \prod_{i=1}^{m} |v_i|^{2k_i}, \tag{150}$$

and

$$\sum_{\substack{V \in B(\Gamma) \\ n(V)>l}} C_{V,\Gamma} V \leq \sum_{d=l+1}^{k} \frac{n(\Gamma)! \left\{ {d \atop n(\Gamma)} \right\} r^{d-n(\Gamma)+1}}{\binom{Tr}{d} d!} \sum_{\mathbf{g} \in \mathcal{M}(k,d)[\Delta]} \binom{k}{\mathbf{g}} \prod_{i=1}^{m} |v_i|^{2g_i}. \tag{151}$$

Proof. We show (150) first. Since we consider the case $n(V) = n(\Gamma) = l$, by Lemma 12, V is product of norms of the vectors $\{|v_i\rangle\}_{i \in [m]}$, and there will be exactly $n(\Gamma)$ different norms in the product, up to multiplicities. The indices $t \in N(\Gamma)$ are one to one with indices in $N(V)$, and the multiplicites will be $\mathbf{k}(\Gamma) = (k_1, \ldots, k_l)$. Thus, we can find as many factors $V \in B(\Gamma)$ with exactly $n(V) = n(\Gamma)$ different norms as there are multinomials of the type

$$x_{i_1}^{k_1} \cdots x_{i_l}^{k_l}, \tag{152}$$

with $i_1, \ldots, i_l \in [m]$ different, that is, vectors $\mathbf{k} \in \mathcal{M}(k)$, with the frequencies equal to $f(\Gamma) = f(\Delta)$. This is exactly the definition of $\mathcal{M}(k,l)[\Delta]$ in (146). If we denote the factor with only norms with indices and multiplicities corresponding to $\mathbf{k}$ by $V(\mathbf{k})$, then

$$\begin{aligned} \sum_{\substack{V \in B(\Gamma) \\ n(V)=l}} C_{V,\Gamma} V &= \sum_{\mathbf{k} \in \mathcal{M}(k,l)[\Delta]} C_{V(\mathbf{k}),\Gamma} \prod_{i=1}^{m} |v_i|^{2k_i} && (f(\Gamma) = f(\mathbf{k})) \\ &= \frac{r}{\binom{Tr}{n(\Delta)}\binom{n(\Delta)}{\mathbf{z}(\Delta)}} \sum_{\mathbf{k} \in \mathcal{M}(k,l)[\Delta]} \prod_{i=1}^{m} |v_i|^{2k_i} && (\text{Lemma 14, and } n(\Gamma) = n(\Delta) = l) \end{aligned}$$

We show now (151). Suppose Γ has multiplicities $\mathbf{k}(\Gamma) = (k_1, \ldots, k_l)$, and V is such that $N(V) = \{1, \ldots, n(V)\} \subset [m]$, where $n(V) > n(\Gamma) = l$. Let $\mathbf{g} = (g_1, \ldots, g_{n(V)})$ be the multiplicities of the indices in V, which satisfy

$$k = \sum_{i=1}^{n(V)} g_i. \tag{153}$$

Using Cauchy-Schwartz in all inner products of V we obtain the bound

$$|V| \leq \prod_{i=1}^{n(V)} |v_i|^{2g_i}. \tag{154}$$

Differently to the previous case, elements $t \in N(\Gamma)$ are not one to one with indices in $N(V)$. However, if $V \in B(\Gamma, \pi)$ for some permutation π there is an assignment that maps each $t \in N(\Gamma)$ to a set $\mathcal{V}_t^\pi \cap N(V)$. In this case, the multiplicity k_t of each $t \in N(\Gamma)$ must be equal to the sum of the multiplicities of the indices of V in $\mathcal{V}_t^\pi \cap N(V)$. That is, there exists a partition $\{\mathcal{A}_i\}_{i=1}^{n(\Gamma)}$ of $N(V)$ such that

$$k_t = \sum_{i \in \mathcal{A}_t} g_i \quad \text{for each} \quad t \in N(\Gamma). \tag{155}$$

This is the domination condition from Definition 11. We can thus upper bound the sum in (151) by a sum over multinomials such as (154) whose multiplicities are dominated by $f(\Gamma) = f(\Delta)$, which is $\mathcal{M}(k,d)[\Delta]$ by definition in (149). We are left to count how many $V \in B(\Gamma)$ have the same multinomial in (154) that upper bounds the factor V. Let $\mathcal{N}$ be this number. Any other factor V' with (154) as upper bound will have the same indices $i \in N(V') = N(V)$ appearing in V' with

some other order. We can upper bound the number of ways to order the indices by the number of ways to partition k items in $n(V)$ labeled classes with multiplicities $\mathbf{g} = (g_1, \ldots, g_{n(V)})$, that is

$$\mathcal{N} \leq \mathcal{N}(\mathbf{g}) = \binom{k}{\mathbf{g}}. \tag{156}$$

The previous argument implies that if $\Gamma \in \bar{\Gamma} \in \Delta$,

$$\begin{aligned}
\sum_{\substack{V \in B(\Gamma) \\ n(V) > l}} C_{V,\Gamma} V &\leq \sum_{d=l+1}^{k} \sum_{\substack{V \in B(\Gamma) \\ n(V)=d}} C_{V,\Gamma} \prod_{i \in N(V)} |v_i|^{2g_i} && \text{(From (154))} \\
&\leq \sum_{d=l+1}^{k} \sum_{\substack{V \in B(\Gamma) \\ n(V)=d}} \frac{n(\Gamma)! \left\{ {n(V) \atop n(\Gamma)} \right\} r^{n(V)-n(\Gamma)+1}}{\binom{Tr}{n(V)} n(V)!} \prod_{i \in N(V)} |v_i|^{2g_i} && \text{(Lemma 14)} \\
&\leq \sum_{d=l+1}^{k} \frac{n(\Gamma)! \left\{ {d \atop n(\Gamma)} \right\} r^{d-n(\Gamma)+1}}{\binom{Tr}{d} d!} \sum_{\mathbf{g} \in \mathcal{M}(k,d)[\Delta]} \mathcal{N}(\mathbf{g}) \prod_{i=1}^{m} |v_i|^{2g_i} \\
&\leq \sum_{d=l+1}^{k} \frac{n(\Delta)! \left\{ {d \atop n(\Delta)} \right\} r^{d-n(\Delta)+1}}{\binom{Tr}{d} d!} \sum_{\mathbf{g} \in \mathcal{M}(k,d)[\Delta]} \binom{k}{\mathbf{g}} \prod_{i=1}^{m} |v_i|^{2g_i} && \text{(From (156))}
\end{aligned}$$

□

Let

$$W_\Delta(T) = \frac{r}{\binom{Tr}{n(\Delta)} \binom{n(\Delta)}{\mathbf{z}(\Delta)}} \tag{157}$$

$$W_{d,\Delta}(T) = \frac{n(\Delta)! \left\{ {d \atop n(\Delta)} \right\} r^{d-n(\Delta)+1}}{\binom{Tr}{d} d!} \quad \text{if} \quad d > n(\Delta). \tag{158}$$

the following lemma characterizes the fact that the leading term of $F(k)$ is independent of T. The following asymptotic in T will be key in the computation.

Lemma 17. *Let $\Delta \in G(k)$, and recall that $|\Delta| = |\Delta|(T)$ is (143) and depends on T. As $T \to \infty$:*

(a) $|\Delta|(T) W_\Delta(T) = \dfrac{1 + O(T^{-1})}{r^{n(\Delta)-1}}$

(b) $|\Delta|(T) W_{d,\Delta}(T) = O\Big(\dfrac{k^k}{r^{n(\Delta)-1} T^{d-n(\Delta)}}\Big)$.

Proof. We show (a) first. From Lemma 14, and the bound in (143), we immediately have

$$\begin{aligned}
|\Delta|(T) W_\Delta(T) &= \binom{T}{n(\Delta)} \frac{r}{\binom{Tr}{n(\Delta)}} = \frac{rT(T-1)\cdots(T-n(\Delta)+1)}{rT(rT-1)\cdots(rT-n(\Delta)+1)} \\
&= \frac{1 + O(T^{-1})}{r^{n(\Delta)-1}}.
\end{aligned} \tag{159}$$

For (b), we again use the bound of Lemma 14, and (143) to obtain

$$\begin{aligned}
|\Delta|(T) W_{d,\Delta}(T) &= \binom{T}{n(\Delta)} \binom{n(\Delta)}{\mathbf{z}(\Delta)} \frac{n(\Delta)! \left\{ {d \atop n(\Delta)} \right\} r^{d-n(\Delta)+1}}{\binom{Tr}{d} d!} \\
&\leq \binom{T}{n(\Delta)} \frac{n(\Delta)! \left\{ {d \atop n(\Delta)} \right\} r^{d-n(\Delta)+1}}{\binom{Tr}{d}} && \text{(Since } \tbinom{n(\Delta)}{\mathbf{z}(\Delta)} \leq d!\text{, and } d > n(\Delta)\text{)} \\
&\leq \frac{T(T-1)\cdots(T-n(\Delta)+1)}{rT(rT-1)\cdots(rT-d+1)} d! \frac{n(\Delta)^d}{n(\Delta)!} r^{d-n(\Delta)+1} && \text{(Lemma 13, } \left\{ {d \atop n(\Delta)} \right\} \leq \tfrac{n(\Delta)^d}{n(\Delta)!}\text{)} \\
&= O\Big(\frac{k^k}{r^{n(\Delta)-1} T^{d-n(\Delta)}}\Big). && \text{(From } \tfrac{n(\Delta)^d}{n(\Delta)!} \leq \tfrac{d^d}{d!}\text{)}
\end{aligned}$$

□

A.5 Estimating the k-th moment

We can now estimate the moments. Expanding $F(k)$, we obtain from the definition in (138) that

$$F(k) = \sum_{\Gamma=\{\ell,B(\ell)\}\in\mathcal{G}(k)} \mathbb{E}\Big[\mathrm{Tr}\Big[\prod_{i=1}^{k} \mathcal{L}_{\ell_i}^{\pi}(\theta)\Big]\Big]. \tag{160}$$

We present the asymptotic expansion of $F(k)$ as $T \to \infty$ in the following lemma.

Lemma 18. *Let the expectation of $F(k)$ in (107) be with respect to a random permutation $\pi \sim \mathrm{Unif}(S_m)$ and phases $\exp(\mathrm{i}\Theta_i) \sim \mathrm{Unif}(S^1)$ for all $i \in [m]$. Then as $T \to \infty$ we have*

$$F(k) = \sum_{l=1}^{k} \frac{1}{r^{l-1}} \sum_{\mathbf{k}\in\mathcal{M}(k,l)} \binom{k}{\mathbf{k}} \prod_{i=1}^{m} |v_i|^{2k_i} + O\Big(\frac{k!k^{2k}r^k}{T}\Big). \tag{161}$$

Proof. We expand

$$\begin{aligned} F(k) &= \sum_{\Gamma\in\mathcal{G}(k)} \sum_{V\in B(\Gamma)} C_{V,\Gamma} V \\ &= \sum_{l=1}^{k} \sum_{\substack{\Gamma\in\mathcal{G}(k)\\ n(\Gamma)=l}} \left(\sum_{\substack{V\in B(\Gamma)\\ n(V)=l}} C_{V,\Gamma} V + \sum_{\substack{\gamma\in B(\Gamma)\\ n(V)>l}} C_{V,\Gamma} V \right) \\ &= I_1 + I_2. \end{aligned} \tag{162}$$

We are left to bound both terms I_1 and I_2 from (162).
Bound of I_1. The following inequalities hold:

$$\begin{aligned} I_1 &= \sum_{l=1}^{k} \sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}} \sum_{\bar{\Gamma}\in\Delta} \sum_{\Gamma\in\bar{\Gamma}} \sum_{\substack{V\in B(\Gamma)\\ n(V)=l}} C_{V,\Gamma} V && \text{(From (131) and (138))} \\ &= \sum_{l=1}^{k} \sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}} \sum_{\bar{\Gamma}\in\Delta} \sum_{\Gamma\in\bar{\Gamma}} \sum_{\mathbf{k}\in\mathcal{M}(k,l)[\Delta]} W_\Delta \prod_{i=1}^{m} |v_i|^{2k_i} && \text{(Lemma 16, and (157))} \\ &= \sum_{l=1}^{k} \sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}} |\Delta||\bar{\Gamma}| \sum_{\mathbf{k}\in\mathcal{M}(k,l)[\Delta]} W_\Delta \prod_{i=1}^{m} |v_i|^{2k_i} \\ &= \sum_{l=1}^{k} \sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}} |\Delta| W_\Delta \binom{k}{\mathbf{k}(\Delta)} \sum_{\mathbf{k}\in\mathcal{M}(k,l)[\Delta]} \prod_{i=1}^{m} |v_i|^{2k_i} && \text{(From (132))} \\ &= \sum_{l=1}^{k} \sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}} |\Delta| W_\Delta \sum_{\mathbf{k}\in\mathcal{M}(k,l)[\Delta]} \binom{k}{\mathbf{k}} \prod_{i=1}^{m} |v_i|^{2k_i} && \text{(From } f(\Delta)=f(\mathbf{k})) \\ &= \sum_{l=1}^{k} \frac{1+O(T^{-1})}{r^{l-1}} \sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}} \sum_{\mathbf{k}\in\mathcal{M}(k,l)[\Delta]} \binom{k}{\mathbf{k}} \prod_{i=1}^{m} |v_i|^{2k_i} && \text{(Lemma 17)} \\ &= \sum_{l=1}^{k} \frac{1}{r^{l-1}} \sum_{\mathbf{k}\in\mathcal{M}(k,l)} \binom{k}{\mathbf{k}} \prod_{i=1}^{m} |v_i|^{2k_i} + O\Big(\frac{r^k}{T}\Big), && (163) \end{aligned}$$

where in (163) we have used (147), and the bound

$$\sum_{\mathbf{k}\in\mathcal{M}(k,l)}\binom{k}{\mathbf{k}}\prod_{i=1}^{m}|v_i|^{2k_i}\leq\sum_{\mathbf{k}\in\mathcal{M}(k)}\binom{k}{\mathbf{k}}\prod_{i=1}^{m}|v_i|^{2k_i}=\Big(\sum_{i=1}^{m}|v_i|^2\Big)^k\overset{(39)}{=}r^k. \tag{164}$$

Bound of I_2. Recall that if $\Gamma\in\bar{\Gamma}\in\Delta$, we have $n(\Delta)=n(\Gamma)$.

$$\begin{aligned}
I_2&=\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}\sum_{\bar{\Gamma}\in\Delta}\sum_{\Gamma\in\bar{\Gamma}}\sum_{\substack{V\in B(\Gamma)\\ n(V)>l}}C_{V,\Gamma}V && \text{(From (131), and (138))}\\
&\leq\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}\sum_{\bar{\Gamma}\in\Delta}\sum_{\Gamma\in\bar{\Gamma}}\sum_{d=l+1}^{k}W_{d,\Delta}\sum_{\mathbf{g}\in\mathcal{M}(k,d)[\Delta]}\binom{k}{\mathbf{g}}\prod_{i=1}^{m}|v_i|^{2g_i} && \text{(Lemma 16)}\\
&\leq\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}|\Delta||\bar{\Gamma}|\sum_{d=l+1}^{k}W_{d,\Delta}\sum_{\mathbf{g}\in\mathcal{M}(k,d)[\Delta]}\binom{k}{\mathbf{g}}\prod_{i=1}^{m}|v_i|^{2g_i}\\
&\leq\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}\sum_{d=l+1}^{k}|\Delta|W_{d,\Delta}\binom{k}{\mathbf{k}(\Delta)}\sum_{\mathbf{g}\in\mathcal{M}(k,d)[\Delta]}\binom{k}{\mathbf{g}}\prod_{i=1}^{m}|v_i|^{2g_i} && \text{(From (132))}\\
&\leq\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}\sum_{d=l+1}^{k}O\Big(\frac{k^k}{r^{l-1}T^{d-l}}\Big)\binom{k}{\mathbf{k}(\Delta)}\sum_{\mathbf{g}\in\mathcal{M}(k,d)[\Delta]}\binom{k}{\mathbf{g}}\prod_{i=1}^{m}|v_i|^{2g_i} && \text{(Lemma 17)}\\
&\leq O\Big(\frac{k!k^k}{T}\Big)\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}\sum_{d=l+1}^{k}\sum_{\mathbf{g}\in\mathcal{M}(k,d)[\Delta]}\binom{k}{\mathbf{g}}\prod_{i=1}^{m}|v_i|^{2g_i} && \text{(From } \tbinom{k}{\mathbf{k}(\Delta)}\leq k!)\\
&\leq O\Big(\frac{k!k^k}{T}\Big)\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}\sum_{\mathbf{g}\in\mathcal{M}(k)}\binom{k}{\mathbf{g}}\prod_{i=1}^{m}|v_i|^{2g_i} && (\sqcup_{d=l+1}^{k}\mathcal{M}(k,d)[\Delta]\subset\mathcal{M}(k))\\
&\leq O\Big(\frac{k!k^k}{T}\Big)\sum_{l=1}^{k}\sum_{\substack{\Delta\in G(k)\\ n(\Delta)=l}}r^k && \text{(From (164))}\\
&\leq O\Big(\frac{k!k^kr^k}{T}\Big)|G(k)|\\
&=O\Big(\frac{k!k^{2k}r^k}{T}\Big), && (165)
\end{aligned}$$

where in (165) we have used that $|G(k)|\leq k^k$. This follows from the fact that each $\Delta\in G(k)$ is in one-to-one correspondance with $f(\Delta)\in\mathfrak{F}(k)$ by Lemma 15, and the number of multiplicities that can be attained is far less than k^k. Combining the terms (163) and (165) yields the claim. □

We are left to estimate the leading term in Lemma 18, which will finally show Lemma 2.

Lemma 19. *Let $\{v_i\}_{i=1}^m\subset\mathbb{C}^r$ be a frame in dimension r satisfying* (39)*. There exists a constant $c>0$ independent of k and r such that*

$$\sum_{l=1}^{k}\frac{1}{r^{l-1}}\sum_{\mathbf{k}\in\mathcal{M}(k,l)}\binom{k}{\mathbf{k}}\prod_{i=1}^{m}|v_i|^{2k_i}\leq crk^k. \tag{166}$$

Proof. Define

$$M_l(k)=\sum_{\mathbf{k}\in\mathcal{M}(k,l)}\binom{k}{\mathbf{k}}\prod_{i=1}^{m}|v_i|^{2k_i}. \tag{167}$$

In particular, from (39) and the fact that $|v_i| \leq 1$ we have that for any $k \geq 1$,

$$M_1(k) = \sum_{i=1}^{m} |v_i|^{2k} \leq r. \tag{168}$$

In (167), note that up to multiplicities, there are exactly l different vector norms appearing in each multinomial. From comparing coefficients, and since all are positive, we can readily see that the following inequality holds for all $l \leq k$,

$$M_l(k) \leq \sum_{\substack{\mathbf{b}=(b_1,\ldots,b_l) \\ \sum_{i=1}^{l} b_i = k, b_i \geq 0}} \binom{k}{\mathbf{b}} \prod_{i=1}^{l} M_1(b_i). \tag{169}$$

Using (168) in (169), and the multinomial coefficient formula

$$\sum_{\substack{\mathbf{b}=(b_1,\ldots,b_l) \\ \sum_{i=1}^{l} b_i = k, b_i \geq 0}} \binom{k}{\mathbf{b}} \leq l^k, \tag{170}$$

yields the bound

$$M_l(k) \leq r^l \sum_{\substack{\mathbf{b}=(b_1,\ldots,b_l) \\ \sum_{i=1}^{l} b_i = k, b_i > 0}} \binom{k}{\mathbf{b}} \leq r^l l^k. \tag{171}$$

Using this last bound, we have the following inequality

$$\begin{aligned} \sum_{l=1}^{k} \frac{1}{r^{l-1}} \sum_{\mathbf{k} \in \mathcal{M}(k,l)} \binom{k}{\mathbf{k}} \prod_{i=1}^{k} |v_i|^{2k_i} &\leq \sum_{l=1}^{k} \frac{1}{r^{l-1}} r^l l^k \\ &\leq r \sum_{l=1}^{k} l^k \\ &\leq crk^k, \end{aligned} \tag{172}$$

where in the last step we have used that there exists a constant $c > 0$ independent of $k \geq 1$ such that

$$\sum_{l=1}^{k} l^k \leq c \int_1^k x^k dx \leq c \frac{k^{k+1}}{k+1} \leq ck^k. \tag{173}$$

□

B Appendix on numerical experiments

Our implementation code in `Python` can be shared upon request. The following are other comments about the implementation.

- *Projection on the tangent space:* Consider $U(\delta) = \exp(\delta A)U$ in (100) and differentiate with respect to δ. If we let $UE_iU^\dagger = P_i(U)$, by evaluating the gradient at $\delta = 0$ we obtain its projection onto the tangent space of $U(n)$ at U, which is

$$P_n(U) = \sum_{i=1}^{n} \Big(\rho P_i(U) \tau P_i(U) - P_i(U) \tau P_i(U) \rho \Big) + \Big(\tau P_i(U) \rho P_i(U) - P_i(U) \rho P_i(U) \tau \Big). \tag{174}$$

  This matrix is skew-symmetric and belongs to the Lie algebra of $U(n)$.

- *Initialization of quantum states:* We initialize ρ by evaluating $W \sim \mathrm{Unif}(U(r))$, the vector $d = (d_1, \ldots, d_r)$ uniformly distributed in the simplex, and setting $\rho = W\ \mathrm{Diag}(d)\ W^\dagger$, where $\mathrm{Diag}(d)$ is the matrix with d in the diagonal and zeros elsewhere. We perform a similar, and independent initialization of τ.
- *Landscape complexity:* From trying different settings, our simulations suggests that there are several local maxima or saddle points in the optimization landscape of (15), and convergence to those is dependent on the initialization.